Neural Computing and Applications
https://doi.org/10.1007/s00521-024-09501-8

ORIGINAL ARTICLE

In-use calibration: improving domain-specific fine-grained few-shot recognition

Minghui Li[1] · Hongxun Yao[1]

Received: 3 April 2023 / Accepted: 14 January 2024

Abstract
Learning to recognize novel visual classes from few samples is challenging but promising. Previous studies have shown that few-shot model tends to overfit and lead to poor generalization performance, which is because it finds a biased distribution based on a few samples. In addition, in agriculture-specific domains, there are more serious research challenges such as imbalanced disease distribution, one-shot representation biases, fine-grained recognition, and granularity shift. As far as we know, this study is the first work on the fine-grained "*Coarse-to-Fine*" few-shot plant disease classification, which classifies "fine-grained novel classes" (***specific to disease severity***) based on "coarse-grained base classes" (***specific to plant species***). A complete two-stage in-use calibration strategy is presented in this paper. Firstly, we propose an attention-based inverse Mahalanobis distance weighted prototype calibration module (AIPCM). By transferring statistics from sample-rich coarse-grained base classes to sample-scarce fine-grained novel classes, we achieve prototype calibration for 1-shot sample and obtain an unbiased distribution in the feature space. Secondly, to generate more reasonable decision boundaries, we propose a prior-driven task-adapted decision boundary calibration module (TDBCM) based on class-covariance metric. The original Euclidean/Cosine distance is updated to the Mahalanobis distance by introducing the prior mean and covariance of the high-dimensional features. Experimental results on several datasets demonstrate that our model outperforms the state-of-the-art (SOTA) models. It can be said that our work is a valuable supplement to the domain-specific agricultural applications.

Keywords Fine-grained classification · Few-shot learning · Visual attention · Prototype calibration · Decision boundary calibration · Class-covariance metric

1 Introduction

Deep learning has been widely used in computer vision tasks such as image classification [1, 2], object detection [3, 4], and image segmentation [5, 6]. These methods typically use large amounts of labeled datasets for training to achieve optimal model performance. However, it is costly to acquire or annotate data to create large training datasets in real-world applications. The model tends to overfit the training sample, resulting in a significant reduction in generalization performance over novel classes when there is too little data. In recent years, research topics focusing on "recognizing new visual categories after seeing some labeled samples" have emerged, which is an important research attempt to make artificial intelligence truly "intelligent". Research on this topic is typically referred to as "*few-shot learning*".

Actually, it is challenging to acquire scarce sample datasets in specific fields. In recent years, smart agriculture has emerged as a strategic development domain in numerous countries. Rapidly detecting plant diseases is crucial for food production safety and sustainability. However, it is extremely limited for annotating plant status data due to low incidence, high-cost of collection, time-consuming, and the need for professional manpower. Agriculture-specific fine-grained few-shot learning emerge

✉ Hongxun Yao
h.yao@hit.edu.cn

Minghui Li
21b903087@stu.hit.edu.cn

[1] Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China

Published online: 19 March 2024

in this situation and become one of the important topics in computer vision research.

The specific domain data exhibits distinct characteristic features, including high disease image collection costs, subtle inter-class differences, higher susceptibility to noise, need for expert annotations, low disease rates, and imbalanced sample distribution. Regarding the aforementioned domain-specific issues, we believe that the fine-grained few-shot disease recognition poses more challenges compared to conventional image classification tasks:

- Limited and imbalanced disease data.
- Distribution bias formed by few-shot samples.
- One-shot disease prototype space construction.
- Fine-grained disease representation and recognition.
- Granularity shift.

Unlike traditional deep learning tasks [1–6], few-shot learning aims to train a model from well-labeled base classes and apply it to unseen novel classes with only a few labeled data. Current few-shot learning methods fall into three categories: optimization-based, metric-based, and generation-based methods.

Optimization-based methods employ gradient descent to rapidly adapt model parameters to novel tasks. For example, the meta-learning algorithm MAML [7] sought an optimal parameter initialization, simplifying fine-tuning for new tasks. The meta-transfer learning algorithm (MTL) [8] employed a deep pre-trained network for transfer learning, retraining a limited parameter subset to ensure the model's transferability without compromising its generalization capabilities. Reference [9] enhanced the robustness of results through joint prediction, which integrated different stages of training upon MTL.

Metric-based methods learn a metric to indicate the similarity relationship between samples. For example, the matching network [10] learned a distance metric to construct the relationships between sample pairs, and to classify unlabeled samples based on the relationship score. Treating the average feature of each class as the prototype, the prototypical network [11] calculated the Euclidean distance between the input sample, and the class prototype, and classified the sample as the closest prototype class. Another approach [12] introduced a prototype correction network using Cosine similarity. It corrected novel class prototypes by addressing intra-class and cross-class deviations. In recent years, most of the SOTA methods have fallen under the umbrella of metric learning.

Generation-based methods aim to increase the number of training samples, which is essentially a data augmentation method. Wang et al. [13] utilized data augmentation to generate diverse samples by altering factors like pose, lighting, and position, thereby enlarging the training dataset. Park and Han et al. [14] transferred the feature variations from base to novel classes by sharing factors across categories, reducing novel class representation bias. Recent research by Xian et al. [15] revealed that the feature space has a lower dimensionality, and processing on sample features can reduce bias. Subsequently, Liu et al. [16] assumed that the features of each category are independent and follow a Gaussian distribution in long-tailed data classification, by which the feature representation of the tail data can be enriched using the intra-class variance in the head data.

Based on the above research, it can be found that most of the previous work focus on developing more powerful models, but pay little attention to the properties and distributions of the data itself. Leveraging prior distributions from base classes and applying them to novel classes sharing akin distributions can enhance the generalization ability of few-shot learning models. Some existing data distribution-based models [16, 17], on the one hand, mainly select the most similar categories from the base classes to adjust the novel class data features, inadvertently disregarding valuable information from alternate categories. On the other hand, the calibrated novel class distribution is mainly used for generating samples within the distribution. While in the final classifier, only simple logistic regression is used for processing, which is essentially a data augmentation method.

In this paper, we explore a novel work on fine-grained few-shot plant disease classification for the smart agriculture field, which classifies "fine-grained novel classes" (***specific to disease severity***) based on "coarse-grained base classes" (***specific to plant species***). First, we propose an attention-based inverse Mahalanobis distance [18] weighted prototype calibration module (AIPCM). The extracted high-dimensional features are processed by power transformation, and similarity weighting is applied to the base classes to calibrate the novel prototype. To obtain an unbiased distribution, we transfer statistics from sample-rich coarse-grained base classes to sample-scarce fine-grained novel classes. In this way, the prototype and distribution calibration for 1-shot sample is achieved. To obtain more reasonable decision boundaries, we introduce a prior-driven task-adapted decision boundary calibration module (TDBCM) based on class-covariance metric. This method is inspired by previous similar works [19–21], introducing class covariance distance as a fundamental metric function. The Mahalanobis distance considers inter-feature correlations through the class-covariance matrix, effectively addressing the Euclidean [11]/Cosine [22, 23] distance's insensitivity to the distribution of intra-class samples with respect to their prototypes. Unlike previous methods, TDBCM fully considers prior distributions from base classes and performs episode-based intra- and inter-class co-computation of the covariance matrix. For few-

shot tasks in specific domains, our approach further optimizes the decision boundary by incorporating prior mean and variance information. Our two-stage calibration procedure (depicted in Fig. 1) yields a precise few-shot nonlinear classifier through prototype and metric function calibration. Finally, we apply our approach to several domain-specific fine-grained datasets to tackle real-world novel or rare plant disease classification tasks.

To summarize, the main contributions of this paper are as follows:

- We propose an attention-based inverse Mahalanobis distance weighted prototype calibration module (AIPCM). On the one hand, more local fine-grained features are discovered using the focal region localization mechanism. On the other hand, compared with existing prototype-based few-shot studies, the proposed method can fully utilize statistics from all base classes, and obtain more accurate novel classes prototype representation through inverse Mahalanobis distance weighting.

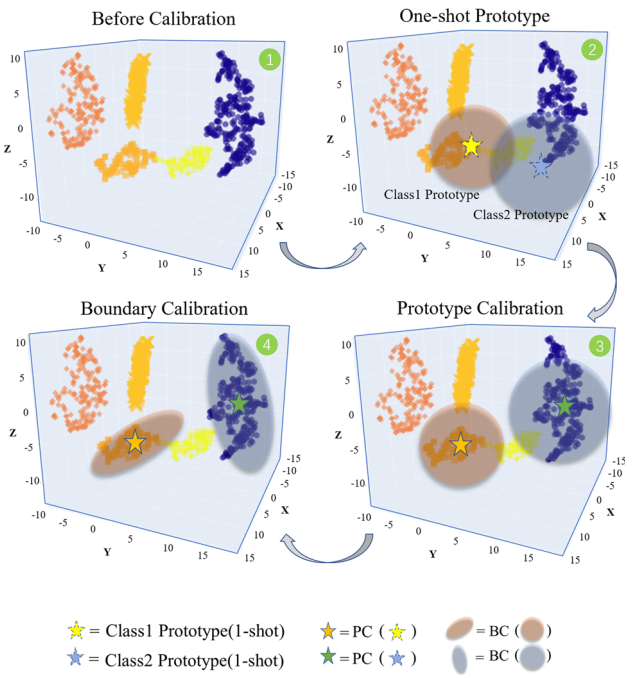


Fig. 1 Complete two-stage in-use calibration strategy: 1-shot prototype calibration and prior-driven task-adapted decision boundary calibration. PC(·) denotes prototype calibration and BC(·) denotes decision boundary calibration. ① shows the t-SNE representations of the five classes in the feature space. The two ★ in 2 represent the 1-shot prototypes of two novel classes. Before calibration, they are located at the edge of the distribution and are not representative. The prototype ★ is calibrated in ③ and located at the center of the class distribution. The covariance distance instead of the Euclidean distance in ④ allows the classification decision boundary to be calibrated and the 1-shot classification task gets better generalization ability

- A prior-driven task-adapted decision boundary calibration module (TDBCM) based on class-covariance metric is introduced to construct more rational decision boundaries. Unlike previous covariance-based methods, TDBCM incorporates prior distributions and performs episode-based intra- and inter-class covariance computation collaboratively to capture inter-feature correlations, addressing the Euclidean distance's insensitivity to the intra-class sample distribution.
- A comprehensive two-stage in-use calibration strategy is presented in this paper to effectively address the representation and distribution bias issues in few-shot learning. Particularly, on three domain-specific fine-grained few-shot datasets, our method achieves optimal performance. It can be said that our work serves as a valuable supplement to the domain-specific application models.

The remaining parts of this paper are organized as follows. Section 2 summarizes related work on few-shot learning, fine-grained classification and plant disease classification. Section 3 provides a detailed description of the proposed AIPCM and TDBCM methods. Then, Sect. 4 presents experimental settings, ablation experiments, and analysis of experimental results on different datasets. Finally, Sect. 5 concludes the paper and provides prospects for future research.

2 Related work

In this section, we briefly introduce related research fields to define and describe our own proposed approach. We introduce the current status of plant disease classification, along with two closely related fields: few-shot learning and fine-grained image classification.

2.1 Plant disease classification

Plant disease classification. Recently, artificial intelligence, particularly deep learning, has rapidly advanced and made inroads across agricultural domains. In 2019, Selvaraj et al. [24] achieved 94.1% accuracy for banana fruit disease detection using Faster-RCNN. Mohanty et al. [25] trained and tested on 54,306 plant disease leaf images from the PlantVillage dataset with AlexNet [26] and GoogleNet [27], achieving the highest accuracy of 99.35%. Brahimi et al. [28] utilized 14,828 tomato disease leaf images for training and testing, resulting in 98.66% and 99.18% accuracy using AlexNet and GoogLeNet models. Recent research [29] proposed a lightweight and cost-effective deep learning architecture, using the proposed DenseNet-121 model to classify leaf images from a dataset named

"PlantDoc." This model achieved fast and efficient recognition, but the overall classification accuracy was only 92.5%.

The quantity of plant disease images utilized in the aforementioned studies suffices for deep learning model training. However, in real-world scenarios, some plant diseases have scarce or merely tens of training data due to low incidence rates and costly image collection, which constrains the application of the above-mentioned methods.

2.2 Few-shot learning

Few-shot learning explores how to improve the performance of a model with limited labeled data. Considering algorithm relevance to our study, we focus on two of them.

Optimization-based few-shot learning. The goal of optimization-based few-shot learning algorithms is to find good initialization parameters. In this way, the model can quickly find the optimal initial value with fine-tuning when identifying novel classes that have never been seen before. Model-agnostic meta-learning (MAML) [30] and Meta-SGD [31] tended to modify the gradient computation algorithm, leading to notable outcomes in just a few gradient steps. Ravi and Larochelle [32] not only acquired effective initial parameters but also developed an LSTM-based [33] optimizer tailored for fine-tuning efficiency. However, fine-tuning is often required during testing, which means there are still model updates when dealing with the target mission, so they are not very effective in 1-shot learning. Momin Abbas et al. [7] proposed an improved MAML algorithm that utilized sharpness-aware minimization to avoid the loss function from getting stuck in a local optimum.

