Few-Shot Classification Model with Fused Dual-Granularity Features

ZHOU Ying, CHEN Yuming*

(College of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China)

Abstract: In few-shot image classification, the number of samples is usually small. Moreover, existing metric learning models cannot well balance local and global information. They are often affected by complex backgrounds and thus cannot focus well on the main subject information. To address these issues, a few-shot classification model named dual-granularity region confusion networks (DGRCN) that integrates region-reorganized images and sub-block features is proposed. The DGRCN model introduces the multi-granularity idea and proposes the region confusion granulation (RCG) method. Through this method, multi-granularity region-reorganized images and multi-granularity sub-block sequences are obtained, which alleviates the problem of insufficient samples. Secondly, picture-level and patch-level features are extracted from the region-reorganized images and sub-block sequences, respectively. An attention mechanism is added to refine the areas of focus and perform multi-granularity feature fusion based on these two types of features. Finally, the classification results of the dual features are adaptively fused, taking into account both global and local key information. DGRCN is experimented on the public datasets miniImageNet, CUB and Stanford Dogs. Compared with the baseline model DN4, the accuracy has been slightly improved. Experimental results show that DGRCN can effectively improve the performance of few-shot classification tasks.

Highlights:

1. The dual-granularity region confusion network (DGRCN) is proposed to adaptively fuse patch-level and picture-level dual-granularity features. Independent distance measurement and weighted classification in dual-granularity feature spaces further align metric results with real distances.
2. A region confusion granulation (RCG) module is proposed to extract patch-level and picture-level features and fully integrate global and local feature correlations.
3. Extensive experiments on miniImageNet, CUB and Stanford Dogs demonstrate the superior accuracy of the proposed method, especially prominent performance improvements on fine-grained datasets.

Key words: few-shot learning; metric learning; data augmentation; feature fusion; granular computing

---

Foundation items: Natural Science Foundation of Xiamen(No.3502Z202473069); Fujian Natural Science Foundation(No.2024J011192); National Natural Science Foundation of China(No.61976183); Fujian Science and Technology Program Project(No.2025E3007).

Received: 2025-03-19; **Revised:** 2025-07-18

***Corresponding author, E-mail**: grcdeep@163.com.

ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 4,Jul. 2026,pp. 1133－1146
DOI:10. 16337/j. 1004-9037. 2026. 04. 015

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

融合双粒度特征的小样本分类模型

周　莹，陈玉明

（厦门理工学院计算机与信息工程学院，厦门 361024）

摘　要：在小样本图像分类中，样本的数量通常较少，且现有的度量学习模型不能够很好地兼顾局部和全局信息，时常受到一些复杂背景的影响，从而不能很好地关注主体信息。针对这些问题，本文提出一种融合区域重组图像及子块特征的小样本分类模型DGRCN（Dual-granularity region confusion networks）。DGRCN算法引入多粒度思想，提出区域重组粒化（Region confusion granulation，RCG）方法，得到多粒度区域重组图像及多粒度子块序列，缓解样本不足问题；其次，分别从区域重组图像及子块序列提取了图像级与局部块级特征，细化注意力以突出关键区域，进行基于两类特征的多粒度特征融合；最后，自适应融合双特征的分类结果，兼顾了全局与局部的重点信息。DGRCN在公开数据集miniImageNet、CUB、Stanford Dogs进行了实验，与基线模型DN4相比准确率均略有提升，实验结果表明DGRCN能有效地提高小样本分类任务的性能。

关键词：小样本学习；度量学习；数据增强；特征融合；粒计算

中图分类号：TP181　　**文献标志码：**A

引用格式：周莹，陈玉明. 融合双粒度特征的小样本分类模型[J]. 数据采集与处理，2026，41(4)：1133-1146. ZHOU Ying，CHEN Yuming. Few-shot classification model with fused dual-granularity features[J]. Journal of Data Acquisition and Processing，2026，41(4)：1133-1146.

引　　言

近年来，深度学习神经网络由于其强大的学习能力和数据表征能力，在许多领域都取得了重大成功。然而，深度学习方法在各种任务上取得优异成绩依赖于大量标记数据及其多次的迭代学习，当数据量过少时，模型会出现效果不佳和过拟合的现象。相比之下，人类具有通过少量数据进行快速学习的能力，为使神经网络拥有这种快速学习的能力，本文提出一种小样本学习[1-2]方法，其能够通过对少量标注样本的训练完成任务。

小样本学习被广泛地运用于许多领域，并迅速发展起来。数据增强[3]、元学习[4]、迁移学习[5]以及度量学习[6]是现有小样本学习常用的4种方法。基于数据增强的方法为解决小样本学习样本不足的问题，或对样本进行变换，或通过生成模型生成与原始数据相似分布的样本。元学习[4]是指学习如何学习的过程，而不是面向单个任务。迁移学习是一种利用以前学到的知识来更快地解决新问题的方法。而度量学习通过特征提取器将样本映射到特征空间，计算样本在特征空间之间的距离。

度量学习[7]被认为是解决小样本分类中一种简单有效的方法。传统的度量学习模型都在图像级特征进行学习，例如孪生网络[8]、匹配网络[9]和原型网络[10]等。图像级别的特征能够提供全局的信息，使其在小样本学习中有效。然而，若只使用一个紧凑的图像级特征，则无法对比图像之间的局部信息。很显然，当样本的数目较少时，使用基于图像级的特征会损失许多有用信息。因此，基于局部描述符的方

基金项目：厦门市自然科学基金（3502Z202473069）；福建省自然科学基金（2024J011192）；国家自然科学基金（61976183）；福建省科技计划项目（2025E3007）。
收稿日期：2025-03-19；**修订日期：**2025-07-18

法因其能有效地关注到局部特征而备受关注。Li等[11-12]提出CovaMNet和DN4(Deep nearest neighbor neural network),皆使用丰富的局部描述符来表示图像。DeepBDC[13]通过深度神经网络来学习数据的特征表示,利用布朗距离协方差(Brownian distance covariance, BDC)来挖掘这些特征之间的内在关系。此外,其他领域也使用了局部信息,充分展示了局部信息的重要性,例如在航空场景分类任务中,MIDC-Net[14]巧妙运用局部语义输入,凭借基于信道空间注意力的多实例池化操作,有效提升分类效果。

度量学习的常规做法是获取完整图像特征,然后衡量特征图中点到点的距离。但该类方法有两类缺陷:一是数据量小,容易过拟合;二是样本差异大,导致泛化能力不强。例如,大部分场景里,每张图像中每个区域的内容不可能完全对应,为此DN4[12]利用局部描述符来计算相似度。局部特征符计算两个特征图中最相似的几个区域,而不是对应位置上的特征相似度。但在某些极端情况下,在真实场景中所识别的绝大部分内容都为背景,而目标主体只占据图像的一小部分,当背景皆为“草地”这类高度相似的信息时,使用该方式仍有可能错过关键信息。

