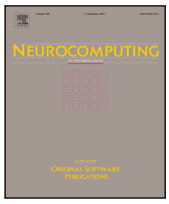
Neurocomputing 636 (2025) 130056

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Drop inherent biases: Multi-level attention calibration for robust cross-domain few-shot classification

Minghui Li [a,b], Jing Jiang [a,b], Hongxun Yao [a,b,c,*]

[a] Faculty of Computing, Harbin Institute of Technology, Harbin, 150001, China
[b] Harbin Institute of Technology - China Mobile 5G Application Innovation Joint Research Institute, Harbin, 150001, China
[c] National Key Laboratory of Smart Farming Technology and Systems, Harbin, 150001, China

ARTICLE INFO

Communicated by Y. Long

Keywords:
Few-shot classification
Cross-domain
Inductive bias
Attention release and reaggregation
Cross alignment

ABSTRACT

Few-shot learning (FSL) is a promising approach for addressing the challenge of classifying novel classes with only limited labeled data. Many few-shot studies have elaborated various task-shared inductive biases (meta-knowledge) to solve such tasks and have achieved impressive performance. However, when there is a domain shift between the training and testing tasks, the learned inductive biases fail to generalize across domains. In this paper, we attempt to suppress and correct inherent discriminative inductive biases from the source domain through source domain attention release and target domain attention reaggregation. We propose a few-shot learning framework, which systematically addresses the large domain shift between base and novel classes. Specifically, the framework consists of three parts: prototype-level attention calibration, feature-level attention calibration for attention release and reaggregation, and loss attention calibration. First, the prototype-level attention calibration module highlights key instances via prototype calibration, reducing the influence of noisy instances in few-shot settings. Second, the feature-level attention calibration module suppresses and corrects erroneous discriminative inductive biases from the source domain through base class attention release and novel class attention reaggregation, respectively. Finally, we incorporate the loss attention calibration module into the loss function to balance the discriminability and diversity of the classification matrix, mitigating the decline in generalization ability caused by erroneous discriminative features during domain shift. We conduct experiments on eight classic few-shot cross-domain datasets. The results demonstrate that, under varying domain shifts, our method improves performance, with average accuracy gains of 0.82% and 1.31% in the 5-way 1-shot and 5-way 5-shot settings, respectively, compared to the existing state-of-the-art (SOTA) method.

1. Introduction

This paper challenges a real-world problem: few-shot learning for cross-domain scenarios. Existing few-shot models can quickly learn to recognize novel classes from a limited number of samples [1–4]. However, in practice, there is often a domain shift between the base and novel classes [5–7]. Compared to domain adaptation, cross-domain few-shot learning (CD-FSL) faces more severe challenges: the extremely limited number of novel class samples is insufficient to alleviate the domain gap between the base and novel classes. In this setting, the domain gap significantly affects the accuracy of novel class recognition. Therefore, learning models with strong cross-domain generalization capability is crucial for CD-FSL.

Under the few-shot setting [8,9], the novel classes are distinct from the base classes (but remain within the same domain), and each novel class has very few labeled samples. By explicitly constructing n-way k-shot classification scenarios during training and testing, many studies have effectively learned cross-task inductive biases [10] and reasoning mechanisms [11–13]. In recent years, state-of-the-art (SOTA) FSL methods [2,3,14,15] have built upon these baseline models with further optimizations, primarily focusing on metric learning and meta learning. Although efficient recognition of rare novel classes within the same domain has been achieved, existing FSL models often perform poorly when encountering with domain shifts between base and novel class tasks. Some highly effective meta-learning models even perform worse than simple fine-tuning models [6,16]. Domain adaptation (DA) typically mitigates domain shift by learning domain-invariant features through adversarial training, aiming to generalize the learned model to different domains. Mainstream DA models [17–21] focus on unsupervised domain adaptation, effectively enhancing the model's generalization ability in domain shift scenarios. However, it requires a large number of unlabeled samples in the target domain for training,

* Corresponding author.
E-mail address: h.yao@hit.edu.cn (H. Yao).

https://doi.org/10.1016/j.neucom.2025.130056
Received 15 December 2024; Received in revised form 2 March 2025; Accepted 15 March 2025
Available online 24 March 2025

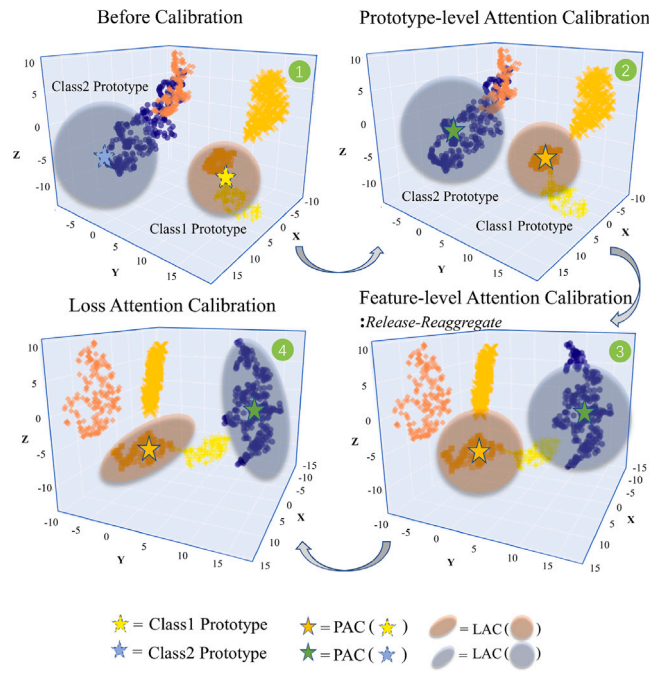


Fig. 1. The three-level calibration process diagram of P-R^2-L. ① shows the t-SNE representation of five classes in feature space, where the two ★ represent the class prototypes of two classes before calibration. Before calibration, due to the influence of noisy support instances, they are located at the edge of the distribution. After applying PAC, the pentagrams ★ are corrected in ② and relocated to the center of the class distribution. In ③, FAC suppresses the inherent distribution and refocuses on discriminative features, reconstructing a feature space with a better distribution. In ④, the *nuclear-norm* constraint replaces the cross-entropy constraint, refining the decision boundary. As a result, the CD-FSL task achieves improved generalization ability.

and the class labels between the source and target domains need to remain consistent. In CD-FSL tasks, the novel classes do not overlap with the base classes, and the training samples are extremely limited. Existing domain adaptation methods are not well-suited to effectively address CD-FSL challenges.

In essence, CD-FSL faces challenges posed by both FSL and DA issues. A simple combination of existing FSL and DA methods does not provide an effective solution. To address domain shift in few-shot settings, many approaches have been proposed in recent years. One type of approach [7,22,23] focuses on constructing complex training scenarios to prevent overfitting to knowledge from a specific domain. For instance, the study in [22] employs a "domain switching learning" strategy to rapidly switch training domains and impose constraints, simulating *n*-way *k*-shot training under cross-domain conditions. Similarly, ATA [23] considers the worst-case scenario of source domain distribution and employs task augmentation techniques to construct "challenging" virtual tasks, increasing the diversity of training tasks and effectively improving the model's robustness against domain shift. Another type of approach [7,24–27] simply integrates DA methods, incorporating adversarial learning and complex feature transformation for optimization. For instance, Tseng et al. [5] train a learnable feature transformation layer to simulate the distribution of image features extracted from tasks across different domains, effectively improving the generalization ability of metric-based models to unseen domains. Hu et al. [26] propose an adversarial feature augmentation (AFA) method that effectively aligns the distribution of target domain data, enhancing cross-domain performance. Although effective, these methods often face stringent learning conditions and complex learning processes during base model training. When transferred to new tasks, they typically require retraining or suffer from overfitting issues.

To avoid the stringent prerequisites and complex learning processes of existing methods, we propose a general, simple, and efficient learning framework, P-R^2-L, for handling source domain inductive biases. The framework suppresses and corrects the inherent erroneous inductive biases from the source domain through simple source domain attention release and target domain attention reaggregation. We implement three core components: the Prototype-level Attention Calibration (PAC) module, the Feature-level Attention Calibration (R^2-FAC) module for attention Release and Reaggregation, and the Loss Attention Calibration (LAC) module. First, we propose a PAC module, which highlights key instances by performing cross-image prototype calibration between the support and query set, reducing the impact of noisy instances in few-shot settings. Second, to suppress and correct erroneous discriminative inductive biases from the source domain, we propose the R^2-FAC module. This module sequentially integrates the base class attention release (BAR) submodule and the novel class attention reaggregation (NAR) submodule. It first weakens the erroneous attention from the source domain and then refocuses fine-grained discriminative information by cross-aligning attention between query and support images. Finally, we incorporate an LAC module into the loss function to balance the discriminability and diversity of the classification matrix, mitigating the decline in generalization ability caused by erroneous discriminative features during domain shift. The calibration process of our P-R^2-L framework is illustrated in Fig. 1. Through three levels of calibration — prototype, feature, and loss — we achieve an accurate cross-domain few-shot classifier.

Our contributions can be summarized as follows:

- We propose a similarity-weighted prototype-level attention calibration (PAC) module. This module weights intra-class instances based on similarity comparisons between support and query images to achieve prototype calibration, effectively mitigating the negative impact of atypical instances on class prototypes in few-shot scenarios.
- A cascaded feature-level attention calibration module (R^2-FAC) is proposed to suppress and correct erroneous discriminative inductive biases from the source domain. This module sequentially integrates the base class attention release (BAR) submodule and the novel class attention reaggregation (NAR) submodule to suppress erroneous inductive biases from the source domain and realign discriminative inductive information in the target domain, significantly improving generalization ability in cross-domain tasks.
- A loss attention calibration (LAC) module based on matrix *nuclear-norm* constraint is introduced. This module leverages the properties of the *nuclear-norm* to balance the discriminability and diversity of the classification matrix, effectively mitigating the degradation of generalization ability brought about by erroneous discriminative features when crossing domains.
- We propose a three-level general CD-FSL framework, P-R^2-L, to effectively address the issues of inductive bias and poor generalization in cross-domain few-shot recognition. Extensive experiments under the standard CD-FSL setting demonstrate that P-R^2-L achieves SOTA performance.

This paper is organized as follows: Section 2 reviews relevant research on FSL, DA, and CD-FSL. Section 3 offers an in-depth introduction to the proposed methods PAC, R^2-FAC, and LAC. Section 4 details the experimental setup, ablation studies, and analyses of results across various cross-domain datasets. Finally, Section 5 summarizes the findings and discusses future research directions.

2. Related work

This section briefly introduces related research areas to define and describe our proposed method. We primarily focus on *cross-domain few-shot learning* (in Section 2.3) and present the current state of research in two closely related fields: *few-shot learning* (in Section 2.1) and *domain adaptation* (in Section 2.2).

2.1. Few-shot learning

Few-shot learning explores how to build models that can effectively generalize to novel classes or tasks with limited labeled data. In recent years, most SOTA methods fall under the category of metric learning. Metric learning-based methods learn a mapping function that transforms data into a new feature space, minimizing the distance between same-class samples while maximizing the distance between different-class samples. Early metric learning works achieve significant breakthroughs by introducing schemes such as cosine similarity [28], class prototypes [29], and adaptive metrics [10]. However, these methods have been proven to perform poorly under CD-FSL settings, often yielding worse results [16] than simple fine-tuning approaches when faced with cross-domain tasks. Previous work [6] provides a comprehensive review of this topic. In recent years, the best FSL models have been built on these baselines, attempting to propose more reasonable metric strategies or conduct in-depth theoretical analyses. Examples include global feature augmentation [3], minimum matching cost [30], joint distribution distance metric [1], negative-margin loss [14], contrastive learning embedding [31], channel importance modulator [4], and hybrid feature fusion [32]. A large body of work has focused on class prototype optimization, such as multimodal prototype completion [33], generating representative prototype samples [34], and class prototype correction [35], among others. However, these methods often rely on fixed prototype generation or correction strategies, which may lead to performance degradation in cross-domain tasks due to instability in prototype modeling or distortion during prototype generation. Recent research introduces APPN [36], which proposes a plug-and-play model-adaptive resizer and the corresponding metric scheme (ASM). It demonstrates that improvements in model metric schemes can significantly enhance model performance. TAD [37] introduces the task attribute distance, which effectively quantifies the correlations between tasks and measures the adaptability challenges of different FSL models to new tasks. However, these FSL models share the common assumption that both base classes and novel classes come from the same domain, and none of them evaluate performance in cross-domain scenarios. They focus solely on performance optimization in classic few-shot scenarios with different classes within the same domain.

2.2. Domain adaptation

Domain adaptation aims to ensure that models trained on a source domain maintain strong performance when applied to a target domain. The basic assumption of domain adaptation is that the source and target domains share the same class and feature spaces, but differ in their data distributions.

Early domain adaptation methods typically rely on adapting shallow classification models to address domain shift issues, such as instance-based adaptation [38] and parameter-based adaptation [39,40].

Recently, influenced by deep neural networks and theoretical analyses [41] by Shai et al. mainstream unsupervised domain adaptation (UDA) methods focus on using CNNs to explore learning domain-invariant feature representations. Among these, metric-based UDA methods map features from the source and target domains to a common reproducing kernel Hilbert space using predefined distance metrics, thereby explicitly reducing the differences between the two domains. Representative methods include DDC [42], which aligns cross-domain feature distributions by minimizing Maximum Mean Discrepancy (MMD). Subsequently, DAN [43] and JAN [44] improve upon DDC [42] by minimizing Multi-Kernel Maximum Mean Discrepancy (MK-MMD) and Joint Maximum Mean Discrepancy (JMMD), respectively. The idea of adversarial domain adaptation originates from GANs [45], which use adversarial training to ensure that the features from the source and target domains significantly overlap in a common feature space, thus achieving domain adaptation. In this setup, the feature extractor and the domain classifier form an adversarial relationship: the domain classifier tries to correctly identify the domain, while the feature extractor aims to make it impossible for the classifier to determine the source of the features. Representative works include DANN [46], CDAN [47], ILA-DA [48], MCD [49], Li [50], LIN [51], PDA-Net [20], AAT [52], and others. Although effective, all UDA methods must ensure that the class label space remains consistent between the source and target domains. Some studies [53] explore scenarios with non-overlapping classes, but they still require partial class overlap and sufficient target domain samples for effective transfer. This differs significantly from the preconditions and task requirements of our cross-domain few-shot scenario.

