Frequency Decomposition to Tap the Potential of Single Domain for Generalization

Hongjing Niu*
sasori@mail.ustc.edu.cn
Qingyue Yang*
yangqingyue@mail.ustc.edu.cn
Pengfei Xia
xpengfei@mail.ustc.edu.cn
Wei Zhang
zw1996@mail.ustc.edu.cn
Bin Li
binli@ustc.edu.cn
Feng Zhao ✉
fzhao956@ustc.edu.cn

MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China, Hefei, China

Abstract

Domain generalization (DG) is essential for general artificial intelligence, enabling models to operate across unseen domains. This paper addresses the challenge of DG when limited to single-source domain training, where identifying domain-invariant features is difficult due to lack of comparative data. We propose that these invariant features are embedded in single-source domain samples and focus on extracting them. Our hypothesis suggests a close relationship between these features and frequency. We introduce a novel method that leverages multiple frequency domains. The approach involves dividing each image's frequency domain into subdomains and extracting features via a dual-branch network. This technique forces the model to learn from a narrowly defined spectrum, enhancing the detection of domain-invariant features that might be overshadowed by easily learned features. Extensive experiments show that frequency decomposition aids in learning complex features and our method surpasses existing single-source domain generalization methods.

1 Introduction

Deep learning has achieved remarkable performance in various domains [13, 14, 20, 30, 31]. However, relying on the assumption that training and testing samples come from independent and identically distributed data poses a limitation on the model's generalization capability. This assumption is often difficult to satisfy in real-world scenarios due to factors like lighting,

✉ Corresponding author, * Equal contribution.

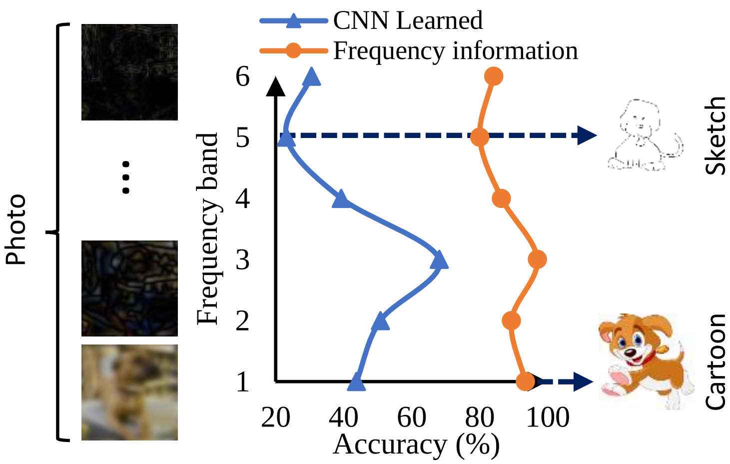


Figure 1: Accuracy of different frequency slices of photo domain image. The orange line with round dots is the accuracy of training with each frequency band and evaluating on itself. It means the classification information of each frequency band in the photo domain. The blue line with triangle dots is the accuracy of training with original images and evaluating on each frequency band. There is a large gap between the two lines which we want to fill. Different domains have different main frequency bands, e.g. cartoon images have the most information in the frequency band 1, and sketch images are concentrated in the frequency band 5. So we want to learn all frequency bands well.

background, and other unpredictable variables. Consequently, the model's performance is significantly compromised [3]. Therefore, the ability to generalize across domains is crucial for machine learning models.

Domain generalization (DG) has gained significant attention. In the multi-source DG setting, models are trained on samples from multiple source domains and evaluated on unseen target domains with different distributions. For example, in image classification tasks, datasets from different domains may vary in acquisition methods or image styles [3]. Approaches in DG aim to learn common features across source domains, such as minimizing the distance between samples of the same categories from different domains [25, 35]. However, collecting and annotating data from multiple domains can be costly, leading to the need for single domain generalization (SDG) approaches.

