

DDSF: Robust Few-Shot Learning via Disentangled Subspaces with Determinantal Point Process

Xulun Ye^{1*} Yifan Mei^{1*} Kun Zhou^{2†} Zelei Wu¹ Jieyu Zhao^{1†}

Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China

²School of Architecture & Urban Planning, Shenzhen University, Shenzhen, China

{yexulun, 2411100090, zhao_jieyu}@nbu.edu.cn, zhoukun@szu.edu.cn, zeleiwu214@gmail.com

Abstract

The performance of mean-based prototypical methods in few-shot learning is frequently compromised by noise and hard positives, where entangled feature representations cause prototype instability. We present a novel “Filter-Repair-Expand” framework grounded in Determinantal Point Process (DPP) theory. The method leverages DPP as its core logic, employing it to estimate sample confidence to filter anomalous samples from the initial set, guide a diffusion process via volume-maximization to enhance the sample representation, and subsequently maximize the volume of synergistic disentangled subspaces, constructing robust and diverse prototype subspaces. Experimental results establish new state-of-the-art performance on multiple benchmarks, demonstrating significant gains in few-shot learning robustness. Code is available at <https://github.com/24sleeping90/DDSF>

1. Introduction

Few-Shot Learning (FSL) aims to enable models to learn new concepts and generalize from a minimal number of labeled examples [8, 34, 37–39, 42], emulating a human-like ability to learn from few instances. Within the FSL paradigm, prototype-based metric learning methods [37, 38, 42] have become a dominant benchmark. However, their core assumption—that the support set is high-quality and representative—is often impractical in real-world scenarios [18, 21, 53]. Inevitable noise in data acquisition, such as label noise or out-of-distribution (OOD) samples, poses a core challenge for robust FSL [18, 21, 30, 53]: the construction of reliable class representations from heavily corrupted support sets.

To address this challenge, current mainstream approaches [21, 53] widely adopt a “Filter-Compress” strat-

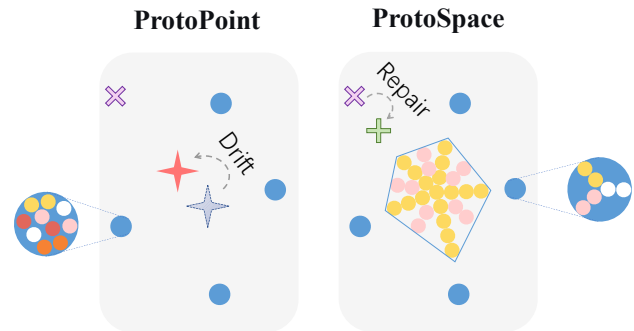


Figure 1. Left: The standard point prototype (star) suffers from drift via outliers (cross) and feature conflation (magnified as mixed colors representing diverse features), causing Representation Confusion. Right: Our paradigm repairs the outlier (plus) and disentangles features (magnified as separated shared and unique colored dots) to build a robust subspace (volume), resolving both issues.

egy. This strategy identifies and removes noisy samples to compress a robust point prototype, a fragile concept (Fig. 1, left). This is manifested in two forms: explicit filtering, which discards samples via outlier detection (e.g., DETA++ [53]), and implicit filtering, which down-weights suspicious samples using attention mechanisms (e.g., TraNFS [21]) and attentive feature regularization [55]). Nevertheless, the robustness of these methods still faces challenges when confronted with more complex or severe noise [4, 25, 29, 56], suggesting that this paradigm has inherent limitations.

The inherent limitations of these methods stem from two intertwined structural flaws. The first is Representation Confusion, intrinsic to prototype networks, where high-variance, non-discriminative shared features are confounded with low-variance, discriminative unique features and compressed into a single mean point. The second is Information Destruction, induced by the Filter-Compress strategy itself. In its pursuit of purity, this strategy indiscriminately discards hard positives [35] along with true

*Equal contribution.

†Corresponding author.

noise, as they are often indistinguishable in a data-scarce setting. This practice not only results in the loss of critical discriminative information but also exacerbates Representation Confusion, which results in the prototype being dominated by the remaining common, high-variance shared features.

To address these challenges, we propose a novel “Filter-Repair-Expand” paradigm, which is grounded in the unifying theory of Determinantal Point Processes (DPP) [17] and implemented through our designed Disentangled-Subspace Framework (DDSF). The framework filters by using DPP’s probabilistic inference to locate high-interference samples, preserving rather than discarding information. It then repairs these located samples by utilizing the DPP volume maximization objective to guide a novel diffusion model, which transforms potential corruptions into effective features and resolves Information Destruction, and finally, after obtaining a purified feature set, DDSF expands the representation via a synergistic disentanglement process: it compels a shared subspace to capture cross-class commonalities via reconstruction, and then applies the DPP objective to expand the volume of the remaining, class-specific unique features. This process constructs a diverse and robust prototype subspace instead of a fragile point (visualized in Fig. 1, right), thereby resolving Representation Confusion.

Our main contributions are summarized as follows:

- We propose the “Filter-Repair-Expand” paradigm and design DDSF, a novel framework to address prototype instability by repairing, rather than discarding, corrupted samples.
- We introduce a synergistic application of Determinantal Point Process (DPP) theory, leveraging it for three distinct purposes: (1) probabilistic anomaly identification (Filter), (2) volume-maximizing guidance for diffusion-based sample repair (Repair), and (3) volume-expansion for constructing disentangled subspaces (Expand).
- Extensive experiments on multiple noise benchmarks demonstrate that DDSF significantly outperforms existing state-of-the-art methods.

2. Related Work

2.1. Few-Shot Learning

Research in few-shot learning primarily follows two dominant directions [39]. The first is optimization-based meta-learning, exemplified by MAML [8], which learns a model initialization optimized for rapid adaptation to new tasks [34]. The second, metric-based learning, aims to learn a feature embedding space where classification relies on distance to class representations. Prototypical Networks [37] established the efficacy of using a simple class mean (a point prototype) as this representation. Subsequent works, such as Matching Networks [42] and Relation Net-

works [38], explored more complex, learned metrics. While effective in clean settings, the performance of these foundational methods degrades significantly when the support set is contaminated [10], as their core representations are easily corrupted by outliers.

