HAP: Harmonized Amplitude Perturbation for Cross-Domain Few-Shot Learning

Wenqian Li[1,2], **Pengfei Fang**[1,2*], **Hui Xue**[1,2*]

[1]School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[2]Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, China
{wenqianli.li, fangpengfei, hxue}@seu.edu.cn

Abstract

Cross-Domain Few-Shot Learning (CD-FSL) remains a significant challenge due to substantial distribution shifts between source and target domains. While prior approaches primarily focus on spatial alignment, they often overlook discrepancies in the frequency domain. In this paper, we reveal frequency band discretization as a key phenomenon, characterized by intra-domain low-frequency dominance, inter-domain amplitude divergence, and limited high-frequency variation. This spectral disharmony biases models toward low-frequency components, leading to spectral collapse. We quantify spectral collapse via the effective rank, a principled measure of spectral diversity. To mitigate spectral collapse, we propose Harmonized Amplitude Perturbation (HAP), a frequency-domain augmentation strategy that perturbs the amplitude spectrum via frequency-aware gains sampled from Harmonized Distributions, while fixing the phase spectrum to maintain semantic integrity. Extensive experiments on both Cross-Domain Few-Shot Image Classification and Object Detection benchmarks demonstrate that HAP effectively increases spectral diversity and consistently improves generalization, outperforming state-of-the-art methods without introducing extra model complexity.

Introduction

Vision models have achieved remarkable progress in tasks such as image classification and object detection (Feng et al. 2025b; Fu et al. 2025; Shi et al. 2025; Xue et al. 2025). However, their performance degrades significantly when generalizing to novel domains. This challenge is especially pronounced in Cross-Domain Few-Shot Learning (CD-FSL), where substantial domain shifts arise from disjoint data and label distributions between source and target domains.

Numerous approaches have been proposed to address this challenge by improving spatial alignment. Many of these methods focus on transforming visual appearance (Hu and Ma 2022; Feng et al. 2025a) or adjusting statistical properties (*e.g.* mean and standard deviation) of images and features (Li, Fang, and Xue 2025; Feng et al. 2025d). While these techniques are effective in aligning spatial distributions between source and target domains, they often neglect discrepancies in the frequency domain, as shown in Figure 1. The frequency discrepancies undermine spectral harmony and significantly hinder cross-domain generalization.

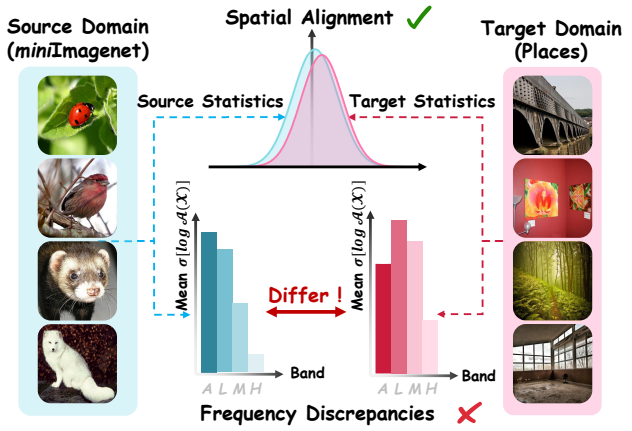

Figure 1: While spatial alignment is achieved, frequency discrepancies persist. "A", "L", "M", and "H" denote All-, Low-, Middle-, and High-Frequency Bands, respectively.

In this paper, we explore this overlooked dimension and uncover a fundamental yet underexplored phenomenon, which we refer to as **frequency band discretization**. As illustrated in Figure 2(a), this phenomenon manifests in three key ways: (1) Intra-domain low-frequency dominance, where within a single domain, spectral energy tends to concentrate in the low-frequency (LF) band, leaving middle-frequency (MF) and high-frequency (HF) components underutilized. (2) Inter-domain amplitude divergence, where across domains, the amplitude distributions differ significantly across all frequency bands. (3) Limited HF variation in source domains, where the source domain often exhibits higher variability in the LF band but suppressed variation in MF and HF bands, as compared to target domains.

This spectral disharmony biases models toward low-frequency components, eventually resulting in **spectral collapse**, a pathological condition where spectral diversity is underutilized, severely impairing generalization. To quantify the spectral collapse, we employ the effective rank (Roy and Vetterli 2007) as a principled metric of spectral diversity. A lower effective rank indicates the spectral energy is

*Corresponding author.

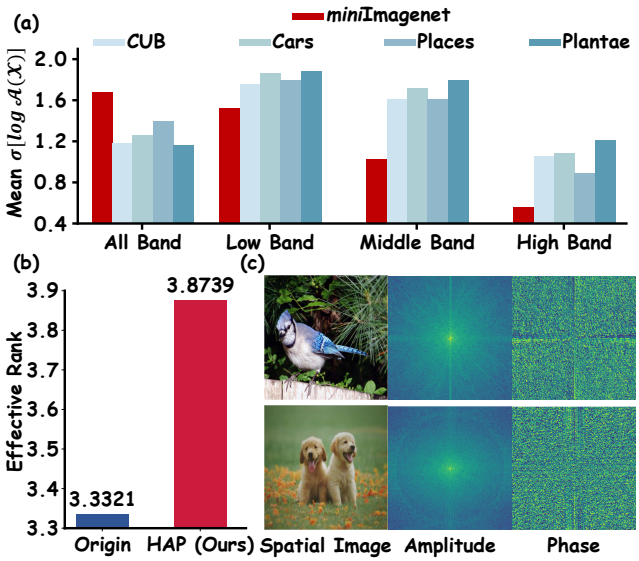

Figure 2: (a) Frequency discretization. (b) Effective rank. (c) Spatial image, amplitude spectrum and phase spectrum.

overly concentrated, thus signifying a higher degree of collapse (Park et al. 2024). As shown in Figure 2(b), our proposed method significantly increases the effective rank of transformed images, indicating a more harmonized utilization of frequency components.

Specifically, to mitigate spectral collapse, we propose **H**armonized **A**mplitude **P**erturbation (**HAP**), a frequency-domain augmentation strategy designed to promote harmonized spectral utilization. As shown in Figure 2(c), HAP transforms a spatial image into the frequency domain via a 2D Fast Fourier Transform (FFT) to obtain amplitude and phase spectra. The amplitude spectrum is perturbed using frequency-aware gains, while the phase spectrum is fixed to maintain semantic integrity. We introduce a Spatial-Frequency (SF) Grid to characterize the relative distance of each frequency component from the spectral center. By scaling statistical properties (*e.g.* mean and standard deviation) of this grid and applying a mapping function, we construct a Harmonized Spatial-Frequency (HSF) Grid. Each pixel in the HSF Grid encodes a frequency-specific standard deviation, from which we model multiple Harmonized Distributions. Frequency-aware gains sampled from these distributions are applied to perturb the amplitude spectrum. Finally, the perturbed amplitude spectrum is recombined with the original phase spectrum via inverse FFT (iFFT), generating an augmented image with enhanced spectral diversity.