Metric-based few-shot learning. Metric learning-based approaches strive to develop an embedding space where samples of the same class are closer and distinct ones are distant. During testing, it is classified by comparing the similarity between the query sample and the class prototype. RelationNet [10] argued that simple metrics cannot measure complex relationships between high-dimensional data well, and therefore introduced an adaptive similarity metric learning module. Reference [34] utilized graph neural networks (GNN) to model the similarity between class prototypes and query samples. The above networks typically ignore the advantages of local features in enhancing model discriminability and adaptability. Therefore, DeepEMD [35] introduced EMD distance for similarity measures while preserving local features, which brought higher computational complexity in both training and testing stages. Adaptive plug-and-play network [36] proposed that the metric function determined the upper limit of the few-shot classification accuracy, and introduced a model-adaptive resizer and adaptive similarity metric, achieving advanced results on multiple datasets.

2.3 Fine-grained image classification

Fine-grained image classification refers to the classification of objects with strong feature similarity, which are usually very similar in appearance, even difficult to be distinguished by humans. The current mainstream methods [37–39] usually start by localizing the most discriminative regions in the image and then using the extracted local features for classification. Recent studies increasingly adopt self-attention models like transformers to address fine-grained challenges. TransFG [40] proposed a novel transformer-based structure, which aggregated all the original attention weights into an attention map to guide the network to efficiently and accurately select discriminative patches. ViT-FOD [41], a fine-grained detection model based on Vision Transformer, decomposed the input image into multiple small patches and encoded these patches with vision transformer to extract a feature representation of the image. Experiments verified that ViT-FOD has strong performance and generalization ability.

In summary, despite the notable accomplishments of deep convolutional neural networks and transformers across various visual tasks [24, 28, 37–39, 41], obtaining discriminative representations remains a challenging problem for fine-grained image classification. In particular, there is still no effective method for classifying fine-grained images using only a few labeled samples. Different from these methods, this paper proposes an attention-based inverse Mahalanobis distance weighted prototype calibration module, which can focus on fine-grained feature information besides obtaining a more accurate class prototype representation. Meanwhile, a distance function based on class-covariance metric is introduced to obtain a more reasonable partition surface. The two-stage in-use calibration approach can eventually be used to solve the fine-grained few-shot image classification problem in a specific domain.

3 Proposed method

Problem definition. According to the standard definition of few-shot classification task [10, 11, 22, 42], we divide the dataset into a base classes dataset $\mathcal{D}_{\text{base}} = \{(x_i, y_i)\}_{i=1}^{N}$ and a novel classes dataset $\mathcal{D}_{\text{novel}}$, where x_i represents an image sample, y_i represents the class label of x_i, $\mathcal{D}_{\text{base}} \bigcap \mathcal{D}_{\text{novel}} = \emptyset$, and N denotes the total number of images. Few-shot learning typically adopts the meta-task for training and evaluation. The most common paradigm

for constructing meta-task is the C-way K-shot Q-query task, where C classes are sampled from the novel set, K labeled images are extracted from each class as training samples, and Q samples are extracted from the remaining images of each class as query images for prediction. The labeled dataset is called support set, and unlabeled data for prediction are called query set. For each meta-task Γ (C-way K-shot Q-query), the support set S and query set Q are defined as follows:

$$\begin{aligned} \mathcal{S} &= \left\{\left(x_i^s, y_i^s\right)\right\}_{i=1}^{N_s} (N_s = K * C) \\ \mathcal{Q} &= \left\{\left(x_j, y_j\right)\right\}_{j=1}^{N_q} \end{aligned} \tag{1}$$

where K denotes K images with labels and C denotes the number of novel classes. A few-shot task defined in this way is called a C-way K-shot setting.

Overall framework. To address the fine-grained few-shot classification problem, this paper proposes a novel two-stage in-use calibration approach. The overview framework, as depicted in Fig. 2, primarily comprises an attention-based inverse Mahalanobis distance weighted prototype calibration module (AIPCM) and a prior-driven task-adapted decision boundary calibration module (TDBCM) aimed at refining the partition surface. AIPCM comprises a hybrid attention module (HAM), a feature Gaussianization module (FGM), and a prototype calibration module (PCM). HAM employs both instance and region attention to focus on global and local information from different perspectives. FGM gaussizes the high-dimensional features to obtain usable feature distributions. The prototype calibration module employs inverse Mahalanobis distance weighting to calibrate the distribution of few-shot prototypes in the novel classes, thereby obtaining prototype representations with smaller biases. TDBCM introduces a covariance-based metric to constrain the decision boundary between classes.

Details of each module will be described below. Section 3.1 presents the implementation details of HAM and FGM, corresponding to ① in Fig. 2. Sections 3.2 A and B summarize the base class statistics and prototype

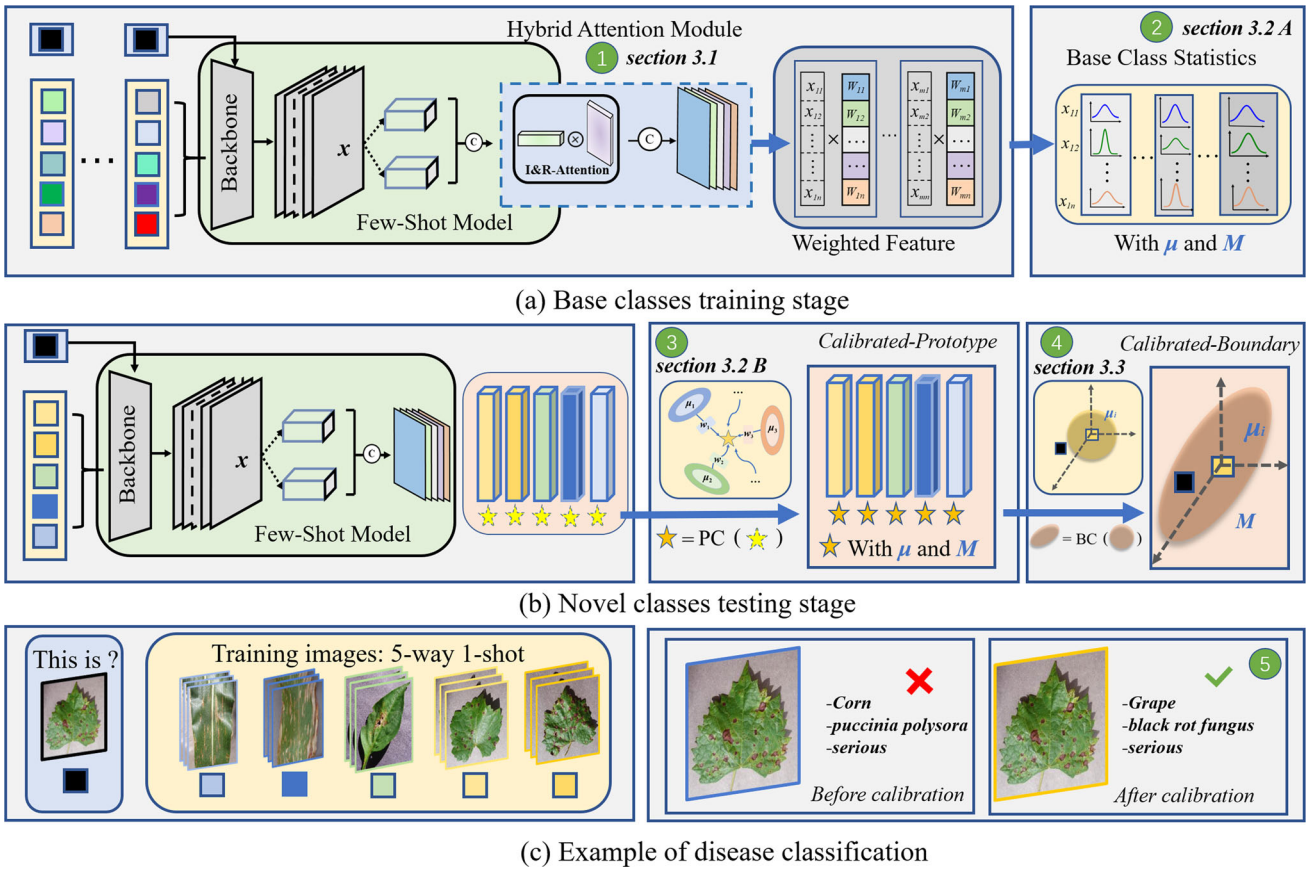


Fig. 2 Overview framework proposed for the 5-way 1-shot fine-grained classification task. **a** The process of collecting base class statistics, including feature weighting and Gaussianization. **b** The evaluation stage for 1-shot sample. We calibrate the prototype representation with PC(·) to obtain a well-positioned prototype and calibrate the decision boundary with informative prior-driven task-adapted BC(·) to obtain an accurate classification decision boundary. **c** An example of 5-way 1-shot task for fine-grained plant disease classification. Current disease classification is only accurate to the disease category; while, our model can accurately identify the disease severity, such as *"Grape-black rot fungus-serious"* in the figure

calibration process, corresponding to 2 and ③ in Fig. 2. Section 3.3 describes TDBCM in detail, corresponding to ④ in Fig. 2.

3.1 Feature pre-processing module

The most commonly used feature encoder backbones for few-shot tasks are Conv-4 [11] and ResNet-12 [11]. However, the results of these networks are not satisfactory for fine-grained few-shot tasks, and models relying on classical backbones as encoders fail to fulfill the demands of fine-grained image recognition. To effectively extract and leverage these fine-grained features, pre-processing is essential. Specifically, we first integrate the convolutional block attention module into the feature encoder to construct a novel HAM. Secondly, to enable the distribution calibration method, we construct FGM to Gaussianize high-dimensional feature vectors.

Focus: hybrid attention module. Inspired by CBAM [43], we integrate the convolutional block attention module into the feature encoder to construct HAM. As shown in Fig. 3, HAM treats the features extracted by the backbone as prior knowledge for the subsequent attention branch. Then, it focuses on the discriminative features from different perspectives using instance attention and region attention. This module does not require separate pre-training, which simplifies the training process. The detailed structure of HAM is illustrated in Fig. 3, encompassing two submodules: The instance attention module and the region attention module, with these two attention submodules being concatenated. HAM is inserted into the backbone network in two ways. For Conv-4, HAM is inserted after each convolutional layer, as depicted in the convolutional attention unit (CAU) in Fig. 4. For ResNet-12, the insertion is done by inserting HAM after the skip connection, as depicted in the residual attention unit (RAU) in Fig. 4. In the feature extraction stage, the base class samples are used for pre-training and the final Softmax layer is removed to obtain a feature extractor for input samples.

Gaussianization: refined power transform. Existing work [16] usually assumes that features from the same class have a specific Gaussian distribution. However, in reality, features extracted by the feature extractor are usually not Gaussian distributed. To make the distribution calibration assumption hold, we first need to Gaussianize high-dimensional feature vectors. The research in [44] shows that power transformation (PT) can make the features better fit the distribution assumption. Motivated by this, we introduce this transformation formula to Gaussianize the input feature.