2.2. Robust Few-Shot Learning

Addressing noise in the support set is a critical challenge for practical FSL [18, 21, 53]. Contamination from label noise [21], out-of-distribution (OOD) samples [18, 30], or complex backgrounds [28] can severely destabilize class prototypes. The dominant strategy to combat this, the “Filter-Compress” paradigm, focuses on purifying the support set before representation construction. This manifests in two forms. Explicit filtering methods employ outlier detection modules to identify and discard anomalous samples [53]. For example, DETA++ [53] uses a contrastive relevance aggregation to compute sample weights and filter noise. Implicit filtering methods use soft-weighting mechanisms, such as attention modules [21] or attentive feature regularization [55], to down-weight the influence of suspected outliers. Despite their progress, these methods face a difficult trade-off in data-scarce scenarios: the filtering modules often cannot distinguish true noise from valuable hard positives [35]. This leads to Information Destruction, the inadvertent rejection of critical samples. This practice not only fails to resolve Representation Confusion but actively exacerbates it by leaving the prototype dominated by common, high-variance shared features [28].

2.3. Generative Models in Few-Shot Learning

Another line of research leverages generative models to augment the limited support data, thereby mitigating overfitting [31, 43, 49]. Early approaches focused on simple transformations or feature-space interpolations [31, 48]. More recent works have employed adversarial task augmentation to create more challenging and diverse meta-training tasks [43, 45]. With the advent of diffusion models [14], recent methods have begun to explore their powerful generative capabilities for FSL. MetaDiff [49], for instance, uses a conditional diffusion model for novel sample generation. Other works leverage the generative priors of large foundation models to synthesize new features or classifiers [54]. These methods primarily use generation to increase the quantity of data.

2.4. Subspace and Disentangled Representations

The vulnerability of point prototypes [37] has motivated a shift toward richer class representations that better capture intra-class variance. This includes modeling classes as Gaussian distributions [9] or as adaptive, low-rank subspaces [1, 36]. Methods like DeepEMD [50] compute structural distances between sets of features. These ap-

proaches offer a more robust class representation than a single mean vector. Parallel efforts in representation learning have demonstrated the value of disentanglement [2, 13], separating common, shared features from unique, discriminative factors.

3. Methodology

To implement the “Filter-Repair-Expand” paradigm, we propose the Determinantal Point Process-guided Disentangled Subspace Framework (DDSF) to construct robust class representations from a noisy support set $\mathcal{S} = \{x_i\}_{i=1}^K$ (where x_i is an individual sample) and a query set $\mathcal{Q}$, ultimately maximizing classification accuracy.

3.1. Framework Overview

As illustrated in Figure 2, our proposed DDSF implements the Filter-Repair-Expand paradigm through three synergistic modules that process the support set $\mathcal{S}$ sequentially. First, all samples pass through the DPP-based Anomaly ID module (Sec 3.3.1), which leverages DPP-based anomaly scores (Sec. 3.3.1) to identify a high-interference candidate set $\mathcal{S}_a$. Second, this anomalous set $\mathcal{S}_a$ is processed by the DPP-Guided Repair module (Sec 3.3.2), which utilizes a DPP-gradient (Eq. 5) to transform corruptions into effective features. Finally, the resulting purified feature set $\mathcal{S}_f$ (i.e., $\mathcal{S}_k \cup \mathcal{S}_r$, where $\mathcal{S}_k$ is the known-clean set and $\mathcal{S}_r$ is the repaired set) enters the Disentangled Subspace Construction module (Sec. 3.4). This module is driven by two cooperative objectives ($\mathcal{L}_{sh}$ for capturing commonality via feature reconstruction, and $\mathcal{L}_{un}$ for maximizing the unique subspace volume) to build the robust shared ($\mathcal{V}_{sh}$) and unique ($\mathcal{V}_{un}$) subspaces for classification.

3.2. Determinantal Point Processes (DPP)

DPP, a class of probabilistic models that simultaneously capture quality and diversity [17], serves as the theoretical foundation for DDSF. Given a kernel matrix $\mathbf{L}$, typically defined as the Gram matrix of features:

$$\mathbf{L}_{ij} = \langle \phi(x_i), \phi(x_j) \rangle \quad (1)$$

where $x_i, x_j \in \mathcal{S}$, and $\phi(\cdot)$ is the feature embedding. If we let $\mathbf{F}$ be the matrix whose rows are the features $\{\phi(x_i)\}_{i=1}^M$, the kernel can be expressed as $\mathbf{L} = \mathbf{F}\mathbf{F}^T$. DPP defines the probability of sampling a subset $\mathcal{Y}$ of size K from M items [16, 17, 22]:

$$P(\mathcal{Y}) = \frac{\det(\mathbf{L}_{\mathcal{Y}})}{\det(\mathbf{L} + \mathbf{I})} \quad (2)$$

where $\mathbf{L}_{\mathcal{Y}}$ is the principal submatrix of $\mathbf{L}$ indexed by $\mathcal{Y}$. The numerator, $\det(\mathbf{L}_{\mathcal{Y}})$, is geometrically equivalent to the squared volume of the parallelotope spanned by the feature vectors in $\mathcal{Y}$ [17].

However, directly applying DPP’s volume maximization to raw features has an inherent limitation. As the determi-

nant value is sensitive to feature norms, its quality term is prone to be dominated by high-magnitude noise or high-variance shared background features ($\mathbf{f}_{sh}$). This bias can lead the model to favor a polluted representation. To circumvent this issue, the DDSF framework applies the DPP principle to purified feature subspaces.

3.3. Filter and Repair

3.3.1. Probabilistic Anomaly Identification

Unlike conventional methods that discard outliers, the filter stage identifies candidate samples for repair, thus avoiding potential information loss. This mechanism leverages dual information dimensions embedded within the Determinantal Point Process (DPP) kernel $\mathbf{L}$ (Eq. 1): diversity and quality. An anomaly is characterized by high diversity (orthogonality to the clean manifold) and low quality (weak cohesion with the class concept).

Sample uniqueness (diversity) is quantified via the DPP marginal probability p_k :

$$p_k = [\mathbf{L}(\mathbf{L} + \mathbf{I})^{-1}]_{kk} \quad (3)$$

where $\mathbf{L}$ is the kernel matrix defined in Eq. 1, and $\mathbf{I}$ is the identity matrix. Concurrently, sample representativeness (quality) is assessed via a cohesion score q_k , which measures the aggregate affinity of a sample with other members in its set:

$$q_k = \sum_{j \neq k} (\mathbf{L}_{kj})^2 \quad (4)$$

where $\mathbf{L}_{kj}$ denotes the similarity between sample k and j .

