CLASS-ADAPTIVE CHARACTERISTIC FUNCTION MATCHING FOR CROSS-DOMAIN FEW-SHOT LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Under domain shift, successful few-shot recognition requires not only transferable source-trained features but also a comparison rule appropriate for novel target classes. Most metric-based approaches to cross-domain few-shot learning (CD-FSL) construct class prototypes from target support examples, yet evaluate all queries using a shared Euclidean or cosine metric. Target supervision therefore adapts the class representatives but leaves the comparison function unchanged across classes. We address this asymmetry with Class-Adaptive Characteristic Function Matching, a plug-and-play metric module that uses target support examples to adapt both class representation and query-class comparison. Our key observation is that the frequency arguments of an empirical characteristic function can parameterize the comparison rule: each frequency induces a nonlinear response to a feature projection, while a finite set of frequencies forms a compact representation of the support class. For each candidate class, a lightweight KL-regularized stochastic generator conditions on its support set and produces a compact set of class-specific frequency arguments. Given a query, we tentatively include it in each candidate support set and measure the resulting change between the original and query-augmented empirical characteristic-function representations. This change serves as the membership score, allowing the same query to be evaluated through a distinct support-conditioned comparison for every novel target class. The proposed module can replace conventional point-based similarities without modifying the feature backbone or the target-domain adaptation procedure of the host method. Experiments across four CD-FSL benchmarks, three base learners, and multiple pretrained backbones demonstrate consistent improvements with only a small parameter overhead. Code and models will be publicly released.

1 INTRODUCTION

Recognition under domain shift depends not only on whether source-trained features remain informative, but also on how those features are compared in the target domain. This issue is particularly important in cross-domain few-shot learning (CD-FSL) Zhou et al. (2023); Zou et al. (2025a), where a model trained on a source domain, typically natural images Vinyals et al. (2016), must recognize novel classes from substantially different domains using only a few labeled examples. Representative target domains include satellite imagery Helber et al. (2019), agricultural imagery Bedi et al. (2021), and medical X-rays Wang et al. (2017). The scarcity of target annotations makes it difficult to relearn either the feature representation or the comparison function from scratch, leaving the model dependent on a feature geometry largely established by the source domain.

A common strategy in CD-FSL is to improve the transferability of the feature extractor while retaining the prototype-based prediction rule used in standard few-shot learning Snell et al. (2017); Vinyals et al. (2016). Each target class is summarized by the mean feature of its support examples, and a query is classified according to its Euclidean or cosine distance from these prototypes. Recent methods improve feature adaptation, task-specific representations, or target-domain generalization Hu et al. (2022); Zhou et al. (2025). Nevertheless, their final prediction commonly relies on a comparison rule shared by all candidate classes. The target support examples therefore determine the class representatives, but have limited influence on how similarity itself is measured.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

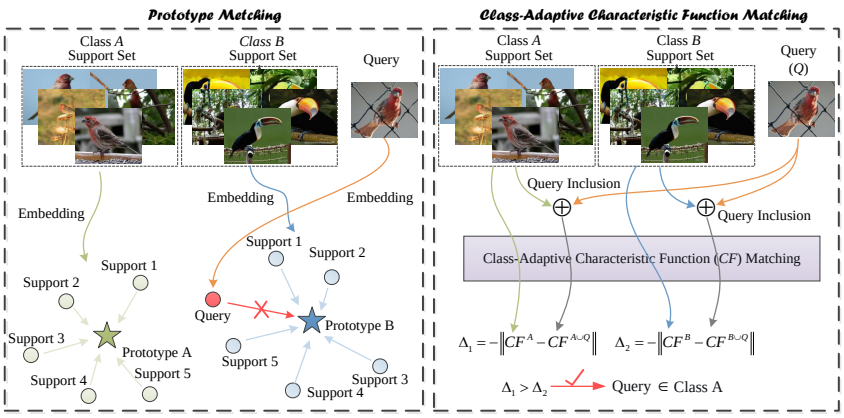


Figure 1: Conceptual comparison between prototype-based matching and our characteristic-function matching. Conventional methods summarize each support class by a single prototype and measure the point-wise distance from the query to that prototype (left). Our method represents the support set through finite evaluations of its empirical characteristic function and assesses query–class compatibility by the representation change induced when the query is tentatively included in the support set (right). The class-adaptive construction of frequency arguments is described in the method section.

Such a class-agnostic comparison can be restrictive under domain shift. The feature directions and discrepancy patterns that distinguish novel target classes may differ from those emphasized by the source-induced feature geometry. Moreover, different target classes within the same episode may require different comparisons. Applying one Euclidean or cosine metric to all classes treats every feature discrepancy in the same manner, regardless of the evidence provided by each support set. Since the support examples are the only labeled observations available from the target domain, we argue that they should be used not only to construct class representations, but also to adapt the comparison rule. This motivates the conceptual transition shown in Fig. 1: from comparing a query with a point prototype under a globally shared metric to assessing its compatibility with a support set through a class-conditioned representation.

Characteristic functions Feuerverger & Mureika (1977) provide a natural mechanism for realizing this idea. A characteristic function represents a distribution through its responses in the frequency domain, while its empirical form can be computed directly from a set of samples without fitting a parametric density model. In our setting, finite evaluations of the empirical characteristic function form a compact nonlinear representation of a support class. More importantly, the frequency arguments determine the feature projections involved in these evaluations and consequently control how feature discrepancies are expressed. Selecting frequency arguments separately for each support class therefore allows the target examples to influence both the resulting class representation and the comparison applied to a query.

Based on this observation, we propose Class-Adaptive Characteristic Function Matching, a plug-and-play metric module for CD-FSL. For each candidate class, a lightweight KL-regularized stochastic generator conditions on its support set and produces a compact set of class-specific frequency arguments. The empirical characteristic-function representation of the class is then constructed using these frequencies. Unlike fixed or globally shared frequency sets, the generated arguments induce support-conditioned nonlinear projections. Consequently, the same query can be evaluated differently for different candidate classes, with each comparison determined by the corresponding target support set. In this way, the proposed module adapts the comparison rule using scarce target supervision without modifying the feature backbone or explicitly aligning the source and target domains.