The main contribution of our paper is:

- We identify and characterize frequency band discretization as a key limitation in CD-FSL. We further reveal that the spectral disharmony introduces learning bias, eventually leading to spectral collapse.
- We introduce the effective rank, a principled metric of spectral diversity to quantify spectral collapse and analyze frequency component utilization.
- We propose Harmonized Amplitude Perturbation (HAP), a plug-and-play frequency-domain augmentation strategy that perturbs the amplitude spectrum with frequency-aware gains, effectively increasing spectral diversity.
- We conduct extensive experiments on both Cross-Domain Few-Shot Image Classification (CD-FSIC) and Object Detection (CD-FSOD) benchmarks. Results show that HAP significantly improves baseline performance and achieves state-of-the-art results.

Related Work

Cross-Domain Few-Shot Learning

Few-Shot Learning (FSL) aims to train the model with limited labeled data. As an extension of FSL, CD-FSL addresses scenario where source and target domains have varying degrees of domain shift (Perera and Halgamuge 2024). Generally, there are two types of approaches to solving CD-FSL problems. The first involves designing additional modules (Feng et al. 2025c; Zou et al. 2024), which increases parameter count and model complexity. The second focuses on data augmentation to expand training data (Fu et al. 2023; Zhou et al. 2023), which is easier and comparably effective. For the second approaches, Zhou *et al*. (Zhou et al. 2023) incorporated local image crops to diversify the data. Zheng (Zheng et al. 2023) transformed a small portion of target images into different augmented images. Moreover, several works have considered the style of features. For example, Fu (Fu et al. 2023) perturbed the original styles to obtain the adversarial styles iteratively. However, these visually-plausible transformation strategies often fail to improve models' robustness against noise concentrated in the HF regions of the frequency spectrum (Yin et al. 2019).

Frequency-based Data Augmentation

Frequency domain analysis has long been recognized as a crucial tool in digital image processing (Lin et al. 2024). Recently, frequency-based augmentation has shown overwhelming performance on various vision tasks (He et al. 2024). These strategies generally manipulate the amplitude or phase spectrum in the frequency domain to introduce informative variations that are often missed in the spatial domain. For example, Wave-SAN (Fu et al. 2022) augments source domain styles by swapping LF components between episodes using wavelet decomposition. While this approach modifies global structure and texture effectively, it focuses solely on LF information and neglects MF and HF discrepancies that are critical for fine-grained representation. FreqAug (Kim et al. 2023) randomly removes specific frequency components. However, this strategy lacks frequency awareness and may discard useful information indiscriminately. DG-UCDIR (Hu, Zhang, and Lee 2023) introduces a two-stage frequency-domain augmentation that modifies both amplitude and phase spectra in the LF domain. Although it enhances LF robustness, it fails to address spectral disharmony across the all-frequency spectrum.

Methodology

Preliminaries

Fourier Transformation. Given an input image $\boldsymbol{I} \in \mathbb{R}^{H\times W\times C}$ where H, W and C denote the height, width, and number of channels respectively, the FFT is applied channel-wise to obtain the frequency representation as follows:

$$\boldsymbol{F}_I(u,v,c) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} \boldsymbol{I}(h,w,c) e^{-j2\pi(\frac{h}{H}u+\frac{w}{W}v)}, \quad (1)$$

where $j^2 = -1$ and u, v denote the spatial frequencies in the horizontal and vertical directions, respectively. The amplitude and phase spectra are computed as:

$$\boldsymbol{\mathcal{A}}_I(u,v,c) = \sqrt{\mathrm{Im}(\boldsymbol{F}_I(u,v,c))^2 + \mathrm{Re}(\boldsymbol{F}_I(u,v,c))^2} \quad (2)$$

and

$$\boldsymbol{\mathcal{P}}_I(u,v,c) = \arctan\Big(\frac{\mathrm{Im}\big(\boldsymbol{F}_I(u,v,c)\big)}{\mathrm{Re}\big(\boldsymbol{F}_I(u,v,c)\big)}\Big), \quad (3)$$

where $\mathrm{Im}(\cdot)$ and $\mathrm{Re}(\cdot)$ represent the imaginary and real parts of $\boldsymbol{F}_I$, respectively. The LF components of the amplitude and phase spectra are shifted to the center of the frequency spectrum by default. The inverse FFT (iFFT) reconstructs the spatial image from its frequency representation as:

$$\boldsymbol{I}(h,w,c) = \frac{1}{H\cdot W}\sum_{u=0}^{H-1}\sum_{v=0}^{W-1} \boldsymbol{F}_I(u,v,c) e^{j2\pi(\frac{h}{H}u+\frac{w}{W}v)}. \quad (4)$$

Spatial-Frequency Grid. To facilitate frequency-aware augmentation, we introduce a Spatial-Frequency (SF) Grid $\boldsymbol{G} \in \mathbb{R}^{H\times W}$, where each entry corresponds to the distance of a spatial coordinate (u, v) from the frequency center:

$$\boldsymbol{G}(u,v) = \sqrt{u^2+v^2}. \quad (5)$$

This grid serves as a spatial reference to characterize and manipulate spectral energy distributions. Following prior studies of natural image spectra, which generally follow a power law of the form $\frac{1}{f^{\varpi}}$ (Swartz et al. 2024), we utilize the SF Grid to differentiate between LF, MF, and HF components and guide frequency-aware perturbations.

Spectral Collapse

In CD-FSL, models exhibit a tendency to overemphasize LF components, leading to an underutilization of MF and HF signals. We term this degradation in spectral diversity as spectral collapse. To quantify spectral collapse, we adopt the effective rank as a principled measure of spectral diversity, inspired by prior work (Park et al. 2024). Given a matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ with singular values $\{\lambda_i\}$, its effective rank is defined as the entropy of its normalized singular values:

$$\rho(\boldsymbol{X}) = -\sum_{i=1}^{min(m,n)} \bar{\lambda}_i \, log \, \bar{\lambda}_i, \quad (6)$$

where $\bar{\lambda}_i = \frac{\lambda_i}{\sum_j \lambda_j}$. A lower effective rank indicates that the spectral energy is overly concentrated in a small number of components, signaling stronger spectral collapse.

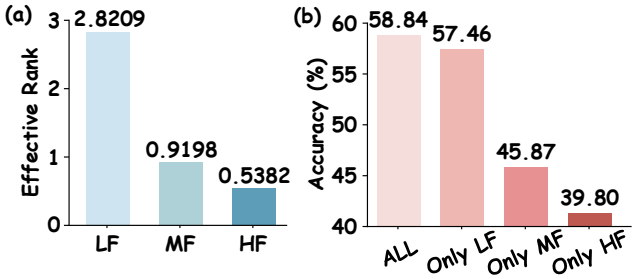

Figure 3: (a) Effective rank when retaining only specific frequency bands. (b) Average classification accuracy when using all-frequency bands versus individual frequency bands.

