Exploring Cross-Domain Few-Shot Classification via Frequency-Aware Prompting

Tiange Zhang , **Qing Cai**[*] , **Feng Gao** , **Lin Qi** and **Junyu Dong**[*]

Faculty of Computer Science and Technology, Ocean University of China

zhangtiange@stu.ouc.edu.cn, {cq, gaofeng, qilin, dongjunyu}@ouc.edu.cn

Abstract

Cross-Domain Few-Shot Learning has witnessed great stride with the development of meta-learning. However, most existing methods pay more attention to learning domain-adaptive inductive bias (meta-knowledge) through feature-wise manipulation or task diversity improvement while neglecting the phenomenon that deep networks tend to rely more on high-frequency cues to make the classification decision, which thus degenerates the robustness of learned inductive bias since high-frequency information is vulnerable and easy to be disturbed by noisy information. Hence in this paper, we make one of the first attempts to propose a Frequency-Aware Prompting method with mutual attention for Cross-Domain Few-Shot classification, which can let networks simulate the human visual perception of selecting different frequency cues when facing new recognition tasks. Specifically, a frequency-aware prompting mechanism is first proposed, in which high-frequency components of the decomposed source image are switched either with normal distribution sampling or zeroing to get frequency-aware augment samples. Then, a mutual attention module is designed to learn generalizable inductive bias under CD-FSL settings. More importantly, the proposed method is a plug-and-play module that can be directly applied to most off-the-shelf CD-FLS methods. Experimental results on CD-FSL benchmarks demonstrate the effectiveness of our proposed method as well as robustly improve the performance of existing CD-FLS methods. Resources at https://github.com/tinkez/FAP_CDFSC.

1 Introduction

Recently, most progress on Few-Shot Learning (FSL) has followed a Meta-Learning pipeline, where the sampled classes of episodic source and target tasks are disjoint but from the same dataset, i.e., they share a similar domain. However, this underlying assumption is not applicable to real-world scenarios since the source and target domains are usually different

---

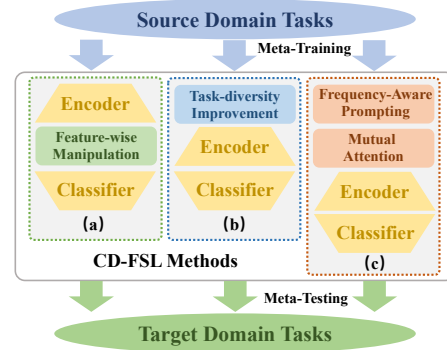[*]Corresponding authors: Qing Cai, Junyu Dong.

Figure 1: Comparison of existing CD-FSL methods with ours. (a) Feature-wise Manipulation methods have changed the backbone while (b) Task-diversity Improvement methods as well as (c) ours haven't. Thus, (c) is compatible with (a) and (b).

in realistic tasks. Besides, recent works [Chen *et al.*, 2019; Guo *et al.*, 2020] have also revealed that most existing meta-learning models suffer performance degradation when facing a domain shift between source and target tasks, and even underperform simple pre-training and fine-tuning paradigms. This intractable issue makes the new research direction, Cross-Domain Few-Shot Learning (CD-FSL) distinguishable from the conventional FSL settings.

CD-FSL aims to generalize FSL models across different domains. Here we focus on the most challenging CD-FSL setting where only one single source domain is available for training. To achieve that, [Tseng *et al.*, 2020] uses the feature-wise transformation layers to manipulate the feature activation with affine parameters to improve the generalization of metric-based inductive bias. This feature-wise transformation approach has influenced some later research on varying the feature distribution for target domain adaptation [Liang *et al.*, 2021; Zhang *et al.*, 2022; Hu and Ma, 2022], but they either have employed extra parameters of weights on backbone networks or suffered from an incompatibility with other models. Apart from the feature-wise manipulation, it is also intuitive that improving the diversity of training tasks also takes a significant role in cross-domain generalization for inductive bias since it's learned from a collection of training tasks. Therefore, [Wang and Deng, 2021] first resorted to solving a worst-case problem around the source task distribution and

proposed Adversarial Task Augmentation (ATA) to construct 'challenging' virtual tasks as a task augmentation technique. Their strategy improves the robustness of various inductive biases learned from meta-learning models and achieves good performance on the CD-FSL benchmarks.

However, when tackling the generalization problem on domain shift, most previous CD-FSL methods neglect an important factor that the frequency components of images also get involved in the robustness of inductive bias learned by existing meta-learning models. Several researchers [Wang *et al.*, 2020a; Geirhos *et al.*, 2018; Li *et al.*, 2021; Rahaman *et al.*, 2019] have already shown that the generalization behaviors of deep neural networks (especially CNNs) are sometimes counter-intuitive to human understandings (e.g., adversarial examples [Nguyen *et al.*, 2015]). One convincing explanation is that deep neural networks are more sensitive to high-frequency components of a natural image which are not easy-perceivable for human beings in recognition tasks [Wang *et al.*, 2020a; Rahaman *et al.*, 2019]. And the experimental observation in [Wang *et al.*, 2020a] also demonstrates a phenomenon that when making the final classification decision, deep-learning models tend to capture these specific high-frequency cues of a specific training domain for a higher accuracy after they learn from the low-frequency information at the early training stage. This phenomenon will guide the meta-learning models to learn a distribution-specific inductive bias towards the high-frequency cues during the meta-training stage on the source domain, and thus limit the performance of cross-domain generalization when meta-testing on a new domain dataset.

Therefore, in order to learn the more robust inductive bias that can be generalized to the cross-domain scenario for existing meta-learning methods, we make one of the first attempts to consider the CD-FSL challenge from a frequency-aware perspective. As shown in Figure 1, in this paper, we try to propose a frequency-aware prompting method with mutual attention to prompt the network either to stress the low-frequency semantic information or weaken the tendency on capturing the distribution-specific high-frequency cues, which aims to learn robust inductive bias and let the meta-learning models perceive more like the human when facing a new recognition task. Concretely, we first decompose the source images into corresponding low-frequency and high-frequency components by applying the Discrete Wavelet Transform (DWT) and constructing the frequency-aware augment samples by modifying high-frequency components. In addition, a mutual attention module is designed to enhance the interaction between different frequency-constructed features, and implicitly stress the low-frequency semantic information or suppress the high-frequency cues. Note that [Fu *et al.*, 2022b] also explores the CD-FSL problem from the frequency viewpoint, but their research focuses on varying the feature distribution on a feature-wise level, whereas our frequency-aware prompting method is used at the task level and can be orthogonal with the most off-the-shelf CD-FSL method. We also employ an adversarial training procedure to make task augmentation, which is similar to [Wang and Deng, 2021]. A KL Divergency loss is also applied to restrict the final class prediction of different frequency-reconstructed inputs from the same source image and help the network converge.