SDG is more challenging as it involves learning enough knowledge from a single source domain to generalize to multiple domains. The absence of comparable information to identify domain invariant features further complicates the task. Methods that find commonalities between source domains, like invariant risk minimization [1], become ineffective. Several interesting studies have explored SDG tasks. For instance, some methods utilize data augmentation, adding various transformations to simulate domain changes, to enhance model generalization [6, 23, 27, 36]. However, ensuring the effectiveness and sufficiency of these modifications remains challenging, limiting their widespread application. More analysis on the related works can be found in the appendix.

In addition to introducing new perturbations, another strategy is to mine effective information within the samples. In our method, we propose a novel approach that decomposes the image and learns all information without relying on pseudo-domain information, distinguishing it from previous methods. The underlying motivation is that domain invariant representations might be present in the images but difficult to learn. Deep learning models often struggle to capture all the features in the data that are beneficial for the task at hand. For instance, they may focus only on learning shortcuts [9], which are decision rules that perform well on standard benchmarks but fail to generalize to real-world scenarios.

We hypothesize that the subpar performance of deep learning models in SDG tasks can be attributed, in part, to their inability to fully learn all the effective features in the provided dataset. To address this, our approach decomposes the image, thereby exposing hidden information. By decomposing the image, we aim to uncover domain invariant features that may have been previously overshadowed.

We employ frequency domain decomposition to break down the image while preserving the essential features of each component. Frequency domain analysis is commonly used for image decomposition, and various frequency-based methods have been applied to DG tasks. For instance, some studies [37] have found that high-frequency information in images captures object edge structures, which tend to be consistent across different domains. Consequently, several approaches have attempted to separate frequencies into domain-related and invariant components from different perspectives and combine the related ones.

To illustrate the feasibility of frequency domain decomposition, we conducted experiments to assess the classification accuracy using different frequency bands. The results are depicted in Figure 1. On one hand, the model trained on each frequency band of images in the "Photo" domain achieved an accuracy of over 80%, indicating that each frequency band contains a significant amount of effective features. On the other hand, the model trained on the original images in the "Photo" domain did not perform well when evaluated on any frequency bands. These observations lead us to conclude that each frequency band contains valuable information in an image, but it is often challenging for the model to fully learn it. Further details and discussions regarding this experiment can be found in Section 4.

Based on the above observations, we propose a simple and effective structure to more fully learn effective features in images. And our contributions can be summarized as follows:

- We found that each frequency component contains effective information that cannot be ignored, but not all of them are learned by deep learning models.
- Based on the observation, we propose a new single source domain generalization method to better learn the effective features contained in each frequency band.

2 Related Work

Domain generalization [29, 33, 43] aims to learn a model on source domains that can perform well on unseen domains. Many methods are proposed to solve domain generalization tasks. Some approaches [10, 17] align source domain distributions by minimizing moments of transformed features between source domains, to learn domain-invariant representation. Li et al. [22] first applied meta-learning to DG and many people [2, 42] followed this work. In these methods, source domains are divided into non-overlapping meta-source and meta-target domains to simulate domain shift. Many methods use data augmentation, such as image transformation[32], random augmentation[38], and feature-based augmentation[44] to solve DG tasks. While Volpi et al. [33] uses adversarial gradients obtained from the classifier to perturb the input images, so the coverage of the training domain is expanded.

Single domain generalization is a more challenging domain generalization task. The task setting is similar to normal DG (multi-source DG) except that only one source domain is available in training the model. So it is no longer possible to obtain domain invariant representation by finding the commonality between source domains. To solve this difficult task, former proposed works put the emphasis on meta-learning and introducing disturbance. Wang et al. [34] decomposes image features into meta features, so as to encode an image without domain information. Qiao and Peng [27] expose the model to domain shift during training via meta-learning, and the synthesis of out-of-domain data is guided by uncertainty assessment. Introducing disturbance, such as adversarial training and expanding the coverage of the training domain, aims to generate out-of-domain data. By training with these samples, which simulate the changed domains, the model can learn domain invariant features and enhance the generalization ability. Qiao et al. [28] and Fan et al. [8] use adversarial training to create fictitious but challenging groups, from which the model can be learned and promoted under the theoretical guarantee.