由此可见,基于度量的方法既不能过度注重局部信息,也不能仅仅使用图像的全局信息。针对以上缺陷,能够关联上下文信息的Transformer[15]得到许多研究者的关注。其中,Wang等[16]使用两个独立网络的Transformer来测量全局与局部特征之间的相似性并融合注意力机制,能有效捕获长距离信息,增强非局部信息。Su等[17]通过构建全局-局部特征的层次架构,对模型在粗粒度与细粒度特征表征的能力进行了优化。Sun等[18]构建了局部子网与全局子网两个子网,并蒸馏局部特征与全局特征,从而增强了模型从不同层面特征中提取关键信息并有效整合的性能。

此外,由于其样本数量有限,模型只容易学习到训练样本的特征,泛化性不强。由此,出现了融合多粒度的小样本模型,通过整合粗粒度的整体特征和细粒度的局部特征,模型能够学习到更具一般性和鲁棒性的特征表示。多尺度学习为其中一种方式。Ding等[19]提出了基于多尺度关系网络。Wang等[20]为解决多尺度特征简单拼接、特征贡献程度不同的问题,提出了基于特征融合的多尺度决策网络。曾武等[21]引入了多图聚合的概念,在原始图像上基于多种特征变换,聚合不同拓展图的特征。Huang等[22]则提出了基于多原型的模型,该模型引入了通道压缩和空间激励(Channel squeeze and spatial excitation,sSE)注意力模块生成多原型,从而减少原型在样本上的不确定性。鉴于多粒度表示的有效性,针对样本数量不足问题,本文引入多粒度图像的概念,对图像进行拓展后得到了多粒度图像,实现局部与全局两个层面的特征提取,从根本上扩充了数据的数量,再采用融合策略对多粒度图像的特征进行处理,能多层次、多角度地提取样本的内容,比其他方式更为高效。

综上所述,小样本图像分类任务样本量不足会导致模型难以拟合真实任务场景,样本中主体信息和背景信息重点不突出,从而致使分类不准确。为此,本文提出一种融合双粒度特征的小样本分类模型DGRCN(Dual-granularity region confusion networks)。该模型基于孪生网络,对样本进行多粒度加工后进行特征提取,得到基于局部块级与图像级的双粒度特征,样本在双粒度特征空间各自进行距离度量,加权得到最终分类结果,所得的双粒度融合特征能为模型提供更多的重点内容,以提高模型的感知能力。

1 本文方法

1.1 问题定义

在小样本学习中,原始数据集被划分成训练集D_{train}、验证集D_{val}与测试集D_{test}。由于样本数量较少,为防止过拟合,使用了元学习的训练方式,元学习包含了元训练、元验证和元测试3个阶段。在元训练阶段,每个元任务包含1 000个子任务,这种方式被称为episode训练策略,每个子任务皆由支持集$D_{\text{train, support}}$与查询集$D_{\text{train, query}}$组成,其中有$D_{\text{train, support}} \cup D_{\text{train, query}} = D_{\text{train}}$。如果支持集包含$N$类,每个类中有$K$个样本,则被称为$N$-way K-shot任务。在验证与测试阶段与元训练阶段的数据组成一致。通过大量的episode,模型被不断训练与优化,最终得到性能较好的模型。

1.2　模型概述

融合双粒度区域特征的小样本分类模型以孪生网络为基础架构,图1展示了两种孪生网络的整体结构。其中,图1(a)为传统孪生网络框架,图1(b)为基于双粒度区域特征的孪生网络框架。与传统框架对比,基于双粒度区域特征的孪生网络框架在区域重组粒化模块的基础上将原始图像转变为多粒度图像与子块序列,使模型能够有效提取图像级与区域块级的双粒度特征。模型的整体框架如图2所示,由区域重组粒化(Region confusion granulation, RCG)模块、双粒度特征提取(Dual-granularity feature extraction,DGFE)模块、自适应注意力特征融合(Adaptive attention feature fusion, AAFF)模块,以及双粒度距离度量(Dual-granularity distance metric,DGDM)模块这4个模块组成,其算法主要步骤如下。

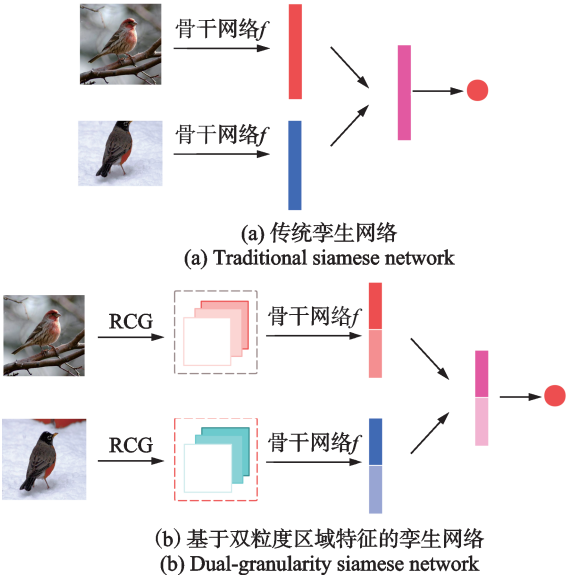


图1　孪生网络框架

Fig.1　Framework of siamese network

(1)区域重组粒化模块将图像切割成若干个区域,得到区域子块序列与区域重组图像,以丰富小样本数据集的多样性。

(2)双粒度特征提取模块使用骨干网络生成多粒度区域子块序列与区域重组图像的特征图,得到基于局部块级的特征与图像级特征。

(3)自适应注意力特征融合模块用来提取重点信息并进行整合以便后续分类。双粒度特征包含更丰富的信息。将不同粒度级别的特征通过线性层映射到同样长度的特征,对多粒度局部级与图像级特征进行注意力区域提取,并加权拼接得到双粒度特征。

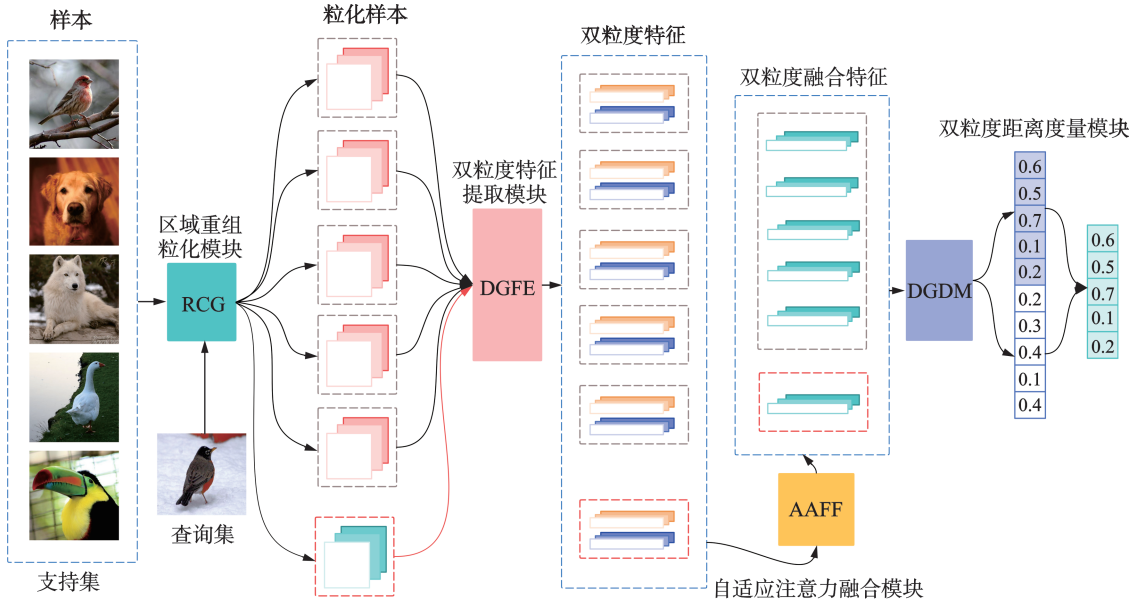


图2　融合双粒度区域重组特征的小样本分类模型

Fig.2　Few shot classification model with fused dual granularity region-reorganized features

