



多尺度特征细化的小样本图像分类

吕伏, 张旭, 张紫扬

引用本文:

吕伏, 张旭, 张紫扬. 多尺度特征细化的小样本图像分类[J]. *智能系统学报*, 2026, 21(3): 675-687.

LYU Fu, ZHANG Xu, ZHANG Ziyang. Multi-scale feature refinement for few-shot image classification[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(3): 675-687.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202505026>

您可能感兴趣的其他文章

基于孪生变分自编码器的小样本图像分类方法

A small-sample image classification method based on a Siamese variational auto-encoder
智能系统学报. 2021, 16(2): 254-262 <https://dx.doi.org/10.11992/tis.201906022>

多感知兴趣区域特征融合的图像识别方法

Image recognition method based on multi-perceptual interest region feature fusion
智能系统学报. 2021, 16(2): 263-270 <https://dx.doi.org/10.11992/tis.201906032>

基于级联宽度学习的多模态材质识别

Cascade broad learning for multi-modal material recognition
智能系统学报. 2020, 15(4): 787-794 <https://dx.doi.org/10.11992/tis.201908021>

基于宽度学习方法的多模态信息融合

Multi-modal information fusion based on broad learning method
智能系统学报. 2019, 14(1): 150-157 <https://dx.doi.org/10.11992/tis.201803022>

多标记学习自编码网络无监督维数约简

Unsupervised dimensionality reduction of multi-label learning via autoencoder networks
智能系统学报. 2018, 13(5): 808-817 <https://dx.doi.org/10.11992/tis.201804051>

基于自编码器的特征迁移算法

Feature transfer algorithm based on an auto-encoder
智能系统学报. 2017, 12(6): 894-898 <https://dx.doi.org/10.11992/tis.201706037>

DOI: 10.11992/tis.202505026

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20260310.1342.002>

多尺度特征细化的小样本图像分类

吕伏^{1,2}, 张旭¹, 张紫扬¹

(1. 辽宁工程技术大学软件学院, 辽宁葫芦岛 125105; 2. 辽宁工程技术大学基础教学部, 辽宁葫芦岛 125105)

摘要: 在小样本图像分类任务中, 特征融合虽有助于提升有限样本的利用效率, 但易引入冗余信息, 影响模型判别能力。为此, 提出一种多尺度特征融合细化网络 (multi-scale feature fusion refinement network, MFRNet), 通过系统性的冗余抑制机制优化特征表达。构建双流注意力融合模块, 结合跨层级特征交互与自适应全局注意力, 实现空间与通道维度的协同融合, 有效抑制初始冗余。引入可分离通道注意力机制, 利用深度可分离卷积解耦空间滤波与通道交互, 进一步压缩冗余信息, 提升特征判别性。设计自适应特征细化网络, 通过双级压缩-释放机制与异构化特征分解, 实现冗余的层级化过滤与判别性特征的保留。在 5 个不同粒度数据集上的实验表明, MFRNet 在两种实验设置下均优于现有主流方法, 展现出显著性能优势。

关键词: 小样本学习; 度量学习; 特征融合; 特征细化; 图像分类; 原型网络; 多尺度特征; 深度学习

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2026)03-0675-13

中文引用格式: 吕伏, 张旭, 张紫扬. 多尺度特征细化的小样本图像分类 [J]. 智能系统学报, 2026, 21(3): 675-687.

英文引用格式: LYU Fu, ZHANG Xu, ZHANG Ziyang. Multi-scale feature refinement for few-shot image classification [J]. CAAI transactions on intelligent systems, 2026, 21(3): 675-687.

Multi-scale feature refinement for few-shot image classification

LYU Fu^{1,2}, ZHANG Xu¹, ZHANG Ziyang¹

(1. School of Software, Liaoning Technical University, Huludao 125105, China; 2. Department of Basic Teaching, Liaoning Technical University, Huludao 125105, China)

Abstract: In few-shot image classification tasks, feature fusion improves the utilization of limited samples but can readily introduce redundant information and degrade the model's discriminative ability. To address this, this paper proposes a multi-scale feature fusion refinement network (MFRNet) that optimizes feature representation through a systematic redundancy suppression mechanism. First, a dual-stream attention fusion module is constructed, combining cross-level feature interaction with adaptive global attention to achieve collaborative fusion across spatial and channel dimensions and effectively suppress initial redundancy. Second, a separable channel attention mechanism is introduced, using deep separable convolutions to decouple spatial filtering from channel interaction, further compressing redundant information and improving feature discriminability. Finally, an adaptive feature refinement network is designed to achieve hierarchical filtering of redundancy and retention of discriminative features through a two-stage compression-release mechanism and heterogeneous feature decomposition. Experiments on five datasets of varying granularity show that MFRNet outperforms existing mainstream methods in both experimental settings, demonstrating significant performance advantages.

