计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.009873] [CSTR: 32024.14.csa.009873]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于多元嵌入增强网络的少样本图像分类算法①

徐　震

(南京信息工程大学 软件学院, 南京 210044)
通信作者: 徐　震, E-mail: 202212210043@nuist.edu.cn

摘　要: 少样本图像分类旨在从有限的标注数据中学习分类器. 尽管现有方法已取得显著进展, 但由于训练样本有限、类内差异过大、类间差异过小, 支持样本与查询样本容易发生混淆, 导致现有方法在提取有用特征和准确区分图像类别方面仍面临挑战. 为了解决这些问题, 我们设计了一种新的多元嵌入增强网络. 该网络轻量且高效, 通过生成一组特征嵌入来表示图像, 而非仅依赖单一的图像级特征. 它能够生成多种层析结构, 从而学习更丰富的特征表示, 减小类内差异并扩大类间差异. 此外, 我们提出了一种基于集合的度量方法, 并结合动态自适应加权机制, 用于衡量查询集和支持集之间的相似度. 实验结果表明, 在 miniImageNet、tieredImageNet 和 CUB 数据集上, 模型表现优异. 在使用 ResNet-12 网络的 1-shot 设置下, 准确率分别达到了 72.22%、75.43% 和 85.02%, 相较于基准模型分别提升了 1.09%、2.93% 和 1.47%.

关键词: 图像分类; 小样本图像分类; 特征集合; 自适应加权

引用格式: 徐震.基于多元嵌入增强网络的少样本图像分类算法.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9873.html

Few-shot Image Classification Algorithm Based on Multi-embedding Enhanced Network

XU Zhen

(School of Software, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: Few-shot image classification aims to learn a classifier from a limited amount of labeled data. Despite significant progress made by existing methods, challenges remain in extracting useful features and accurately classifying images due to the limited number of training samples, large intra-class variance, and small inter-class variance, which lead to confusion between support and query samples. To address these issues, this study proposes a novel multi-embedding enhanced network. This lightweight and efficient network represents images by generating a set of feature embeddings, rather than relying solely on single-image-level features. It is capable of generating various hierarchical structures to learn richer feature representations, thereby reducing intra-class variance and increasing inter-class variance. In addition, the study proposes a set-based metric combined with a dynamic self-adaptive weighting mechanism to measure the similarity between query and support sets. Experimental results demonstrate the excellent performance of the proposed model on the miniImageNet, tieredImageNet, and CUB datasets. Using a 1-shot setting in the ResNet-12 network, the model achieves accuracies of 72.22%, 75.43%, and 85.02%, respectively, outperforming the baseline models by 1.09%, 2.93%, and 1.47%.
Key words: image classification; few-shot image classification; feature set; adaptive weighting

在计算机视觉领域, 深度卷积神经网络[1–3]已广泛应用于图像分类、目标检测和语义分割等任务, 并取得了卓越的性能. 然而, 大多数图像分类任务[4–6]通常需要依赖庞大的标注数据集进行模型训练, 其在小规

① 收稿时间: 2024-11-28; 修改时间: 2024-12-17; 采用时间: 2025-01-07; csa 在线出版时间: 2025-03-24

模数据集上的学习能力仍有待提升. 因此, 当可用的标注数据极其有限时, 这些方法往往难以取得理想的效果.

相比之下, 从少量示例中学习是一项对人类至关重要的技能. 例如, 儿童在学习识别图像时, 仅需一个或几个示例便能掌握对某类物体的辨识能力. 少样本图像分类方法[7–9]正是受到这一现象的启发, 其核心任务是从极少量的示例中学习知识, 并将其有效迁移到新的类别中.

近年来, 基于度量学习的少样本学习方法[10,11]已成为主流. 这些方法通过学习样本之间的度量, 在嵌入空间中计算不同样本之间的相似度度量. 然而, 我们观察到, 同一类别的物体在支持集和查询集之间可能表现出显著的图像级外观差异, 而不同类别的物体之间可能具有相似的外观. 例如, 同一类别的狗展现出完全不同的外观, 并且背景也不同, 而不同类别 (狗和狼) 之间的外观则相似, 背景也相似.

在这项工作中, 我们提出了一种有效且鲁棒的方法来解决这些问题. 仅使用图像级表示使得准确区分支持集和查询样本变得具有挑战性. 因此, 本文的核心思想是以更有意义的方式表示每张图像, 而不仅使用一个图像级特征. 为此, 提出使用一组特征嵌入, 使得网络能够从图像的不同视角聚合更丰富、更有用的特征. 然而, 这也引出了一个新问题: 如何衡量两组图像嵌入 (查询集和支持集) 之间的相似度或不相似度? 为此, 提出了一种基于加权集合度量的方法来解决这个问题. 此外, 为了抑制负面特征并促进对最终分类结果有重大影响的正面特征, 我们引入了一种无监督方法, 根据基于集合度量的特征生成一个动态自适应的权重. 实验结果表明, 所提出的方法能够从查询图像和支持图像中探索更多有用的信息, 从而有助于提高分类结果.

1 相关工作

1.1 少样本图像分类

近年来, 少样本学习逐渐成为研究热点. 当前, 该领域的研究主要分为 3 类: 基于数据增强方法、基于梯度的方法和基于度量的方法. 1) 数据增强方法主要用于直接解决分类数据不足的问题. 它既可以通过迁移学习利用外部数据引入额外样本, 也可以在内部对标记图像或其特征表示应用各种变换, 或者使用未标记数据来实现小样本图像分类. Bateni 等人[12]提出了一种转换元学习方法, 该方法利用未标记实例提升少量图像分类的性能. 该方法结合了基于马氏距离的正则化软 k 均值聚类过程与改进的神经自适应特征提取器, 通过未标记数据提升测试时的分类精度. 在换向少样本学习任务中, 该方法通过给定一组支持样本共同预测查询样本的标签. 2) 基于梯度的方法[13]通过梯度下降算法, 使模型能够更快速、更准确地适应新任务和新数据, 从而增强模型的泛化能力和性能. 其核心目标是从各种少样本任务中提取通用知识, 促进对新任务的快速适应. 具体而言 Rusu 等人[14]提出了隐嵌入优化 (LEO) 方法, 该方法通过学习模型参数的数据依赖潜在生成表示, 并在低维潜在空间中执行基于梯度的元学习, 有效解决了在极低数据域的高维参数空间中操作的实际困难. LEO 将基于梯度的自适应过程与底层模型参数的高维空间解耦, 提升了少量学习和快速适应的能力. 3) 基于度量的方法[10,11]则主要关注在特征空间内精确衡量样本间的相似性或距离, 通过最小化类内距离并最大化类间距离来提升分类效果.