(4)双粒度距离度量模型对融合后的双粒度特征分别进行相似度度量,最终使用一个自适应层对两个级别的相似度进行融合得到结果。

1.2.1 区域重组粒化模块

在小样本图像分类任务中,存在着有些图像的全局结构相似而局部细节不同的状况,自然语言处理在面对类似的情况,会将语序打乱,从而迫使神经网络学习更有辨识度的局部信息。为此,区域重组粒化模块将图像划分成多个区域,再随机打乱图像块的位置后再重组,随后得到重组图像与切割后的子块,如图3所示。区域重组粒化模块将图像信息细化,从而提取更有用的特征。

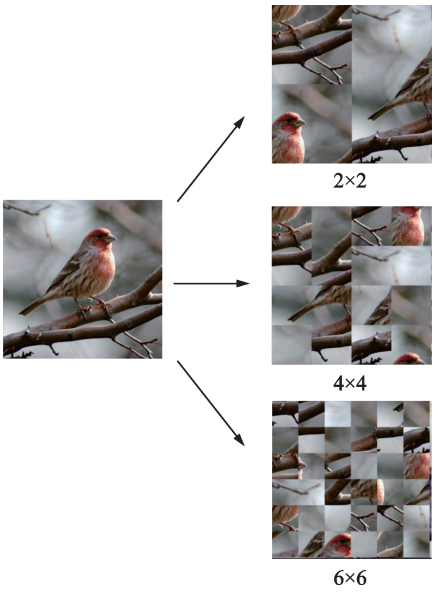


图3 区域重组粒化结果

Fig.3 Results of region confusion granulation

首先,给定一张图像$\boldsymbol{I}$,利用网格分块法均匀地将图像划分成的子区域,表示为$R(i,j)$,i表示为行,j为列($1\leqslant i,j\leqslant n$)。将子区域按序列铺平,也可将每个子区域表示成$R(k)$,$k$满足$1\leqslant k\leqslant n^2$。

通过生成一个长度为n^2的一维向量$\boldsymbol{q}$,$\boldsymbol{q}$中所有元素是唯一的且每个元素都来自一个给定范围的随机序列,该向量中的第k个元素取值r,记作$\boldsymbol{q}_k$,其中k满足$1\leqslant k\leqslant n^2$。$r$值是在区间$(1,n^2)$中等概率选取的一个随机数,概率公式为

$$\boldsymbol{q}_k=r \quad r\sim\mathcal{U}(1,n^2) \quad 1\leqslant r\leqslant n^2 \tag{1}$$

$$p(r,n^2)=\frac{1}{n^2} \quad r=1,2,\cdots,n^2 \tag{2}$$

将所有子区域根据$\boldsymbol{q}$中的随机序列进行重新排序,得到打乱后的子块,即

$$k=\boldsymbol{q}_k,\bar{\boldsymbol{I}}_{n,k}=R(k)=R(\boldsymbol{q}_k) \tag{3}$$

1.2.2 双粒度特征提取模块

多粒度图像能促使模型提取到更加丰富的语义特征,且基于局部图像块与全局图像能够兼顾细节与整体信息。对于每张图像$\boldsymbol{I}$,经过RCG模块的处理,图像$\boldsymbol{I}$被粒化为k张混乱粒度不同的图像,得到多粒度图像与多粒度子块序列。具体而言,多粒度图像可表示为$\{\bar{\boldsymbol{I}}_{n_i}\}_{i=1}^{k}=\{\mathrm{Random}(\boldsymbol{I},n)\}_{i=1}^{k}$,其中$n=\{n_i\}_{i=1}^{k}$是用来表达混乱粒度的粒子,每个粒度的子块序列表示为$\{\boldsymbol{I}_{n_i,j}\}_{j=1}^{n_i^2}$。将多粒度图像与多粒度图像子块送入骨干网络$f$,得到多粒度图像特征$\{f(\bar{\boldsymbol{I}}_{n_i})\}_{i=1}^{k}$与多粒度图像子块特征$\{f(\{\boldsymbol{I}_{n_i,j}\}_{j=1}^{n_i^2})\}_{i=1}^{k}$。每个粒度的图像特征可表示为

$$\boldsymbol{T}_{\mathrm{picture}}^{n_i}=f(\bar{\boldsymbol{I}}_{n_i}),\quad \boldsymbol{T}_{\mathrm{patch}}^{n_i}=f(\{\boldsymbol{I}_{n_i,j}\}_{j=1}^{n_i^2}) \tag{4}$$

双粒度特征的提取与融合过程如图4所示。

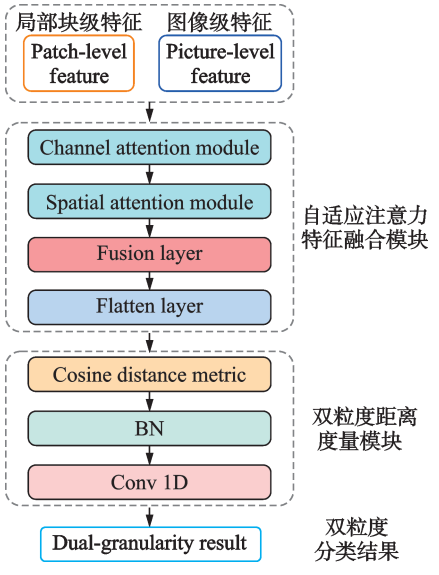


图4 双粒度特征融合与分类结构图

Fig.4 Structure of dual ganularity feature fusion and classification

1.2.3　自适应注意力特征融合模块

考虑到简单拼接多粒度特征图会导致维度变大,影响小样本学习效率,且不同混乱粒度的特征在深度神经网络中的贡献程度不同,从而提出一个自适应注意力特征融合模块。

自适应注意力特征融合模块使用了卷积块注意力模块(Convolutional block attention module,CBAM)来区分特征图上不同信息的重要性,沿着通道与空间两个维度对特征图进行加权,使得主体部分的权重较大,以此增强网络对有效信息部分的识别和利用。具体地,将原始特征进行全局平均池化和最大池化,将所得特征 F 送入多层感知机(Multi-layer perceptron,MLP)后进行相加后,再经Sigmoid函数得到通道注意力权重,与原始特征相乘后得到基于通道注意力的特征。同样地,将特征 F' 进行通道维度平均池化与最大池化,拼接后经卷积层与Sigmoid函数,再与特征 F' 相乘得到最后的特征 F''。将多粒度特征 $T_{\text{picture}}^{n_i}$ 与 $T_{\text{patch}}^{n_i}$ 送入模块,按式(5)和式(6)进行处理,得到的 $A_{\text{picture}}^{n_i}$ 与 $A_{\text{patch}}^{n_i}$ 能够提取到不同粒度图像的关键信息,有