Overall, our contribution can be summarized as followed:

- We consider the most challenging single source domain generalization of the Cross-Domain Few-Shot classification from a frequency-aware perspective. To learn more robust domain-adaptive inductive bias for existing meta-learning models, we proposed a frequency-aware prompting method to simulate human perception when facing a new recognition task. The augment samples are elaborately constructed using the decomposed frequency components to prompt the meta-training process.

- A special mutual attention module is designed to facilitate the information interaction across different frequency-reconstructed features and thus promote the network to implicitly stress the related low-frequency semantic information which helps with generalization.

- Experimental results on two CD-FSL benchmarks [Guo *et al.*, 2020; Tseng *et al.*, 2020] reveal that our method can effectively improve the average classification accuracy in CD-FSL settings when compared to previous methods, and also can be used as a plug-and-play method for most existing meta-learning models.

2 Related Works

In this section, we briefly introduce the previous related works on Few-Shot Learning (FSL) and extended Cross-Domain Few-Shot Learning (CD-FSL).

2.1 Few-Shot Learning

Few-Shot Learning [Shu *et al.*, 2018; Wang *et al.*, 2020b; Lu *et al.*, 2020; Koch *et al.*, 2015] has been widely studied, especially in the Meta-Learning manner [Vilalta and Drissi, 2002; Hospedales *et al.*, 2021; Sun *et al.*, 2019; Triantafillou *et al.*, 2019; Vanschoren, 2018]. Most recent works of meta-learning methods proposed for FSL problems can be roughly divided into two types: 1) Optimization-based methods [Finn *et al.*, 2017; Nichol *et al.*, 2018; Ravi and Larochelle, 2017] aim to learn a generalizable model initialization that can be quickly adapted to target novel tasks with limited labeled samples. 2) Metric-based methods [Vinyals *et al.*, 2016; Sung *et al.*, 2018; Snell *et al.*, 2017; Satorras and Estrach, 2018; Oreshkin *et al.*, 2018] try to construct a metric function in the embedding space and perform classification on a novel task by comparing the similarity between query images and limited labeled support images through this metric. Among these two types, the metric-based methods have drawn more attention from researchers when considering the performance and simplicity. However, what is common to all of these FSL solutions is that they assume base and novel classes are from the same domain. Due to this limitation, the Cross-Domain Few-Shot Learning (CD-FSL) benchmark is proposed to focus on the domain shift problem.

2.2 Cross-Domain Few-Shot Learning

As mentioned, CD-FSL [Guo *et al.*, 2020; Tseng *et al.*, 2020; Du *et al.*, 2022; Li *et al.*, 2022; Oh *et al.*, 2022; Yuan *et al.*, 2022] has been trying to address a more challenging and
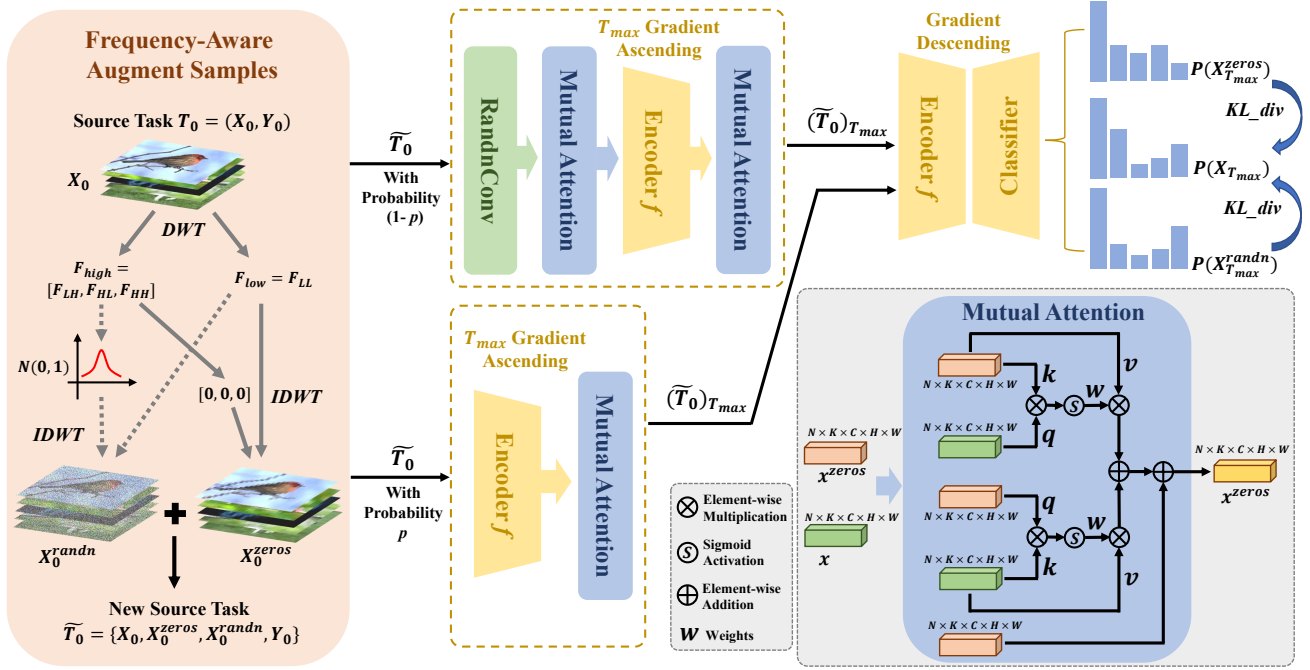
Figure 2: The overall pipeline of the proposed Frequency-Aware Prompting method. DWT represents the Discrete Wavelet Transform, $IDWT$ represents the Inverse Discrete Wavelet Transform, $N(0, 1)$ represents the Normal Distribution Sampling, $[0, 0, 0]$ represents zeroing, $P(.)$ represents the network prediction confidences, and KL_div represents the KL divergence loss. An illustration of the Mutual Attention Module is also provided. Detailed explanations of our proposed Frequency-Aware Augment Samples and Mutual Attention are described in subsections 3.3 and 3.4 respectively.