As illustrated in Figure 3(a), the effective rank when retaining only LF components is significantly higher than that of MF and HF components, indicating that LF features dominate the learned representations. We further conduct an ablation study in which individual frequency components are selectively retained. As shown in Figure 3(b), the model retains high classification accuracy when only LF components are preserved, indicating its strong reliance on coarse-grained features. In contrast, performance degrades substantially when the model is restricted to either MF or HF components, both of which are known to encode more fine-grained structures. These results suggest that the model underutilizes the MF and HF components, which are critical for recognizing transferable patterns under distribution shifts. Overall, our analysis reveals a pronounced spectral collapse in current models, which disproportionately favor LF components while largely ignoring HF cues.

The Proposed HAP

In this section, we first give an overview of our proposed HAP. Then, we elaborate on each component in detail.

Overview. As depicted in Figure 4, our approach HAP is designed to effectively perturb the source images in the frequency domain with Harmonized Distributions. HAP first transforms an input image $\boldsymbol{I}$ by applying FFT and introduces Spatial-Frequency (SF) Grid $\boldsymbol{G}$. Scaling the statistical properties of $\boldsymbol{G}$, we then project the masked SF grid $\boldsymbol{G}_M$ to obtain the Harmonized Spatial-Frequency (HSF) Grid $\boldsymbol{G}_H$ by employing a designed mapping function. Using the value at each pixel of $\boldsymbol{G}_H$ as the standard deviation, several Harmonized Distributions $\boldsymbol{\Phi}$ are modeled. The original amplitude spectrum is perturbed by applying the harmonized gains, randomly sampled from these Harmonized Distributions, along with Gaussian noise $\boldsymbol{n}$. Then, the augmented image $\hat{\boldsymbol{I}}$ is generated by applying iFFT to the perturbed amplitude spectrum and the fixed phase spectrum.

Harmonized Spatial-Frequency Grid. To construct the Harmonized Spatial-Frequency (HSF) Grid $\boldsymbol{G}_H$, we adopt a two-stage process designed to refine and redistribute the spatial-frequency information.

In the first stage, we compute the global statistical prop-

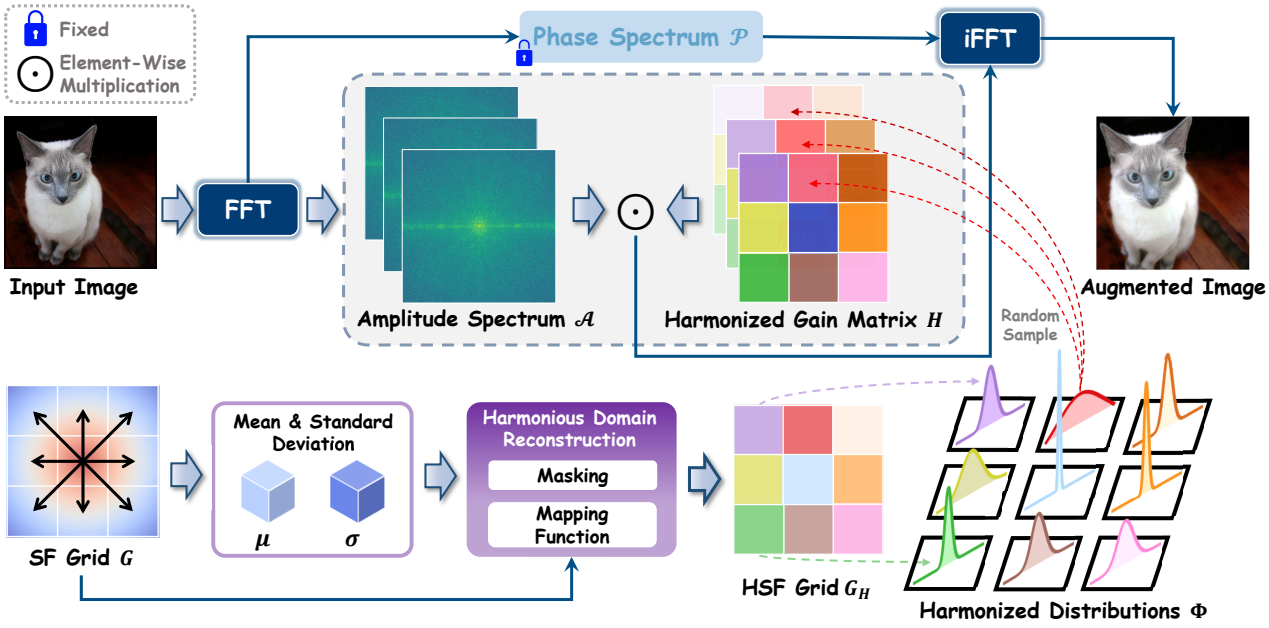

Figure 4: Overview of our proposed **HAP**. "SF" is an abbreviation for Spatial-Frequency. "HSF" is an abbreviation for Harmonized Spatial-Frequency. The input image is transformed via FFT to obtain amplitude and phase spectra. A Harmonized Spatial-Frequency Grid guides the generation of frequency-aware gains, which perturb the amplitude spectrum.

erties (*e.g.* mean and standard deviation) of $\boldsymbol{G}$ as:

$$\boldsymbol{\mu} = \frac{1}{HW}\sum_{u=1}^{H}\sum_{v=1}^{W}\boldsymbol{G}(u,v) \tag{7}$$

and

$$\boldsymbol{\sigma} = \sqrt{\frac{1}{HW}\sum_{u=1}^{H}\sum_{v=1}^{W}(\boldsymbol{G}(u,v)-\boldsymbol{\mu})^2 + \epsilon}, \tag{8}$$

where ϵ is a small value to ensure the numerical stability. Subsequently, we introduce a binary mask $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{H\times W}$ to suppress the LF components and emphasize the MF and HF components, whose values are set to zero for all regions except these near the frequency center:

$$\boldsymbol{\mathcal{M}}(u,v) = \begin{cases} 1, & \text{if } \boldsymbol{G}(u,v) \le \tau \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where τ is the LF threshold (Wang et al. 2023). The masked SF Grid $\boldsymbol{G}_M$ is then obtained as:

$$\boldsymbol{G}_M = (\mathbf{1} - \boldsymbol{\mathcal{M}}) \odot \boldsymbol{G}. \tag{10}$$

In the second stage, we apply a mapping function to remap $\boldsymbol{G}_M$ to a smoothly distributed HSF grid $\boldsymbol{G}_H$, enhancing frequency diversity while preserving consistency with the original spectral statistics. The mapping is defined as:

$$\boldsymbol{G}_H(u,v) = \left(\rho \cdot \exp\left(-\frac{(\boldsymbol{G}_M(u,v) - \alpha\cdot\boldsymbol{\mu})^2}{2(\beta\cdot\boldsymbol{\sigma})^2}\right)\right)^{\gamma} + \varepsilon, \tag{11}$$

where ρ, γ and ε are hyper-parameters that control the smoothness and balance of the frequency distribution. The parameters α and β are scale factors applied to $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.