To robustly distinguish anomalies from hard positives (which may exhibit high p_k but also high q_k), these metrics are fused into a final anomaly score $s_k = p_k/q_k$. A data-driven thresholding process, governed by a global hyperparameter τ , then partitions the support set $\mathcal{S}$ based on these scores. Samples with a score s_k indicative of an anomaly are assigned to $\mathcal{S}_a$, while the remainder form $\mathcal{S}_k$. $\mathcal{S}_a$ is then forwarded to the repair module.

3.3.2. DPP-Guided Feature Optimization

The core of this module is a parameter-free DDPM [14] reverse process guided by a DPP gradient. This design avoids introducing a large auxiliary network, resulting in a lightweight and flexible optimization process. Its effectiveness stems from the injected guidance loss $\mathcal{L}_g$ (Eq. 5).

$$\mathcal{L}_g(\mathbf{f}_{t-1}) = -\log \det(\mathbf{K}_{\{\mathbf{f}_{t-1}\} \cup \mathcal{S}_k}) \quad (5)$$

$$\boldsymbol{\mu}'(\mathbf{f}_t, t) = \boldsymbol{\mu}(\mathbf{f}_t, t) - \gamma \boldsymbol{\Sigma}_t \nabla_{\mathbf{f}} \mathcal{L}_g \quad (6)$$

where $\mathcal{L}_g$ (Eq. 5) computes the collective volume of the feature-to-be-repaired $\mathbf{f}_{t-1}$ and $\mathcal{S}_k$, with $\mathbf{K}$ being the corresponding Gram matrix. In Eq. 6, $\boldsymbol{\mu}$ is the standard mean of the DDPM reverse process, $\nabla_{\mathbf{f}} \mathcal{L}_g$ is the gradient of the volume loss which points in the direction that maximizes collective diversity. $\boldsymbol{\mu}'$ is the guided mean, and γ is the guidance scale.

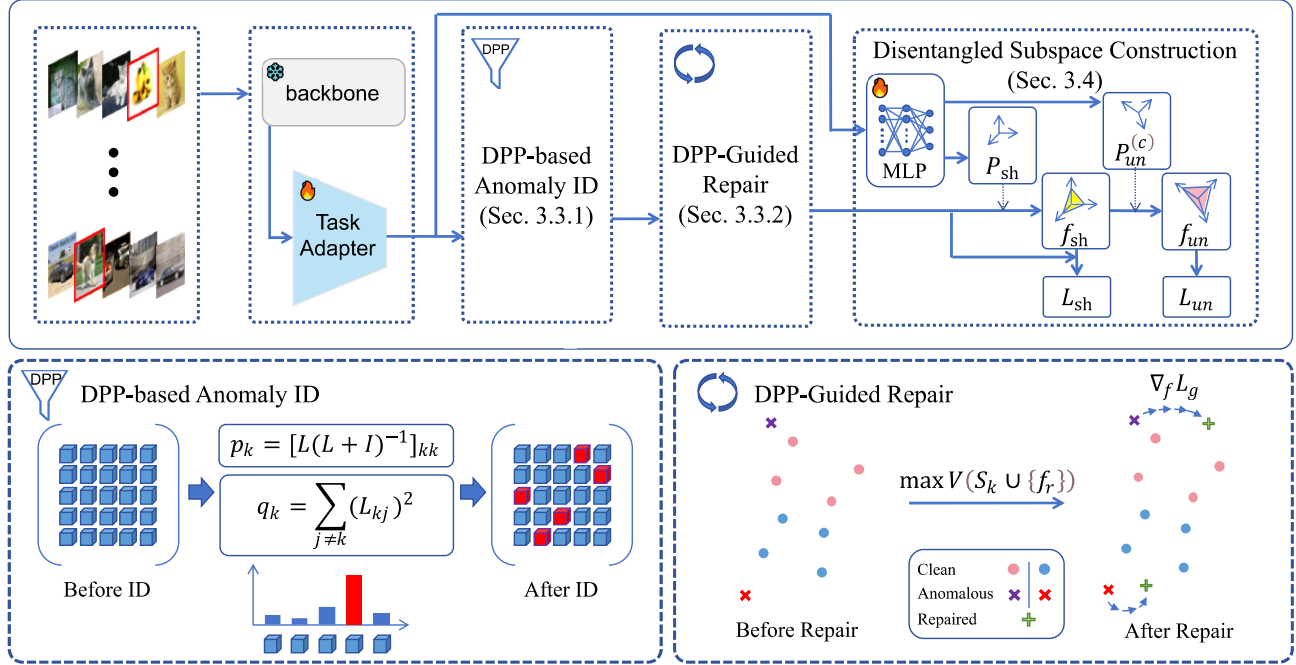


Figure 2. DDSF architecture, illustrating the Filter-Repair-Expand paradigm. It sequentially performs: (1) DPP-based anomaly identification (Sec. 3.3.1), (2) DPP-guided repair (Sec. 3.3.2), and (3) disentangled subspace construction (Sec. 3.4).

In each iteration, this gradient (Eq. 6), derived from the DPP volume objective [17], updates the noisy feature, not only aiming to remove contamination but also optimizing it towards a direction that maximizes the collective diversity with (S_k) . Consequently, the initial anomalous feature is transformed into an effective feature that contributes to the robust class representation.

