

Article

Revisiting the Transferability of Few-Shot Image Classification: A Frequency Spectrum Perspective

Min Zhang ^{1,2}, Zhitao Wang ² and Donglin Wang ^{2,*}

¹ College of Computer Science & Technology, Zhejiang University, Hangzhou 310027, China; zhangmin@westlake.edu.cn

² School of Engineer, Westlake University, Hangzhou 310030, China; wangzhitao@westlake.edu.cn

* Correspondence: wangdonglin@westlake.edu.cn

Abstract: Few-shot learning, especially few-shot image classification (FSIC), endeavors to recognize new categories using only a handful of labeled images by transferring knowledge from a model trained on base categories. Despite numerous efforts to address the challenge of deficient transferability caused by the distribution shift between the base and new classes, the fundamental principles remain a subject of debate. In this paper, we elucidate **why** a decline in performance occurs and **what** information is transferred during the testing phase, examining it from a frequency spectrum perspective. Specifically, we adopt causality on the frequency space for FSIC. With our causal assumption, non-causal frequencies (e.g., background knowledge) act as confounders between causal frequencies (e.g., object information) and predictions. Our experimental results reveal that different frequency components represent distinct semantics, and non-causal frequencies adversely affect transferability, resulting in suboptimal performance. Subsequently, we suggest a straightforward but potent approach, namely the *Frequency Spectrum Mask* (FRSM), to weight the frequency and mitigate the impact of non-causal frequencies. Extensive experiments demonstrate that the proposed FRSM method significantly enhanced the transferability of the FSIC model across nine testing datasets.

Keywords: few-shot learning; few-shot image classification; frequency spectrum; distribution shift; causal perspective



Citation: Zhang, M.; Wang, Z.; Wang, D. Revisiting the Transferability of Few-Shot Image Classification: A Frequency Spectrum Perspective. *Entropy* **2024**, *26*, 473. <https://doi.org/10.3390/e26060473>

Academic Editor: Frank Nielsen

Received: 22 April 2024

Revised: 27 May 2024

Accepted: 29 May 2024

Published: 29 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Few-shot image classification (FSIC) endeavors to identify unlabeled images within the query set belonging to novel classes, leveraging knowledge acquired from base classes with a few labeled images in the support set [1,2]. Recently, numerous methodologies have emerged to address the challenge of few-shot image classification, which are mainly divided as follows. (1) **Fine-tuning-based methods** [3–5] tackle the problem by *learning to transfer*. They follow the standard machine learning or transfer learning [6–8] procedure to pretrain transferable knowledge and test-tune the knowledge in FSIC episodes sampled from novel classes. (2) **Metric-based methods** [9,10] solve the problem by *learning to compare*. They calculate the similarity between the query with the unlabeled data and the support set with labeled images. Finally, (3) **meta-based methods** [11–13] address the problem by *learning to learn*. They learn a good model initialization for fast adaptation to novel classes.

When distribution shifts exist between the base (or training dataset) and novel classes (or testing dataset), it is common for the transferability of the model trained on the base classes to diminish. In Figure 1, we scrutinize the reasons behind performance declines resulting from distribution shifts by employing T-SNE [14] visualization. This analysis is conducted using Resnet12 trained on the *miniImageNet*-train dataset. In Figure 1a–d, we can observe that Figure 1a demonstrates superior class clustering performance. This can be attributed to the training model using a training dataset that closely resembled the testing dataset during the training phase.

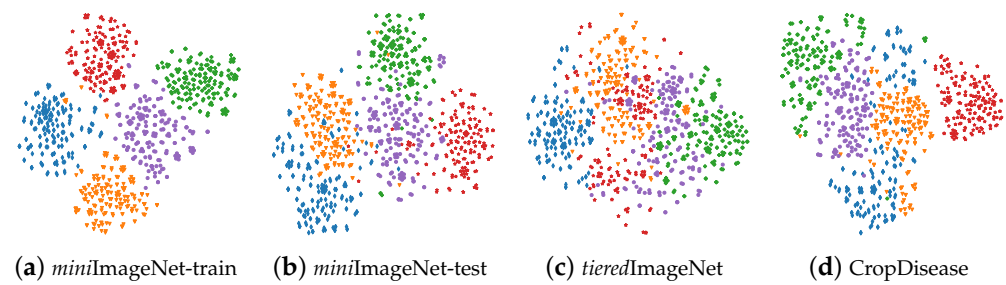


Figure 1. T-SNE [14] visualization of four datasets utilizing a model pretrained on the training set of *miniImageNet* (i.e., *miniImageNet-train*). Here, *miniImageNet-test* means the testing data of *miniImageNet*. From left to right, the performance of the class cluster declines, which can be attributed to the impoverished similarity between the training and testing datasets.