Harmonized Gain Matrix. To enable frequency-aware perturbations, we construct a set of Gaussian distributions, referred to as Harmonized Distributions $\boldsymbol{\Phi}$. For each spatial frequency coordinate (u,v), a Gaussian distribution is defined with a fixed mean of 1 and a standard deviation determined by the corresponding value in $\boldsymbol{G}_H$. Formally, the Harmonized Distributions $\boldsymbol{\Phi}$ are defined as:

$$\boldsymbol{\Phi}(u,v) = \mathcal{N}\left(\mathbf{1}, \boldsymbol{G}_H^2(u,v)\right). \tag{12}$$

Since higher frequencies generally correspond to larger values in $\boldsymbol{G}_H$, they are assigned greater standard deviations, leading to stronger perturbations and promoting more diverse HF variations.

From these distributions, we construct the Harmonized Gain Matrix $\boldsymbol{H} \in \mathbb{R}^{H\times W\times C}$ by randomly sampling gain values at each spatial frequency and channel:

$$\boldsymbol{H}(u,v,c) \sim \boldsymbol{\Phi}(u,v), \tag{13}$$

where each element of $\boldsymbol{H}$ represents a frequency-aware gain applied to the amplitude spectrum at coordinate (u,v) for channel c.

Augmented Images. The original amplitude spectrum is thus perturbed with $\boldsymbol{H}$ along with additional Gaussian noise $\boldsymbol{n} \sim \mathcal{N}(0, 0.01) \in \mathbb{R}^{H\times W\times C}$, formulated as:

$$\hat{\boldsymbol{\mathcal{A}}}_I = \boldsymbol{\mathcal{A}}_I \odot \boldsymbol{H} + \boldsymbol{n}. \tag{14}$$

The final augmented image $\hat{\boldsymbol{I}}$ can be obtained by:

$$\hat{\boldsymbol{I}} = \mathrm{iFFT}(\hat{\boldsymbol{\mathcal{A}}}_I e^{-j\boldsymbol{\mathcal{P}}_I}). \tag{15}$$

Method	FT	ChestX	ISIC	EuroSAT	CropDisease	CUB	Cars	Places	Plantae	Aver.
GNN	✗	22.00±0.46	32.02±0.66	63.69±1.03	64.48±1.08	45.69±0.68	31.79±0.51	53.10±0.80	35.60±0.56	43.55
GNN + Ours	✗	**23.01±0.35**	**35.68±0.60**	**66.56±0.80**	**67.05±0.78**	**45.73±0.67**	**33.18±0.54**	**54.09±0.78**	**38.15±0.62**	**45.43**
ATA	✗	22.10±0.20	33.21±0.40	61.35±0.50	67.47±0.50	45.00±0.50	33.61±0.40	53.57±0.50	34.42±0.40	43.84
ATA + Ours	✗	**22.35±0.33**	**37.20±0.55**	**65.66±0.72**	**68.89±0.77**	**46.99±0.69**	**34.07±0.57**	**54.87±0.76**	**36.51±0.59**	**45.82**
StyleAdv	✗	22.64±0.35	33.96±0.57	70.94±0.82	74.13±0.78	48.49±0.72	34.64±0.57	58.58±0.83	41.13±0.67	48.06
StyleAdv + Ours	✗	**23.20±0.34**	**38.05±0.63**	**73.19±0.82**	**74.41±0.77**	47.71±0.70	**35.50±0.59**	57.71±0.78	**41.14±0.62**	**48.86**
FLoR	✗	23.11±0.31	38.11±0.51	62.95±0.65	71.34±0.64	49.99±0.62	37.41±0.54	49.75±0.63	40.12±0.59	46.60
FLoR + Ours	✗	**23.70±0.33**	**38.54±0.52**	**64.99±0.64**	**71.89±0.67**	**52.01±0.63**	**38.05±0.55**	**51.06±0.62**	**42.20±0.60**	**47.81**
GNN	✓	22.13±0.49	34.64±0.62	66.18±1.01	73.43±1.07	43.56±0.72	35.12±0.59	51.77±0.79	38.12±0.55	45.62
GNN-FT + Ours	✓	**23.01±0.35**	**35.65±0.53**	**70.62±0.74**	**77.61±0.69**	**45.73±0.67**	**35.94±0.58**	**55.27±0.79**	**38.15±0.62**	**47.75**
ATA-FT	✓	22.15±0.20	34.94±0.40	68.62±0.50	75.41±0.50	46.23±0.50	37.15±0.40	54.18±0.50	37.38±0.40	47.01
ATA-FT + Ours	✓	**22.74±0.34**	**37.20±0.55**	**68.97±0.52**	**75.92±0.48**	**46.99±0.69**	**37.85±0.40**	**54.87±0.76**	**38.19±0.39**	**47.84**
StyleAdv-FT	✓	22.64±0.35	35.76±0.52	72.92±0.75	80.69±0.28	48.49±0.72	35.09±0.55	58.58±0.83	41.13±0.67	49.41
StyleAdv-FT + Ours	✓	**23.20±0.34**	**38.08±0.50**	**74.55±0.79**	**80.82±0.28**	**48.96±0.70**	**36.57±0.59**	**59.02±0.78**	**41.16±0.62**	**50.30**
FLoR-FT	✓	23.12±0.36	38.82±0.53	69.13±0.80	84.04±0.32	50.01±0.70	38.14±0.58	50.68±0.81	40.20±0.68	49.27
FLoR-FT + Ours	✓	**23.78±0.33**	38.58±0.50	**69.81±0.78**	**84.19±0.30**	**52.22±0.70**	**38.57±0.56**	**51.11±0.82**	**42.25±0.66**	**50.06**

Table 1: Classification accuracy of 5-way 1-shot tasks on CD-FSL based on ResNet-10. "**FT**" with ✓ means fine-tuning is used, vice versa. "**Aver.**" means "Average Accuracy". The improved results are marked in **bold**.