Keywords: few-shot learning; metric learning; feature fusion; feature refinement; image classification; prototypical networks; multi-scale feature; deep learning

深度学习在数据充沛的监督学习范式下已实现接近人类水平的图像分类性能, 但高度的数据依赖性导致其在样本稀缺的场景中面临严峻挑

战。参数空间与数据量的失配易引发过拟合, 并且有限的监督信号难以支撑可靠泛化边界的构建。相较之下, 人类展现的小样本认知能力揭示了先验知识引导的归纳机制优势。小样本学习 (few-shot learning, FSL) 通过知识迁移与元学习框架建立了从有限样本到普适规律的映射机制, 现

收稿日期: 2025-05-26. 网络出版日期: 2026-03-10.

基金项目: 国家自然科学基金面上项目 (51874166, 52274206);
国家自然科学基金青年基金项目 (51904144).

通信作者: 吕伏. E-mail: 38458786@qq.com.

已成为提升模型样本特征利用效率的主要研究方向之一^[1-3]。

小样本学习的目标是构建具备任务泛化能力的模型,使其通过少量样本即可快速适应新任务^[4]。该领域面临的核心问题是如何从高维稀疏数据中解耦出任务不变性特征,以突破经验风险最小化框架的样本依赖瓶颈。

元学习(meta-learning)通过模拟任务分布来构建双层的优化范式,在任务层面学习可迁移的归纳偏置,在参数层面解耦任务不可知与任务特定组件。其核心思想是让模型在训练过程中积累跨任务的通用知识,从而在面对新任务时仅需少量样本或微调即可高效完成学习。与传统机器学习直接拟合数据不同,元学习以任务为训练单元。通过“学习如何学习”的机制,将先验知识的概念引入小样本学习当中,并成为指导方法。经典的元学习方法主要包含 2 类:直接建立输入和预测值之间映射的方法,如记忆增强神经网络^[5]和 SNAIL(simple neural attentive learner)^[6]以及优化初始化参数的方法,如 MAML(model-agnostic meta-learning)^[7]和 Reptile^[8]。

小样本学习中的先验知识主要分为 3 种类型:初始模型参数、优化策略和度量空间。根据先验知识的不同,元学习方法可以分成 3 类:基于模型的方法、基于优化的方法和基于度量的方法^[4]。目前技术演进呈现明显的度量导向趋势。基于度量学习的框架通过支持集-查询集样本的语义对齐,在保持线性时间复杂度的同时实现端到端可微分类。该类别的代表方法有:匹配网络^[9](matching network)通过比较不同的特征提取器提取查询集和支持集样本特征的余弦相似度来完成分类;原型网络^[10](prototypical networks, PN)引入原型概念来表示支持集样本,通过比较查询集样本和原型类间的欧氏距离来实现图像分类;关系网络^[11](relation networks)通过对特征提取和非线性度量矩阵同时进行学习,将学习到的特征向量串联起来并利用关系模块计算得到分类结果。

目前基于度量学习的方法多使用全局特征来表示整个图像,并且假设在一个可见类样本上训练得到的全局特征仍然可以适用于另一个不可见类。例如, GNN(graph neural network)将图神经网络引入小样本学习中学习一个全局特征的度量^[12]。相较于全局特征表示,低层次的局部表示更加适合在不同的类中共享,更容易迁移到其他任务中。

局部描述符^[13](local description, LD)是指经由 CNN(convolutional neural network)处理后得到的三维特征图的每个空间位置特征。使用局部描

述符的经典方法有 DN4(deep nearest neighbor neural network)^[14]和 DMN4(discriminative mutual nearest neighbor neural network)^[15]。它们将全局表示替换成局部描述符,并使用 k 近邻显式地利用局部描述符。还有一些方法,通过融合相邻局部描述符增强局部特征表达,来提高小样本分类效果,如邻域融合特征增强网络^[16]和 MKGRE(multi-scale knowledge-guided region enhancement method)^[17]。此外,还有一些方法,同时采用图像全局特征和局部描述符计算查询样本到类的相似度,如 TAPNet(time series classification model with attentional prototype network)^[18]、TALDSNet(task-aware adaptive local descriptors selection network)^[19]、FFSNet(feature fusion similarity network)^[20]、SAFFNet(self-attention-based feature fusion network)^[21]和 CAN(cross attention network)^[22]等。

然而,图像全局与局部描述的融合特征通常包含大量冗余信息。这种冗余不仅增加了计算复杂度,还可能导致模型在有限的样本下过拟合,降低其分类性能。冗余信息的存在使得模型难以有效区分关键特征,从而影响其在不可见类上的泛化能力。因此,如何减少冗余并提取更具判别性的局部特征是提升小样本分类效果的关键挑战之一。

在图像重建模型中的特征细化方法能够有效地剔除特征中的冗余信息,提升图像质量。受图像重建方法的启发,将特征细化模块引入融合模块之后,提出了一种多尺度特征融合细化小样本图像分类网络(multi-scale feature fusion refinement network, MFRNet)。其主要贡献为:1)基于内容引导注意力融合方法,构建双流注意力融合模块(dual-stream attention fusion, DSAF),在融合不同尺度全局与局部细节特征的同时降低冗余信息;2)设计可分离通道注意力(separable channel attention, SCA),解耦空间滤波与通道交互,进一步抑制空间冗余;3)构建自适应特征细化网络(adaptive feature refinement network, AFRN),实现双级压缩-释放机制,完成冗余信息的层级化过滤与判别性特征保留。