3.4. Disentangled Subspace Expansion

After obtaining the purified feature set $\mathcal{S}_f$, we proceed to the expand phase. This stage implements our synergistic subspace learning, grounded in the hypothesis (Eq. 7) that a feature $\mathbf{f}$ can be decomposed. To achieve this, we parameterize the shared subspace $\mathcal{V}_{sh}$ and each class-unique subspace $\mathcal{V}_{un}^{(c)}$ as two distinct, learnable linear projection modules (visible as the MLP in Fig. 2), denoted $\mathbf{P}_{sh}$ and $\mathbf{P}_{un}^{(c)}$. Our disentanglement process, which enforces this separation, is then driven by two cooperative objectives that jointly structure the feature space [2, 12, 13].

$$\mathbf{f} = \mathbf{f}_{sh} + \mathbf{f}_{un} + \epsilon \quad (7)$$

Operationally, based on the learnable projections $\mathbf{P}_{sh}$ and $\mathbf{P}_{un}^{(c)}$ defined above, these components are modeled as $\mathbf{f}_{sh} = \mathbf{P}_{sh}\mathbf{f}$ and $\mathbf{f}_{un} = \mathbf{P}_{un}^{(c)}\mathbf{f}$. First, we introduce the Shared Reconstruction Loss ($\mathcal{L}_{sh}$), designed to expand the shared subspace $\mathcal{V}_{sh}$ to form a comprehensive basis that explains

the commonalities across all classes. This objective is formulated as the mean squared reconstruction error between the purified features $\mathbf{f} \in \mathcal{S}_f$ and their projections onto the shared subspace $\mathbf{P}_{sh}$:

$$\mathcal{L}_{sh} = \frac{1}{|\mathcal{S}_f|} \sum_{\mathbf{f} \in \mathcal{S}_f} \|\mathbf{f} - \mathbf{P}_{sh}\mathbf{f}\|_2^2 \quad (8)$$

By minimizing $\mathcal{L}_{sh}$, the optimizer is compelled to expand the expressive capacity of $\mathcal{V}_{sh}$ to cover and reconstruct the universal features present in all classes [2, 15] (e.g., the general concept of dog in a multi-species task).

Concurrently, the Unique Volume Loss ($\mathcal{L}_{un}$) is designed to expand the class-specific unique subspaces $\mathcal{V}_{un}^{(c)}$ to capture the discriminative information not explained by the shared subspace. We first compute the residual features $\mathbf{f}_{res}^{(c)}$ for each class c , representing the portion of the signal not captured by $\mathcal{V}_{sh}$:

$$\mathbf{f}_{res}^{(c)} = \mathbf{f}^{(c)} - \mathbf{P}_{sh}\mathbf{f}^{(c)}, \quad \text{for } \mathbf{f}^{(c)} \in \mathcal{S}_f^{(c)} \quad (9)$$

$\mathcal{L}_{un}$ then operates on these residuals by maximizing their collective volume, applying the DPP objective [17]. This loss is defined as the negative log-determinant of the Gram matrix $\mathbf{G}_{res}^{(c)}$ computed from these residual features:

$$\mathcal{L}_{un} = - \sum_{c=1}^N \log \det(\mathbf{G}_{res}^{(c)} + \epsilon \mathbf{I}) \quad (10)$$

Minimizing $\mathcal{L}_{un}$ is equivalent to maximizing the log-

determinant, which geometrically corresponds to maximizing the volume spanned by the residuals [17]. This objective forces each $\mathcal{V}_{un}^{(c)}$ to expand and model all remaining class-specific, discriminative information (e.g., the unique facial patterns of a husky). These two losses thus operate synergistically: $\mathcal{L}_{sh}$ expands to capture commonality via reconstruction, while $\mathcal{L}_{un}$ expands to capture the unique, discriminative residuals via volume maximization.

3.5. Classification and Joint Optimization

3.5.1. Subspace-Reconstruction Classifier

Based on our disentanglement hypothesis (Eq. 7), the final classifier eschews conventional point-based distances [37, 42]. Given an (undisentangled) query feature $\mathbf{f}_q$, we predict its label $\hat{y}$ by finding the class c whose total reconstruction subspace $\mathbf{P}_{total}^{(c)} = \mathbf{P}_{sh} + \mathbf{P}_{un}^{(c)}$ yields the minimum reconstruction error:

$$\hat{y} = \underset{c \in \{1, \dots, N\}}{\operatorname{argmin}} \|\mathbf{f}_q - (\mathbf{P}_{sh} + \mathbf{P}_{un}^{(c)})\mathbf{f}_q\|_2^2 \quad (11)$$

This classification mechanism directly applies the training paradigm at inference. Instead of proximity, it evaluates $\mathbf{f}_q$'s compatibility with each learned subspace. A correct class c should explain the query with minimal reconstruction error using its combined bases ($\mathbf{P}_{sh}$ and $\mathbf{P}_{un}^{(c)}$).

3.5.2. Joint Training Objective

To optimize the framework, our total loss function $\mathcal{L}_{total}$ combines the primary classification loss with our two disentanglement regularizers:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda_{sh}\mathcal{L}_{sh} + \lambda_{un}\mathcal{L}_{un} \quad (12)$$

$\mathcal{L}_{cls}$ (a cross-entropy loss [7, 11] on logits from negative reconstruction errors) supervises the classifier (Eq. 11). $\mathcal{L}_{sh}$ (Eq. 8) and $\mathcal{L}_{un}$ (Eq. 10) are synergistic regularizers. $\mathcal{L}_{sh}$ captures commonality by minimizing reconstruction error, while $\mathcal{L}_{un}$ captures uniqueness by maximizing residual volume. λ_{sh} and λ_{un} are balancing hyperparameters.

4. Experiments

We perform comprehensive experiments to evaluate the efficacy of our proposed DDSF framework. The evaluation is structured to first benchmark DDSF against state-of-the-art (SOTA) methods (Sec. 4.2), followed by a systematic ablation study (Sec. 4.3) to scrutinize each pillar of the Filter-Repair-Expand paradigm. We conclude with analyses on noise robustness (Sec. 4.4) and hyperparameter sensitivity (Sec. 4.5).

4.1. Experimental Setup

Datasets. Our primary evaluation is conducted on Meta-Dataset [41], a large-scale, high-diversity FSL benchmark. To stringently assess noise robustness, we adopt the OOD-polluted Meta-Dataset protocol [53], which simulates label

contamination by replacing a portion (e.g., 10%, 30%, 50%, 70%) of support samples with out-of-distribution (OOD) samples.

Task Protocol. We adopt the varied-way 10-shot task sampling protocol from DETA++ [53] for fair comparison. In this protocol, the number of classes (ways) per task varies, while the number of support samples is fixed at 10.