$$F'=\sigma(\text{MLP}(\text{AvgPool}(F))+\text{MLP}(\text{MaxPool}(F)))\otimes F \tag{5}$$

$$F''=\sigma(f([\text{AvgPool}(F'),\text{MaxPool}(F')]))\otimes F' \tag{6}$$

其次,由于不同粒度的划分会产生不同数量的子块,对于拉直后的局部级特征,在各通道上的特征长度长短不一,而各粒度的图像级特征长度相同,需对局部级特征进行调整。具体地,先将 $A_{\text{picture}}^{n_i}$ 与 $A_{\text{patch}}^{n_i}$ 拉直展平成由 M 个 C 维的局部描述符组成特征 $D_{\text{picture}}^{n_i}\in\mathbf{R}^{C\times M_1}$ 和 $D_{\text{patch}}^{n_i}\in\mathbf{R}^{C\times M_2}$,再使用线性层将 $D_{\text{patch}}^{n_i}\in\mathbf{R}^{C\times M_2}$ 映射到 $\mathbf{R}^{C\times M_1}$ 的空间中,此时局部块级特征长度与图像级相同。

最后,自适应学习各混乱粒度层的占比可以有效缓解特征维度过大问题,对于包含更多特征信息的粒度层,将分配到更大的权重。具体地,设每个粒度的特征权重为 $\lambda=\{\lambda_1,\lambda_2,\cdots,\lambda_k\}$,满足所有粒度的权重之和为1,每个粒度的初始权重为

$$\lambda_i=\frac{1/n_i}{\sum_{i=1}^{k}1/n_i} \tag{7}$$

式中 n_i 为多粒度图像对应的混乱粒度,初始的混乱程度越高占比越小。权重参数被放入网络中,随着网络结果回传,不断地自适应调整。设 λ^{picture} 与 λ^{patch} 分别为图像级与局部块级对应的多粒度权重参数,则多粒度融合特征 F_{picture} 和 F_{patch} 为

$$F_{\text{picture}}=\sum_{i=1}^{k}\lambda_i^{\text{picture}}\cdot F_{\text{picture}}^{n_i},\ F_{\text{patch}}=\sum_{i=1}^{k}\lambda_i^{\text{patch}}\cdot F_{\text{patch}}^{n_i} \tag{8}$$

1.2.4　双粒度距离度量模块

多粒度子块特征提取得到局部块级特征,迫使网络关注到一些局部细节,但包含主体信息的子块可能会被切割,多混乱粒度重组图像提取得到图像级特征,但重组图像的子块边缘进行卷积等操作之后可能携带不必要的信息。然而,局部块级特征不包含连接子块边缘的信息,图像级的特征有几率保存主体信息的完整性,因此,为提高类空间的多样性,对双粒度特征进行距离度量后再决策能得到更好的效果。

首先,将支持集特征与查询集特征表示为式(9)所示的局部描述符,再通过式(10)分别计算双粒度特征中 F_{S} 与 F_{Q} 局部描述符的余弦相似度,即

$$F_{\text{Q}}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{M_1}],F_{\text{S}}=[\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_{M_1}] \tag{9}$$

$$\cos(\boldsymbol{x}_i,\boldsymbol{y}_j)=\frac{\boldsymbol{x}_i^{\text{T}}\boldsymbol{y}_j}{\|\boldsymbol{x}_i\|\ \|\boldsymbol{y}_j\|} \tag{10}$$

其次,计算特征相似度。对于 F_{Q} 中的每个局部描述符 $\boldsymbol{x}_i$,在每类 F_{S} 中找到最相似的 m 个局部描述

符,进行累加得到关于$\boldsymbol{x}_i$的相似度。类似地,累加M_1个描述符各自的相似度,利用式(12)得到最终结果。F_{Q}关于F_{S}的相似度公式为

$$\mathrm{MaxDis}(F_{\mathrm{Q}},F_{\mathrm{S}},m)=\sum_{i=1}^{M_1}\sum_{j=1}^{m}\cos(\boldsymbol{x}_i,\boldsymbol{y}_j) \tag{11}$$

$$\mathrm{Dis}(F_{\mathrm{Q}},F_{\mathrm{S}})=\frac{\mathrm{MaxDis}(F_{\mathrm{Q}},F_{\mathrm{S}},m)}{M_1\times m} \tag{12}$$

最后,融合双粒度特征。使用批归一化层,对其进行标准化处理,然后使用可学习的二维权重[w_1, w_2]进行加权融合,在训练过程中权重会根据两类特征的重要程度进行调整。最终,计算出查询集与支持集中每个类之间的相似度,以此对查询集进行分类,表达式为

$$\mathrm{Dis}_{\mathrm{fusion}}(F_{\mathrm{Q}},F_{\mathrm{S}})=w_1\cdot\mathrm{Dis}(F_{\mathrm{Q}}^{\mathrm{picture}},F_{\mathrm{S}}^{\mathrm{picture}})+w_2\cdot\mathrm{Dis}(F_{\mathrm{Q}}^{\mathrm{patch}},F_{\mathrm{S}}^{\mathrm{patch}}) \tag{13}$$

式中:$F_{\mathrm{Q}}^{\mathrm{picture}}$和$F_{\mathrm{S}}^{\mathrm{picture}}$分别为图像级的查询集与支持集特征,$F_{\mathrm{Q}}^{\mathrm{patch}}$和$F_{\mathrm{S}}^{\mathrm{patch}}$为局部块级的查询集与支持集特征,$w_1$和$w_2$分别为图像级特征权重与局部块级特征权重。

2 实验结果及分析

2.1 数据集

本实验选取了常用的小样本数据集miniImageNet,细粒度数据集CUB与Stanford Dogs进行对比实验,数据集的具体情况如表1所示。mini-ImageNet数据集为图像数据集ImageNet的子集,由100个类组成,每个类有600张大小为84像素×84像素的图像。其中64个类用于训练,6个类用于验证,20个类用于测试。CUB是加州理工学院提出来的一个用于分类任务的细粒度图像集,共包含11 788张大小为84像素×84像素的图像,200种类别。CUB的130个类用于训练,20个类用于验证,50个类用于测试。Stanford Dogs包含来自世界各地的120种犬类图像,图像及其标注都来自ImageNet数据集,适用于精细图像分类任务。数据集的70个类用于训练,20个类用于验证,另外30个类用于测试。

表1 数据集描述

Table 1 Descriptions of datasets

数据集	(训练/验证/测试)/类	(训练/验证/测试)/个
mini-ImageNet	64/16/20	38 400/9 600/12 000
CUB	130/20/50	7 615/1 195/2 978
Stanford Dogs	70/20/30	12 165/3 312/5 103

2.2 实验环境

本实验的环境为Ubuntu22.04,64位操作系统,使用Python3.9与Pytorch2.1.0框架,均在显存为40 GB的A100-PCIE显卡上运行。实验使用Conv-64与Resnet-12作为骨干网络,Adam算法作为优化器,交叉熵损失进行梯度更新,初始学习率设为1×10^{-4},采用余弦退火的方式进行学习率的调整。