1 小样本分类问题和原型网络

在小样本图像分类任务中,通常采用 N-way、K-shot 的基本设置^[21]。在小样本学习过程中,所有数据集均划分为训练集、验证集和测试集,且这 3 个数据集的类别无重叠。小样本分类包括一个训练阶段,在一个较大的数据集上学习模型;

一个适应阶段, 在有限的标记样本下, 将学习到的模型适应于先前未知的任务^[23]。

在度量学习众多方法中, 原型网络通过原型向量的解耦设计实现动态扩展能力。新增类别时仅需基于少量样本计算原型向量, 无需调整特征提取网络参数, 支持增量式学习和零样本场景下的类别扩展, 避免传统模型全参数微调的计算冗余, 具有较大的可扩展性。通过引入不同的注意力机制, 可以增强原型的表达, 如混合注意力原型网络^[24]和类注意力原型网络^[25]。

并且其通过支持集样本特征均值计算原型, 使单样本异常值对整体原型的干扰随样本量增加呈指数衰减。这使其可以引入注意力加权原型或基于密度的鲁棒聚合, 进一步抑制噪声样本影响。

故本文使用原型网络的方法进行分类, 即使使用均值计算每个类别的支持集对应的类原型中心:

$$P_{s_i} = \frac{\sum_{j=1}^K F(I_j)}{K}, P_{s_i} \in \mathbf{R}^{C \times H \times W} \quad (1)$$

式中: $F(\cdot)$ 表示特征提取网络的计算过程; I 表示支持集的某个类别, 根据 N-way K-shot 设, I 有 N 个类别, 每个类别有 K 个数据; I_j 表示属于第 I 类的第 j 个数据。第 I 支撑集中的类原型中心可以由方程得到, 其他支撑集类原型中心的求解方法与第 I 类类似。

2 多尺度特征融合细化网络

针对小样本图像分类中特征融合引发的冗余干扰与判别性退化问题, 本文提出多尺度特征融合细化网络 (MFRNet), 其由双流注意力融合模块 (DSAF)、可分离通道注意力 (SCA)、自适应特征细化网络 (AFRN) 与相似度度量模块 4 部分构成, 如图 1 所示, 通过系统性冗余抑制策略实现特征表达的优化。

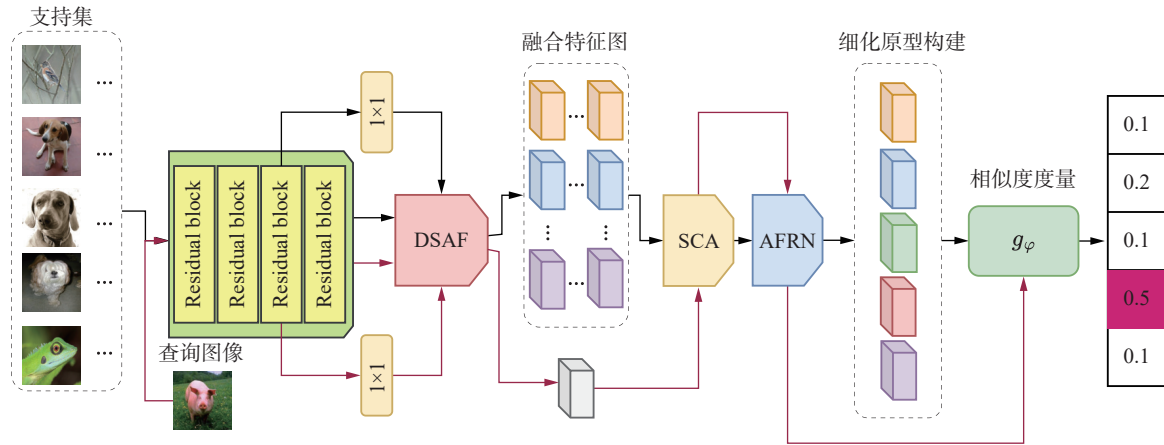


图 1 多尺度特征融合细化网络整体框架

Fig. 1 Overall framework of multi-scale feature fusion and refinement network

MFRNet 采用三阶段级联优化架构, 核心逻辑在于对冗余信息的层级递进式抑制与判别性特征的逐步精细化提取, 各模块协同工作且形成“粗筛-细滤-精炼”的完整处理流水线。前一模块的输出直接作为后一模块的输入, 处理目标从宏观到微观、从粗略到精细逐层递进。

DSAF 模块作为前端融合模块, 采用双流注意力融合机制, 将嵌入网络提取的支持集与查询集多尺度特征输入自适应全局注意力模块, 通过动态分配空间-通道维度权重, 在融合全局上下文与局部细节特征的同时初步降低冗余信息; 但其输出仍包含大量空间维度与通道维度的噪声。

SCA 模块承担承上启下的作用, 接收 DSAF 输出的融合特征后, 通过深度可分离卷积解耦空间滤波与通道交互, 以轻量化架构专注于抑制空间与通道维度上的广泛冗余; 同时对 DSAF 未能完全消除的分散噪声进行精细化处理, 为后续特