Implementation Details and Baselines. All our DDSF models (including ablations) use the Swin Transformer (SwinT) [26] backbone. Our primary comparison is against the current SOTA, DETA++ [53], which utilizes the identical SwinT backbone. We also benchmark DDSF against a comprehensive suite of SOTA methods (see Table 1), whose results are reported from their original publications. These baselines cover the main paradigms in FSL and utilize diverse pre-trained models:

- **Adapter-based methods:** This includes TSA [20], which integrates a residual adapter with URL [23] (w/ResNet-18), and eTT [47], which attaches a self-attention adapter to DINO [3] (w/ViT-Small).
- **Full fine-tuning methods:** We compare against full-FT applied to various modern backbones, including MoCo [12] (w/ResNet-50), CLIP [33] (w/ResNet-50 & ViT-B/32), and DeiT [40] (w/ViT-Small).
- **Other recent methods:** We include metric-learning (ProtoNet [37]), distribution-based (SemFew [51]), generative (MetaDiff [49]), and other strong competitors [6, 19, 24, 44, 46] utilizing backbones like ResNet-12 [4] and Visformer-T [5].

This diverse set of baselines ensures a thorough evaluation of the framework's robustness and generalization.

Evaluation Metrics. We report the mean classification accuracy (ACC, %) across 600 randomly sampled test tasks, along with the 95% confidence interval [41, 53].

4.2. Comparison with State-of-the-Art Methods

We first benchmark DDSF against existing SOTA methods on the OOD-polluted Meta-Dataset. To ensure a fair and direct comparison, our primary analysis focuses on methods using the identical SwinT backbone.

As shown in Table 1, DDSF significantly outperforms all baselines across all noise levels.

- **Fundamental Generalization Advantage:** At 0% noise, DDSF (91.9%) achieves a +5.4% absolute gain over DETA++ (86.5%) on the identical SwinT backbone. This superior performance demonstrates the efficacy of the purify-then-expand representation learning paradigm.
- **Robustness Advantage:** The performance gap widens significantly as noise increases. From 0% to 50% noise,

Table 1. Main SOTA Comparison on OOD-polluted Meta-Dataset. We compare our DDSF against standard baselines and recent state-of-the-art methods. All results are mean accuracy (%). Our method outperforms all competitors, especially in the direct comparison on the same SwinT backbone.

Method	Backbone	0% Noise	10% Noise	30% Noise	50% Noise	70% Noise
TSA (2022) [20]	URL	75.5	71.4	64.4	56.0	36.7
eTT (2022) [47]	DINO	80.9	76.7	67.5	50.4	38.1
full-FT	MoCo	73.0	66.9	62.8	54.1	34.8
full-FT	CLIP	75.3	70.1	64.9	54.0	35.7
full-FT	DeiT	82.5	78.0	75.4	63.8	43.2
AM3 (2019) [46]	ResNet-12	74.1	68.4	62.5	50.8	33.7
FGFL (2023) [6]	ResNet-12	80.1	74.8	69.9	55.8	36.7
MetaDiff (2024) [49]	ResNet-12	74.5	69.2	63.5	51.5	34.1
SemFew (2024) [51]	ResNet-12	81.5	78.6	73.5	61.0	44.9
SemFew-Trans (2024) [51]	SwinT	81.9	77.2	71.2	57.9	40.6
FedFSL-CFRD (2025) [44]	ResNet-12	65.0	60.6	56.8	50.8	42.5
ECER-FSL (2025) [24]	ResNet-12	84.4	78.1	73.3	61.2	42.8
ECER-FSL (2025) [24]	Visformer-T	85.2	79.8	74.3	62.8	43.8
LDC (2025) [19]	CLIP	85.2	79.2	74.2	64.3	44.6
full-FT	SwinT	81.3	75.4	70.7	58.5	41.1
DETA++ (2025) [53]	SwinT	86.5	80.9	77.3	64.2	47.0
Ours (DDSF)	SwinT	91.9 (± 0.43)	89.9 (± 0.50)	85.3 (± 0.65)	75.9 (± 0.88)	61.6 (± 1.09)

DETA++ drops 22.3 points, whereas DDSF drops only 16.0 points. At the 70% noise setting, DDSF (61.6%) maintains a +14.6% lead over the SOTA (47.0%), highlighting the framework’s high resilience.

A more comprehensive analysis covering low-shot settings, real-world cross-domain generalization, model efficiency, and disentanglement visualization can be found in our supplementary materials.

Backbone Generality. Table 2 evaluates DDSF’s performance across five different modern backbones, demonstrating its wide applicability and consistently high performance across all tested architectures. In the clean (0% noise) setting, DDSF’s accuracy ranges from 89.1% (ResNet-50) to a peak of 93.1% (CLIP). This robustness is maintained in the 30% noise setting, where all backbones still perform well above 82.2%. Notably, the transformer-based architectures (CLIP, SwinT, ViT) show particularly strong results, with CLIP achieving the highest performance overall (93.1% at 0% and 86.1% at 30% noise), confirming our framework’s compatibility and effectiveness.

4.3. Ablation Study

We conduct a systematic component-additive ablation study to analyze our DDSF framework. We first validate the Filter-Repair pipeline by dissecting the components of the DPP-based anomaly score (Sec 3.3.1) and the repair module (Sec 3.3.2). We then validate the Expand phase by analyzing the effect of our disentanglement loss (Sec 3.4). Table 3 summarizes the results.

Table 2. Performance of our DDSF framework across various backbone architectures. All results are reported on the OOD-polluted Meta-Dataset under 0% and 30% noise levels.

Backbone	0% Noise	30% Noise
Swin-T [26]	91.9 (± 0.43)	85.3 (± 0.65)
ViT-B/32 (CLIP) [33]	93.1 (± 0.31)	86.1 (± 0.55)
ViT-S (DINO) [3]	90.0 (± 0.53)	84.2 (± 0.71)
ConvNeXt-B [27]	90.4 (± 0.52)	83.9 (± 0.61)
ResNet-50 [11]	89.1 (± 0.49)	82.2 (± 0.76)

Analysis of Filter-Repair Components. We first evaluate the Filter and Repair pipeline by analyzing rows (A) through (E) in Table 3. The baseline model (Idx A) with no filtering yields 77.1% accuracy in the 30% noise setting. We then analyze our DPP-based anomaly score $s_k = p_k/q_k$ (Sec 3.3.1) by dissecting its components. Using only the diversity metric p_k (Eq. 3) provides a +2.3% gain (Idx B). Concurrently, using only the quality/cohesion metric q_k (Eq. 4) yields a similar +2.4% gain (Idx C). Our full proposed score s_k (Idx D), which fuses both metrics, achieves 80.7%, demonstrating that the fusion of quality and diversity (+3.6% gain) is more robust than either metric alone. Finally, activating the DPP Repair (Idx E) module—repairing the outliers identified by s_k —provides the largest boost in this phase of +2.7% (to 83.4%). This underscores the efficacy of repairing features rather than discarding them (Sec 3.3.2).