practical scenario where the source (base classes) and target (novel classes) image domains are dissimilar, as well as their labels are disjoint. This problem was preliminarily discussed in [Chen *et al.*, 2019] when conducting an analysis of existing meta-learning methods in the cross-domain setting. [Guo *et al.*, 2020] then extend the study to a Broader Study of Cross-Domain Few-Shot Learning (BSCD-FSL) benchmark which consists of four image datasets, including agriculture (CropDisease), satellite (EuroSAT), dermatology (ISIC), and radiological (ChestX) images, along with a decreasing domain similarity to ImageNet. [Tseng *et al.*, 2020] also proposed a benchmark with four datasets and use the feature-wise transformation layer to simulate various feature distributions during the training stage and thus encourage existing metric-based meta-learning models to generalize to these unseen domains. Similarly, [Liang *et al.*, 2021] then introduces a noise-enhanced supervised autoencoder (NSAE) which can teach the model to capture broader variations of feature distributions and boost the cross-domain generalize capability.

Different from focusing on diversifying feature distributions, [Wang and Deng, 2021] develops an Adversarial Task Augmentation (ATA) algorithm to construct inductive bias adaptive challenging tasks by solving a worst-case problem. This task augmentation method tries to improve the generalization capability of existing meta-learning models from the perspective that how well the inductive bias (a.k.a meta-knowledge) is learned depends on the training task diversity. It is intuitive that a wider task distribution space would cover the domain gap between source and target domains.

There are also other researches [Phoo and Hariharan, 2021; Fu *et al.*, 2021; Das *et al.*, 2022; Fu *et al.*, 2022a] that aim to adapt to the novel target domain by leveraging additional unlabelled data [Phoo and Hariharan, 2021] or a few labeled data [Fu *et al.*, 2021; Das *et al.*, 2022], whereas in this paper we consider that there is no access to any target data.

3 Methodology

3.1 Problem Formulation

For every meta-learning task T consisting of a support set T_s and a query set T_q, it is generally called a N-way K-shot situation when T_s contains N classes with K labeled samples in each class. And T_q consists of Q query samples sharing the same classes as in T_s. Then it can be defined as $T = (T_s, T_q)$, where $T_s = (x_i^s, y_i^s)_{i=1}^{N \times K}$ and $T_q = (x_j^q, y_j^q)_{j=1}^{Q}$ are sampled from a task distribution $\mathcal{T}_0$. As for cross-domain few-shot classification, due to the domain shift scenario, the meta-training dataset (e.g. mini-ImageNet [Vinyals *et al.*, 2016]) is different from the meta-testing dataset (e.g. CUB [Welinder *et al.*, 2010]). Here we only consider the setting where only one source domain dataset (mini-Imagenet) is used for training while testing on several other different datasets. But notice that every combination of support set T_s and a query set T_q for a meta-learning task is from the same domain.

3.2 Overall Mechanism

Figure 2 shows the overall pipeline of our proposed Frequency-Aware Prompting method. Here we first give

an overall explanation of our Frequency-Aware Prompting Mechanism and the training strategy. Our method takes a dual-branch framework, where the decision is made by a hyper-parameter probability p. A method for constructing frequency-aware augment samples is designed at the task level. The difference between the two branches is that a random convolution [Lee *et al.*, 2020] layer is maintained from [Wang and Deng, 2021] for feature embedding in one branch, and followed by the mutual attention module before the next step. The max iteration number T_{max} of gradient ascending serves for the adversarial training. **Other details can be viewed at the Algorithm. 1 in supplements.**

3.3 Frequency-Aware Augmentation

To prompt the meta-training with frequency information, we design a method of generating frequency-aware augment samples by reconstructing the high-frequency components. This manner includes two meanings. On the one hand, the extra samples will be reunited along with the original samples to form a new source task, thus improving the task diversity for meta-training. On the other hand, these prompt samples aim to weaken the tendency of networks that focus too much on high-frequency cues while paying less attention to low-frequency information.

To be specific, for a source task $T_0 = (X_0, Y_0)$, we conduct Discrete Wavelet Transformation (DWT) on X_0 with the 2D Haar wavelets, and get four frequency subbands denote as $F_{LL}, F_{LH}, F_{HL}, F_{HH}$. We consider the low-frequency components as $F_{low} = F_{LL}$ and keep it unchanged while performing two different operations on the high-frequency components $F_{high} = [F_{LH}, F_{HL}, F_{HH}]$ as (1) replacing the F_{high} with the same size tensor of zeros; (2) replacing the F_{high} with the same size of tensor consisting of random numbers sampled from the normal distribution. For the first operation, we can get reconstructed X_0^{zeros} after applying Inverse Discrete Wavelet Transformation (IDWT) and X_0^{randn} in the same way for the second operation. By doing so, we get a new source task $\widetilde{T_0} = \{X_0, X_0^{zeros}, X_0^{randn}, Y_0\}$ as mentioned before, where the labels Y_0 are same.

Compared to the original T_0, this new $\widetilde{T_0}$ is more diverse as a meta-training source task. Besides, the task distribution $\widetilde{\mathcal{T}_0}$ it represents is more faithfully than $\mathcal{T}_0$ since it consists of 'noisy' samples near the distribution boundary. Note that for images from the same category, these frequency-aware samples will not introduce semantic drift since we keep the low-frequency components unchanged. This can implicitly prompt the network to rely more on low-frequency information to make the classification decision and learned robust inductive bias which can perform better in the later cross-domain generalization than previous.

By using the gradient ascent process with early stopping (for T_{max} iterations) as the adversarial training based on:

$$\widetilde{T_0} = \{X_0, X_0^{zeros}, X_0^{randn}, Y_0\}, \tag{1}$$

which is sampled from the wider source task distribution $\widetilde{\mathcal{T}_0}$, we can get corresponding virtual 'challenging' tasks as:

$$(\widetilde{T_0})_{T_{max}} = \{X_{T_{max}}, X_{T_{max}}^{zeros}, X_{T_{max}}^{randn}, Y_0\}, \tag{2}$$

where $\widetilde{T}$ can represent a virtual 'challenging' task sampled from a distribution $\widetilde{\mathcal{T}}$ which is ρ distance away from $\widetilde{\mathcal{T}_0}$.