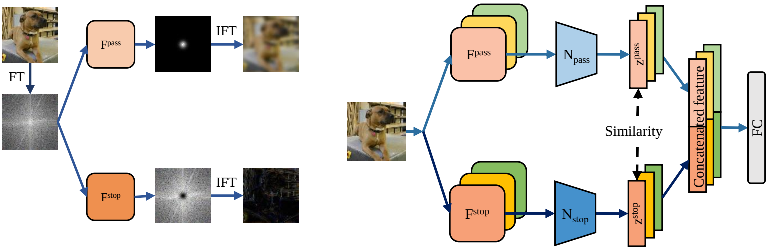
Research based on data augmentation adds noise to an image or its feature to let the model learn domain invariant features. Cugu et al. [6] uses multiple visual corruptions to alter the training images to simulate new domains. Li et al. [23] and Wang et al. [36] generate the extended domains by using comparative learning or mutual information to guarantee the safety and effectiveness of the extended domain. The implicit assumption of these methods is that the introduced invariance is effective for domain generalization. Unlike those data augmentation methods, our approach adds nothing to the image but decomposes samples into slices. By fully learning them, the model can learn domain invariant features. **Frequency domain** options are incorporated into deep learning methods for enhancing the robustness and generalization capability of the model. Many works focused on image frequency processing. Jeon et al. [16] changes the low frequency component of the image to affect the image style and keeps the high frequency component unchanged to maintain the shape information. By using this data augmentation, the model aims to learn domain invariant features. Chen et al. [5] supposes that humans recognize images more through phase information. So they propose a method to keep the phase of the image unchanged and add disturbance to the amplitude to increase the robustness of CNN (Convolutional Neural Networks). Guo et al. [11] uses low-frequency perturbations on the image for adversarial attacks.

On the other hand, some methods convert features to frequency domain for processing. Lin et al. [24] transforms the feature maps of different network layers into the frequency domain. Then a mask is generated to enhance the domain invariant frequency components and suppress the components that are not conducive to generalization. Guo and Ouyang [12] learn an effective frequency range for the features of each convolution layer to improve the convergence and robustness of CNN. Our method decomposes images from the frequency domain and aims at learning them all rather than supposing that there are frequency components with good generalization.

3 Method

3.1 Problem Definition

The SDG task encompasses various application scenarios and challenges. During training, only data from a single source domain is available. However, in the testing phase, the model is evaluated on multiple target domains that were not seen during training.



(a) FT & IFT on images (b) The proposed framework

Figure 2: The overall framework of the proposed method. (a) The frequency slices of the image. The image is divided into two slices in the frequency domain by a pair of complementary filters. (b) the overall framework. The image is processed by a pair of filters for each frequency band, then input into two neural networks. The similarity between the two features extracted from the image is calculated.

The SDG task involves diverse domains that share a common category space for image classification. However, their styles and acquisition methods differ. To address this, extracting common features from the source domain becomes crucial. With only one source domain available, it is challenging to determine the best-performing features for generalization. Thus, our proposed method aims to extract as many effective features as possible from different frequency components to enhance single source domain generalization.

3.2 Frequency Slice of Image

As shown in Figure 2, we apply Fourier transformation (FT) to the three RGB channels of the image $\boldsymbol{x}$ to transform the image from the spatial domain to the frequency domain. Then we use a Gaussian bandpass filter F^{pass} to take out part of the frequency band of the image. Meanwhile, there is a complementary bandstop filter F^{stop}, which stop frequency is consistent with the pass frequency of the bandpass filter. For this pair of bandpass and bandstop filters,

$$H^{pass}(\boldsymbol{x}) + H^{stop}(\boldsymbol{x}) = \mathbf{1}, \tag{1}$$

where $H^{pass}(\cdot)$ denotes the frequency response of $F^{pass}(\cdot)$ and $H^{stop}(\cdot)$ denotes the frequency response of $F^{stop}(\cdot)$. Finally, we use inverse Fourier transformation (IFT) to transform the frequency components back to the spatial domain:

$$\boldsymbol{x}_i^{pass} = IFT(FT(\boldsymbol{x}) \circ F_i^{pass}) \tag{2}$$

$$\boldsymbol{x}_i^{stop} = IFT(FT(\boldsymbol{x}) \circ F_i^{stop}) \tag{3}$$

where $\boldsymbol{x}_i^{pass}$ and $\boldsymbol{x}_i^{stop}$ denote the frequency slice of the image, $\circ$ denotes Hadamard product.