To avoid the dominance of features with large variances in the existing methods, we add a unit variance projection operation to the original commonly used PT method, Tukey's Transformation Ladder [44]. Assuming $\mathbf{v}$ is the feature extracted from $\mathcal{D}_{\text{novel}}$, the refined power

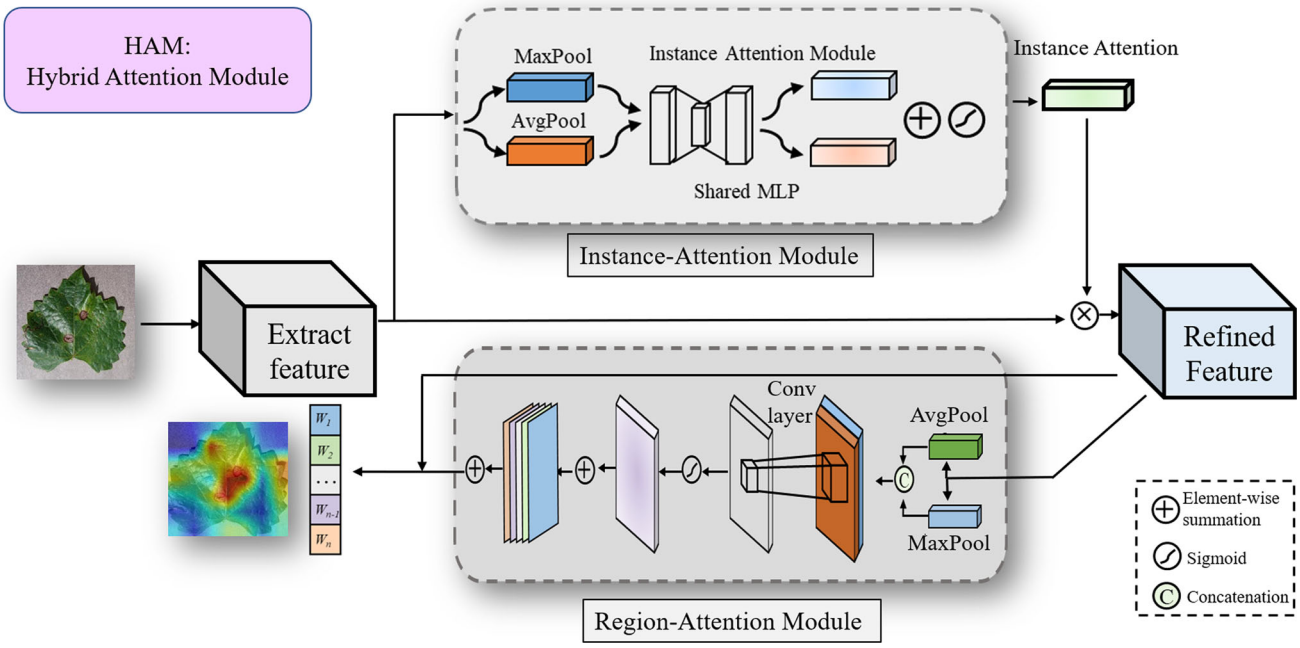


Fig. 3 Hybrid attention module (HAM). The image features are weighted by the instance attention module and region attention module, respectively, to obtain high-dimensional feature vectors. HAM is integrated into the commonly used backbone networks Conv-4 and ResNet-12 for few-shot learning, forming a novel attentive feature encoder

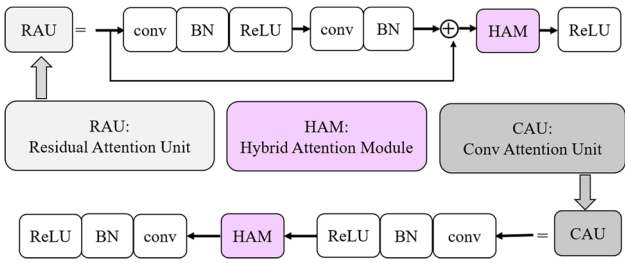


Fig. 4 Diagram of residual attention unit (RAU) and convolutional attention unit (CAU). HAM is inserted differently in Conv-4 and ResNet-12. In RAU, HAM is added after the skip connection. In CAU, HAM is inserted between two adjacent convolutional layers

transformation (R-PT) formula for feature $\mathbf{v}$ can be expressed as follows:

$$f(\mathbf{v}) = \begin{cases} \dfrac{(\mathbf{v}+\epsilon)^{\beta}}{\left\|(\mathbf{v}+\epsilon)^{\beta}\right\|_2} & \text{if } \beta \neq 0 \\ \dfrac{\log(\mathbf{v}+\epsilon)}{\|\log(\mathbf{v}+\epsilon)\|_2} & \text{if } \beta = 0 \end{cases} \tag{2}$$

where β is a hyperparameter used to adjust the distribution skewness. When β=1, the distribution is almost unaffected, and as β changes, the skewness of the distribution changes accordingly. $\epsilon = 1e-6$ is a hyperparameter to ensure that $\mathbf{v}+\epsilon$ is always positive.

R-PT can alleviate distribution skewness, facilitating features to better fit a typical Gaussian distribution. The unit variance projection allows features to be scaled to a uniform scale, avoiding the dominance of features with large variances.

3.2 Inverse-MD weighted prototype calibration module

As depicted in Fig. 1, the prototype distribution is typically not accurate enough for 1-shot tasks. The 1-shot sample is likely to occur at the edge of the distribution, making the representative prototype less typical. In general, existing prototype calibration models exhibit two shortcomings. On the one hand, current models utilize the mean value of the top-n base classes closest to the novel classes to calibrate few-shot prototype, while ignoring the similarity information in the remaining base classes. On the other hand, most approaches neglect the correlation between feature dimensions when measuring feature similarity and simply use Euclidean/Cosine distance to compare sample similarity. To address the above issues, we propose AIPCM which makes full use of all the learned base class statistics to calibrate the novel class prototypes.

A. Base class statistics.

Each feature dimension of each class after feature Gaussianization follows a Gaussian-like distribution. As shown in Fig. 2a, for all samples in base classes, the feature vectors are extracted by the attention-based feature extractor and then Gaussianized using FGM. Subsequently, the mean and variance of all samples within each class are calculated. The module 2 in the upper right corner of Fig. 2 represents the base classes statistics module. For a few-shot learning task Γ, we define the Gaussianized high-dimensional feature vector as x, the class prototype mean of each class of the support set as μ, k denotes the k-th class, and S_k^{Γ} denotes all samples belonging to class k. The mean $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{M}_k$ are calculated as follows:

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{|S_k^{\Gamma}|} x_i}{|S_k^{\Gamma}|} \tag{3}$$

$$\boldsymbol{M}_k = \frac{1}{|S_k|-1} \sum_{x_i \in S_k} (x_i - \mu_k)(x_i - \mu_k)^{\mathrm{T}} \tag{4}$$

Figure 5 shows the detailed distribution calculation process for class k.

B. Inverse-MD prototype calibration.

Similar classes typically have similar feature representations with mean and variance. We can use the mean and variance of the Gaussianized high-dimensional vectors in the base classes to calibrate the prototype distribution of few-shot samples. The computed mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{M}_k$ can be stored as the prior distribution of the base classes for future use.

To fully leverage the prior base class distribution, we propose a prototype calibration method based on similarity weighting. We first calculate the Mahalanobis distance [18] between the few-shot prototypes and the feature mean in each base class. Base classes exhibiting smaller distances to the novel class samples will be assigned greater weights in the prototype representation. Here, we opt for the Mahalanobis distance over the conventional Euclidean distance. The application of the Mahalanobis distance takes into account the correlation among the sample feature dimensions.

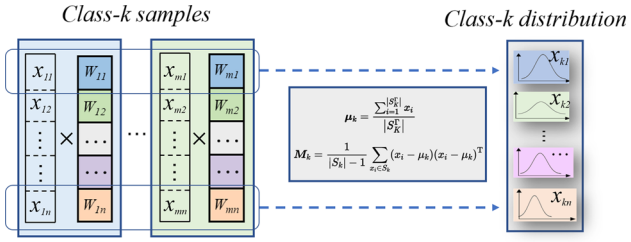


Fig. 5 Base class distribution statistics. The left side shows the high-dimensional weighted vectors of all samples belonging to class k. The right side shows the statistical results of the distribution of class k, from which we can obtain the mean and variance of the data in each feature dimension

For a multivariate variable x with mean μ and covariance matrix $\boldsymbol{Q}$ (specific to the task Γ and class k) (The detailed formula of $\boldsymbol{Q}$ is shown in Sect. 3.3), its covariance distance can be represented as:

$$D_k(x) = \sqrt{(x-\mu)^{\mathrm{T}}(\boldsymbol{Q})^{-1}(x-\mu)} \tag{5}$$

Then, the similarity coefficients are calculated using the Mahalanobis distances, and the similarity coefficients are normalized to obtain the weight coefficients. The similarity coefficient m_k and the weight coefficient w_k can be expressed as follows:

$$m_k = \frac{1}{\sqrt{(x-\mu_k)^{\mathrm{T}}(\boldsymbol{Q})^{-1}(x-\mu_k)}} \tag{6}$$

$$w_k = \frac{m_k}{\sum_{k=1}^{|S_k^{\Gamma}|} m_k} \tag{7}$$

Finally, the few-shot prototypes in the novel classes are represented by weighting all base class prototypes as follows:

$$p = \sum_{k=1}^{|S_k^{\Gamma}|} w_k \cdot \mu_k \tag{8}$$

For N novel classes, there are N calibrated prototype distributions, avoiding bias caused by few-shot samples. Finally, we can obtain a set of statistics for the calibrated distribution. The brief operation of AIPCM is shown in Fig. 6. As shown in Fig. 7, the dark yellow ★ in the right figure is the calibrated prototype.

3.3 Prior-driven task-adapted decision boundary calibration module

After obtaining the calibrated novel classes prototype and distribution, we can further refine the decision boundary using prior distribution. After investigation, we find that existing few-shot learning methods typically use squared Euclidean distance [11] or Cosine distance [22, 23] as the distance metric after complex nonlinear mapping. The Euclidean/Cosine distance is indeed a good choice for tasks with unknown sample distributions, but when the distribution can be quantified, using Mahalanobis distance with the covariance matrix incorporated into the distance metric is a better choice. Compared to the Mahalanobis distance, Euclidean/Cosine distance suffers a notable limitation: it overlooks the correlation among feature dimensions. The Mahalanobis distance incorporates feature dimension

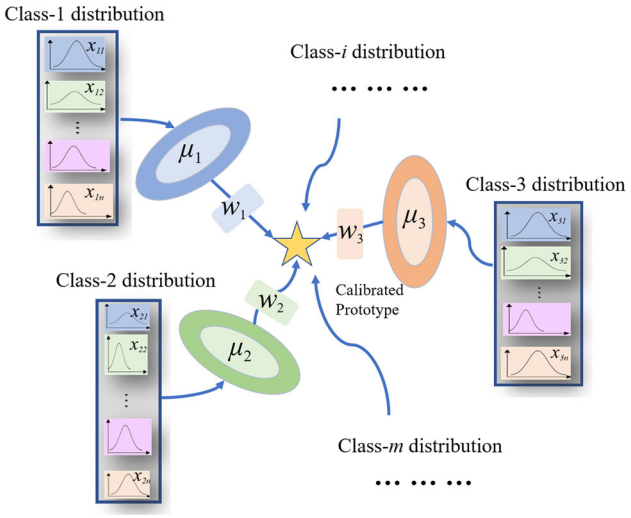


Fig. 6 Attention-based inverse Mahalanobis distance weighted prototype calibration module (AIPCM). We calculate covariance distances between the query image and all base class prototypes, obtaining similarity coefficients and weight coefficients for each class. This allows us to fully utilize all base class information to calibrate the prototype representation of 1-shot query image

correlations via the covariance matrix during distance computation, effectively addressing the limitation of the Euclidean distance's insensitivity to intra-class sample distribution. The class decision boundaries with Mahalanobis distance and Euclidean distance are shown in Fig. 8.

In this paper, we propose a prior-driven task-adapted decision boundary calibration module (TDBCM). TDBCM differs from previous methods [19–21] by fully incorporating prior distribution information from base classes and performing episode-based intra- and inter-class covariance matrix computation. Compared to other covariance measurement methods, our approach has four main characteristics: Intra-class and inter-class covariance collaboration; Episode-based covariance training; Prior-driven covariance matrix calculation; Interpretability and distribution quantification. In the distance metric part, the similarity metric is performed by replacing the Euclidean/Cosine distance with the Mahalanobis distance. The right module ④ in Fig. 2b shows the calibrated decision boundary.

For a few-shot learning task Γ, we define the Gaussianized high-dimensional feature vector as x, the class prototype mean of each class of the support set as μ, and k denotes the k-th class. The Mahalanobis distance d_k can be expressed as:

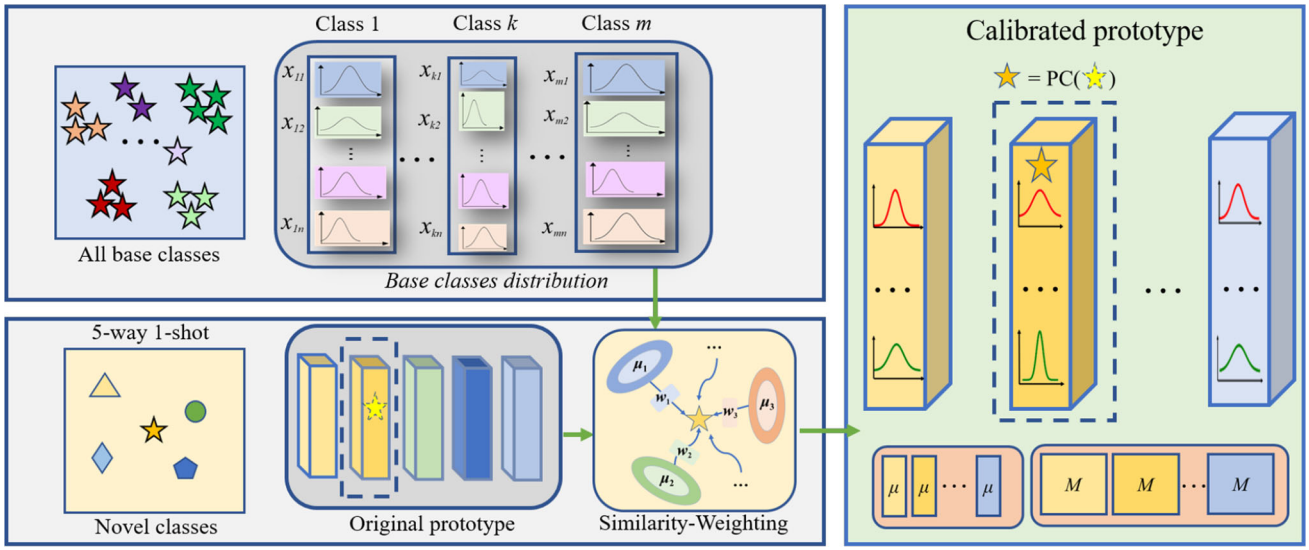


Fig. 7 Illustration of AIPCM usage. The Base class statistics in the upper left part of the figure are stored for use. After feature extraction, a 5-way 1-shot task yields 5 initial prototypes. We calculate the covariance distance for each prototype with all base class prototypes separately and obtain the final calibrated prototype by similarity weighting. The calibrated prototype contains distribution information such as mean μ and covariance M