在分类模块的相似度度量中,选用DN4框架中的余弦相似度度量函数,需找到查询集图像在每个支持集中k个最相似的局部描述符。在元训练阶段,采用了30个epoch,每个epoch构造10 000个episodic,每类使用15个查询图像。在元测试阶段,每个epoch只构造1 000个episodic。选取训练阶段中验证集准确率最高的模型,对其进行5个epoch的测试,取其平均值和95%的置信度作为最后的结果。

2.3 对比实验

为验证所提出的DGRCN的有效性,对比了DGRCN在5-way 1-shot与5-way 5-shot任务与其他经典模型的性能。如表2所示,在5-way 1-shot与5-way 5-shot任务里,DGRCN均取得最优结果。在5-way 1-shot任务上,与经典的基于图像级特征的网络Siamese Net、Matching Nets、Prototypical Nets和Relation Net对比,分别高出了6.79%、11.96%、7.47%和5.48%;而在5-way 5-shot任务中,高出了

10.88%、17.50%、6.28%和7.94%。由此可见,对比基于图像级特征的网络,基于双粒度特征的网络能使准确率有大幅度的提升。而基于局部特征符的CovaMNet、DN4、LMPNet和ADM,相较于其中表现最佳的ADM,在5-way 1-shot与5-way 5-shot任务里也分别高出了0.93%和0.27%。对比元学习方法与基于自监督学习理论的SSL-ProtoNet,DGRCN也能取胜。实验结果表明,基于双粒度的方法比基于图像级或局部特征符的方法更有效,丰富的图像特征信息使得模型效果更佳。

除此之外,DGRCN也应用在小样本细粒度图像数据集中,表3与表4分别展示了CUB与Stanford Dogs数据集的实验结果。通常情况下,细粒度图像分类数据集中的各类别之间差异更小,例如在哈士奇与阿拉斯加之间,轮廓相似,局部特征不同。同一大类中不同小类之间的区分,更需要注重细节。DGRCN在两个细粒度数据集上,仍获得了很好的结果。由表3可知,在5-way 1-shot和5-way 5-shot任务上都展现了极大的优

表2　miniImageNet数据集5 way 1-shot/5-shot对比实验

Table 2　5-way 1-shot and 5-shot comparison on miniImageNet dataset

模型	骨干网络	5-way accuracy/%	
		1-shot	5-shot
Siamese Net[8]	Conv-64	49.23±0.17	61.93±0.18
Matching Net[9]	Conv-64	43.56±0.84	55.31±0.73
ProtoNet[10]	Conv-64	48.45±0.96	66.53±0.51
Relation Net[23]	Conv-64	50.44±0.82	65.32±0.70
CovaMNet[11]	Conv-64	51.19±0.76	67.65±0.63
DN4[12]	Conv-64	51.24±0.74	71.02±0.64
LMPNet[24]	Conv-64	49.87±0.20	68.81±0.16
ADM[25]	Conv-64	54.26±0.63	72.54±0.50
MAML[26]	Conv-64	47.89±1.20	47.89±1.20
MeTAL[27]	Conv-64	52.63±0.37	70.52±0.29
GAP[28]	Conv-64	54.86±0.85	71.55±0.61
CFS-space[29]	Conv-64	55.15±0.84	69.29±0.69
HFFCR[30]	Conv-64	51.79±0.85	65.74±0.72
SSL-ProtoNet[31]	Conv-64	52.58±0.45	70.87±0.36
Ours(DGRCN)	Conv-64	**56.01±0.31**	**72.81±0.17**

表3　CUB数据集5 way 1-shot/5-shot对比实验

Table 3　5-way 1-shot and 5-shot comparison on CUB dataset

模型	骨干网络	5-way accuracy/%	
		1-shot	5-shot
Matching Net[9]	Conv-64	60.06±0.88	74.57±0.73
ProtoNet[10]	Conv-64	50.67±0.88	75.06±0.67
Relation Net[23]	Conv-64	63.94±0.92	77.87±0.64
CovaMNet[11]	Conv-64	52.42±0.76	63.76±0.64
DN4[12]	Conv-64	53.15±0.84	81.90±0.60
LMPNet[24]	Conv-64	53.50±0.24	66.60±0.19
CFRN[32]	Conv-64	64.21±0.72	78.93±0.41
CFS-space[29]	Conv-64	62.73±0.89	76.54±0.62
AGLRs[33]	Conv-64	69.34±0.70	84.72±0.42
Ours(DGRCN)	Conv-64	**69.86±0.33**	**88.92±0.11**
GLFA[34]	Resnet-12	74.01±0.40	87.02±0.27
DeepEMD[35]	Resnet-12	75.65±0.83	88.69±0.50
MIAN[36]	Resnet-12	66.20±0.99	82.30±0.58
KLSANet[37]	Resnet-12	74.94±0.43	88.92±0.41
Ours(DGRCN)	Resnet-12	**76.37±0.31**	**89.52±0.69**

表4　Stanford Dogs数据集5 way 1-shot/5-shot对比实验

Table 4　5-way 1-shot and 5-shot comparison on Stanford Dogs dataset

模型	骨干网络	5-way accuracy/%	
		1-shot	5-shot
Matching Net[9]	Conv-64	46.10±0.86	59.79±0.72
ProtoNet[10]	Conv-64	40.81±0.83	61.58±0.71
Relation Net[23]	Conv-64	47.35±0.88	66.20±0.74
CovaMNet[11]	Conv-64	49.10±0.76	63.04±0.65
DN4[12]	Conv-64	45.73±0.76	66.33±0.62
LMPNet[24]	Conv-64	51.17±0.22	64.06±0.18
AGLRs[33]	Conv-64	58.85±0.69	75.82±0.49
DLG[38]	Conv-64	47.77±0.86	67.07±0.72
RaPSPNet[39]	Conv-64	55.77±0.89	73.58±0.66
Ours(DGRCN)	Conv-64	**59.46±0.39**	**78.18±0.35**
TST_MFL[18]	Resnet-12	67.84±0.82	81.72±0.61
HelixFormer[40]	Resnet-12	65.92±0.49	80.65±0.36
PaCL[41]	Resnet-12	59.76±0.70	77.50±0.48
KLSANet[37]	Resnet-12	64.43±0.81	81.07±0.31
Ours(DGRCN)	Resnet-12	**71.76±0.23**	**84.22±0.42**

势。以Conv-64为骨干网络时，CUB数据集比表现最好的AGLRs高出了0.52%和4.20%，而在Stanford Dogs数据集上，则高出了0.61%与2.32%。对于同样聚焦于局部语义的KLSANet，尤其在Stanford Dogs数据集上有显著的提升，相比于表3与表4中的其他模型，则有大幅度的提升。以Resnet-12为骨干网络时，虽然也在对比中比绝大部分模型效果更好，但是相较于Conv-64来说，准确率的提升较小，这可能是因为双粒度特征所提供的丰富的特征细节较为具体，使用复杂度比较低的骨干网络就能达到比较好的效果。

