# AmpNorm: An Effective Style Normalization Method for Single Domain Generalization

Jingyu Hu [1], Tao Zhong [1], Xilin He [1], Weicheng Xie [1*], Siyang Song [2], Linlin Shen [1]

[1] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
[2] University of Leicester, Leicester LE1 7RH, UK
[*] wcxie@szu.edu.cn

**Abstract: Deep neural networks have demonstrated remarkable efficacy in numerous computer vision tasks. However, due to the training and testing sets of data coming from different domains, the domain gap limits the performances of deep neural networks. To enhance the generalization performance of deep neural networks, we provide a unique amplitude normalization (AmpNorm) approach to decrease the domain gap from the frequency domain perspective. Specifically, ours AmpNorm collects the amplitude spectrum from the source domain during training and then converts the images from the unseen target domain into those with a style similar to the source domain during testing. Our AmpNorm is simple yet effective, as well as plug-and-play, which is readily implemented into the majority of single-domain generalization (SDG) methods. Extensive results on three public benchmarks demonstrate that our AmpNorm can greatly improve these models' performance on the invisible target domain.**

**Keywords: Frequency domain, Style normalization, Domain Generalization**

## 1. Introduction

Deep neural networks have demonstrated remarkable efficacy in numerous computer vision tasks [6]. Conventional deep neural networks are often trained based on the assumption that training and testing data are identically and independently distributed (i.e., they are independent and identically distributed (i.i.d.)). The assumption, however, is not always valid in practical implementations because of the scenarios' intricate data collecting limitations. When the testing data are collected from out-of-distribution (OOD) domains, the trained models' performance will drastically decline. [2]. Therefore, the domain generalization (DG) challenge [27], which tries to train a model with good generalization on unseen target domains, has attracted increasing attention from academic and industry fields.

Previous methods [3,12,16] were widely suggested to solve the domain generalization challenge based on multi-source domains, where the domain gap can be reduced by simply aggregating the data from multiple domains [13]. However, given the limitations of the scenarios' intricate data collecting, this method could not work well in real-world applications. Therefore, SDG methods have been widely explored [4,23,26], which need only one domain of data for training. Previous solutions for the challenging SDG task are mainly based on data manipulation [4,7], which aims to generate more diverse samples to expand the training sample space. However, because there is little variation in the styles specific to the samples from a single source domain, the produced samples are not sufficiently varied, which restricts the performances of models.

To eliminate the style diversity of the samples, it is revealed that the amplitude of the frequency domain can approximately represent the style of an image [19,20,22]. We conducted a toy experiment to study the relationship between the image amplitude and its style in Fig. 1. It shows that the

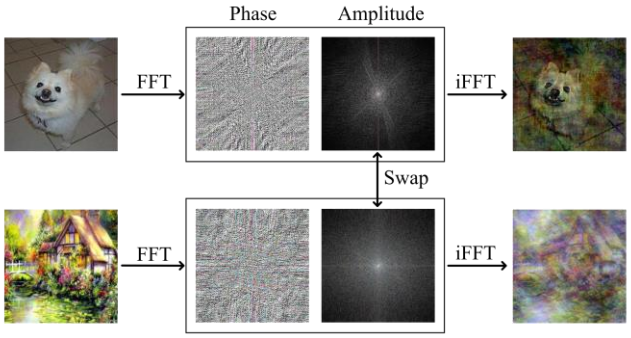



***Fig. 1.*** *Amplitude and phase of the frequency domain can approximate image style and content, respectively. If we swap the amplitudes of two images, their style will be swapped while their content will remain unchanged.*

content of the amplitude-exchanged sample is similar to the provided phase spectrum of the original sample, and the style of the amplitude-exchanged sample is similar to the provided amplitude spectrum. Motivated by the aforementioned observation, we suggest an innovative amplitude normalization (AmpNorm) approach to reduce the domain gap from the perspective of the frequency domain. Specifically, the images of the unseen target domain are converted into images with a style comparable to the source domain during testing, depending on the amplitude spectrum collected from the source domain during training. This paper's primary contributions are summed up as follows:

- We suggest an innovative normalization (AmpNorm), which is one of the pioneering works of converting the target domain style to the source domain style in the frequency domain.
- By using the collected amplitude spectrum of the source domain to replace that of the train and test samples, our AmpNorm can well reduce the domain gap to enhance the trained model's generalization performance.