To convert the support-set representation into a query classification score, we introduce a query-induced change mechanism. For each candidate class, we first construct its empirical characteristic-function representation and then recompute the representation after tentatively including the query in the support set. Both representations are evaluated using the same class-specific frequencies. Their difference measures how strongly the query perturbs the representation of that class, with a smaller change indicating greater query–class compatibility. This formulation avoids the need to estimate a separate distribution for an individual query while preserving a set-level comparison between the query and each candidate support class.

The proposed module can replace conventional Euclidean or cosine similarities in existing CD-FSL pipelines without changing their feature extractors, training objectives, or target-domain adaptation procedures. We evaluate it on four CD-FSL benchmarks using three representative base learners and multiple pretrained backbones. The results demonstrate consistent performance improvements across different target domains and host methods, with only a small additional parameter and computational overhead.

In summary, our main contributions are threefold: (1) We introduce class-adaptive characteristic-function matching for CD-FSL, enabling target support examples to determine both the class representation and the comparison rule used for query classification. (2) We develop a KL-regularized stochastic frequency generator that produces class-specific frequency arguments from limited support examples, together with a query-induced representation-change score for measuring query–class compatibility. (3) We demonstrate the plug-and-play applicability of the proposed module across four CD-FSL benchmarks, three representative base learners, and multiple pretrained backbones, achieving consistent improvements with modest additional overhead.

2 METHODOLOGY

2.1 PROBLEM FORMULATION AND OVERVIEW

In CD-FSL, a model trained on a source domain $\mathcal{D}_s$ is evaluated on a target domain $\mathcal{D}_t$, whose classes are disjoint from those in the source domain. Each target episode contains N novel classes indexed by $c \in \{1, \ldots, N\}$. For class c, its support set is denoted by

$$\mathcal{S}_c = \{(x_i^c, c)\}_{i=1}^{K_c}, \tag{1}$$

where x_i^c is the i-th support image and K_c is the number of labeled support examples. The complete support set is $\mathcal{S} = \bigcup_{c=1}^{N} \mathcal{S}_c$, and $\mathcal{Q}$ denotes the query set. Standard CD-FSL benchmarks adopt balanced N-way K-shot episodes, where $K_c = K$ for every candidate class.

Let f_θ denote the feature extractor of the host CD-FSL method. The support and query features are

$$\mathbf{f}_i^c = f_\theta(x_i^c), \qquad \mathbf{f}_q = f_\theta(x_q), \qquad \mathbf{f}_i^c, \mathbf{f}_q \in \mathbb{R}^d, \tag{2}$$

where d is the feature dimension. We denote the support feature set of class c by $\mathcal{F}_c = \{\mathbf{f}_i^c\}_{i=1}^{K_c}$.

Prototype-based methods summarize each candidate class by

$$\mathbf{p}_c = \frac{1}{K_c} \sum_{i=1}^{K_c} \mathbf{f}_i^c \tag{3}$$

and assign a query according to

$$\widehat{y}_q = \arg\min_c d\left(\mathbf{f}_q, \mathbf{p}_c\right), \tag{4}$$

where $d(\cdot, \cdot)$ is typically Euclidean or cosine distance. The prototype depends on the target support class, whereas the comparison function remains shared across all candidate classes. Under domain shift, such a class-agnostic comparison may be restrictive because the feature directions useful for distinguishing different target classes can vary.

We address this limitation by allowing each target support class to determine its own comparison. As illustrated in Fig. 2, a stochastic frequency generator first produces a compact set of class-specific frequency arguments from each support set. These frequencies define a finite empirical characteristic-function representation of the support class. Given a query, we tentatively include it in each candidate support set and measure the resulting representation change. Since the frequency set is conditioned on the candidate support class, the same query is evaluated through a different comparison for every class.

2.2 FINITE EMPIRICAL CHARACTERISTIC REPRESENTATION

Let $\mathbf{f} \in \mathbb{R}^d$ be a random feature vector with probability distribution P. Its characteristic function is defined as

$$\Psi(\boldsymbol{\omega}; P) = \mathbb{E}_{\mathbf{f} \sim P}\left[e^{j\langle \boldsymbol{\omega}, \mathbf{f} \rangle}\right], \qquad \boldsymbol{\omega} \in \mathbb{R}^d, \tag{5}$$

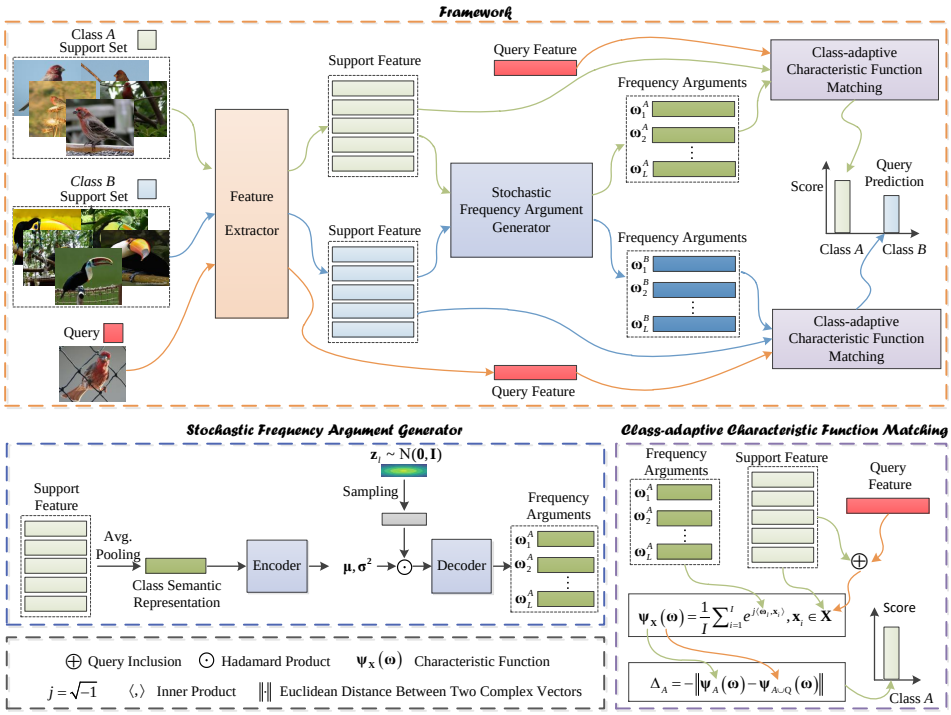


Figure 2: Overview of Class-Adaptive Characteristic Function Matching. A feature extractor encodes support and query images. For each candidate class, the stochastic frequency argument generator summarizes its support features and produces a compact set of class-specific frequencies. These frequencies define the empirical characteristic-function representation used to evaluate that class. Query compatibility is measured by the representation change induced when the query is tentatively included in the support set. The frequencies are generated from the original support set and remain fixed during query inclusion.