The two filters decompose the image into two complementary parts about the pass frequency band, which contain specific image information about this frequency component. The pair of filters helps the model extract features from specific frequency bands.

3.3 Overall Framework

Our training process is shown in Figure 2. Firstly, we apply FT to the input image to get the spectrogram. Then the model has two branches, the upper **pass branch** and the lower **stop**

branch, which means the filters in each branch are bandpass filters and bandstop filters.

In the pass branch, K Gaussian bandpass filters $F^{pass}_{i=1\sim K}$ decompose the image into K pieces in the frequency domain. The pass frequencies of $F^{pass}_{i=1\sim K}$ satisfy:

$$\sum_{i=1}^{K} H_i^{pass}(\boldsymbol{x}) = \mathbf{1} \tag{4}$$

where $H_i^{pass}(\cdot)$ denotes the frequency response of F_i^{pass}. Then the frequency pieces are transformed back to the spatial domain and become frequency slices $\boldsymbol{x}^{pass}_{i=1\sim K}$ as Section 3.2. After that, the frequency slices $\boldsymbol{x}_i^{pass}$ are send to a neural network $N_{pass}(\cdot)$ to extract features $\boldsymbol{z}_i^{pass}$. The frequency slices only have the information of their own frequency bands and let the network learn information of every frequency band precisely. In this way, the network can fully learn the information of the image, including the domain invariant information.

In the stop branch, a series of Gaussian bandstop filters $F^{stop}_{i=1\sim K}$ are used. Their stop frequency is consistent with the pass frequency of the pass branch as the Equation 1 illustrated. The stop frequencies of F_i^{stop} satisfy:

$$\frac{1}{K-1}\sum_{i=1}^{K} H_i^{stop}(\boldsymbol{x}) = \mathbf{1}, K \geq 2, \tag{5}$$

where $H_i^{stop}(\cdot)$ denotes the frequency response of F_i^{stop}. After the filters F_i^{stop}, the frequency slices $\boldsymbol{x}_i^{stop}$ are send to a neural network $N_{stop}(\cdot)$ as the pass branch and become features $\boldsymbol{z}_i^{stop}$. Except for the frequency slices, the original image is also sent to the two neural networks to provide information from the whole frequency band aspect. Since the two branches aim to extract effective features of the same object from different frequency bands, the features should be similar to some extent. So the consistency loss $L_{cons}(\boldsymbol{z}_i^{pass}, \boldsymbol{z}_i^{stop})$ is calculated for the features. We use cosine similarity loss as the L_{cons} so that the two branches can extract features which exist in a pair of non-overlapping frequency bands.

Finally, each pair of features $\boldsymbol{z}_i^{pass}$ and $\boldsymbol{z}_i^{stop}$ are concatenated together and input to a fully connected layer $FC(\cdot)$ to obtain the classification results $\widehat{y}_i$. Between the true label y and every predicted label $\widehat{y}_i$, classify loss $L_{cls}(y, \widehat{y}_i)$ is calculated to ensure the correct classification. We use cross-entropy loss as the L_{cls}. The whole loss function is

$$L = \sum_{i=1}^{K} L_{cls}(y, \widehat{y}_i) + \alpha L_{cons}(\boldsymbol{z}_i^{pass}, \boldsymbol{z}_i^{stop}) \tag{6}$$

where α is the trade-off hyper-parameter.

In test phase, the input image is directly input to the two networks N_{pass} and N_{stop} without filters. Then features are combined and passed to the liner layer to get the prediction.