- Our AmpNorm's superiority is confirmed using three widely used benchmarks, and the results demonstrate that it can significantly enhance the model's performance on the previously undiscovered target domain.

## 2. Related Work

### 2.1. Domain Generalization

Domain Generalization (DG) intends to improve the performance of deep neural networks from the source domain to the unseen target domain. In terms of the number of source domains, DG can be categorized into multi-source and single-source. SDG task is more difficult since it is hard to find consistent patterns across different source domains for the unseen domains [1]. Data manipulation methods [4,14] are used to deal with the task of SDG by manipulating the input samples to help deep neural networks learn general representations. In this paper, we suggest an innovative data manipulation approach, which converts the images of the unseen target domain into those with a style similar to the source domain.

### 2.2. Style-based Method for DG

Previous researches [9,18,19,20,22,29,30] have shown that the style and content cues of the image are particular to the amplitude spectrum and the phase spectrum of the frequency domain, respectively. Therefore, many researchers [4,8,29] suggested introducing noise into the amplitude spectrum of the frequency domain to generate more varied samples. [29] resorted to the amplitude mix (AM) and amplitude swap (AS) approaches. [4] proposed phase scaling, constant amplitude, and high pass filter to generate augmentation samples for source domain training. However, the generated samples lack sufficient diversity due to the limited number and style of the single-source domain samples. In this paper, we suggest an innovative amplitude normalization (AmpNorm) approach to reduce the domain gap from the perspective of the frequency domain, i.e. transferring the amplitude spectrum's style from the source domain to the unseen target domain.

## 3. Methodology

### 3.1. Preliminary

**Fourier Transform** Given an image $x \in \mathbb{R}^{H\times W}$ with a single channel, the Fourier transform of $\mathcal{F}(x)$ at the coordinate of $(u, v)$ for converting $x$ from the spatial domain to the frequency domain is formulated as:

$$x_z(u,v) = \mathcal{F}(x)(u,v) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} x(h,w)e^{-i2\pi\left(\frac{h}{H}u+\frac{w}{W}v\right)} \quad (1)$$

where $i$ is imaginary unit and the height and width of the image are denoted by $H, W$, respectively. The amplitude and phase spectrum are formulated as:

$$\begin{aligned}\mathcal{A}(x)(u,v) &= \sqrt{\mathcal{R}^2(x)(u,v) + \mathcal{I}^2(x)(u,v)},\\ \mathcal{P}(x)(u,v) &= \arctan\left(\frac{\mathcal{I}(x)(u,v)}{\mathcal{R}(x)(u,v)}\right)\end{aligned} \quad (2)$$

where $\mathcal{I}(x)$ and $\mathcal{R}(x)$ represent imaginary the and real parts of $\mathcal{F}(x)$, respectively. Moreover, $\mathcal{F}(x)$ can put back together using $\mathcal{A}(x)$ and $\mathcal{P}(x)$ as:

$$\begin{aligned}\mathcal{F}(x)(u,v) = \mathcal{A}(x)(u,v)\cos\big(\mathcal{P}(x)(u,v)\big)\\ +i\mathcal{A}(x)(u,v)\sin\big(\mathcal{P}(x)(u,v)\big)\end{aligned} \quad (3)$$

For RGB images, $\mathcal{A}(x)$ and $\mathcal{P}(x)$ are calculated in the channel-wise manner.