where $\boldsymbol{\omega}$ is a d-dimensional frequency argument, $j = \sqrt{-1}$ is the imaginary unit, and $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product. The expectation is taken with respect to $\mathbf{f} \sim P$, and $\Psi(\boldsymbol{\omega}; P) \in \mathbb{C}$ is the complex-valued response of P at frequency $\boldsymbol{\omega}$.

Given a finite feature set $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{I}$, the empirical characteristic function is

$$\widehat{\Psi}(\boldsymbol{\omega}; \mathcal{X}) = \frac{1}{I} \sum_{i=1}^{I} e^{j\langle \boldsymbol{\omega}, \mathbf{x}_i \rangle}, \tag{6}$$

where $I = |\mathcal{X}|$ is the cardinality of the feature set. This estimator can be computed directly from the observed features without fitting a parametric density model.

In practice, evaluating the characteristic function over its continuous frequency domain is infeasible. We instead consider a finite set of L frequency arguments,

$$\Omega = \{\boldsymbol{\omega}_\ell\}_{\ell=1}^{L}, \qquad \boldsymbol{\omega}_\ell \in \mathbb{R}^d. \tag{7}$$

The resulting finite empirical characteristic representation is

$$\boldsymbol{\Psi}(\mathcal{X}; \Omega) = \left[\widehat{\Psi}(\boldsymbol{\omega}_1; \mathcal{X}), \ldots, \widehat{\Psi}(\boldsymbol{\omega}_L; \mathcal{X})\right]^\top \in \mathbb{C}^L, \tag{8}$$

where $(\cdot)^\top$ denotes vector transpose.

Each coordinate of this representation applies a nonlinear complex response to the projection $\langle \boldsymbol{\omega}_\ell, \mathbf{x}_i \rangle$. The frequency arguments therefore determine the feature directions and variation scales involved in the representation. Selecting Ω is consequently not only a numerical approximation problem; it determines how feature discrepancies are expressed and compared. This observation motivates generating a different frequency set for each target support class.

2.3 Stochastic Frequency Argument Generator

A fixed or globally shared frequency set applies the same Fourier projections to every candidate class. We instead generate

$$\Omega_c = \{\boldsymbol{\omega}_\ell^c\}_{\ell=1}^{L} \tag{9}$$

from the support features $\mathcal{F}_c$ of each class c, where $\boldsymbol{\omega}_\ell^c \in \mathbb{R}^d$ denotes its ℓ-th frequency argument.

Following the framework in Fig. 2, we first aggregate the support features into a class semantic representation:

$$\mathbf{h}_c = \frac{1}{K_c} \sum_{i=1}^{K_c} \mathbf{f}_i^c \in \mathbb{R}^d. \tag{10}$$

Instead of directly predicting all Ld frequency values, an encoder E_η, parameterized by η, maps $\mathbf{h}_c$ to the parameters of an r-dimensional Gaussian latent distribution:

$$\left(\boldsymbol{\mu}_c, \log \boldsymbol{\sigma}_c^2\right) = E_\eta(\mathbf{h}_c), \tag{11}$$

where $\boldsymbol{\mu}_c \in \mathbb{R}^r$ and $\boldsymbol{\sigma}_c \in \mathbb{R}_+^r$. The resulting class-conditioned latent distribution is

$$q_\eta(\widetilde{\mathbf{z}} \mid \mathbf{h}_c) = \mathcal{N}\left(\widetilde{\mathbf{z}}; \boldsymbol{\mu}_c, \mathrm{diag}(\boldsymbol{\sigma}_c^2)\right), \tag{12}$$

where $\mathrm{diag}(\boldsymbol{\sigma}_c^2)$ denotes a diagonal covariance matrix.

For each frequency index $\ell \in \{1, \ldots, L\}$, we independently sample a standard Gaussian noise vector

$$\mathbf{z}_\ell^c \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_r) \tag{13}$$

and obtain a class-conditioned latent vector through reparameterization:

$$\widetilde{\mathbf{z}}_\ell^c = \boldsymbol{\mu}_c + \boldsymbol{\sigma}_c \odot \mathbf{z}_\ell^c, \tag{14}$$

where $\mathbf{I}_r$ is the r-dimensional identity matrix and $\odot$ denotes element-wise multiplication. A shared decoder G_ξ, parameterized by ξ, maps the latent vector to a frequency argument:

$$\boldsymbol{\omega}_\ell^c = G_\xi\left(\widetilde{\mathbf{z}}_\ell^c\right) \in \mathbb{R}^d. \tag{15}$$

Repeating this procedure L times yields Ω_c. The shared decoder transforms samples from the low-dimensional class-conditioned latent space into frequency arguments, while the support representation controls the location and scale of the sampling distribution. The generator is optimized end-to-end through episodic classification, enabling the produced frequencies to support discrimination among novel classes.

2.4 Class-Adaptive Characteristic Function Matching

Support Representation. Given the generated frequency set Ω_c, the empirical characteristic function of class c at its ℓ-th frequency is

$$\Psi_\ell^c = \widehat{\Psi}\left(\boldsymbol{\omega}_\ell^c; \mathcal{F}_c\right) = \frac{1}{K_c} \sum_{i=1}^{K_c} e^{j\langle \boldsymbol{\omega}_\ell^c, \mathbf{f}_i^c \rangle}, \qquad \ell = 1, \ldots, L. \tag{16}$$

Collecting the L responses gives the finite representation

$$\boldsymbol{\Psi}^c = [\Psi_1^c, \ldots, \Psi_L^c]^\top \in \mathbb{C}^L. \tag{17}$$

Every support feature is evaluated at every class-specific frequency before the responses are averaged.

