

# 采用特征图增强原型的小样本图像分类方法

许华杰<sup>1,2,3,4+</sup>, 梁书伟<sup>1</sup>

1. 广西大学 计算机与电子信息学院, 南宁 530004
  2. 广西多媒体通信与网络技术重点实验室, 南宁 530004
  3. 广西高校并行分布与智能计算重点实验室, 南宁 530004
  4. 广西智能数字服务工程技术研究中心, 南宁 530004
- + 通信作者 E-mail: hjxu2009@163.com

**摘要:**在基于度量学习的小样本图像分类方法中,由于标注样本的稀缺,仅用支持集样本得到的类原型往往难以代表整个类别的真实分布;同时,同类样本间也可能在多个方面存在较大差异,较大的类内差异可能使样本特征偏离类别中心。针对上述可能严重影响图像分类性能的问题,提出一种采用特征图增强原型的小样本图像分类方法(FMEP)。首先,用余弦相似度从查询集样本特征图中选择部分相似特征加入类原型中,得到更具代表性的特征图增强原型;其次,对相似的查询集样本特征进行聚合,缓解类内差异大导致的问题,使同类样本的特征分布更接近;最后,用在特征空间中与真实类别分布都更接近的特征图增强原型和聚合查询特征进行相似度比较得到更优的分类结果。所提方法在 MiniImageNet、TieredImageNet、CUB-200 和 CIFAR-FS 等常用的小样本图像分类数据集上进行了实验,结果表明所提方法获得了比基线模型更优的分类性能,同时也优于同类型的小样本图像分类方法。

**关键词:**小样本学习;图像分类;度量学习;特征图增强原型;余弦相似度

**文献标志码:**A **中图分类号:**TP391

## Few-Shot Image Classification Method with Feature Maps Enhancement Prototype

XU Huajie<sup>1,2,3,4+</sup>, LIANG Shuwei<sup>1</sup>

1. College of Computer and Electronic Information, Guangxi University, Nanning 530004, China
2. Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China
3. Guangxi Key Laboratory of Parallel, Distributed and Intelligent Computing, Nanning 530004, China
4. Guangxi Intelligent Digital Services Research Center of Engineering Technology, Nanning 530004, China

**Abstract:** Due to the scarcity of labeled samples, the class prototype obtained by support set samples is difficult to represent the real distribution of the whole class in metric-based few-shot image classification methods. Meanwhile, samples of the same class may also have large difference in many aspects and the large intra-class bias may make the sample features deviate from the class center. Aiming at the above problems that may seriously affect the performance, a few-shot image classification method with feature maps enhancement prototype (FMEP) is proposed. Firstly, this paper selects some similar features of the query set sample feature maps with cosine similarity and adds them to class prototypes to obtain more representative prototypes. Secondly, this paper aggregates similar features of the query

**基金项目:**国家自然科学基金(71963001);广西科技计划项目(2017AB15008);崇左市科技计划项目(FB2018001)。

This work was supported by the National Natural Science Foundation of China (71963001), the Science and Technology Plan Project of Guangxi (2017AB15008), and the Science and Technology Plan Project of Chongzuo (FB2018001).

**收稿日期:**2023-02-07 **修回日期:**2023-04-18

set to alleviate the problem caused by large intra-class bias and makes features distribution of the same class closer. Finally, this paper compares enhanced prototypes and aggregated features which are both closer to real distribution to get better results. The proposed method is tested on four commonly used few-shot classification datasets, namely MiniImageNet, TieredImageNet, CUB-200 and CIFAR-FS. The results show that the proposed method can not only improve the performance of the baseline model, but also obtain better performance compared with the same type of methods.

**Key words:** few-shot learning; image classification; metric learning; feature maps enhancement prototype; cosine similarity

在图像分类任务上,深度学习技术凭借着数亿级别的参数模型和数百万甚至是数亿级别的标注数据在图像分类任务上的准确率早已远超人类<sup>[1]</sup>。深度学习技术的出现得益于海量的数据,大大推动了人工智能领域的发展。然而这种依赖于大量标注数据的学习方式,与人类的学习方式还相差甚远。受到人类学习方式的启发,小样本学习(few-shot learning, FSL)正是为了实现让模型能够从少量的标注样本中学习新类别而提出的研究方向,其中用于学习的新类别标注样本集合通常被称为支持集(support set),待分类的新类别样本集合则被称为查询集(query set)。目前小样本图像分类方法可以大致分为两类:基于微调的方法<sup>[2-3]</sup>和基于度量学习的方法<sup>[4-11]</sup>。基于微调的方法使用支持集中的少量的标注样本对模型参数进行调整;基于度量学习的方法通常在特征空间中度量支持集样本特征和查询集样本特征距离,并以距离大小作为分类的依据<sup>[12]</sup>。

小样本学习所面向的任务存在标注数据稀缺的问题,而待分类的数据数量较多。Snell等提出的ProtoNet<sup>[5]</sup>是基于度量学习方向的经典代表方法,其将每类支持集样本在特征空间表示的均值作为该类别的原型,通过对比查询样本与类原型间的距离将查询样本进行分类。然而ProtoNet通过少量样本得到的原型往往难以代表整个类别的真实分布;查询集中同类别样本间也可能在形状、大小、背景等多方面存在较大差异,较大的类内差异使查询集样本特征可能偏离类别中心。针对上述可能严重影响图像分类性能的问题,本文提出一种采用特征图增强原型的小样本图像分类方法(few-shot image classification method with feature maps enhancement prototype, FMEP)。首先,利用查询集样本对类原型进行增强,即从查询集特征图中筛选出与类原型相似的特征并加入到原始类原型,得到的特征图增强原型相比原始类原型在特征空间上更具代表性;其次,对于

相似的查询集样本特征进行聚合,使特征空间中部分相距较远的特征更靠近类别中心,缓解类内差异较大导致的问题;最后,用与真实类别分布更为接近的特征图增强原型与聚合后的查询集样本特征进行相似度比较得到最终分类结果。实验结果表明所提方法优于同类型的小样本图像分类方法。