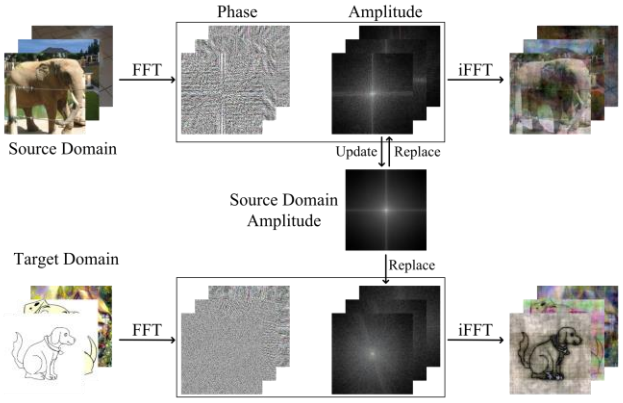



***Fig. 2.*** *Overview of our amplitude normalization. We update the source domain amplitude by the EMA of the source domain from the amplitude of the frequency domain samples to replace the amplitude of source and target domain samples and then generate samples with a similar style to the source domain.*

### 3.2. Amplitude Normalization

In this section, we introduce our amplitude normalization (AmpNorm) method that converts samples from any unseen domain into samples with similar styles to the source domain. Fig. 2 shows the overview of our AmpNorm.

Since the amplitude and phase of the image in the frequency domain can surrogate the style and content of the image, respectively, we customize the style of the image by regularizing the amplitude of this image. We extract the amplitude, i.e. $x_a$ and the phase, i.e. $x_p$ of the frequency domain samples $x_z$ as:

$$\begin{cases} x_a = \mathcal{A}(x_z), \\ x_p = \mathcal{P}(x_z). \end{cases} \quad (4)$$

We collect the amplitude spectrum of the source domain during training to represent the style of the source domain in order to transferring the image style from the source domain to the unseen target domain. Specifically, the

domain consistent samples $\bar{x}$ encoded the source domain style can be obtained as:

$$\bar{x}_z = s_a^S \cos(x_p) + x_a^S \sin(x_p), \quad \bar{x} = \mathcal{F}^{-1}(\bar{x}_z). \tag{5}$$

where $\mathcal{F}^{-1}(\cdot)$ denotes the inverse Fourier transform, $x_a^S$ is the source domain amplitude, which is updated iteratively by the exponential moving average (EMA) of the source domain amplitude as:

$$x_{a^{(k)}}^S = (1-\lambda) x_{a^{(k-1)}}^S + \lambda x_{a^{(k)}} \tag{6}$$

where $x_{a^{(k)}}^S$ is the source domain amplitude $x_a^S$ in the $k$-th iteration, and $x_{a^{(k)}}$ is the amplitude specific to the mini-batch samples in the $k$-th iteration. Finally, we feed the domain consistent samples $\bar{x}$ into network $f$ as:

$$\bar{y} = f(\bar{x}, \theta) \tag{7}$$

where $\theta$ records the parameters of $f$. The classification accuracy is evaluated using the cross-entropy loss as:

$$\ell = \mathcal{L}(\bar{y}, y) \tag{8}$$

where $y$ is true label of the sample $x$. For clarity, Pseudo-code of our AmpNorm is shown in Algorithm 1.

**Algorithm 1** Pseudocode of AmpNorm, PyTorch-like

```
# f: the backbone network
# domain_amp: the source domain's amplitude
# λ: the momentum of the exponential moving average

for x, y in loader:  # load a batch of data
    x_r = AmpNorm(x)
    y_r = f(x_r)  # (N, K) # Eq. (7)
    L = CrossEntropyLoss(y_r, y)  # Eq. (8)
    L.backward()  # back-propagate
    update(f)  # update network parameters

def AmpNorm(x):  # Amplitude Normalization
    x_z = fft(x)  # Fast Fourier transform (Eq. (1))
    x_a, x_p = decompose(x_z)  # calculate amplitude and phase (Eq. (2))
    if training:  # update source domain amplitude only if in training stage
        domain_amp = (1 - λ) * domain_amp + λ * mean(x_a, dim=0) # Eq. (6)
    x_rz = compose(domain_amp, x_p)  # replace amplitude # Eq. (3)
    x_r = ifft(x_rz)  # Inverse Fast Fourier transform
    return x_r
```

## 4. Experiments

We first introduce the datasets and experimental settings in this section. Secondly, we compare the suggested method with the state-of-the-art (SOTA) approaches on three benchmarks. Lastly, we conduct ablation studies and visualization experiments to analyze the proposed method.