2.3. Cross-domain few-shot learning

In CD-FSL tasks, novel classes and base classes come from different domains, with no overlap in class labels and extremely limited visible samples. In [6], extensive experiments on various FSL methods under cross-domain setting demonstrate that FSL fails to effectively handle significant domain shift issues. Clearly, due to the completely different task settings, existing domain adaptation methods cannot effectively address the CD-FSL problem. The targeted CD-FSL model and benchmark were first introduced in FWT [5]. Subsequent research mainly covers three types of approaches: methods learning generalized features [5,12,54,55], methods utilizing auxiliary networks [27,56,57], and pretraining-based methods [58–62]. In recent years, task augmentation methods and adversarial learning approaches that integrate pre-adaptation have gained significant attention. The typical ATA method [23] addresses the worst-case scenario of source task distribution, proposing an adversarial task augmentation strategy that generates inductive bias-adaptive challenging tasks, which can be conveniently applied to various meta-learning models. PCS [63] proposes an end-to-end prototypical cross-domain self-supervised learning framework that captures the semantic structure of categories in the data through intra-domain prototype contrastive learning and performs feature alignment through cross-domain prototype self-supervision. However, its performance heavily depends on the effectiveness of cross-domain modeling, and the implementation is relatively complex. FLoR [7] extends the analysis of loss landscapes from the parameter space to the representation space, employing normalization techniques to smooth sharp minima in the representation space, thereby achieving long-range flatness of the minima and enhancing transferability. Recent research proposes the ANIL [64] method, which improves the feature reuse and adaptability of meta-learning in few-shot cross-domain fault diagnosis by optimizing the inner-loop structure and introducing an adaptive loss function. GCC-FSL [65] enhances feature representation by using graph convolution to align the data distributions of the source and target domains, addressing the domain shift problem in cross-domain few-shot classification. PKEMTL [66] addresses the few-shot fault diagnosis problem under varying working conditions by introducing sequence tracking and self-supervised tasks for data augmentation, combined with multi-scale feature encoding and adaptive information fusion. ADAPTER [67] tackles the cross-domain few-shot learning problem under large domain shifts by employing a bidirectional cross-attention mechanism and the DINO training method, surpassing existing methods in the classic benchmark. Although these methods demonstrate some effectiveness, they often face strict learning conditions and complex learning processes during base model training. When transferred to new tasks, they typically require retraining or suffer from overfitting issues. We aim to achieve efficient few-shot domain generalization through a simple learning setup and framework (see Fig. 2). In summary, despite significant achievements in few-shot learning and domain adaptation within their respective fields [4,20,37,52], there remains a lack of simple and efficient learning methods for cross-domain few-shot tasks. Unlike existing methods, this paper proposes a simple and efficient few-shot

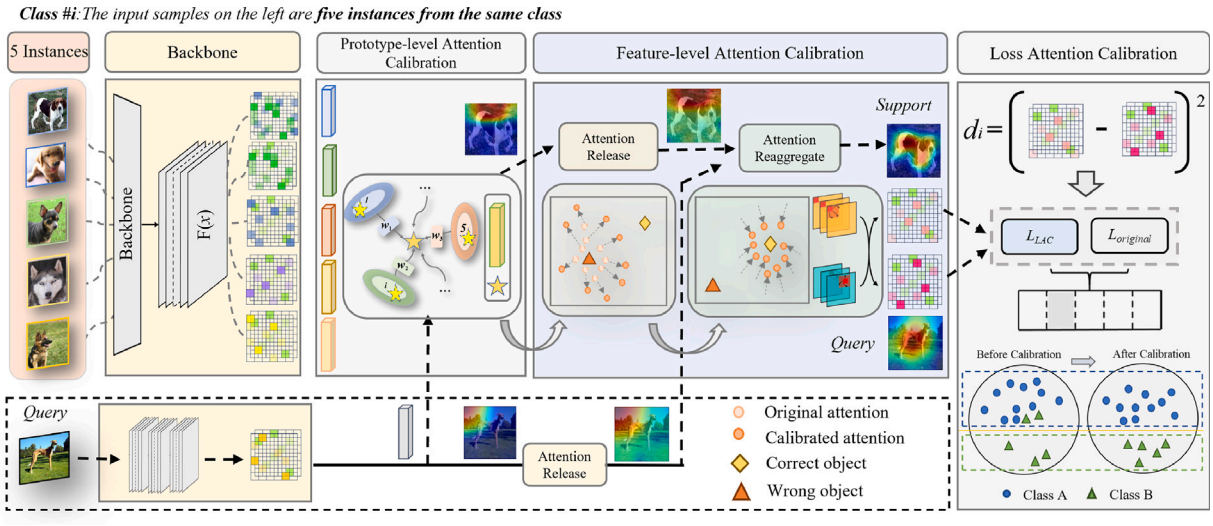


Fig. 2. The complete P-R^2-L framework diagram. We construct numerous 5-way 5-shot episodic training tasks, taking one of the meta-tasks as an example. ① PAC evaluates and reweights the importance of instances within a class to highlight key instances. ② FAC tandemly accesses attention release and attention reaggregation submodules to achieve source-domain erroneous discriminative information suppression and target-domain discriminative information relocalization, respectively. ③ By constraining the matrix *nuclear-norm*, LAC mitigates the loss function's over-concern with discriminative information and balances the discriminability and diversity of the classification results.

cross-domain learning framework (P-R^2-L) based on the straightforward idea of suppressing erroneous inherent biases from the source domain. It systematically explores how correcting the erroneous attention from the source domain at different stages can enhance the model's cross-domain generalization performance.

3. Proposed method

Formal Problem Definition. CD-FSL aims to recognize novel classes C^{novel} in the target domain using only a few training samples, by leveraging knowledge from base classes C^{base} in the source domain. Notably, $C^{novel} \cap C^{base} = \emptyset$, and there is a domain gap between these two class sets. The model is initially trained on a base-class dataset $D^{base} = \{x_i, y_i\}_{i=1}^{N}$, where $y_i \in C^{base}$. Training at this stage typically minimizes cross-entropy loss:

$$\mathcal{L} = \mathcal{L}_{cls}(G(x_i), y_i),$$

where $G(x_i) = h(g(x_i))$ outputs classification probabilities, consisting of a feature extractor $g(\cdot)$ and a classifier $h(\cdot)$. The feature extractor $g(\cdot)$ is then transferred to the novel-class dataset $D^{novel} = \{x_i^u, y_i^u\}_{i=1}^{N^u}$, where $y_i^u \in C^{novel}$. Due to limited training samples (1 or 5 per class), learning novel classes is challenging. For fair evaluation, current methods [7,22,23] employ n-way k-shot episodes for training and testing. Each episode consists of a support set $S = \{x_{i,j}^s, y_{i,j}^s\}_{i,j=1}^{n,k}$ for training and a query set $Q = \{x_i^q\}_{i=1}^{N_q^u}$ for evaluation. The model's prediction for a query sample x_i^q is:

$$\hat{y}_i^u = \arg\max G^u(x_i^q).$$

The model evaluates performance by sampling multiple tasks from the novel classes in the target domain and calculating the average accuracy.

Overall Framework. To address the CD-FSL problem, this paper proposes a novel P-R^2-L few-shot learning framework. As illustrated in Fig. 2, our method mainly consists of a PCA module, an R^2-FAC module, and an LAC module. Specifically, PAC enhances the contribution of key instances in n-shot tasks through prototype calibration, reducing the impact of atypical instances. In R^2-FAC, the base class attention release submodule and the novel class attention reaggregation submodule are integrated to suppress incorrect source domain discriminative inductive biases caused by domain shift and reaggregate fine-grained discriminative information in the target domain. LAC effectively balances the discriminability and diversity of classification results through matrix *nuclear-norm* constraint, mitigating the loss of generalization caused by incorrect cross-domain discriminative features. The entire network operates end-to-end, and the details of each component will be described below.

3.1. Prototype-level attention calibration module

For few-shot tasks, when the representation of a noisy instance differs significantly from others, this may result in a large deviation in the class prototype. As shown in Fig. 1, under the influence of atypical intra-class instances, the class prototype is likely to appear at the edge of the distribution, making it less representative. Existing models [4,7,27,37] typically use the average of all samples directly as the few-shot prototype, which can result in significant prototype bias under the influence of atypical instances. In this work, we propose a support-query cross-image PAC module to focus more attention on instances related to the query and reduce the impact of noise. We believe that, given a query, not all instances are equal and should be weighted according to their level of representativeness.

PAC performs prototype-level attention calibration based on the similarity weighting between support and query images. We first calculate the cosine distance between intra-class instances and the query image. Instances that are closer to the query sample will be assigned greater weights in the prototype representation. The prototype calibration process of the novel class is shown in Fig. 3, where *Instance-i* represents different instance representations within the support set, P is the calibrated prototype representation, and w_i represents the weight coefficients of different instances within the same class.

Specifically, cosine distance [13,28] is used to measure the similarity between the support instances of the novel class and the query image, and this similarity is normalized to obtain the similarity score m and the weight coefficient w. We represent the k instances in the nth class as: $S = \{\mathbf{s}_1^n, \mathbf{s}_2^n, \ldots, \mathbf{s}_k^n\}$. Then, the cosine distance between the ith query image $\mathbf{q}_i$ and the jth instance in the nth class can be expressed as:

$$D_{i,j} = 1 - \frac{\left\langle \mathbf{q}_i, \mathbf{s}_j^n \right\rangle}{\left\| \mathbf{q}_i \right\| \left\| \mathbf{s}_j^n \right\|}, \tag{1}$$

where, $\langle \cdot, \cdot \rangle$ refers to the dot product between two vectors, and $\| \cdot \|$ represents the L_2-norm.

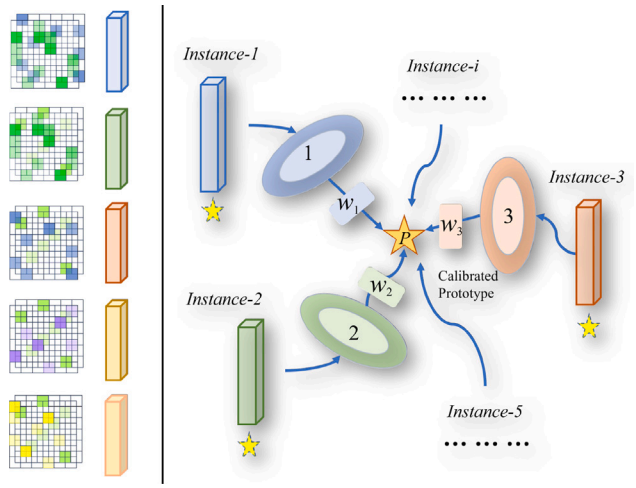


Fig. 3. The PAC module. We compute the cosine distance between the query image and all the intra-class instances in the support set to get the similarity coefficient and weighting coefficient of each instance with respect to the current query image, and reweight the key instances and noise instances.

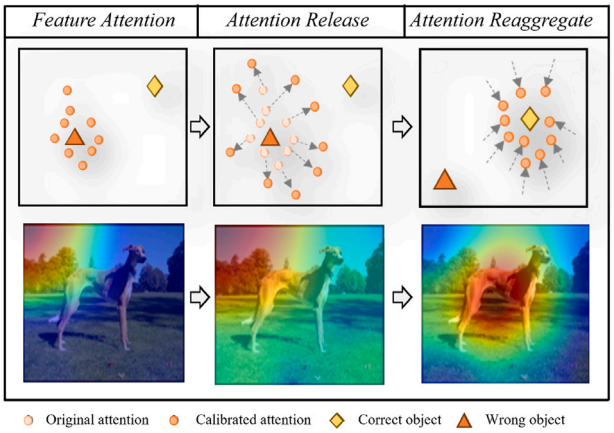


Fig. 4. Visualization of the attention transformation during R²-FAC calibration. The upper part of the figure briefly illustrates the movement trend of the attention factor: when attention is released, the attention factors are evenly scattered to the whole map region; when attention is reaggregated, the attention factors refocus on new discriminative regions. The lower part of the figure visualizes the attention heatmap before attention calibration, after attention release, and after attention reaggregation, respectively.

Then, the similarity coefficient is calculated using the computed cosine distance, and the similarity coefficient is normalized to obtain the weight coefficient. The similarity coefficient $m_{i,j}$ and weight coefficient $w_{i,j}$ can be expressed as follows:

$$m_{i,j} = \frac{1}{1 - \frac{\langle \mathbf{q}_i, \mathbf{s}_j^n \rangle}{\|\mathbf{q}_i\| \, \|\mathbf{s}_j^n\|}}, \tag{2}$$

$$w_{i,j} = \frac{e^{m_{i,j}}}{\sum_{j=1}^{k} e^{m_{i,j}}}. \tag{3}$$

Finally, the class prototype of the nth class is represented by weighting all instance representations as follows:

$$p_n = \sum_{j=1}^{k} w_{i,j} \cdot \mathbf{s}_j^n. \tag{4}$$

There are N calibrated prototype distributions for the N novel classes, which helps avoid bias caused by noisy samples. As shown in Fig. 3, the dark yellow pentagon represents the calibrated prototypes.

3.2. R^2-feature attention calibration module

In this section, we provide detailed information about R²-FAC (see Fig. 4), including (1) the base class attention release (BAR) module and (2) the novel class attention reaggregation (NAR) module. An overview of our proposed R²-FAC is shown in Fig. 5, with the main contributions being: suppression of source domain discriminative inductive biases during testing, and the reaggregation strategy for novel class discriminative inductive biases based on cross-alignment. These contributions will be discussed in detail below.

3.2.1. Base class attention release module

Large-scale and long-distance domain shifts often lead to significant changes in discriminative features in the novel domain. Existing models [7,22,23,27] typically fail to detect this in time and instead complete the testing phase tasks based on the experience learned from the base classes. We approach this from another perspective: if these discriminative features change from training to testing, and the image feature representations fail to respond appropriately to this change, there should be a specific feature transformation that corrects the feature representations, leading to performance improvement. Our proposed solution is to suppress the discriminative inductive biases from the source domain, thereby reducing the impact of incorrect inductive biases.

Let $\boldsymbol{I}$ be the input image. After feature extraction using the function $g(\cdot)$, we obtain the vector $\mathbf{v} = g(\boldsymbol{I}) \in \mathbb{R}^m$, where the ith dimension is defined as the ith component of the feature, i.e., $\{v_i\}_{i=1}^m$ represents the set of all m dimensions. We define the following transformation function to apply the transformation to each dimension of the feature:

$$\vec{\psi}_\alpha(v_i) = \alpha\sqrt{v_i + \epsilon} + (1-\alpha)\sigma(v_i), \quad i = 1, 2, \ldots, m, \tag{5}$$

where α is a weighting coefficient, $\sqrt{v_i + \epsilon}$ represents the square root transformation, and a small positive number ϵ is introduced to ensure numerical stability, especially when v_i approaches 0, to avoid gradient explosion. $\sigma(v_i)$ is the sigmoid activation function. If $v_i <= 0$, we directly set $\vec{\psi}_\alpha(v_i) = 0$.

For the feature vector $\mathbf{v}$, we can express the transformation for each dimension as follows:

$$\vec{\boldsymbol{\psi}}_\alpha(\mathbf{v}) = \left[\vec{\psi}_\alpha(v_1), \vec{\psi}_\alpha(v_2), \ldots, \vec{\psi}_\alpha(v_m)\right], \tag{6}$$

here, each dimension v_i undergoes the same non-linear transformation $\vec{\psi}_\alpha$, and ϵ ensures numerical stability, preventing gradient divergence when $v_i \to 0$.

During the testing phase, this function is applied dimension-wise to each feature representation of the image. That is, when this transformation is applied, all image features for the target classification task are transformed, regardless of whether they are in the support set or query set.