## 1 相关工作

在基于度量学习的小样本图像分类方法方面,MatchingNet<sup>[4]</sup>是较早提出的方法,该方法结合注意力机制计算查询集样本与支持集中每个样本的相似性。随后提出具有代表性的方法是ProtoNet<sup>[5]</sup>,其通过均值计算得到每类样本的原型,通过对比类原型与查询集样本的距离对查询集样本进行分类。RelationNet<sup>[6]</sup>使用卷积神经网络层“学习”如何度量查询集样本和支持集样本的相似度,而不是用常规的距离度量函数(比如欧式距离、余弦相似度)。Meta-Baseline<sup>[7]</sup>是近期出现的具有代表性的方法,其分类方式与ProtoNet一样使用余弦相似度将类原型与查询集样本进行比较,不同之处在于其使用两阶段的训练方式,却能极大提高分类性能。MSFA(multi-scale feature aggregation)<sup>[8]</sup>提出一种多尺度特征聚合的方法,利用多尺度的样本特征获得更为准确的特征表示,并提出一种加权距离,从而完成更为准确的距离度量。RENet(relational embedding network)<sup>[9]</sup>提出自关联信息模块,使特征更关注于物体以及交叉信息提取模块,使查询特征与支持特征进行交互后能更好定位到待分类的物体。DeepBDC(deep Brownian distance covariance)<sup>[10]</sup>提出新的距离度量函数——深度布朗距离,其能够以较小的计算开销刻画特征向量之间的相关性和独立性。MCL(mutual centralized learning)<sup>[11]</sup>提出互中心化学习,能够关注到图像局部特征的稠密表示,更好地发掘查询图像与支持图像的关系。

为了缓解小样本学习中标注数据稀缺的问题,一种可行的方式是将支持集学习到的知识推广到查询集中未标注数据中,并利用查询集中的有用信息进行学习<sup>[13]</sup>,这类方法也被称之为直推式(transductive)方法,而前述的ProtoNet、Meta-Baseline等方法则都属于归纳式(inductive)方法。Ren等<sup>[14]</sup>通过计算查询集特征与类原型的距离得到样本属于该类的置信度,将置信度作为权重加入到类原型中,并设计了三种引入查询集样本更新类原型的策略。CAN(cross attention network)<sup>[15]</sup>是一种基于交叉注意力网络的方法,利用查询集特征和类原型的相关性作为权重更新类原型。BD-CSPN(bias diminishing cosine similarity based prototypical network)<sup>[16]</sup>使用伪标签生成策略为查询样本预测类别,将概率值作为权重加到类原型中,并在新的类原型中将支持样本与类中心的距离作为权重赋予支持样本。上述三种方法共同之处在于将查询集特征加入到类原型中,从而获得更具代表性的原型。但是它们的做法都是获得查询集特征图后直接使用全局平均池化(global average pooling, GAP)获取特征,这样的做法可能会引入一定的噪声,因为在一张图像当中除了标注的物体之外,还存在大量的无关背景或者是其他类别的物体,这些无关的信息都会被一起加入到类别原型中,对最终分类效果产生不良影响。本文方法与上述三种方法的不同之处在于:只从特征图中筛选出与类原型最相似的部分用于增强类原型。所提方法与三种方法相比较,虽然增加了一定的计算量,但既可以使类原型更具代表性,又减少了查询集特征图中无关信息对于类原型的不良影响。由于图神经网络(graph neural network, GNN)在表示特征间关系方面的优势,一些方法采用GNN作为距离度量函数:TPN(transductive propagation network)<sup>[17]</sup>将支持集特征和查询集特征作为GNN的节点,并提出标签传播算法将标签从支持集特征传播到查询集特征以实现分类;TGAN(transductive graph-attention network)<sup>[18]</sup>设计图注意力模块用于更好探索查询集特征与支持集特征的关系,同样使用标签传播算法进行分类。TEAM(transductive episodic-wise adaptive metric)<sup>[19]</sup>设计一种不同任务下自适应调整的距离度量(episodic-wise adaptive metric, EAM),并采用双向相似性度量(bi-directional similarity, Bi-SIM)来计算分类概率,两种度量计算过程都使用了查询集特征。CGC(curvature generation in curved spaces)<sup>[20]</sup>等创造性地提出

一种不同任务下自感知的曲率空间(task-aware curved embedding space),能够利用样本的类内信息和类间上下文信息生成更具有代表性的类别原型。 $\alpha$ -TIM(transductive information maximization based on  $\alpha$ -divergence)<sup>[21]</sup>提出一种基于 $\alpha$ 散度的互信息损失,能够有效应对类别分布的偏差。本文所提方法仅用余弦相似度作为距离度量函数,不仅避免了复杂的网络结构,而且可以较小的计算开销获得较好的性能。

## 2 采用特征图增强原型的小样本图像分类方法

### 2.1 方法概要

在小样本图像分类任务当中,数据集被分为训练集  $D_{\text{train}}$ 、支持集  $S$  和查询集  $Q$ 。其中训练集包含的样本都是标注样本且数量较多,训练集样本所属类别通常被称为基础类  $C_{\text{base}}$ ;支持集和查询集是测试集  $D_{\text{test}}$  中少量标注样本和大量无标注样本的集合,它们所包含的类别被称为稀有类  $C_{\text{novel}}$ 。稀有类与基础类完全不同,即  $C_{\text{base}} \cap C_{\text{novel}} = \emptyset$ 。小样本图像分类任务的目标就是让神经网络从给定的训练集和支持集的标注样本中进行学习,从而使其能够正确地对查询集无标注样本进行分类。一般来说,在测试阶段,构造  $T$  个  $N$ -way  $K$ -shot 元任务,即每个元任务从测试集中选择  $N$  个类,每个类选取  $K$  个标注样本作为支持集,并选取每类一定的剩余样本作为查询集。例如,在常见的元任务设置中,将  $N$  设置为 5,将  $K$  设置为 1 或 5,每类选取 15 个样本作为查询集样本,那么支持集样本的数量为  $1 \times 5 = 5$  或  $5 \times 5 = 25$ ,查询集样本的数量为  $5 \times 15 = 75$ 。

在分类模型训练阶段,先用训练集样本训练得到特征提取器  $f_{\theta}$ 。在测试阶段的一个元任务当中,首先将支持集样本  $x^S$  和查询集样本  $x^Q$  送入特征提取器  $f_{\theta}$  得到特征图  $M^S \in \mathbb{R}^{c \times h \times w}$  和  $M^Q \in \mathbb{R}^{c \times h \times w}$ ,其中  $c$ 、 $h$ 、 $w$  分别表示特征图的通道数、长和宽。然后使用 GAP 将特征图转换为特征向量  $F^S \in \mathbb{R}^{c \times 1 \times 1}$  和  $F^Q \in \mathbb{R}^{c \times 1 \times 1}$ ,即:

$$F = \text{GAP}(f_{\theta}(x)) \quad (1)$$

将支持集特征向量  $F^S$  中每类特征累加后取均值得到的特征向量称之为支持原型  $P$ 。支持集中属于类别  $c$  的支持原型  $P_c$  的计算公式为:

$$P_c = \frac{1}{K} \sum_{i=1}^K F_{i,c}^S \quad (2)$$

最后根据查询特征  $F^Q$  与支持原型  $P$  的相似度对查询集样本进行分类。查询集样本  $x^Q$  属于类别  $c$  的概率计算公式为：

$$p(y=c|x^Q)=\frac{\exp(\langle F^Q, P_c \rangle)}{\sum_{c'=1}^N \exp(\langle F^Q, P_{c'} \rangle)} \quad (3)$$