### *4.1. Datasets and Experimental Settings*

Three benchmarks are used to evaluate ours AmpNorm, i.e. PACS [12], Office-Home [25] and DomainNet [21]. All the three benchmarks are used in object recognition task for evaluating the domain generalization performance.

**PACS** [12] contains four domains, i.e. Photo, Art painting, Sketch and Cartoon. PACS contains7 categories with 9,991 images.

**Office-Home** [25] contains four domains, i.e. Art, Product, Real-World and Clipart. It contains 65 categories with 15,500 images.

**DomainNet** [21] contains six domains, i.e. Quickdraw, Painting, Infograph, Real, Clipart and Sketch. It contains 345 categories with 0.6 million images.

ResNet18 [6] serves as the backbone network pre-trained on ImageNet [5], and empirical risk minimization (ERM) [24] is used as the baseline. Adam [10] optimizer with momentum of 0.9 and weight decay of 5e-4 are used. The initial learning rate is set to 5e-4 and the cosine annealing learning rate scheduler [17] is used for 120 epochs. The size of the input image is $224 \times 224$ and the batch size is 64. For the AmpNorm module, the momentum of the exponential moving average $\lambda$ of 0.1 (Eq. (6)), and the amplitude (Eq. (5)) of the source domain is initialized to 1.

### *4.2. Comparison with SOTA methods*

The test accuracies of AmpNorm and SOTA methods for SDG on PACS dataset are reported in Table 1. It shows that our AmpNorm improves the performance of recent single-domain methods in different proportions. For example, AmpNorm improves the performance of ERM by 5.51% and that of DAC by 1.97%. Notably, our AmpNorm can largely improve the performance of the model, while using Photo and Sketch as the source domains. This may be because that the Photo domain contains more rich style information and the Sketch domain have a highly similar style, AmpNorm can sufficiently leverage these style cues. For Cartoon and Art painting domains, due to the artistic processing of the image, the style information is more complex and difficult to recognize, the performance of AmpNorm is not as good as the other two domains.

Table 2 shows the results of Office-Home dataset. It clearly shows that our AmpNorm can improve the performance of the model after integrating AmpNorm into the baseline and other SOTA domain generalization methods. For example, AmpNorm improves the performance of ERM by 2.30% and that of DAC by 1.59%.

The results of DomainNet dataset are reported in Table 3. We can see that our AmpNorm can facilitate the

**Table 1** SDG classification accuracies (%) with ResNet18 as the backbone on the PACS dataset, where each SOTA is used as the baseline. We use a source domain for model training and the remaining three domains for model evaluation. The strong labels denote the higher performance levels. Red italics indicate the best performance.

| Method | Photo | Art Painting | Cartoon | Sketch | Average |
|---|---|---|---|---|---|
| EMR [24] | 39.54 | 65.92 | 60.96 | 40.58 | 51.75 |
| EMR w/ AmpNorm | **54.05** | **66.79** | **61.67** | **46.54** | **57.26** (5.51 ↑ ) |
| AugMix [7] | 55.44 | 69.19 | 70.36 | 51.95 | 61.73 |
| AugMix w/ AmpNorm | **58.68** | **70.64** | **70.48** | **53.08** | **63.22** (1.49 ↑ ) |
| L2D [28] | 59.55 | ***74.95*** | 72.05 | 50.70 | 64.31 |
| L2D w/ AmpNorm | **60.91** | 74.24 | ***73.41*** | **54.92** | **65.87** (1.56↑) |
| ACVC [4] | 58.40 | **72.09** | **72.38** | 51.68 | 63.64 |
| ACVC w/ AmpNorm | **61.88** | 71.03 | 72.05 | **52.00** | **64.24** (0.60 ↑ ) |
| DSU [15] | 53.06 | 72.21 | **66.58** | 46.58 | 59.61 |
| DSU w/ AmpNorm | **59.11** | **72.60** | 65.93 | **47.81** | **61.36** (1.75 ↑ ) |
| DAC [11] | 56.17 | 73.06 | 70.66 | 56.92 | 64.20 |
| DAC w/ AmpNorm | ***62.31*** | **73.87** | **71.10** | ***57.40*** | ***66.17*** (2.51 ↑ ) |