To further analyze the behavior of the transformation during optimization, we compute the gradient of the transformation function with respect to v_i. For $v_i \geq 0$, the gradient is:

$$\frac{\partial \vec{\psi}_\alpha(v_i)}{\partial v_i} = \alpha \frac{1}{2\sqrt{v_i + \epsilon}} + (1-\alpha)\sigma(v_i)(1-\sigma(v_i)), \tag{7}$$

here, ϵ is a small positive number that ensures smooth processing, preventing the gradient from diverging as v_i approaches 0. The first term $\frac{1}{2\sqrt{v_i+\epsilon}}$ ensures that as $v_i \to 0$, the gradient does not become infinite, while the second term is the derivative of $\sigma(v_i)$.

We adopt a zeroing strategy when $v_i \leq 0$, with the core motivation of ensuring computational stability, enhancing feature sparsity and balancing feature values. Negative features may carry non-discriminative information in certain cases or even disrupt feature alignment. Zeroing effectively removes these unstable factors, allowing the model to focus more on high-confidence discriminative features. Additionally, this approach helps reduce disparities between feature values, mitigating

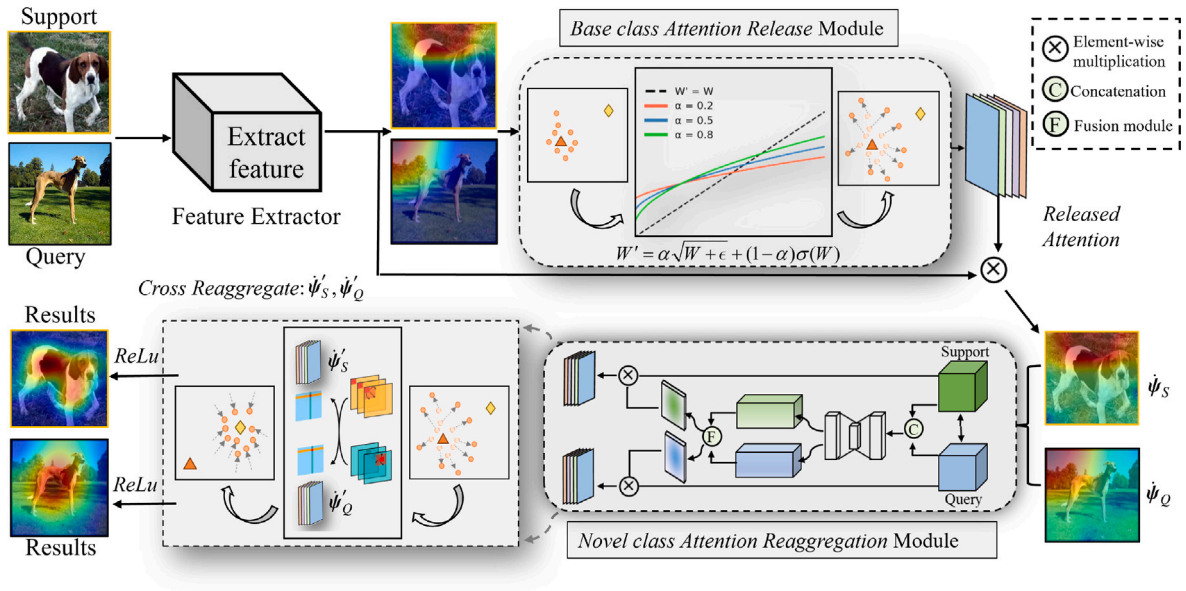


Fig. 5. The R^2-FAC module. R^2-FAC consists of base class attention release (BAR) and novel class attention reaggregation (NAR) submodules connected in series. The BAR submodule is shown on the upper part of the figure, which suppresses high-magnitude features and appropriately amplifies low-magnitude features to mitigate erroneous inductive bias from the source domain. Subsequently, the transformed feature maps $\dot{\psi}_S$ and $\dot{\psi}_Q$ are fed into the NAR submodule on the lower part for discriminative feature alignment across images to obtain accurate novel class cross-attention feature maps $\dot{\psi}'_S$ and $\dot{\psi}'_Q$.

the impact of extreme values and leading to a more uniform feature distribution, thereby improving the stability of cross-domain feature matching. Compared to alternative methods such as soft thresholding or learnable functions, our zeroing strategy achieves a better balance between computational efficiency, model convergence speed, and task generalization ability.

While direct zeroing can enhance computational stability and feature sparsity, it may also lead to information loss, affect gradient propagation, and potentially introduce negative impacts on cross-task adaptability and feature distribution. In some tasks, negative features may still play a crucial role. Therefore, future work could explore learnable functions or adaptive thresholding strategies as alternatives to direct zeroing to mitigate its negative effects.

One noticeable effect of this function is that it smooths the feature distribution: it suppresses high-amplitude features while appropriately amplifying low-amplitude features. The BAR module in Fig. 5 clearly illustrates this phenomenon, where we plot the feature amplitude transformation process under various choices of α. The black line represents the original distribution, while other colors represent the transformed distributions. The transformed distribution becomes more uniform. Larger feature magnitudes imply that the model places more emphasis on those features. After base class attention release, incorrect inductive biases caused by the source domain are suppressed.

3.2.2. Novel class attention reaggregation module

The NAR module structure is illustrated in the lower part of Fig. 5. For input images — support image $\boldsymbol{I}_S$ and query image $\boldsymbol{I}_Q$ — we denote the resulting feature maps after {BAR, NAR} as {$\dot{\boldsymbol{\psi}}_S$ and $\dot{\boldsymbol{\psi}}_Q$, $\dot{\boldsymbol{\psi}}'_S$ and $\dot{\boldsymbol{\psi}}'_Q$}.

As shown in Fig. 5, The first step computes a correlation map between $\dot{\boldsymbol{\psi}}_S$ and $\dot{\boldsymbol{\psi}}_Q$, which then guides the generation of the cross-alignment map. To accomplish this, the feature maps $\dot{\boldsymbol{\psi}}_S$ and $\dot{\boldsymbol{\psi}}_Q$ are reshaped into matrices of size $\mathbb{R}^{C\times m}$, where:

$$\dot{\boldsymbol{\psi}}_S = \left[\vec{\psi}_S^1, \vec{\psi}_S^2, \ldots, \vec{\psi}_S^m\right], \quad \dot{\boldsymbol{\psi}}_Q = \left[\vec{\psi}_Q^1, \vec{\psi}_Q^2, \ldots, \vec{\psi}_Q^m\right], \tag{8}$$

with $m = H \times W$ representing the number of spatial positions. Here, W, H, and C represent the width, height, and number of channels of the feature map, respectively. The terms $\vec{\psi}_S^i$ and $\vec{\psi}_Q^i$ refer to feature vectors at the ith spatial position of $\dot{\boldsymbol{\psi}}_S$ and $\dot{\boldsymbol{\psi}}_Q$. The cosine similarity between these vectors is then calculated to form a semantic relevance matrix:

$$\mathrm{CA}_{i,j} = \frac{\left\langle \vec{\psi}_Q^i, \vec{\psi}_S^j \right\rangle}{\left\|\vec{\psi}_Q^i\right\| \left\|\vec{\psi}_S^j\right\|}, i, j = 1, \ldots, m, \tag{9}$$

here, $\langle \cdot, \cdot \rangle$ refers to the dot product between two vectors, and $\| \cdot \|$ represents the L_2-norm. The matrix $\mathrm{CA}_{i,j}$ captures the local correlation between support and query features.

Learning Meta-Fusion Embedding. The meta-fusion layer generates cross-alignment maps by aligning corresponding positions according to the correlation matrix $\mathrm{CA}_{i,j}$. Taking the support feature alignment as an example, the goal is to reweight the spatial position coefficients of $\dot{\boldsymbol{\psi}}_S$ to align with the query feature $\dot{\boldsymbol{\psi}}_Q$. For this, each row of the correlation matrix is normalized to sum to 1, ensuring that it acts as a weight vector for $\dot{\boldsymbol{\psi}}_S$. The normalization is performed as follows:

$$\overline{\mathrm{CA}}_{i,j} = \frac{\exp(\mathrm{CA}_{i,j})}{\sum_{t=1}^{HW} \exp(\mathrm{CA}_{t,j})}. \tag{10}$$

The aligned support feature $\dot{\boldsymbol{\psi}}'_S \in \mathbb{R}^{C\times HW}$ is computed by multiplying the normalized correlation matrix $\overline{\mathrm{CA}}$ by the transpose of $\dot{\boldsymbol{\psi}}_S$ (The calculation for the query feature $\dot{\boldsymbol{\psi}}'_Q$ is similar):

$$\dot{\boldsymbol{\psi}}'_S = \left(\overline{\mathrm{CA}} \cdot \dot{\boldsymbol{\psi}}_S^T\right)^T. \tag{11}$$

Algorithm 1 provides the complete execution flow of the attention release and reaggregation mechanisms, which do not rely on accurate cross-domain shift modeling but instead address the channel bias issue in few-shot cross-domain tasks by optimizing key features. The BAR module releases erroneous attention in the source domain through a smoothing function, while the NAR module refocuses the key information of the target domain through cross-image attention alignment, optimizing the learning of target domain features. Our framework relies on an adaptive attention mechanism, dynamically adjusting feature weights to enhance the model's generalization ability. Even in cases where domain shift modeling is insufficient, it can still effectively focus on task-relevant features. The core computational processes in Eqs. (5) and (9)–(11) focus on feature smoothing and alignment, not depending on accurate cross-domain shift modeling, and are capable

Algorithm 1: Pseudo-code for **Attention Release and Reaggregation Strategy**

Require: Support image $\boldsymbol{I}_S$, query image $\boldsymbol{I}_Q$
Require: Weight coefficient α, stability constant ϵ
for *each support image* $\boldsymbol{I}_S$ *and query image* $\boldsymbol{I}_Q$ **do**
  Extract feature vectors v_i from $\boldsymbol{I}_S$
  for *each component* $i = 1, 2, \ldots, m$ **do**
    Apply feature transformation (Attention Release) as:
      $\vec{\psi}_\alpha(v_i) = \alpha\sqrt{v_i + \epsilon} + (1-\alpha)\sigma(v_i)$
    If $v_i \le 0$, set $\vec{\psi}_\alpha(v_i) = 0$
  end for
  Update feature vector as:
    $\dot{\boldsymbol{\psi}}_\alpha(\mathbf{v}) = [\vec{\psi}_\alpha(v_1), \ldots, \vec{\psi}_\alpha(v_m)]$
end for
for *each support map* $\dot{\boldsymbol{\psi}}_S$ *and query map* $\dot{\boldsymbol{\psi}}_Q$ **do**
  Reshape feature maps into matrices :
    $\dot{\boldsymbol{\psi}} = \left[\vec{\psi}^1, \vec{\psi}^2, \ldots, \vec{\psi}^m\right]$
  Compute cross-alignment matrix (CA):
    $\mathrm{CA}_{i,j} = \frac{\langle \vec{\psi}_Q^i, \vec{\psi}_S^j \rangle}{\|\vec{\psi}_Q^i\| \|\vec{\psi}_S^j\|}, \quad i, j = 1, 2, \ldots, m$
  Normalize the correlation matrix:
    $\overline{\mathrm{CA}}_{i,j} = \frac{\exp(\mathrm{CA}_{i,j})}{\sum_{t=1}^{HW} \exp(\mathrm{CA}_{t,j})}$
  Reaggregate discriminative features as:
    $\dot{\boldsymbol{\psi}}' = \left(\overline{\mathrm{CA}} \cdot \dot{\boldsymbol{\psi}}^T\right)^T$
end for
Return aligned features $\dot{\boldsymbol{\psi}}'_S$ and $\dot{\boldsymbol{\psi}}'_Q$.

of optimizing feature selection and attention allocation without perfect shifts modeling, thereby improving the model's adaptability across different tasks.

3.3. Loss attention calibration module

In CD-FSL tasks, it is normal for certain classes to dominate among randomly selected B batches of samples, while other categories contain few or even no samples. In such cases, models using traditional loss functions tend to rely on "discriminative information" to classify samples near the decision boundary into the majority class [68–70]. The continued convergence towards the majority class reduces prediction diversity, which is detrimental to cross-domain prediction accuracy. To achieve a more reasonable decision boundary, we introduce matrix *nuclear-norm* constraint in the loss function to balance the discriminability and diversity of the classification results.

Measuring Discriminability with *F-norm*. In traditional supervised learning, training with a sufficient number of labeled samples can lead to a well-distributed class representation and robust prediction performance. However, in few-shot scenarios, the data density near the decision boundary is high, especially for classes with fewer samples, which tend to be classified into the majority class [69,70]. To improve the prediction results in fine-grained few-shot classification, common methods [69–71] enhance discriminability by minimizing cross-entropy. Additionally, some methods [72] enhance classification performance by maximizing the *F-norm* of the classification matrix, which constrains misclassification behavior. Let us assume the model's prediction matrix for a batch of data is $\mathbf{M}$, where B and L represent the batch size and the number of classes, respectively. The optimization objective for such methods is usually:

$$\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^{B}\sum_{j=1}^{L} |M_{i,j}|^2}. \tag{12}$$

Since the *F-norm* of the classification matrix and the cross-entropy H(**M**) exhibit opposite monotonicity, both play similar roles when used as classification loss functions. Maximizing the *F-norm* has an effect similar to minimizing the cross-entropy.

Mathematically, the matrix *F-norm* can be constrained by the matrix *nuclear-norm* [73–75]:

$$\|\mathbf{M}\|_* \le \sqrt{\min(B, L)} \cdot \|\mathbf{M}\|_F \le \sqrt{\min(B, L)} \cdot B. \tag{13}$$

Therefore, optimizing the matrix *nuclear-norm* to its maximum can also drive the *F-norm* toward its maximum, thereby effectively enhancing the model's discriminability.

Measuring Diversity with Matrix Rank. Diversity can be approximately expressed as the number of predicted classes in the batch matrix—more predicted classes indicate greater response diversity. Considering the linear correlation of probability distributions across different classes, if two probability distributions belong to different classes, they will differ significantly and be linearly independent, whereas if they belong to the same class, they will be approximately linearly correlated. The number of predicted classes corresponds to the maximum number of linearly independent vectors in the matrix, which is the rank of the matrix. In other words, in the prediction of a randomly sampled batch of data, constraining the rank of the classification matrix $\mathbf{M}$ to be maximized can prevent the model's predictions from collapsing into the majority category. However, optimizing the matrix rank is an NP-Hard problem, making it challenging to train directly. Mathematically, the *nuclear-norm* serves as a convex approximation of its rank, allowing us to indirectly constrain the matrix's diversity by imposing a constraint on its *nuclear-norm*.

Loss for *Nuclear-norm* Maximization. Through the above analysis, we can conclude that maximizing the *F-norm* and the rank of the matrix can respectively enhance the matrix's discriminability and diversity, respectively. Since the *F-norm* and *nuclear-norm* can mathematically constrain each other and exhibit the same monotonicity, and the *nuclear-norm* is a convex approximation of the matrix rank, constraining the *nuclear-norm* alone can effectively balance the discriminability and diversity of the classification results.