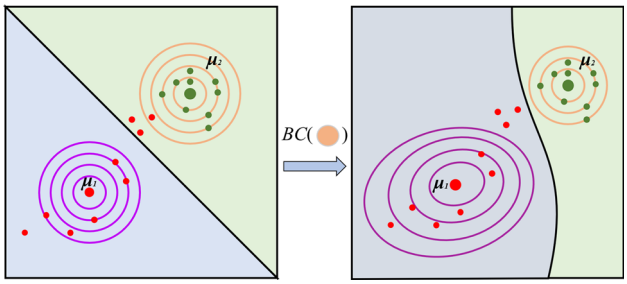


Fig. 8 Examples of classification results with different distance measures. The Euclidean distance (left) decision boundary is centered on the class mean μ, forming an equidistant decision boundary in each dimension, and wrong classification results are obtained. The Mahalanobis distance (right) takes into account the feature dimension covariance in forming the decision boundary and yields a more reasonable partition surface. BC(·) represents the operator for decision boundary calibration

$$d_k(x,\mu) = \frac{1}{2}(x-\mu)^T \left(\boldsymbol{Q}_k^{\Gamma}\right)^{-1}(x-\mu) \tag{9}$$

where $d_k(x,\mu)$ denotes the covariance distance, and its covariance matrix is denoted as $\boldsymbol{Q}_k^{\Gamma}$, which represents the covariance matrix specific to task Γ and class k. Since the number of samples in the support set in few-shot learning is much smaller than the feature dimension, a regularization method is used to calculate $\boldsymbol{Q}_k^{\Gamma}$ to ensure invertibility.

$$\boldsymbol{Q}_k^{\Gamma} = \lambda_k^{\Gamma}\boldsymbol{M}_k^{\Gamma} + \left(1-\lambda_k^{\Gamma}\right)\boldsymbol{M}^{\Gamma} + \beta\boldsymbol{I} \tag{10}$$

In Eq. (10), $\boldsymbol{I}$ denotes the identity matrix; β denotes the matrix scaling factor, and $\beta = 0.5$ in this experiment; $\boldsymbol{Q}_k^{\Gamma}$ is a weighted combination of two covariance matrices, where $\boldsymbol{M}_k^{\Gamma}$ is the intra-class covariance matrix, and $\boldsymbol{M}^{\Gamma}$ is the inter-class covariance matrix; $\boldsymbol{M}^{\Gamma}$ is obtained by estimating the eigenvalues x_i for all $x_i \in S_k^{\Gamma}$.

$$\boldsymbol{M}_k^{\Gamma} = \frac{1}{|S_k^{\Gamma}|-1}\sum_{x_i \in S_k^{\Gamma}}(x_i-\mu_k)(x_i-\mu_k)^{\mathrm{T}} \tag{11}$$

where S_k^{Γ} denotes the samples belonging to class k in the support set, and $\boldsymbol{M}_k^{\Gamma}$ is defined as the zero matrix in the 1-shot case. For $\boldsymbol{M}^{\Gamma}$, it is calculated in the same way as $\boldsymbol{M}_k^{\Gamma}$, except that $\boldsymbol{M}^{\Gamma}$ is defined as the covariance matrix of all classes of task Γ for all samples in S^{Γ}. The scale factor λ_k^{Γ} is calculated by the following equation.

$$\lambda_k^{\Gamma} = |S_k^{\Gamma}|/\left(|S_k^{\Gamma}|+1\right) \tag{12}$$

For the few-shot learning task Γ with C-way N-shot, when N= 1, $\boldsymbol{Q}_k^{\Gamma} = 0.5\boldsymbol{M}^{\Gamma} + \beta\boldsymbol{I}$, where $\boldsymbol{Q}_k^{\Gamma}$ only depends on β and $\boldsymbol{M}^{\Gamma}$; when N is larger than 1, $\boldsymbol{Q}_k^{\Gamma}$ is gradually determined by $\boldsymbol{M}_k^{\Gamma}$, $\boldsymbol{M}^{\Gamma}$ and β together, and the larger N is, the greater the influence of the intra-class covariance matrix $\boldsymbol{M}_k^{\Gamma}$ on $\boldsymbol{Q}_k^{\Gamma}$. The full procedure of the proposed algorithm can be obtained from Algorithm 1.

Algorithm 1 Algorithm procedure for few-shot

Input: Training set $\mathcal{D}_{\text{base}} = \{(x_i, y_i)\}_{i=1}^{N}$, novel set $\mathcal{D}_{\text{novel}}$
Output: Calibrated prototypes P, Calibrated decision boundaries E

Feature weighting for $\{(x_i)\}_{i=1}^{N}$
Feature Gaussian-like process for $\{(x_i)\}_{i=1}^{N}$
for $\{(x_i, y_i)\}_{i=1}^{N}$ **do**
    Calculate the mean $\boldsymbol{\mu}_k$ for each class
    Calculate the covariance $\boldsymbol{M}_k$ for each class
    Store the statistics
end for
for $episode = 1, 2, ...,$ **do**
    Initialize covariance distance $D_k(x)$
    Calculate $D_k(x)$ as Eq.(5)
    Calculate similarity m_k as Eq.(6)
    Calculate weight coefficient w_k as Eq.(7)
    Obtain the calibrated distribution P
    Calculate $\boldsymbol{M}^{\Gamma}$ and $\boldsymbol{M}_k^{\Gamma}$ as Eq.(11)
    Calculate scale factor λ_k^{Γ} as Eq.(12)
    Calculate $\boldsymbol{Q}_k^{\Gamma}$ as Eq.(10)
    Calculate Mahalanobis distance D_k
    Obtain the calibrated decision boundary E
end for

4 Experimental result and analysis

In this paper, we focus on intra-domain fine-grained classification tasks with similar distributions. In particular, we will explore a novel "*Coarse-to-Fine*" plant disease classification task (base classes: coarse-grained → novel classes: fine-grained). Our proposed method demonstrates robust migration generalization across closely distributed datasets within the same domain. In this section, we will answer some questions as follows.

(1) How does our model perform in the module ablation experiment? Are all modules necessary and effective? (Sect. 4.3)
(2) How does our novel prototype and boundary calibration strategy fare on generic datasets (e.g., *mini*-ImageNet and *CUB*) compared to state-of-the-art methods? (Sect. 4.4)
(3) How effective is it in agriculture-specific fine-grained classification tasks (e.g., *FPV*)? (Sect. 4.5 "*Fine-to-Fine*" task)
(4) How does our proposed methodology perform on the agriculture-specific "Coarse-to-Fine" task? (Sect. 4.5 "*Coarse-to-Fine*" task)

4.1 Experiments

To evaluate the model effects, we conduct experiments on datasets of varying granularity. We select *mini*-ImageNet [45] as a representative coarse-grained dataset and *CUB* [46] as a representative fine-grained dataset. These datasets are commonly employed in few-shot tasks, allowing for a fair assessment of performance disparities between our approach and SOTA methods.

In addition, we apply the proposed model to fine-grained plant disease classification. The classification performance in this field is evaluated based on the *FPV* dataset, a finer-grained dataset containing 10 species and 61 diseases, each labeled in the format "*plant disease severity*." As shown in Fig. 9, the inner circle represents its coarse-grained classification and the outer circle represents its fine-grained classification.

In the ablation experiments, we compare the performance differences when HAM, AIPCM, and TDBCM appear individually and in combination, and verify the effectiveness of each module in fine-grained few-shot tasks.

For the method comparison, we experiment with 5-way 1-shot or 5-way 5-shot classification settings on the above datasets under uniform parameters. *mini*-ImageNet contains 100 different classes with image size of $84 \times 84 \times 3$. Following the common splitting method in previous studies [11, 22, 47], we divide the dataset into 64 base classes, 16 validation classes, and 20 novel classes. *CUB* contains 200 different species of birds. Also following the previous splitting method [11, 22, 47], we divide the dataset into 100 base classes, 50 validation classes, and 50 novel classes.

The domain-specific datasets are applied to two classification tasks: the "*Fine-to-Fine*" task (Task-1) and the "*Coarse-to-Fine*" task (Task-2). In *FPV*, 30 classes are selected as base classes, 15 classes as validation classes, and 15 classes as novel classes for the general Task-1. The novel "*Coarse-to-Fine*" task starts by dividing the dataset evenly into two disjoint subsets. Subset A serves as the base classes and is divided into 10 classes according to plant species. Subset B, the novel classes, is divided into 60 classes according to disease severity. In Sect. 4.5 Task-2, we select 20 of these classes as novel classes for experiments.

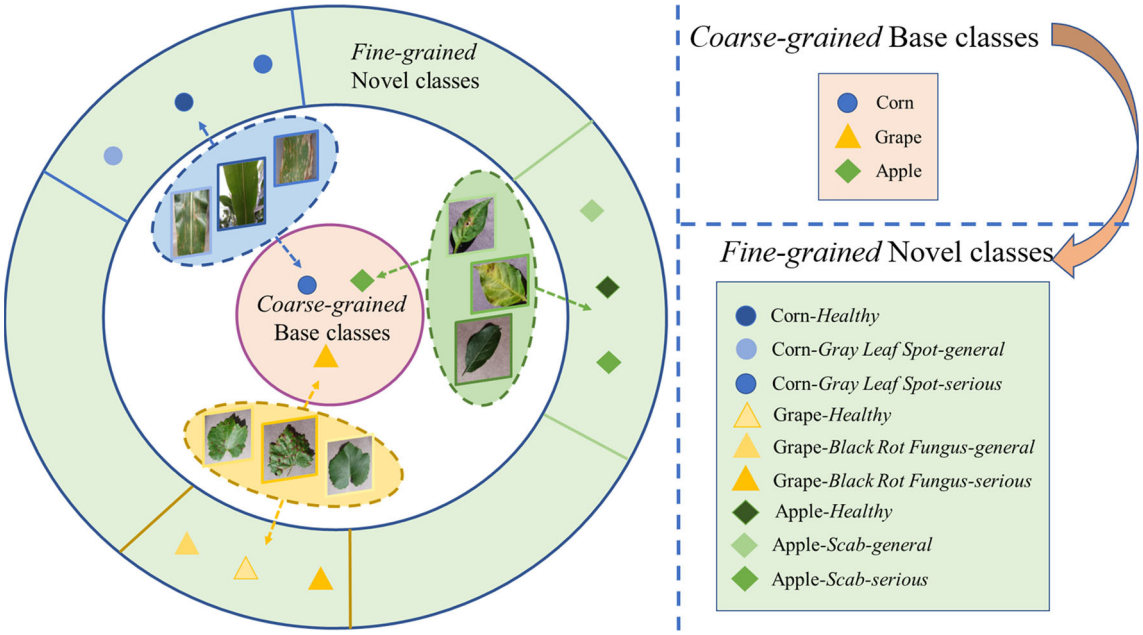


Fig. 9 "*Coarse-to-Fine*" task. The inner circles are coarse-grained base class data, which are labeled in the format of "*plant species*." The outer circle shows the fine-grained annotation of each image, which is in the format of "*plant disease severity*." In this novel task, the base class samples are visible with labels accurate only to plant species, and the novel class data with invisible labels need to be classified to fine-grained disease severity

4.2 Implementation details

Details. We use ResNet-12 [1] as the feature extractor. Pre-training is performed using base class samples. Once the training is completed, the parameters of this feature extractor are frozen unchangeably and the final Softmax layer is removed to obtain the feature extractor. All images are resized to a uniform input size of 84 × 84 and the momentum is set to 0.9. The initial learning rate is set to 0.001, and halving is performed every 20 epochs. Experiments are conducted on NVIDIA 2080 Ti GPU based on the PyTorch framework.

Evaluation metrics. The evaluation metrics align with the mainstream metrics [11, 22]. The accuracy figures presented in the table are all top-1 accuracy. We report the classification accuracy of different methods on *mini*-ImageNet, *CUB*, and *FPV* under 5-way 1-shot and 5-way 5-shot experimental settings, respectively.