其中,  $\langle \cdot, \cdot \rangle$  表示相似性度量函数。从式(3)中可知, 如果  $F^Q$  与其对应类别的支持原型  $P_c$  在特征空间中与真实类别分布都越接近, 那么  $F^Q$  与  $P_c$  的相似度就会远大于  $F^Q$  与其他类别支持原型的相似度, 计算得到的  $F^Q$  属于对应类别  $c$  的概率也随之越高, 分类的结果也会越准确。

本文提出采用特征图增强原型的小样本图像分类方法(FMEP), 框架图如图1所示, 其主要步骤如下：

(1)提取基础特征。将支持集和查询集图像送入训练好的卷积神经网络特征提取器中, 得到支持集特征图 and 查询集特征图, 每类的支持集特征图根据式(1)和式(2)计算后得到支持原型。

(2)获得特征图增强原型。如图1中支持原型增强模块(support prototype enhancement module, SPEM)所示, 为了获得更具代表性的支持原型, 用伪标签策略为查询集特征图分配伪标签, 根据伪标签计算特征图各部分与对应支持原型的相似度, 仅选择最相似的部分加入支持原型中得到特征图增强原型。

(3)获得聚合查询集特征。如图1中查询特征聚合模块(query feature aggregation module, QFAM)所

示, 为了缓解类内差异大导致的问题, 计算查询集特征与其他查询集特征的相似度, 选择前  $k$  个最相似特征累加得到聚合查询集特征。

(4)获取分类结果。最后用余弦相似度对特征图增强原型和聚合查询集特征进行相似度比较得到更为准确的分类结果, 因为它们在特征空间上都更为接近真实类别分布。查询集样本  $x^Q$  属于类别  $c$  的概率计算公式为：

$$p(y=c|x^Q)=\frac{\exp(\tau \cdot \cos(\tilde{F}^Q, \tilde{P}_c))}{\sum_{c'=1}^N \exp(\tau \cdot \cos(\tilde{F}^Q, \tilde{P}_{c'}))} \quad (4)$$

其中,  $\cos(a, b) = a \cdot b / \|a\| \|b\|$ ,  $\tilde{P}_c$  为特征图增强原型,  $\tilde{F}^Q$  为聚合查询集特征,  $\tau$  为温度系数。以下主要对步骤(2)所涉及的特征图增强原型以及步骤(3)所涉及的查询特征聚合做进一步介绍。

### 2.2 特征图增强原型

由于小样本学习所面临的任务通常存在标注数据稀缺的问题, 仅用少量支持样本学习得到的支持原型对大量查询集样本进行正确分类是很困难的, 这也是小样本学习面临的长期困境。在现有的与FMEP较为类似的直推式方法中<sup>[14-16]</sup>, 普遍采用带伪标签的查询集特征对支持原型进行增强。但此类方法仍存在的问题: 一是伪标签可能存在部分错误; 二是图像中除了存在标注物体类别的信息之外, 还存在大量的无关背景和其他物体的信息, 这些干扰信息也都会被加入到支持原型中, 引入了较多的噪声, 对最终分类结果产生不良影响。在基于度量学习的小样本图像分割方法(few-shot segmentation)<sup>[22]</sup>中, 通过用类原型与查询集特征图进行相似度度量, 相

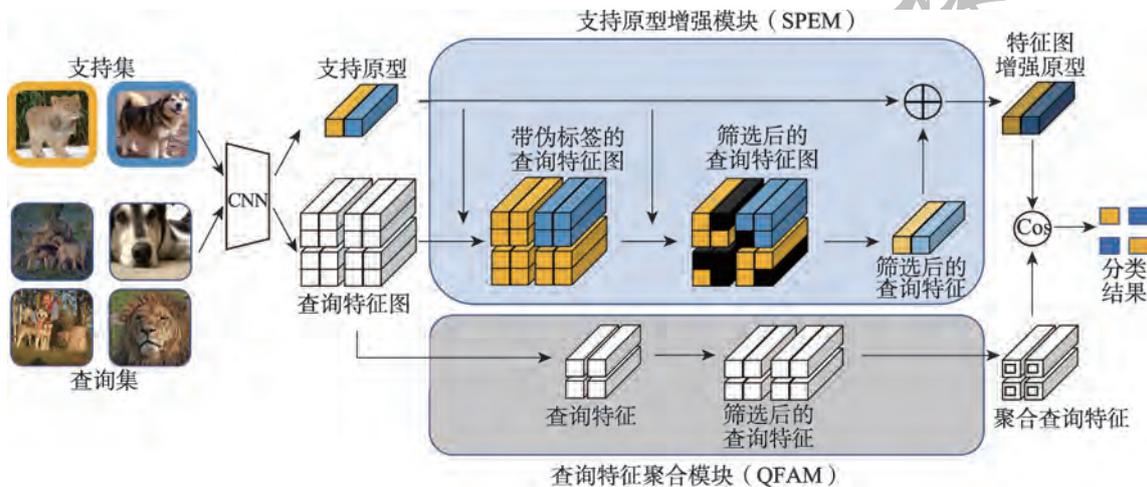


图1 FMEP方法框架图

Fig.1 Framework of FMEP method

似度矩阵中值较大的区域往往就是待分割类别物体所处的位置。受此类方法启发,如果仅将查询集特征图中与支持原型余弦相似度较大的区域加入到支持原型中以获得特征图增强原型,既可以达到使支持原型更接近类别真实分布的目的,又减少了查询集特征图中干扰信息对于支持原型的影响。以上做法在本文所提方法中以支持原型增强模块的方式加以实现,流程如图2所示。以类别为  $c$  的支持原型  $P_c$  为例,其流程主要包括以下5个步骤:

(1)将全部类别的支持原型和查询集特征  $F^Q$  根据式(3)获取全部查询集样本的伪标签  $l$ ,进而得到带伪标签的查询集特征图  $\hat{M}^Q$ 。

(2)计算  $\hat{M}^Q$  中属于类别  $c$  的特征图  $\hat{M}_c^Q$  与  $P_c$  在空间上相似度矩阵  $A_c \in \mathbb{R}^{h \times w}$ ,计算公式为:

$$A_c = \text{softmax}(\tau \cdot \cos(\hat{M}_c^Q, P_c)) \quad (5)$$

(3)根据相似度阈值  $t$ ,将相似度矩阵  $A_c$  中大于阈值  $t$  的元素设置为1,其余为0。阈值  $t$  的作用是为了筛选  $\hat{M}_c^Q$  中与  $P_c$  中相似的区域,可以通过调节  $t$  的大小来增多或减少加入  $P_c$  中的特征,  $t$  的具体数值设置将在后续实验部分进行详细讨论。设  $a_{ij}$  为  $A_c$  中的元素,则经过阈值筛选的相似度矩阵  $\tilde{A}_c \in \mathbb{R}^{h \times w}$  计算公式为:

$$\tilde{A}_c = \begin{cases} 1, & a_{ij} > t \\ 0, & a_{ij} \leq t \end{cases} \quad (6)$$