Method	FT	ChestX	ISIC	EuroSAT	CropDisease	CUB	Cars	Places	Plantae	Aver.
GNN	✗	25.27±0.46	43.94±0.67	83.64±0.77	87.96±0.67	62.25±0.65	44.28±0.63	70.84±0.65	52.53±0.59	58.84
GNN + Ours	✗	**26.32±0.37**	**48.26±0.57**	**84.88±0.58**	**89.49±0.45**	**62.49±0.67**	**47.61±0.59**	**73.07±0.64**	**56.54±0.63**	**61.08**
ATA	✗	24.32±0.40	44.91±0.40	83.75±0.40	90.59±0.30	66.22±0.50	49.14±0.40	75.48±0.40	52.69±0.40	60.89
ATA + Ours	✗	**25.93±0.36**	**51.30±0.54**	**86.67±0.53**	**90.85±0.47**	**66.80±0.68**	**49.49±0.64**	74.64±0.63	**53.87±0.64**	**62.44**
StyleAdv	✗	26.07±0.37	45.77±0.51	86.58±0.54	93.65±0.39	68.72±0.67	50.13±0.68	77.73±0.62	61.52±0.68	63.77
StyleAdv + Ours	✗	**26.39±0.36**	**51.56±0.55**	**89.57±0.48**	**93.91±0.37**	66.62±0.70	**50.75±0.62**	**77.84±0.60**	60.85±0.65	**64.69**
FLoR	✗	26.71±0.45	51.44±0.49	80.87±0.48	89.02±0.41	70.39±0.55	53.43±0.60	68.51±0.54	55.81±0.57	62.02
FLoR + Ours	✗	**27.47±0.35**	**51.80±0.50**	**81.95±0.48**	**89.06±0.41**	**71.35±0.52**	**53.68±0.62**	**69.52±0.52**	**58.21±0.62**	**62.88**
GNN	✓	25.97±0.41	48.11±0.65	81.61±0.70	89.25±0.51	64.14±0.77	51.23±0.72	71.78±0.73	56.45±0.61	61.07
GNN-FT + Ours	✓	**26.32±0.37**	**53.41±0.58**	**90.70±0.45**	**95.53±0.32**	**70.86±0.64**	**57.51±0.72**	**77.95±0.62**	**60.78±0.65**	**66.63**
ATA-FT	✓	25.08±0.20	49.79±0.40	89.64±0.30	95.44±0.20	69.83±0.50	54.28±0.50	76.64±0.40	58.08±0.40	64.85
ATA-FT + Ours	✓	**26.04±0.36**	**53.68±0.57**	**90.21±0.43**	**95.56±0.33**	**70.09±0.68**	**54.58±0.70**	**77.70±0.64**	**58.89±0.37**	**65.84**
StyleAdv-FT	✓	26.24±0.35	53.05±0.54	91.64±0.43	96.51±0.28	70.90±0.63	56.44±0.68	79.35±0.61	64.10±0.64	67.28
StyleAdv-FT + Ours	✓	**27.07±0.38**	**54.09±0.56**	**92.37±0.38**	**96.90±0.32**	**71.27±0.64**	**58.16±0.72**	78.46±0.63	**64.81±0.64**	**67.89**
FLoR-FT	✓	26.77±0.39	56.72±0.55	83.06±0.46	92.33±0.28	73.40±0.65	57.21±0.72	72.33±0.66	61.12±0.62	65.37
FLoR-FT + Ours	✓	**27.51±0.36**	**56.80±0.54**	**83.91±0.44**	**92.40±0.28**	**74.01±0.64**	**57.27±0.73**	**73.12±0.62**	**61.29±0.62**	**65.79**

Table 2: Classification accuracy of 5-way 5-shot tasks on CD-FSL based on ResNet-10.

Experiments

Datasets

CD-FSIC. We use two widely-used benchmarks: BSCD-FSL (Guo et al. 2020) and mini-CUB (Tseng et al. 2020). Both benchmarks share the same source domain, *mini*ImageNet. BSCD-FSL includes four target domains: ChestX (Wang et al. 2017), ISIC (Tschandl, Rosendahl, and Kittler 2018), EuroSAT (Helber et al. 2019) and CropDisease (Mohanty, Hughes, and Salathé 2016). The mini-CUB benchmark consists of another four target domains: CUB (Wah et al. 2011), Cars (Krause et al. 2013), Places (Zhou et al. 2017) and Plantae (Van Horn et al. 2018).

CD-FSOD. We follow the CD-ViTO benchmark (Fu et al. 2024). The model is fine-tuned on six target domains: ArTaxOr (Drange 2019), Clipart1k (Inoue et al. 2018), DIOR (Li et al. 2020), DeepFish (Saleh et al. 2020), NEU-DET (Song and Yan 2013) and UODD (Drange 2019).

Models and Baselines

Models. For image classification, we use ResNet-10 (He et al. 2016) as the backbone and GNN as the classifier. For CD-FSOD, we adopt ViT-B/14 as the backbone, following CD-ViTO (Fu et al. 2024).

	Method	Venue	ChestX	ISIC	EuroSAT	CropDisease	CUB	Cars	Places	Plantae	Aver.
1-shot	APR-S	ICCV21	21.94±0.34	32.36±0.53	60.29±0.76	63.25±0.80	41.89±0.64	31.43±0.54	52.43±0.74	35.39±0.59	42.37
	DSU	ICLR22	22.35±0.36	31.43±0.51	64.55±0.83	64.73±0.81	**47.74±0.72**	31.61±0.53	**54.81±0.81**	37.19±0.61	44.30
	PASTA	ICCV23	21.95±0.34	32.41±0.55	64.09±0.82	66.20±0.78	39.04±0.60	30.80±0.51	50.28±0.71	35.85±0.56	42.58
	AFA	CVPR24	22.14±0.33	32.35±0.51	64.80±0.78	63.26±0.81	42.82±0.65	30.20±0.51	50.22±0.74	33.56±0.56	42.42
	EntAug	ECCV24	22.42±0.33	33.84±0.55	63.41±0.79	62.86±0.81	42.04±0.64	30.86±0.50	52.18±0.74	35.92±0.56	42.94
	VIPAug	AAAI24	22.29±0.34	34.61±0.58	63.74±0.79	63.21±0.80	44.02±0.67	32.07±0.56	52.89±0.75	35.24±0.58	43.51
	PhysAug	AAAI25	22.91±0.34	33.28±0.55	64.41±0.78	64.20±0.83	45.44±0.68	33.11±0.54	53.99±0.75	36.89±0.60	44.28
	HAP(Ours)		**23.01±0.35**	**35.68±0.60**	**66.56±0.80**	**67.05±0.78**	45.73±0.67	**33.18±0.54**	54.09±0.78	**38.15±0.62**	**45.43**
	Method	**Venue**	**ChestX**	**ISIC**	**EuroSAT**	**CropDisease**	**CUB**	**Cars**	**Places**	**Plantae**	**Aver.**
5-shot	APR-S	ICCV21	24.91±0.35	44.80±0.52	83.47±0.60	84.70±0.58	61.06±0.67	44.69±0.66	71.45±0.64	53.08±0.64	58.52
	DSU	ICLR22	25.02±0.36	45.19±0.52	80.30±0.63	86.30±0.56	**67.94±0.66**	45.65±0.63	**75.17±0.64**	54.31±0.62	59.99
	PASTA	ICCV23	24.70±0.36	45.85±0.54	83.08±0.60	87.23±0.53	59.07±0.66	47.01±0.58	68.02±0.63	51.80±0.61	58.35
	AFA	CVPR24	24.38±0.35	43.12±0.52	82.54±0.56	83.60±0.62	60.13±0.67	41.75±0.59	68.42±0.65	48.20±0.64	56.52
	EntAug	ECCV24	25.98±0.38	45.76±0.54	82.10±0.60	87.25±0.55	60.52±0.69	44.76±0.67	71.32±0.67	52.59±0.63	58.79
	VIPAug	AAAI24	25.26±0.35	43.95±0.50	80.86±0.55	84.38±0.56	59.83±0.67	42.75±0.60	69.34±0.65	50.12±0.62	57.06
	PhysAug	AAAI25	25.62±0.37	45.75±0.55	82.87±0.64	88.81±0.54	60.71±0.68	44.35±0.61	72.95±0.65	55.61±0.66	59.58
	HAP(Ours)		**26.32±0.37**	**48.26±0.57**	**84.88±0.58**	**89.49±0.45**	62.49±0.67	**47.61±0.59**	73.07±0.64	**56.54±0.63**	**61.08**

Table 3: Comparisons to different augmentation methods based on ResNet-10. The best results are marked in **bold**.