4 Experiment

4.1 Datasets and implementation details.

Datasets. To evaluate the proposed method, we conduct experiments over a widely used SDG benchmark PACS[21]. Since the styles vary greatly between domains, it is a challenging benchmark. PACS contains four domains: **P**hoto, **A**rt painting, **C**artoon, and **S**ketch. It has a total of 7 categories and 9991 images. In the task of single-domain generalization, the common datasets are PACS and VLCS. And PACS is the hardest and the most widely used dataset among them. The domain shift (difference) between PACS domains is the largest, and the frequency with different performances is more likely to appear after frequency decomposition. For VLCS, all domains of it are photo, and the importance of frequency bands between domains is close, so the significance of frequency division is not great. So we chose PACS as the benchmark dataset.

Implementation details. In comparison with other SDG methods, we follow their settings. We use the images in the Photo domain as the training and validation set, and the model is tested on all the samples of each other domains: art painting (A), cartoon (C), and sketch (S). The average accuracy is also calculated for comparison. Fast Fourier transform is utilized to transform the image into the frequency domain. Six Gaussian bandpass filters are used to decompose the image into six frequency slices. Their center frequencies and bandwidths are (0,6), (7,8), (20,20), (40,20), (60,20), and (92,44), respectively. This division is an empirical value to ensure that each frequency component contains certain effective information. At the same time, six Gaussian bandstop filters are used in the stop branch, and their center frequency and bandwidth are consistent with those of the bandpass filters. During training, two corresponding frequency slices will be input to different branches of the framework for training at the same time. In addition to the frequency slices, the original image is input to both branches for learning the context information between frequency bands. We use ResNet18 [13] as the backbone, and the pre-trained parameters on ImageNet1K [7] provided by PyTorch [26] are used. The shape of each branch output feature is 512. And the final used linear layer size is input 1024 and output 7. We use Adam [19] optimizer. The initial learning rate is 0.0001 and adjusted by the cosine annealing algorithm. The weight decay is 0 by default. Unless otherwise stated, the epoch is set to 100, and the batch size is set to 32. We tested the influence of different values of the hyper-parameter α on the generalization performance and took the $\alpha = 5$ corresponding to the highest precision as the final hyper-parameter. All experiments are implemented with PyTorch and run on an NVIDIA Tesla V100 GPU.

4.2 Inadequate Learning of Effective Frequency Slices

Frequency slices in Photo domain. We assume that for a deep learning model trained under an experienced risk minimization (ERM) strategy, some frequency components of images that may have a positive effect on image classification may not be learned adequately.

To verify this idea, we first designed an experiment to observe whether each frequency component of the image is effective and whether it has been learned. And the results can be found in Figure 3. Specifically, we split the photo images into six frequency slices (F1$\sim$F6) according to Equation 2, where F1 represents the lowest frequency, and F6 is the highest frequency band. We have designed three training settings. Firstly, we trained a DNN on the original photo images and then tested it on the frequency slices to explore whether the model could learn each frequency component. Higher accuracy on each frequency slice indicates more adequately the frequency component is learned. We also tested the model with original photo images as a contrast. Then we trained and tested the model on each separated frequency slice to explore whether each frequency component was effective for the classification task. Higher accuracy on each frequency slice indicates the more effective the frequency component is. Finally, the proposed method was used to verify whether it is helpful for learning various frequency components.

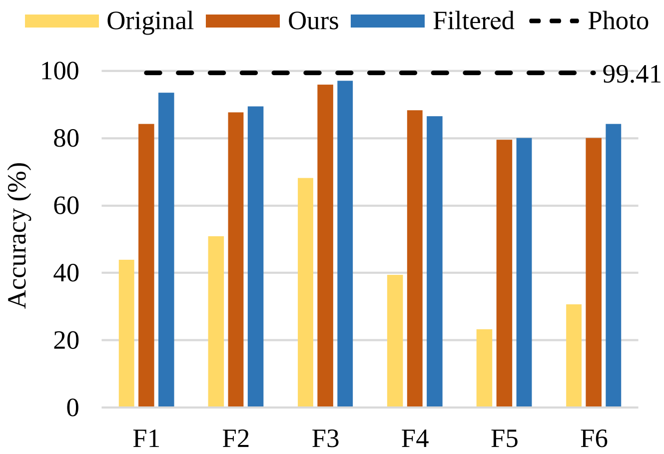


Figure 3: Accuracy (%) histogram of 3 different training settings. The testing is conducted on the frequency slices. The yellow bars represent training on the original photo samples, the red bars represent our method, and the blue bars represent training separately on each frequency component. F1 $\sim$ F6 means testing on each frequency component with F1 lowest and F6 highest. The dotted line is trained and tested on original photo images.