Table 3. Component-additive ablation study of the DDSF framework. We analyze the contribution of each module defined in Sec 3.3.1 (Filter) and Sec 3.4 (Expand). Models (A-E) serve as the baseline pipeline, while the full model (F) additionally integrates the DPP-based volume loss ($\mathcal{L}_{un}$). Accuracy (%) is reported on 0% (clean) and 30% (noisy) settings.

Idx	Components					Accuracy (%)	
	Filter (p_k)	Filter (q_k)	Filter (s_k)	Repair (DPP)	Expand ($\mathcal{L}_{un}$)	0% Noise	30% Noise
(A)	×	×	×	×	×	84.4	77.1
(B)	✓	×	×	×	×	86.3	79.4
(C)	×	✓	×	×	×	86.2	79.5
(D)	×	×	✓	×	×	87.3	80.7
(E)	×	×	✓	✓	×	89.9	83.4
(F)	×	×	✓	✓	✓	91.9	85.3

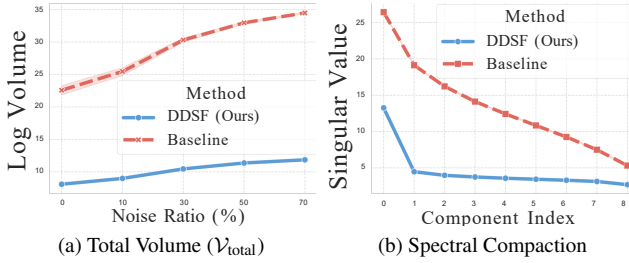


Figure 3. Ablation analysis of the feature space. (a) Log Volume: DDSF’s purified feature volume remains low and stable against noise, unlike the baseline. (b) Spectral Compaction: DDSF’s singular values decay sharply, indicating successful purification and compaction of shared/noise variance.

Analysis of Disentangled Expansion Loss. We next validate the Expand mechanism (Sec 3.4). Models (A) through (E) all utilize a standard orthogonal loss (ortho). To create the Full DDSF (Idx F), we replace this standard ortho loss with our proposed DPP-based Unique Volume Loss ($\mathcal{L}_{un}$). This single change provides another significant performance increase, boosting accuracy by +1.9% (from 83.4% to 85.3%) in the noisy setting. This direct comparison between (E) and (F) confirms that the DPP-driven expansion (maximizing residual volume) outperforms standard orthogonal constraints. Furthermore, this module also provides the largest gain in the clean setting (+2.0% gain from Idx E to F), yielding the peak 0% noise performance of 91.9% and demonstrating its universal benefit.

Figure 3 provides qualitative evidence for this Expand mechanism. The analysis of total feature volume (Fig. 3a) and spectral compaction (Fig. 3b) validates our hypothesis. The baseline exhibits high, unstable feature volume that increases with noise. In contrast, the full DDSF model maintains a low, stable volume and sharp spectral decay indicates the feature space has been successfully purified and compacted by the synergistic Expand losses ($\mathcal{L}_{sh}$ and $\mathcal{L}_{un}$).

Table 4. Performance vs. 30% OOD and 30% SLS Noise (SwinT).

Method	Noise Type	Accuracy (ACC, %)
DETA++ [53]	OOD Noise	77.3
	SLS Noise	75.1
Ours (DDSF)	OOD Noise	85.3
	SLS Noise	83.2

4.4. Robustness Across Noise Types

We evaluate robustness against OOD noise (geometrically separable) and Symmetric Label Swapping (SLS) noise (semantically confusing) [52]. Table 4 shows that DDSF maintains exceptionally high performance (83.2%) even against the more difficult SLS noise. This validates that the DPP-based probabilistic identification mechanism captures semantic inconsistency, not just simple geometric distance.

4.5. Hyperparameter Sensitivity

Impact of Loss Weights (λ_{sh} , λ_{un}). DDSF’s sensitivity to its two core loss weights, λ_{sh} and λ_{un} , is analyzed in Figure 4. The resulting heatmap demonstrates strong robustness across a wide range of hyperparameter settings, showing a large high-performance region (e.g., > 90% accuracy) where λ_{sh} and λ_{un} are appropriately balanced (e.g., $\lambda_{un} \geq 0.1$). Notably, even the lowest recorded accuracy in the grid (at $\lambda_{sh} = 0.01$, $\lambda_{un} = 0.01$) is 88.11%, indicating that the framework’s performance is stable and not reliant on sensitive manual tuning.

Impact of Anomaly Threshold (τ). We analyze the sensitivity of the anomaly threshold τ (Sec 3.3.1), which defines the minimum credible gap width required for data-driven partitioning. A τ value that is too high (e.g., $\tau \geq 0.5$) forces the model to be overly conservative, only partitioning on extremely obvious gaps and thus failing to detect most noise; this pollutes wen. Conversely, a τ value that is too low (e.g., $\tau = 0.1$) is slightly too permissive, potentially over-detecting anomalies on minor fluctuations and deplet-

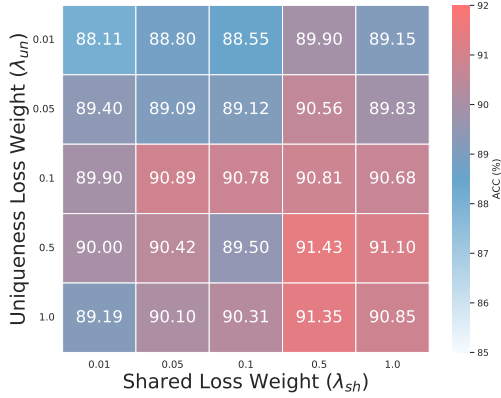


Figure 4. Sensitivity analysis for λ_{sh} and λ_{un} . The large high-performance region (warm colors) demonstrates our framework is robust to the specific choice of weights.

Table 5. Sensitivity analysis for the anomaly threshold τ (Sec 3.3.1). We report mean accuracy (%) on 30% OOD and 30% SLS noise settings.