Recently, the majority of works have been dedicated to addressing the issue of distribution shift [15–17]. However, these works predominantly concentrated on **enhancing transferability** by introducing diverse regularizations or larger model architectures, often leading to increased memory and time consumption in exchange for marginal performance improvements [18,19]. The primary challenge of these works stems from a lack of clarity regarding **why** FSIC performance diminishes and **what** information undergoes transferring during the distribution shifts. In this paper, we seek to elucidate the underlying mechanism behind the remarkable transferability observed in the FSIC problem. Therefore, we approached the problem by constructing a causal graph of FSIC from a frequency spectrum perspective, motivated by both theoretical and experimental evidence suggesting that the frequency space contains more distinguishable semantic information than the feature space, as established in Figure 2 and prior studies [20,21]. Figure 2 showcases the average amplitudes of the eigenfrequencies across four testing datasets using a pretrained model, vividly illustrating pronounced distinctions in the frequency space among testing datasets.

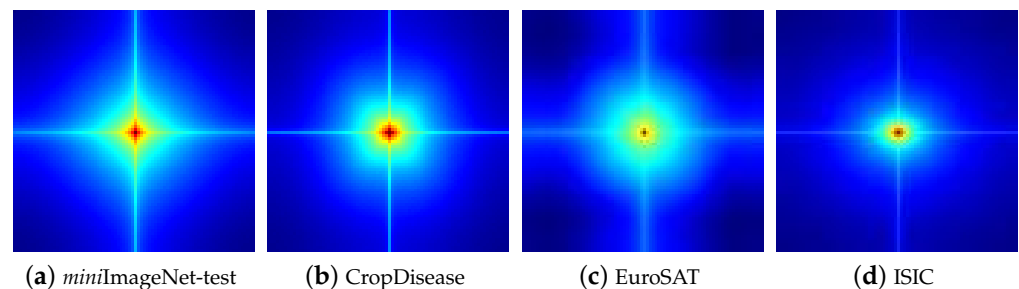


Figure 2. The average amplitudes of the eigenfrequencies across four datasets, applying a model pretrained on the training set of *miniImageNet* (i.e., *miniImageNet-train*). Here, *miniImageNet-test* means the testing data of *miniImageNet*. Proximity to the center signifies a low-frequency component, while being far from the center indicates a high-frequency component.

In this paper, we construct a causal graph of FSIC within the frequency space which delineates the relationships among the causal frequency (e.g., object information), non-causal frequency (e.g., background knowledge), and prediction. As illustrated in Figure 3a, our causal assumption posits the non-causal frequency as a confounder between the causal frequency and prediction [22]. For example, consider a scenario where the dog images in the training data feature grass backgrounds. This scenario poses a challenge during the testing phase when encountering the dog image with a water background, as the inconsistent background information hampers recognition. Consequently, the presence of **confounding information stands as a primary factor** influencing the performance of FSIC under distribution shifts. Notably, the **causal frequency, representing domain-invariant knowledge, should be effectively transferred** from training to testing datasets.

We introduce a straightforward yet potent method called the Frequency Spectrum Mask (FRSM) to weight the influences of causal frequency components. Through empirical validation, we benchmark the FRSM against state-of-the-art (SOTA) methods across nine FSIC datasets, demonstrating its efficacy. Our primary contributions can be outlined as follows:

- We clarify **why** there is a decline in performance and **what** information is transferred during the distribution shift in the few-shot image classification task from a frequency spectrum perspective.
- We adopt a causal perspective of few-shot image classification to demonstrate that non-causal frequencies impact transferability and introduce a straightforward yet efficient method, the FRSM, to weight frequencies.
- The experimental results indicate that the proposed FRSM method achieves superior performance compared with representative state-of-the-art methods in the few-shot image classification task.

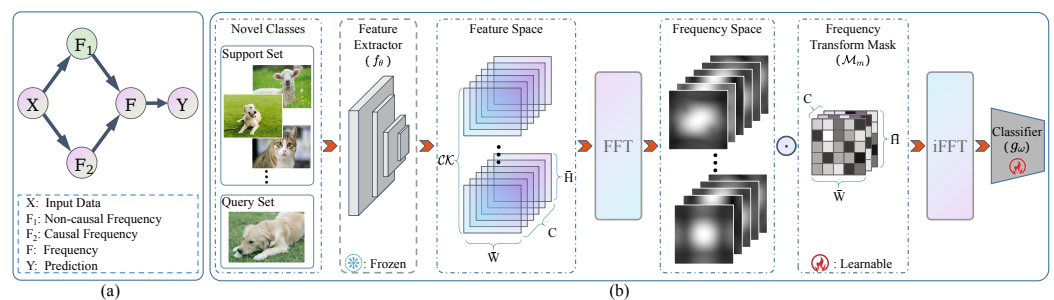


Figure 3. (a) A causal look at FSIC from the frequency spectrum perspective. (b) The test-tuning process. The FRSM weights the frequency learning from novel classes in the testing or fine-tuning phase, which uses a frozen feature extractor pretrained on base classes.