4.3 Ablation studies

During the ablation experiment, ResNet-12 is used as the basic backbone of the feature encoder. For 5-way 5-shot setting, each episode consists of 5 classes, and each class contains 5 support samples. As shown in Table 1, to further validate the effectiveness of HAM, FGM+PCM, and TDBCM, we apply these three modules individually or in combination to the baseline model and conduct experiments on three datasets.

As shown in Table 1, in the classification task, if the module is added then mark ($\surd$) the appropriate place in the table, otherwise leave it blank. Several marks ($\surd$) means that several modules are used at the same time. Since the PCM module must rely on FGM to operate, they are used in conjunction. When the modules are combined, FGM and PCM do not appear separately.

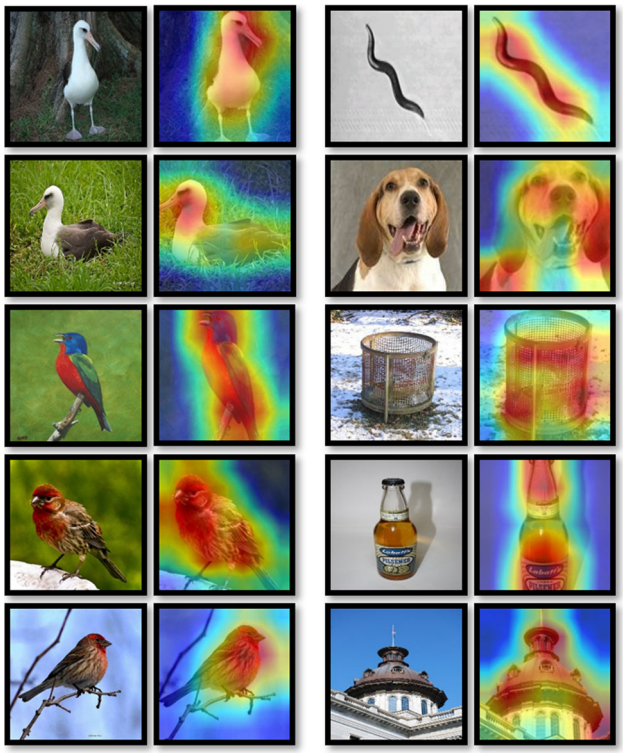
The influence of HAM. The second row of Table 1 shows the classification results with the addition of HAM alone. Experiments show that HAM has good generalization on all three datasets and can effectively improve the classification performance of the original model. In particular, the classification accuracy has been improved by 0.72% and 1.05% on the fine-grained datasets *CUB* and *FPV*, respectively. This indicates that the attention mechanism is very helpful for fine-grained feature extraction, and focusing on the most relevant features through HAM is important. We visualize the focusing effect of some samples after HAM processing, and the attention heat map is shown in Fig. 10.

The influence of R-PT. Comparing rows 3 and 4, 6 and 7, 9 and 10 in Table 1, it is evident that FGM (R-PT) + PCM consistently yields better results. Particularly on the coarse-grained dataset *mini*-ImageNet, the improvement from the R-PT method is more prominent compared to the fine-grained dataset. This effect could potentially stem

Table 1 Ablation study of 5-way 5-shot on few-shot classification

Baseline	HAM	FGM(PT) +PCM	FGM(R-PT) +PCM	TDBCM	*mini*-ImageNet	*CUB*	*FPV*
√					83.01 ± 0.33	90.10 ± 0.23	88.84 ± 0.40
√	√				84.00 ± 0.34	90.82 ± 0.44	89.89 ± 0.37
√		√			83.61 ± 0.33	91.00 ± 0.30	89.59 ± 0.35
√			√		83.86 ± 0.36	91.04 ± 0.31	89.67 ± 0.37
√				√	84.41 ± 0.35	91.13 ± 0.31	90.07 ± 0.43
√	√	√			84.30 ± 0.33	91.08 ± 0.32	89.93 ± 0.35
√	√		√		84.50 ± 0.30	91.20 ± 0.28	90.09 ± 0.39
√	√			√	85.08 ± 0.28	91.23 ± 0.29	90.31 ± 0.41
√	√	√		√	85.28 ± 0.33	91.37 ± 0.30	90.70 ± 0.44
√	√		√	√	**85.52 ± 0.29**	**91.43 ± 0.28**	**90.85 ± 0.37**

The distance metric of the baseline is Cosine distance. The best results are displayed in boldface (mean ± S.D.%). Numbers are in percentage (%). The mark (√) indicates that the module is used



a. *CUB* b. *mini*-ImageNet

Fig. 10 Visualization of attention heat maps generated by HAM. **a** shows the focal region on the fine-grained dataset *CUB*, **b** indicates the focal region on *mini*-Imagenet

from the constraints of the R-PT method on large intra-class variances in coarse-grained data.

The influence of FGM+PCM. It is employed during instance predictions within novel classes. Focusing on the first and fourth rows in Table 1, we can observe that there is a significant performance improvement on *FPV* and *CUB*, but a smaller improvement on *mini*-ImageNet. We can probably infer that the possible reason is that our model assumes that the base classes and the novel classes have similar distributions. In the case of *mini*-ImageNet, a generalized dataset with limited class similarity, the feature distribution transfer-based AIPCM exhibits diminished performance. However, in datasets *CUB* and *FPV*, where the similarity criterion is fulfilled, the prototype calibration module demonstrates substantial enhancement in effectiveness.

The influence of TDBCM. TDBCM is a similarity distance metric module that fully considers the data correlation problem among dimensions in high-dimensional features to calibrate the decision boundary of nonlinear classifiers. Focusing on the first and fifth rows in Table 1, we observe that the separate presence of TDBCM leads to performance enhancement. The baseline employs the Cosine distance as its distance metric. The performance improvement results from substituting the Cosine distance with the Mahalanobis distance, which effectively utilizes the base classes feature distribution in the distance metric. In addition, we compare the 5-way *n*-shot classification accuracy (%) on *mini*-ImageNet and *CUB* with Euclidean distance and Mahalanobis distance. As shown in Fig. 11, where the green bars indicate the Euclidean distance and the yellow bars indicate the Mahalanobis distance. It is evident that the covariance metric exhibits improved classification accuracy across varying numbers of shots, with accuracy improvements more pronounced for smaller shot numbers. In the last row of Table 1, the simultaneous use of the three proposed modules resulted in the best performance.

In summary, the series of ablation experiments demonstrate the effectiveness and strong generalization ability of the proposed modules.

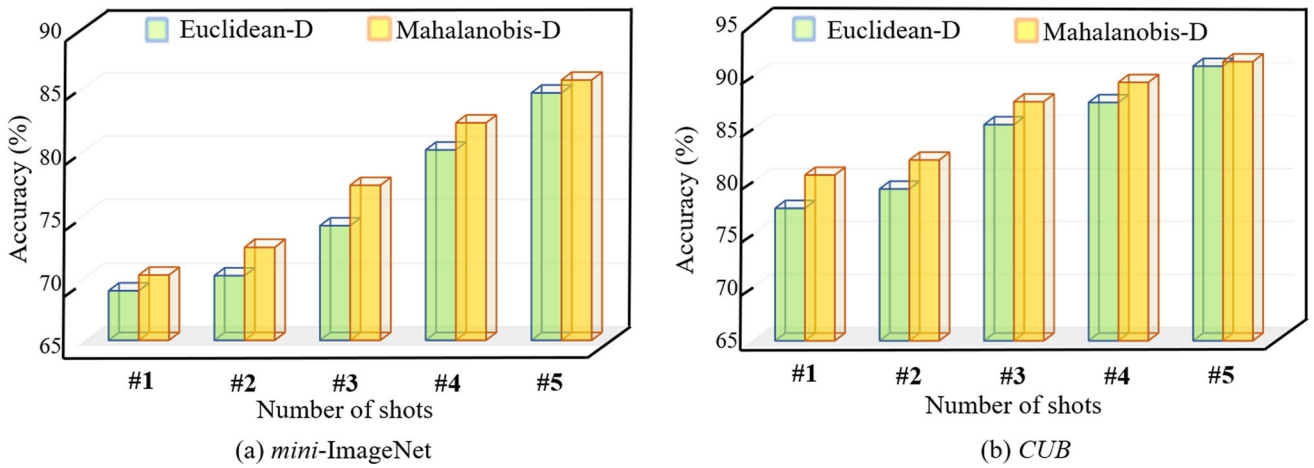


Fig. 11 5-way *n*-shot classification accuracy (%) on *mini*-ImageNet and *CUB* with different distance metrics. Green indicates Euclidean distance and yellow indicates Mahalanobis distance. **a** shows the experimental results in *mini*-ImageNet, **b** shows the experimental results in *CUB*

4.4 Comparison with state-of-the-art

To evaluate the effectiveness of the proposed method, we compare it with classical and advanced few-shot algorithms. We will perform a series of experiments on the common few-shot datasets *mini*-ImageNet and *CUB* in *n*-way *k*-shot experimental setting to validate the performance of our method on general datasets. Table 2 shows the comparison results between our approach and the state-of-the-art methods on *mini*-ImageNet and *CUB*.

For each dataset, we train the feature extractor using base class samples and evaluate the model performance using novel class samples. We try the commonly used feature extraction backbone networks Conv-4 and ResNet-12 respectively [11, 22, 30]. As shown in Table 3, the performance of the lightweight Conv-4 is slightly inferior to the complex ResNet-12. Given our model's sensitivity to fine-grained feature extraction, the following experiments are conducted based on ResNet-12, which is more capable of feature extraction.

Table 2 presents the classification results on *mini*-ImageNet and *CUB* datasets. As observed in Table 2, our proposed method achieves better classification results compared to the optimization-based and metric-based few-shot learning methods. Experiments on *mini*-ImageNet show that the proposed model achieves 70.75% accuracy under the 5-way1-shot experimental setting, which is a 0.91% improvement over the sub-optimal method STANet [59]. In the 5-way 5-shot experimental setting, the model's accuracy reaches 85.52%, marking a 0.64% improvement over the sub-optimal approach. Compared with *mini*-ImageNet, the *CUB* dataset, where all classes belong to the same domain and the feature similarity of each class is higher, is more suitable for our approach. On *CUB* dataset, the accuracy of the proposed model reaches 80.68% and 91.43% for 5-way 1-shot and 5-way 5-shot settings, respectively, with 1.08% and 0.71% improvement compared to SOTA method.

Based on all the experimental results, we can conclude that:

(1) Compared to *mini*-ImageNet, our model has better performance on *CUB*, where the inter-class distribution is more similar.
(2) Compared to 5-way 5-shot setting, our model has a stronger performance improvement under 5-way 1-shot setting.

This phenomenon can be attributed to the fact that, in contrast to metric-based algorithms such as DeepBDC [55] and optimization-based models like MAML [30], our proposed method effectively leverages the prior feature information from all base classes, resulting in more reasonable and accurate prototype representations for novel classes. The *CUB* dataset contains only one kind of organism, bird, with similar feature distribution among different classes. The excellent performance under 1-shot setting is also due to this distribution transfer between base and novel classes. Prototype calibration and decision boundary calibration enable the proposed model to better handle this 1-shot classification task.

4.5 Domain-specific tasks

For domain-specific real-world tasks such as agricultural plant disease classification, large-scale sample collection is not feasible due to the indeterminable nature of disease samples, which often require labeling by experienced agricultural experts. Currently, there are some open-source datasets available online, but their classification granularity

Table 2 Comparison of the state-of-the-art few-shot classification algorithms on the *mini*-ImageNet and *CUB* dataset

Methods	*mini*-ImageNet		*CUB*	
	5way1shot	5way5shot	5way1shot	5way5shot
Optimization-based				
MAML [30] (2017)	57.40 ± 0.47	72.42 ± 0.65	70.44 ± 0.55	85.50 ± 0.33
E^3BM [9] (2020)	64.45 ± 0.34	81.04 ± 0.53	78.22 ± 0.61	89.34 ± 0.35
EMO [48] (2023)	69.15 ± 0.34	84.13 ± 0.25	–	–
Generation-based				
MVT [14] (2020)	–	67.67 ± 0.70	–	80.33 ± 0.60
TriNet [49] (2019)	58.12 ± 1.37	76.92 ± 0.69	69.61 ± 0.46	84.10 ± 0.30
Metric-based				
Baseline [50] (2019)	60.00 ± 0.44	80.55 ± 0.31	71.85 ± 0.46	88.09 ± 0.25
Baseline++ [50] (2019)	63.25 ± 0.44	81.67 ± 0.30	75.25 ± 0.45	89.85 ± 0.23
Meta-Baseline [23] (2020)	64.17 ± 0.45	81.41 ± 0.31	78.16 ± 0.43	90.04 ± 0.23
Neg-Margin [51] (2020)	61.70 ± 0.46	78.03 ± 0.33	78.14 ± 0.46	90.00 ± 0.24
FEAT [52] (2020)	66.78 ± 0.20	82.05 ± 0.14	77.53 ± 0.83	89.79 ± 0.28
BML [53] (2021)	67.04 ± 0.63	83.63 ± 0.29	77.21 ± 0.63	90.45 ± 0.36
DeepEMD [35] (2020)	65.91 ± 0.82	82.41 ± 0.56	75.65 ± 0.63	88.69 ± 0.50
MCL [54] (2022)	67.36 ± 0.20	83.63 ± 0.20	-	–
DeepBDC [55] (2022)	67.83 ± 0.43	84.45 ± 0.29	79.01 ± 0.42	90.42 ± 0.17
SetFeat12 [56] (2022)	68.32 ± 0.62	82.71 ± 0.46	79.60 ± 0.80	90.48 ± 0.44
DeepEMD v2 [57] (2022)	68.77 ± 0.29	84.13 ± 0.53	–	–
IAM [58] (2023)	67.95 ± 0.19	84.86 ± 0.13	78.28 ± 0.22	90.72 ± 0.12
STANet [59] (2023)	69.84 ± 0.47	84.88 ± 0.30	-	-
Ours	**70.75 ± 0.41**	**85.52 ± 0.29**	**80.68 ± 0.43**	**91.43 ± 0.28**