3.4 Mutual Attention Module

To facilitate the information interaction across different frequency-reconstructed features, we proposed a Mutual Attention Module for either stressing the correlated features or suppressing other unrelated information on extra generated features of x^{zeros} and x^{randn} with their original image feature x. As shown in the Figure 2, we take x^{zeros} and x as examples for better illustration. The attention module takes two same-size features as inputs, where for every attention operation there is one input used as the query and the other is the key tensor. The attention calculation is mutual to both inputs and the weights are applied to the corresponding value tensor before they are merged. A residual tensor is also added to modify the final outputs.

Figure. 2 has shown where to use this attention module. Overall, we consider the features of an original image $x \in \mathbb{R}^{N \times K \times C \times H \times W}$ as an anchor, meanwhile modifying two frequency-manipulated feature $x^{zeros} \in \mathbb{R}^{N \times K \times C \times H \times W}$ and $x^{randn} \in \mathbb{R}^{N \times K \times C \times H \times W}$ based on x. The underlying motivation is that the low-frequency features have been preserved and stressed, whereas other uncorrelated frequency features are suppressed. Therefore, the network is implicitly prompted to pay more attention to low-frequency information which is more useful for generalization across different domains. Note that the proposed Mutual Attention Modules do not change the dimension of both inputs and outputs.

3.5 Training and Inference

Network Training: Our Frequency-Aware Prompting method is a bi-level optimization process in the meta-training stage, which is similar to [Wang and Deng, 2021]. For the inner loop, it tries to find a 'challenging' input based on the new source task for the outer loop by maximizing the loss function, while the outer loop updates the model parameters (learned inductive bias) to minimize the loss function on the new challenging input. And the proposed Frequency-Aware Prompting Augmentation is employed to help improve the diversity of tasks in the inner loop stage. During the meta-training stage, the objective function is first composed of three initial Cross-Entropy (CE) FSL losses from $X_{T_{max}}$, $X_{T_{max}}^{zeros}$ and $X_{T_{max}}^{randn}$ that:

$$\mathcal{L}_0 = L((X_{T_{max}}, Y_0); \theta) = CE(X_{T_{max}}, Y_0),$$
$$\mathcal{L}_{zeros} = L((X_{T_{max}}^{zeros}, Y_0); \theta) = CE(X_{T_{max}}^{zeros}, Y_0), \tag{3}$$
$$\mathcal{L}_{randn} = L((X_{T_{max}}^{randn}, Y_0); \theta) = CE(X_{T_{max}}^{randn}, Y_0),$$

and then to restrict the network prediction and help the network to converge, we also proposed to use the KL divergence loss as the special regularization in training. Specifically, there exist two KL divergence losses defined as:

$$\mathcal{L}_{zeros}^{kl} = KLDivergence(P(X_{T_{max}}^{zeros}), P(X_{T_{max}})),$$
$$\mathcal{L}_{randn}^{kl} = KLDivergence(P(X_{T_{max}}^{randn}), P(X_{T_{max}})), \tag{4}$$

where $P(.)$ represents the distribution of the network prediction scores for classification.

Overall, the training loss can be summarized as:

$$\mathcal{L}_{train} = \mathcal{L}_0 + \mathcal{L}_{zeros} + \mathcal{L}_{randn} + \mathcal{L}_{zeros}^{kl} + \mathcal{L}_{randn}^{kl} . \quad (5)$$

Network Testing: During the meta-testing stage of the network, for each episodic task of images sampled from the target domain, we only feed the original images as inputs to the baseline meta-learning backbone with all additional processes discarded, and get their prediction. No frequency-aware augment samples as well as the mutual attention module are used. The class with the highest probability will be considered as the classification results.

Further Explanation of Methodology: Due to the limited pages, a more detailed explanation of the Methodology can be found in our supplementary materials.

4 Experiments

In this section, we evaluate the proposed Frequency-Aware Prompting method on several CD-FSL benchmark datasets and baseline methods. Following we first describe the experiments settings and the implementation details. Then we present the evaluation results and ablation studies.

4.1 Experiments Settings

Datasets: We conduct extensive experiments under the very strict cross-domain few-shot learning settings where only one single source dataset is used for meta-training, i.e., mini-ImagneNet [Vinyals *et al.*, 2016]. There are two CD-FSL benchmarks used in the meta-testing phase for evaluation. One is from the FWT benchmark [Tseng *et al.*, 2020] which includes CUB [Welinder *et al.*, 2010], Cars [Krause *et al.*, 2013], Places [Zhou *et al.*, 2017], and Plantae [Van Horn *et al.*, 2018]. The other one is the BSCD-FSL benchmark [Guo *et al.*, 2020] which consists of ChestX, ISIC, EuroSAT, and CropDisease. For all experiments, we select the model with the best validation accuracy on mini-ImageNet for the next testing on eight target datasets.

Implementation Details: We conduct our experiments under the open-source CD-FSL framework provided by [Wang and Deng, 2021]. For a fair comparison, ResNet-10 is selected as the feature extractor and the optimizer is Adam with a fixed learning rate $\alpha = 0.001$. The iteration number for early stopping is kept as $T_{max} = 5$ in all experiments and the gradient ascending step takes a learning rate β from $\{20, 40, 60, 80\}$. We maintain the Random Convolution layer applied in [Wang and Deng, 2021] with a probability $(1 - p)$, where p is selected from $\{0.5, 0.6, 0.7\}$. The filter size k is randomly sampled from the candidate pool $\mathcal{K} = \{1, 3, 5, 7, 11, 15\}$ and the Xavier normal distribution [Glorot and Bengio, 2010] is used to initialize the layer weights. Both stride and padding sizes are determined to keep the image size unchanged. Note that we follow the strategy of ATA which uses a pre-trained feature extractor by minimizing the standard cross-entropy classification loss on the 64 training classes in the mini-ImageNet dataset. Evaluation results are obtained in the 5-way 1-shot/5-shot settings with 2000 randomly sampled episodic tasks with 16 query samples per class, and report the average accuracy (%) with a 95% confidence interval. All experiments reported are conducted on the RTX 2080Ti GPUs.