2. Related Work

2.1. Few-Shot Image Classification

Most works have been proposed in recent years to solve the FSIC problem. These methods are broadly categorized into three groups: (1) *fine-tuning-based methods* employ a non-episodic paradigm to pretrain the model on base classes and then test-tune the pretrained model on novel classes. For example, Baseline and Bselene++ [3] propose using the fully connected layer or cosine distance to calculate the classification loss, respectively. SKD [23] further employs rotation-based self-supervision learning during pretraining to enhance the feature extraction process. (2) *Metric-based methods* aim to learn a discriminative embedding space by leveraging a learned distance metric. For example, ProtoNet [10] proposes performing classification by measuring the Euclidean distance between the query and prototype representation of each class. RelationNet [24] proposes calculating the distance between the query and prototype by leveraging the relational module. (3) *Meta-based methods* focus on learning how to optimize a model through bi-level optimization. In particular, MAML [11] proposes quickly adapting to novel classes with only a few labeled images and a small number of gradient updates. LEO [12] proposes learning a stochastic latent space from high-dimensional parameters. Our method tends toward fine-tuning based methods due to their simplicity and effectiveness.

2.2. Frequency Spectrum Learning

In conventional image processing, frequency analysis has been widely studied for years. Recently, it has been set forth to be incorporated into deep learning methods. Some researchers have been working on analyzing and understanding some behaviors of deep neural networks. The frequency spectrum has garnered considerable attention from researchers due to its utility in analyzing and comprehending the behavior and interpretability of deep neural networks. For example, in [25], the authors observed that

convolutional neural networks (CNNs) exhibit a pronounced bias toward recognizing textures rather than shapes. Some research efforts have aimed at enhancing the generalization capability of CNNs by leveraging insights from the frequency spectrum. Notably, FACT [26] proposes forcing the model to capture phase information under the assumption that phase information is more robust to distribution shifts. FSDR [27] improves the generalization capability through randomizing images in a frequency space, preserving domain-invariant frequency components while randomizing domain-variant ones exclusively. Diverging from prior research, ours centers on analyzing the transferability of the FSIC problem across base and novel classes through frequency analysis.

3. Methodology

In this section, we begin by delineating the problem formulation of few-shot image classification, setting the stage for a detailed exploration of this challenge. Following this, we adopt a novel approach by examining the issue through the lens of causal analysis from a frequency spectrum perspective. This allows us to identify key frequency components that significantly influence classification outcomes, especially when limited data are available. Lastly, we introduce a straightforward yet potent technique: the frequency spectrum mask (FRSM). This method strategically adjusts the weights of frequency components to enhance the model's capability to emphasize the most pertinent features for classification. This refinement ultimately leads to improved accuracy and robustness in few-shot learning scenarios. By implementing the FRSM, we aim to mitigate the impact of irrelevant frequency noise and amplify the contribution of crucial frequencies.

3.1. Few-Shot Image Classification

Several approaches have been proposed to tackle the few-shot image classification (FSIC) problem. Among these, fine-tuning-based methods have emerged as particularly effective due to their simplicity and notable efficacy. Therefore, we developed the FRSM based on fine-tuning techniques. These methods typically adhere to the standard transfer learning procedure, which comprises two phases: the **pretraining phase**, using the training set $\mathcal{D}_{train}$, and the **test-tuning phase**, employing a testing set $\mathcal{D}_{test}$. Additionally, a validation set $\mathcal{D}_{val}$ is utilized for model selection during the test-tuning phase.

During the pretraining phase, the entire training set $\mathcal{D}_{train}$ is employed to pretrain the feature extractor and classifier on base classes $\mathcal{C}_{base}$ using the standard cross-entropy loss $\mathcal{L}_{ce}$ as follows:

$$\arg \min_{\theta, \omega} \sum_{i=1}^N \mathcal{L}_{ce}(g_{\omega}(z_i), y_i), \quad \text{s.t. } z_i = f_{\theta}(x_i), \quad (1)$$

where f and g are the feature extractor and classifier, parameterized by θ and ω , respectively. The pairs $\{x_i, y_i\}$ represent images and their corresponding labels from the training set, denoted as $\mathcal{D}_{train} = \{x_i, y_i\}_{i=1}^N$, where $x_i \in \mathbb{R}^{3 \times H \times W}$, N is the total number of images and H and W represent the height and width of the images, respectively, with the one-hot labeling vector $y_i \in Y = \{0, 1\}$.

During the test-tuning phase, we sampled a $\mathcal{C}$ -way $\mathcal{K}$ -shot episode from the testing set $\mathcal{D}_{test}$ containing novel classes $\mathcal{C}_{novel}$. Each episode $\mathcal{T}$ comprises a support set $\mathcal{S}$ and a query set $\mathcal{Q}$, denoted as $\mathcal{T} = \langle \mathcal{S}, \mathcal{Q} \rangle$, where $\mathcal{S} = \{x_i, y_i\}_{i=1}^{\mathcal{CK}}$ represents the support samples and $\mathcal{Q} = \{x_i, y_i\}_{i=1}^{\mathcal{CM}}$ denotes the query samples, with $\mathcal{CK}$ denoting the number of support samples and $\mathcal{CM}$ denoting the number of query samples. A new $\mathcal{C}$ -class classifier will be relearned based on $\mathcal{S}$ every time. Basically, the pretrained embedding parameter θ is fixed to avoid overfitting because there are limited labeled data in $\mathcal{S}$. Once the novel classifier is learned, the labels of $\mathcal{Q}$ can be predicted.