	Method	ArTaxOr	Clipart	DIOR	DeepFish	NEU-DET	UODD	Aver.
1-shot	DE-ViT-FT	12.4	10.6	9.2	3.2	1.8	1.9	6.5
	+ Ours	**13.9**	**11.4**	**10.0**	**13.7**	**2.6**	**2.1**	**9.0**
	CD-ViTO	13.6	13.6	15.7	13.2	4.0	2.4	10.4
	+ Ours	**15.5**	13.2	**15.9**	**13.3**	**4.5**	**2.5**	**10.8**
5-shot	DE-ViT-FT	35.6	32.2	19.1	20.7	6.2	5.1	19.8
	+ Ours	**36.8**	**33.5**	**20.1**	**21.7**	**6.3**	**5.2**	**20.6**
	CD-ViTO	43.1	36.3	22.8	20.2	10.7	7.4	23.4
	+ Ours	**43.7**	**36.4**	**23.0**	**20.3**	**11.1**	**7.5**	**23.7**
10-shot	DE-ViT-FT	42.8	37.4	21.8	19.5	9.5	4.9	22.7
	+ Ours	**43.9**	**37.5**	**23.0**	**20.4**	**9.8**	**5.2**	**23.3**
	CD-ViTO	54.8	40.4	27.4	19.5	13.1	6.2	26.9
	+ Ours	**55.9**	**40.8**	**27.4**	**20.2**	**14.8**	**7.0**	**27.7**

Table 4: mAP on CD-FSOD benchmark based on ViT-B/14.

Baselines. For image classification, we evaluate our method on four representative approaches: GNN (Garcia and Bruna 2018), StyleAdv (Fu et al. 2023) and FLoR (Zou et al. 2024). For object detection, we adopt two competitive baselines: DE-ViT-FT (Zhang et al. 2025) and CD-ViTO (Fu et al. 2024). Additionally, we compare our strategy against several other spatial and frequency-based methods including APR-S (Chen et al. 2021), DSU (Li et al. 2022), PASTA (Chattopadhyay et al. 2023), AFA (Vaish, Wang, and Strisciuglio 2024), EntAugment (EntAug) (Yang, Shen, and Zhao 2024), VIPAug (Lee, Lee, and Myung 2024) and PhysAug (Xu et al. 2025).

Experimental Results

HAP outperforms SOTA CD-FSIC methods. As shown in Table 1 and 2, HAP consistently improves all baselines and achieves state-of-the-art performance under both 5-way 1-shot and 5-shot settings. On average, HAP brings gains of +1.88%, +1.98%, +0.80% and +1.21% for GNN, ATA, StyleAdv and FLoR, respectively. Notably, GNN improves across *all* domains. Significant per-domain gains include ATA (+4.31% on ISIC, +4.09% on EuroSAT), StyleAdv (+4.09% on ISIC) and FLoR (+2.04% on EuroSAT, +2.02% on CUB). Similar trends hold in the 5-shot setting, with large boosts on ISIC and EuroSAT across baselines.

HAP Outperforms Existing Augmentation Methods. We compare HAP with SOTA augmentation methods. As shown in Table 3, HAP achieves the highest average accuracy in both 1-shot and 5-shot settings, outperforming all competitors. Unlike prior frequency-based methods that leave spectral gaps, HAP effectively harmonizes amplitude spectra while preserving spatial structure, validating its superiority.

HAP outperforms SOTA CD-FSOD methods. Table 4 shows that incorporating HAP consistently improves both baselines across all shot settings. On average, HAP brings highest gains of +2.5 mAP for DE-ViT-FT and +0.8 mAP for CD-ViTO, with the significant improvement observed in the challenging 1-shot scenario (+2.5 and +0.4 mAP, respectively). These results demonstrate HAP's effectiveness in enhancing generalization without modifying the backbone.

Further Analysis

Effectiveness of the Harmonized Amplitude Spectrum. Figure 5(a) presents visualizations of augmented images and their amplitude spectra. Figure 5(b) reports the effective rank. It can be observed that existing methods lead to degraded HF content and limited spectral diversity. In contrast, HAP preserves more informative MF and HF components, as reflected by the brighter spectral regions. This results in a higher effective rank, suggesting that HAP effectively mitigates spectral collapse and enriches the frequency space.

Reduction of Frequency Discrepancies. Figure 5(c) illustrates the Mean Squared Error (MSE) between the log-amplitude spectra of the source and target domains across multiple datasets. A lower MSE value implies better alignment in the frequency domain. Compared to the baseline, HAP significantly reduces the spectral discrepancies.

Analysis of Perturbation Strength Hyperparameters.