2.4 消融实验

2.4.1 超参数k与不同粒度组合对实验的影响

为了证明多粒度组合的有效性，在miniImageNet数据集的5-way 1-shot任务上进行消融实验，局部描述符数量k取值5。主要选取0、2、3、4、6这5个混乱粒度进行实验，其中混乱粒度为0表示不对图像进行处理，即原图像。

表5的实验结果表明，混乱粒度为0的组合准确率更高，比不包含粒度为0的组合，几乎都提高了2%以上的准确率。此外，参考[0,2,4]与[0,2,4,6]的结果，以及[2,4,6]与[2,3,4,6]的结果，后者的准确率均高于前者，可知当粒度组合中的绝大部分粒度都相同时，粒度组合个数越多，其准确率也就越高。由此可见，在粒化过程中，图像会被切割打乱，每个粒度的特征图所表达的特征信息各不相同，粒度组合越多，其包含信息也就越多。而打乱图像也会对图像的信息造成一定的破坏，因此当原图像特征加入多粒度融合时，能够有效缓解因为打乱图像而造成的信息损失，分类效果更佳。

表5 不同粒度组合的消融实验结果

Table 5 Results of different granularity combinations

粒度组合	1-shot accuracy/%
[0,2]	54.97
[0,4]	54.31
[0,6]	54.10
[0,2,4]	55.18
[0,2,4,6]	**55.19**
[0,3,6]	54.54
[2,4,6]	51.73
[2-4,6]	52.39

在余弦相似度度量中，对于每张查询图像，需在支持集中找到最相似的k个局部描述符进行相似度计算。如何选取k值对查询样本与支持样本之间的相似度计算十分关键。为此，本文选取4个k值，在粒度组合数目n为2和3的条件下进行实验，表6给出了在其条件下miniImageNet的实验结果。由表6可知，k值的选取对实验结果的影响大概在2%以内，对于粒度组合为2的情况，k取1时效果最好，说明此时多粒度特征所包含的信息较少，选取较大的k值会有不必要的信息加入相似度计算；粒度组合取3时，k取5时效果最佳，此时粒度信息较多时需要选用较大的k值才能达到比较好的效果。总体而言，粒度数目n与k的选值并非越大越好，过小的k值可能导致性能不够稳定，过大k值可能会引入不必要的信息，因此应合理选择n与k值的搭配。

表6 超参数k的选取对实验的影响

Table 6 Effect of hyperparameter k on experiment %

DGRCN	k=1	k=3	k=5	k=7
[0,2](n=2)	**56.01**	55.26	54.97	54.68
[0,2,4](n=3)	53.99	54.00	55.18	54.12

2.4.2 不同注意力模块对实验结果的影响

为了评估注意力模块对DGRCN的影响，本文进行了基于注意力机制的消融实验，注意力机制作为一个即插即用的模块，被添加在基础特征提取器之后。本文实验均在同粒度组合、双粒度特征的基础下进行miniImageNet的5-way 1-shot任务，结果如表7所示。可以看到，加入CBAM模块的表现最优，比无注意力时提高了0.20%，而加入自注意力模块反而比仅使用原始特征的效果更差，这可能是因为在

RCG模块中,原始的空间信息已被打乱,而自注意力机制更擅长捕捉图像的局部像素与全局像素之间的依赖关系,混乱的空间信息打乱了这种依赖关系的提取。而CBAM更专注于增强模型的特征提取能力,能够加强重要的特征信息,从而能够忽略乱序所带来的噪声。

表7　不同注意力机制的消融实验结果

Table 7　Results of different attention mechanisms

No-attention	Self-attention	CBAM	Accuracy/%
√	—	—	54.98
—	√	—	51.76
—	—	√	**55.18**

2.4.3　不同特征模块对实验结果的影响

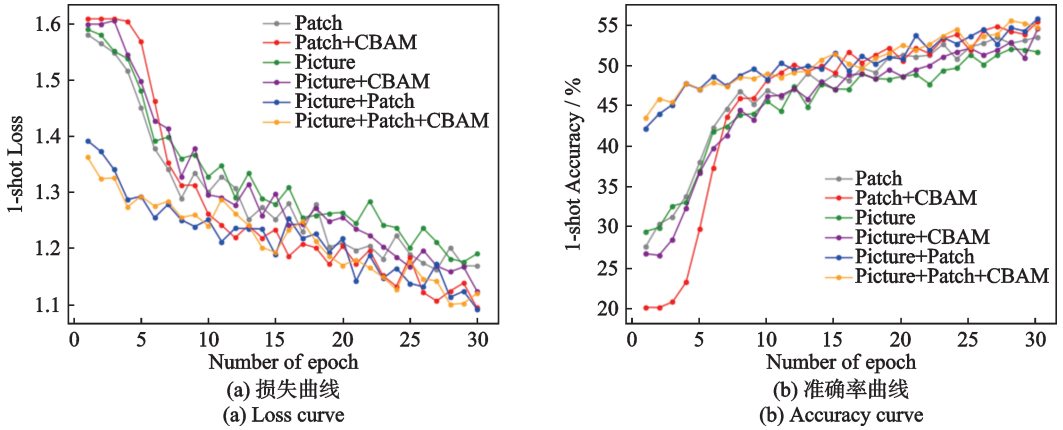
在表7对注意力机制的消融实验中,CBAM的效果最好。为了探讨图像级特征和局部块级特征的影响,实验均使用同样的粒度组合进行双粒度特征融合。由表8可知,仅使用图像级特征的结果皆低于局部块级特征的结果,而基于双粒度特征的准确率最高。同时,尽管基于图像级特征的模型在效果上不如局部块级特征的模型,但图像级特征的结果仍保持在一个相对较好的水平,说明打乱空间信息不会对最终的分类效果产生不利影响,而将图像切分成子块后再进行特征提取,能使原本有可能被网络忽略的一些细节信息被发现,从而达到更好的效果。

表8　不同特征选择的消融实验结果

Table 8　Results of different feature selections

模块	注意力	Accuracy/%
Picture	—	52.05
Picture	+CBAM	53.81
Patch	—	52.86
Patch	+CBAM	54.92
Picture+Patch	—	54.98
Picture+Patch	+CBAM	**55.18**

图5为几种特征提取方式在30个epoch测试阶段的Loss与Accuracy的波动。由图5可知,基于双粒度特征的2种模型的Loss均低于其余4个模型,最终都能够以较小的损失值收敛,在每个epoch中,Accuracy也略高于其余4个模型。在0～5轮次的epoch中,基于双粒度特征的准确率在最开始就能够达到40%～50%,且模型的Loss趋于稳定。而基于局部块级与图像级的准确率在0～5轮次的epoch中快速增长,在5～10轮次才达到比较稳定的高准确率。由对比图可得知,基于双粒度特征的模型有收敛快和稳定性好的特点。综上所述,双粒度特征结合了两者的优势,使网络在忽略整体关联性的同时,能够专注于寻找更有利的局部区域。此外,它还能减轻通过打乱拼接所产生的粒度图像的拼接边缘对整体效果的负面影响。双粒度特征也使模型能够更快地收敛,并迅速稳定在较高水准的性能。