3.2. A Causal Graph of FSIC

We used a casual graph of the FSIC to model and construct a structural causal model (SCM) in Figure 3a. It shows the causality relationship among five variances: the input

data X , non-causal frequency F_1 , causal frequency F_2 , frequency F , and prediction or label Y , where the link from one variable to another indicates the cause-and-effect relationship (cause $\rightarrow$ effect). We will now list the following explanations for our proposed SCM:

- $F_2 \leftarrow X \rightarrow F_1$. The variable F_2 signifies the causal frequency that genuinely represents the inherent characteristic of the input data X , such as the object details in an image. Conversely, F_1 indicates the non-causal frequency typically resulting from biases in the data or superficial patterns, such as the background details of an image. Given that F_2 and F_1 coexist within the input data X , these causal relationships are established.
- $F_2 \rightarrow F \leftarrow F_1$. The variable F corresponds to the frequency of the input data X ; that is, $F = FFT(f_\theta(X))$, with FFT denoting the fast Fourier transformation. To produce F , the traditional learning approach uses both the non-causal frequency F_1 and the causal frequency F_2 to extract the discriminative features.
- $F \rightarrow Y$. The primary aim of learning via the frequency spectrum is to ascertain the attributes of the input data X . The classifier g_ω will determine the prediction Y based on the frequency F ; specifically, $Y = g_\omega(iFFT(F))$, where $iFFT$ stands for the inverse fast Fourier transformation.

When scrutinizing our proposed SCM, we acknowledge the role of the non-causal frequency F_1 as a confounding factor between F_2 and Y , stemming from the existence of a backdoor path between F_2 and Y ; specifically, $F_2 \leftarrow X \rightarrow F_1 \rightarrow Z \rightarrow Y$. Even if F_2 lacks a direct connection to Y , the presence of this backdoor path can lead F_2 to exhibit a spurious correlation with Y , resulting in erroneous predictions based on the non-causal frequency F_1 instead of the causal frequency F_2 . Hence, it is imperative to enhance the transferability of FSIC by adjusting the frequencies and mitigating the influence of the non-causal frequency. In this paper, we propose a straightforward yet effective method, the FRSM, for assigning weights to each frequency.

3.3. Frequency Spectrum Mask

In this section, we introduce our FRSM to weight each frequency. Figure 3b illustrates the overall learnable process of the FRSM. Note that the FRSM is used in the test-tuning phase with novel classes $\mathcal{C}_{novel}$, which uses a frozen feature extractor f_θ pretrained on the base classes $\mathcal{C}_{base}$. The reason for this is that the distribution shift between the base and novel classes makes it impossible for a feature extractor pretrained on the base classes to extract features suitable for novel classes. Therefore, we needed to weight the extracted features to meet the requirements of the novel classes. Following previous works [20,26], the FRSM was developed in the frequency space, as the frequency space has more information than the feature space.

In the test-tuning phase, firstly, a $\mathcal{C}$ -way $\mathcal{K}$ -shot episode $\mathcal{T} = \langle \mathcal{S}, \mathcal{Q} \rangle$ is sampled from the novel classes $\mathcal{C}_{novel}$. Then, for each episode, the few labeled support images $\mathcal{S} = \{x_i, y_i\}_{i=1}^{\mathcal{K}}$ are used to pre-learn a new classifier g_ω and our FRSM $\mathcal{M}_m$ parameterized by m , based on the frozen feature extractor f_θ . Finally, the unlabeled query images $\mathcal{Q}$ are used to evaluate the performance of the test-tuned classifier g_ω and FRSM $\mathcal{M}_m$. Next, we give a detailed description of the test-tuning phase.

We used the frozen f_θ to extract the features Z_S of each image in $\mathcal{S}$ (i.e., $Z_S = \{z_i\}_{i=1}^{\mathcal{K}}$ with $z_i = f_\theta(x_i)$ and $z_i \in \mathbb{R}^{C \times \bar{H} \times \bar{W}}$, where C is the channel and $\bar{H}$ and $\bar{W}$ are the height and width of the feature representations, respectively). For each image feature, we transformed it into frequency space to find the corresponding frequency representation z_i^F with a fast Fourier transformation (FFT), which is formulated as follows:

$$\begin{aligned} z_i^F(\bar{h}^F, \bar{w}^F) &= FFT(z_i(\bar{h}, \bar{w})) \\ &= \sum_{\bar{h}=0}^{\bar{H}-1} \sum_{\bar{w}=0}^{\bar{W}-1} z_i(\bar{h}, \bar{w}) e^{-j2\pi(\bar{h}^F \frac{\bar{h}}{\bar{H}} + \bar{w}^F \frac{\bar{w}}{\bar{W}})}, \end{aligned} \quad (2)$$