征精加工奠定基础。

AFRN 模块作为后端精炼模块, 结合部分卷积 (PConv) 与层级压缩-释放机制, 对经 SCA 净化后的特征开展深层判别性挖掘; 通过异构化特征分解、动态通道-空间注意力校准及双级机制, 最终实现对顽固性冗余的层级化过滤与核心判别特征的精准保留及增强。

2.1 双流注意力融合模块

在小样本图像分类的训练与测试过程中, 受限于情节训练阶段支持样本数量的不足, 类别判别能力受到限制。为提升分类准确性, 亟须在有限样本条件下提取更具代表性的特征信息, 以增强类别的语义表征能力。融合特征能够反映图像在不同尺度下的多层次语义信息, 提升语义表达能力。然而, 传统特征融合方法普遍缺乏跨层次、跨维度的动态交互机制, 难以构建鲁棒且具有区分性的多尺度语义表示。

为解决上述问题,本文在内容引导注意力融合模块^[26]的基础上提出了双流注意力融合模块(DSAF)。需要特别说明的是,此处所述“去除跳跃连接”仅针对融合模块中的跨层特征相加操作,目的是抑制浅层冗余信息直接干扰高层语义特征表达。而主干特征提取网络(ResNet-12)中的残差连接结构未做任何修改,仍完整保留,以保障深层网络训练的稳定性与特征流动性。传统特征融合方法通过跳跃连接传递多尺度特征,虽能缓解梯度消失问题,但在小样本场景中,浅层特征的噪声和冗余信息会通过跳跃连接直接干扰高层语义特征的表达。并且跳跃连接在数据充足时能加速收敛,但在小样本条件下,浅层特征的局部噪声与高层特征的语义目标可能因直接相加而产生特征冲突。本文提出的 DSAF 模块通过移除跳跃连接,切断低层次噪声的传播路径,迫使模型通过自适应全局注意力(AGA)动态筛选有价值的跨层次特征,以构建高效跨层次特征交互机制。

在注意力机制方面,针对小样本分类场景,剔除输入特征与注意力输出间的残差路径,并在通道注意力中引入全局最大池化和全局平均池化的串联操作,通过多尺度统计特征聚合增强通道表征,避免冗余信息对权重分配的干扰,构建自适应全局注意力(AGA)跨维度动态分配权重。

其中空间注意力基于全局上下文信息定位关键区域,量化特征图各位置的空间显著性;通道注意力则通过跨空间聚合提取通道级统计特征,挖掘高响应通道的语义价值。该机制通过动态加权融合策略,既保留局部区域的细粒度信息,又强化全局语义的类间区分性,最终构建兼顾判别力与泛化能力的多尺度特征表示。

双流注意力融合方法(DSAF)的整体框架如图2所示。DSAf 模块首先对多尺度特征进行叠加融合,并将其输入 AGA 生成通道与空间注意力权重 W_s 和 W_c 。AGA 具体步骤如图3,首先为了得到空间权重 W_s 和通道权重 W_c , AGA 将输入的多尺度特征输入到并行的空间注意力和通道注意力中。空间注意力通过并行的全局平均池化和全局最大池化获取输入特征的空间信息,再通过一个 7×7 卷积得到空间权重。而通道注意力通过串行的一个全局最大池化和全局平均池化获取特征的通道信息,并且为了减少参数数量和限制模型复杂度采用了两个连续的 1×1 卷积进行处理,即第 1 个 1×1 卷积将通道维度从 C 降低到 C/r (r 是指减速比),第 2 个 1×1 卷积将其扩展回 C 。令 $X \in \mathbf{R}, C \times H \times W$ 表示正在进行的输入特征, W_s 和 W_c 可以表示为

$$W_c = C_{1 \times 1}(\max(0, C_{1 \times 1}(\text{GAP}(X_{\text{GMP}}^c)))) \quad (2)$$

$$W_s = C_{7 \times 7}([X_{\text{GAP}}^s, X_{\text{GMP}}^s]) \quad (3)$$

式中: $\max(0, x)$ 表示 ReLU 激活函数, $C_{k \times k}(\cdot)$ 表示核大小是 $k \times k$ 的卷积层, $[\cdot, \cdot]$ 表示通道级联操作, $\text{GAP}(\cdot)$ 表示全局平均池化操作。 X_{GMP}^c 、 X_{GAP}^s 和 X_{GMP}^s 分别表示经过跨通道维度的全局最大池化操作处理的特征、跨空间维度的全局平均池化操作处理的特征和跨空间维度全局最大池化操作处理的特征。