Based on Figure 3 there are some observations.

- Compared with the testing accuracy on the original images (dotted line), models trained on original images can hardly achieve comparable accuracy on separated frequency slices (yellow bars). It reveals that even if the model has achieved high classification accuracy, it did not learn all frequency components adequately. And it is more obvious for the high-frequency. The possible reasons are CNN's preference to learn low-frequency components first [39], and most of the energy in images are concentrated in low-frequency [40].

- The accuracies of training with each separated frequency slice (blue bar) are much higher and very close to the results on the original images. It is a strong evidence supporting each slice contains a considerable number of effective information.

- Model trained with our approach (red bar) achieves a performance close to training with separated frequency slices (blue bar) for each frequency band. By comparing the three groups of bars, it can be concluded that our method significantly improves the model's learning of effective information in all the frequency bands.

Different domains do not share similar main frequency bands. According to the last experiment, we found that for a single domain, all the frequency slices contain effective information for the classification task, but they are not fully learned. Furthermore, we propose two extension problems, which will affect the design idea of the method. (1) Is this a special case of some domains? (2) For different domains, whether the distributions of effective information in the frequency bands are similar.

We extend the experiment to all four domains of the PACS dataset, and the experimental results are shown in Table 1. It is trained and tested with each separated frequency band. Higher accuracy indicates more effective information for classification is contained in the frequency band.

Table 1: Accuracy (%) of different domain frequency slices. They are all trained with a specific frequency band of a domain and tested with the samples in the same distribution of themselves. The best result of each domain is in **bold faces**. In comparison, the accuracy of random guess is 14.29%.

Domain	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**
Photo	93.57	89.47	**97.08**	86.55	80.12	84.21
Art painting	77.2	76.17	**83.42**	74.61	63.73	63.73
Cartoon	**96.61**	90.68	93.64	91.53	82.63	84.75
Sketch	90.7	92.96	**95.98**	94.72	93.22	94.97

According to the experimental results, the answers to the previous two questions can be given. (1) It is not a special case that the effective information of each frequency band has not been fully learned. It exists widely in all the tested domains. (2) Different domains do not share similar main frequency bands. For example, Cartoon has the most effective information in the lowest frequency band because the image in this domain is represented by a lot of colors which are related to low-frequency. Meanwhile, the main frequency band of Sketch has a higher frequency because lines are more high-frequency than gradient colors. Therefore, it is important to fully learn the effective information of each frequency band for domain generalization tasks. Since the effective information distribution of each frequency band is different for domains, it is more reasonable to learn on different frequency slices than simply assign weights to frequency bands.

4.3 Evaluation of Single Domain Generalization

Table 2: SDG accuracy (%) on PACS. One domain is used to train the model, and other domains are used for testing. And the accuracy is the average of the accuracy of the three domains. The best results are in **bold faces**, and the second best results are <u>underlined</u>.

Method	**P**	**A**	**C**	**S**	**Avg.**
ERM [33]	42.2	70.9	76.5	53.10	60.7
MixStyle [44]	41.2	61.9	71.5	32.2	51.7
EFDMix [41]	42.5	63.2	73.9	38.1	54.4
RSC [15]	41.63	70.67	75.08	47.25	58.66
SelfReg [18]	43.46	72.59	76.56	45.76	59.59
L2D [36]	52.29	<u>76.91</u>	<u>77.88</u>	53.66	65.18
ASR [8]	<u>54.6</u>	76.7	**79.3**	**61.6**	<u>68.1</u>
ours	**64.52**	**79.91**	77.63	<u>57.68</u>	**68.94**

PACS is a widely used dataset for SDG tasks. We follow the most common strategy, leave-one-out, to test the SDG performance of the proposed approach. Specifically, each domain is used as the source domain, in turn. Meanwhile, other domains are used as the target domain. Then the average result of a group of experiments is calculated to measure the SDG capability. The results are shown in Table 4.3.