where $(\bar{h}, \bar{w}) \in (\bar{H}, \bar{W})$ is the height and width pair in the feature space and $(\bar{h}^F, \bar{w}^F) \in (\bar{H}^F, \bar{W}^F)$ is the corresponding height and width pair in the frequency spectrum space. Following the literature [26], the frequency representation z_i^F includes the amplitude z_i^A and phase z_i^P components. Then, the two components z_i^A and z_i^P are calculated as follows:

$$\begin{aligned} z_i^A(\bar{h}^F, \bar{w}^F) &= \left[R^2(z_i)(\bar{h}^F, \bar{w}^F) + I^2(z_i)(\bar{h}^F, \bar{w}^F) \right]^{1/2}, \\ z_i^P(\bar{h}^F, \bar{w}^F) &= \arctan \left[\frac{I(z_i)(\bar{h}^F, \bar{w}^F)}{R(z_i)(\bar{h}^F, \bar{w}^F)} \right], \end{aligned} \quad (3)$$

where $R(z_i)$ and $I(z_i)$ represent the real and imaginary parts of $z_i^F(\bar{h}^F, \bar{w}^F)$, respectively. The operator $\arctan$ is the inverse tangent function. Since z_i^A represents the magnitude of the frequency, it is crucial to assign appropriate weights to it for accurate representation and analysis. To achieve this, we propose using our proposed FRSM, denoted as $\mathcal{M}_m$. By applying this method, we can effectively weight z_i^A , thereby enhancing the precision. The transformed frequency components are formulated as shown in the equation below:

$$\hat{z}_i^A(\bar{h}^F, \bar{w}^F) = \mathcal{M}_m \odot z_i^A(\bar{h}^F, \bar{w}^F), \quad (4)$$

where $\mathcal{M}_m \in \mathbb{R}^{C \times \bar{H} \times \bar{W}}$ is initialized to one and $\odot$ is the element-wise product. The weighted amplitude $\hat{z}_i^A$ is combined with the original phase z_i^P to form a new frequency $\hat{z}_i^F$:

$$\hat{z}_i^F(\bar{h}^F, \bar{w}^F) = \hat{z}_i^A(\bar{h}^F, \bar{w}^F) * e^{-jz_i^P(\bar{h}^F, \bar{w}^F)}. \quad (5)$$

The new frequency representation $\hat{z}_i^F$ can be transferred to the original feature space via an inverse fast Fourier transformation (iFFT), which can be formulated as follows:

$$\begin{aligned} \hat{z}_i(\bar{h}, \bar{w}) &= iFFT(\hat{z}_i^F(\bar{h}^F, \bar{w}^F)) \\ &= \frac{1}{\bar{H} \cdot \bar{W}} \sum_{\bar{h}=0}^{\bar{H}-1} \sum_{\bar{w}=0}^{\bar{W}-1} \hat{z}_i^F(\bar{h}^F, \bar{w}^F) e^{j2\pi(\bar{h}^F \frac{\bar{h}}{\bar{H}} + \bar{w}^F \frac{\bar{w}}{\bar{W}})}. \end{aligned} \quad (6)$$

The new feature representation $\hat{z}_i$ is used to replace the previous feature representation z_i in Equation (1). Then, the cross-entropy loss $\mathcal{L}^{ce}$ is calculated to update the new classifier g_ω and our proposed FRSM $\mathcal{M}_m$.

4. Experiment

In this section, the research questions guide the running of experiments as follows. **Q1:** Why does the transferability decrease as the distribution shifts between the base and novel classes increase? **Q2:** What information should be transferred from the base classes to the novel classes? **Q3:** How effective is our FRSM for few-shot image classification tasks in all testing datasets?

4.1. Experimental Set-Up

Datasets. In our experiments, for the training dataset $\mathcal{D}_{train}$, we selected the train split of *miniImageNet* [9] to pretrain the feature extractor due to the diversity and complexity of its images. The training dataset was also named *miniImageNet-train*. For the testing dataset $\mathcal{D}_{test}$, we chose nine benchmarks, including the subsets of ILSVRC-2012 [28] (i.e., the test split of *miniImageNet*, named *miniImageNet-test*, and *tieredImageNet*), all evaluation datasets of cross-domain few-shot learning (CDFSL) [29] (i.e., EuroSAT, CropDisease, ChestX, and ISIC), and the evaluation benchmarks proposed by Tseng et al. [15] (i.e., CUB, Cars and Plantae). All images in the above datasets were resized to 84×84 pixels, and data augmentation was used. For more details, please refer to LibFSL [30] and Table 1.

Table 1. Summary of testing datasets we used in this paper. For each dataset, we picked 5 classes, and we show illustrative images.

Dataset	Samples	Number of Classes	Number of Samples
<i>tieredImageNet</i>		160	206,209
CUB-200-2011		200	11,788
ISIC		7	10,015
CropDisease		38	43,456
Cars		196	16,185
EuroSAT		10	27,000
Plantae		69	26,650
ChestX		7	25,848

Baseline. To evaluate the effectiveness of our FRSM method, we selected seven classic and effective methods for comparison. They included Baseline [3], Baseline++ [3], and SKD [23] from the fine-tuning-based methods, ProtoNet [10] and RelationNet [24] from the metric-based methods, and MAML [11] and LEO [12] from the meta-based methods, where SKD and LEO recognized the novel classes based on the pretrained model in the training phase.