(4)用  $\tilde{A}_c$  对  $\hat{M}_c^Q$  进行筛选得到筛选后的查询集特征图  $\tilde{M}_c^Q$ :

$$\tilde{M}_c^Q = \tilde{A}_c \otimes \hat{M}_c^Q \quad (7)$$

(5)将  $\tilde{M}_c^Q$  加入  $P_c$  中获得特征图增强原型  $\tilde{P}_c$ :

$$\tilde{P}_c = P_c + \sum_{i=1}^{n_c} \text{GAP}(\tilde{M}_{i,c}^Q) \quad (8)$$

其中,  $n_c$  表示查询集样本中被预测为类别  $c$  的个数。

对于所有类别重复步骤(2)~(5)即可获得所有类别的特征图增强原型  $\tilde{P}$ 。

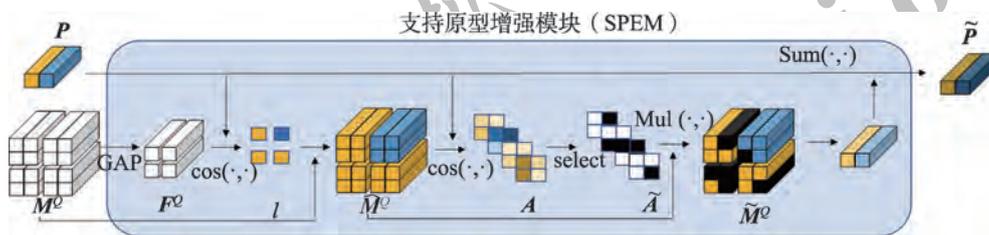


图2 支持原型增强模块处理流程图

Fig.2 Flow diagram of support prototype enhancement module

## 2.3 查询特征聚合

查询集中样本数量相较于支持集多,即使是同类的图像也可能在形状、大小、背景等诸多方面存在较大差异;同时特征提取器仅在基础类上进行充分训练,其提取得到的查询集特征可能存在部分远离真实类别分布的特征,不利于正确分类。在前述支持原型增强模块中,采用多个来自同一类别特征的平均分布替代有偏的单一特征分布,从而更接近真实类别分布;此外,在实践中5-shot的情况下( $K=5$ )分类准确率也远高于1-shot的情况。基于以上两方面的考虑,将查询集特征之间的余弦相似度最大的  $k$  个特征视为同一类别的特征,并将其聚合得到聚合查询特征,其在特征空间中的距离更为接近真实类别分布中心。以上做法在本文所提方法中以查询特征聚合模块的方式加以实现,流程如图3所示,包括以下3个步骤:

(1)根据式(1)从查询集特征图  $M^Q$  中计算得到查询特征  $F^Q$ 。

(2)计算查询特征  $F^Q$  之间的余弦相似度,得到自相似度矩阵  $A^Q \in \mathbb{R}^{|Q| \times |Q|}$ ,其中  $|Q|$  表示一个元任务中查询集样本的数量,  $A^Q$  的计算公式为:

$$A^Q = \cos(F^Q, F^Q) \quad (9)$$

(3)对于查询样本特征  $F_i^Q$ ,根据  $A^Q$  选择  $k$  个最为相似的特征进行聚合。  $A^Q$  中第  $i$  行第  $j$  列的数值表示第  $i$  个查询特征  $F_i^Q$  与第  $j$  个查询特征  $F_j^Q$  的相似度,若用  $\text{topk}(A_i^Q)$  表示与  $F_i^Q$  最为相似的  $k$  个查询

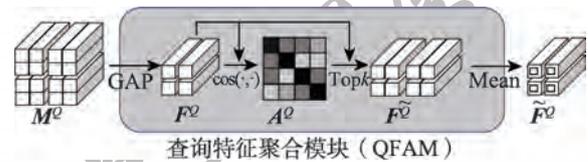


图3 查询特征聚合模块处理流程图

Fig.3 Flow diagram of query feature enhancement module

特征的索引集合,那么聚合特征  $\tilde{F}_i^Q$  计算公式为:

$$\tilde{F}_i^Q = \frac{1}{k+1} \left( F_i^Q + \sum_{j=1}^{topk(A_i^Q)} F_j^Q \right) \quad (10)$$

对于所有查询特征重复步骤(3)即可得到所有的聚合查询特征  $\tilde{F}^Q$ 。但由计算相似度得到的这  $k$  个相似特征并不能确保一定属于同一类,因此聚类数  $k$  的值不能设置过大,过大将会引入较多噪声,  $k$  的具体设置也将在后续实验部分进行详细讨论。

### 3 实验及其结果分析

#### 3.1 实验数据集及实验设置

本文所提 FMEP 方法在 MiniImageNet<sup>[4]</sup>、Tiered-ImageNet<sup>[14]</sup>、CUB-200 和 CIFAR-FS<sup>[23]</sup> 四个常用的小样本图像分类数据集上进行了验证。其中 MiniImageNet 和 TieredImageNet 都是大规模图像数据集 ImageNet<sup>[1]</sup> 的子集; CUB-200 是由加州理工学院提供的细粒度鸟类图像数据集,通常细粒度数据集具有类间相似和类内差异大的特点,这使得细粒度图像分类任务更加具有挑战性<sup>[24]</sup>; CIFAR-FS 由 Bertinetto 等<sup>[23]</sup> 提出,来源于 CIFAR-100 数据集。需要注意的是,与深度学习中常规有监督图像分类方法不同,小样本图像分类方法中使用的训练集、验证集和测试集往往类别数量不同且类别间并无交集。遵循主流的小样本图像分类方法<sup>[2-11,14-21]</sup> 的常规做法将数据集划分为训练集、测试集和验证集。(1)训练集的类别数较多且样本数也较多,用于训练具有良好表征能力的特征提取器。(2)方法往往并不关心在训练集类别上的分类准确率,而是注重将学习得到的旧类别(训练集类别)的知识应用于新类别(验证集、测试集类别)

的能力,以达到在小样本场景下只用少数标注样本即可学习到新类别分类信息的目的。因此,小样本图像分类方法通常以在测试集类别上的分类准确率作为性能评判的标准:通过在测试集上构建  $T$  个  $N$ -way  $K$ -shot 元任务来模拟小样本的场景,其中标注样本的集合为支持集,待分类样本的集合为查询集,最终计算  $T$  个元任务下查询集样本的平均分类准确率。(3)验证集的作用是在训练的过程中挑选分类准确率最高的特征提取器,通过在验证集上构建  $T'$  个  $N$ -way  $K$ -shot 元任务(通常  $T' \ll T$ ),选择  $T'$  个元任务下查询集样本平均分类准确率最高的特征提取器作为测试阶段使用的最终模型。