**Table 2** SDG classification accuracies (%) with ResNet18 as backbone on the Office-Home dataset, where each SOTA is used as the baseline. We use a source domain for model training and the remaining three domains for model evaluation.

| Method | Art | Clipart | Product | Real | Average |
|---|---|---|---|---|---|
| EMR [24] | 44.90 | 36.46 | 35.12 | 51.47 | 41.99 |
| EMR w/ AmpNorm | **45.52** | **36.96** | **39.90** | **54.41** | **44.20** (2.21 ↑ ) |
| AugMix [7] | 50.81 | 48.41 | 45.60 | 57.28 | 50.53 |
| AugMix w/ AmpNorm | **51.23** | **49.26** | **48.33** | **60.27** | **52.27** (1.74 ↑ ) |
| L2D [28] | 51.93 | **46.04** | 47.95 | 57.53 | 50.86 |
| L2D w/ AmpNorm | **52.27** | 46.03 | **49.09** | **61.04** | **52.11** (1.25↑) |
| ACVC [4] | 52.15 | 51.97 | 48.35 | 59.18 | 52.91 |
| ACVC w/ AmpNorm | **52.19** | ***52.52*** | **50.31** | **62.98** | ***54.50*** (1.59 ↑ ) |
| DSU [15] | 49.45 | 42.99 | 44.41 | 57.20 | 48.51 |
| DSU w/ AmpNorm | **50.43** | **44.30** | **47.21** | **59.79** | **50.43** (1.92 ↑ ) |
| DAC [11] | ***52.84*** | 47.89 | 48.38 | 61.19 | 52.57 |
| DAC w/ AmpNorm | 52.46 | **48.73** | ***51.47*** | ***62.59*** | **53.81** (1.24 ↑ ) |

**Table 3** The performances of the SOTAs on the DomainNet benchmark for SDG with ResNet18 as backbone, where each SOTA is used as the baseline. We use a source domain for model training and the remaining five domains for model evaluation

| Method | Clipart | Infograph | Painting | Quickdraw | Real | Sketch | Average |
|---|---|---|---|---|---|---|---|
| EMR [24] | 22.17 | 17.22 | 23.68 | 2.92 | 19.33 | 22.84 | 18.03 |
| EMR w/ AmpNorm | **23.38** | **18.86** | **25.33** | **6.09** | **21.94** | **23.83** | **19.90** (1.87 ↑) |
| AugMix [7] | 24.73 | 18.38 | 23.87 | 5.52 | 21.05 | 24.69 | 19.71 |
| AugMix w/ AmpNorm | **25.17** | **19.87** | **25.80** | **8.05** | **23.41** | ***25.74*** | **21.34**(1.63↑) |
| L2D [28] | 23.05 | 18.38 | 23.14 | 5.09 | 21.12 | 23.83 | 19.10 |
| L2D w/ AmpNorm | **25.38** | **19.06** | **26.59** | **7.05** | ***24.19*** | **24.78** | **21.18**(2.08↑) |
| ACVC [4] | 25.86 | 19.60 | 23.65 | 8.11 | 22.46 | 24.26 | 20.66 |
| ACVC w/ AmpNorm | ***26.72*** | ***20.80*** | **25.20** | ***9.06*** | **23.88** | **24.97** | ***21.77*** (1.11 ↑) |
| DSU [15] | 23.62 | 18.88 | 25.86 | 5.85 | 21.27 | 23.11 | 19.76 |
| DSU w/ AmpNorm | **25.57** | **19.45** | ***26.76*** | **8.15** | **22.97** | **24.47** | **21.33** (1.57 ↑) |
| DAC [11] | 24.09 | 18.10 | 23.73 | 5.99 | 22.69 | 24.99 | 19.93 |
| DAC w/ AmpNorm | **24.87** | **19.73** | **24.55** | **7.47** | **22.91** | **25.38** | **20.82** (0.89 ↑) |