Furthermore, we show the accuracy tested on other domains when **P**hoto domain, as the most common domain, is used as the source domain. And the results are shown in Table 4.3.

Single domain generalization on PACS. We trained our method with each domain in the PACS dataset. It can be seen that our approaches generally outperform other methods. An interesting observation is that compared with the previous method, our method performs particularly well on P, while the improvement on S is relatively small. A possible reason is that there may be some similarity between the effective information of frequency slices. For the Photo domain, the difference between frequency slices is larger. Thus, our approaches can learn more effective information. In contrast, the difference between frequency slices is smaller for the Sketch domain.

Performance on different target domains. Table 4.3 shows the evaluation of P→ACS. Our method achieved state-of-the-art results on this challenging benchmark. The result shows that our method can greatly enhance the generalization capability of the model in testing, which is due to better learning of the effective features of each frequency.

Table 3: SDG accuracy (%) on PACS. Models are trained on photo and test on other domains. The best results are in **bold faces** and the second best results are <u>underlined</u>.

Method	**A**	**C**	**S**	**Avg.**
ERM [33]	54.43	42.74	42.02	46.39
JiGen [4]	54.98	42.62	40.62	46.07
RSC [15]	56.26	39.59	47.13	47.66
ADA [33]	58.72	45.58	48.26	50.85
M-ADA [28]	<u>58.96</u>	44.09	49.96	51.00
L2D [36]	56.26	51.04	58.42	55.24
MetaCNN [34]	54.05	**53.58**	<u>63.88</u>	<u>57.17</u>
ours	**66.41**	<u>53.07</u>	**74.10**	**64.52**

The improvement in Sketch is particularly obvious, which is over 10% higher than MetaCNN, the last state-of-the-art method. The reason might be that sketches are made of lines without color, so most of the effective information is contained in the high-frequency component, which is easily ignored in the training process of general models. Our method has solved this problem well and achieved good results.

5 Conclusion

In this paper, we first share an observation that each frequency component contains effective information that cannot be ignored, but not all of them are learned by deep learning models. Based on the observation, we proposed a new domain generalization method, which can better learn the effective features contained in each frequency band. Through frequency decomposing, our approach can better learn efficient features from all the frequency bands. Sufficient experiments from multiple angles indicate that our approach outperforms the state-of-the-art single domain generalization methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants U19B2044 and the Anhui Provincial Natural Science Foundation under Grant 2108085UD12. We acknowledge the support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC.

References

[1] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019.

[2] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. *Advances in Neural Information Processing Systems*, 31:1–11, 2018.

[3] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79(1):151–175, 2010.

[4] Fabio M Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2229–2238, 2019.

[5] Guangyao Chen, Peixi Peng, Li Ma, Jia Li, Lin Du, and Yonghong Tian. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 458–467, 2021.

[6] Ilke Cugu, Massimiliano Mancini, Yanbei Chen, and Zeynep Akata. Attention consistency on visual corruptions for single-source domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4165–4174, 2022.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[8] Xinjie Fan, Qifei Wang, Junjie Ke, Feng Yang, Boqing Gong, and Mingyuan Zhou. Adversarially adaptive normalization for single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8208–8217, 2021.

[9] Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, 2020.

[10] Muhammad Ghifary, David Balduzzi, W Bastiaan Kleijn, and Mengjie Zhang. Scatter component analysis: A unified framework for domain adaptation and domain generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(7): 1414–1430, 2016.

[11] Chuan Guo, Jared S Frank, and Kilian Q Weinberger. Low frequency adversarial perturbation. *arXiv preprint arXiv:1809.08758*, 2018.

[12] Weiyu Guo and Yidong Ouyang. Robust learning with frequency domain regularization. *arXiv preprint arXiv:2007.03244*, 2020.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[14] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.