传统方法通常由 2 个部分组成: 特征提取器和分类器. 在原型网络[10]中, 每个类别由一个原型向量表示, 该向量是该类别内所有样本特征向量的平均值. 特征提取器用于为每个样本生成单一的特征向量. 在推理阶段, 原型网络[10]通过计算测试样本与各类别原型之间的欧几里得距离或余弦相似度, 判断其距离最近或相似度最高的类别原型, 并以此作为分类结果.

我们的研究方法与这一流程密切相关. 然而, 与传统方法不同, 我们并不为每张图像提取单一特征向量, 而是生成一个特征集合. 同时, 在度量阶段, 我们提出了一种基于加权集合度量的方法, 此外, 为了抑制负面特征并促进对最终分类结果有重大影响的正面特征, 我们引入了一种无监督方法, 根据基于集合度量的特征生成一个动态自适应的权重.

1.2 注意力机制

注意力机制在多个领域得到了广泛应用, 其灵活性和有效性使其成为众多深度学习模型的重要组成部分. 我们将其转化为一个嵌入在原始特征提取器中的多元嵌入扩展模块. 与现有工作相比, 所提的多元嵌入扩展模块既独立又轻量, 不会显著增加网络的总参数量. 本文将这一模块嵌入到特征提取器的每一层中. 此外, 为了增强模型从低层到高层整合语义信息的能力、提升对输入数据语义结构的理解, 并减少信息丢失的风险, 采用了一种自适应机制, 在将特征图输入多元嵌入扩展模块之前, 先对来自不同层的信息进行聚合.

1.3 特征集合度量

特征集合在计算机视觉领域备受关注, 其重要性已被众多研究所证实. Huang 等人[15]通过引入多尺度特征提取实现语义分割. 然而, 我们的特征集应用旨在解决小样本分类问题. 在少样本学习中, 我们的方法仅聚焦于特征集. Li 等人[16]通过基于局部描述符的图像到类度量替代了传统的图像级特征度量. 相比之下, 本文更注重为每个单独样本提取特征集. 此外, 我们的方法与 DeepEMD[17]也有相似之处, 它使用地球移动距离进行跨多个裁剪图像表示的通用数据增强. 与此不同, 本文专注于为每个单独样本提取特征集.

关于基于集合的度量, 本文工作与原型网络[10]相关, 该方法使用余弦相似度计算两个特征集之间的距离. 在基于集合的度量中, 我们的实验揭示了从单张图像中提取的多个特征向量对分类结果的贡献存在差异. Zhu 等人[18]探讨了直接和无监督方法在小样本学习中的潜在应用, 并使用无监督度量学习获取有助于小样本分类任务的判别子空间. 本文中, 采用无监督方法评估样本之间的相似度, 以生成集合的权重.

2 网络架构

2.1 问题定义

我们采用当前流行的少样本分类方法, 将问题转化为 N-way K-shot 场景. 一个从未知任务分布 $P(T)$ 中抽样的 N-way K-shot 任务 T 由两个集合组成: 一个是训练集 (支持集) $S=\{(x_i,y_i)\}_{i=0}^{N\times K}$ 包含少量标注样本; 另一个是测试集 (查询集) $Q=\{(x_i,y_i)\}_{i=0}^{J}$ 包含一些未标注样本. 其中, x_i 表示从新颖的类 C_{novel} 中采样的图像, $y_i\in C_{\text{novel}}$ 是 x_i 的标签, N 表示 S 中的类数, K 表示 S 中每个类的图像数, J 表示 Q 中图像的数量. 我们的目标是学习一个好的分类器, 通过支持集 S 来对查询集 Q 进行分类.

2.2 整体框架

本文提的整体框如图 1 所示.