Query-Induced Matching. For query feature $\mathbf{f}_q$, its response under the frequency set of class c is

$$\boldsymbol{\Phi}_q^c = \left[e^{j\langle \boldsymbol{\omega}_1^c, \mathbf{f}_q \rangle}, \ldots, e^{j\langle \boldsymbol{\omega}_L^c, \mathbf{f}_q \rangle}\right]^\top \in \mathbb{C}^L. \tag{18}$$

Since Ω_c is class-specific, the same query produces a different response vector for each candidate class.

We tentatively include the query in the support feature set:

$$\mathcal{F}_{c,q} = \mathcal{F}_c \cup \{\mathbf{f}_q\}. \tag{19}$$

The frequency set Ω_c is generated from the original support set $\mathcal{F}_c$ and remains fixed after query inclusion. The empirical characteristic representation of the augmented set is

$$\begin{aligned} \mathbf{\Psi}_q^c &= \mathbf{\Psi}\left(\mathcal{F}_{c,q}; \Omega_c\right) \\ &= \frac{K_c \mathbf{\Psi}^c + \mathbf{\Phi}_q^c}{K_c + 1}. \end{aligned} \tag{20}$$

We define the membership score of query q with respect to class c as

$$\Delta_q^c = -\left\| \mathbf{\Psi}_q^c - \mathbf{\Psi}^c \right\|_2, \tag{21}$$

where the Euclidean norm of a complex vector $\mathbf{a} \in \mathbb{C}^L$ is

$$\|\mathbf{a}\|_2 = \sqrt{\sum_{\ell=1}^{L} |a_\ell|^2}, \tag{22}$$

and $|\cdot|$ denotes the complex modulus. A query that induces a smaller change in the support representation receives a larger membership score.

Relation to Direct Fourier Matching. Substituting the query-augmented representation into the membership score gives

$$\begin{aligned} \Delta_q^c &= -\left\| \frac{K_c \mathbf{\Psi}^c + \mathbf{\Phi}_q^c}{K_c + 1} - \mathbf{\Psi}^c \right\|_2 \\ &= -\frac{1}{K_c + 1} \left\| \mathbf{\Phi}_q^c - \mathbf{\Psi}^c \right\|_2. \end{aligned} \tag{23}$$

This expression is the general query-induced score for arbitrary class-specific support cardinalities. It corresponds to a support-size-weighted distance between the query response and the support representation in the class-specific Fourier space.

Under the balanced N-way K-shot protocol used by standard CD-FSL benchmarks, $K_c = K$ for all candidate classes. The factor $1/(K+1)$ is therefore class-independent, and

$$\arg\max_c \Delta_q^c = \arg\min_c \left\| \mathbf{\Phi}_q^c - \mathbf{\Psi}^c \right\|_2. \tag{24}$$

Thus, in the common balanced-shot setting, query-induced matching reduces to direct query-to-class Fourier matching up to a common scaling factor. This reduction concerns only the query-inclusion operation: both $\mathbf{\Phi}_q^c$ and $\mathbf{\Psi}^c$ remain dependent on the class-specific frequency set Ω_c. Different candidate classes therefore continue to evaluate the same query through different support-conditioned comparisons. When the support cardinalities vary across classes, the factor $1/(K_c+1)$ remains part of the general query-induced score.

2.5 Episodic Learning and Inference

During source-domain episodic training, the membership scores are used as classification logits:

$$p(y_q = c \mid x_q, \mathcal{S}) = \frac{e^{\Delta_q^c}}{\sum_{r=1}^{N} e^{\Delta_q^r}}, \tag{25}$$

where y_q denotes the query label. The classification loss is

$$\mathcal{L}_{\text{cls}} = -\frac{1}{|\mathcal{Q}|} \sum_{(x_q, y_q) \in \mathcal{Q}} \log p(y_q \mid x_q, \mathcal{S}). \tag{26}$$

We regularize each class-conditioned latent distribution toward a standard Gaussian prior:

$$\mathcal{L}_{\text{KL}} = \sum_{c=1}^{N} D_{\text{KL}}\left(q_\eta(\widetilde{\mathbf{z}} \mid \mathbf{h}_c) \,\|\, \mathcal{N}(\widetilde{\mathbf{z}}; \mathbf{0}, \mathbf{I}_r)\right), \tag{27}$$

where $D_{\mathrm{KL}}(\cdot\|\cdot)$ denotes the Kullback–Leibler divergence. The overall training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{cls}} + \lambda \mathcal{L}_{\mathrm{KL}}, \tag{28}$$

where $\lambda \geq 0$ controls the regularization strength. The classification loss optimizes the frequency generator for episodic discrimination, while the KL term regularizes the class-conditioned latent space.

At target-domain evaluation, the adaptation procedure of the host CD-FSL method is retained. After any host-specific adaptation, one frequency set Ω_c is generated from each original target support class and reused when evaluating its queries. The predicted label of query x_q is

$$\widehat{y}_q = \arg \max_{c \in \{1,\ldots,N\}} \Delta_q^c. \tag{29}$$

The proposed module therefore changes only the query–class comparison, allowing scarce target support supervision to determine a class-specific comparison without modifying the feature backbone or the target-domain adaptation procedure of the host method.

3 Experiments

3.1 Experimental Setup

Datasets and Evaluation Protocol. Following prior work Guo et al. (2020); Zhou et al. (2023); Zou et al. (2025a); Li et al. (2025); Ma et al. (2025), we use mini-ImageNet Vinyals et al. (2016), containing 64 base classes, as the source domain for meta-training. Evaluation is conducted on four target domains with disjoint classes and substantial domain shifts: EuroSAT Helber et al. (2019) for remote sensing, ISIC Tschandl et al. (2018) for dermoscopic lesion recognition, ChestX Wang et al. (2017) for chest radiography, and CropDisease Bedi et al. (2021) for plant disease recognition. For each target domain, we sample 600 5-way episodes under the 1-shot and 5-shot settings, with 15 queries per class. Each method uses the support set according to its original adaptation protocol, and we report the mean query accuracy over all episodes.