[15] Zeyi Huang, Haohan Wang, Eric P Xing, and Dong Huang. Self-challenging improves cross-domain generalization. In *European Conference on Computer Vision*, pages 124–140. Springer, 2020.

[16] Seogkyu Jeon, Kibeom Hong, Pilhyeon Lee, Jewook Lee, and Hyeran Byun. Feature stylization and domain-aware contrastive learning for domain generalization. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 22–31, 2021.

[17] Xin Jin, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. Feature alignment and restoration for domain generalization and adaptation. *arXiv preprint arXiv:2006.12009*, 2020.

[18] Daehee Kim, Youngjun Yoo, Seunghyun Park, Jinkyu Kim, and Jaekoo Lee. Selfreg: Self-supervised contrastive regularization for domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9619–9628, 2021.

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[21] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5542–5550, 2017.

[22] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pages 1–8, 2018.

[23] Lei Li, Ke Gao, Juan Cao, Ziyao Huang, Yepeng Weng, Xiaoyue Mi, Zhengze Yu, Xiaoya Li, and Boyang Xia. Progressive domain expansion network for single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 224–233, 2021.

[24] Shiqi Lin, Zhizheng Zhang, Zhipeng Huang, Yan Lu, Cuiling Lan, Peng Chu, Quanzeng You, Jiang Wang, Zicheng Liu, Amey Parulkar, et al. Deep frequency filtering for domain generalization. *arXiv preprint arXiv:2203.12198*, 2022.

[25] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5715–5725, 2017.

[26] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32:1–12, 2019.

[27] Fengchun Qiao and Xi Peng. Uncertainty-guided model generalization to unseen domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6790–6800, 2021.

[28] Fengchun Qiao, Long Zhao, and Xi Peng. Learning to learn single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12556–12565, 2020.

[29] Zheyan Shen, Jiashuo Liu, Yue He, Xingxuan Zhang, Renzhe Xu, Han Yu, and Peng Cui. Towards out-of-distribution generalization: A survey. *arXiv preprint arXiv:2108.13624*, 2021.

[30] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[31] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[32] Riccardo Volpi and Vittorio Murino. Addressing model vulnerability to distributional shifts over image transformation sets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7980–7989, 2019.

[33] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. *Advances in Neural Information Processing Systems*, 31:1–11, 2018.

[34] Chaoqun Wan, Xu Shen, Yonggang Zhang, Zhiheng Yin, Xinmei Tian, Feng Gao, Jianqiang Huang, and Xian-Sheng Hua. Meta convolutional neural networks for single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4682–4691, 2022.

[35] Jindong Wang, Wenjie Feng, Yiqiang Chen, Han Yu, Meiyu Huang, and Philip S Yu. Visual domain adaptation with manifold embedded distribution alignment. In *Proceedings of the 26th ACM International Conference on Multimedia*, pages 402–410, 2018.

[36] Zijian Wang, Yadan Luo, Ruihong Qiu, Zi Huang, and Mahsa Baktashmotlagh. Learning to diversify for single domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 834–843, 2021.

[37] Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14383–14392, 2021.

[38] Zhenlin Xu, Deyi Liu, Junlin Yang, Colin Raffel, and Marc Niethammer. Robust and generalizable visual representation learning via random convolutions. *arXiv preprint arXiv:2007.13003*, 2020.

[39] Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*, 2019.

[40] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020.

[41] Yabin Zhang, Minghan Li, Ruihuang Li, Kui Jia, and Lei Zhang. Exact feature distribution matching for arbitrary style transfer and domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8035–8045, 2022.

[42] Yuyang Zhao, Zhun Zhong, Fengxiang Yang, Zhiming Luo, Yaojin Lin, Shaozi Li, and Nicu Sebe. Learning to generalize unseen domains via memory-based multi-source meta-learning for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6277–6286, 2021.

[43] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization in vision: A survey. *arXiv preprint arXiv:2103.02503*, 2021.

[44] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. *arXiv preprint arXiv:2104.02008*, 2021.