Experimental details. Following the literature [30], we adopted two different feature extractors: Conv64F (see Table 2) and ResNet12 (see Figure 1). In the test-tuning phase, we used the pretrained feature extractor from [23]. We utilized the SGD optimizer with a momentum of 0.9 and a weight decay of $1e-3$. The learning rates for the re-learned classifier and our FRSM was initialized as $1e-2$ and 3, respectively. For each example of testing data, we randomly sampled 600 episodes, and each episode contained 5 classes. Each class had 5 or 10 labeled images (support set) and an additional 15 unlabeled images (query set) for performance evaluation, formulating the 5 way 5 shot or 10 shot episode.

Table 2. Few-shot image classification average accuracy (%) on eight testing datasets under the 5 way 5 shot or 5 way 10 shot settings. The model was trained on *miniImageNet*. The best results are bold.

Model	<i>tieredImageNet</i>		CUB-200-2011		CropDisease		EuroSAT		Average	
	5 Shot	10 Shot	5 Shot	10 Shot	5 Shot	10 Shot	5 Shot	10 Shot	5 Shot	10 Shot
Baseline [3]	65.49%	71.72%	55.41%	63.80%	81.21%	87.84%	70.12%	76.99%	68.06%	75.09%
Baseline++ [3]	68.68%	72.62%	53.64%	59.41%	60.60%	68.59%	55.12%	60.09%	59.51%	65.18%
SKD [23]	66.93%	72.70%	55.59%	64.28%	80.36%	87.46%	70.85%	77.18%	68.44%	75.41%
ProtoNet [10]	71.81%	76.17%	57.65%	64.34%	80.15%	85.76%	69.03%	73.10%	69.66%	74.84%
RelationNet [24]	70.16%	64.86%	55.31%	51.18%	61.57%	53.50%	51.56%	38.57%	59.65%	52.03%
GCLR [31]	70.98%	73.12%	55.78%	60.89%	68.56%	70.45%	60.87%	64.57%	64.05%	67.26%
AGNN [32]	72.34%	74.56%	55.89%	62.31%	76.45%	75.68%	70.45%	72.96%	68.78%	71.38%
MAML [11]	68.27%	72.49%	53.09%	59.30%	76.25%	82.43%	65.86%	70.18%	65.87%	71.10%
LEO [12]	70.47%	74.61%	56.79%	62.11%	65.76%	69.24%	59.16%	61.67%	63.04%	66.91%
FRSM (Ours)	67.62%	72.57%	55.03%	63.82%	82.55%	88.93%	74.75%	79.85%	69.99%	76.29%

Table 2. Cont.

Model	Plantae		Cars		ISIC		ChestX		Average	
	5 Shot	10 Shot	5 Shot	10 Shot	5 Shot	10 Shot	5 Shot	10 Shot	5 Shot	10 Shot
Baseline [3]	47.96%	55.22%	42.67%	50.11%	42.00%	49.05%	24.73%	26.60%	39.34%	45.25%
Baseline++ [3]	43.72%	49.63%	36.92%	42.11%	41.11%	45.24%	23.44%	24.84%	36.30%	40.45%
SKD [23]	49.59%	57.05%	41.78%	49.23%	41.11%	47.10%	23.22%	25.56%	38.92%	44.74%
ProtoNet [10]	49.82%	56.08%	41.07%	47.13%	41.96%	48.64%	25.03%	26.52%	39.47%	44.59%
RelationNet [24]	42.26%	39.47%	36.11%	33.18%	32.63%	28.25%	23.23%	22.82%	33.56%	30.93%
GCLR [31]	49.23%	54.67%	39.87%	44.09%	35.68%	39.01%	23.45%	25.78%	37.06%	40.89%
AGNN [32]	50.45%	55.78%	40.78%	48.60%	23.45%	46.90%	24.10%	27.01%	34.70%	44.57%
MAML [11]	43.85%	48.82%	38.54%	43.09%	42.38%	46.18%	23.34%	23.78%	37.03%	40.47%
LEO [12]	48.32%	53.45%	38.86%	43.32%	36.22%	38.47%	22.94%	23.59%	36.59%	39.71%
FRSM (Ours)	51.25%	58.80%	43.66%	51.04%	43.45%	49.45%	25.12%	27.28%	40.87%	46.64%

4.2. Experimental Results

In this section, we run experiments to answer the following questions.

Q1. The transferability from base to novel classes. The experimental results in Figure 4 aim to answer Q1. In Figure 4, we show the similarity of the testing and training datasets according to the amplitude ratio. The testing datasets' similarity to the training dataset was ordered as follows: *miniImageNet*-test > Cars > Plantae > EuroSAT > ISIC > ChestX. The similarity is known to affect the transferability of the training dataset features into the testing datasets. To this end, we found that the datasets' similarity and few-shot difficulty simultaneously led to performance degradation.