4.2 Evaluation Results

Here in this subsection, we present the evaluation results with CD-FSL benchmarks and the baselines.

Comparison with State-of-the-arts: We follow the settings in [Wang and Deng, 2021] to make a fair comparison on eight datasets with several stat-of-the-arts CD-FSL methods. Specifically, we integrate the Frequency-Aware Prompting (FAP) into three meta-learning baselines: GNN [Satorras and Estrach, 2018], RelationNet [Sung *et al.*, 2018], and TPN [Ma *et al.*, 2020]. Then compare the performance improvements on the baselines using ours with FWT [Tseng *et al.*, 2020], LRP [Sun *et al.*, 2021], and ATA [Wang and Deng, 2021]. Note that LRP can only be applied to metric-based meta-leaning models such as GNN and RelationNet. All the methods presented select ResNet-10 as the feature extractor. Quantitative results of 1-shot and 5-shot classification average accuracies as well as their 95% confidence interval are shown in Table 2, and the parameters of ours for comparison are shown in Table 1. As presented, our method achieves considerable accuracy gaining on three baselines, and in most cases achieves the best or second-best performance over benchmark datasets. Also, we observe that our method can provide the most impressive results on the GNN baseline, which means that GNN is more robust when applied with our frequency-aware method under the CD-FSL settings. Furthermore, there are also some performance limitations on the target dataset such as the satellite images (EuroSAT).

baseline	setting	lr β	T_{max}	p
GNN	1-shot	40	5	0.5
GNN	5-shot	40	5	0.5
RelationNet	1-shot	80	5	0.5
RelationNet	5-shot	80	5	0.5
TPN	1-shot	20	5	0.5
TPN	5-shot	40	5	0.5

Table 1: The parameters of our method in the Table 2 presentation.

Evaluation on Robustness and Frequency Perception: In order to verify the robustness of the inductive bias learned from our Frequency-Aware Prompting method, we also provide the evaluation results of meta-testing with different frequency-reconstructed images in Table 3. To be specific, instead of meta-testing on tasks sampled from the eight novel target datasets, we turn to change the frequency cues of original meta-testing samples into new frequency-reconstructed tasks. For a target domain novel task $T_n = (X_n, Y_n)$, we design three types of new target domain task as (1) replacing the F_{high} with zeros and get $T_n^{zeros} = (X_n^{zeros}, Y_n)$; (2) replacing the F_{high} with random numbers sampled from the normal distribution and get $T_n^{randn} = (X_n^{randn}, Y_n)$; (3) directly add the Gaussian noise to images and get the $T_n^{noise} = (X_n^{noise}, Y_n)$. The purpose is to test the robustness of the learned inductive bias when facing a new task with frequency-disturbed cues or noise under the cross-domain settings and its tendency on frequency perception compared to the original meta-testing.

From the results, we learn that the low-frequency components take the most cues for semantic information since us-

5-shot	ChestX	ISIC	EuroSAT	CropDisease	Cars	CUB	Places	Plantae
GNN (*baseline*)	23.87±0.2	42.54±0.4	78.69±0.4	83.12±0.4	43.70±0.4	62.87±0.5	70.91±0.5	48.51±0.4
+FWT	24.28±0.2	40.87±0.4	78.02±0.4	87.07±0.4	46.19±0.4	64.97±0.5	70.70±0.5	49.66±0.4
+LRP	**24.53±0.3**	44.14±0.4	77.14±0.4	86.15±0.4	46.07±0.4	62.86±0.5	71.38±0.5	50.31±0.4
+ATA	24.32±0.4	**44.91±0.4**	**83.75±0.4**	**90.59±0.3**	**49.14±0.4**	**66.22±0.5**	**75.48±0.4**	**52.69±0.4**
+FAP (*ours*)	**25.31±0.2**	**47.60±0.4**	82.52±0.4	**91.79±0.3**	**50.20±0.4**	**67.66±0.5**	74.98±0.4	**54.54±0.4**
RelationNet (*baseline*)	24.07±0.2	38.60±0.3	65.56±0.4	72.86±0.4	40.46±0.4	56.77±0.4	64.25±0.4	42.71±0.3
+FWT	23.95±0.2	38.68±0.3	69.13±0.4	75.78±0.4	40.18±0.4	**59.77±0.4**	65.55±0.4	44.29±0.3
+LRP	**24.28±0.2**	**39.97±0.3**	67.54±0.4	74.21±0.4	41.21±0.4	57.70±0.4	65.35±0.4	43.70±0.3
+ATA	**24.43±0.2**	**40.38±0.3**	**71.02±0.4**	**78.20±0.4**	**42.95±0.4**	**59.36±0.4**	**66.90±0.4**	**45.32±0.4**
+FAP (*ours*)	23.71±0.2	39.14±0.4	69.66±0.3	78.23±0.4	42.86±0.4	57.78±0.4	65.97±0.4	45.19±0.3
TPN (*baseline*)	22.17±0.2	45.66±0.3	77.22±0.4	81.91±0.5	44.54±0.4	63.52±0.4	71.39±0.4	50.96±0.4
+FWT	21.22±0.1	36.96±0.4	65.69±0.5	70.06±0.7	34.03±0.4	58.18±0.5	66.75±0.5	43.20±0.5
+ATA	**23.60±0.2**	**45.83±0.3**	79.47±0.3	88.15±0.5	**46.95±0.4**	**65.31±0.4**	**72.12±0.4**	**55.08±0.4**
+FAP (*ours*)	**24.15±0.2**	44.58±0.3	**80.24±0.3**	**88.34±0.3**	**47.38±0.4**	64.17±0.4	72.05±0.4	53.58±0.4
1-shot	ChestX	ISIC	EuroSAT	CropDisease	Cars	CUB	Places	Plantae
GNN (*baseline*)	21.94±0.2	30.14±0.3	54.61±0.5	59.19±0.5	31.72±0.4	44.40±0.5	52.42±0.5	33.60±0.4
+FWT	22.00±0.2	30.22±0.3	55.53±0.5	60.74±0.5	32.25±0.4	**45.50±0.5**	53.44±0.5	32.56±0.4
+LRP	**22.11±0.2**	30.94±0.3	54.99±0.4	59.23±0.4	31.46±0.4	43.89±0.5	52.28±0.5	33.20±0.4
+ATA	**22.10±0.2**	**33.21±0.4**	**61.35±0.5**	**67.47±0.5**	**33.61±0.4**	45.00±0.5	**53.57±0.4**	**34.42±0.4**
+FAP (*ours*)	**22.36±0.2**	**35.63±0.4**	**62.96±0.5**	**69.97±0.5**	**34.69±0.4**	**46.91±0.5**	**54.41±0.5**	**36.73±0.4**
RelationNet (*baseline*)	21.95±0.2	30.53±0.3	49.08±0.4	53.58±0.4	30.09±0.3	41.27±0.4	48.16±0.5	31.23±0.3
+FWT	21.79±0.2	30.38±0.3	53.53±0.4	57.57±0.5	30.45±0.3	**43.33±0.4**	**49.92±0.5**	32.57±0.3
+LRP	**22.11±0.2**	**31.16±0.3**	50.99±0.4	55.01±0.4	30.48±0.3	41.57±0.4	48.47±0.5	32.11±0.3
+ATA	**22.14±0.2**	**31.13±0.3**	**55.69±0.5**	**61.17±0.5**	**31.79±0.3**	**43.02±0.4**	**51.16±0.5**	**33.72±0.3**
+FAP (*ours*)	21.56±0.2	30.55±0.3	**53.72±0.4**	**57.97±0.5**	**31.10±0.3**	41.76±0.4	49.61±0.5	**32.98±0.3**
TPN (*baseline*)	21.05±0.2	**35.08±0.4**	63.90±0.5	68.39±0.6	32.42±0.4	48.03±0.4	56.17±0.5	**37.40±0.4**
+FWT	20.46±0.1	29.62±0.3	52.68±0.6	56.06±0.7	26.50±0.3	44.24±0.5	52.45±0.5	32.46±0.4
+ATA	**21.67±0.2**	**34.70±0.4**	**65.94±0.5**	**77.82±0.5**	**34.18±0.4**	**50.26±0.5**	**57.03±0.5**	**39.83±0.4**
+FAP (*ours*)	21.56±0.2	33.63±0.4	62.62±0.5	76.11±0.5	34.39±0.4	**50.56±0.5**	**57.34±0.5**	37.44±0.4