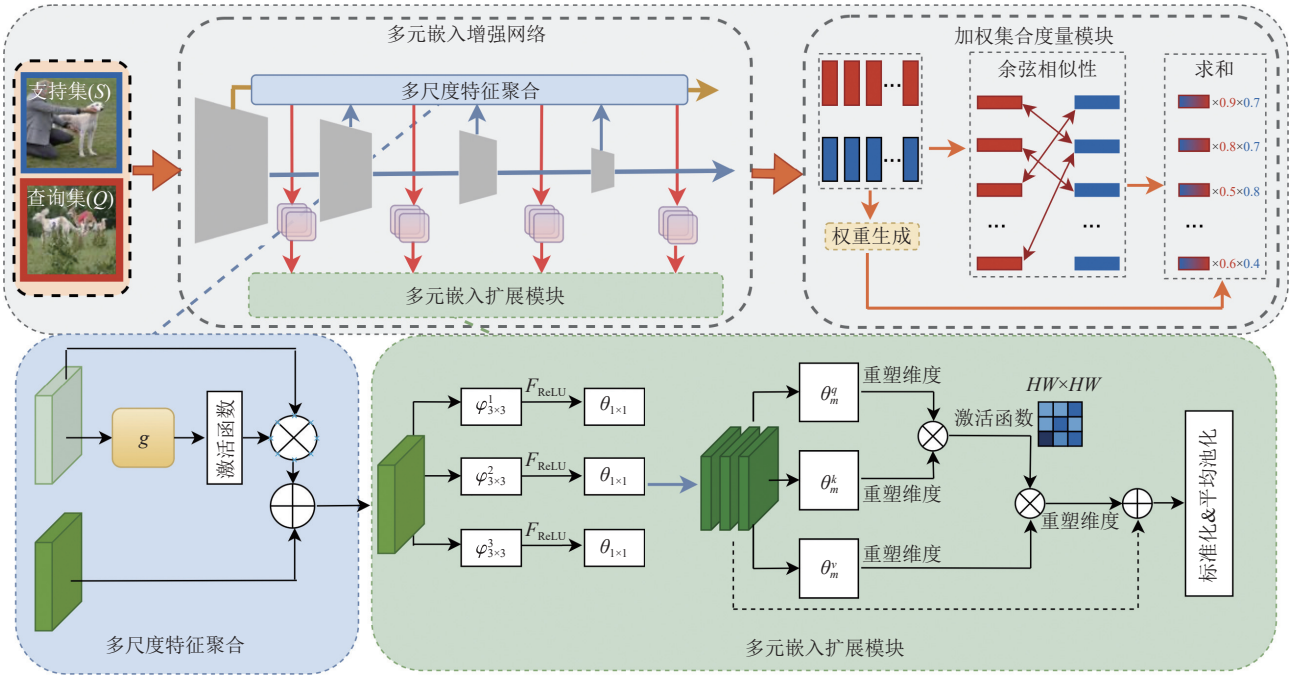


图 1 本文模型框架图

该方法由两个主要部分组成: 多元嵌入增强网络和加权集合度量模块. 多元嵌入增强网络是一个轻量级模型, 包含两个核心模块: 多尺度特征聚合和多元嵌入扩展. 其中, 多尺度特征聚合通过整合不同尺度的特征, 使模型获得更丰富的特征表达. 多元嵌入扩展模块通过自注意力机制从图像中提取特征集, 而非单一特征向量, 从而增强特征表达能力. 随后, 特征集合匹配模块使用无监督权重学习的方法, 通过综合考量特征之间的相似性和相异性, 为特征生成权重.

3 网络实现细节与算法

3.1 多元嵌入增强网络

多元嵌入增强网络如图 1 所示, 其目标是将传统

的单个特征向量提取策略转变为特征集合的提取. 该网络由两部分组成: 第 1 部分是多尺度特征聚合模块, 通过自适应方式动态融合不同尺度的特征; 第 2 部分是多元嵌入扩展模块, 主要用于生成更多的嵌入特征.

3.1.1 多尺度特征聚合模块

为了让多元嵌入增强网络捕获足够丰富且多样化的特征表示, 我们集成了一个高效的空间多阶段特征聚合块 (如图 1 所示). 具体而言, 除了第 1 个卷积块外, 我们聚合了两种类型的特征映射: 当前块之前的低级特征映射 $f_l \in \mathbb{R}^{c_l \times h_l \times w_l}$ 和当前块的高级特征映射 $f_h \in \mathbb{R}^{c_h \times h_h \times w_h}$, 其中 c、h 和 w 分别表示通道数、高度和宽度.

首先, 我们通过平均池化调整低级特征的宽度和高度, 使其与高级特征对齐. 接着, 为了自适应地将低级特征聚合到高级特征中, 我们利用卷积层学习 f_l 的权重 w_l. 随后, 通过 f_l 和 w_l 的矩阵乘法以及激活函数计算低级特征的聚合结果. 最后, 将低级特征 f_l 与高级特征 f_h 通过矩阵相加完成融合:

$$f_h = \sigma(w_l(avg(f_l))) \cdot f_l + f_h \tag{1}$$

3.1.2 多元嵌入扩展模块

为了生成更多的嵌入, 我们设计了多元嵌入扩展模块, 通过使用特征集合代替单一的特征向量来表示图像. 该模块被嵌入到骨干网络的每一层, 从而生成更多的嵌入 (如图 1 所示).

具体而言, 对于每一层嵌入的多元嵌入扩展模块, 我们首先使用 3 个不同空洞率 (1, 2, 3) 的 3×3 空洞卷积层 $\varphi^1_{3\times3}$, $\varphi^2_{3\times3}$, $\varphi^3_{3\times3}$ 获取更多尺度的特征. 然后通过 ReLU 激活层 F_{ReLU} 来提高它的非线性表示能力. 接着, 再使用一个 1×1 卷积层 $\theta_{1\times1}$ 对获得的特征图进行处理, 让其维度保持一致:

$$Z_{b_m} = \theta_{1\times1}\left(concat\left(\varphi^1_{3\times3}(f_h), \varphi^2_{3\times3}(f_h), \varphi^3_{3\times3}(f_h)\right)\right) \tag{2}$$

其中, Z_{b_m} 是所有生成的嵌入被按第 1 个维度拼接, 得到的下一阶段的输入. 其中, b 表示骨干网络中的卷积块, m 表示我们特征集合中特征的个数.

接下来, 使用自注意力机制将 Z_{b_m} 转换为最终的输出 h_m. 具体来说, 我们将 Z_{b_m} 输入到一个卷积层中, 得到 3 个参数化元素: $q(Z_{b_m}|\theta^q_m)$、$k(Z_{b_m}|\theta^k_m)$ 和 $v(Z_{b_m}|\theta^v_m)$. 然后, 首先使用两个参数化元素 $q(Z_{b_m}|\theta^q_m)$ 和 $k(Z_{b_m}|\theta^k_m)$ 计算注意图 a_m.

$$a_m = Softmax\left(q(Z_{b_m}|\theta^q_m) \cdot k(Z_{b_m}|\theta^k_m)/\sqrt{d_k}\right) \tag{3}$$

其中, $a_m \in \mathbb{R}^{p\times c\times h\times w}$ 是 Z_{b_m} 的注意力得分, $\sqrt{d_k}$ 是缩放因子. 然后我们让 a_m 和 $v(Z_{b_m}|\theta^v_m)$ 做乘法运算并和 Z_{b_m} 做加法运算得到最终的输出 h_m:

$$h_m = Z_{b_m} + a_m \cdot v(Z_{b_m}|\theta^v_m) \tag{4}$$

其中, $h_m \in \mathbb{R}^{p\times c\times h\times w}$ 是我们通过多元嵌入扩展模块生成的一组嵌入.

3.2 加权集合度量模块

为了抑制负面特征并增强对最终分类结果具有重要影响的正面特征, 从而更有效地利用集合度量进行图像分类, 我们提议为集合中的每个特征分配权重. 如图 2 所示, 我们通过综合考虑特征之间的相似性和差异性, 动态地为集合中的特征分配权重.