Q2. The information from base to novel classes. We conducted experiments on testing datasets, with the results shown in Figure 2. We found that the high-frequency components in different datasets had more similarity than the low-frequency components. Therefore, compared with the low-frequency components, the high-frequency components contributed to the transferability more.

Q3. The performance of the FRSM. To answer Q3, we conducted experiments on eight datasets. From Table 2, which reports the average classification accuracy, we have the following findings. (1) Our proposed FRSM achieved the best performance in most settings. This is because our FRSM assigned a large weight to the causal frequency and a low weight to the non-causal one, thereby avoiding the influence of the confounder on the transferability. (2) Surprisingly, our FRSM was weaker than the baseline for relatively similar testing data (e.g., *tieredImageNet*). The possible reason for this is that the confounder is generally simpler. For example, learning the backgrounds of images is easier than learning the foreground, and this can help the transferability from similar base classes to novel classes. Our FRSM suppressed the confounder and led to performance degradation. Nevertheless, we should focus more on testing datasets with larger distribution shifts, because this case is more in line with real-world environments.

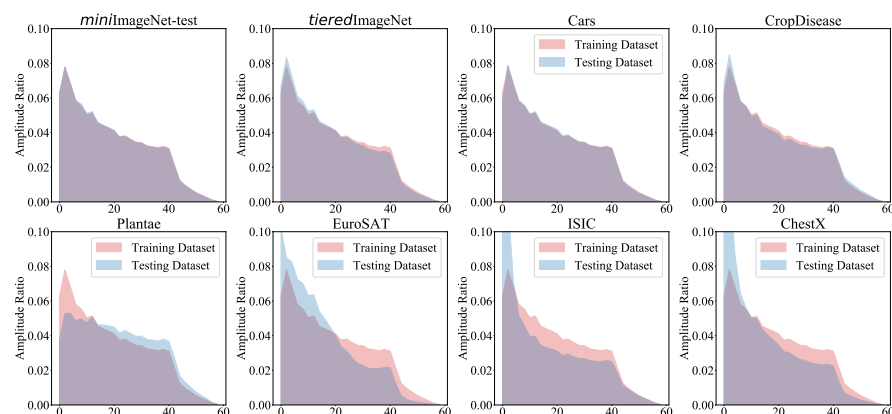


Figure 4. The average amplitude ratio between six testing datasets and the training dataset (i.e., *miniImageNet*-train).

5. Conclusions

In this paper, we explored the underlying reasons for the decline in performance and transferability encountered when facing distribution shifts between base and novel classes. We approached this investigation from the perspective of frequency spectrum analysis. Furthermore, we clarified **which** information ought to be transferred from base to novel classes. Through a causal perspective on FSIC, we delved into the phenomenon and demonstrated that non-causal frequencies can profoundly influence transferability as confounding factors. Therefore, we proposed a straightforward yet effective method, the FRSM, to dynamically weigh each frequency using a learnable paradigm. This approach helps mitigate the influence of confounders and enhances transferability. Extensive experiments demonstrated that our proposed FRSM method achieved new state-of-the-art results.

Author Contributions: Conceptualization, M.Z.; methodology, M.Z.; writing—original draft, M.Z.; writing—review and editing, Z.W. and D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the NSFC General Program (Grant No. 62176215).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.* **2020**, *53*, 1–34. [[CrossRef](#)]
- Thrun, S.; Pratt, L. *Learning to Learn*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
- Chen, W.Y.; Liu, Y.C.; Kira, Z.; Wang, Y.C.F.; Huang, J.B. A closer look at few-shot classification. In Proceedings of the International Conference on Learning Representations, ICLR, New Orleans, LA, USA, 6–9 May 2019.
- Chen, Y.; Liu, Z.; Xu, H.; Darrell, T.; Wang, X. Meta-baseline: Exploring simple meta-learning for few-shot learning. In Proceedings of the International Conference on Computer Vision, ICCV, Montreal, QC, Canada, 10–17 October 2021.
- Tian, Y.; Wang, Y.; Krishnan, D.; Tenenbaum, J.B.; Isola, P. Rethinking few-shot image classification: A good embedding is all you need? In Proceedings of the European Conference on Computer Vision, ECCV, Glasgow, UK, 23–28 August 2020.
- Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2020**, *109*, 43–76. [[CrossRef](#)]
- Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
- Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [[CrossRef](#)]
- Vinyals, O.; Blundell, C.; Lillicrap, T.; Kavukcuoglu, K.; Wierstra, D. Matching networks for one shot learning. In Proceedings of the Advances in Neural Information Processing Systems, NeurIPS, Barcelona, Spain, 5–10 December 2016.
- Snell, J.; Swersky, K.; Zemel, R.S. Prototypical networks for few-shot learning. In Proceedings of the NeurIPS, Long Beach, CA, USA, 4–9 December 2017.
- Finn, C.; Abbeel, P.; Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Proceedings of the ICML, Sydney, Australia, 6–11 August 2017.
- Rusu, A.A.; Rao, D.; Sygnowski, J.; Vinyals, O.; Pascanu, R.; Osindero, S.; Hadsell, R. Meta-Learning with Latent Embedding Optimization. In Proceedings of the International Conference on Learning Representations, ICLR, New Orleans, LA, USA, 6–9 May 2019.
- Raghu, A.; Raghu, M.; Bengio, S.; Vinyals, O. Rapid Learning or Feature Reuse? Towards Understanding the Effectiveness of MAML. In Proceedings of the International Conference on Learning Representations, ICLR, Addis Ababa, Ethiopia, 30 April 2020.
- Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
- Tseng, H.Y.; Lee, H.Y.; Huang, J.B.; Yang, M.-H. Cross-Domain Few-Shot Classification via Learned Feature-Wise Transformation. In Proceedings of the International Conference on Learning Representations, ICLR, Addis Ababa, Ethiopia, 30 April 2020.
- Wang, H.; Deng, Z.H. Cross-Domain Few-Shot Classification via Adversarial Task Augmentation. In Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI, Montreal, QC, Canada, 19–27 August 2021.
- Sun, J.; Lapuschkin, S.; Samek, W.; Zhao, Y.; Cheung, N.M.; Binder, A. Explanation-guided training for cross-domain few-shot classification. In Proceedings of the International Conference on Pattern Recognition, ICPR, Milan, Italy, 10–15 January 2021.