Finally, for a randomly sampled batch of B samples, with the classification matrix represented as $\mathbf{M}(X)$, the loss function under the *nuclear-norm* constraint can be expressed as:

$$\mathcal{L}_{LAC} = -\frac{1}{B}\|\mathbf{M}(X)\|_*. \tag{14}$$

Fig. 6 compares the effects of entropy minimization and *nuclear-norm* constraint when processing the same classification matrix. For the hard-to-classify sample in the last row (with a probability distribution of [0.5, 0.5]), if entropy minimization is used to further optimize the model, the result tends to favor the majority class (shifting from [0.5, 0.5] to [0.6, 0.4]), with *Class 1* gaining the upper hand. However, if the matrix *nuclear-norm* constraint is used, *Class 2* gains the upper hand (shifting from [0.5, 0.5] to [0.4, 0.6]), effectively enhancing both the diversity and accuracy of the prediction results.

In CD-FSL tasks, the decline in generalization ability to novel classes caused by large domain shifts is often attributed to the inherent inductive bias of the source domain pretrained model (e.g., feature and channel bias). A significant amount of research has shown that discarding irrelevant features or channels from the source domain can effectively alleviate this issue and improve cross-domain generalization. Our method, through the BAR and NAR mechanisms in the R^2-FAC module, dynamically adjusts attention to suppress the inherent errors from the source domain's bias and optimize target domain feature selection. With this mechanism, the model can effectively mitigate the inherent inductive bias caused by large domain shifts between the source and target domains, enhancing the representational power of the target domain features. Furthermore, the P-R^2-L framework calibrates at three levels — prototype, feature, and loss — to reduce the impact of noisy instances, correct the inductive bias in the source domain, and

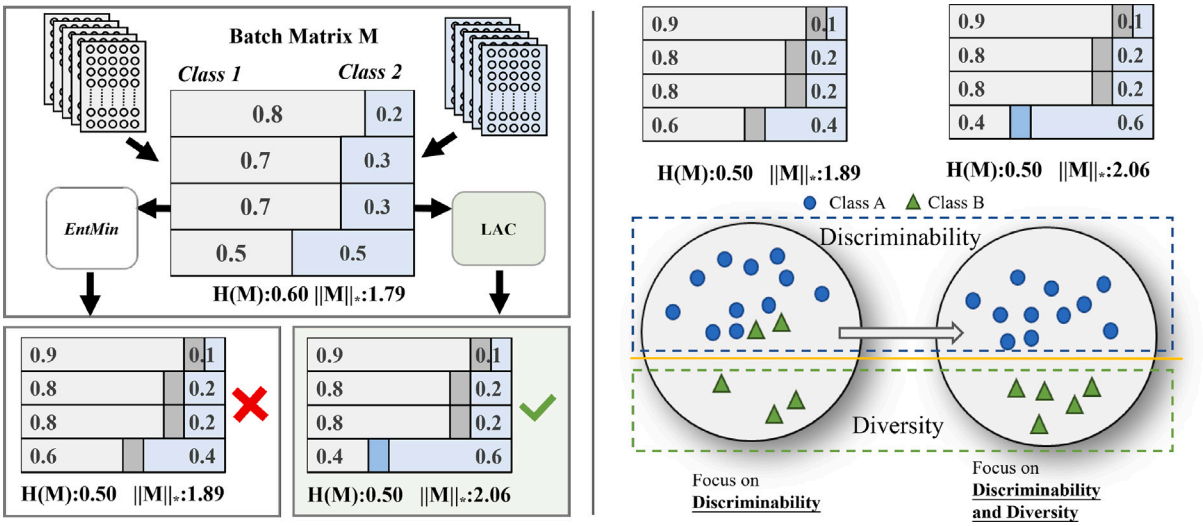


Fig. 6. Comparison of *nuclear-norm* constraint and entropy minimization (*EntMin*) on hard-to-classify samples. The figure assumes two classes, *Class 1* and *Class 2*, with **M** as the classification matrix. After processing with LAC and *EntMin*, new classification matrices are obtained. For the last row of hard-to-classify samples (with probability distribution [0.5,0.5]), the matrix *nuclear-norm* constraint (LAC) gives the minority *Class 2* an advantage (from [0.5, 0.5] to [0.4, 0.6]), improving classification diversity and accuracy. The dark regions indicate an increase in the variable, where dark gray (blue) signifies an increase in the gray (blue) variable. H(**M**) denotes the entropy value, while $\|\mathbf{M}\|_*$ represents the value of the *nuclear-norm*.

balance the discriminability and diversity. This enhances the model's adaptability and generalization performance in unseen domain shifts.

For the widely concerning issue of negative transfer in domain adaptation, our framework effectively addresses this problem through multiple modules. First, the BAR submodule in R^2-FAC reduces attention to irrelevant features in the source domain, suppressing the negative influence from the source domain and reducing the risk of negative transfer. Second, the NAR submodule minimizes interference from the inherent biases of the source domain by refocusing on the discriminative features of the target domain. Finally, the LAC module balances the diversity and discriminability of classification features within the loss function, preventing the misclassification of minority novel class instances as the dominant class, thus mitigating negative transfer caused by the source domain's dominant features.

4. Experimental result and analysis

In this paper, we primarily focus on CD-FSL tasks. For the proposed method, we will answer the following questions in the experimental section:

- **Q1.** In the ablation study, does each module demonstrate its necessity and advancement? (Section 4.3)
- **Q2.** How does our model's performance vary under different degrees of domain shift? (Sections 4.4 and 4.5)
- **Q3.** How does our model's performance vary with different numbers of support instances? (Sections 4.4 and 4.5)
- **Q4.** What does the visualization of the key results look like after processing the critical module? (Section 4.6)

4.1. Experimental settings

Dataset Overview. This paper primarily focuses on the classical CD-FSL problem in a single-source domain setting. To properly evaluate the cross-domain generalization performance, we follow the settings used in previous methods [6,7,16], using 64 base classes from *mini*-ImageNet [28] as the source domain and eight commonly used datasets — CUB [76], Cars [77], Places [78], Plantae [79], ChestX [80], ISIC [81], EuroSAT [82], and CropDisease [83] — as the target domains. As shown in Table 1, the first four datasets are benchmarks summarized in [5], consisting of natural images with various attributes and relatively small domain shifts, which we refer to as natural *near*-domain datasets. The latter four datasets are summarized in [6], originating from distinct fields such as medicine, remote sensing, and agriculture, and exhibit significant domain shifts, which we refer to as extreme *distant*-domain datasets. Experiments adopt the leave-one-out method, selecting the weights with the highest accuracy from the *mini*-ImageNet validation set for model evaluation in the target domains.

Table 1
Detailed information about the cross-domain datasets.

Domain shift	Dataset	Description	Classes
Natural *near-domain*	CUB [76]	Fine-grained birds	50
	Cars [77]	Fine-grained cars	49
	Plantae [79]	Plantae images	50
	Places [78]	Scene images	19
Extreme *distant-domain*	CropDiscases [83]	Crop disease images	38
	EuroSAT [82]	Satellite images	10
	ISIC [81]	Skin lesion images	7
	ChestX [80]	X-ray images	7

Benchmarking Methods. We conduct extensive evaluations of P-R^2-L on multiple SOTA benchmarks to verify its superiority. First, we compare P-R^2-L with classical few-shot learning methods [11,12,28] that utilize episodic training strategies. Then, we compare P-R^2-L with the optimal CD-FSL methods [7,23,26,27,83,84] that explore domain alignment and feature transformation. Additionally, as demonstrated by Guo et al. [6], traditional pretraining and fine-tuning methods outperform meta-learning methods in few-shot settings when domain shifts occur. In light of this, we further compare P-R^2-L with fine-tuning-based methods [6,23,24,84]. Overall, we thoroughly validated the effectiveness and superiority of our approach in addressing CD-FSL challenges compared to various types of methods.

4.2. Implementation details

Training settings. We uniformly use ResNet-10 [6,23,84] as the backbone network for feature extraction. The backbone's pretrained parameters are obtained by training the model on the source domain *mini*-ImageNet [28]. Following the protocol in [27], all input images are resized to 224 × 224. The momentum of the Adam optimizer is set to 0.9, with an initial learning rate of 0.001. For hyperparameters, we

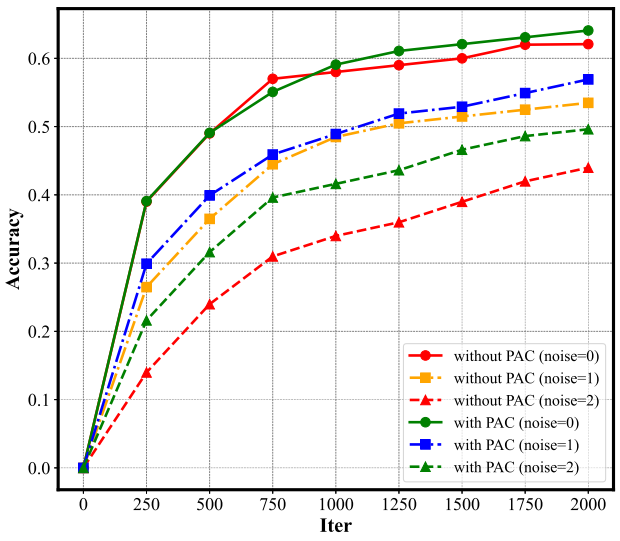


Fig. 7. Ablation study on PAC. Comparison of cross-domain experimental accuracy before and after applying PAC under different numbers (0,1,2) of noise instances. With the influence of PAC, the more the number of noise instances, the more significant the improvement in accuracy.

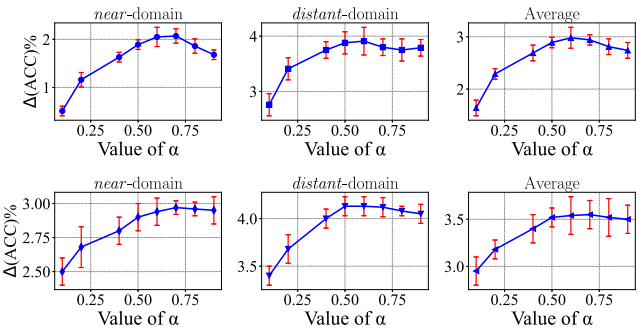


Fig. 8. Ablation study on α-values. Variation in accuracy with different values of the hyperparameter α across *near*-domain, *distant*-domain, and *full*-domain datasets. The optimal α value corresponding to the best accuracy varies with different levels of domain shift, generally falling within the range of 0.5 to 0.7. The upper and lower parts of the figure show the experimental results for the 5-way 1-shot and 5-way 5-shot settings, respectively.

set $\alpha = 0.6$. We evaluate the model on 600 randomly sampled episodes (with 15 query samples per class) under the 5-way 1-shot/5-shot setting and report the average accuracy (%). The experiments are conducted using the open-source PyTorch framework on an NVIDIA 2080 Ti GPU. By detailing our choices in optimization settings and parameters, we ensure the reproducibility of the experiments.

Evaluation metrics. The evaluation metrics of the model are consistent with the mainstream CD-FSL evaluation metrics [27,84]. For each target domain, we randomly sample 600 n-way k-shot 15-query tasks and record the average accuracy of these sampled tasks. The accuracy values shown in the tables represent the model's *top*-1 accuracy. We report the classification accuracy of different methods across 8 target domain datasets under both 5-way 1-shot and 5-way 5-shot experimental settings, respectively.

4.3. Ablation studies

We conduct ablation experiments with ResNet-10 [6,23,84] as the feature extractor in the 5-way 5-shot setting. As shown in Table 2, to further verify the effectiveness of each design in the framework, we apply PAC, BAR/NAR, R^2-FAC, and LAC individually or in combination to the baseline model and conduct experiments across eight cross-domain datasets. To conveniently demonstrate the model's performance under different degrees of domain shifts, we divide the datasets into *near*-domain, *distant*-domain, and *full*-domain categories and report the average classification accuracy for each.

In Table 2, a checkmark ($\surd$) is placed in the corresponding position if the module is used; otherwise, the space is left blank. Multiple checkmarks ($\surd$) indicate the simultaneous use of several modules. The R^2-FAC module is composed of both the BAR and NAR submodules.

The influence of PAC. To validate the module's impact on classification performance, we design multiple comparison experiments under 5-shot settings. We vary the number of noisy samples in the support set to 0, 1, and 2 to examine whether the module can effectively highlight key instances. Fig. 7 shows the average classification accuracy across 8 datasets, where the model with prototype calibration consistently performs better under varying levels of noisy sample interference. The higher the noise ratio, the more significant the performance improvement, which aligns with our expectation that PAC highlights key instances and mitigates the effect of atypical intra-class samples in few-shot settings.

The influence of the α value. In Fig. 8, we show how the hyperparameter α in Eq. (5) affects cross-domain few-shot classification performance. On both *near*-domain and *distant*-domain datasets, accuracy first increases and then decreases as α grows. The optimal α value varies with the degree of domain shift, typically ranging between 0.5 and 0.7. For datasets with larger domain shifts, smaller optimal α-values are usually observed. This is reasonable, as Eq. (5) indicates that smaller α-values help smooth attention, allowing the attention release function to effectively correct misaligned discriminative inductive biases from the source domain when the domain shift is large.

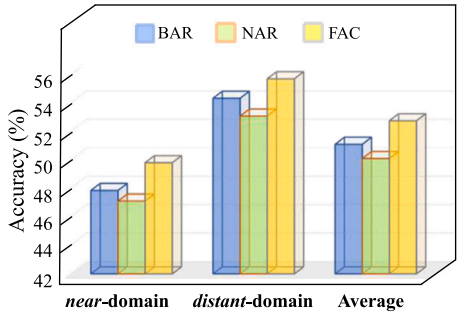
The influence of BAR/NAR. The R^2-FAC module consists of the BAR and NAR submodules. Fig. 9 shows the classification results after adding BAR, NAR, and R^2-FAC under both 1-shot and 5-shot settings. Models using each submodule individually perform better than those without any attention modules, demonstrating the effectiveness of both submodules. Table 2 presents the classification results when BAR appears alone or in combination with other modules across different test sets. In all cases, models using the BAR submodule alone outperform those without attention modules, confirming its effectiveness. Notably, performance improvements are more significant on datasets with larger domain shifts. BAR effectively weakens attention to discriminative features from the source domain, dispersing attention elsewhere. This characteristic alleviates erroneous inductive biases from the source domain in scenarios with large domain shifts, leading to improved classification accuracy.

The influence of FAC. The fourth row of Table 2 shows the classification results after adding the complete FAC. The experiments demonstrate that R^2-FAC exhibits strong generalization across all eight datasets and effectively improves the original model's classification performance. Comparing the results before and after adding NAR, it is evident that models with reaggregated attention perform better than those using only BAR. NAR aligns discriminative information across images, enabling more refined fine-grained feature extraction. By concatenating both modules, R^2-FAC suppresses and corrects incorrect inductive biases from the source domain, achieving optimal performance.