Implementation Details. We integrate our method into three representative metric-based approaches: ProtoNet Snell et al. (2017), PMF Hu et al. (2022), and CD-CLS Zou et al. (2025a). The proposed module replaces their original similarity function without altering the feature backbone or target-domain adaptation procedure. All host-specific hyperparameters and optimization settings follow the corresponding implementations. Both E_η and G_ξ are single-layer MLPs, where d denotes the backbone output dimension (e.g., $d = 384$ for ViT-Small) and the latent dimension is set to $r = 128$. We use $L = 1024$ frequency arguments and set the KL coefficient λ to 0.3 for 1-shot tasks and 0.01 for 5-shot tasks. The generator is trained for 100 epochs using Adam with a learning rate of 0.01. All experiments are conducted on an NVIDIA RTX 3090 GPU.

Table 1: Comparison with state-of-the-art methods on 5-way 1-shot cross-domain few-shot learning.

Method	Pre-train	Backbone	ChestX	ISIC	EuroSAT	CropDisease	Avg.
MEM Walsh & Osman (2023)	DINOv1	ViT-Small	22.76	32.97	68.11	81.11	51.24
StyleAdv Fu et al. (2023)	DINOv1	ViT-Small	22.92	33.05	72.15	81.22	52.34
FLoR Zou et al. (2024)	DINOv1	ViT-Small	23.26	35.49	73.09	83.55	53.85
SVasP Li et al. (2025)	DINOv1	ViT-Small	22.68	34.49	75.51	83.98	54.17
DAMIM Ma et al. (2025)	DINOv1	ViT-Small	23.38	36.35	73.61	83.90	54.31
SG-FSL Zou et al. (2025b)	DINOv1	ViT-Small	23.01	34.39	72.82	83.38	53.40
FMA-Net Zhou et al. (2025)	DINOv1	ViT-Small	23.13	34.48	72.08	82.13	52.96
ProtoNet Snell et al. (2017)	DINOv1	ViT-Small	22.57	33.29	72.42	81.02	52.33
ProtoNet+Ours	DINOv1	ViT-Small	22.78	34.69	73.02	82.35	53.21
PMF Hu et al. (2022)	DINOv1	ViT-Small	21.73	30.36	70.74	80.79	50.91
PMF+Ours	DINOv1	ViT-Small	22.93	35.37	73.04	81.88	53.31
CD-CLS Zou et al. (2025a)	DINOv1	ViT-Small	23.39	35.56	74.97	84.54	54.62
CD-CLS+Ours	DINOv1	ViT-Small	23.61	36.68	75.71	85.24	55.31

3.2 Comparison with State-of-the-Art Methods

As shown in Tables 1 and 2, our module improves the average accuracy of all three host methods in both settings. In 1-shot evaluation, it improves ProtoNet from 52.33% to 53.21%, PMF from 50.91% to 53.31%, and CD-CLS from 54.62% to 55.31%, corresponding to gains of 0.88, 2.40, and 0.69 percentage points, respectively. In 5-shot evaluation, the corresponding gains are 1.37, 2.74, and 0.88 points. CD-CLS equipped with our module achieves the highest average accuracy among all compared methods, reaching 55.31% in the 1-shot setting and 68.42% in the 5-shot setting.

Table 2: Comparison with state-of-the-art methods on 5-way 5-shot cross-domain few-shot learning.

Method	Pre-train	Backbone	ChestX	ISIC	EuroSAT	CropDisease	Avg.
MEM Walsh & Osman (2023)	DINOv1	ViT-Small	26.67	47.38	86.49	93.74	63.57
StyleAdv Fu et al. (2023)	DINOv1	ViT-Small	26.97	47.73	88.57	94.85	64.53
FLoR Zou et al. (2024)	DINOv1	ViT-Small	27.02	53.06	90.75	96.47	66.83
SVasP Li et al. (2025)	DINOv1	ViT-Small	26.77	51.62	90.55	96.17	66.28
DAMIM Ma et al. (2025)	DINOv1	ViT-Small	27.82	54.86	91.18	96.34	67.55
SG-FSL Zou et al. (2025b)	DINOv1	ViT-Small	26.84	48.56	90.55	96.11	65.52
FMA-Net Zhou et al. (2025)	DINOv1	ViT-Small	26.25	50.55	89.99	95.81	65.65
ProtoNet Snell et al. (2017)	DINOv1	ViT-Small	25.76	48.63	88.88	94.75	64.51
ProtoNet+Ours	DINOv1	ViT-Small	26.42	51.89	89.12	96.07	65.88
PMF Hu et al. (2022)	DINOv1	ViT-Small	27.27	50.12	85.98	92.96	64.08
PMF+Ours	DINOv1	ViT-Small	26.56	54.31	90.29	96.10	66.82
CD-CLS Zou et al. (2025a)	DINOv1	ViT-Small	27.66	54.69	91.53	96.27	67.54
CD-CLS+Ours	DINOv1	ViT-Small	27.91	56.97	91.83	96.98	68.42

These results demonstrate that class-adaptive characteristic-function matching is complementary to different CD-FSL pipelines and improves query–class comparison without changing their feature backbones or adaptation procedures.

Table 3: Ablation study on 5-way cross-domain few-shot learning.

Method	ChestX		ISIC		EuroSAT		CropDisease		Average	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline	22.57	25.76	33.29	48.63	72.42	88.88	81.02	94.75	52.33	64.51
Fixed Random Frequency	22.73	25.97	34.06	48.90	71.46	88.62	80.21	94.04	52.14	64.38
Ours w/o Class-Adaptive CF Matching	22.76	23.77	34.62	49.42	72.74	87.73	81.07	94.76	52.79	64.41
Ours w/o Stochastic Frequency Generator	21.72	24.64	30.06	43.59	61.49	86.42	66.2	91.94	44.87	61.43
Ours w/o $\mathcal{L}_{\text{KL}}$	22.42	25.89	32.75	49.96	65.20	87.11	76.36	95.11	49.18	64.52
Ours	**22.78**	**26.42**	**34.69**	**51.89**	**73.02**	**89.12**	**82.35**	**96.07**	**53.21**	**65.88**

3.3 Ablation Study

Our central idea is to adapt query–class comparison under domain shift through class-adaptive characteristic-function matching. Its plug-and-play effectiveness is evaluated on three CD-FSL host methods in Tables 1 and 2. Table 3 further examines the overall formulation and its three key components: class-adaptive matching, stochastic frequency generation, and KL regularization. All ablations use ProtoNet with a DINOv1-pretrained ViT-Small backbone.