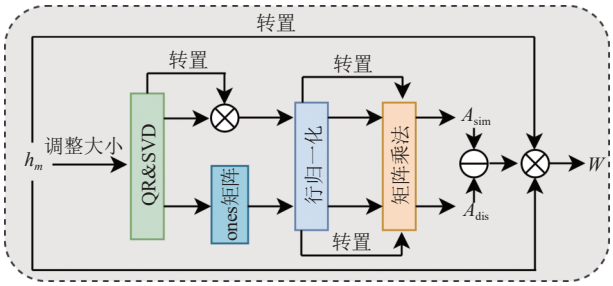


图 2 无监督权重生成

3.2.1 不相似矩阵

在以往的研究中, 研究者们倾向于基于相似度矩阵进行特征加权. 然而, 我们在权重学习过程中融合了相似度矩阵和不相似度矩阵的影响. 具体来说, 在一个 N-way K-shot 样本任务中, 每个类别有 B 个查询样本, 在我们的基于集合的度量中, 为了评估每个特征向量 (总共 M 个) 在每个样本特征集中的贡献, 我们将问题转化为一个 $(K+B+M)N$-way 1-shot 样本问题, 其中对角线代表相同的实体. 最终, 我们得到不相似度矩阵:

$$A_{dis} = (1/K+B+M)ee^{\mathrm{T}} - I \tag{5}$$

其中, e 是一个维度为 $(K+B+M)$ 的向量, 元素全部为 1, I 表示单位矩阵.

3.2.2 相似矩阵

在低秩表示中, 每个数据点 x_i 被描述为其他点的线性组合, 即 $x_i = z_{ij}x_j$, 使用表示系数 $\left(|z_{ij}| + |z_{ji}|\right)/2$ 来衡量 x_i 和 x_j 之间的相似性. 我们的目标是利用数据中的内在相关结构来评估样本对之间的相似度矩阵. 因此, 我们将低秩表示应用于以下最小秩优化问题:

$$\arg\min_{Z}\left\|Z-h_m h_m^{\mathrm{T}}Z\right\|_F^2,\quad \text{s.t. } \operatorname{rank}(Z)=K \tag{6}$$

式 (6) 通过两个步骤求解: 1) $Z=VV^{\mathrm{T}}$, 其中 V 是从特征 h_m 的奇异值分解中得到的; 2) 对于 V, 保留 K 个绝对值最大的元素. 我们的相似度矩阵可以表示为$A_{\text{sim}}=|Z|-\operatorname{diag}(|Z|)$.

通过上述步骤, 我们得到了相似矩阵 A_{sim} 和不相似矩阵 A_{dis}. 根据特征之间的相似性和差异性, 为每个特征嵌入生成了动态自适应权重:

$$W=h_m(\alpha A_{\text{dis}}-A_{\text{sim}})h_m^{\mathrm{T}} \tag{7}$$

其中, h_m 表示通过多元嵌入增强网络获取的特征集, A_{dis} 和 A_{sim} 是两个具有对立效果的度量: 前者是一个不相似度矩阵, 后者是一个相似度矩阵. 参数α用于平衡这两者的影响.

接下来我们先解释如何使用先前提取的特征集合进行图像分类. 遵循原型网络[10]和 SetFeat[19]的原则, 我们将查询特征集与每个类别支持集中的对应特征集进行比较, 以推断查询集的类别. 具体而言, 我们采用基于集合的度量, 记作$d_{\text{set}}=\left(X_q,S_n\right)$, 用来量化集合之间的距离. 在这里, X_q 表示查询集, S_n 表示支持集的原型. 在本文中, 我们使用 $h_m(X_q)$ 来表示由多元嵌入增强网络提取的特征嵌入集, 使用$h_m(s)=1/|s|=\sum_{x\in S}h_m\left(X_q\right)$来表示从查询集中提取的原型. 在这种情况下, 我们使用负余弦相似度函数来计算查询集中的特征嵌入与支持集质心之间最小距离的总和:

$$d_{\text{set}}=\sum^{M}\min d\left(h_i\left(X_q\right),\left(\overline{h_j}s\right)_n\right) \tag{8}$$

其中, d_{set} 计算了每个集合内特征向量之间的距离和求和.

通过式 (7) 获得权重之后, 我们的加权基于集合的度量可以表示为如下:

$$d_{w\_set}=\sum\min d\left(W\cdot h_i\left(x_q\right),\left(W\cdot\bar{h}_j s\right)_n\right) \tag{9}$$

4 实验分析

4.1 实验环境和参数

为了与先前的工作进行实验比较, 本文方法采用了 ResNet-12 作为主干网络, 并严格按照原型网络[10]的规范进行了实现. 训练的过程中采用随机梯度 (SGD) 下降优化器, 学习率为 0.001, 损失函数为 MSELoss, 迭代次数为 600. 本文方法使用了 MoCo 中的数据增强, 这些数据增强包括随机裁剪、随机颜色抖动、随机水平翻转、随机灰度转换和随机模糊增强. 最后, 在测试阶段使用 5 个随机种子进行测试, 以消除实验误差. 本文模型采用 PyTorch 框架, 并在一台 NVIDIA RTX 3090 24 GB 服务器上运行所有实验.

4.2 数据集

我们的实验是在几个广泛使用的少样本学习基准数据集上进行的, 包括 miniImageNet、tieredImageNet 和 CUB. miniImageNet 数据集包含 100 个类别; 根据原型网络[10]中的配置, 我们使用了 64 个类别用于训练, 16 个类别用于验证, 剩余的 20 个类别用于评估模型性能. tieredImageNet 分类遵循一个分层结构, 按语义区分基础类和新类. 我们采用了 SetFeat[19]中提出的划分方法, 其中包含 20 个超类作为基础集, 6 个超类作为验证集, 8 个超类作为新类集. CUB 是一个精细粒度分类数据集, 涵盖了 200 种鸟类物种. 根据 SetFeat[19]中的设置, 100 个类别用于基础类, 50 个用于评估, 剩余的 50 个用于新类. CIFAR-FS 数据集是一个最近提出的小样本图像分类基准数据集, 源自 CIFAR. 它包含所有 100 个类别, 并进一步随机划分为 64 个训练类别, 16 个验证类别和 20 个测试类别. 每个类别包含 600 张 32×32 大小的图像. FC100 数据集是另一个基于 CIFAR 的小样本分类数据集, 其主要思想与 tieredImageNet 相似, 100 个类别被分为 20 个超类, 每个超类由 5 个标准类别组成. 这些超类分别被划分为 12 个、4 个、4 个用于训练、验证和测试.