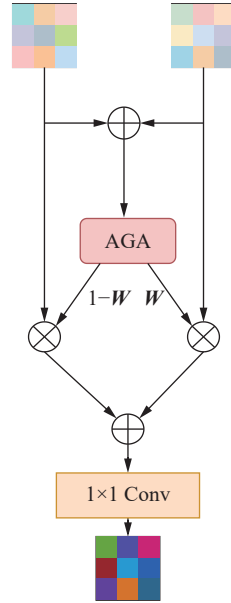


图 2 双流注意力特征融合模块
Fig. 2 Dual-stream attention features fusion module

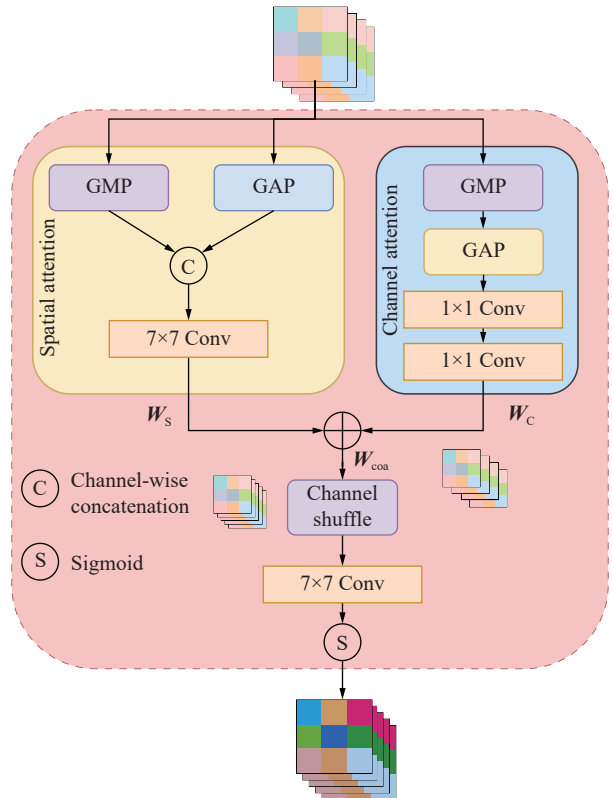


图 3 自适应全局注意力
Fig. 3 Adaptive global attention

然后通过一个矩阵加计算将 W_C 和 W_S 进行融合, 公式为

$$W_{coa} = W_C + W_S \quad (4)$$

在 AGA 的最后我们利用通道洗牌和一组 7×7 卷积得到最终的精化 W :

$$W = \sigma(GC_{7 \times 7}(CS(W_{coa}))) \quad (5)$$

式中: σ 表示 Sigmoid 操作, $CS(\cdot)$ 表示通道 shuffle 操作, $GC_{k \times k}(\cdot)$ 表示一个卷积核大小为 $k \times k$ 的组卷积层。

AGA 引导模型关注每个通道的显著区域, 可以使模型在数据有限的情况下依然能够学到丰富且稳健的特征表示^[26]。

在获得 AGA 生成的权重后, 基于该权重对原始特征进行跨维度加权融合, 重构特征表达, 进一步强化关键区域的语义表征能力。最后融合特征经由 1×1 卷积进一步抽象, 生成具有更强判别能力的多尺度融合表示 F_{fuse} 。这一过程可表示为

$$F_{fuse} = C_{1 \times 1}(F_1 \cdot W + F_2 \cdot (1 - W)) \quad (6)$$

式中: F_1 和 F_2 表示经由特征提取模块得到支持集和查询集不同层次的特征。

经双流注意力融合模块处理后, 可获取多尺度的全局-局部融合特征。该特征既能够保障语义信息的有效表达, 又可初步降低部分冗余信息。

2.2 可分离通道注意力

尽管 DSAF 模块已实现多尺度特征的初步融合与部分冗余的抑制。但在特征融合过程中, 浅层特征和深层特征、全局特征和局部特征的交叉融合, 易导致低语义噪声对高层特征表达产生干扰。在此背景下, 该模块输出的融合图像特征中, 仍存在大量分布的分散空间维度冗余和通道维度冗余。

为进一步精细化处理 DSAF 模块输出特征中的冗余信息, 本文引入可分离通道注意力机制 (SCA)。该机制作为连接 DSAF 与后续 AFRN 模块的关键桥梁承上启下。其设计目标并非局限于简单的特征增强, 而是通过对融合特征进行空间维度与通道维度上的解耦操作实现特征净化, 为 AFRN 模块的深度特征细化提供更“干净”的输入。传统的通道注意力机制 (如 SENet^[27]) 主要通过全连接结构建模通道间的依赖关系, 但忽略了空间维度上的冗余处理。针对该问题, SCA 引入深度可分离卷积, 将通道注意力过程解耦为两个独立阶段, 从而分别实现空间滤波与通道融合。

首先利用深度可分离卷积对每个通道独立进行空间滤波, 降低通道内部的空间噪声干扰; 随

后采用 1×1 卷积建模通道间的依赖关系, 从中削弱冗余通道的影响, 突出具有判别性的特征通道。通过从通道内部噪声抑制与通道间冗余关联削弱两个维度协同优化, SCA 显著提升了注意力机制的表示能力。

此外, SCA 采用深度可分离卷积与 1×1 卷积的级联结构, 实现对通道维度的动态响应调节, 在保持特征判别性的同时, 有效控制参数规模与计算开销, 进一步提升融合特征的有效性与模型整体的计算效率。如图 4 所示。