Overall Matching Formulation. We construct a fixed-random-frequency variant that evaluates characteristic-function responses using a single randomly sampled frequency set shared by all classes and fixed throughout training. This variant preserves characteristic-function matching while removing frequency learning and support conditioning. Averaged across domains, this variant underperforms the full model in both shot settings, demonstrating the benefit of learning and adapting frequency arguments to the target support classes.

Class-Adaptive Matching. *Ours w/o Class-Adaptive CF Matching* replaces the class-specific frequency sets with a globally learnable set shared by all classes. It retains characteristic-function matching and frequency optimization but uses the same comparison rule for every class. Its lower average performance across target domains confirms the importance of adapting frequency arguments to individual support classes.

Stochastic Frequency Generator. *Ours w/o Stochastic Frequency Generator* replaces the stochastic generator with a parameter-matched deterministic alternative. Directly predicting the complete $L \times d$ frequency matrix (e.g., 1024×384) would require a prohibitively large output layer; we therefore adopt a low-rank amortized parameterization for a fair comparison. Despite retaining class-adaptive matching, this variant yields substantially lower average accuracy across target domains, particularly in the 1-shot setting. This highlights the difficulty of deterministically predicting a high-dimensional frequency set from limited support information. In contrast, our stochastic generator models a low-dimensional, support-conditioned distribution, enabling more robust frequency generation under data scarcity.

KL Regularization. *Ours w/o* $\mathcal{L}_{\text{KL}}$ removes the KL term while retaining the other components of the stochastic frequency generator. Its lower average performance across target domains confirms the importance of regularizing the support-conditioned latent distribution.

Table 4: Evaluating our method across pretrained backbones on 5-way CD-FSL.

Method	Pre-train	ChestX		ISIC		EuroSAT		CropDisease		Average	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline	DINOv1	22.57	25.76	33.29	48.63	72.42	88.88	81.02	94.75	52.33	64.51
Ours	DINOv1	**22.78**	**26.42**	**34.69**	**51.89**	**73.02**	**89.12**	**82.35**	**96.07**	**53.21**	**65.88**
Baseline	DINOv3	22.10	24.74	34.48	49.88	70.43	86.85	90.18	96.40	54.30	64.47
Ours	DINOv3	**22.72**	**26.20**	**35.24**	**53.73**	**72.57**	**90.17**	**90.73**	**97.33**	**55.32**	**66.86**
Baseline	CLIP	20.82	22.35	30.21	41.11	65.61	81.92	77.90	92.83	48.64	59.55
Ours	CLIP	**21.42**	**23.50**	**33.15**	**49.32**	**68.27**	**87.44**	**82.55**	**95.57**	**51.35**	**63.96**

3.4 Further Analysis

Robustness Across Pretrained Backbones. We evaluate whether the proposed module depends on a particular feature backbone by extending ProtoNet to DINOv3 Siméoni et al. (2025) and CLIP Radford et al. (2021). We use ViT-Small for DINOv1 Caron et al. (2021) and ViT-Base for DINOv3 and CLIP. As shown in Table 4, our method improves every target domain under all backbones and shot settings. These results demonstrate compatibility across different backbone scales and both self-supervised and vision-language pretraining.

Robustness Across Support Sizes. Beyond the standard 1-shot and 5-shot settings, we evaluate 3-shot, 10-shot, and 15-shot episodes using ProtoNet with a DINOv1-pretrained backbone. As shown in Table 5, our method improves the baseline at every support size and on all target domains. The 1-shot improvement reflects the benefit of class-specific frequency arguments, while additional support examples enable the empirical characteristic representation to aggregate multiple feature responses. These results demonstrate robustness across a broad range of support sizes.

Table 5: Evaluating our method across support sizes.

	ChestX					ISIC				
Method	**1-shot**	**3-shot**	**5-shot**	**10-shot**	**15-shot**	**1-shot**	**3-shot**	**5-shot**	**10-shot**	**15-shot**
Baseline	22.57	24.97	25.76	27.88	28.85	33.29	44.11	48.63	53.92	56.70
Ours	**22.78**	**25.52**	**26.42**	**28.90**	**29.97**	**34.69**	**46.47**	**51.89**	**57.77**	**60.53**
	EuroSAT					**CropDisease**				
Method	**1-shot**	**3-shot**	**5-shot**	**10-shot**	**15-shot**	**1-shot**	**3-shot**	**5-shot**	**10-shot**	**15-shot**
Baseline	72.42	86.15	88.88	91.63	92.39	81.02	92.22	94.75	96.16	96.91
Ours	**73.02**	**86.21**	**89.12**	**92.08**	**92.97**	**82.35**	**93.36**	**96.07**	**97.79**	**98.45**

Efficiency Analysis. We evaluate efficiency on 5-way 1-shot episodes using PMF Hu et al. (2022), a ViT-Small backbone, and an NVIDIA RTX 3090 GPU. Table 6 reports the time per episode, including adaptation and inference, total parameters, and average accuracy across the four target domains. Our module adds only 0.0753M parameters, corresponding to a 0.35% increase, while improving average accuracy by 2.40 percentage points. The runtime increases from 9.56 to 11.71 seconds per episode, indicating a small parameter overhead and a moderate computational cost.

Table 6: **Evaluating the efficiency of our method.**

Method	**Time (S)** ↓	**Params (M)** ↓	**Acc. (%)** ↑
PMF Hu et al. (2022)	9.56	21.6657	50.91
PMF+Ours	11.71	21.7410	53.31

4 Related Work

Cross-Domain Few-Shot Learning. Existing CD-FSL methods mitigate domain shift through style augmentation Fu et al. (2023); Li et al. (2025), masked or reconstruction-based learning Walsh & Osman (2023); Ma et al. (2025), representation and optimization regularization Zou et al. (2024; 2025b), and prototype or pretrained-model adaptation Zhou et al. (2023; 2024; 2025); Zou et al. (2025a). These methods mainly improve feature transfer or target-domain adaptation, while many retain prototype-based classification under a shared Euclidean or cosine metric. Our work is complementary, using each target support class to adapt the query–class comparison itself.