实验过程中四个图像数据集的划分遵循主流小样本图像分类方法的做法<sup>[2-11,14-21]</sup>: MiniImageNet 和 CIFAR-FS 都有 100 个类别,每个类别都有 600 张图像,分别选取 64、16、20 类作为训练集、验证集和测试集; TieredImageNet 包含 34 个大类别,每个大类又包含 10~30 个小类别,共 608 个类别,每个类别的样本数量不等,分别选择 20、6、8 个大类别作为训练集、验证集和训练集; CUB-200 包含 200 种鸟类的 11 788 幅图像,分别选择 100、50、50 类作为训练集、验证集和测试集。

实验采用的特征提取器为 ResNet-12 网络模型<sup>[25]</sup>,选取 Meta-Baseline<sup>[7]</sup> 作为实验的基线模型,所采用的训练方法及各项参数的设置基本遵循 Meta-Baseline 中的做法:训练分为两阶段进行,第一阶段用随机初始化后的 ResNet-12 和线性分类器在训练集上使用交叉熵损失函数进行训练,在单 GPU 上设置 batch 大小为 64,共训练 200 个批次,其中在 MiniImageNet 上训练模型时的损失收敛曲线如图 4(a)所示;第二阶

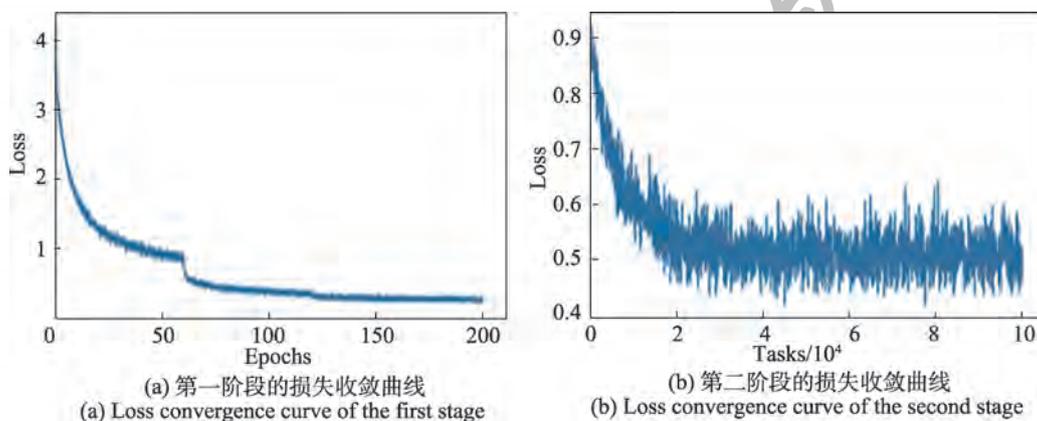


图4 两阶段训练过程的损失收敛曲线

Fig.4 Loss convergence curves for two-stage training

段仍在训练集类别上随机构建 100 000 个  $N$ -way  $K$ -shot 元任务, 这里的  $N$  设置为 5,  $K$  设置为 1 或 5, 每类选取 15 个样本作为查询集, 故每个元任务共有 75 个查询集样本, 模仿测试阶段的分类过程但不使用线性分类器, 只使用上一步训练好的 ResNet-12 模型进行度量学习, 其中 1-shot 的设置下在 MiniImageNet 上训练模型时的损失收敛曲线如图 4(b) 所示。最终测试阶段使用的是在验证集上具有最高分类准确率的模型, 在测试集上共构建 2 000 个  $N$ -way  $K$ -shot 元任务, 查询集样本数量设置与训练阶段一样, 采用 Top-1 平均准确率衡量(即 2 000 个元任务中所有任务准确率的平均值)模型分类性能。

支持原型增强模块中用到的相似度阈值  $t$  和查询特征聚合模块中用到的聚类数  $k$  是 FMEP 方法中的两个重要超参数, 因此有必要通过实验探究这两个超参数不同的设置对于分类准确率的影响。图 5 展示了在不同  $t$  和  $k$  组合的情况下, FMEP 在 MiniImageNet 数据集 5-way 1-shot 设置下分类平均准确率的变化情况。

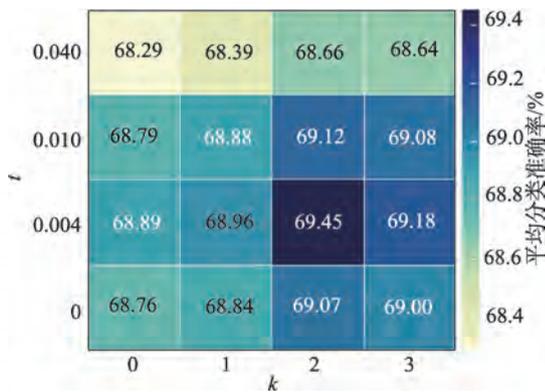


图5 不同阈值  $t$  和聚类数  $k$  组合下的平均分类准确率

Fig.5 Average classification accuracies according to different  $t$  and  $k$

实验中  $t$  的初始值选取为 0.04, 选取的依据为: 在 ResNet-12 作为特征提取器及输入图片大小为  $84 \times 84$  的情况下, 设输出特征图的长和宽均为 5, 即相似度矩阵大小为  $5 \times 5$ , 取均值即  $1/(5 \times 5) = 0.04$ 。当  $t$  从 0.04 一直降到 0 时 ( $t$  为 0 表示将所有伪标签查询特征图都加入到与其对应的支持原型中), 从图 5 中可以观察到分类准确率先上升而后下降。其原因在于, 随着阈值  $t$  的逐渐减小, 加入到支持原型的特征会变得更多, 使得支持原型逐渐接近真实分布从而获得较高分类准确率; 但是如果将特征图中的所有特征都加入到支持原型, 特征图中与支持类别无关

的特征, 比如背景、干扰物等会对支持原型产生不良的影响, 从而导致分类准确率反而降低。同理, 当聚类数  $k$  从 0 到 4 ( $k$  为 0 代表不对查询特征进行增强) 进行变化时, 也观察到了分类准确率先上升而后下降的现象。根据以上分析, 最终在后续所有实验中, 折中选择相似度阈值  $t$  设置为 0.004 和聚类数  $k$  设置为 2 的超参数设置组合。

### 3.2 对比实验

为了测试方法的性能, 通过对比实验将所提出的 FMEP 分别与 ProtoNet<sup>[4]</sup>、Meta-Baseline<sup>[7]</sup> 两种归纳式以及与其同属于直推式的 TPN<sup>[17]</sup>、TEAM<sup>[19]</sup>、CAN<sup>[15]</sup>、BD-CSPN<sup>[16]</sup>、 $\alpha$ -TIM<sup>[21]</sup>、CGC<sup>[20]</sup>、TGAN<sup>[18]</sup> 等七种方法进行比较, 主要选择同类方法(直推式)进行对比。