Table 2: Quantitative results with state-of-the-arts on CD-FSL benchmark datasets. GNN, RelationNet, and TPN are three baselines. The results of other methods are borrowed from [Wang and Deng, 2021]. The training parameters of our results on each baseline presented here are demonstrated in the previous subsection for clearness. The **Bold** for the best and the second-best results.

ing T_n^{zeros} for meta-testing can already get a considerable classification accuracy in CD-FSL benchmarks compared to using original tasks T_n. Besides, the disturbance performed on high-frequency components actually influences the final classification results, which leads to the performance degradation. But our learned inductive bias can be more robust to these disturbances, suggesting that our frequency-aware mechanism prompts the model to concentrate on the most important low-frequency cues. And the testing results presented by using T_n^{noise} also demonstrate the robustness of inductive bias learned by our method in CD-FSL settings.

4.3 Ablation Studies

Effect of Gradient Ascend Learning Rate β: In all the above experiments, we keep the max iteration number for early stopping as $T_{max} = 5$ and the probability of dual-branch decision as $p = 0.5$ unchanged. Thus, the most important hyper-parameter which influenced the training process is the learning rate β. Here in this ablation study, we try to explore the effect of the learning rate β used in the gradient ascend process, which aims to construct a virtual 'challenging' task $(\widetilde{T_0})_{T_{max}}$ based on the frequency-aware augmented task $\widetilde{T_0}$ for gradient descending optimization.

Experiments are conducted on both 5-way 5-shot and 5-way 1-shot settings with the learning rate β selected from
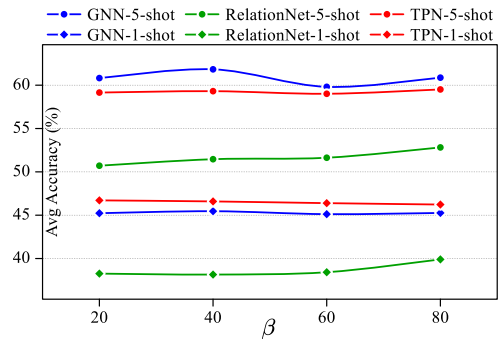

Figure 3: Average accuracy on 5-way 5-shot/1-shot tasks over eight CD-FSL datasets on three baselines with the varying learning rate β of gradient ascending process used in the adversarial training. Other parameters are fixed as $T_{max} = 5$ and $p = 0.5$ for all stages.

the set of $\{20, 40, 60, 80\}$ as mentioned before. Other meta-training parameters are fixed as $T_{max} = 5$ and $p = 0.5$ following the previous experiments. Average classification results over eight cross-domain few-shot learning datasets are reported in Figure 3. From it we can see that different baselines prefer various β. GNN and TPN perform better on a small β while the RelationNet needs a larger one.

Effect of Probability p: The average accuracy on eight

5-shot	ChestX	ISIC	EuroSAT	CropDisease	Cars	CUB	Places	Plantae
GNN + FAP (T_n)	25.31±0.2	47.60±0.4	82.52±0.4	91.79±0.3	50.20±0.4	67.66±0.5	74.98±0.4	54.54±0.4
GNN + FAP (T_n^{zeros})	25.14±0.2	48.02±0.4	82.32±0.4	91.06±0.3	49.77±0.4	67.67±0.5	74.97±0.4	53.74±0.4
GNN + FAP (T_n^{randn})	24.36±0.2	46.84±0.4	81.63±0.4	90.47±0.3	50.27±0.4	65.83±0.5	74.25±0.4	53.54±0.4
GNN + FAP (T_n^{noise})	23.18±0.2	42.11±0.3	71.43±0.4	79.84±0.5	47.11±0.4	54.96±0.5	62.00±0.5	48.60±0.4
GNN + ATA (T_n)	24.38±0.2	46.31±0.4	83.86±0.4	90.16±0.3	48.99±0.4	64.65±0.5	74.47±0.4	52.28±0.4
GNN + ATA (T_n^{zeros})	24.97±0.2	46.85±0.4	83.33±0.4	90.34±0.3	49.63±0.4	65.10±0.5	73.68±0.5	51.59±0.4
GNN + ATA (T_n^{randn})	22.82±0.2	38.00±0.3	72.94±0.4	80.60±0.5	45.02±0.4	47.63±0.4	56.82±0.5	43.26±0.4
GNN + ATA (T_n^{noise})	22.97±0.2	37.23±0.3	71.42±0.5	75.95±0.5	43.50±0.4	44.46±0.4	54.74±0.4	40.90±0.4