τ	0.1	0.2	0.3	0.4	0.5	0.6
OOD Noise	84.5	84.7	85.1	84.6	83.9	80.8
SLS Noise	82.6	82.7	83.3	82.9	82.3	79.8

Table 6. Impact of explicit foreground masking. We compare DDSF against a variant (DDSF + Mask) that applies all losses only to foreground features identified by a U-2-Net.

Method	0% Noise	30% OOD	30% SLS
Ours (DDSF)	91.9	85.3	83.2
DDSF + Mask	92.1	85.2	83.5

ing $\mathcal{S}_k$. Table 5 reports the performance as τ is varied. The results show a clear performance peak at $\tau = 0.3$ (85.1% OOD, 83.3% SLS) and demonstrate high robustness within a wide stable range [0.1-0.4]. Performance degrades as the threshold becomes too conservative ($\tau \geq 0.5$).

4.6. Robustness to Background Variance

To evaluate robustness to background variance, we test a “DDSF + Mask” variant. This variant utilizes a pre-trained U-2-Net [32] to generate foreground masks, subsequently applying all framework losses ($\mathcal{L}_{sh}$, $\mathcal{L}_{un}$, and $\mathcal{L}_{cls}$) only to features within these masked regions. As shown in Table 6, the introduction of this explicit masking module yields only negligible performance fluctuations (ranging from -0.1% to +0.3%) across the tested settings. This result suggests that the core disentanglement mechanism is already inherently robust to background variance. The synergistic operation of $\mathcal{L}_{sh}$, which captures common, non-discriminative features (including shared backgrounds) via reconstruction, and $\mathcal{L}_{un}$, which expands the volume of unique, discrimina-

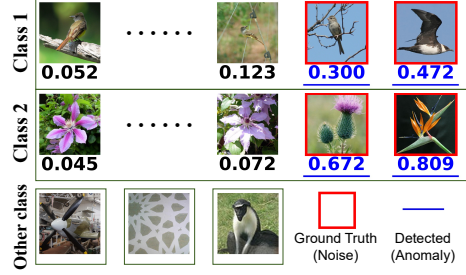


Figure 5. Qualitative analysis of the DPP-based anomaly identification module (Sec 3.3.1). We visualize computed $s_k = p_k/q_k$ scores from a representative 10-shot task, showing a bird class (class 1, top) and flower class (class 2, middle) support set. Samples are sorted by score (left-to-right). Ground-truth noise (red boxes) consistently receives high anomaly scores, validating the filter’s effectiveness.

tive features (i.e., the unique subspace), effectively achieves implicit background invariance, obviating the need for an additional, computationally expensive masking module.

4.7. Qualitative Analysis of Anomaly Identification

To provide qualitative insight into the DPP-based anomaly identification (Sec. 3.3.1), we visualize the filter process on a representative OOD-polluted 10-shot task in Figure 5. For each class, the 10 support samples are strictly sorted by their calculated anomaly score ($s_k = p_k/q_k$) from left to right.

The visualization clearly demonstrates the efficacy of the scoring mechanism. Clean, typical samples (e.g., the birds in class 1 and flowers in class 2) are consistently assigned low s_k scores (low diversity p_k , high quality q_k) and are preserved. Conversely, all ground-truth OOD noise (indicated by red borders) are assigned high scores (e.g., $s_k > 0.30$ in class 1 and $s_k > 0.67$ in class 2). As shown by the blue underlines, the detections highly correspond with the ground truth in both classes. This result validates the robustness of the Filter stage, which provides a reliable set of candidates $\mathcal{S}_a$ for the subsequent Repair module (Sec 3.3.2).

5. Conclusion

We present DDSF, a robust few-shot learning framework introducing a “Filter-Repair-Expand” paradigm to address representation instability from corrupted data. The framework leverages Determinantal Point Process theory to synergistically perform probabilistic anomaly identification, guide a diffusion-based repair, and maximize the volume of disentangled, class-specific subspaces. This approach transforms high-interference samples into effective features and constructs robust prototype subspaces, resolving information destruction and representation confusion of prior methods. Extensive experiments on multiple noise benchmarks demonstrate DDSF significantly outperforms state-of-the-art methods, establishing new standards for FSL robustness.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant 62471266, 62006131, 62071260; the National Natural Science Foundation of Zhejiang Province under Grant LQ24F020001; and the Ningbo Major Research and Development Plan Project (No.2023Z224).

References

- [1] Peyman Bateni, Raghav Goyal, Vaden Masrani, Frank Wood, and Leonid Sigal. Improved few-shot visual classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14493–14502, 2020. 2
- [2] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013. 3, 4
- [3] 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*, pages 9650–9660, 2021. 5, 6
- [4] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*, 2019. 1, 5
- [5] Zhengsu Chen, Lingxi Xie, Jianwei Niu, Xuefeng Liu, Longhui Wei, and Qi Tian. Visformer: The vision-friendly transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 589–598, 2021. 5
- [6] Hao Cheng, Siyuan Yang, Joey Tianyi Zhou, Lanqing Guo, and Bihan Wen. Frequency guidance matters in few-shot learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11814–11824, 2023. 5, 6
- [7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5
- [8] 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. PMLR, 2017. 1, 2
- [9] Stanislav Fort. Gaussian prototypical networks for few-shot learning on omniglot. *arXiv preprint arXiv:1708.02735*, 2017. 2
- [10] Micah Goldblum, Liam Fowl, and Tom Goldstein. Adversarially robust few-shot learning: A meta-learning approach. *Advances in Neural Information Processing Systems*, 33: 17886–17895, 2020. 2
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5, 6
- [12] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020. 4, 5
- [13] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*, 2017. 3, 4
- [14] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 3
- [15] D. P. Kingma and M. Welling. Auto-encoding variational bayes, 2013. *arXiv preprint arXiv:1312.6114*. 4
- [16] Alex Kulesza and Ben Taskar. Learning determinantal point processes for document summarization. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 427–434, 2011. 3
- [17] Alex Kulesza, Ben Taskar, et al. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2–3):123–286, 2012. 2, 3, 4, 5
- [18] D. Le, K. D. Nguyen, K. Nguyen, Q.-H. Tran, R. Nguyen, and B.-S. Hua. Poodle: Improving few-shot learning via penalizing out-of-distribution samples. In *NeurIPS*, pages 23942–23955, 2021. 1, 2
- [19] Shuo Li, Fang Liu, Zehua Hao, Xinyi Wang, Lingling Li, Xu Liu, Puhua Chen, and Wenping Ma. Logits deconfusion with clip for few-shot learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 25411–25421, 2025. 5, 6
- [20] 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. 5, 6
- [21] Kevin J Liang, Samrudhdi B Rangrej, Vladan Petrovic, and Tal Hassner. Few-shot learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9089–9098, 2022. 1, 2
- [22] Hui Lin and Jeff Bilmes. A class of submodular functions for document summarization. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 510–520, 2011. 3
- [23] L. Liu, W. Hamilton, G. Long, J. Jiang, and H. Larochelle. A universal representation transformer layer for few-shot image classification. In *ICLR*, 2021. 5
- [24] Mushui Liu, Fangtai Wu, Bozheng Li, Ziqian Lu, Yunlong Yu, and Xi Li. Envisioning class entity reasoning by large language models for few-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 18906–18914, 2025. 5, 6
- [25] Xinwei Liu, Xiaojun Jia, Jindong Gu, Yuan Xun, Siyuan Liang, and Xiaochun Cao. Does few-shot learning suffer from backdoor attacks? In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 19893–19901, 2024. 1
- [26] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 5, 6