FMEP 与其他对比方法在 MiniImageNet、Tiered-ImageNet、CUB-200 以及 CIFAR-FS 4 个数据集上的平均分类准确率结果如表 1 所示。出于公平考虑, 大部分方法采用的特征提取器是 ResNet-12, 例外的是:  $\alpha$ -TIM 所有结果采用的是 ResNet-18, 用“†”进行标注; CGC 部分结果采用的是 BigResNet-12 (ResNet-12 卷积层通道数为 64、128、256、512, BigResNet-12 卷积层通道数为 64、160、320、640), 用“‡”进行标注。表 1 中 Meta-Baseline 方法是本文的基线模型, 带“\*”的方法表示, 其所对应的所有准确率结果都是由论文复现得到的, 均略优于原论文中报告的结果; 表 1 中所列的 TPN 结果来源于 CAN 论文对比实验中给出的准确率结果, 其余方法的结果则直接来源于所对应的论文。“±”后的数字表示平均准确率在置信水平为 95% 的情况下的误差界限, 该误差界限的计算方法来自于概率统计学中总体均值的区间估计: 假设  $T$  个元任务的平均准确率为  $\bar{p}$ , 标准差为  $\sigma$ , 则  $\bar{p}$  的误差界限计算公式为  $z_{\alpha/2} \times \frac{\sigma}{\sqrt{T}}$ , 其中在置信水平为 95%

的情况下  $z_{\alpha/2}$  的取值为 1.96。在 MiniImageNet 数据集上, FMEP 相比于基线模型 Meta-Baseline, 分类准确率在 5-way 1-shot (在表中简称 5w1s) 和 5-way 5-shot (在表中简称 5w5s) 的设置下分别提升了 5.21 个百分点和 2.11 个百分点; 同特征提取器的情况下, 相比于对比方法中最好的 TGAN 方法和 CAN 方法, 准确率分别高出了 0.72 个百分点和 2.75 个百分点; 在 5-shot 设置下比使用 ResNet-18 的  $\alpha$ -TIM 准确率高 0.89 个百分点。在 TieredImageNet 数据集上, 与基线模型相比, 分类准确率在 1-shot 和 5-shot 的设置下分别提升了 5.28 个百分点和 1.01 个百分点; 同特征提取器

表1 在4个常用数据集上的平均分类准确率  
Table 1 Average classification accuracies on four popular datasets

单位：%

Method	Model	MiniImageNet		TieredImageNet		CUB-200		CIFAR-FS	
		5w1s	5w5s	5w1s	5w5s	5w1s	5w5s	5w1s	5w5s
Inductive	ProtoNet <sup>*[14]</sup>	56.13±0.45	75.70±0.33	59.11±0.52	75.30±0.42	74.35±0.48	88.50±0.25	57.67±0.51	82.65±0.35
	Meta-Baseline <sup>*[17]</sup>	64.24±0.46	81.28±0.31	72.32±0.50	85.18±0.36	78.18±0.43	90.04±0.23	73.82±0.47	86.45±0.33
Transductive	TPN <sup>[17]</sup>	59.46	75.65	58.68	74.26	—	—	—	—
	TEAM <sup>[19]</sup>	60.07	75.90	—	—	80.16	87.17	70.43	81.25
	CAN <sup>[15]</sup>	67.19±0.55	80.64±0.35	69.89±0.51	84.23±0.37	—	—	—	—
	BD-CSPN <sup>[16]</sup>	65.94	79.23	76.17	85.70	<b>84.90</b>	90.22	—	—
	$\alpha$ -TIM <sup>[21]</sup>	67.40†	82.50†	74.40†	<b>86.60†</b>	75.70†	89.80†	—	—
	CGC <sup>[20]</sup>	66.73±0.22	80.57±0.14	77.19±0.24‡	86.18±0.15‡	—	—	76.80±0.70‡	86.40±0.50‡
	TGAN <sup>[18]</sup>	68.73	75.61	70.63	85.22	—	—	—	—
	FMEP(Ours)	<b>69.45±0.53</b>	<b>83.39±0.30</b>	<b>77.60±0.55</b>	86.19±0.39	84.89±0.45	<b>91.56±0.22</b>	<b>79.82±0.50</b>	<b>87.66±0.33</b>

的情况下,相比于对比方法中最好的BD-CSPN方法,准确率分别高出了1.43个百分点和0.49个百分点;与使用参数量更大特征提取器的方法( $\alpha$ -TIM和CGC)准确率基本持平。在CUB-200数据集上,与基线模型相比,分类准确率在1-shot和5-shot的设置下分别提升了6.71个百分点和1.52个百分点;相比于对比方法中最好的BD-CSPN方法,在1-shot的设置下基本相同,在5-shot的设置下高出了1.34个百分点。在CIFAR-FS数据集上,与基线模型相比,分类准确率在1-shot和5-shot的设置下分别提升了6.00个百分点和1.21个百分点;相比于对比方法中最好的CGC方法,准确率分别高出了3.02个百分点和1.26个百分点。可见,在4个数据集的5-way 1-shot和5-way 5-shot的设置下,与对比方法相比FMEP都获得了最高的分类准确率。从表1的实验结果中还可以观察到在1-shot的设置下准确率的提升幅度往往要比5-shot的设置下大,表明当支持集中每类样本数只有1时,获得的支持原型与真实类别分布差异较大,因此使用相同类别的查询集特征对其进行增强,可以获得较大幅度的性能提升;而5-shot的情况下,支持集中每类样本有5张图片,获得的支持原型特征分布已经比较接近真实类别分布,因此提升的幅度有限。

### 3.3 消融实验

本文提出的特征图增强原型和聚合查询特征通过支持原型增强模块和查询特征聚合模块实现。为了验证对基线模型的这两处改进的有效性,开展消融实验用于测试这两个模块各自对整体方法的贡献以及两个模块之间的相容性,实验在MiniImageNet数据集上5-way 1-shot和5-way 5-shot两个设置下分

别进行。

消融实验结果如表2所示:在仅有SPEM情况下,方法准确率相对于基线模型在1-shot和5-shot设置下分别提升了4.74个百分点和1.93个百分点;在仅有QFAM的情况下,准确率在1-shot和5-shot设置下分别提升了1.11个百分点和0.35个百分点;同时使用两个模块的情况下,准确率在1-shot和5-shot设置下分别提升了5.21个百分点和2.11个百分点。消融实验结果说明,SPEM和QFAM模块都对提升方法的分类准确率有所贡献,并且二者结合比单独使用任意一个模块表现更好,验证了所提出这两个模块的有效性和它们之间的相容性。

表2 不同模块对平均分类准确率的影响  
Table 2 Effect of different modules on average classification accuracy 单位：%