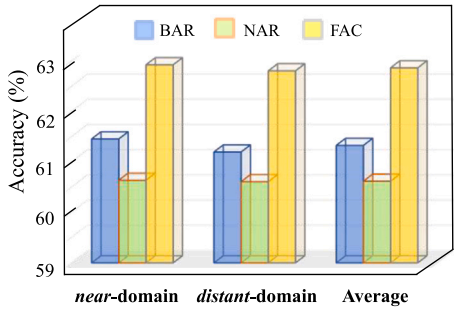
The influence of LAC. LAC is introduced as a constraint on the classification loss function. Focusing on rows 1 and 5, rows 2 and 8, and rows 7 and 10 in Table 2, we observe that the application of LAC consistently yields better results. Models with LAC show an average accuracy improvement of 1.14% on *near*-domain datasets and 1.06% on *distant*-domain datasets compared to the baseline. This effect may be attributed to LAC's ability to correct the boundaries of difficult-to-classify samples. By constraining the *nuclear-norm* of the classification matrix, LAC balances the discriminability and diversity of classification results, effectively mitigating the decline in generalization ability caused by

Table 2
Ablation study of 5-way 5-shot tasks trained with the *mini*-ImageNet dataset. The best results are displayed in **boldface** (mean ± S.D.%). Numbers are in percentage (%). The checkmark (√) indicates that the module is used.

Baseline	PAC	BAR	R^2-FAC	LAC	*Ave:near*-domain	*Ave:distant*-domain	Average
√					58.50 ± 0.31	57.10 ± 0.27	57.80 ± 0.29
√	√				59.41 ± 0.30	58.20 ± 0.31	58.81 ± 0.31
√		√			61.46 ± 0.34	61.20 ± 0.26	61.33 ± 0.30
√			√		62.93 ± 0.28	62.81 ± 0.30	62.87 ± 0.28
√				√	59.64 ± 0.35	58.16 ± 0.31	58.90 ± 0.34
√	√	√			61.54 ± 0.41	62.11 ± 0.26	61.83 ± 0.29
√	√		√		63.66 ± 0.30	63.11 ± 0.25	63.39 ± 0.27
√	√			√	59.90 ± 0.36	58.42 ± 0.29	59.16 ± 0.32
√	√	√		√	63.67 ± 0.28	63.24 ± 0.25	63.46 ± 0.26
√	√		√	√	**64.34 ± 0.35**	**63.81 ± 0.31**	**64.08 ± 0.32**



(a) Results for 1-shot settings



(b) Results for 5-shot settings

Fig. 9. Results of ablation experiments for BAR, NAR, and FAC. Blue, green, and yellow respectively represent the classification accuracy when BAR, NAR, and FAC are applied individually. (a) Classification accuracy on datasets with different domain shifts under 1-shot settings. (b) Classification accuracy on datasets with different domain shifts under 5-shot settings.

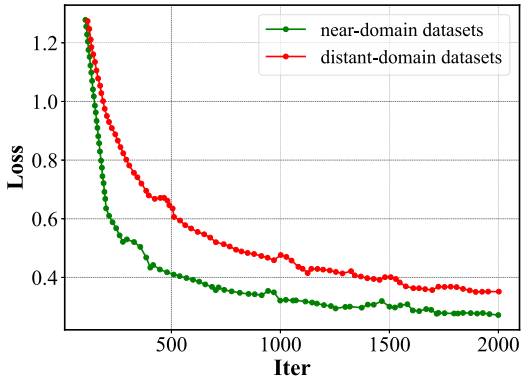


Fig. 10. Loss of our model on both *near*-domain and *distant*-domain datasets.

Table 3
Comparison of model parameter count and complexity. PC: Parameter Count; GFLOPs: Floating-point Operations; GIT: GPU (NVIDIA 2080 Ti) Inference Time for a task. The backbone network is set to ResNet-12 for both.

Methods	PC	5-way 1-shot		5-way 5-shot	
		GFLOPs	GIT	GFLOPs	GIT
ProtoNet	8.04M	101.550	0.00960 s	126.938	0.0130 s
Ours	8.29M	101.909	0.00971 s	127.519	0.0156 s

erroneous discriminative features during domain shifts. Based on the loss curves during training (Fig. 10), we observe stable convergence for both the *near*-domain dataset (green curve) and the *distant*-domain dataset (red curve). Specifically, both curves exhibit a rapid decrease in the early stages of training and stabilize in the later stages, indicating that our model optimizes stably and converges effectively, validating the stability and reliability of the model across different datasets. Notably, the loss for the *distant*-domain dataset decreases more slowly, suggesting that the learning process is more challenging, likely due to the larger distributional gap between the data and the target task.

Complexity Comparison. The additional parameter count introduced by the P-R^2-L framework is mainly dominated by the R^2-FAC module, and the overall complexity is $O(k \times m + m^2)$. The additional computational complexity of the framework is $O(k \times m + m \times H \times W + B \times L)$, where k is the number of classes, m is the number of samples per class, H and W are the height and width of the feature map, and B and L are the batch size and the number of classes. To illustrate the model's complexity, we report the parameter count (PC), floating-point operations (GFLOPs), and GPU inference time (GIT) for the 5-way 1-shot and 5-way 5-shot tasks. As shown in Table 3, although P-R^2-L slightly increases the computational cost and inference time compared to the classic network ProtoNet [29], its parameter count is only 0.25M larger than that of ProtoNet, which makes the model more complex and leads to a significant performance improvement. In terms of computational overhead, P-R^2-L introduces only a small increase in computation (approximately 0.35% for 1-shot and 0.46% for 5-shot).

According to Table 2, the combination of PAC, R^2-FAC, and LAC yields the highest performance. In summary, the ablation experiments confirm the modules' effectiveness and demonstrate their strong generalization capability.

4.4. Numerical results: Comparison with state-of-the-art

To validate the effectiveness of the proposed method, in this section, we compare P-R^2-L with classical FSL methods [11,12,28] based on episodic training strategies and the optimal CD-FSL methods [7,23,26,27,83,84] exploring domain alignment and feature transformation. To ensure fairness, we group these experiments based on whether fine-tuning (FT) or transductive learning (TR) is used.

The experiments are conducted under the n-way k-shot setting, with the test data being the eight target domain datasets mentioned earlier. To analyze the model's performance under varying degrees of domain shift, we divide the datasets into natural *near*-domain datasets (CUB [76], Cars [77], Places [78], and Plantae [79]) and extreme

Table 4
Few-shot results with different settings of backbones (Conv-4 and ResNet-10). The best results are displayed in **boldface** (mean ± S.D.%). Numbers are in percentage (%).

Methods	Backbones	*Ave:near*-domain		*Ave:distant*-domain		Average	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
GNN+AFA [26]	Conv-4	$30.35_{\pm 0.4}$	$49.04_{\pm 0.3}$	$40.18_{\pm 0.4}$	$54.57_{\pm 0.4}$	$35.27_{\pm 0.4}$	$51.81_{\pm 0.3}$
	ResNet-10	$42.98_{\pm 0.5}$	$62.00_{\pm 0.4}$	$46.49_{\pm 0.4}$	$59.64_{\pm 0.4}$	$44.74_{\pm 0.4}$	$60.82_{\pm 0.4}$
LDP-net [27]	Conv-4	$32.18_{\pm 0.4}$	$50.06_{\pm 0.3}$	$40.98_{\pm 0.5}$	$55.69_{\pm 0.3}$	$36.58_{\pm 0.5}$	$52.88_{\pm 0.3}$
	ResNet-10	$43.19_{\pm 0.3}$	$61.16_{\pm 0.4}$	$47.93_{\pm 0.4}$	$61.54_{\pm 0.4}$	$45.56_{\pm 0.3}$	$61.35_{\pm 0.4}$
Ours	Conv-4	$\mathbf{33.71_{\pm 0.4}}$	$\mathbf{52.04_{\pm 0.4}}$	$\mathbf{42.43_{\pm 0.5}}$	$\mathbf{57.11_{\pm 0.3}}$	$\mathbf{38.07_{\pm 0.5}}$	$\mathbf{54.58_{\pm 0.4}}$
	ResNet-10	$\mathbf{45.82_{\pm 0.3}}$	$\mathbf{64.34_{\pm 0.4}}$	$\mathbf{50.44_{\pm 0.3}}$	$\mathbf{63.81_{\pm 0.3}}$	$\mathbf{48.13_{\pm 0.3}}$	$\mathbf{64.08_{\pm 0.4}}$

Table 5
Classification accuracy (%) of 5-way 1-shot/5-shot tasks on *near*-domain datasets, trained with the *mini*-ImageNet dataset. **TR** stands for exploiting the full data of FSL task. Numbers are in percentage (%). The best results are highlighted in **bold** (mean ± S.D.%).

Model	TR	CUB		Cars		Places		Plantae	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MatchingNet [28] (2016)	✗	$35.89_{\pm 0.5}$	$51.37_{\pm 0.8}$	$30.77_{\pm 0.5}$	$38.99_{\pm 0.6}$	$49.86_{\pm 0.8}$	$63.16_{\pm 0.8}$	$32.70_{\pm 0.6}$	$46.53_{\pm 0.7}$
GNN [11] (2017)	✗	$44.40_{\pm 0.5}$	$62.87_{\pm 0.5}$	$31.72_{\pm 0.4}$	$43.70_{\pm 0.4}$	$52.42_{\pm 0.5}$	$70.91_{\pm 0.5}$	$33.60_{\pm 0.4}$	$48.51_{\pm 0.4}$
GNN+ATA [23] (2021)	✗	$45.00_{\pm 0.5}$	$66.22_{\pm 0.5}$	$33.61_{\pm 0.4}$	$49.14_{\pm 0.4}$	$53.57_{\pm 0.5}$	$75.48_{\pm 0.4}$	$34.42_{\pm 0.4}$	$52.69_{\pm 0.4}$
PCS [63] (2021)	✗	$43.44_{\pm 0.4}$	$66.10_{\pm 0.6}$	$34.42_{\pm 0.5}$	$51.26_{\pm 0.4}$	$54.03_{\pm 0.4}$	$73.65_{\pm 0.4}$	$36.02_{\pm 0.5}$	$54.12_{\pm 0.3}$
MN+AFA [26] (2022)	✗	$41.02_{\pm 0.4}$	$59.46_{\pm 0.4}$	$33.52_{\pm 0.4}$	$46.13_{\pm 0.4}$	$\mathbf{54.66_{\pm 0.5}}$	$68.87_{\pm 0.4}$	$37.60_{\pm 0.4}$	$52.43_{\pm 0.4}$
GNN+AFA [26] (2022)	✗	$46.86_{\pm 0.5}$	$68.25_{\pm 0.5}$	$34.25_{\pm 0.4}$	$49.28_{\pm 0.5}$	$54.04_{\pm 0.6}$	$\mathbf{76.21_{\pm 0.5}}$	$36.76_{\pm 0.4}$	$54.26_{\pm 0.4}$
LDP-net [27] (2023)	✗	49.82	70.39	35.51	52.84	53.82	72.90	39.84	58.49
FLoR [7] (2024)	✗	49.99	70.39	37.41	53.43	53.18	72.31	40.10	55.80
Ours	✗	$\mathbf{50.44_{\pm 0.3}}$	$\mathbf{71.05_{\pm 0.3}}$	$\mathbf{38.20_{\pm 0.4}}$	$\mathbf{54.38_{\pm 0.4}}$	$53.72_{\pm 0.3}$	$73.41_{\pm 0.4}$	$\mathbf{40.92_{\pm 0.3}}$	$\mathbf{58.51_{\pm 0.3}}$
TPN [12] (2018)	✓	$48.30_{\pm 0.4}$	$63.52_{\pm 0.4}$	$32.42_{\pm 0.4}$	$44.54_{\pm 0.4}$	$56.17_{\pm 0.5}$	$71.39_{\pm 0.4}$	$37.40_{\pm 0.4}$	$50.96_{\pm 0.4}$
TPN+ATA [23] (2021)	✓	$50.26_{\pm 0.5}$	$65.31_{\pm 0.4}$	$34.18_{\pm 0.4}$	$46.95_{\pm 0.4}$	$57.03_{\pm 0.5}$	$72.12_{\pm 0.4}$	$39.83_{\pm 0.4}$	$55.08_{\pm 0.4}$
TPN+AFA [26] (2022)	✓	$50.85_{\pm 0.4}$	$65.86_{\pm 0.4}$	$38.43_{\pm 0.4}$	$47.89_{\pm 0.4}$	$60.29_{\pm 0.5}$	$72.81_{\pm 0.4}$	$40.27_{\pm 0.4}$	$55.67_{\pm 0.4}$
RDC [84] (2022)	✓	$48.68_{\pm 0.5}$	$64.36_{\pm 0.4}$	$38.26_{\pm 0.5}$	$52.15_{\pm 0.4}$	$59.53_{\pm 0.5}$	$73.24_{\pm 0.4}$	$\mathbf{42.29_{\pm 0.5}}$	$\mathbf{57.50_{\pm 0.4}}$
FLoR [7] (2024)	✓	55.35	70.83	38.86	53.55	60.94	73.88	41.61	56.28
Ours	✓	$\mathbf{55.42_{\pm 0.5}}$	$\mathbf{71.14_{\pm 0.5}}$	$\mathbf{39.16_{\pm 0.4}}$	$\mathbf{53.99_{\pm 0.4}}$	$\mathbf{61.13_{\pm 0.5}}$	$\mathbf{74.24_{\pm 0.4}}$	$42.23_{\pm 0.4}$	$57.17_{\pm 0.4}$

Table 6
Classification accuracy (%) of 5-way 1-shot/5-shot tasks on *distant*-domain datasets, trained with the *mini*-ImageNet dataset. **TR** stands for exploiting the full data of FSL task. Numbers are in percentage (%). The best results are highlighted in **bold** (mean ± S.D.%).