Numbers are in percentage (%). The best results are highlighted in bold (mean ± S.D.%)

Table 3 Few-shot results with different settings of backbones (Conv-4 and ResNet-12)

Methods	Backbones	*mini*-ImageNet		*CUB*	
		5way1shot	5way5shot	5way1shot	5way5shot
Relation N [10] (2018)	Conv-4	49.69 ± 0.43	68.14 ± 0.35	–	–
	ResNet-12	54.12 ± 0.46	71.31 ± 0.37	73.22 ± 0.48	86.94 ± 0.28
Baseline [50] (2019)	Conv-4	46.06 ± 0.39	65.83 ± 0.35	47.73 ± 0.41	68.77 ± 0.38
	ResNet-12	60.00 ± 0.44	80.55 ± 0.31	71.85 ± 0.46	88.09 ± 0.25
Baseline++ [50] (2019)	Conv-4	51.16 ± 0.43	67.99 ± 0.36	62.01 ± 0.49	77.72 ± 0.36
	ResNet-12	63.25 ± 0.44	81.67 ± 0.30	75.25 ± 0.45	89.85 ± 0.23
Meta-Baseline [23] (2020)	Conv-4	51.35 ± 0.42	66.99 ± 0.37	58.98 ± 0.47	75.77 ± 0.37
	ResNet-12	64.17 ± 0.45	81.41 ± 0.31	78.16 ± 0.43	90.04 ± 0.23
Neg-Margin [51] (2020)	Conv-4	51.15 ± 0.42	67.32 ± 0.35	64.08 ± 0.48	80.69 ± 0.34
	ResNet-12	61.70 ± 0.46	78.03 ± 0.33	78.54 ± 0.46	90.19 ± 0.24
Ours	Conv-4	51.71 ± 0.41	68.64 ± 0.35	65.45 ± 0.47	81.91 ± 0.34
	ResNet-12	**70.75 ± 0.41**	**85.52 ± 0.29**	**80.68 ± 0.43**	**91.43 ± 0.28**

The best results are displayed in boldface (mean ± S.D.%). Numbers are in percentage (%)

usually only reaches the disease category. The *FPV* dataset targeted in this paper classifies the categories to disease severity, which is of great significance for solving practical agricultural problems. Therefore, this section focuses on the experiments and discussions for *FPV* dataset. To thoroughly validate our model's superiority in agricultural domain, besides *FPV*, we additionally include two datasets, *PV* from laboratory settings and *PlantDoc* from real agricultural fields, as our experimental datasets.

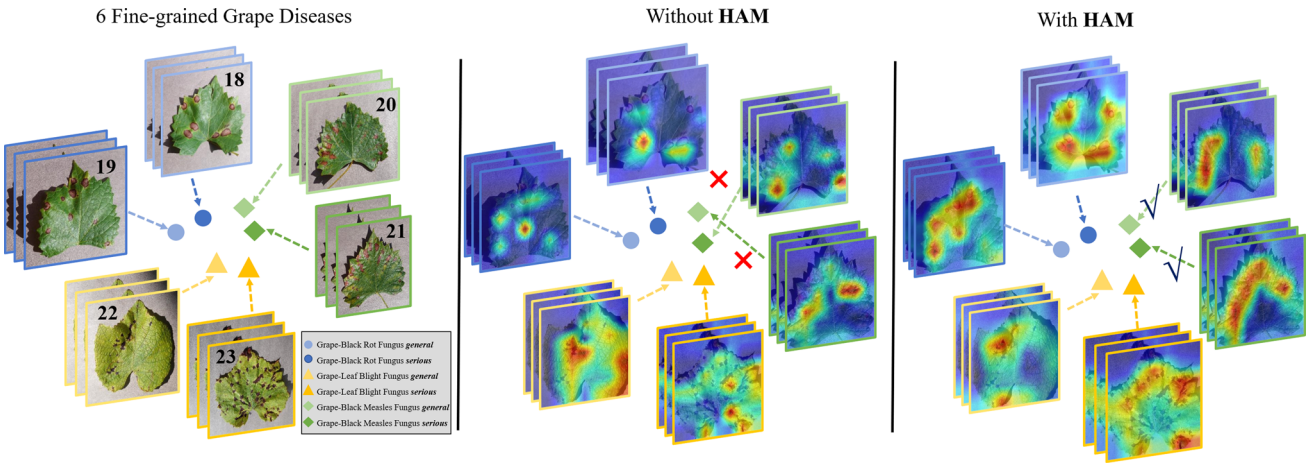


Fig. 12 Visualization of the fine-grained focal region localization on the *FPV* dataset. The figure shows 6 fine-grained disease samples and their attention heat map (second column without HAM, third column with HAM). Different shapes represent different diseases (e.g., ● represents Grape-black bot fungus). The color represents the disease severity (e.g., light blue and dark blue represent Grape-black bot fungus-*general* and -*serious*, respectively). ✗ indicates a misclassification and ✔ indicates a correct classification

We propose two classification tasks for domain-specific datasets, Task-1 is a "*Fine-to-Fine*" task and Task-2 is a "*Coarse-to-Fine*" task. The base class samples for Task-1 are fine-grained data (e.g., data with disease labels), and the novel class samples are data with the same granularity level. Task-2 is a completely novel task with coarse-grained data (e.g., plant species) for the base classes and fine-grained data (e.g., disease severity) for the novel classes.

PV is comprised of 54,306 plant disease images, covering 38 disease categories across 14 plant species. Image labels follow the format "*plant disease*." For the "*Fine-to-Fine*" task, this dataset is divided into 15 base classes, 10 validation classes, and 13 novel classes. For the "*Coarse-to-Fine*" task, we first evenly divide the dataset into two disjoint subsets. Subset A is divided into 14 classes based on plant species and serves as the coarse-grained base classes. Subset B is categorized based on plant diseases, and 20 classes are selected from it as novel classes.

PlantDoc consists of images captured under real cultivation conditions, encompassing 27 disease categories and covering 13 species. In total, the dataset contains 2,598 images. To perform the "*Fine-to-Fine*" task, we divide this dataset into 12 base classes, 7 validation classes, and 8 novel classes. For the "*Coarse-to-Fine*" task, we first evenly divide the dataset into two disjoint subsets. Subset A is divided into 13 classes based on plant species and serves as the coarse-grained base classes. Subset B is categorized based on plant diseases, and 20 classes are selected from it as novel classes.

FPV is a fine-grained plant disease dataset with further class refinement based on *PV*. The dataset is labeled to disease severity, and the label composition format is "*plant disease severity*." It contains 10 plant species and 61 plant diseases. Since the number of samples for two diseases is less than five, we exclude them from the dataset splitting. For the "*Fine-to-Fine*" task, we select 30 of these diseases as base classes, 14 as validation classes, and 15 as novel classes. For the "*Coarse-to-Fine*" task, we first evenly divide the dataset into two disjoint subsets. Subset A is divided into 14 classes based on plant species and serves as the coarse-grained base classes. Subset B is categorized based on plant diseases, and 20 classes are selected from it as novel classes.

Figure 9 shows some dataset examples, with the coarse-grained base class samples annotated as "*plant species*" in the inner circle and the fine-grained novel class samples annotated as "*plant disease severity*" in the outer circle. In this novel "*Coarse-to-Fine*" task, the base class samples are visible with labels accurate only to plant species; while, the novel classes data with invisible labels need to be classified into fine-grained disease severity. The image-label data are derived from *FPV*.

Task-1: "*Fine-to-Fine.*"

The "*Fine-to-Fine*" task is the same as the experimental setting on *mini*-ImageNet and *CUB*, where the model is trained on base classes and then transferred to novel classes with the same granularity for application.

In this section, a series of experiments will be conducted on *PlantDoc* and *FPV* with 5-way 1-shot and 5-way 5-shot experimental settings. Unlike other few-shot disease classification tasks, this paper focuses on fine-grained few-shot classification tasks specific to the disease severity for the first time.

As shown in Table 4, for coarse-grained *PlantDoc* (compared to *FPV*), our approach outperforms other SOTA methods in both 1-shot and 5-shot settings. Specifically, for *PlantDoc* collected from real agricultural scenes, the 5-shot

Table 4 Comparison of the state-of-the-art few-shot classification algorithms on domain-specific datasets (Task-1: *Fine-to-Fine*)

Methods	*FPV* (*fine-grained* → *fine-grained*)		*PlantDoc* (*fine-grained* → *fine-grained*)	
	5way1shot	5way5shot	5way1shot	5way5shot
Optimization-based				
MAML [30] (2017)	69.96 ± 0.46	82.84 ± 0.40	74.25 ± 0.33	84.82 ± 0.22
E^3BM [9] (2020)	78.02 ± 0.42	88.34 ± 0.40	82.43 ± 0.40	90.66 ± 0.25
Metric-based				
Baseline [50] (2019)	71.96 ± 0.37	87.25 ± 0.22	76.78 ± 0.50	87.96 ± 0.17
Baseline++ [50] (2019)	76.11 ± 0.40	88.73 ± 0.31	80.46 ± 0.51	89.82 ± 0.14
Meta-Baseline [23] (2020)	78.25 ± 0.41	88.76 ± 0.26	81.61 ± 0.40	90.12 ± 0.23
Neg-Margin [51] (2020)	78.06 ± 0.46	88.48 ± 0.44	81.24 ± 0.37	89.95 ± 0.17
FEAT [52] (2020)	76.25 ± 0.41	88.02 ± 0.24	80.11 ± 0.37	90.40 ± 0.30
BML [53] (2021)	77.21 ± 0.63	89.33 ± 0.29	81.62 ± 0.35	90.74 ± 0.37
DeepEMD [35] (2020)	76.69 ± 0.47	87.92 ± 0.34	80.41 ± 0.33	90.04 ± 0.20
DeepBDC [55] (2022)	79.00 ± 0.52	89.45 ± 0.26	82.87 ± 0.28	91.50 ± 0.23
SetFeat12 [56] (2022)	79.21 ± 0.50	89.07 ± 0.28	83.10 ± 0.28	91.72 ± 0.12
DeepEMD v2 [57] (2022)	78.86 ± 0.46	89.20 ± 0.28	83.02 ± 0.37	92.11 ± 0.30
Ours	**81.03 ± 0.44**	**90.85 ± 0.37**	**84.15 ± 0.30**	**93.20 ± 0.17**

The best results are highlighted in bold (mean ± S.D.%). Numbers are in percentage (%)

Table 5 Comparison of the state-of-the-art few-shot classification algorithms on domain-specific datasets (Task-2: *Coarse-to-Fine*)

Methods	*FPV* (*coarse* → *fine*)		*PlantDoc* (*coarse* → *fine*)		*PV* (*coarse* → *fine*)	
	5way1shot	5way5shot	5way1shot	5way5shot	5way1shot	5way5shot
Optimization-based						
MAML [30] (2017)	63.67 ± 0.47	73.38 ± 0.37	72.32 ± 0.37	80.04 ± 0.22	75.64 ± 0.37	83.37 ± 0.40
E^3BM [9] (2020)	74.02 ± 0.44	82.79 ± 0.34	78.86 ± 0.35	86.00 ± 0.27	79.83 ± 0.30	87.80 ± 0.34
Metric-based						
Baseline [50] (2019)	69.03 ± 0.41	81.44 ± 0.20	76.65 ± 0.33	84.86 ± 0.21	76.09 ± 0.28	86.03 ± 0.22
Baseline++ [50] (2019)	73.26 ± 0.51	82.31 ± 0.34	77.96 ± 0.35	85.72 ± 0.21	78.68 ± 0.27	87.60 ± 0.20
Meta-Baseline [23] (2020)	73.83 ± 0.55	83.45 ± 0.29	78.53 ± 0.36	86.22 ± 0.22	79.64 ± 0.35	88.00 ± 0.30
Neg-Margin [51] (2020)	72.18 ± 0.52	82.64 ± 0.33	77.94 ± 0.29	86.04 ± 0.27	79.22 ± 0.29	87.66 ± 0.27
FEAT [52] (2020)	71.03 ± 0.47	81.92 ± 0.29	77.61 ± 0.31	85.79 ± 0.30	78.33 ± 0.31	87.57 ± 0.25
BML [53] (2021)	71.61 ± 0.55	82.87 ± 0.30	78.05 ± 0.42	86.27 ± 0.31	78.90 ± 0.41	88.12 ± 0.19
DeepEMD [35] (2020)	71.33 ± 0.44	81.00 ± 0.31	77.16 ± 0.37	85.62 ± 0.25	78.62 ± 0.40	87.08 ± 0.42
DeepBDC [55] (2022)	74.90 ± 0.48	82.66 ± 0.27	78.94 ± 0.37	86.55 ± 0.25	80.11 ± 0.37	88.40 ± 0.25
SetFeat12 [56] (2022)	75.16 ± 0.48	82.93 ± 0.30	79.22 ± 0.35	86.80 ± 0.24	80.35 ± 0.25	88.49 ± 0.31
DeepEMD v2 [57] (2022)	74.44 ± 0.45	83.26 ± 0.28	79.01 ± 0.33	86.43 ± 0.27	80.08 ± 0.28	88.14 ± 0.26
Ours	**76.51 ± 0.30**	**84.16 ± 0.26**	**80.26 ± 0.36**	**87.73 ± 0.24**	**81.43 ± 0.32**	**89.43 ± 0.24**

The best results are highlighted in bold (mean ± S.D.%). Numbers are in percentage (%)

accuracy is improved by 1.09% compared to the sub-optimal DeepEMD v2. In 1-shot tasks, our method demonstrates improvements of 2.54%, 1.28%, 1.05%, and 1.13% when compared to Meta-Baseline [23], DeepBDC [55], SetFeat12 [56], and DeepEMD v2 [57] respectively.