Adaptive Metrics and Meta-Generation. Metric-based few-shot methods typically classify queries using prototypes in a learned embedding space Snell et al. (2017); Vinyals et al. (2016), while TADAM Oreshkin et al. (2018) introduces task-conditioned metric adaptation. Generative meta-learning further uses support examples to produce classifier weights, feature transformations, or adaptation parameters Gidaris & Komodakis (2019); Tseng et al. (2019); Rusu et al. (2019); Gordon et al. (2019). Our generator instead produces frequency arguments separately for each candidate class, allowing the same query to be evaluated through different support-conditioned comparisons.

Characteristic Functions and Fourier Features. Empirical characteristic functions provide sample-based Fourier representations and have been applied to generative modeling and distribution comparison Feuerverger & Mureika (1977); Ansari et al. (2020). Random Fourier features use finite spectral projections to approximate shift-invariant kernels Rahimi & Recht (2007). Most closely related, MetaVRF Zhen et al. (2020) and MetaKernel Du et al. (2022) learn task-specific spectral distributions for few-shot kernel learning. In contrast, our method generates a separate frequency set for each candidate class and measures query compatibility through the induced change in its empirical characteristic-function representation. This class-specific formulation adapts Fourier comparison to CD-FSL and can directly replace the shared metric in existing pipelines.

5 Conclusion

This work investigated how query–class comparison can be adapted when few-shot recognition is transferred across domains. We developed Class-Adaptive Characteristic Function Matching, which replaces a shared point-based metric with support-conditioned comparisons defined by class-specific frequency arguments. Experiments validate the SoTA performance of our plug-and-play framework.

AI USE STATEMENT

In this work, we used generative AI tools solely for manuscript preparation, including translation, language polishing, and refinement of wording, sentence structure, and presentation. We did not use generative AI to formulate mathematical claims or proofs, to generate experimental data, or to produce any reported results, and the remaining disclosure tasks are not applicable to this work.

All AI-assisted text was reviewed and revised by the authors. The datasets, source images, and ground-truth labels used in this work were not generated by AI. We take full responsibility for the final content of this work, including its text, claims, results, and artifacts.

ETHICS STATEMENT

This work uses only publicly available datasets collected under their original licenses; no human-subject experiments were conducted. Models trained on natural-image source domains may inherit pretraining biases and perform unevenly across target domains or patient populations. Our method improves average accuracy but does not by itself guarantee fairness. We declare no conflicts of interest or external sponsorship.

REPRODUCIBILITY STATEMENT

The main text specifies the backbones, the class-adaptive matching module, the stochastic frequency generator, the query-induced score, and the training objective. Implementation details, the optimizer, and the episode sampling protocol, are given in the Experimental Setup and Appendix, together with sensitivity analyses, confidence intervals, and efficiency results. Source code and evaluation scripts will be released upon acceptance.

REFERENCES

Abdul Fatir Ansari, Jonathan Scarlett, and Harold Soh. A characteristic function approach to deep implicit generative modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7478–7487, 2020.

Punam Bedi, Pushkar Gole, and Sumit Kumar Agarwal. 18 using deep learning for image-based plant disease detection. *Internet Of things and machine learning in agriculture*, pp. 369–402, 2021.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

Yingjun Du, Haoliang Sun, Xiantong Zhen, Jun Xu, Yilong Yin, Ling Shao, and Cees GM Snoek. Metakernel: Learning variational random features with limited labels. *IEEE transactions on pattern analysis and machine intelligence*, 46(3):1464–1478, 2022.

Andrey Feuerverger and Roman A Mureika. The empirical characteristic function and its applications. *The annals of Statistics*, pp. 88–97, 1977.

Yuqian Fu, Yu Xie, Yanwei Fu, and Yu-Gang Jiang. Styleadv: Meta style adversarial training for cross-domain few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24575–24584, 2023.

Spyros Gidaris and Nikos Komodakis. Generating classification weights with gnn denoising autoencoders for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 21–30, 2019.

Jonathan Gordon, John Bronskill, Sebastian Nowozin, Matthias Bauer, and Richard E Turner. Meta-learning probabilistic inference for prediction. In *7th International Conference on Learning Representations, ICLR 2019*, 2019.

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*, pp. 124–141. Springer, 2020.

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. 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, 2019.

Shell Xu Hu, Da Li, Jan Stühmer, Minyoung Kim, and Timothy M Hospedales. Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9068–9077, 2022.

Wenqian Li, Pengfei Fang, and Hui Xue. Svasp: Self-versatility adversarial style perturbation for cross-domain few-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 15275–15283, 2025.

Ran Ma, Yixiong Zou, Yuhua Li, and Ruixuan Li. Reconstruction target matters in masked image modeling for cross-domain few-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 19305–19313, 2025.

Boris Oreshkin, Pau Rodríguez López, and Alexandre Lacoste. Tadam: Task dependent adaptive metric for improved few-shot learning. In *Advances in Neural Information Processing Systems*, pp. 721–731, 2018.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007.

Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. In *International Conference on Learning Representations*, 2019.

Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pp. 4077–4087, 2017.

Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018.

Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang, and Ming-Hsuan Yang. Cross-domain few-shot classification via learned feature-wise transformation. In *International Conference on Learning Representations*, 2019.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in neural information processing systems*, pp. 3630–3638, 2016.

R Walsh and I Osman. Masked embedding modeling with rapid domain adjustment for few-shot image classification. *IEEE Transactions on Image Processing*, 32:4907–4920, 2023.

Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. 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 conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.

Xiantong Zhen, Haoliang Sun, Yingjun Du, Jun Xu, Yilong Yin, Ling Shao, and Cees Snoek. Learning to learn kernels with variational random features. In *International Conference on Machine Learning*, pp. 11409–11419. PMLR, 2020.

Fei Zhou, Peng Wang, Lei Zhang, Wei Wei, and Yanning Zhang. Revisiting prototypical network for cross domain few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20061–20070, 2023.