Model	TR	CropDiseases		EuroSAT		ISIC		ChestX	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MatchingNet [28] (2016)	✗	$57.57_{\pm 0.5}$	$73.26_{\pm 0.5}$	$54.19_{\pm 0.5}$	$67.50_{\pm 0.5}$	$29.62_{\pm 0.3}$	$32.98_{\pm 0.3}$	$22.30_{\pm 0.2}$	$22.85_{\pm 0.2}$
GNN [11] (2017)	✗	$59.19_{\pm 0.5}$	$83.12_{\pm 0.4}$	$54.61_{\pm 0.5}$	$78.69_{\pm 0.4}$	$30.14_{\pm 0.3}$	$42.54_{\pm 0.4}$	$21.94_{\pm 0.2}$	$23.87_{\pm 0.2}$
GNN+ATA [23] (2021)	✗	$67.45_{\pm 0.5}$	$90.59_{\pm 0.3}$	$61.35_{\pm 0.5}$	$83.75_{\pm 0.4}$	$33.21_{\pm 0.4}$	$44.91_{\pm 0.4}$	$22.10_{\pm 0.2}$	$24.32_{\pm 0.4}$
PCS [63] (2021)	✗	$66.04_{\pm 0.4}$	$89.94_{\pm 0.4}$	$62.14_{\pm 0.3}$	$83.98_{\pm 0.4}$	$31.96_{\pm 0.5}$	$42.65_{\pm 0.5}$	$21.83_{\pm 0.4}$	$25.63_{\pm 0.3}$
MN+AFA [26] (2022)	✗	$60.71_{\pm 0.5}$	$80.07_{\pm 0.4}$	$61.28_{\pm 0.5}$	$69.63_{\pm 0.5}$	$32.32_{\pm 0.3}$	$39.88_{\pm 0.3}$	$22.11_{\pm 0.2}$	$23.18_{\pm 0.2}$
GNN+AFA [26] (2022)	✗	$67.61_{\pm 0.5}$	$88.06_{\pm 0.3}$	$63.12_{\pm 0.5}$	$\mathbf{85.58_{\pm 0.4}}$	$33.21_{\pm 0.3}$	$46.01_{\pm 0.4}$	$22.92_{\pm 0.2}$	$25.02_{\pm 0.2}$
LDP-net [27] (2023)	✗	69.64	89.40	**65.11**	82.01	33.97	48.06	23.01	26.67
FLoR [7] (2024)	✗	73.64	91.25	62.90	80.87	38.11	51.44	23.11	26.70
Ours	✗	$\mathbf{74.60_{\pm 0.3}}$	$\mathbf{92.36_{\pm 0.4}}$	$63.79_{\pm 0.3}$	$82.12_{\pm 0.3}$	$\mathbf{39.17_{\pm 0.3}}$	$\mathbf{52.74_{\pm 0.3}}$	$\mathbf{24.19_{\pm 0.3}}$	$\mathbf{28.02_{\pm 0.3}}$
TPN [12] (2018)	✓	$68.39_{\pm 0.6}$	$81.91_{\pm 0.5}$	$63.90_{\pm 0.5}$	$77.22_{\pm 0.4}$	$35.08_{\pm 0.4}$	$45.66_{\pm 0.3}$	$21.05_{\pm 0.2}$	$22.17_{\pm 0.2}$
TPN+ATA [23] (2021)	✓	$77.82_{\pm 0.5}$	$88.15_{\pm 0.5}$	$65.94_{\pm 0.5}$	$79.47_{\pm 0.3}$	$34.70_{\pm 0.4}$	$45.83_{\pm 0.3}$	$21.67_{\pm 0.2}$	$23.60_{\pm 0.2}$
TPN+AFA [26] (2022)	✓	$72.44_{\pm 0.6}$	$85.69_{\pm 0.4}$	$66.17_{\pm 0.4}$	$80.12_{\pm 0.4}$	$34.25_{\pm 0.4}$	$46.29_{\pm 0.3}$	$21.69_{\pm 0.1}$	$23.47_{\pm 0.2}$
RDC [84] (2022)	✓	$79.72_{\pm 0.5}$	$88.90_{\pm 0.3}$	$65.58_{\pm 0.5}$	$77.15_{\pm 0.4}$	$32.33_{\pm 0.3}$	$41.28_{\pm 0.3}$	$22.77_{\pm 0.2}$	$25.91_{\pm 0.2}$
FLoR [7] (2024)	✓	85.95	92.32	70.96	82.04	39.78	52.16	22.92	26.27
Ours	✓	$\mathbf{86.35_{\pm 0.4}}$	$\mathbf{93.24_{\pm 0.3}}$	$\mathbf{71.57_{\pm 0.4}}$	$\mathbf{82.98_{\pm 0.4}}$	$\mathbf{40.35_{\pm 0.4}}$	$\mathbf{53.24_{\pm 0.5}}$	$\mathbf{23.64_{\pm 0.3}}$	$\mathbf{27.48_{\pm 0.3}}$

distant-domain datasets (ChestX [80], ISIC [81], EuroSAT [82], and CropDisease [83]). Tables 5 and 6 present the experimental results for 5-way 1-shot and 5-way 5-shot settings across different domain shifts.

To verify the generalizability and applicability of our method to different backbone networks, we experimented with commonly used feature extraction backbones [28,29], Conv-4 and ResNet-10. As shown in Table 4, by comparing the results from these different backbone networks, we observe that our proposed method performs competitively on both the lightweight Conv-4 and the more complex ResNet-10. Given our model's sensitivity to fine-grained feature extraction, all subsequent experiments will be conducted using ResNet-10, which is better suited for feature extraction.

Results for *Near*-Domain Datasets. For datasets with smaller domain shifts, such as CUB, Cars, Places, and Plantae, our method generally outperforms other SOTA methods [7,23,26,27,84] in both 1-shot and 5-shot settings. Specifically, in 5-way tasks, compared to the second-best methods FLoR [7], our model achieves accuracy improvements of 0.66%, 0.95%, and 2.71% on CUB, Cars, and Plantae, respectively. However, compared to GNN+AFA [26], the accuracy on Places decreases. We attribute this to the characteristics of the base class attention release submodule. Since Places overlaps with the source domain *mini*-ImageNet, it is classified as clearly *near*-domain data. Due to the suppression of discriminative features similar to those in the source domain by the attention release module, the accuracy on this dataset has decreased. CUB and Cars are classic fine-grained classification datasets, and the strong performance on these two datasets demonstrates the excellent fine-grained classification ability of our model. Combined with the ablation experiments (Table 2), it is evident that this performance improvement is mainly due to the attention release and reaggregation modules, which effectively correct and extract fine-grained information for challenging fine-grained classification tasks.

Table 7
Classification accuracy (%) of 5-way 1-shot/5-shot tasks on *near*-domain datasets, trained with the *mini*-ImageNet dataset. **FT** stands for fine-tuning on target domain, **TR** stands for exploiting the full data of FSL task. Numbers are in percentage (%). The best results are highlighted in **bold** (mean ± S.D.%).

Model	FT	TR	CUB		Cars		Places		Plantae	
			1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Fine-tuning [6] (2020)	✓	✗	$43.53_{\pm 0.4}$	$63.76_{\pm 0.4}$	$35.12_{\pm 0.4}$	$51.21_{\pm 0.4}$	$50.57_{\pm 0.4}$	$70.68_{\pm 0.4}$	$38.77_{\pm 0.4}$	$56.45_{\pm 0.4}$
NSAE [24] (2021)	✓	✗	–	$68.51_{\pm 0.8}$	–	$59.41_{\pm 0.7}$	–	$71.02_{\pm 0.7}$	–	$59.55_{\pm 0.7}$
FLoR [7] (2024)	✓	✗	50.01	73.39	38.13	57.21	53.61	72.37	40.20	61.11
Ours	✓	✗	$\mathbf{50.13_{\pm 0.4}}$	$\mathbf{73.75_{\pm 0.4}}$	$\mathbf{38.62_{\pm 0.3}}$	$\mathbf{58.26_{\pm 0.4}}$	$\mathbf{54.04_{\pm 0.5}}$	$\mathbf{73.50_{\pm 0.4}}$	$\mathbf{40.92_{\pm 0.3}}$	$\mathbf{62.43_{\pm 0.3}}$
TPN+ATA [23] (2021)	✓	✓	$51.89_{\pm 0.5}$	$70.14_{\pm 0.4}$	$38.07_{\pm 0.4}$	$55.23_{\pm 0.4}$	$57.26_{\pm 0.5}$	$73.87_{\pm 0.4}$	$40.75_{\pm 0.4}$	$59.02_{\pm 0.4}$
RDC [84] (2022)	✓	✓	$51.20_{\pm 0.5}$	$67.77_{\pm 0.4}$	$39.13_{\pm 0.5}$	$53.75_{\pm 0.5}$	$61.50_{\pm 0.6}$	$74.65_{\pm 0.4}$	$\mathbf{44.33_{\pm 0.6}}$	$60.63_{\pm 0.4}$
FLoR [7] (2024)	✓	✓	55.94	74.06	40.01	57.98	61.27	74.25	41.70	61.70
Ours	✓	✓	$\mathbf{56.35_{\pm 0.5}}$	$\mathbf{74.82_{\pm 0.5}}$	$\mathbf{40.80_{\pm 0.3}}$	$\mathbf{59.43_{\pm 0.3}}$	$\mathbf{61.95_{\pm 0.3}}$	$\mathbf{75.24_{\pm 0.5}}$	$42.72_{\pm 0.4}$	$\mathbf{63.00_{\pm 0.4}}$

Table 8
Classification accuracy (%) of 5-way 1-shot/5-shot tasks on *distant*-domain datasets, trained with the *mini*-ImageNet dataset. **FT** stands for fine-tuning on target domain, **TR** stands for exploiting the full data of FSL task. Numbers are in percentage (%). The best results are highlighted in **bold** (mean ± S.D.%).

Model	FT	TR	CropDiseases		EuroSAT		ISIC		ChestX	
			1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Fine-tuning [6] (2020)	✓	✗	$73.43_{\pm 0.5}$	$89.84_{\pm 0.3}$	$66.17_{\pm 0.5}$	$81.59_{\pm 0.3}$	$34.60_{\pm 0.3}$	$49.51_{\pm 0.3}$	$22.13_{\pm 0.2}$	$25.37_{\pm 0.2}$
NSAE [24] (2021)	✓	✗	–	$93.14_{\pm 0.5}$	–	$83.96_{\pm 0.6}$	–	$54.04_{\pm 0.6}$	–	$27.10_{\pm 0.4}$
FLoR [7] (2024)	✓	✗	84.04	92.33	69.13	83.06	38.81	56.74	23.12	26.77
Ours	✓	✗	$\mathbf{84.76_{\pm 0.4}}$	$\mathbf{93.19_{\pm 0.4}}$	$\mathbf{70.10_{\pm 0.3}}$	$\mathbf{84.61_{\pm 0.3}}$	$\mathbf{39.65_{\pm 0.3}}$	$\mathbf{57.67_{\pm 0.3}}$	$\mathbf{24.14_{\pm 0.3}}$	$\mathbf{28.38_{\pm 0.2}}$
TPN+ATA [23] (2021)	✓	✓	$82.47_{\pm 0.5}$	$93.56_{\pm 0.2}$	$70.84_{\pm 0.5}$	$85.47_{\pm 0.3}$	$35.55_{\pm 0.4}$	$49.83_{\pm 0.3}$	$22.45_{\pm 0.2}$	$24.74_{\pm 0.2}$
RDC [84] (2022)	✓	✓	$86.33_{\pm 0.5}$	$93.55_{\pm 0.3}$	$71.57_{\pm 0.5}$	$84.67_{\pm 0.3}$	$35.84_{\pm 0.4}$	$49.06_{\pm 0.3}$	$22.27_{\pm 0.2}$	$25.48_{\pm 0.2}$
FLoR [7] (2024)	✓	✓	86.30	93.60	71.38	83.76	41.67	57.54	23.12	26.89
Ours	✓	✓	$\mathbf{87.00_{\pm 0.3}}$	$\mathbf{94.53_{\pm 0.4}}$	$\mathbf{71.87_{\pm 0.5}}$	$\mathbf{84.81_{\pm 0.5}}$	$\mathbf{42.43_{\pm 0.4}}$	$\mathbf{58.63_{\pm 0.3}}$	$\mathbf{23.94_{\pm 0.2}}$	$\mathbf{28.34_{\pm 0.2}}$

In 1-shot tasks, our method also achieves SOTA performance on the CUB, Cars, and Plantae datasets, with improvements of 0.45%, 0.79%, and 0.82% compared to the respective second-best method. Comparing the 1-shot and 5-shot experiments within the same target domain, we find that performance improvements in 1-shot tasks are relatively modest compared to 5-shot results. Referring to the ablation experiments (Table 2), we can attribute part of this performance decline to the prerequisites for the PAC module. Specifically, the prototype can only be corrected when there is more than one instance per class, so the module is inactive in 1-shot tasks. For experiments under TR settings, our method achieves SOTA results across all datasets in 5-shot tasks, with a 0.51% improvement in average accuracy compared to the second-best method, FLoR [7]. In 1-shot tasks, our method also performs well, though the accuracy improvement is less pronounced compared to the 5-shot setting.

Results for *Distant*-Domain Datasets. Compared to the aforementioned four natural *near*-domain datasets, extreme cross-domain datasets like ChestX, ISIC, EuroSAT, and CropDisease exhibit greater domain shifts, making generalization more challenging. As shown in Table 6, it is evident that our proposed model also achieves SOTA performance on extreme *distant*-domain datasets in both 1-shot and 5-shot settings. In 5-way tasks, the accuracy reaches 28.02%, 52.74%, and 92.36% for the ChestX, ISIC, and CropDisease datasets, respectively, showing improvements of 1.32%, 1.30%, and 1.11% compared to the second-best method. Consistent with the observations on *near*-domain datasets, the performance improvement in the 1-shot tasks on *distant*-domain datasets is also relatively modest. Combining this with the ablation experiments (Table 2), we can confirm that PAC effectively reduces the negative impact of atypical intra-class instances in multi-instance tasks. Under TR settings, our method achieves SOTA results across all datasets in both the 1-shot and 5-shot tasks, with an average accuracy improvement of 0.58% and 1.13%, respectively, compared to the suboptimal method, FLoR [7]. The extensive results across different target domains and experimental settings support our viewpoint that the well-designed cross-domain learning framework P-R^2-L is more suitable for CD-FSL tasks than existing complex techniques that explore domain alignment and feature transformation.

Qualitative Findings of Few-Shot Classification. From our experimental observations, we can draw four key conclusions:

- **(1)** Our model shows more significant performance improvements over other models on fine-grained tasks.
- **(2)** Compared to *near*-domain tasks, our model is better suited for *distant*-domain tasks with larger domain shifts.
- **(3)** Compared to single-instance tasks, our model shows more pronounced improvements in multi-instance scenarios.
- **(4)** Our model achieves optimal performance under both transductive and inductive settings.