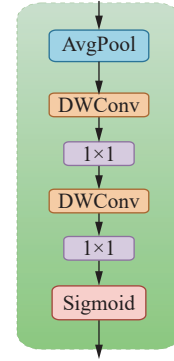


图 4 可分离通道注意力
Fig. 4 Separable channel attention

SCA 首先利用平均池化压缩空间维度信息, 生成通道级特征。可以表示为

$$F_{pool} = P_{avg}(F_{fuse}) \quad (7)$$

其中 P_{avg} 表示平均池化操作。

其次, 通过两个连续的深度可分离卷积建模通道间非线性依赖关系公式为

$$F_w = C_{1 \times 1}(DWConv_{1 \times k_b}(F_{pool})) \quad (8)$$

$$F_h = C_{1 \times 1}(DWConv_{k_b \times 1}(F_w)) \quad (9)$$

最后, 基于 Sigmoid 函数生成注意力掩码 A , 实现特征重标定。可以表示为

$$A = \text{Sigmoid}(\text{Conv}_{1 \times 1}(F_h)) \quad (10)$$

$$F_{attn} = (A \odot F_{fuse}) \oplus F_{fuse} \quad (11)$$

式中: Sigmoid 函数保证了注意力掩码 A 在值域 $(0, 1)$ 内, \odot 表示元素乘法, \oplus 表示元素求和, F_{attn} 是增强特征。

2.3 自适应特征细化网络

尽管 SCA 模块借助解耦机制有效抑制了空间与通道域中的广泛性冗余信息, 但融合特征中仍残留部分深层且复杂的冗余模式, 同时还存在着与任务关联度较低的次要特征。

鉴于特征细化前馈网络在抑制深层冗余、增强判别性特征表达方面的显著优势, 本文以该网络为基础, 引入部分卷积和深度可分离卷积, 进一步设计出自适应特征细化网络 (AFRN)。AFRN

接收经由 SCA 净化后的特征,其核心目标不再是处理广泛分布的噪声,而是进行更深层次的“精加工”。通过双级压缩-释放机制与异构分支处理,该网络可实现冗余信息的层级化过滤与判别性特征的最终筛选与强化,进而完成特征优化的最后环节。

AFRN 通过双级压缩-释放机制实现冗余信息的层级化过滤,并以多阶段渐进式处理策略促进特征表达的优化。首先采用部分卷积^[28]对通道维度进行粗粒度过滤,从中选择性激活关键特征子集,实现初步的信息筛选。随后引入双分支差异化处理路径,其中语义分支侧重于全局语义信息的保持,而细节分支则聚焦于局部结构特征的提取,从而构建空间与通道维度协同作用的细粒度过滤机制。最终,通过特征重组模块实现跨维度的信息交互与融合,完成从局部纹理到全局语义的逐层特征提炼,相较于单级过滤策略,能够更精准地区分有效特征与冗余噪声。

在压缩阶段,AFRN 采用部分卷积与深度可分离卷积对浅层特征进行初步压缩,有效抑制无效信息向深层网络的传递。其中,部分卷积仅作用于输入特征的部分通道,从机制上引导模型关注关键通道信息;而深度可分离卷积则进一步压缩空间维度上的冗余信息。在释放阶段,模块通过残差连接机制回引压缩过程中获得的判别性特征,有效避免了重要信息的丢失,保障特征表达的完整性与判别性。如图 5 所示。

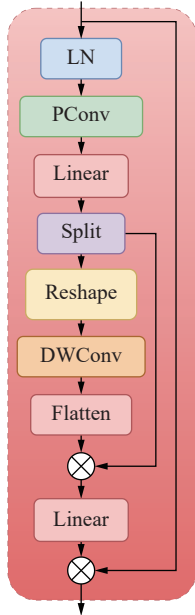


图 5 自适应特征细化网络

Fig. 5 Adaptive features refinement network

AFRN 首先通过一个层归一化、部分卷积和线性层,然后通过一个 GELU 激活函数将来自

SCA 的特征初步处理,公式为

$$\mathbf{F}'_{\text{attn}} = \text{GELU}(\mathbf{W}_1 \text{PConv}(\text{LN}(\mathbf{F}_{\text{attn}}))) \quad (12)$$

式中: \mathbf{W}_1 为经过线性层产生的线性投影, $\text{LN}(\cdot)$ 表示层归一化, $\text{PConv}(\cdot)$ 表示部分卷积。

然后经过基于通道的 Split 分片将初步处理后的特征分成语义增强和细节增强两个分支。语义增强分支专注于提取与类别相关的全局语义信息;细节增强分支聚焦于保留具有判别性的局部细节特征,公式为

$$[\mathbf{F}'_{\text{attn1}}, \mathbf{F}'_{\text{attn2}}] = \mathbf{F}'_{\text{attn}} \quad (13)$$

对语义增强分支 $\mathbf{F}'_{\text{attn1}}$ 保留了经层归一化后的平滑语义,通过线性层增强类别相关性;细节增强分支 $\mathbf{F}'_{\text{attn2}}$ 通过 Reshape 和深度可分离卷积提取局部细节。两个分支的特征通过残差连接进行融合,既保留了输入的完整性,又增强了关键特征的表达,公式为

$$\mathbf{F}'_{\text{attnr}} = \mathbf{F}'_{\text{attn1}} \otimes \text{F}(\text{DWConv}(\text{R}(\mathbf{F}'_{\text{attn2}}))) \quad (14)$$