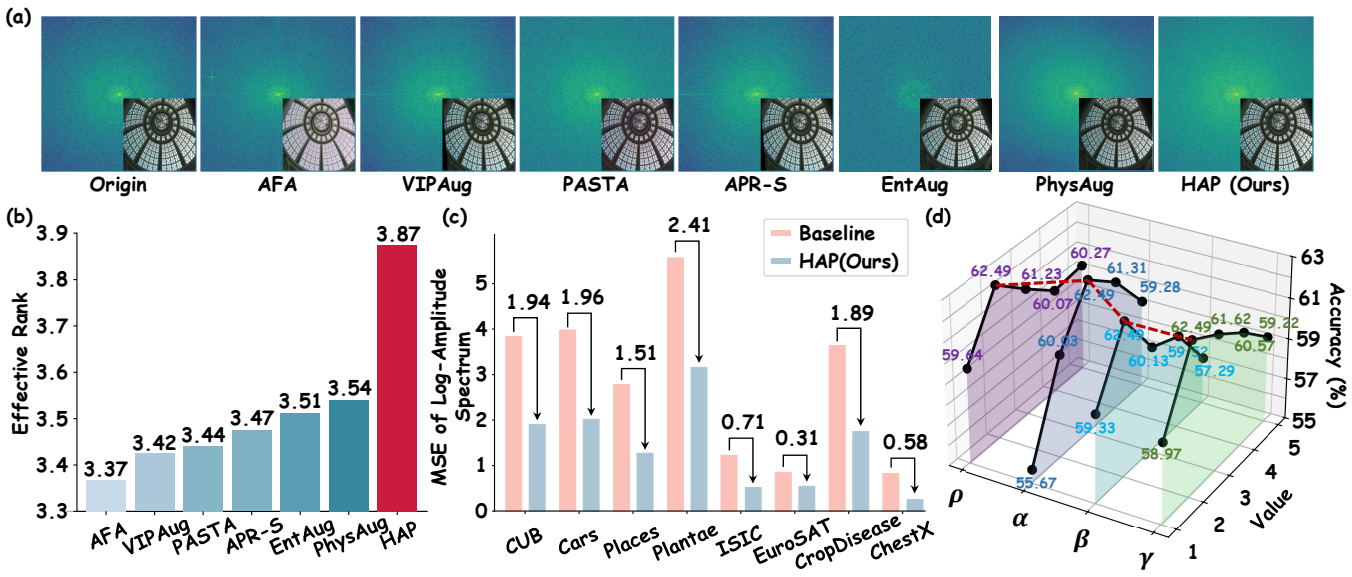

Figure 5: (a) Visualizations of augmented samples and their amplitude spectra under different methods. (b) Effective rank comparison. (c) MSE between log-amplitude spectra of source and target domains. (d) Ablation on HAP hyperparameters.

Attack ε	CUB	Cars	Places	Plantae	Aver.
5e-1	56.63±0.69	41.57±0.57	67.22±0.62	47.18±0.59	53.15
5e-2	60.43±0.70	43.68±0.62	71.48±0.64	51.60±0.62	56.80
5e-3(Ours)	62.49±0.67	**47.61±0.59**	73.07±0.64	**56.54±0.63**	**59.93**
5e-4	59.86±0.68	46.85±0.62	**73.59±0.64**	54.26±0.65	58.64
5e-5	**63.85±0.69**	45.01±0.65	72.24±0.64	51.97±0.62	58.27
5e-6	58.75±0.66	43.81±0.62	70.24±0.84	50.84±0.61	55.91

Table 5: Effects of different attack parameters ε.

Method	CUB	Cars	Places	Plantae	Aver.
5-way 1-shot					
GD	36.26±0.67	28.32±0.52	31.96±0.76	25.15±0.62	30.42
UD	41.49±0.65	30.70±0.50	51.69±0.75	34.79±0.55	39.67
HD(Ours)	**45.73±0.67**	**33.18±0.54**	**54.09±0.78**	**38.15±0.62**	**42.79**
5-way 5-shot					
GD	51.11±0.55	36.64±0.55	56.02±0.59	41.45±0.52	46.31
UD	58.67±0.65	43.42±0.63	68.46±0.62	50.37±0.63	55.23
HD(Ours)	**62.49±0.67**	**47.61±0.59**	**73.07±0.64**	**56.54±0.63**	**59.93**

Table 6: Comparison of Gaussian (GD), Uniform (UD) and Our Harmonized Distributions (HD) to model the gains.

Each curve in Figure 5(d) corresponds to varying one parameter while keeping the others fixed at their default values ($\rho = 2$, $\alpha = 3$, $\beta = 2$, and $\gamma = 2$). As the parameter value increases, accuracy first improves and then declines. This observation indicates that moderate perturbation enhances generalization by enriching spectral features, whereas excessive perturbation may distort semantic content. The results suggest that HAP benefits from carefully balanced frequency-aware perturbations.

Impact of Different Attack Parameters. We examine how varying the attack parameter ε, which controls perturbation strength, affects performance. The default setting is $\varepsilon = 5e - 3$. As shown in Table 5, larger values cause severe feature occlusion, while smaller ones introduce insufficient perturbation, both leading to degraded performance.

Effect of the Harmonized Distributions. HAP perturbs the amplitude spectrum by sampling gains from Harmonized Distributions (HD). For comparison, we also consider Gaussian Distributions (GD), denoted as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma})$, and Uniform Distributions (UD), denoted as $U(-\boldsymbol{\sigma}, \boldsymbol{\sigma})$. As shown in Table 6, GD causes notable performance degradation due to source-domain bias, while UD also underperforms HD. The proposed HD achieves the best results, validating its effectiveness in ensuring balanced, frequency-aware perturbations and improving generalization.

Conclusion

In this work, we identify frequency band discretization as a fundamental bottleneck for Cross-Domain Few-Shot Learning, driven by low-frequency dominance and limited utilization of mid- and high-frequency components. Unlike existing methods that mainly rely on spatial augmentations, we show that such strategies cannot eliminate this spectral disharmony. To address this, we propose Harmonized Amplitude Perturbation (HAP), a simple yet effective plug-and-play frequency-domain augmentation that perturbs the amplitude spectrum using a Harmonized Spatial-Frequency Grid while preserving thr phase spectrum to maintain semantic integrity. HAP substantially improves spectral diversity and model generalization without adding extra modules or training overhead. Extensive experiments on both classification and object detection benchmarks confirm its superiority, achieving consistent gains and SOTA performance across diverse Cross-Domain settings.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62476056 and 62306070) and the Social Development Science and Technology Project of Jiangsu Province (No. BE2022811). This work was also supported in part by the Southeast University Start-Up Grant for New Faculty under Grant 4009002309. Furthermore, the work was also supported by the Big Data Computing Center of Southeast University. This work was also supported by "the Fundamental Research Funds for the Central Universities(2242025K30024)".

References

Chattopadhyay, P.; Sarangmath, K.; Vijaykumar, V.; and Hoffman, J. 2023. Pasta: Proportional amplitude spectrum training augmentation for syn-to-real domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 19288–19300.

Chen, G.; Peng, P.; Ma, L.; Li, J.; Du, L.; and Tian, Y. 2021. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 458–467.

Drange, G. 2019. Arthropod taxonomy orders object detection dataset.

Feng, F.; Wang, J.; Yang, X.; and Geng, X. 2025a. Learngene: Inheritable "genes" in intelligent agents. *Artificial Intelligence*, 104421.

Feng, F.; Xie, Y.; Shi, R.; Shen, J.; Wang, J.; and Geng, X. 2025b. ECO: Evolving Core Knowledge for Efficient Transfer. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*.

Feng, F.; Xie, Y.; Wang, J.; and Geng, X. 2025c. WAVE: Weight Templates for Adaptive Initialization of Variable-sized Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4819–4828.

Feng, F.; Xie, Y.; Yang, X.; Wang, J.; and Geng, X. 2025d. Redefining <creative> in dictionary: Towards an enhanced semantic understanding of creative generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18444–18454.

Fu, Y.; Qiu, X.; Ren, B.; Fu, Y.; Timofte, R.; Sebe, N.; Yang, M.-H.; Van Gool, L.; Zhang, K.; Nong, Q.; et al. 2025. NTIRE 2025 challenge on cross-domain few-shot object detection: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1048–1069.

Fu, Y.; Wang, Y.; Pan, Y.; Huai, L.; Qiu, X.; Shangguan, Z.; Liu, T.; Fu, Y.; Van Gool, L.; and Jiang, X. 2024. Cross-domain few-shot object detection via enhanced open-set object detector. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 247–264. Springer.