Compared to *PlantDoc*, *FPV* exhibits higher intra-class variance and lower inter-class variance, rendering recognition more challenging. On *FPV*, our proposed model achieves accuracies of 81.03% and 90.85% in 1-shot and 5-shot settings, respectively, which improves by 1.82% and

1.40% compared to the second-best method. The experiments in Table 4 indicate that our model effectively adapts to the characteristics of fine-grained tasks and demonstrates superior performance across multiple datasets in specific domains. This experiment is similar to other fine-grained datasets (e.g., *CUB*) experiments, so we only conduct a brief analysis of the results here.

Task-2: "*Coarse-to-Fine.*"

The "*Coarse-to-Fine*" task is a novel disease classification task proposed in this paper specifically for the agriculture field. As shown in Fig. 9, the inner circle represents the classes encountered during model training, which are labeled only to plant species without requiring high-cost disease annotations. The outer circle represents the novel classes we need to identify, which are fine-grained classes specific to plant disease severity. In the experiment, we first evenly divide the dataset into two disjoint subsets. Subset A is divided into n classes according to plant species and is used as coarse-grained base classes. Subset B is classified according to disease severity, and 20 of these classes are selected as novel classes.

Table 5 shows the classification results under this experimental setup. It is evident that compared to the classification results at the same granularity, there is a certain degree of degradation in accuracy. This phenomenon can be attributed to the alteration in class granularity between the base and novel classes, where granularity significantly impacts the classification task. However, we can also see that our model exhibits substantial enhancement over the existing state-of-the-art (SOTA) method within the "*Coarse-to-Fine*" setting. The experimental results under 5-way 5-shot settings reveal that our method achieves accuracy improvements of 0.90%, 0.94% and 0.93% on *FPV*, *PV* and *PlantDoc*, respectively, compared to the second-best model. Our method outperforms the current SOTA few-shot classification methods on the 5-shot settings. It is well known that "the fewer images selected in each category (such as 1-shot), the higher the requirement for model optimization capability". The experimental results of our model under 5-way 1-shot are also improved by different degrees. Specifically, on the *FPV* dataset of primary interest, our method improves by 1.61%, 1.35% and 2.07% compared with DeepBDC [55], SetFeat12 [56], and DeepEMD v2 [57], respectively, and shows superiority in the 1-shot classification task.

In summary, our model exhibits superior performance over other SOTA methods [55–57] in both the 1-shot and 5-shot settings, particularly in the 1-shot case. The reason for this phenomenon is that as the number of support sets decreases, the available information for each category becomes more limited. However, compared with other few-shot learning methods, our model makes more effective use of the information provided by the base classes. Through the efficient utilization of base class distribution and optimization of decision boundaries, higher classification accuracy is ultimately obtained in the 1-shot setting. The significance of this task is that we can achieve the identification of few-shot hard-to-annotate fine-grained data based on easily labeled domain-specific coarse-grained data.

Attention visualization (Task-2): In addition, we visualize the focal regions of some samples in *FPV*. We select some samples from six fine-grained grape disease classes and generate heat maps showcasing feature weights. The first column in Fig. 12 displays the original grape leaves, which are meticulously categorized into six diseases (including Grape-Black Rot Fungus-*general*, Grape-Black Rot Fungus-*serious*, etc.). The second and third columns exhibit the visualized heat map before and after adding HAM, respectively. It can be observed that our model shows different heat map distributions on each class, and extracts different key features. The key object regions captured by our method contribute to extracting essential discriminative features.

Feature map visualization (Task-2): To better demonstrate how the hybrid attention method affects feature extraction. We visualize the feature embedding space in select convolutional layers, as shown in Fig. 13. The same as before, we compress multiple channels into a single channel for feature map visualization. From Fig. 13, we can see that HAM helps outline the object from the background, preserving more details and suppressing background noise.

Confusion matrix and t-SNE visualization (Task-2): The confusion matrix comparing our method with the baseline model on *FPV* is presented in Fig. 14. Class 17 represents the healthy grape class. Class 18, 19, 20, 21, 22, and 23 are six representative classes of fine-grained grape diseases, which are identical to the six disease categories shown in Fig. 12. It can be seen that our method greatly improves the classification accuracy of class 20 and class 21, and reduces the possibility of class 17 being misclassified as class 18. We perform t-SNE dimensionality reduction for the high-dimensional representations of easy-to-classify samples and hard-to-classify samples. For the convenience of viewing, we only show the dimensionality reduction results for some samples. Figure 15 shows the dimensionality reduction results for the easy-to-classify samples, corresponding to the 10 coarse-grained base classes images (Labeled as plant species). Figure 16 shows the t-SNE visualization for the 7 fine-grained novel classes presented in the confusion matrix, where each class has very similar feature representations.

Case study (Task-2): The ✗ in Fig. 12 displays the most challenging class pairs for classification. For instance,

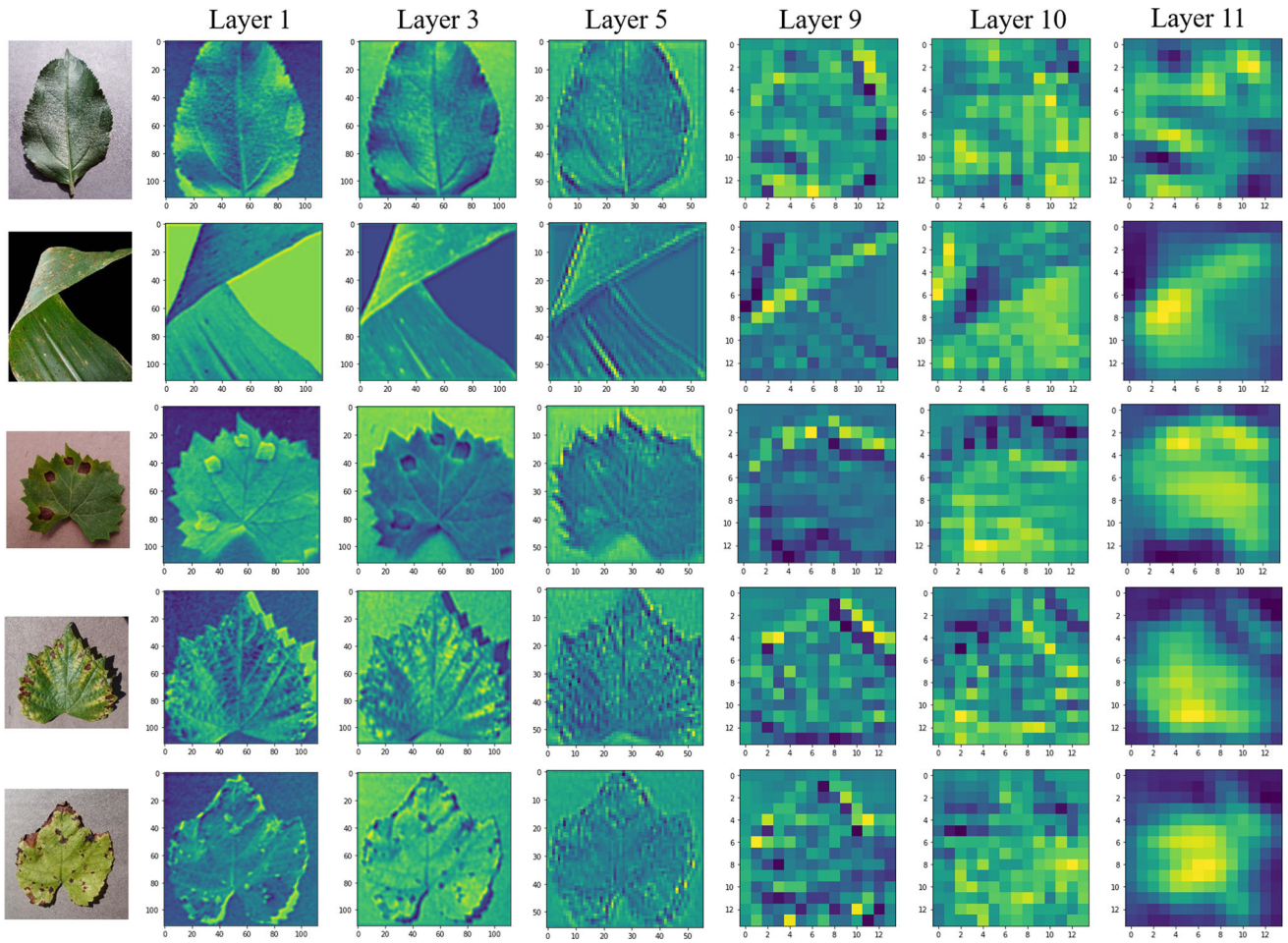
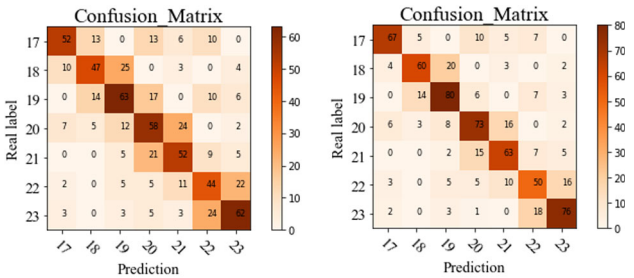


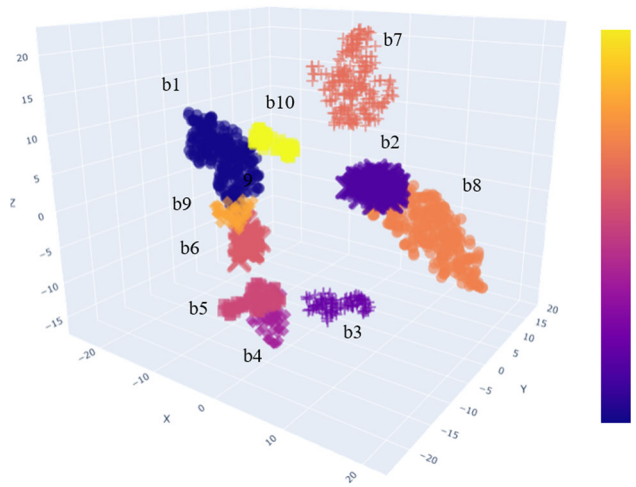
Fig. 13 Visualization of feature maps under different layers. Different columns represent the visual feature maps of different layers. Different rows represent different disease samples respectively



(a). Confusion matrix with baseline (b). Confusion matrix with our model

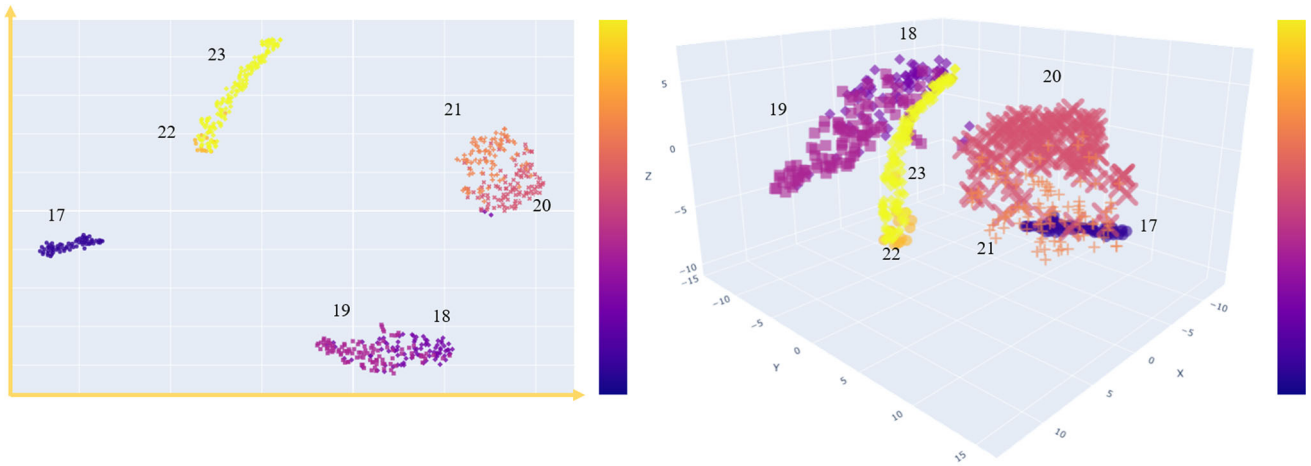
Fig. 14 Confusion matrix of the baseline and our model. 17, 18, 19, 20, 21, 22, and 23 are seven representative classes of fine-grained grape diseases. Each column in the matrix represents the prediction result. Each row represents the real label. The 7 fine-grained classes shown in the figure are consistent with the classes in Figs. 12 and 16

samples from class 20 are usually misclassified as class 21. As can be seen in Fig. 12, there is very little difference between these classes. This similarity even confuses agricultural experts to distinguish them.