(a) 损失曲线　(a) Loss curve　　(b) 准确率曲线　(b) Accuracy curve

图5　不同特征的损失曲线与准确率曲线对比图

Fig.5　Loss curve and accuracy curve comparison chart of different features

2.5 复杂度分析

为验证DGRCN的运行效率，在相同的代码框架与硬件环境下，基于miniImageNet数据集在5-way 1-shot设置下，开展了多粒度与双分支度量机制对整体训练开销的对比实验。所有模型均采用Conv-64作为骨干网络，比较了参数量、FLOPs等模型复杂度指标，结果如表9所示。由表9可知，虽然DGRCN对比经典模型参数量与FLOPs相较于经典模型DN4与ProtoNet有一定的提升，但DGRCN在$n=2$的情况实现了56.01%的准确率，对比DN4与ProtoNet有明显提升，其运行时间仅增加至0.11 s，仍在可接受范围内。

表9 不同模型的复杂度参数对比

Table 9 Comparison of complexity parameters between different models

模型	参数/10^6	FlOPs/10^9	运行内存/MB	运行时间/s	Accuracy/%
DN4	0.11	8.89	2 702	0.07	51.24
ProtoNet	0.11	8.98	2 676	0.06	48.85
DGRCN($n=2$)	0.29	27.61	3 936	0.11	**56.01**
DGRCN($n=3$)	0.29	45.75	5 908	0.19	55.18

2.6 可视化分析

图6展示了miniImageNet数据集在5-way 1-shot任务和5-way 5-shot任务的t-SNE可视化。每行的左中右子图分别展示了原始分布、基于局部块级特征与基于图像级特征在空间的分布情况。由图6可见，样本在经过特征提取之前，在空间的分布比较散乱，在经过特征提取之后，每个类别在空间的分布比较集中。在度量学习中，同类别的样本在空间中的距离较近，使得样本在分类阶段能被更好地区分，DGRCN的双粒度特征方法在两个特征空间上进行度量，模型从不同的角度学习数据的特征，避免了少量异常样本的偶然性，从而使模型更加精准可靠。

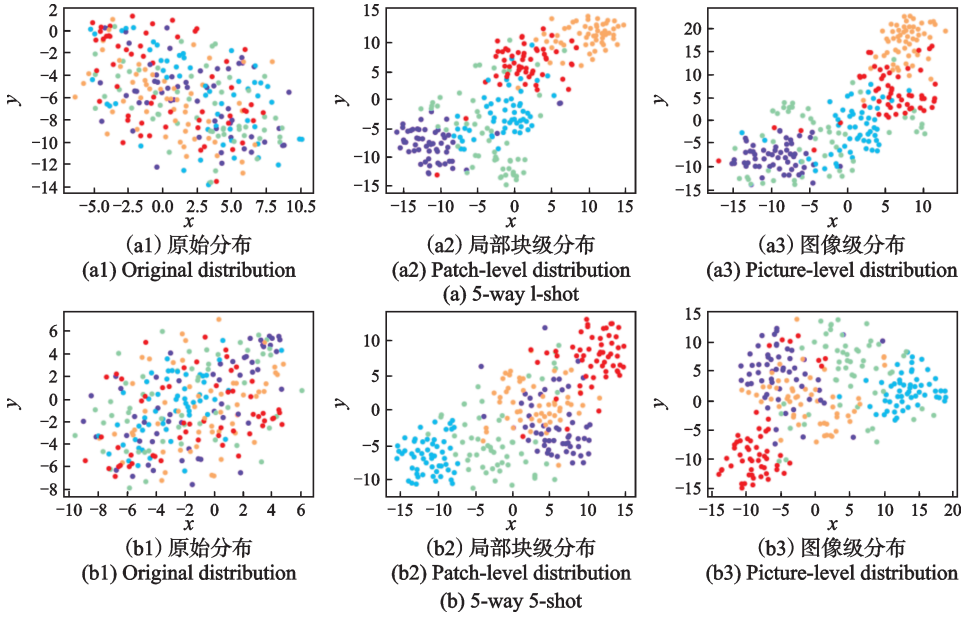
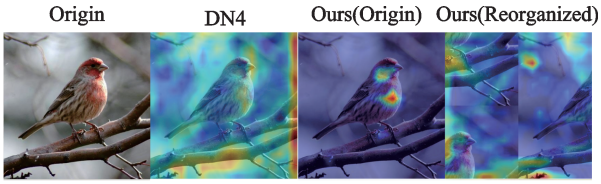


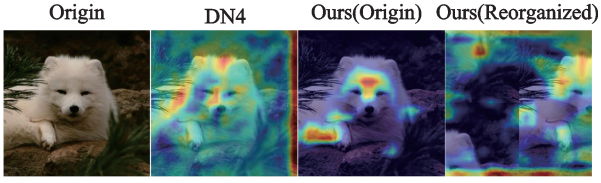
图6 miniImageNet数据集的t-SNE可视化

Fig.6 t-SNE visualization of the miniImageNet dataset

图7展示了Grad-CAM的可视化对比图。图7的第1列显示了原始图像,第2列是DN4模型的热力图与图像的叠加图,最后两列是来自于DGRCN中混乱粒度为2时图像的热力图。由可视化图像可知,与传统的DN4模型相比,该模型更侧重于图像的主要特征,而忽略了图像中大部分背景。由此可知,DGRCN提取模型有效信息的能力更强。



(a) 鸟
(a) Bird



(b) 狐狸
(b) Fox

图7　Grad-CAM的可视化对比图

Fig.7　Visual comparison charts of Grad-CAM

3　结束语

针对现有小样本分类模型方法存在数据量不足且特征提取不够完全的缺陷,本文提出了一种融合双粒度特征的小样本分类模型DGRCN。该模型使用区域重组粒化模块,使样本不足的问题得到了缓解,同时提出了一种双粒度特征的提取方式,充分突出了特征中需要被重点关注的部分,为不同粒度的信息分配了不同的权重,从而提取出有效信息。实验表明,DGRCN方法具有竞争力,在经典数据集和细粒度数据集上均有不错的表现。但本文模型仍有需要改进的空间,例如,区域重组粒化模块打乱了图像的空间位置,后续可以考虑加入子块的位置信息,或是通过一个能够携带子块原始位置信息的矩阵,还原空间位置。此外,由于DGRCN提取的特征包含比较多的信息,也可借助粒计算理论对双粒度特征进行决策以提高效率。

参 考 文 献:

[1]　赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述[J]. 软件学报, 2021, 32(2): 349-369.
ZHAO Kailin, JIN Xiaolong, WANG Yuanzhuo. Survey on few-shot learning[J]. Journal of Software, 2021, 32(2): 349-369.

[2]　LAKE B, SALAKHUTDINOV R, TENENBAUM J. Human-level concept learning through probabilistic program induction [J]. Science, 2015, 350(6266): 1332-1338.

[3]　SCHWARTZ E, KARLINSKY L, SHTOK J, et al. Delta-encoder: An effective sample synthesis method for few-shot object recognition[C]//Proceedings of the 31st Conference on Neural Information Processing Systems. Montreal, Canada: Curran Associates, Inc., 2018: 2850-2860.