SPEM	QFAM	MiniImageNet	
		5w1s	5w5s
×	×	64.24	81.28
√	×	68.98(+4.74)	83.21(+1.93)
×	√	65.35(+1.11)	81.63(+0.35)
√	√	69.45(+5.21)	83.39(+2.11)

### 3.4 可视化实验

为了分析所提方法产生效果背后的原因,在MiniImageNet数据集5-way 1-shot的设置下开展实验,对一个元任务中所有的聚合查询特征、支持原型和特征图增强原型用T分布随机近邻嵌入(t-distribution stochastic neighbor embedding, t-SNE)降至二维后在平面图上画出,如图6所示。图6中符号“•”代表聚合查询特征;“×”代表使用基线模型得到的支持原

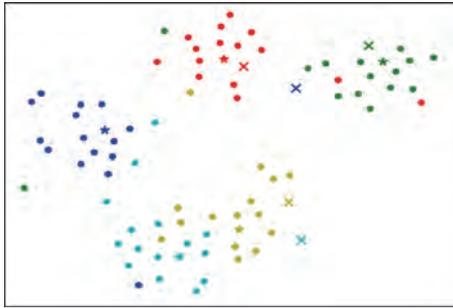


图6 特征与原型的t-SNE可视化

Fig.6 t-SNE visualization of features and prototypes

型;“★”代表特征图增强原型,相同颜色的符号代表属于同一类别。在图6中可以清晰地看到,对于蓝色和青色所代表的类别,其支持原型与查询特征相比,更像是一个个“离群点”,不能有效地代表整个真实类别的分布;而特征图增强原型相比于支持原型,大部分在特征空间上与其对应类别查询特征的聚类中心接近。实验结果表明本文所提出的特征图增强原型相比原始的支持原型在特征空间上更具代表性,有助于提升小样本图像分类方法的性能。

同样在 MiniImageNet 数据集 5-way 1-shot 的设置下使用支持原型与查询特征对应的原始图像进行可视化实验以测试小样本图像分类的效果,实验中根据式(4)计算得到的相似度矩阵  $A$ 。经过上采样并恢复到图像大小后与查询集图像叠加,形成了类似于类别激活图(class activation maps)<sup>[26]</sup>的相似度图。其中的一个样例如图7所示,图中颜色偏红的部分代表查询集样本与支持原型相似的部分。图7中第一

行“Support”表示5个类别的支持集图像,图像上方的文字代表支持集图像对应的标签;第二行“Meta-Baseline”中的图像由1张待分类的查询集图像和该查询集图像分别与5张支持集图像计算得到的相似度图叠加形成(以下简称为叠加图),图像上方的文字表示其真实标签,下方的数字表示其与对应支持集图像的相似度得分,加粗字体表示最高得分,所有得分均由基线模型计算得到;第三行“FMEP”所表示的叠加图和相似度得分由本文所提方法得到。由图7可见:在基线模型中,模型认为查询集图像与支持集图像中“black footed ferret”类别最为相似(相似度得分0.630在所有类别中得分最高),因此将其误分类为“black footed ferret”类别,但其真实类别应为“lion”;而在FMEP中,成功将该张查询集图片正确分类为“lion”。待分类查询集图像的内容是一只狮子正在追捕一头水牛,但由于其对应的正确类别“lion”的支持集图像只有一张,且该支持集图像与查询集图像中的“lion”在大小、形状上都有较大差异,加之查询集图像中存在的干扰物体(水牛)与支持集类别中的“black footed ferret”(黑脚雪貂)视觉上较为相似,导致基线模型对其误分类。

图6和图7所进行的可视化实验说明仅由1张支持集图像得到的支持原型,往往会存在无法在特征空间上有效代表真实类别分布的问题,而在FMEP中提出的特征图增强原型,通过其他正确分类的查询集特征图对支持原型进行有效增强,一定程度上缓解了由于支持集图像数量稀少导致的此类问题,从而提高小样本图像分类的准确率。

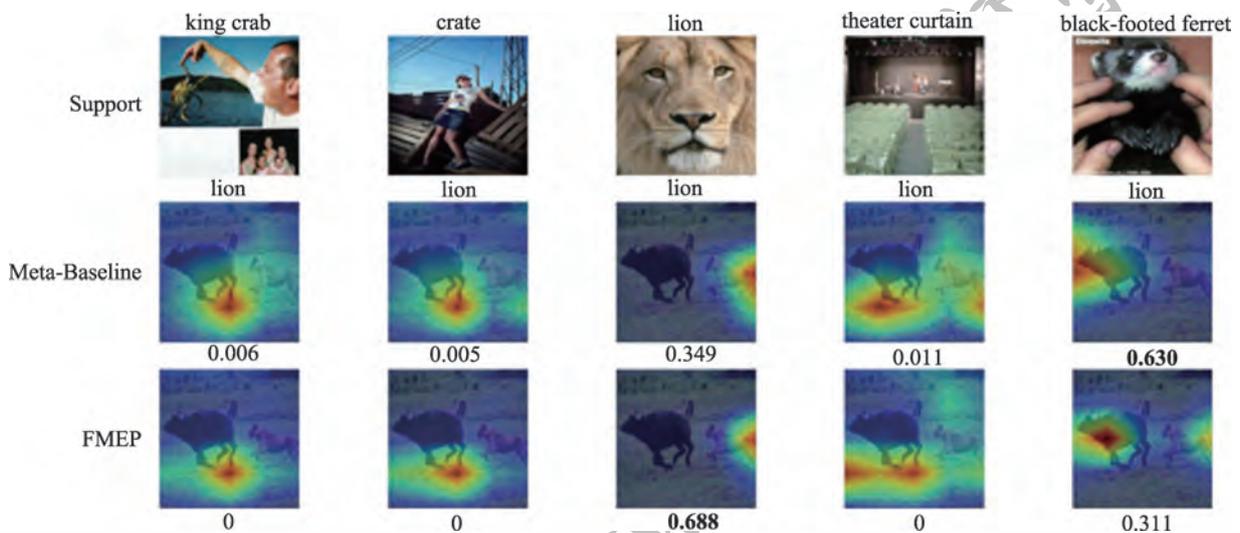


图7 相似度矩阵可视化样例

Fig.7 Example of similarity matrix visualization

## 4 结束语

针对基于度量学习小样本图像分类方法由于支持集样本稀少而导致的支持原型分布难以代表真实类别分布的问题以及由于查询集样本类内差距大导致的部分特征分布偏离类别中心的问题, 本文提出一种采用特征图增强原型的小样本图像分类方法。方法中通过余弦相似度从查询集特征图中筛选出与支持原型最为相似的部分加入到支持原型中, 既可丰富支持原型的特征表达, 又可在一定程度上减少查询集特征图中无关特征的干扰; 此外, 使用多个相似查询集样本特征分布替代单一样本特征的有偏分布, 有效缩短查询集特征分布与真实类别特征分布之间的距离; 最后, 将在特征空间中与真实类别分布都更为接近的特征图增强原型和聚合查询特征进行相似度比较得到最终的分类型结果。在四个常用的小样本图像分类数据集 MiniImageNet、TieredImageNet、CUB-200 和 CIFAR-FS 上的实验结果表明, 本文方法不仅可获得比基线模型更高的分类准确率, 也优于目前主流的小样本图像分类方法。在未来的工作中, 可以考虑探索比余弦相似度更为动态的、精确的特征筛选方法, 进一步提高小样本图像分类的准确率。