baseline and the SOTA methods to improve their performances. For example, AmpNorm improves the performance of ERM by 1.87% and improves that of DAC by 0.89%.

**Table 4** The running time (ms) of the baseline and our AmpNorm based on different backbones.

| Backbone | ResNet18 | ResNet50 | ViT-B | Swim-B |
|---|---|---|---|---|
| Baseline | 15.0 | 46.1 | 151.8 | 151.7 |
| AmpNorm | 17.7(+2.2) | 48.8(+2.7) | 153.5(+1.7) | 152.5(+0.8) |

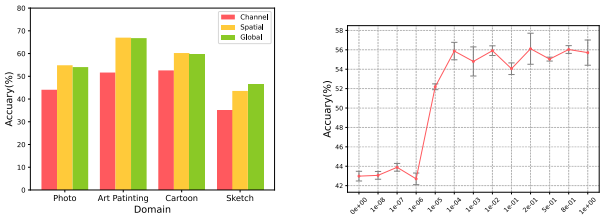
All the aforementioned comparisons demonstrate the effectiveness of our AmpNorm in enhancing generalization performance across domains. Meanwhile, our AmpNorm is plug-and-play and can be seamlessly incorporated into SOTA methods without additional runtime overhead, as shown in Table 4.

### *4.3. Ablation Study*

**Amplitude Dimension.** To study the performance of our AmpNorm against the dimension of the source domain amplitude, i.e. Eq. (5), we present the results in Fig. 3(a). The channel amplitude $\in \mathbb{R}^{C}$ the spatial amplitude $\in \mathbb{R}^{H\times W}$ and the global amplitude $\in \mathbb{R}^{C\times H\times W}$ mean the amplitude of the frequency domain in the channel, spatial, and global dimensions, respectively. The channel amplitude has only $C = 3$ parameters but the performance is unsatisfactory, while the global amplitude has the most parameters and achieves the best performance. This is probably that the channel amplitude is too simple to represent the style information of the source domain, while the global amplitude can represent rich style information of the source domain.

**Performance sensitivity against $\lambda$ in Exponential Moving Average (EMA).** The outcomes that study the influence of the $\lambda$ of EMA on the AmpNorm are shown in Fig. 3(b). It is evident that this hyperparameter has little influence on the performance of the model within a certain range. When $\lambda$ is set as 0.0, the source domain amplitude is not updated, and the performance is the worst. When $\lambda$ is set as 1.0, the source domain amplitude belongs to the last mini-batch in the training process. A suitable $\lambda$ can help the model collect source domain amplitude quickly in the early training stage, and stabilize the model training in the later stage. The trained model has a certain robustness against the style of the source domain, the value of $\lambda$ has minimal impact on the generalization performance of the model even if it is set as 1.0.

**Dynamic $\lambda$ of EMA in Eq. (6).** While a large $\lambda$ can make the model collect source domain amplitude quickly, this is not conducive to the stable training of the model. Therefore, we consider a dynamic $\lambda$ strategy for the EMA (Eq. (6)) of the AmpNorm, i.e. $\lambda$ is initialized to 1.0 and decays to 0.0 based on the cosine annealing strategy. And $\lambda$ keeps to 0.0 in the remaining training iterations, which means that the source domain amplitude is not updated. The results are shown in Table 5, where the dynamic iteration is the number of iterations of the cosine annealing strategy. The results demonstrate that the dynamic $\lambda$ strategy can enhance the performance of the trained model to a certain extent.