Table 3: Robustness evaluation results on different reconstructed meta-testing tasks. T_n represents the original target domain task, T_n^{zeros}, T_n^{randn}, and T_n^{noise} are new target domain meta-testing tasks that described in previous for details. Both two methods are meta-trained with the same parameters as gradient ascending learning rate β of 40, T_{max} of 5, and the p with 0.5.

datasets of our methods is shown in Table 4. In order to make a fair comparison through different baselines (GNN, RealtionNet, TPN), we set hyper-parameter p as 0.5 unchanged. However, as p grows, the training consumption is increasing according to Algorithm 1 in Supplementary Material. Therefore, keeping p as 0.5 is the better choice to get a performance and effectiveness trade-off in our experiments.

p	baseline	lr β	T_{max}	5-shot acc	1-shot acc
0.5	GNN	40	5	61.82	45.46
0.6	GNN	40	5	60.83	45.90
0.7	GNN	40	5	61.36	46.15

Table 4: Average accuracy variation over eight CD-FSL datasets with the changing probability p.

Effect of Mutual Attention Module: We also explore the improvement provided by the Mutual Attention Module designed in our method. Ablation experiments are conducted on both 5-way 5-shot and 5-way 1-shot settings with fixed meta-training parameters as $\beta = 40$, $T_{max} = 5$, and $p = 0.5$. All baselines and the ablative models are trained in the same manner for a fair comparison. Figure 4 shows the average classification results over eight cross-domain few-shot learning datasets with the settings of using baseline model, using frequency-aware prompting method without mutual attention modules, and using our complete method respectively. From the figure we can learn that either with or without the mutual attention modules, the average classification results of the frequency-aware prompting method on eight datasets still achieve higher accuracy than original baseline models.

T-SNE Visualization: Figure 5 provides the t-SNE visualization of baseline model GNN and GNN+FAP (ours) on both mini-ImageNet (source) and CropDisease (target) datasets. Concretely, the representations encoded by the feature extractor are projected into a 2D space. For each dataset, 5 categories of 100 samples are randomly sampled, and they are fed into two models samely. Different colors indicate different categories. From Figure 5, we can learn that compared to GNN, the distribution of each category is indeed more dispersed in our method as expected, hence making it easier to recognize the target categories.

Computational Burden Explanation: Generally, the proposed method mainly focuses on frequency-aware augmentation when compared with other methods. Therefore, it didn't introduce too much computational burden.
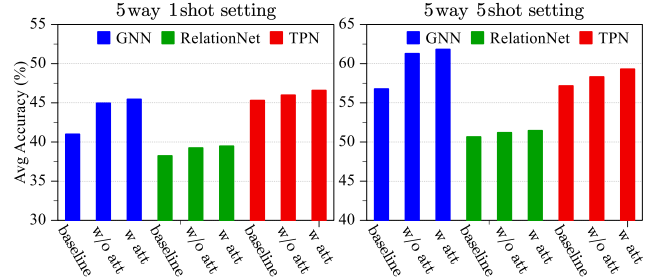

Figure 4: Average accuracy on 5-way 5-shot/1-shot tasks over eight CD-FSL datasets on three baselines. It respectively shows the results of the original baseline, without mutual attention modules (w/o att) and our complete Frequency-Aware Prompting method (w att).
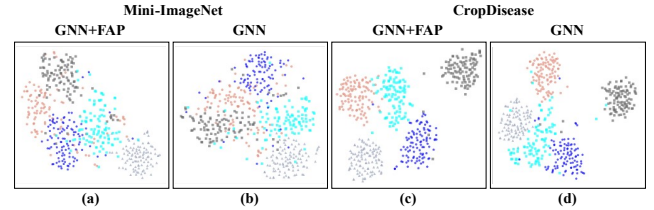

Figure 5: The t-SNE visualization. (a), (b) denote mini-ImageNet while (c), (d) represent CropDisease dataset.

Supplements: The full paper with supplementary materials can be found in the resources link for further explanation.

5 Conclusion

In this paper, we try to explore the cross-domain few-shot learning problem from a frequency-aware perspective. For this purpose, we propose a frequency-aware prompting method with mutual attention to simulating the human perception of selecting frequency cues when facing a new task. Concretely, the generated frequency-aware augment samples can improve the diversity of the meta-training episodic task and prompt the meta-training process during training. The mutual attention modules can facilitate the information interaction across different frequency-reconstructed features to either stress or suppress the corresponding information. Experimental results on the benchmarks show the effectiveness of our method on CD-FSL settings and the robustness of inductive bias learned from existing meta-learning baselines.

Acknowledgments

References

[Chen *et al.*, 2019] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Wang, and Jia-Bin Huang. A closer look at few-shot classification. In *International Conference on Learning Representations*, 2019.

[Das *et al.*, 2022] Debasmit Das, Sungrack Yun, and Fatih Porikli. Confess: A framework for single source cross-domain few-shot learning. In *International Conference on Learning Representations*, 2022.

[Du *et al.*, 2022] Yingjun Du, Xiantong Zhen, Ling Shao, and Cees G. M. Snoek. Hierarchical variational memory for few-shot learning across domains. In *International Conference on Learning Representations*, 2022.

[Finn *et al.*, 2017] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135, 2017.

[Fu *et al.*, 2021] Yuqian Fu, Yanwei Fu, and Yu-Gang Jiang. Meta-fdmixup: Cross-domain few-shot learning guided by labeled target data. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 5326–5334, 2021.

[Fu *et al.*, 2022a] Yuqian Fu, Yanwei Fu, Jingjing Chen, and Yu-Gang Jiang. Generalized meta-fdmixup: Cross-domain few-shot learning guided by labeled target data. *IEEE Transactions on Image Processing*, 31:7078–7090, 2022.