式中: $\text{R}(\cdot)$ 和 $\text{F}(\cdot)$ 分别为 Reshape 和 Flatten 操作; $\text{DWConv}(\cdot)$ 为深度可分离卷积操作。

最后,通过线性变换与 GELU 激活函数对特征进行非线性映射,并采用残差连接机制,将处理结果与初始输入进行逐元素相乘,生成细化后的特征表示,从而增强关键信息的表征能力并抑制冗余干扰,公式为

$$\mathbf{F}'_{\text{out}} = \mathbf{F}'_{\text{attn}} \otimes \text{GELU}(\mathbf{W}_2 \mathbf{F}'_{\text{attnr}}) \quad (15)$$

式中 \mathbf{W}_2 为经由第二个线性层产生的线性投影。

通过以上操作,AFRN 能够在全局和局部层面筛选特征,让模型专注于分类相关的部分。并且通过提取那些具有代表性的特征来增强特征表示,在精简冗余信息的同时实现信息流动。

3 实验与分析

3.1 数据集

本文在 5 个小样本图像分类数据集上对所提出的方法进行了评估,分别为 TieredImageNet、MiniImageNet、CUB-200-2011、Stanford Dogs 和 Stanford Cars。TieredImageNet^[15] 为 ILSVRC-2012 的子集,亦为流的小样本分类数据集,包含 34 个高层类别及共计 779 195 张图像,具备更强的跨类别泛化性。MiniImageNet^[9] 同样源自 ILSVRC-2012,包含 100 个类别及 60 000 张图像,广泛用于评估模型在小样本设定下的分类性能。CUB-200-2011 是小样本分类任务中广泛使用的基准数据集,包含 200 个鸟类子类别,共计 11 788 张图像^[29]。Stanford Dogs^[30] 包含 120 个犬种,共约 20 580 张

图像, 背景复杂、姿态多变, 适用于评估模型在复杂条件下的辨别能力。Stanford Cars^[31] 提供 196 个具体车型 (含品牌和年份) 共 16 185 张图像, 适合研究细粒度的车辆识别。

3.2 实验设置

本文在 5-way 1-shot 和 5-way 5-shot 小样本分类设置下对所提出的 MFRNet 方法进行了性能评估。实验在基于 PyTorch 框架的 RTX3090 GPU 远程服务器上进行。为确保与现有方法的公平对比, 实验选用了小样本分类任务中最常使用的两种主干网络结构: ResNet-12 和 Conv-4。训练阶段将所有图像统一调整为固定尺寸, 以适配模型输入要求。

训练设置: 训练集、验证集与测试集均采用一致的参数配置。所有实验均采用随机梯度下降 (SGD) 优化算法对模型参数进行更新, 初始学习率设定为 0.1。第 50 次迭代后, 学习率每隔

10 次迭代衰减为原值的一半; 动量系数设置为 0.9, 权重衰减系数为 0.000 5, 其他超参数均采用默认设置。

评估设置: 在与现有小样本图像分类方法进行对比时, 从测试集中随机采样任务, 采用与训练阶段一致的 5-way 1-shot 和 5-way 5-shot 实验设置进行评估。模型在每次评估中运行 5 个 epoch, 每个 epoch 含 100 个独立任务, 并以其平均分类准确率作为最终评价指标。为进一步验证所提模型的性能, 并系统比较与其他少样本分类方法的性能差异, 本文设计了对比实验与消融实验, 以全面展示模型的有效性与关键模块的贡献。

3.3 对比实验

为了评估 MFRNet 性能, 本小节首先和其余 7 个小样本图像分类方法在两个粗粒度数据集 MiniImageNet 和 TieredImageNet 上进行比较, 结果如表 1 所示。

表 1 粗粒度图像对比实验结果
Table 1 Coarse-grained image comparison experimental results

方法	Conv-4				ResNet-12			
	MiniImageNet		TieredImageNet		MiniImageNet		TieredImageNet	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MatchingNet	44.32	54.51	45.56	55.12	63.08	75.99	68.53	80.61
PN	50.72	67.32	53.31	71.33	62.67	77.88	68.48	83.46
RelationNet	49.62	64.42	53.55	70.52	65.91	82.41	71.16	86.03
DeepEMD	51.52	64.71	51.22	65.83	65.91	82.41	71.16	86.03
DN4	50.52	70.38	52.63	72.63	65.35	81.10	69.62	83.41
DMN4	55.73	74.19	57.06	74.15	67.15	84.20	71.86	85.76
TALDS-Net	56.78	74.63	57.54	75.79	67.08	83.82	71.02	85.23
QSFormer	55.82	75.16	58.12	76.31	65.44	80.06	72.74	85.66
bdc+var+invar	57.28	74.89	57.09	75.72	68.61	84.71	73.37	88.26
PatchProto+tSF	57.25	73.37	59.75	74.58	69.27	84.41	72.48	85.14
MFRNet(本文)	57.56	76.33	58.42	77.25	70.49	88.96	75.42	89.85