18. Oh, J.; Kim, S.; Ho, N.; Kim, J.H.; Song, H.; Yun, S.Y. Understanding Cross-Domain Few-Shot Learning Based on Domain Similarity and Few-Shot Difficulty. In Proceedings of the Advances in Neural Information Processing Systems, NeurIPS, New Orleans, LA, USA, 28 November–9 December 2022.
19. Islam, A.; Chen, C.F.R.; Panda, R.; Karlinsky, L.; Feris, R.; Radke, R.J. Dynamic distillation network for cross-domain few-shot recognition with unlabeled data. In Proceedings of the Advances in Neural Information Processing Systems, NeurIPS, Virtual, 6–14 December 2021.
20. Long, Y.; Zhang, Q.; Zeng, B.; Gao, L.; Liu, X.; Zhang, J.; Song, J. Frequency domain model augmentation for adversarial attack. In Proceedings of the European Conference on Computer Vision, ECCV, Tel Aviv, Israel, 23–27 October 2022.
21. Xu, K.; Qin, M.; Sun, F.; Wang, Y.; Chen, Y.K.; Ren, F. Learning in the frequency domain. In Proceedings of the Computer Vision and Pattern Recognition, CVPR, Seattle, WA, USA, 13–19 June 2020.
22. Yue, Z.; Zhang, H.; Sun, Q.; Hua, X.S. Interventional few-shot learning. In Proceedings of the Advances in Neural Information Processing Systems, NeurIPS, Virtual, 6–12 December 2020.
23. Rajasegaran, J.; Khan, S.; Hayat, M.; Khan, F.S.; Shah, M. Self-supervised knowledge distillation for few-shot learning. In Proceedings of the British Machine Vision Conference, BMVC, Virtual, 22–25 November 2021.
24. Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P.H.S.; Hospedales, T.M. Learning to compare: Relation network for few-shot learning. In Proceedings of the Computer Vision and Pattern Recognition, CVPR, Salt Lake City, UT, USA, 18–22 June 2018.
25. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.A.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In Proceedings of the International Conference on Learning Representations, ICLR, New Orleans, LA, USA, 6–9 May 2019.
26. Xu, Q.; Zhang, R.; Zhang, Y.; Wang, Y.; Tian, Q. A Fourier-based Framework for Domain Generalization. In Proceedings of the Computer Vision and Pattern Recognition, CVPR, Nashville, TN, USA, 20–25 June 2021.
27. Huang, J.; Guan, D.; Xiao, A.; Lu, S. Fsd: Frequency space domain randomization for domain generalization. In Proceedings of the Computer Vision and Pattern Recognition, CVPR, Nashville, TN, USA, 20–25 June 2021.
28. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
29. Guo, Y.; Codella, N.C.; Karlinsky, L.; Codella, J.V.; Smith, J.R.; Saenko, K.; Rosing, T.; Feris, R. A Broader Study of Cross-Domain Few-Shot Learning. In Proceedings of the European Conference on Computer Vision, ECCV, Glasgow, UK, 23–28 August 2020.
30. Li, W.; Wang, Z.; Yang, X.; Dong, C.; Tian, P.; Qin, T.; Huo, J.; Shi, Y.; Wang, L.; Gao, Y.; et al. Libfewshot: A comprehensive library for few-shot learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 14938–14955. [[CrossRef](#)] [[PubMed](#)]
31. Zhong, X.; Gu, C.; Ye, M.; Huang, W.; Lin, C.W. Graph complemented latent representation for few-shot image classification. *IEEE Trans. Multimed.* **2022**, *25*, 1979–1990. [[CrossRef](#)]
32. Cheng, H.; Zhou, J.T.; Tay, W.P.; Wen, B. Graph neural networks with triple attention for few-shot learning. *IEEE Trans. Multimed.* **2023**, *25*, 8225–8239. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.