## 参考文献:

- [1] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [2] CHEN W Y, LIU Y C, KIRA Z, et al. A closer look at few-shot classification[C]//*Proceedings of the 2018 International Conference on Learning Representations*, Vancouver, Apr 30-May 3, 2018.
- [3] LIU B, CAO Y, LIN Y, et al. Negative margin matters: understanding margin in few-shot classification[C]//LNCS 12349: *Proceedings of the 16th European Conference on Computer Vision*, Glasgow, Aug 23-28, 2020. Cham: Springer, 2020: 438-455.
- [4] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[C]//*Proceedings of the 30th International Conference on Neural Information Processing Systems*, Barcelona, Dec 5-12, 2016: 3637-3645.
- [5] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, Dec 4-9, 2017: 4080-4090.
- [6] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: relation network for few-shot learning[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Jun 18-23, 2018. Piscataway: IEEE, 2018: 1199-1208.
- [7] CHEN Y, LIU Z, XU H, et al. Meta-baseline: exploring simple meta-learning for few-shot learning[C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*, Montreal, Oct 10-17, 2021. Piscataway: IEEE, 2021: 9042-9051.
- [8] 曾武, 毛国君. 多尺度特征聚合的小样本学习方法[J]. *计算机工程与应用*, 2023, 59(15): 151-159.  
ZENG W, MAO G J. Few-shot learning method for multi-scale feature aggregation[J]. *Computer Engineering and Applications*, 2023, 59(15): 151-159.
- [9] KANG D, KWON H, MIN J, et al. Relational embedding for few-shot classification[C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*, Montreal, Oct 10-17, 2021. Piscataway: IEEE, 2021: 8802-8813.
- [10] XIE J, LONG F, LV J, et al. Joint distribution matters: deep Brownian distance covariance for few-shot classification [C]//*Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, Jun 18-22, 2022. Piscataway: IEEE, 2022: 7962-7971.
- [11] LIU Y, ZHANG W, XIANG C, et al. Learning to affiliate: mutual centralized learning for few-shot classification[C]// *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, Jun 18-22, 2022. Piscataway: IEEE, 2022: 14391-14400.
- [12] 刘鑫, 周凯锐, 何玉琳, 等. 基于度量的小样本分类方法研究综述[J]. *模式识别与人工智能*, 2021, 34(10): 909-923.  
LIU X, ZHOU K R, HE Y L, et al. Survey of metric-based few-shot classification[J]. *Pattern Recognition and Artificial Intelligence*, 2021, 34(10): 909-923.
- [13] 安胜彪, 郭昱岐, 白宇, 等. 小样本图像分类研究综述[J]. *计算机科学与探索*, 2023, 17(3): 511-532.  
AN S B, GUO Y Q, BAI Y, et al. Survey of few-shot image classification research[J]. *Journal of Frontiers of Computer Science and Technology*, 2023, 17(3): 511-532.
- [14] REN M, TRIANTAFILLOU E, RAVI S, et al. Meta-learning for semi-supervised few-shot classification[C]//*Proceedings of the 2018 International Conference on Learning Representations*, Vancouver, Apr 30-May 3, 2018.
- [15] HOU R, CHANG H, MA B, et al. Cross attention network for few-shot classification[C]//*Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, Dec 8-14, 2019: 4003-4014.
- [16] LIU J, SONG L, QIN Y. Prototype rectification for few-shot

- learning[C]/LNCS 12349: Proceedings of the 16th European Conference on Computer Vision, Glasgow, Aug 23-28, 2020. Cham: Springer, 2020: 741-756.
- [17] LIU Y, LEE J, PARK M, et al. Learning to propagate labels: transductive propagation network for few-shot learning[C]// Proceedings of the 2019 International Conference on Learning Representations, New Orleans, May 6-9, 2019.
- [18] PAN L, LIU W. Transductive graph-attention network for few-shot classification[C]// Proceedings of the 2022 16th IEEE International Conference on Signal Processing, Beijing, Oct 21, 2022. Piscataway: IEEE, 2022: 190-195.
- [19] QIAO L, SHI Y, LI J, et al. Transductive episodic-wise adaptive metric for few-shot learning[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Oct 27-Nov 2, 2019. Piscataway: IEEE, 2019: 3603-3612.
- [20] GAO Z, WU Y, JIA Y, et al. Curvature generation in curved spaces for few-shot learning[C]// Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, Montreal, Oct 10-17, 2021. Piscataway: IEEE, 2021: 8691-8700.
- [21] VEILLEUX O, BOUDIAF M, PIANTANIDA P, et al. Realistic evaluation of transductive few-shot learning[C]// Advances in Neural Information Processing Systems 34, Dec 6-14, 2021: 9290-9302.
- [22] FAN Q, PEI W, TAI Y W, et al. Self-support few-shot semantic segmentation[C]/LNCS 13679: Proceedings of the 17th European Conference on Computer Vision, Tel Aviv, Oct 23-27, 2022. Cham: Springer, 2022: 701-719.
- [23] BERTINETTO L, HENRIQUES J F, TORR P, et al. Meta-learning with differentiable closed-form solvers[C]// Proceedings of the 2019 International Conference on Learning Representations, New Orleans, May 6-9, 2019.
- [24] 李祥霞, 吉晓慧, 李彬. 细粒度图像分类的深度学习方法[J]. 计算机科学与探索, 2021, 15(10): 1830-1842.  
LI X X, JI X H, LI B. Deep learning method for fine-grained image categorization[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(10): 1830-1842.
- [25] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 770-778.
- [26] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Jun 27-30, 2016. Washington: IEEE Computer Society, 2016: 2921-2929.



**许华杰** (1974—), 男, 广西南宁人, 博士, 副教授, CCF高级会员, 主要研究方向为人工智能、机器视觉。

**XU Huajie**, born in 1974, Ph.D., associate professor, CCF senior member. His research interests include artificial intelligence and machine vision.



**梁书伟** (1996—), 男, 广西玉林人, 硕士研究生, CCF学生会员, 主要研究方向为机器视觉、小样本学习。

**LIANG Shuwei**, born in 1996, M.S. candidate, CCF student member. His research interests include machine vision and few-shot learning.