***(a)*** *Dimension of the amplitude* ***(b)*** *λ of the EMA*

***Fig. 3.*** *Ablation study of (a) the amplitude dimension and (b) the λ of the EMA (Eq. (6)) on the PACS dataset with ResNet18.*

**Table 5** Ablation study of our AmpNorm on the number of dynamic iterations of λ in EMA (Eq. (6)) with ResNet18 on the PACS dataset.

| Dynamic Iterations | Photo | Art Painting | Cartoon | Sketch | Average |
|---|---|---|---|---|---|
| static | 55.04 | 67.99 | 61.00 | 44.70 | 57.18 |
| 1 | 54.69 | 67.45 | 59.54 | 43.29 | 56.24 |
| 300 | 54.72 | **68.49** | 59.93 | 44.59 | 56.93 |
| 500 | 55.64 | 68.11 | 60.83 | **45.45** | **57.51** |
| 1000 | 55.67 | 65.88 | 60.06 | 43.02 | 56.16 |
| 4000 | 54.69 | 66.75 | **60.96** | 44.27 | 56.66 |
| 12000 | **55.75** | 67.29 | 60.79 | 43.40 | 56.81 |

### *4.4. Visualization*

To study the style-normalized samples by our AmpNorm, we show those from four domains of the PACS dataset that are converted into different styles by our AmpNorm, as shown in Fig. 4. We can see that the samples of the unseen target domain can be converted into samples with a style similar to the source domain by our AmpNorm. For example, the samples of the Sketch domain (4th, 8th columns) can be converted into samples with a style similar to the Photo domain (4th row). We can see that our AmpNorm can transfer the source domain style information to the target domain.


***Fig. 4.*** *Visualization of the AmpNorm on the PACS dataset for the domains of Art Painting, Cartoon, Photo, Sketch.*

## 5. Conclusion

Inspired by the frequency domain amplitude is related to the image style, this paper proposed a simple yet effective method, i.e. Amplitude Normalization (AmpNorm), which explores this amplitude to surrogate the style cues for reducing the domain gap and enhancing the generalization performance of the learned model. In contrast to existing data manipulation methods, our AmpNorm does not generate new samples but converts the unseen target domain samples into samples that resemble the source domain in terms of style. Our AmpNorm is easy to implement, and plug-and-play to most existing methods, and can largely improve their generalization capacity. Extensive experiments on three benchmarks of SDG tasks demonstrate that our AmpNorm can greatly improve the state-of-the-art approaches to achieve better performance.

## 6. Acknowledgements

The work was supported by the Natural Science Foundation of China under grants no. 62276170, 82261138629, the Science and Technology Project of Guangdong Province under grants no. 2023A1515011549, 2023A1515010688, the Science and Technology Innovation Commission of Shenzhen under grant no. JCYJ20220531101412030.

## 7. References

[1] Blanchard, G., Lee, G., Scott, C.: 'Generalizing from several related classification tasks to a new unlabeled sample', Advances in neural information processing systems 24, 2011

[2] Carlucci, F.M., D'Innocente, A., Bucci, S., et al.: 'Domain generalization by solving jigsaw puzzles', In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019

[3] Choi, S., Kim, et al.: 'Meta batch-instance normalization for generalizable person re-identification', In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition., 2021, pp. 3425–3435

[4] Cugu, I., Mancini, M., Chen, Y., Akata, et al.: 'Attention consistency on visual corruptions for single-source domain generalization', In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. June 2022, pp. 4165–4174

[5] Deng, J., Dong, W., Socher, R., et al.: 'Imagenet: A large-scale hierarchical image database', In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009, pp. 248–255