[Fu *et al.*, 2022b] Yuqian Fu, Yu Xie, Yanwei Fu, Jingjing Chen, and Yu-Gang Jiang. Wave-san: Wavelet based style augmentation network for cross-domain few-shot learning. *arXiv preprint arXiv:2203.07656*, 2022.

[Geirhos *et al.*, 2018] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2018.

[Glorot and Bengio, 2010] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feed-forward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.

[Guo *et al.*, 2020] Yunhui Guo, Noel C Codella, Leonid Karlinsky, James V Codella, John R Smith, Kate Saenko, Tajana Rosing, and Rogerio Feris. A broader study of cross-domain few-shot learning. In *European Conference on Computer Vision*, 2020.

[Hospedales *et al.*, 2021] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5149–5169, 2021.

[Hu and Ma, 2022] Yanxu Hu and Andy J Ma. Adversarial feature augmentation for cross-domain few-shot classification. In *European Conference on Computer Vision*, 2022.

[Koch *et al.*, 2015] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *International Conference on Machine Learning, Deep Learning Workshop*, 2015.

[Krause *et al.*, 2013] Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 554–561, 2013.

[Lee *et al.*, 2020] Kimin Lee, Kibok Lee, Jinwoo Shin, and Honglak Lee. Network randomization: A simple technique for generalization in deep reinforcement learning. In *International Conference on Learning Representations*, 2020.

[Li *et al.*, 2021] Yingwei Li, Qihang Yu, Mingxing Tan, Jieru Mei, Peng Tang, Wei Shen, Alan Yuille, and cihang xie. Shape-texture debiased neural network training. In *International Conference on Learning Representations*, 2021.

[Li *et al.*, 2022] Wei-Hong Li, Xialei Liu, and Hakan Bilen. Cross-domain few-shot learning with task-specific adapters. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7161–7170, 2022.

[Liang *et al.*, 2021] Hanwen Liang, Qiong Zhang, Peng Dai, and Juwei Lu. Boosting the generalization capability in cross-domain few-shot learning via noise-enhanced supervised autoencoder. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9424–9434, 2021.

[Lu *et al.*, 2020] Jiang Lu, Pinghua Gong, Jieping Ye, and Changshui Zhang. Learning from very few samples: A survey. *arXiv preprint arXiv:2009.02653*, 2020.

[Ma *et al.*, 2020] Yuqing Ma, Shihao Bai, Shan An, Wei Liu, Aishan Liu, Xiantong Zhen, and Xianglong Liu. Transductive relation-propagation network for few-shot learning. In *International Joint Conference on Artificial Intelligence*, volume 20, pages 804–810, 2020.

[Nguyen *et al.*, 2015] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 427–436, 2015.

[Nichol *et al.*, 2018] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.

[Oh *et al.*, 2022] Jaehoon Oh, Sungnyun Kim, Namgyu Ho, Jin-Hwa Kim, Hwanjun Song, and Se-Young Yun. Understanding cross-domain few-shot learning based on domain similarity and few-shot difficulty. In *Advances in Neural Information Processing Systems*, 2022.

[Oreshkin *et al.*, 2018] Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. *Advances in Neural Information Processing Systems*, 31, 2018.

[Phoo and Hariharan, 2021] Cheng Perng Phoo and Bharath Hariharan. Self-training for few-shot transfer across extreme task differences. In *International Conference on Learning Representations*, 2021.

[Rahaman *et al.*, 2019] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pages 5301–5310, 2019.

[Ravi and Larochelle, 2017] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International Conference on Learning Representations*, 2017.

[Satorras and Estrach, 2018] Victor Garcia Satorras and Joan Bruna Estrach. Few-shot learning with graph neural networks. In *International Conference on Learning Representations*, 2018.

[Shu *et al.*, 2018] Jun Shu, Zongben Xu, and Deyu Meng. Small sample learning in big data era. *arXiv preprint arXiv:1808.04572*, 2018.

[Snell *et al.*, 2017] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, 30, 2017.

[Sun *et al.*, 2019] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2019.

[Sun *et al.*, 2021] Jiamei Sun, Sebastian Lapuschkin, Wojciech Samek, Yunqing Zhao, Ngai-Man Cheung, and Alexander Binder. Explanation-guided training for cross-domain few-shot classification. In *2020 25th International Conference on Pattern Recognition*, pages 7609–7616, 2021.

[Sung *et al.*, 2018] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018.

[Triantafillou *et al.*, 2019] Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples. *arXiv preprint arXiv:1903.03096*, 2019.

[Tseng *et al.*, 2020] Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang. Cross-domain few-shot classification via learned feature-wise transformation. *International Conference on Learning Representations*, 2020.

[Van Horn *et al.*, 2018] Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8769–8778, 2018.

[Vanschoren, 2018] Joaquin Vanschoren. Meta-learning: A survey. *arXiv preprint arXiv:1810.03548*, 2018.

[Vilalta and Drissi, 2002] Ricardo Vilalta and Youssef Drissi. A perspective view and survey of meta-learning. *Artificial Intelligence Review*, 18:77–95, 2002.

[Vinyals *et al.*, 2016] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in Neural Information Processing Systems*, 29, 2016.

[Wang and Deng, 2021] Haoqing Wang and Zhi-Hong Deng. Cross-domain few-shot classification via adversarial task augmentation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 1075–1081, 8 2021.

[Wang *et al.*, 2020a] Haohan Wang, Xindi Wu, Zeyi Huang, and Eric P Xing. High-frequency component helps explain the generalization of convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8684–8694, 2020.

[Wang *et al.*, 2020b] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys*, 53(3):1–34, 2020.

[Welinder *et al.*, 2010] Peter Welinder, Steve Branson, Takeshi Mita, Catherine Wah, Florian Schroff, Serge Belongie, and Pietro Perona. Caltech-ucsd birds 200. 2010.

[Yuan *et al.*, 2022] Wang Yuan, Zhizhong Zhang, Cong Wang, Haichuan Song, Yuan Xie, and Lizhuang Ma. Task-level self-supervision for cross-domain few-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3215–3223, 2022.

[Zhang *et al.*, 2022] Ji Zhang, Jingkuan Song, Lianli Gao, and Hengtao Shen. Free-lunch for cross-domain few-shot learning: style-aware episodic training with robust contrastive learning. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2586–2594, 2022.

[Zhou *et al.*, 2017] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1452–1464, 2017.