3D t-SNE visualization

Fig. 15 t-SNE visualization of 10 coarse-grained base classes (b1-apple, b2-cherry, b3-corn, b4-grape, b5-citrus, b6-peach, b7-pepper, b8-potato, b9-strawberry, b10-tomato). The figure shows the 3D t-SNE visualization. The "b + number" in the figure represent class numbers



(a). 2D t-SNE visualization

(b). 3D t-SNE visualization

Fig. 16 t-SNE visualization of 7 fine-grained grape disease classes. **a** shows the 2D t-SNE visualization, and **b** shows the 3D t-SNE visualization. The numbers in the figure represent category numbers

To summarize the above experimental phenomena, few-shot classification still faces the problem of fine-grained classification. Compared with the results of existing methods, the model proposed in this paper achieves significant performance improvement under different task settings. This fully demonstrates the advantages of the proposed method:

- Through the use of AIPCM and TDBCM, we obtain the ability to capture critical regions and efficiently use the available feature distributions in the base class samples to obtain more reasonable prototype representations and decision boundaries. Our method has important value in fine-grained classification tasks where the inter-class distribution is similar.
- For domain-specific applications where data acquisition is difficult, such as the fine-grained agricultural disease classification, our model has strong applicability and the most significant effect improvement is observed on *FPV*. The few-shot approach proposed in this paper is of great importance in smart agriculture fine-grained disease classification.

5 Conclusion

Originating from real-world demand, the research work focuses on the field of smart agriculture and explores a novel few-shot fine-grained disease classification task. We first propose an attention-based inverse Mahalanobis distance weighted prototype calibration module (AIPCM), which calibrates the novel class prototype representation by weighting the similarity of the prior distribution. By transferring the statistical information of base classes with sufficient samples to the fine-grained novel classes with fewer samples, it achieves prototype and distribution calibration of 1-shot data. To calibrate the decision boundary of the novel classes distribution, we introduce the Mahalanobis distance based on class-covariance metric instead of the commonly used Euclidean/Cosine distance, effectively utilizing the mean and covariance of high-dimensional features of base classes. Experimental results on few-shot datasets *mini*-ImageNet and *CUB* demonstrate that the proposed model achieves the best classification performance. In particular, we address the few-shot classification problem of natural images on open-source fine-grained plant disease dataset, and explore a novel "*Coarse-to-Fine*" plant disease classification task (base classes: coarse-grained → novel classes: fine-grained). This attempt offers a practical solution to domain-specific real-world applications.

Acknowledgements This work was supported by the National Science and Technology Major Project (2021ZD0110901).

Data availability Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Declarations

Conflict of interest The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
2. Tu Z, Talebi H, Zhang H, Yang F, Milanfar P, Bovik A, Li Y (2022) Maxvit: multi-axis vision transformer. In: Computer vision–ECCV 2022: 17th European conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV. Springer, pp 459–479
3. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, 28
4. Saavedra D, Banerjee S, Mery D (2021) Detection of threat objects in baggage inspection with X-ray images using deep learning. Neural Comput Appl 33:7803–7819
5. Dong Z, He Y, Qi X, Chen Y, Shu H, Coatrieux J-L, Yang G, Li S (2022) MNet: rethinking 2D/3D networks for anisotropic medical image segmentation. arXiv preprint arXiv:2205.04846
6. Rasi D, Deepa S (2022) Hybrid optimization enabled deep learning model for colour image segmentation and classification. Neural Comput Appl 34(23):21335–21352
7. Abbas M, Xiao Q, Chen L, Chen P-Y, Chen T (2022) Sharp-maml: sharpness-aware model-agnostic meta learning. arXiv preprint arXiv:2206.03996
8. Sun Q, Liu Y, Chua T-S, Schiele B (2019) Meta-transfer learning for few-shot learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 403–412
9. Liu Y, Schiele B, Sun Q (2020) An ensemble of epoch-wise empirical bayes for few-shot learning. In: Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16. Springer, pp 404–421
10. Sung F, Yang Y, Zhang L, Xiang T, Torr PH, Hospedales TM (2018) Learning to compare: relation network for few-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1199–1208
11. Snell J, Swersky K, Zemel R (2017) Prototypical networks for few-shot learning. In: Advances in neural information processing systems, 30
12. Liu J, Song L, Qin Y (2020) Prototype rectification for few-shot learning. In: Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16. Springer, pp 741–756
13. Wang Y-X, Girshick R, Hebert M, Hariharan B (2018) Low-shot learning from imaginary data. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 7278–7286
14. Park S-J, Han S, Baek J-W, Kim I, Song J, Lee HB, Han J-J, Hwang SJ (2020) Meta variance transfer: learning to augment from the others. In: International conference on machine learning. PMLR, pp 7510–7520
15. Xian Y, Lorenz T, Schiele B, Akata Z (2018) Feature generating networks for zero-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 5542–5551
16. Liu J, Sun Y, Han C, Dou Z, Li W (2020) Deep representation learning on long-tailed data: a learnable embedding augmentation perspective. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 2970–2979
17. Yang S, Liu L, Xu M (2021) Free lunch for few-shot learning: distribution calibration. arXiv preprint arXiv:2101.06395
18. Galeano P, Joseph E, Lillo RE (2015) The Mahalanobis distance for functional data with applications to classification. Technometrics 57(2):281–291
19. Bąk S, Charpiat G, Corvee E, Bremond F, Thonnat M (2012) Learning to match appearances by correlations in a covariance metric space. In: Computer vision–ECCV 2012: 12th European conference on computer vision, Florence, Italy, October 7–13, 2012, Proceedings, Part III 12. Springer, pp 806–820
20. Mensink T, Verbeek J, Perronnin F, Csurka G (2013) Distance-based image classification: generalizing to new classes at near-zero cost. IEEE Trans Pattern Anal Mach Intell 35(11):2624–2637
21. Kamal IM, Bae H, Liu L (2022) Metric learning as a service with covariance embedding. arXiv preprint arXiv:2211.15197
22. Vinyals O, Blundell C, Lillicrap T, Wierstra D, et al (2016) Matching networks for one shot learning. In: Advances in neural information processing systems, 29
23. Chen Y, Wang X, Liu Z, Xu H, Darrell T (2020) A new meta-baseline for few-shot learning
24. Selvaraj MG, Vergara A, Ruiz H, Safari N, Elayabalan S, Ocimati W, Blomme G (2019) AI-powered banana diseases and pest detection. Plant Methods 15(1):1–11
25. Mohanty SP, Hughes DP, Salathé M (2016) Using deep learning for image-based plant disease detection. Front Plant Sci 7:1419
26. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1–9
28. Brahimi M, Boukhalfa K, Moussaoui A (2017) Deep learning for tomato diseases: classification and symptoms visualization. Appl Artif Intell 31(4):299–315
29. Chakraborty A, Kumer D, Deeba K (2021) Plant leaf disease recognition using fastai image classification. In: 2021 5th international conference on computing methodologies and communication (ICCMC). IEEE, pp 1624–1630
30. Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. PMLR, pp 1126–1135
31. Li Z, Zhou F, Chen F, Li H (2017) Meta-sgd: learning to learn quickly for few-shot learning. arXiv preprint arXiv:1707.09835
32. Ravi S, Larochelle H (2017) Optimization as a model for few-shot learning. In: International conference on learning representations
33. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
34. Garcia V, Bruna J (2017) Few-shot learning with graph neural networks. arXiv preprint arXiv:1711.04043
35. Zhang C, Cai Y, Lin G, Shen C (2020) Deepemd: few-shot image classification with differentiable earth mover's distance and structured classifiers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 12203–12213
36. Li H, Li L, Huang Y, Li N, Zhang Y (2023) An adaptive plug-and-play network for few-shot learning. arXiv preprint arXiv:2302.09326
37. Fu J, Zheng H, Mei T (2017) Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 4438–4446
38. Sun X, Xv H, Dong J, Zhou H, Chen C, Li Q (2020) Few-shot learning for domain-specific fine-grained image classification. IEEE Trans Ind Electron 68(4):3588–3598
39. Wei X-S, Luo J-H, Wu J, Zhou Z-H (2017) Selective convolutional descriptor aggregation for fine-grained image retrieval. IEEE Trans Image Process 26(6):2868–2881

40. He J, Chen J-N, Liu S, Kortylewski A, Yang C, Bai Y, Wang C (2022) Transfg: a transformer architecture for fine-grained recognition. In: Proceedings of the AAAI conference on artificial intelligence, vol 36. pp 852–860
41. Zhang Z-C, Chen Z-D, Wang Y, Luo X, Xu X-S (2022) Vit-fod: a vision transformer based fine-grained object discriminator. arXiv preprint arXiv:2203.12816
42. Zhu L, Yang Y (2018) Compound memory networks for few-shot video classification. In: Proceedings of the European conference on computer vision (ECCV). pp 751–766
43. Woo S, Park J, Lee J-Y, Kweon IS (2018) Cbam: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp 3–19
44. Tukey JW (1977) Exploratory data analysis, vol 2. Reading, MA
45. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, pp 248–255
46. Wah C, Branson S, Welinder P, Perona P, Belongie S (2011) The caltech-ucsd birds-200-2011 dataset
47. Ravi S, Larochelle H (2016) Optimization as a model for few-shot learning. In: International conference on learning representations
48. Du Y, Shen J, Zhen X, Snoek CG (2023) EMO: episodic memory optimization for few-shot meta-learning. arXiv preprint arXiv:2306.05189
49. Chen Z, Fu Y, Zhang Y, Jiang Y-G, Xue X, Sigal L (2019) Multi-level semantic feature augmentation for one-shot learning. IEEE Trans Image Process 28(9):4594–4605
50. Chen W-Y, Liu Y-C, Kira Z, Wang Y-CF, Huang J-B (2019) A closer look at few-shot classification. arXiv preprint arXiv:1904.04232
51. Liu B, Cao Y, Lin Y, Li Q, Zhang Z, Long M, Hu H (2020) Negative margin matters: understanding margin in few-shot classification. In: European conference on computer vision. Springer, pp 438–455
52. Ye H-J, Hu H, Zhan D-C, Sha F (2020) Few-shot learning via embedding adaptation with set-to-set functions. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 8808–8817
53. Zhou Z, Qiu X, Xie J, Wu J, Zhang C (2021) Binocular mutual learning for improving few-shot classification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp 8402–8411
54. Liu Y, Zhang W, Xiang C, Zheng T, Cai D, He X (2022) Learning to affiliate: mutual centralized learning for few-shot classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 14411–14420
55. Xie J, Long F, Lv J, Wang Q, Li P (2022) Joint distribution matters: deep brownian distance covariance for few-shot classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 7972–7981
56. Afrasiyabi A, Larochelle H, Lalonde J-F, Gagné C (2022) Matching feature sets for few-shot image classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 9014–9024
57. Zhang C, Cai Y, Lin G, Shen C (2022) Deepemd: differentiable earth mover's distance for few-shot learning. IEEE Trans Pattern Anal Mach Intell 45(5):5632–5648
58. Lee S, Moon W, Seong HS, Heo J-P (2023) Task-oriented channel attention for fine-grained few-shot classification. arXiv preprint arXiv:2308.00093
59. Lai J, Yang S, Wu W, Wu T, Jiang G, Wang X, Liu J, Gao B-B, Zhang W, Xie Y, et al (2023) SpatialFormer: semantic and target aware attentions for few-shot learning. arXiv preprint arXiv:2303.09281