[6] He, K., Zhang, X., Ren, S., et al.: 'Deep residual learning for image recognition', In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. June 2016, 770–778

[7] Hendrycks, D., Mu, N., Cubuk, E.D., et al.: 'AugMix: A simple data processing method to improve robustness and uncertainty', Proceedings of the International Conference on Learning Representations (ICLR) , 2020

[8] Huang, J., Guan, D., Xiao, A., et al.: 'Fsdr: Frequency space domain randomization for domain generalization', In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, pp. 6891–6902

[9] Jin, X., Lan, C., Zeng, W., et al.: 'Style normalization and restitution for generalizable person re-identification', In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, pp. 3143–3152

[10] Kingma, D.P., Ba, J. 'Adam: A method for stochastic optimization', arXiv preprint arXiv:1412.6980, 2014
[11] Lee, S., Bae, J., Kim, H.Y. 'Decompose, adjust, compose: Effective normalization by playing with frequency for domain generalization', In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 2023, pp. 11776–11785
[12] Li, D., Yang, Y., Song, Y.Z., et al.: 'Deeper, broader and artier domain generalization', In: Proceedings of the IEEE international conference on computer vision. 2017, pp. 5542–5550
[13] Li, D., Zhang, J., Yang, Y., et al.: 'Episodic training for domain generalization', In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019, pp. 1446–1455
[14] Li, L., Gao, K., Cao, J., et al.: 'Progressive domain expansion network for single domain generalization', In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, pp. 224–233
[15] Li, X., Dai, Y., Ge, Y., et al.: 'Uncertainty modeling for out-of-distribution generalization', In: International Conference on Learning Representations. 2022
[16] Li, Y., Tian, X., Gong, M., et al.: 'Deep domain generalization via conditional invariant adversarial networks', In: Proceedings of the European conference on computer vision (ECCV). 2018, pp. 624–639
[17] Loshchilov, I., Hutter, F. 'Sgdr: Stochastic gradient descent with warm restarts', arXiv preprint arXiv:1608.03983. 2016
[18] Nam, H., Lee, H., Park, J., et al.: 'Reducing domain gap by reducing style bias', In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, pp. 8690–8699
[19] Oppenheim, A., Lim, J., Kopec, G., et al.: 'Phase in speech and pictures', In: ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing. 1979, vol. 4, pp. 632–637. IEEE
[20] Oppenheim, A.V., Lim, J.S. 'The importance of phase in signals', Proceedings of the IEEE. 1981, 69(5), 529–541
[21] Peng, X., Bai, Q., Xia, X., et al.: 'Moment matching for multi-source domain adaptation', In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). October 2019
[22] Piotrowski, L.N., Campbell, F.W. 'A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase', 1982, Perception 11(3), 337–346
[23] Qu, S., Pan, Y., Chen, G., et al.: Modality-agnostic debiasing for single domain generalization, ArXiv abs/2303.07123. 2023
[24] Vapnik, V. 'Statistical learning theory', 1998
[25] Venkateswara, H., Eusebio, J., Chakraborty, S., et al.: 'Deep hashing network for unsupervised domain adaptation', In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 2017
[26] Wan, C., Shen, X., Zhang, Y., et al.: 'Meta convolutional neural networks for single domain generalization', In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 2022, pp. 4682–4691
[27] Wang, J., Lan, C., Liu, C., et al.: 'Generalizing to unseen domains: A survey on domain generalization', IEEE Transactions on Knowledge and Data Engineering. 2022
[28] Wang, Z., Luo, Y., Qiu, R., et al.: 'Learning to diversify for single domain generalization', In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). October 2021, pp. 834–843
[29] Xu, Q., Zhang, R., Zhang, Y., et al.: 'A fourier-based framework for domain generalization', In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, pp. 14383–14392
[30] Y. Tian, Z. Wen, W. Xie, et al.: 'Outlier-Suppressed Triplet Loss with Adaptive Class-Aware Margins for Facial Expression Recognition', 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 46-50