4.3 实验细节

考虑到我们的多元嵌入增强网络不可避免地增加了参数数量, 为了确保性能提升不仅仅是由于参数的增加, 我们相应地减少了骨干特征提取器中卷积核的数量. 具体来说, 在 Conv4-512 骨干中, 我们的方法采用了一组卷积核, 将其数量减少到 96/128/160/200, 总计 162.3 万个参数, 而 Conv4-512 方法则有 159.1 万个参数. 同样地, 在 ResNet-12 中, 我们的方法包括 128/150/180/200 个卷积核, 总计 1295.4 万个参数, 而 ResNet-12 则有 1 242.4 万个参数. 关于 Conv4-64[20], 由于其已经具有最小的参数量, 进一步减少参数会导致训练不稳定. 因此, 我们的方法虽然具有更多的参数, 所以我们在 Conv4-64 中人工增加了参数, 结果表明我们的方法仍然显著优于它. 从图 3 中可以看出, 我们的方法在这两种骨干网络上都达到了最新的性能. 我们的多元嵌入网络在其自注意力模块中使用卷积注意力. 我们优化其查询、键和值, 采用单深度卷积和批量归一化. 特

征集的输出维度配置为与特征提取器最后一层的通道数匹配, 以便无缝连接.

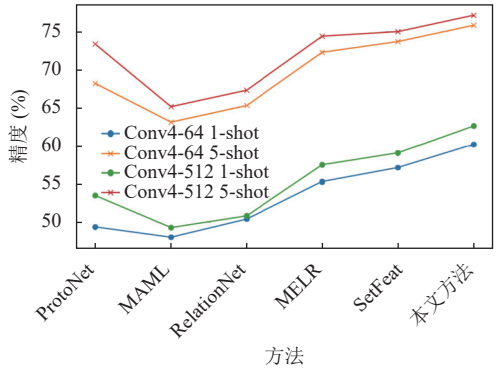


图 3 在 miniImageNet 上评估 Conv4-64 和 Conv4-512 骨干网络的性能

Conv4-64 使用 Adam 优化器进行预训练 (学习率为 0.001, 权重衰减为 5×10^{-4}), 批量大小固定为 64. 相比之下, ResNet-12 使用 Nesterov 动量优化器 (初始学习率为 0.1, 动量为 0.9, 权重衰减为 5×10^{-4}). 数据归一化和增强按照 Xu 等人[21]的指南进行. 在整个元训练阶段, 所有架构均使用 SGD 优化器, 验证集有助于优化 SGD 优化器的调度.

4.4 对比实验

表 1 展示了网络结构为 ResNet-12 时, 在 miniImageNet 和 tieredImageNet 上的 5-way 任务的精度, 最好的结果使用加粗表示, 次好的结果使用下划线表示.

表 1 在 miniImageNet 和 tieredImageNet 数据集上, 本文方法与其他方法的精度比较 (%)

方法	miniImageNet		tieredImageNet	
	1-shot	5-shot	1-shot	5-shot
DMF (CVPR 2021) [21]	67.76	82.71	71.89	85.96
RENet (ICCV 2021) [22]	67.60	82.58	71.61	85.28
MixtFSL (ICCV 2021) [23]	63.98	82.04	70.97	86.16
DeepBDC (CVPR 2022) [24]	67.34	84.46	72.34	87.31
APP2S (AAAI 2022) [20]	66.25	83.42	72.00	86.23
SetFeat (CVPR 2022) [19]	68.32	82.71	73.63	87.59
TALDS-NET[25]	67.89	84.31	71.34	86.12
FGFL (ICCV 2023) [26]	69.14	86.01	73.21	87.21
CORL (WACV 2023) [27]	65.74	83.03	73.84	86.76
HELA (WACV 2024) [28]	68.20	<u>86.70</u>	<u>72.50</u>	<u>87.60</u>
DCPNet (TNNLS 2024) [29]	<u>71.13</u>	86.44	—	—
本文方法	**72.22**	**88.02**	**75.43**	**89.53**

在 miniImageNet 上, 我们的方法明显优于其他最新技术, 特别是在 1-shot 和 5-shot 任务中, 我们的准确率分别比 DCPNet 高出 1.09% 和 1.58%. 在 tieredImageNet 上, 与 ELMOS[30]相比, 我们的方法在 1-shot 任务上提高了 1.59%, 在 5-shot 任务上提高了 1.55%. 表 2 则展示了我们提出的多元嵌入增强网络在细粒度分类数据集上的评估结果, 可以看出, 我们的方法再次超越了所有现有的先进技术.

表 2 在 CUB 数据集上与其他方法的准确度比较 (%)

方法	网络结构	1-shot	5-shot
MixtFSL (ICCV 2021) [23]	ResNet-18	73.94	86.01
Neg-Margin (ECCV 2020) [31]	ResNet-18	72.66	89.40
RENet (ICCV 2021) [22]	ResNet-12	79.49	91.11
SetFeat (CVPR 2022) [19]	ResNet-12	79.60	90.48
BAVARDAGE (PMLR 2023) [30]	ResNet-12	82.00	90.70
APP2S (AAAI 2022) [20]	ResNet-12	77.64	90.43
DeepBDC (CVPR 2022) [24]	ResNet-18	83.55	<u>93.82</u>
LRD (ICLR 2021) [32]	ResNet-12	79.56	90.67
FRN (CVPR 2021) [33]	ResNet-12	<u>83.55</u>	92.92
本文方法	ResNet-12	**85.02**	**94.23**

为了评估其在 Conv4-64 和 Conv4-512 骨干网络上的有效性, 我们在 miniImageNet 数据集上进行了测试如图 3 中的折线图所示, 我们的方法在这两个骨干网络上都达到了当前的最先进性能. 从图 4 可以看出, 我们的方法在 FC100 数据集上也表现出优越性能.