4.5. Numerical results: Comparison with fine-tuning

As mentioned in [6,16], in the case of domain shifts, methods based on pretraining and fine-tuning are more effective than few-shot learning approaches like metric learning and meta learning. In this section, in addition to classical FSL methods and CD-FSL methods, we further include comparisons between advanced fine-tuning methods and our model's performance. For a fair comparison, we use the same initialized ResNet-10 as the backbone network for feature extraction. For the fine-tuning models, a fully connected layer is used as the classification head, while for our model, the P-R^2-L framework is built on the backbone. Following the new comparison protocol proposed in previous work [23], for a given target task T: n-way k-shot q-query, where $q = 15$ refers to the pseudo-samples generated for each class based on the support samples using the data augmentation method from [84]. For the fine-tuning models, during training, 15 pseudo-samples per class are generated from the support samples using data augmentation methods for model fine-tuning in the testing phase. As per convention [6], the learning rate of the SGD optimizer is set to 0.01, with momentum set to 0.9. For our method, we use the same support and query samples as the fine-tuning method, with the Adam optimizer initialized at a learning rate of 0.001. Both the fine-tuning methods and our approach are trained for 50 epochs under the 5-way 1-shot/5-shot setting to obtain the final model. Since the training data is consistent for both models, this ensures a fair comparison. As shown in Tables 7 and 8, under the 5-way setting, our model achieves improvements of 1.61%, 0.93%, 1.55%, and 0.86% over the second-best fine-tuning method on the *distant*-domain datasets ChestX, ISIC, EuroSAT, and CropDisease, respectively, as shown in Table 2. These improvements are significantly more pronounced compared to the gains observed on the *near*-domain

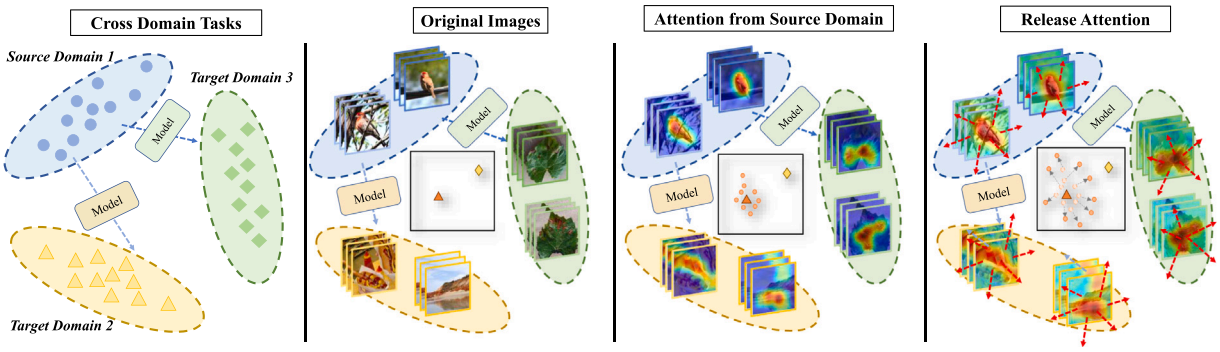


Fig. 11. Heatmap visualization before and after attention release. Source domain: CUB, target domains: *mini*-Imagenet and FPV. The figure displays the attention heatmap transformations for 2 source domain samples and 4 target domain samples (column 3 indicates before attention release and column 4 indicates after attention release). Under the influence of erroneous inductive bias from the source domain, the source domain samples accurately focus on discriminative regions, while images in the target domain all localize to the erroneous discriminative features. Red arrows indicate the direction of attention spread.

datasets {1.32%, 1.13%, 1.05%, and 0.36%}. Similarly, for transductive learning (TR) and single-instance (1-shot) scenarios, we can draw conclusions similar to those in Section 4.4. Therefore, we will not reiterate them here, but will briefly compare the experimental results.

4.6. Visualization of key results

To visually observe the performance changes brought by the proposed modules, in this section, we conduct a visual analysis of key processes and results. To more clearly showcase the experimental effects under tasks of varying granularity, we additionally introduce a finer-grained FPV disease dataset (which requires classification down to the severity of the disease). Its labeling format is "*plant-disease-severity*", which represents a finer granularity compared to datasets like CropDisease, where the labeling format is simply "*plant-disease*".

Visualization of Attention Release. We visualize the feature transformation diagrams for *near*-domain and *distant*-domain datasets. Box plots are used to illustrate the range of feature values on both *near*-domain and *distant*-domain datasets. In Fig. 13, the red area represents the feature values before transformation, while the blue area represents those after transformation. From the main range of the box plots and some of the outliers, it is evident that, compared to the feature values before transformation, the gap between the maximum and minimum values of the transformed features has narrowed, and the feature values are more concentrated in appropriate ranges. The corrected attention is redistributed relatively evenly from the focal areas to the entire image. This indicates that after attention release, the feature distribution is more uniform, allowing the model to focus more precisely on key features and avoid over-relying on specific extreme features. Additionally, after attention release, the model redistributes its focus from overly concentrated areas to various regions of the image, demonstrating a more comprehensive and detailed attention when processing the input. This is especially evident when handling *distant*-domain data, where important information is spread across more relevant areas, avoiding the previous over-concentration on certain specific domains.

Heatmap Visualization Before and After Attention Release. We also present the changes in heatmaps before and after attention release. Using CUB as the source domain, and *mini*-ImageNet and FPV as two target domains for testing, the second column in Fig. 11 shows typical example images from different datasets, while the third column illustrates incorrect attention heatmaps caused by the source domain's erroneous discriminative inductive bias. It can be observed that, without attention release, the attention in the source domain images could accurately focus, while other target domain images mistakenly concentrated on unimportant areas due to incorrect discriminative features, leading to incorrect feature selection. For instance, in some cases, the model wrongly focused on background noise or non-critical regions, likely due to the over-influence of source domain data features on the discriminative process for the target domain data. After applying BAR, we obtain the released attention heatmaps shown in the fourth column. The attention values across the entire image are more balanced, and the focus shifts from the incorrect regions, evenly dispersing across different areas.

Heatmap Visualization Before and After Attention Reaggregation. Similarly, using four target domain samples shown in Fig. 12 as examples, we visualize the heatmap transformation from attention release to cross-image attention reaggregation. During attention reaggregation, each support sample undergoes semantic alignment with a query sample to achieve fine-grained information localization across images. In each subplot, the second column shows the attention heatmap after attention release, and after cross-alignment, the final column displays the reaggregated attention map. It can be observed that our model demonstrates the correct heatmap distribution across different target domains, accurately extracting key features. Especially in some complex scenarios, the model can semantically align based on the characteristics of the target sample, accurately extracting important fine-grained features, rather than relying solely on global features. This process demonstrates the model's attentiveness to feature learning and its reinforcement of key information, while also diminishing irrelevant background features or noise regions, further enhancing the model's accuracy and reliability.

t-SNE Visualization. We perform t-SNE dimensionality reduction on the high-dimensional representations of source and target domain samples. As shown in Fig. 14, samples from different classes are represented by dots of different colors. An effective model should ensure that sample points from different classes remain well-separated, while those within the same class are closely clustered. We use *mini*-ImageNet as the source domain and test on three target domains with varying degrees of granularity: CIFAR, CUB, and FPV. As shown in Fig. 14, our model produces more reasonable and easily classifiable distributions across target domains with different granularity levels. Due to the strong fine-grained feature extraction capability of our model, it is able to maintain compact intra-class distributions and relatively noticeable inter-class distances, even on the extremely fine-grained FPV dataset.

Fine-Grained Difficult Case Study and Confusion Matrix. We conduct a difficult case analysis on the extremely fine-grained FPV dataset (with classifications down to disease severity). The confusion matrix comparing our method to the baseline model is shown in Fig. 15. Classes 18, 19, 20, 21, 22, and 23 represent 6 representative categories of fine-grained grape diseases. In the confusion matrix, the classification results for all these hard-to-classify fine-grained classes show significant improvement. Among them, classes 20 and 21, as the most challenging cases, are often difficult for agricultural experts to distinguish. Under our framework, the classification accuracy for this group of hard-to-classify classes improves significantly.

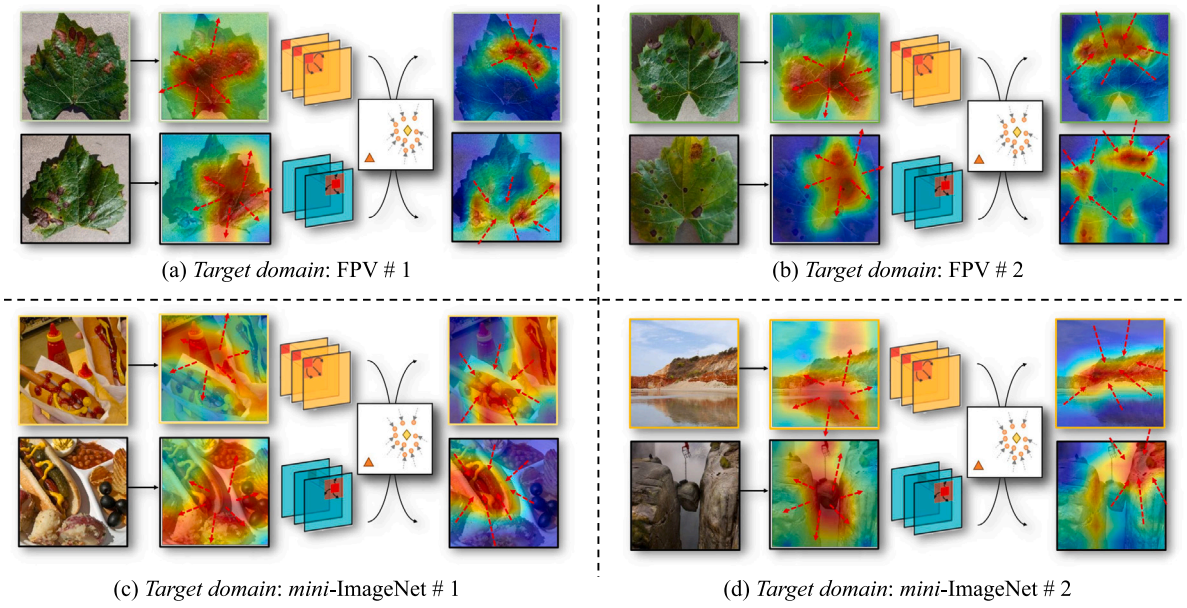


Fig. 12. Heatmap visualization before and after attention reaggregation. (a) (b) Visualization of the attention heatmap transformations on the target domain dataset FPV. (c) (d) Visualization of the attention heatmap transformations on the target domain dataset *mini*-Imagenet. In each subplot, the second column represents the heatmaps after attention release, and the last column shows the heatmaps after attention reaggregation. Red arrows indicate the direction of attention spread.

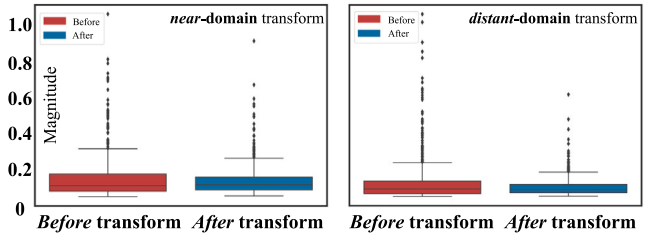


Fig. 13. Box plot illustrating the transformation of feature magnitudes before and after attention release. The left panel shows the transformation for the *near*-domain dataset, while the right panel displays the transformation for the *distant*-domain dataset. The red area represents the feature values before transformation, and the blue area represents the feature values after transformation.

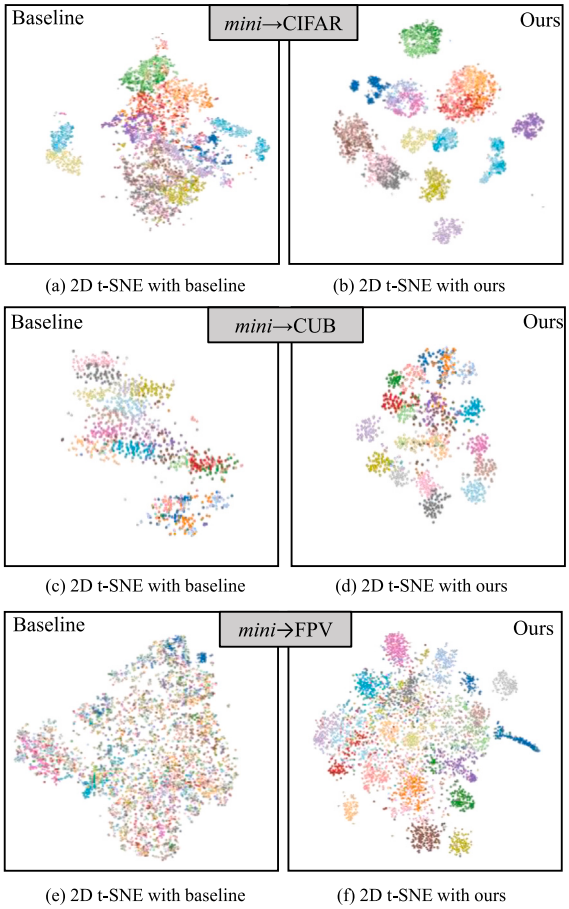
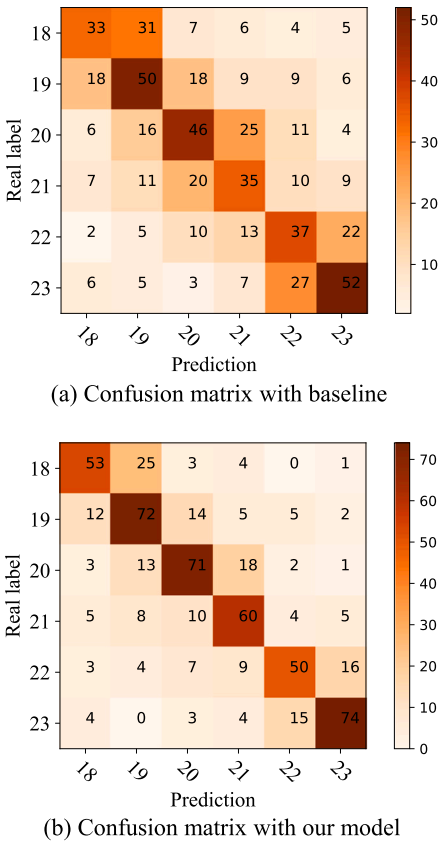


Fig. 14. t-SNE visualization at different granularity levels. The *mini*-Imagenet is used as the source domain, and the three datasets CIFAR, CUB, and FPV with different levels of granularity are used as the target domains, respectively. The left side shows the dimensionality reduction results from other method, while the right side displays the results from our method.

5. Conclusion

In this paper, we examine the issue of poor cross-domain generalization in few-shot learning models from the perspective of suppressing erroneous inductive biases from the source domain. We propose a three-level attention calibration framework, P-R^2-L, to address this issue. First, the PAC module evaluates the importance of instances within a class and reweights them to highlight key instances, thereby reducing the impact of noisy instances. Second, an R^2-FAC module is proposed, which sequentially integrates a BAR submodule and a NAR submodule. This setup suppresses erroneous discriminative inductive biases from the source domain and refocuses discriminative information in the target domain, significantly enhancing the cross-domain generalization ability of few-shot models. Finally, to mitigate the cross-domain model's overemphasis on discriminative information, we introduce an LAC module based on matrix *nuclear-norm* constraint, which effectively balances the discriminability and diversity of classification results. Extensive experiments on eight CD-FSL datasets with varying degrees of domain shift demonstrate the superiority of the proposed method. The visualization analysis of key processes on fine-grained datasets intuitively demonstrates the effectiveness of the proposed modules from an interpretability perspective. Nevertheless, the current method has limitations, especially when faced with extreme domain shifts, which may affect the model's performance. Additionally, the model's



(a) Confusion matrix with baseline

(b) Confusion matrix with our model

Fig. 15. Confusion matrices for the baseline and ours. The numbers 18, 19, 20, 21, 22, and 23 on the x-axis represent six representative fine-grained disease classes of grapes. Each column in the matrix indicates the predicted results, while each row represents the true labels.

computational efficiency and complexity are also challenges that need attention. Although our method improves accuracy, there is still a certain computational overhead. Future work could reduce computational complexity by optimizing the network structure and adopting more efficient training strategies. Furthermore, stronger cross-domain adaptation techniques, such as adversarial learning and unsupervised learning, could be explored to enhance adaptation to unseen domains, and integrating more backbone networks could strengthen the robustness of the method. We hope that these ideas and analyses will inspire researchers to better understand the essence of the CD-FSL problem.

CRediT authorship contribution statement

Minghui Li: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jing Jiang:** Visualization, Validation. **Hongxun Yao:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hongxun Yao reports financial support was provided by National Science and Technology. Hongxun Yao reports a relationship with National Major Science and Technology Projects of China that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Science Foundation of China under Grant 62476069.