Fu, Y.; Xie, Y.; Fu, Y.; Chen, J.; and Jiang, Y.-G. 2022. Wave-san: Wavelet based style augmentation network for Cross-Domain Few-Shot Learning. *arXiv preprint arXiv:2203.07656*.

Fu, Y.; Xie, Y.; Fu, Y.; and Jiang, Y.-G. 2023. Styleadv: Meta style adversarial training for Cross-Domain Few-Shot Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 24575–24584.

Garcia, V.; and Bruna, J. 2018. Few-shot learning with graph neural networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 1–12.

Guo, Y.; Codella, N. C.; Karlinsky, L.; Codella, J. V.; Smith, J. R.; Saenko, K.; Rosing, T.; and Feris, R. 2020. A broader study of Cross-Domain Few-Shot Learning. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 124–141.

He, K.; Si, C.; Lu, Z.; Huang, Y.; Wang, L.; and Wang, X. 2024. Frequency-enhanced data augmentation for vision-and-language navigation. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 36.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Helber, P.; Bischke, B.; Dengel, A.; and Borth, D. 2019. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7): 2217–2226.

Hu, C.; Zhang, C.; and Lee, G. H. 2023. Unsupervised feature representation learning for domain-generalized cross-domain image retrieval. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 11016–11025.

Hu, Y.; and Ma, A. J. 2022. Adversarial feature augmentation for cross-domain few-shot classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 20–37.

Inoue, N.; Furuta, R.; Yamasaki, T.; and Aizawa, K. 2018. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5001–5009.

Kim, J.; Kim, T.; Shim, M.; Han, D.; Wee, D.; and Kim, J. 2023. Frequency selective augmentation for video representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 37, 1124–1132.

Krause, J.; Stark, M.; Deng, J.; and Fei-Fei, L. 2013. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 554–561.

Lee, I.; Lee, W.; and Myung, H. 2024. Domain Generalization with Vital Phase Augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, 2892–2900.

Li, K.; Wan, G.; Cheng, G.; Meng, L.; and Han, J. 2020. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS journal of photogrammetry and remote sensing*, 159: 296–307.

Li, W.; Fang, P.; and Xue, H. 2025. SVasP: Self-Versatility Adversarial Style Perturbation for Cross-Domain Few-Shot Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, 15275–15283.

Li, X.; Dai, Y.; Ge, Y.; Liu, J.; Shan, Y.; and DUAN, L. 2022. Uncertainty Modeling for Out-of-Distribution Generalization. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Lin, Z.; Gao, Y.; Yang, Y.; and Sang, J. 2024. Revisiting visual model robustness: a frequency long-tailed distribution view. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 36.

Mohanty, S. P.; Hughes, D. P.; and Salathé, M. 2016. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7: 215232.

Park, G. Y.; Jung, C.; Lee, S.; Ye, J. C.; and Lee, S. W. 2024. Self-supervised debiasing using low rank regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12395–12405.

Perera, R.; and Halgamuge, S. 2024. Discriminative Sample-Guided and Parameter-Efficient Feature Space Adaptation for Cross-Domain Few-Shot Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 23794–23804.

Roy, O.; and Vetterli, M. 2007. The effective rank: A measure of effective dimensionality. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 606–610. IEEE.

Saleh, A.; Laradji, I. H.; Konovalov, D. A.; Bradley, M.; Vazquez, D.; and Sheaves, M. 2020. A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific reports*, 10(1): 14671.

Shi, R.; Feng, F.; Xie, Y.; Wang, J.; and Geng, X. 2025. FAD: Frequency Adaptation and Diversion for Cross-Domain Few-Shot Learning. *arXiv preprint arXiv:2505.08349*.

Song, K.; and Yan, Y. 2013. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science*, 285: 858–864.

Swartz, A.; Skelton, A. E.; Mather, G.; Bosten, J. M.; Maule, J.; and Franklin, A. 2024. The perceived beauty of art is not strongly calibrated to the statistical regularities of real-world scenes. *Scientific Reports*, 14(1): 19368.

Tschandl, P.; Rosendahl, C.; and Kittler, H. 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(1): 1–9.

Tseng, H.-Y.; Lee, H.-Y.; Huang, J.-B.; and Yang, M.-H. 2020. Cross-domain few-shot classification via learned feature-wise transformation. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 1–14.

Vaish, P.; Wang, S.; and Strisciuglio, N. 2024. Fourier-basis functions to bridge augmentation gap: Rethinking frequency augmentation in image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 17763–17772.

Van Horn, G.; Mac Aodha, O.; Song, Y.; Cui, Y.; Sun, C.; Shepard, A.; Adam, H.; Perona, P.; and Belongie, S. 2018. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8769–8778.

Wah, C.; Branson, S.; Welinder, P.; Perona, P.; and Belongie, S. 2011. The caltech-ucsd birds-200-2011 dataset. Technical report, California Institute of Technology.

Wang, S.; Veldhuis, R.; Brune, C.; and Strisciuglio, N. 2023. What do neural networks learn in image classification? a frequency shortcut perspective. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 1433–1442.

Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; and Summers, R. M. 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2097–2106.

Xu, X.; Yang, J.; Shi, W.; Ding, S.; Luo, L.; and Liu, J. 2025. PhysAug: A Physical-guided and Frequency-based Data Augmentation for Single-Domain Generalized Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, 21815–21823.

Xue, H.; An, Y.; Qin, Y.; Li, W.; Wu, Y.; Che, Y.; Fang, P.; and Zhang, M.-L. 2025. Towards few-shot learning in the open world: a review and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence(TPAMI)*.

Yang, S.; Shen, F.; and Zhao, J. 2024. EntAugment: Entropy-driven adaptive data augmentation framework for image classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 197–214. Springer.

Yin, D.; Gontijo Lopes, R.; Shlens, J.; Cubuk, E. D.; and Gilmer, J. 2019. A fourier perspective on model robustness in computer vision. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 32.

Zhang, X.; Liu, Y.; Wang, Y.; and Boularias, A. 2025. Detect Everything with Few Examples. In *Conference on Robot Learning*, 3986–4004. PMLR.

Zheng, H.; Wang, R.; Liu, J.; and Kanezaki, A. 2023. Cross-Level Distillation and Feature Denoising for Cross-Domain Few-Shot Classification. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 1–12.

Zhou, B.; Lapedriza, A.; Khosla, A.; Oliva, A.; and Torralba, A. 2017. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6): 1452–1464.

Zhou, F.; Wang, P.; Zhang, L.; Wei, W.; and Zhang, Y. 2023. Revisiting prototypical network for cross domain few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 20061–20070.

Zou, Y.; Liu, Y.; Hu, Y.; Li, Y.; and Li, R. 2024. Flatten Long-Range Loss Landscapes for Cross-Domain Few-Shot Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 23575–23584.