[4]　李凡长, 刘洋, 吴鹏翔, 等. 元学习研究综述[J]. 计算机学报, 2021, 44(2): 422-446.
LI Fanchang, LIU Yang, WU Pengxiang, et al. A survey on recent advances in meta-learning[J]. Chinese Journal of Computers, 2021, 44(2): 422-446.

[5]　SUN Q, LIU Y, CHUA T, et al. Meta-transfer learning for few-shot learning[C]//Proceedings of the IEEE/CVF Conference

on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Press, 2019: 403-412.

[6] XING E, JORDAN M, RUSSELL S, et al. Distance metric learning with application to clustering with side-information[J]. Advances in Neural Information Processing Systems, 2002, 15: 505-512.

[7] 吕佳, 曾梦瑶, 董保森. 双路径合作的原型矫正小样本分类模型[J]. 计算机科学与探索, 2024, 18(3): 693-706.
LYU Jia, ZENG Mengyao, DONG Baosen. Prototype rectification few-shot classification model with dual-path cooperation[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(3): 693-706.

[8] KOCH G, ZEMEL R, SALAKHUTDINOV R. Siamese neural networks for one-shot image recognition[C]//Proceedings of the 32nd International Conference on Machine Learning. Paris, France: JMLR, 2015: 1-30.

[9] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[C]//Proceedings of the 29th Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates, Inc., 2016: 3630-3638.

[10] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[C]//Proceedings of the 30th Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates, Inc., 2017: 4077-4087.

[11] LI W, XU J, HUO J, et al. Distribution consistency based covariance metric networks for few-shot learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI Press, 2019, 33(1): 8642-8649.

[12] LI W, WANG L, XU J, et al. Revisiting local descriptor based image-to-class measure for few-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Press, 2019: 7260-7268.

[13] XIE J, LONG F, LV J, et al. Joint distribution matters: Deep Brownian distance covariance for few-shot classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE Press, 2022: 7972-7981.

[14] BI Q, QIN K, ZHANG H, et al. Local semantic enhanced convnet for aerial scene recognition[J]. IEEE Transactions on Image Processing, 2021, 30: 6498-6511.

[15] VASWANI A, SHAZHIE S, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 30th Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates, Inc., 2017: 5998-6008.

[16] WANG X, WANG X, JIANG B, et al. Few-shot learning meets transformer: Unified query-support transformers for few-shot classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(12): 7789-7802.

[17] SU Y, ZHAO H, ZHENG Y, et al. Few-shot learning with multi-granularity knowledge fusion and decision-making[J]. IEEE Transactions on Big Data, 2024, 10(4): 486-497.

[18] SUN Z, ZHENG W, GUO P, et al. TST_MFL: Two-stage training based metric fusion learning for few-shot image classification[J]. Information Fusion, 2025, 113: 102611.

[19] DING Y, TIAN X, YIN L, et al. Multi-scale relation network for few-shot learning based on meta-learning[C]//Proceedings of the International Conference on Computer Vision Systems. San Francisco, USA: IEEE Press, 2019: 343-352.

[20] WANG X, MA B, YU Z, et al. Multi-scale decision network with feature fusion and weighting for few-shot learning[J]. IEEE Access, 2020, 8: 92172-92181.

[21] 曾武, 毛国君. 基于多图特征聚合的小样本学习方法[J]. 计算机科学, 2023, 50(6A): 220400029-10.
ZENG Wu, MAO Guojun. Few-shot learning method based on multi-graph feature aggregation[J]. Computer Science, 2023, 50(6A): 220400029-10.

[22] HUANG H, WU Z, LI W, et al. Local descriptor-based multi-prototype network for few-shot learning[J]. Pattern Recognition, 2021, 116: 107935.

[23] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: Relation network for few-shot learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Press, 2018: 1199-1208.

[24] HUANG H, WU Z, LI W, et al. Local descriptor-based multi-prototype network for few-shot learning[J]. Pattern Recognition, 2021, 116: 107935.

[25] LI W, WANG L, HUO J, et al. Asymmetric distribution measure for few-shot learning[C]//Proceedings of the 29th International Joint Conference on Artificial Intelligence. Montreal, Canada: IJCAI Organization, 2021: 2957-2963.

[26] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//Proceedings of the

34th International Conference on Machine Learning. Sydney, Australia: JMLR, 2017: 1126-1135.

[27] BAIK S, CHOI J, KIM H, et al. Meta-learning with task-adaptive loss function for few-shot learning[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE Press, 2021: 9465-9474.

[28] KANG S, HWANG D, EO M, et al. Meta-learning with a geometry-adaptive preconditioner[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE Press, 2023: 16080-16090.

[29] SONG B, ZHU H, WANG B, et al. Class feature sub-space for few-shot classification[J]. Applied Intelligence, 2024, 54(19): 9177-9194.

[30] JIA X, MAO Y, PAN Z, et al. Few-shot learning based on hierarchical feature fusion via relation networks[J]. International Journal of Approximate Reasoning, 2024, 170: 109186.

[31] LIM J, LIM K, LEE C, et al. SSL-ProtoNet: Self-supervised learning prototypical networks for few-shot learning[J]. Expert Systems with Applications, 2024, 238: 122173.

[32] SONG B, ZHU H, BI Y. A conditioned feature reconstruction network for few-shot classification[J]. Applied Intelligence, 2024, 54(8): 6592-6605.

[33] ABDELAZIZ M, ZHANG Z. Learn to aggregate global and local representations for few-shot learning[J]. Multimedia Tools and Applications, 2023, 82(21): 32991-33014.

[34] SHI B, LI W, HUO J, et al. Global-and local-aware feature augmentation with semantic orthogonality for few-shot image classification[J]. Pattern Recognition, 2023, 142: 109702.

[35] ZHANG C, CAI Y, LIN G, et al. Deepemd: Differentiable earth mover′s distance for few-shot learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5632-5648.

[36] QIN Z, WANG H, MAWULI C, et al. Multi-instance attention network for few-shot learning[J]. Information Sciences, 2022, 611: 464-475.

[37] SUN Z, WANG Z, GUO P. KLSANet: Key local semantic alignment network for few-shot image classification[J]. Neural Networks, 2024, 178: 106456.

[38] CAO S, WANG W, ZHANG J, et al. A few-shot fine-grained image classification method leveraging global and local structures[J]. International Journal of Machine Learning and Cybernetics, 2022, 13(8): 2273-2281.

[39] ZHANG W, ZHAO Y, GAO Y, et al. Re-abstraction and perturbing support pair network for few-shot fine-grained image classification[J]. Pattern Recognition, 2023, 148: 110158.

[40] ZHANG B, YUAN J, LI B, et al. Learning cross-image object semantic relation in transformer for few-shot fine-grained image classification[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal: ACM Press, 2022: 2135-2144.

[41] WANG C, FU H, MA H. PaCL: Part-level contrastive learning for fine-grained few-shot image classification[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal: ACM Press, 2022: 6416-6424.

作者简介:

周　莹(2001-),女,硕士研究生,研究方向:粒计算、小样本学习等。

陈玉明(1977-),**通信作者**,男,博士,教授,博士生导师,研究方向:机器学习、粗糙集以及粒计算等,E-mail:grcdeep@163.com。

(编辑:刘彦东)