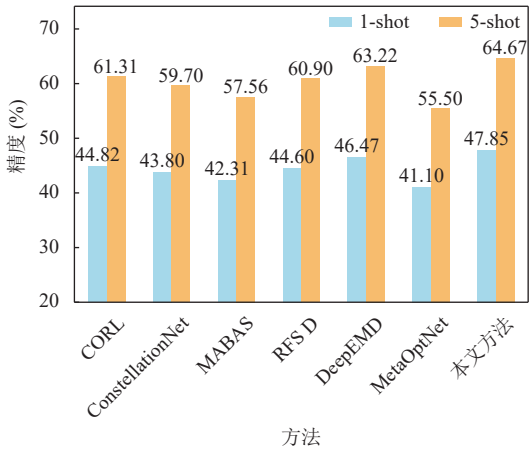


图 4 在 FC100 上与最先进的方法比较

4.5 消融实验

为了进一步分析本文方法, 我们致力于探索决策机制的替代设计, 并更好地理解多元嵌入网络为何能够实现更优的性能和准确度. 为此, 在此进行了广泛的消融研究. 具体而言, 我们不仅分析了各个模块缺失时模型的表现, 还考察了用其他替代模块替换这些模块的效果. 通过这些对比研究, 我们旨在揭示每个模块在整体模型中的重要性和作用. 此外, 还比较了 5-shot 1-way

和 5-shot 5-way 设置的结果, 以评估不同配置对模型性能的影响. 这些研究将帮助我们进一步优化模型设计, 确保其在实际应用中的出色表现.

4.5.1 增强版 Conv4-64

在前文提到通过减少骨干网络中卷积核的数量, 以确保在骨干网络中加入我们的双重注意力映射器不会显著增加网络参数的总数. 然而, 在 Conv4-64 上执行这一操作时, 由于每个骨干网络块只有一个包含 64 个卷积核的层, 这导致了较差的泛化性能. 因此, 我们通过在卷积块后引入 3 个全连接层 (维度分别为 512、160 和 64) 来改变 Conv4-64 的结构. 这一调整使参数数量增加到 0.239M, 与本文方法的 0.238M 参数相符. 结果如表 3 所示. 尽管增强版 Conv4-64 在性能上相较于基准 Conv4-64 有所改进, 但其改进程度明显低于所提方法的表现.

表 3 在 miniImageNet 和 CUB 数据集上的消融实验

方法	miniImageNet		CUB	
	1-shot	5-shot	1-shot	5-shot
ProtoNet	49.42	68.20	68.23	84.03
ProtoNet*	49.98	69.53	69.11	85.27
MAML	49.33	65.17	55.92	72.09
MAML*	49.85	66.43	57.31	73.98
本文方法	60.11	74.23	75.65	88.42

注: *表示我们使用增强版Conv4-64的实现

4.5.2 加特征集合的使用频率

为了确定是否所提取的特征对于最终的分类都有效, 我们分析了在集合度量中使用的特征集合中每个特征向量的使用频率. 图 5 展示了在 miniImageNet 数据集上进行 600 轮测试时, 每个映射器在每个查询类中提供最小原型-查询距离的频率, 该结果是通过 ResNet-12 模型获得的, 并在 600 个 5-way 1-shot 的实验回合中取平均值. 这些结果表明, 尽管早期的特征更频繁地被激活, 但所有特征在分类过程中始终保持有效. 这进一步验证了基于集合的表示的有效性.

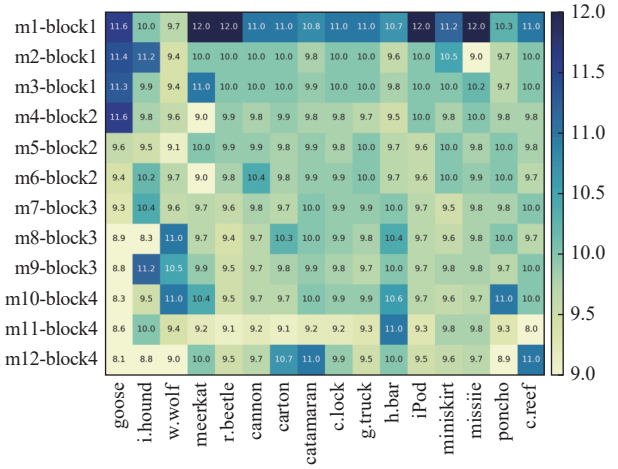
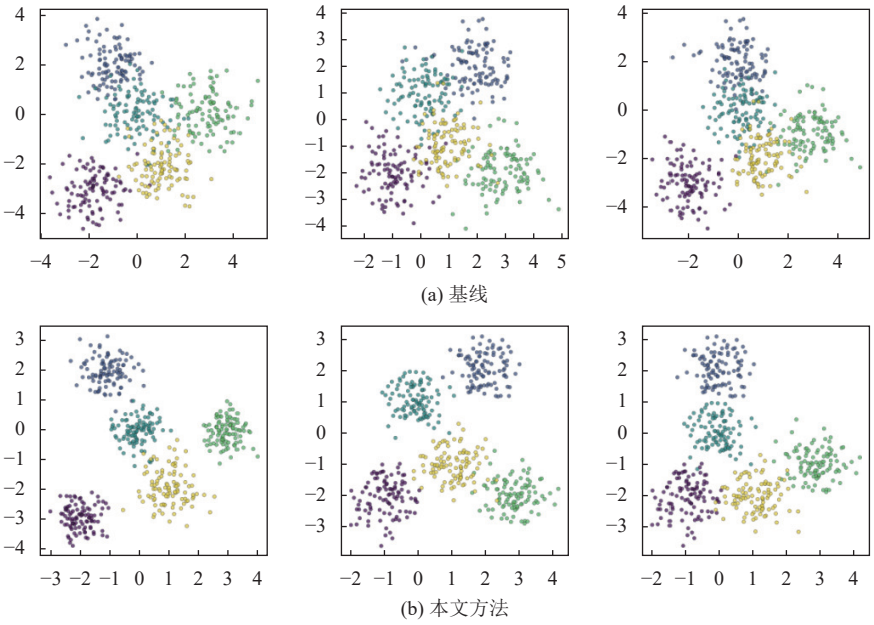


图 5 特征集合中每个特征向量的使用百分比 (纵轴) 在 miniImageNet 数据集的 16 个验证类别 (横轴) 上的选择情况

为了更好地进行可视化, 在图 6 中展示了本文方法和原型网络的 t-SNE 可视化, 其中每个类的基本圆圈是通过在评估阶段从支持集随机选择计算得到的. 为了更好地区分不同类别, 同一类的样本应该更加紧凑, 同一类样本之间的距离应该更大.