- [27] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022. 6
- [28] Xu Luo, Longhui Wei, Liangjian Wen, Jinrong Yang, Lingxi Xie, Zenglin Xu, and Qi Tian. Rectifying the shortcut learning of background for few-shot learning. *Advances in Neural Information Processing Systems*, 34:13073–13085, 2021. 2
- [29] Xu Luo, Hao Wu, Ji Zhang, Lianli Gao, Jing Xu, and Jingkuan Song. A closer look at few-shot classification again. In *International Conference on Machine Learning*, pages 23103–23123. PMLR, 2023. 1
- [30] Atsuyuki Miyai, Qing Yu, Go Irie, and Kiyoharu Aizawa. Locoop: Few-shot out-of-distribution detection via prompt learning. *Advances in Neural Information Processing Systems*, 36:76298–76310, 2023. 1, 2
- [31] Renkun Ni, Micah Goldblum, Amr Sharaf, Kezhi Kong, and Tom Goldstein. Data augmentation for meta-learning. In *International Conference on Machine Learning*, pages 8152–8161. PMLR, 2021. 2
- [32] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern recognition*, 106:107404, 2020. 8
- [33] 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*, pages 8748–8763. PmLR, 2021. 5, 6
- [34] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International conference on learning representations*, 2017. 1, 2
- [35] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 1, 2
- [36] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtaash Harandi. Adaptive subspaces for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4136–4145, 2020. 2
- [37] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017. 1, 2, 5
- [38] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales. Learning to compare: Relation network for few-shot learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018. 1, 2
- [39] Sebastian Thrun and Lorien Pratt, editors. *Learning to Learn*. Springer, Boston, MA, 1998. 1, 2
- [40] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, pages 10347–10357. PMLR, 2021. 5
- [41] 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. 5
- [42] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016. 1, 2, 5
- [43] Haoqing Wang, Huiyu Mai, Yuhang Gong, and Zhi-Hong Deng. Towards well-generalizing meta-learning via adversarial task augmentation. *Artificial Intelligence*, 317:103875, 2023. 2
- [44] Shanfeng Wang, Jianzhao Li, Zaitian Liu, Yourun Zhang, and Maoguo Gong. Fedfsl-cfrd: Personalized federated few-shot learning with collaborative feature representation disentanglement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 21243–21251, 2025. 5, 6
- [45] Yichen Wu, Long-Kai Huang, and Ying Wei. Adversarial task up-sampling for meta-learning. *Advances in Neural Information Processing Systems*, 35:31102–31115, 2022. 2
- [46] Chen Xing, Negar Rostamzadeh, Boris Oreshkin, and Pedro O O Pinheiro. Adaptive cross-modal few-shot learning. *Advances in neural information processing systems*, 32, 2019. 5, 6
- [47] Chengming Xu, Siqian Yang, Yabiao Wang, Zhanxiong Wang, Yanwei Fu, and Xiangyang Xue. Exploring efficient few-shot adaptation for vision transformers. *arXiv preprint arXiv:2301.02419*, 2023. 5, 6
- [48] Huaxiu Yao, Linjun Zhang, and Chelsea Finn. Meta-learning with fewer tasks through task interpolation. *arXiv preprint arXiv:2106.02695*, 2021. 2
- [49] Baoquan Zhang, Chuyao Luo, Demin Yu, Xutao Li, Huiwei Lin, Yunming Ye, and Bowen Zhang. Metadiff: Meta-learning with conditional diffusion for few-shot learning. In *Proceedings of the AAAI conference on artificial intelligence*, pages 16687–16695, 2024. 2, 5, 6
- [50] Chi Zhang, Yujun Cai, Guosheng Lin, and Chunhua Shen. Deepemd: Differentiable earth mover’s distance for few-shot learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5632–5648, 2022. 2
- [51] Hai Zhang, Junzhe Xu, Shanlin Jiang, and Zhenan He. Simple semantic-aided few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28588–28597, 2024. 5, 6
- [52] Ji Zhang, Lianli Gao, Xu Luo, Hengtao Shen, and Jingkuan Song. Deta: Denoised task adaptation for few-shot learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11541–11551, 2023. 7
- [53] Ji Zhang, Jingkuan Song, Lianli Gao, Nicu Sebe, and Heng Tao Shen. Reliable few-shot learning under dual noises. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 1, 2, 5, 6, 7
- [54] Renrui Zhang, Xiangfei Hu, Bohao Li, Siyuan Huang, Hanqiu Deng, Yu Qiao, Peng Gao, and Hongsheng Li. Prompt, generate, then cache: Cascade of foundation models makes strong few-shot learners. In *Proceedings of the IEEE/CVF*

- conference on computer vision and pattern recognition*, pages 15211–15222, 2023. [2](#)
- [55] Xingyu Zhu, Shuo Wang, Jinda Lu, Yanbin Hao, Haifeng Liu, and Xiangnan He. Boosting few-shot learning via attentive feature regularization. In *Proceedings of the AAAI conference on artificial intelligence*, pages 7793–7801, 2024. [1](#), [2](#)
- [56] 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*, pages 23575–23584, 2024. [1](#)