Data availability

Data will be made available on request.

References

[1] J. Xie, F. Long, J. Lv, Q. Wang, P. Li, Joint distribution matters: Deep Brownian distance covariance for few-shot classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 7972–7981.
[2] J. Lai, S. Yang, W. Wu, T. Wu, G. Jiang, X. Wang, J. Liu, B.-B. Gao, W. Zhang, Y. Xie, et al., SpatialFormer: Semantic and target aware attentions for few-shot learning, 2023, arXiv preprint arXiv:2303.09281.
[3] B. Shi, W. Li, J. Huo, P. Zhu, L. Wang, Y. Gao, Global-and local-aware feature augmentation with semantic orthogonality for few-shot image classification, Pattern Recognit. 142 (2023) 109702.
[4] W. Wu, Y. Shao, C. Gao, J.-H. Xue, N. Sang, Query-centric distance modulator for few-shot classification, Pattern Recognit. (2024) 110380.
[5] H.-Y. Tseng, H.-Y. Lee, J.-B. Huang, M.-H. Yang, Cross-domain few-shot classification via learned feature-wise transformation, 2020, arXiv preprint arXiv:2001.08735.
[6] Y. Guo, N.C. Codella, L. Karlinsky, J.V. Codella, J.R. Smith, K. Saenko, T. Rosing, R. Feris, A broader study of cross-domain few-shot learning, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16, Springer, 2020, pp. 124–141.
[7] Y. Zou, Y. Liu, Y. Hu, Y. Li, R. 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, 2024, pp. 23575–23584.
[8] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.
[9] X. Wang, X. Wang, B. Jiang, B. Luo, Few-shot learning meets transformer: Unified query-support transformers for few-shot classification, IEEE Trans. Circuits Syst. Video Technol. (2023).
[10] F. Sung, Y. Yang, L. Zhang, T. Xiang, P.H. Torr, T.M. Hospedales, Learning to compare: Relation network for few-shot learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1199–1208.
[11] V. Garcia, J. Bruna, Few-shot learning with graph neural networks, 2017, arXiv preprint arXiv:1711.04043.
[12] Y. Liu, J. Lee, M. Park, S. Kim, E. Yang, S.J. Hwang, Y. Yang, Learning to propagate labels: Transductive propagation network for few-shot learning, 2018, arXiv preprint arXiv:1805.10002.
[13] Y. Chen, X. Wang, Z. Liu, H. Xu, T. Darrell, A new meta-baseline for few-shot learning, 2020.
[14] B. Liu, Y. Cao, Y. Lin, Q. Li, Z. Zhang, M. Long, H. Hu, Negative margin matters: Understanding margin in few-shot classification, in: European Conference on Computer Vision, Springer, 2020, pp. 438–455.
[15] F. Zhou, L. Zhang, W. Wei, Meta-generating deep attentive metric for few-shot classification, IEEE Trans. Circuits Syst. Video Technol. 32 (10) (2022) 6863–6873.
[16] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C.F. Wang, J.-B. Huang, A closer look at few-shot classification, 2019, arXiv preprint arXiv:1904.04232.
[17] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2009) 1345–1359.
[18] Y. Chen, W. Li, C. Sakaridis, D. Dai, L. Van Gool, Domain adaptive faster r-cnn for object detection in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3339–3348.
[19] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7167–7176.
[20] J. Xie, W. Ko, R.-X. Zhang, B. Yao, Physics-augmented deep learning with adversarial domain adaptation: Applications to STM image denoising, 2024, arXiv preprint arXiv:2409.05118.
[21] J. Chen, Z. Zhang, L. Li, B. Shahrasbi, A. Mishra, Contrastive adversarial training for unsupervised domain adaptation, 2024, arXiv preprint arXiv:2407.12782.
[22] Z. Hu, Y. Sun, Y. Yang, Switch to generalize: Domain-switch learning for cross-domain few-shot classification, in: International Conference on Learning Representations, 2022.
[23] H. Wang, Z.-H. Deng, Cross-domain few-shot classification via adversarial task augmentation, 2021, arXiv preprint arXiv:2104.14385.
[24] H. Liang, Q. Zhang, P. Dai, J. Lu, Boosting the generalization capability in cross-domain few-shot learning via noise-enhanced supervised autoencoder, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9424–9434.

[25] J. Zhang, J. Song, L. Gao, H. Shen, Free-lunch for cross-domain few-shot learning: Style-aware episodic training with robust contrastive learning, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 2586–2594.

[26] Y. Hu, A.J. Ma, Adversarial feature augmentation for cross-domain few-shot classification, in: European Conference on Computer Vision, Springer, 2022, pp. 20–37.

[27] F. Zhou, P. Wang, L. Zhang, W. Wei, Y. Zhang, Revisiting prototypical network for cross domain few-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 20061–20070.

[28] O. Vinyals, C. Blundell, T. Lillicrap, D. Wierstra, et al., Matching networks for one shot learning, Adv. Neural Inf. Process. Syst. 29 (2016).

[29] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, Adv. Neural Inf. Process. Syst. 30 (2017).

[30] C. Zhang, Y. Cai, G. Lin, C. Shen, Deepemd: Differentiable earth mover's distance for few-shot learning, IEEE Trans. Pattern Anal. Mach. Intell. 45 (5) (2022) 5632–5648.

[31] C. Liu, Y. Fu, C. Xu, S. Yang, J. Li, C. Wang, L. Zhang, Learning a few-shot embedding model with contrastive learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 8635–8643.

[32] S. Qiu, W. Yang, M. Yang, Hybrid feature collaborative reconstruction network for few-shot fine-grained image classification, 2024, arXiv preprint arXiv:2407.02123.

[33] B. Zhang, X. Li, Y. Ye, Z. Huang, L. Zhang, Prototype completion with primitive knowledge for few-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3754–3762.

[34] J. Xu, H. Le, Generating representative samples for few-shot classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 9003–9013.

[35] S. Yang, L. Liu, M. Xu, Free lunch for few-shot learning: Distribution calibration, 2021, arXiv preprint arXiv:2101.06395.

[36] H. Li, L. Li, Y. Huang, N. Li, Y. Zhang, An adaptive plug-and-play network for few-shot learning, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, IEEE, 2023, pp. 1–5.

[37] M. Hu, H. Chang, Z. Guo, B. Ma, S. Shan, X. Chen, Understanding few-shot learning: Measuring task relatedness and adaptation difficulty via attributes, Adv. Neural Inf. Process. Syst. 36 (2024).

[38] M. Dudík, S. Phillips, R.E. Schapire, Correcting sample selection bias in maximum entropy density estimation, Adv. Neural Inf. Process. Syst. 18 (2005).

[39] J. Yang, R. Yan, A.G. Hauptmann, Cross-domain video concept detection using adaptive svms, in: Proceedings of the 15th ACM International Conference on Multimedia, 2007, pp. 188–197.

[40] M. Wang, W. Deng, Deep visual domain adaptation: A survey, Neurocomputing 312 (2018) 135–153.

[41] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J.W. Vaughan, A theory of learning from different domains, Mach. Learn. 79 (2010) 151–175.

[42] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, T. Darrell, Deep domain confusion: Maximizing for domain invariance. arxiv 2014, 2019, arXiv preprint arXiv:1412.3474.

[43] M. Long, Y. Cao, J. Wang, M. Jordan, Learning transferable features with deep adaptation networks, in: International Conference on Machine Learning, PMLR, 2015, pp. 97–105.

[44] M. Long, H. Zhu, J. Wang, M.I. Jordan, Deep transfer learning with joint adaptation networks, in: International Conference on Machine Learning, PMLR, 2017, pp. 2208–2217.

[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).

[46] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, V. Lempitsky, Domain-adversarial training of neural networks, J. Mach. Learn. Res. 17 (59) (2016) 1–35.

[47] M. Long, Z. Cao, J. Wang, M.I. Jordan, Conditional adversarial domain adaptation, Adv. Neural Inf. Process. Syst. 31 (2018).

[48] A. Sharma, T. Kalluri, M. Chandraker, Instance level affinity-based transfer for unsupervised domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 5361–5371.

[49] K. Saito, K. Watanabe, Y. Ushiku, T. Harada, Maximum classifier discrepancy for unsupervised domain adaptation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3723–3732.

[50] S. Li, F. Lv, B. Xie, C.H. Liu, J. Liang, C. Qin, Bi-classifier determinacy maximization for unsupervised domain adaptation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 8455–8464.

[51] L. Chen, H. Chen, Z. Wei, X. Jin, X. Tan, Y. Jin, E. Chen, Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 7181–7190.

[52] K. Wang, Y. Shen, M. Lauer, Adversarial attacked teacher for unsupervised domain adaptive object detection, 2024, arXiv preprint arXiv:2408.09431.

[53] K. Saito, D. Kim, S. Sclaroff, K. Saenko, Universal domain adaptation through self supervision, Adv. Neural Inf. Process. Syst. 33 (2020) 16282–16292.

[54] N. Dvornik, C. Schmid, J. Mairal, Selecting relevant features from a multi-domain representation for few-shot classification, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16, Springer, 2020, pp. 769–786.

[55] L. Liu, W. Hamilton, G. Long, J. Jiang, H. Larochelle, A universal representation transformer layer for few-shot image classification, 2020, arXiv preprint arXiv:2006.11702.

[56] P. Bateni, R. Goyal, V. Masrani, F. Wood, L. Sigal, Improved few-shot visual classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14493–14502.

[57] P. Bateni, J. Barber, J.-W. Van de Meent, F. Wood, Enhancing few-shot image classification with unlabelled examples, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 2796–2805.

[58] S.X. Hu, D. Li, J. Stühmer, M. Kim, T.M. Hospedales, Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 9068–9077.

[59] C. Doersch, A. Gupta, A. Zisserman, Crosstransformers: spatially-aware few-shot transfer, Adv. Neural Inf. Process. Syst. 33 (2020) 21981–21993.

[60] T. Adler, J. Brandstetter, M. Widrich, A. Mayr, D. Kreil, M.K. Kopp, G. Klambauer, S. Hochreiter, Cross-domain few-shot learning by representation fusion, 2020.

[61] W.-H. Li, X. Liu, H. Bilen, Cross-domain few-shot learning with task-specific adapters, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 7161–7170.

[62] C.P. Phoo, B. Hariharan, Self-training for few-shot transfer across extreme task differences, 2020, arXiv preprint arXiv:2010.07734.

[63] X. Yue, Z. Zheng, S. Zhang, Y. Gao, T. Darrell, K. Keutzer, A.S. Vincentelli, Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13834–13844.

[64] H. Shao, X. Zhou, J. Lin, B. Liu, Few-shot cross-domain fault diagnosis of bearing driven by task-supervised ANIL, IEEE Internet Things J. (2024).

[65] Z. Ye, J. Wang, T. Sun, J. Zhang, W. Li, Cross-domain few-shot learning based on graph convolution contrast for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. (2024).

[66] Z. Lei, P. Zhang, Y. Chen, K. Feng, G. Wen, Z. Liu, R. Yan, X. Chen, C. Yang, Prior knowledge-embedded meta-transfer learning for few-shot fault diagnosis under variable operating conditions, Mech. Syst. Signal Process. 200 (2023) 110491.

[67] N. Paeedeh, M. Pratama, M.A. Ma'sum, W. Mayer, Z. Cao, R. Kowlczyk, Cross-domain few-shot learning via adaptive transformer networks, Knowl.-Based Syst. 288 (2024) 111458.

[68] Y. Grandvalet, Y. Bengio, Semi-supervised learning by entropy minimization, Adv. Neural Inf. Process. Syst. 17 (2004).

[69] J. Song, C. Shen, Y. Yang, Y. Liu, M. Song, Transductive unbiased embedding for zero-shot learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1024–1033.

[70] J. Zhuo, S. Wang, S. Cui, Q. Huang, Unsupervised open domain recognition by semantic discrepancy minimization, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 750–759.

[71] Y. Zou, Z. Yu, X. Liu, B. Kumar, J. Wang, Confidence regularized self-training, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 5982–5991.

[72] S. Cui, S. Wang, J. Zhuo, L. Li, Q. Huang, Q. Tian, Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3941–3950.

[73] M. Fazel, Matrix Rank Minimization with Applications (Ph.D. thesis), Ph.D. thesis, Stanford University, 2002.

[74] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (3) (2010) 471–501.

[75] N. Srebro, J. Rennie, T. Jaakkola, Maximum-margin matrix factorization, Adv. Neural Inf. Process. Syst. 17 (2004).

[76] C. Wah, S. Branson, P. Welinder, P. Perona, S. Belongie, The caltech-ucsd birds-200–2011 dataset, 2011.

[77] J. Krause, M. Stark, J. Deng, L. Fei-Fei, 3D object representations for fine-grained categorization, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 554–561.

[78] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: A 10 million image database for scene recognition, IEEE Trans. Pattern Anal. Mach. Intell. 40 (6) (2017) 1452–1464.

[79] G. Van Horn, O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, S. Belongie, The inaturalist species classification and detection dataset, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8769–8778.

[80] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. Summers, Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: IEEE CVPR, Vol. 7, sn, 2017.

[81] N. Codella, V. Rotemberg, P. Tschandl, M.E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, et al., Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic), 2019, arXiv preprint arXiv:1902.03368.
[82] P. Helber, B. Bischke, A. Dengel, D. Borth, Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 12 (7) (2019) 2217–2226.
[83] S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection, Front. Plant Sci. 7 (2016) 1419.
[84] P. Li, S. Gong, C. Wang, Y. Fu, Ranking distance calibration for cross-domain few-shot learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 9099–9108.

Minghui Li received the B.S. degree from Civil Aviation University of China in 2018 and the M.S. degree from Sun Yat-sen University in 2021. He is currently pursuing the Ph.D. degree with the Faculty of Computing, Harbin Institute of Technology, Harbin, China. His research interests include computer vision and deep learning, especially focusing on few-shot learning, transfer learning, and domain adaptation.

Jing Jiang is currently a master's candidate in the Faculty of Computing at Harbin Institute of Technology, China. He received a BS degree from Harbin Institute of Technology in 2023. His research interests focus on transfer learning, multimodal learning, panoramic image understanding, and semantic segmentation.

Hongxun Yao received the B.S. and M.S. degrees in computer science from Harbin Shipbuilding Engineering Institute, Harbin, China, in 1987 and 1990, respectively, and the Ph.D. degree in computer science from Harbin Institute of Technology, in 2003. Currently, she is a professor at the Faculty of Harbin Institute of Technology, a recipient of the Ministry of Education Fund for "the New Century Excellent Talent" in China in 2005, and won the honor title of "enjoy special government allowances expert" in Heilongjiang Province, China. Prof. Yao has mainly researched computer vision intelligence, multimedia data analysis and understanding, affective computing, etc. She has been the executive director of China Society of Image and Graphics, and the director of the Affective Computing and Understanding Special Committee in CSIG. She has published more than 200 papers and achieved one Best Paper Award from ICIMCS 2016. She also has won 1 First Prize and 2 Second Prizes of the Provincial Natural Science and Technology Award. She has published 6 books and owns 16 national invention patents.