(a) 基线

(b) 本文方法

图 6 基于支持集计算的 t-SNE 图

4.5.3 加权集度量

为了展示加权集度量的有效性, 图 7 使用条形图展示了在 4 个数据集上使用加权和未加权集度量分别取得的准确率. 可以看到, 我们的加权集度量在所有 3 个数据集上均优于直接使用集度量. 特别是在 tiered-ImageNet 的 5-shot 分类中, 它提升了 2.66%.

4.5.4 我们方法的收敛速度.

近年来, 一些研究表明, 结合标准的迁移学习和第 2 阶段的元训练可以获得良好的性能. 初始阶段包括标准的预训练, 使用全连接层将多元增强网络的输出转换为 C 类进行分类. 因此, 与传统的 MAML 方法相比, 我们采用了一个两阶段的过程, 使用加权集度量来训练本文方法. 在这里, 我们展示了在第 1 阶段预训练过程中观察到的损失和收敛变化. 图 8 展示了所提方法在 CUB 数据集上进行 1-shot 和 5-shot 后的收敛趋势, 每次迭代包含两个任务.

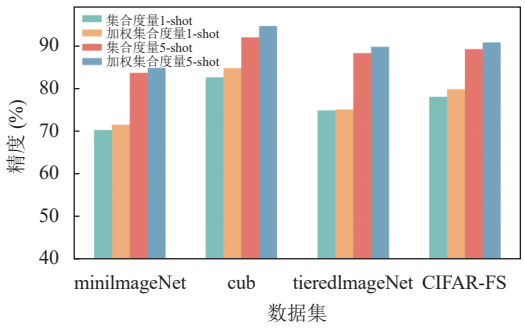
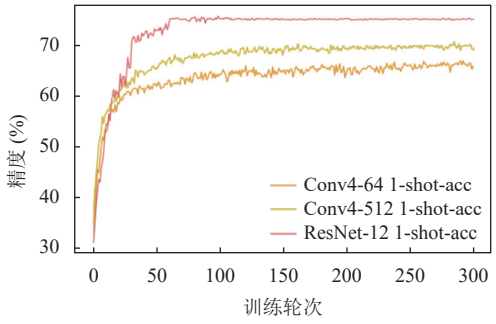
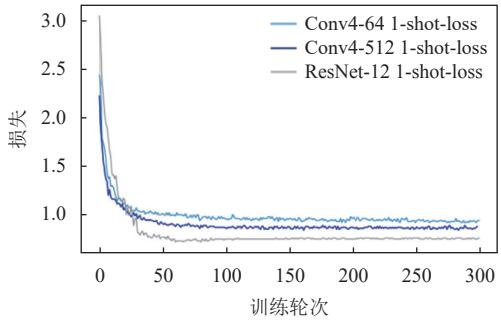


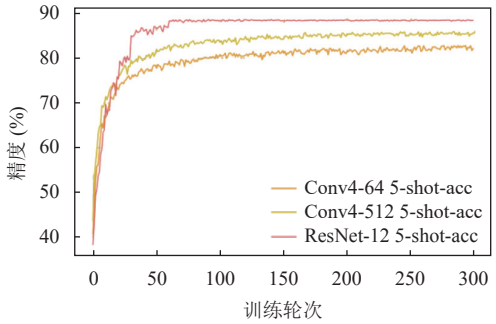
图 7 比较加权集合度量和集合度量在 4 个数据集上 1-shot 和 5-shot 任务的性能



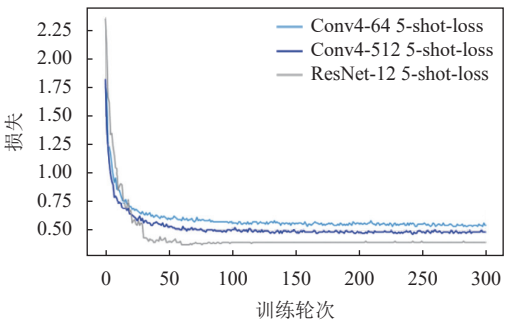
(a) 1-shot精度



(b) 1-shot损失



(c) 5-shot精度



(d) 5-shot损失

图 8 CUB 数据集上进行 1-shot 迭代和 5-shot 迭代的损失精度变化趋势

从曲线的变化可以看出, 我们的方法即使在预训练阶段也取得了令人满意的结果. 这表明我们的方法在模型训练的早期阶段展现了良好的性能趋势. 通过分析这些曲线, 我们观察到模型的性能在预训练过程中逐渐提高, 证明了它在捕捉数据特征和优化模型参数方面的有效性. 这些结果为进一步的微调和模型优化提供了坚实的基础, 确保了后续任务中的稳定可靠表现. 图 8 中也展示了我们方法在 CUB 数据集上进行 1-shot 和 5-shot 后的损失函数变化, 每次迭代包含两个任务. 这个可视化展示了我们方法在迭代过程中的损失演变, 展示了它在训练过程中的表现. 通过观察损失曲线的变化, 我们可以清晰地看到模型在每个训练阶段的性能.

4.5.5 估指标

为了更准确地评估我们的模型在小样本分类任务中的表现, 我们在 miniImageNet 数据集上选择 16 个类别进行 AP 和 mAP 计算. 如图 9 所示.

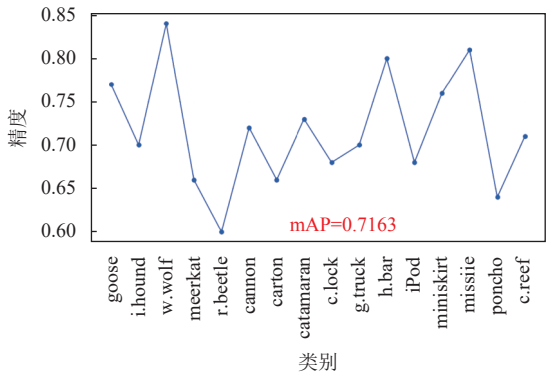


图 9 在 miniImageNet 上计算 AP 和 mAP

4.6 局限性

我们在原始骨干网络中嵌入了固定数量的多元嵌入扩展模块来提取一组特征嵌入. 虽然我们尝试过增加特征嵌入的数量, 但实验结果表明, 增加特征嵌入的数量需要相应地减少过滤器的数量, 这可能会导致网络参数不足和欠拟合. 因此, 在更大的骨干网络上增加特征嵌入数量可能具有潜在价值. 另一个值得进一步研究的方面是加权集度量. 目前, 我们通过合成特征之间的相似性和差异性为集合中的不同特征表示分配不同的权重, 它可能仍然缺乏适应不同任务的灵活性. 未来的研究可以探索更灵活的生成权重的方法.