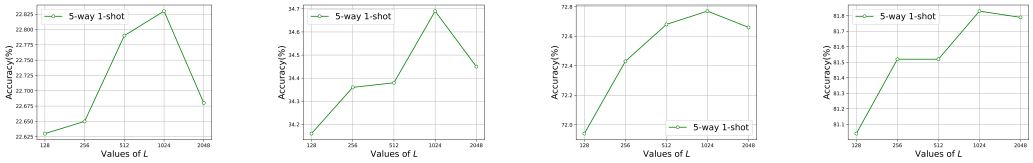
Fei Zhou, Peng Wang, Lei Zhang, Zhenghua Chen, Wei Wei, Chen Ding, Guosheng Lin, and Yanning Zhang. Meta-exploiting frequency prior for cross-domain few-shot learning. *Advances in Neural Information Processing Systems*, 37:116783–116814, 2024.

Fei Zhou, Peng Wang, Lei Zhang, Wei Wei, Chen Ding, Guosheng Lin, and Yanning Zhang. Towards effective foundation model adaptation for extreme cross-domain few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4582–4593, 2025.

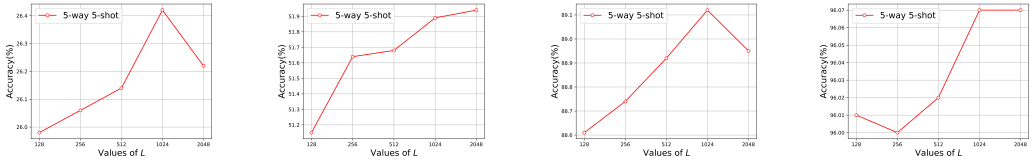
Yixiong Zou, Yicong Liu, Yiman Hu, Yuhua Li, and Ruixuan Li. Flatten long-range loss landscapes for cross-domain few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23575–23584, 2024.

Yixiong Zou, Shuai Yi, Yuhua Li, and Ruixuan Li. A closer look at the cls token for cross-domain few-shot learning. *Advances in Neural Information Processing Systems*, 37:85523–85545, 2025a.

Yunyu Zou, Yishu Liu, Jun Liang, and Bingzhi Chen. Sg-fsl: Cross-domain few-shot learning with style-decoupled augmentation and gradient-conflict adjustment. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 8759–8768, 2025b.

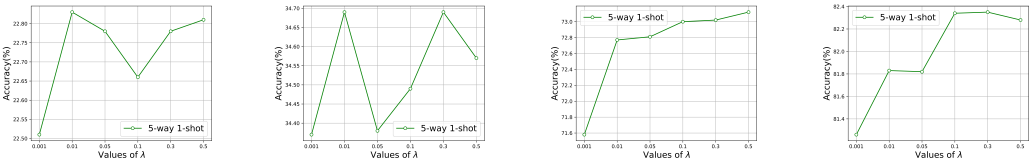


(a) ChestX (b) ISIC (c) EuroSAT (d) CropD.

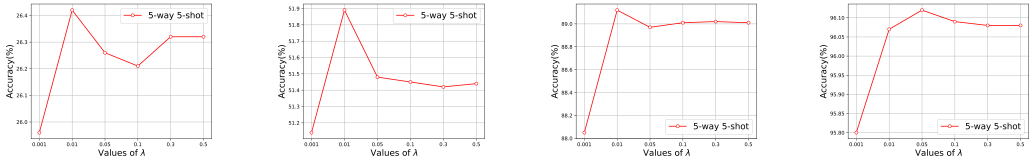


(e) ChestX (f) ISIC (g) EuroSAT (h) CropD.

Figure 3: Sensitivity to the number of frequency arguments.



(a) ChestX (b) ISIC (c) EuroSAT (d) CropD.



(e) ChestX (f) ISIC (g) EuroSAT (h) CropD.

Figure 4: Sensitivity to the KL coefficient.

A Appendix

Overview. This appendix presents additional analyses of the proposed method, including sensitivity to the frequency count L and KL coefficient λ, confidence-interval evaluation, and a discussion of limitations.

Number of Frequency Arguments L. We evaluate $L \in \{128, 256, 512, 1024, 2048\}$ under both 1-shot and 5-shot settings. As shown in Fig. 3, performance generally improves as L increases to 1024, while further increasing it to 2048 yields limited or inconsistent gains. Considering both accuracy and computational cost, we set $L = 1024$ for all domains and shot settings.

KL Coefficient λ. We evaluate $\lambda \in \{0.001, 0.01, 0.05, 0.1, 0.3, 0.5\}$ under both shot settings. As shown in Fig. 4, $\lambda = 0.3$ achieves the best average 1-shot performance, whereas $\lambda = 0.01$ performs best in the 5-shot setting. The stronger regularization required in the 1-shot setting reflects the limited support information available for conditioning the latent distribution. We use the selected value consistently across all target domains within each shot setting.

Confidence-Interval Analysis. To assess robustness to episodic sampling, we report mean accuracy with 95% confidence intervals over 600 episodes. As shown in Table 7, our method improves the mean accuracy over the baseline (i.e., ProtoNet) across all domains and both shot settings, while maintaining comparable confidence-interval widths. This indicates that the improvements do not come at the cost of increased episodic variability.

Limitations. Our method increases the per-episode runtime from 9.56 to 11.71 seconds due to class-specific frequency generation and query–class evaluation. Moreover, visual support features

Table 7: Confidence-interval comparison on 5-way CD-FSL. Results are reported as mean accuracy (%) with 95% confidence intervals over 600 episodes.

Method	ChestX		ISIC		EuroSAT		CropDisease		Average	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline	22.57 $\pm$ 0.41	25.76 $\pm$ 0.45	33.29 $\pm$ 0.56	48.63 $\pm$ 0.60	72.42 $\pm$ 0.50	88.88 $\pm$ 0.45	81.02 $\pm$ 0.51	94.75 $\pm$ 0.41	52.33	64.51
Ours	**22.78** $\pm$ 0.42	**26.42** $\pm$ 0.44	**34.69** $\pm$ 0.58	**51.89** $\pm$ 0.51	**73.02** $\pm$ 0.49	**89.12** $\pm$ 0.41	**82.35** $\pm$ 0.47	**96.07** $\pm$ 0.34	**53.21**	**65.88**

may provide limited conditioning in extremely low-shot settings. We leave efficient frequency generation and the incorporation of semantic or vision–language priors for future work.