5 结论与展望

本文介绍了多元嵌入增强网络在少样本分类中的应用. 它摒弃了传统的基于图像级特征表示的方法, 采用一组特征嵌入来表示图像样本, 这与基于图像级特征表示的分类方法有本质区别. 为了提取一组特征嵌入, 我们通过在常用于少样本学习的 3 个骨干网络中嵌入多元嵌入增强网络来进行提取一组特征. 这个模块依赖于骨干网络学习, 以捕捉图像中不同尺度的特征. 为了增强特征集的提取, 我们使用一个多尺度聚合模块, 在将特征图送入多元嵌入扩展模块之前, 融合低层次特征的信息. 在分类过程中, 我们使用基于集合的度量, 从支持集样本中推断给定查询的类别. 可以观察到, 理想情况下, 算法应该通过对负特征赋予较小的权重来适应, 从而确保无论它们与任何特征嵌入的匹配如何, 它们对整体距离的贡献保持最小. 因此, 为了抑制负特征并促进对最终分类结果影响较大的正特征, 引入了一种无监督方法, 基于集合度量为每个特征嵌入生成一个动态自适应权重. 本文方法在 miniImageNet、tieredImageNet 和 CUB 数据集上取得了最新的成果.

参考文献

1 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述. 计算机学报, 2017, 40(6): 1229–1251. [doi: 10.11897/SP.J.1016.2017.01229]

2 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述. 计算机学报, 2020, 43(5): 755–780. [doi: 10.11897/SP.J.1016.2020.00755]

3 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络. 自动化学报, 2016, 42(9): 1300–1312.

4 杨真真, 匡楠, 范露, 等. 基于卷积神经网络的图像分类算法综述. 信号处理, 2018, 34(12): 1474–1489.

5 马永杰, 刘培培. 基于 DenseNet 进化的卷积神经网络图像分类算法. 激光与光电子学进展, 2020, 57(24): 241001.

6 张珂, 冯晓晗, 郭玉荣, 等. 图像分类的深度卷积神经网络模型综述. 中国图象图形学报, 2021, 26(10): 2305–2325.

7 刘颖, 雷研博, 范九伦, 等. 基于小样本学习的图像分类技术综述. 自动化学报, 2021, 47(2): 297–315.

8 Chen D, Chen YF, Li YH, *et al*. Self-supervised learning for few-shot image classification. Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto: IEEE, 2021. 1745–1749.

9 Li XX, Yang XC, Ma ZY, *et al*. Deep metric learning for few-shot image classification: A review of recent developments. Pattern Recognition, 2023, 138: 109381. [doi: 10.1016/j.patcog.2023.109381]

10 Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 4080–4090.

11 Sung F, Yang YX, Zhang L, *et al*. Learning to compare: Relation network for few-shot learning. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 1199–1208.

12 Bateni P, Barber J, Van de Meent JW, *et al*. Enhancing few-shot image classification with unlabelled examples. Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022. 1597–1606.

13 Zhao YR, Gao XT, Shumailov I, *et al*. Rapid model architecture adaption for meta-learning. Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022. 1360.

14 Rusu AA, Rao D, Sygnowski J, *et al*. Meta-learning with

latent embedding optimization. Proceedings of the 7th International Conference on Learning Representations. New Orleans, 2019.

15 Huang HM, Lin LF, Tong RF, *et al*. UNet 3+: A full-scale connected UNet for medical image segmentation. Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona: IEEE, 2020. 1055–1059.

16 Li WB, Wang L, Xu JL, *et al*. Revisiting local descriptor based image-to-class measure for few-shot learning. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7253–7260.

17 Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision, 2000, 40(2): 99–121. [doi: 10.1023/A:1026543900054]

18 Zhu H, Koniusz P. EASE: Unsupervised discriminant subspace learning for transductive few-shot learning. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 9068–9078.

19 Afrasiyabi A, Larochelle H, Lalonde JF, *et al*. Matching feature sets for few-shot image classification. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 9004–9014.

20 Ma RK, Fang PF, Drummond T, *et al*. Adaptive poincaré point to set distance for few-shot classification. Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI, 2022. 1926–1934.

21 Xu CM, Fu YW, Liu C, *et al*. Learning dynamic alignment via meta-filter for few-shot learning. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 5178–5187.

22 Kang D, Kwon H, Min J, *et al*. Relational embedding for few-shot classification. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 8802–8813.

23 Afrasiyabi A, Lalonde JF, Gagné C. Mixture-based feature space learning for few-shot image classification. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9021–9031.

24 Xie JT, Long F, Lv JM, *et al*. Joint distribution matters: Deep brownian distance covariance for few-shot classification. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 7962–7971.

25 Qiao Q, Xie Y, Zeng ZY, *et al*. TALDS-Net: Task-aware adaptive local descriptors selection for few-shot image classification. Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Seoul: IEEE, 2024. 3750–3754.

26 Cheng H, Yang SY, Zhou JT, *et al*. Frequency guidance matters in few-shot learning. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 11780–11790.

27 He J, Kortylewski A, Yuille A. CORL: Compositional representation learning for few-shot classification. Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2023. 3879–3888.

28 Lee GY, Dam T, Poenar DP, *et al*. HELA-VFA: A hellinger distance-attention-based feature aggregation network for few-shot classification. Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2024. 2162–2172.

29 Xu RH, Jiang KZ, Qi LL, *et al*. DCPNet: Distribution calibration prototypical network for few-shot image classification. IEEE Access, 2024, 12: 67036–67045. [doi: 10.1109/ACCESS.2024.3398134]

30 Hu YQ, Pateux S, Gripon V. Adaptive dimension reduction and variational inference for transductive few-shot classification. Proceedings of the 26th International Conference on Artificial Intelligence and Statistics. Valencia: PMLR, 2023. 5899–5917.

31 Liu B, Cao Y, Lin YT, *et al*. Negative margin matters: Understanding margin in few-shot classification. Proceedings of the 16th European Conference. Glasgow: Springer, 2020. 438–455.

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

33 Wertheimer D, Tang LM, Hariharan B. Few-shot classification with feature map reconstruction networks. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 8008–8017.

(校对责编: 张重毅)