注: 加粗数据表示性能最优。

这些对比的方法包含: 基于全局特征表示度量学习的 PN^[10]、MatchingNet^[9]、RelationNet^[11]; 基于局部特征表示度量学习的 DeepEMD^[32]、DMN4^[15] 和 DN4^[14]; 基于特征优化的 TALDS-Net^[19]; 基于 Transformer 的方法 QSFormer^[33] 和 PatchProto+tSF^[34]; 基于自监督方法 bdc+var+invar^[35]。

从表 1 可以看出, 本文提出的 MFRNet 方法在 MiniImageNet 和 TieredImageNet 两个小样本分类数据集上, 在不同的特征提取结构 (Conv-4 和 ResNet-12) 以及 1-shot 和 5-shot 设置下均显著优于现有对比方法。在所有设置的平均性能上,

MFRNet 分别比当前最优的对比方法平均提升了 0.56%(1-shot) 和 1.99%(5-shot), 相较于次优方法的提升幅度分别达到 1.29% 和 3.17%, 体现了该方法的广泛适应性和稳定性能提升。

与典型的基于全局特征表示的度量学习方法 Prototypical Networks(PN) 进行对比, MFRNet 在两个数据集的 1-shot 和 5-shot 设置上平均分别提升了 8.33% 和 8.45%, 这充分表明, 相较于仅使用全局特征的方式, 引入局部特征描述对于提升小样本分类性能更为关键。

本文方法也显著优于当前表现较优的融合型

方法,如 TALDS-Net。与 TALDS-Net 相比, MFR-Net 在两个数据集上 1-shot 和 5-shot 的平均准确率分别提升了 1.64% 和 2.58%。尤其在 ResNet-12 提取器下的 MiniImageNet 设置中, MFRNet 达到 70.49%(1-shot) 和 88.96%(5-shot), 相比 TALDS-Net 的 67.08% 和 83.82%, 分别提升 3.41% 和 5.14%, 表现出更强的特征表达能力。值得注意的是, 尽管 TALDS-Net 已引入局部特征处理策略, MFR-Net 在此基础上进一步通过特征剥离与细化模块强化了语义区分能力, 有效提升了小样本下的类别可分性。

与基于 Transformer 的方法 QSFormer 和 PatchProto+tSF 进行对比时, MFRNet 同样表现出了优势。MFRNet 在 MiniImageNet 和 TieredImageNet 数据集上的 1-shot 和 5-shot 准确率分别比 QSFormer 平均提高了 6.55% 和 3.87%。虽然以 Conv-4 为提取网络时相较于 PatchProto+tSF 方法在 TieredImageNets 数据集的 1-shot 设置上降低了 1.33%, 但在其他设置中均有所提升。这表明, 尽管 Transformer 结构擅长建模全局依赖关系, 但 MFRNet 通过引入局部特征描述和细化模块, 能更好地处理复杂的局部特征, 尤其是在数据量少

的情况下。

与基于自监督学习的方法, 如 bdc+var+invar 进行对比时, MFRNet 同样展现了其性能优势。Bdc+var+invar 在 5-shot 设置下的 TieredImageNet 上达到了 88.26%, 但 MFRNet 在 MiniImageNet 和 TieredImageNet 数据集上的 1-shot 和 5-shot 的准确率分别提升了 1.97% 到 2.92%。这些结果表明, 尽管自监督方法通过无监督学习生成有效的特征表示, MFRNet 通过多尺度特征细化和注意力机制, 能够在少样本情况下进一步提升特征的判别性和语义表征能力, 从而提供了更优的分类性能。

综上, 本文提出的 MFRNet 在粗粒度数据集和设置下均展现出显著的性能优势, 验证了融合全局与局部信息并对特征进行细化处理对于提升小样本分类性能的重要性与有效性。

本文还在 3 个细粒度图像数据集 Stanford Dogs、Stanford Cars 和 CUB-200-2011 上进行对比实验, 分析模型的性能。实验结果如表 2 所示。对比的方法包含: PN、DN4、TALDS-Net、BiFRN^[36]、C2-Net^[37]、LDCA^[38]、DLDA^[39]、BSFA^[40]、bdc+var+invar、QSFormer、ViTFSL-Baseline^[41] 和 HelixFormer^[42] 方法。

表 2 细粒度图像对比实验结果
Table 2 Fine-grained image comparison experimental results

方法	Embedded	Stanford Dogs		Stanford Cars		CUB-200-2011	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
PN		72.25	86.17	72.23	86.98	71.98	86.64
DN4		45.41	63.51	59.84	88.65	64.02	82.97
TALDS-Net		—	—	—	—	88.32	94.90
BiFRN		72.54	85.86	88.48	96.34	74.36	88.64
C2-Net	ResNet12	75.50	87.65	88.96	95.16	—	—
BSFA		—	—	88.93	95.2	82.27	90.76
QSFormer		71.22	82.48	78.63	88.32	75.74	86.52
bdc+var+invar		73.49	84.83	81.24	90.07	84.41	94.37
HelixFormer		65.41	80.92	78.97	92.11	79.34	91.25
MFRNet(本文)		76.54	90.52	89.34	95.32	89.34	95.52
ViTFSL-Baseline	Transformer	74.24	85.48	80.83	89.69	76.50	89.95
MatchingNet		35.82	47.53	34.81	44.74	45.32	59.56
PN		65.59	82.19	64.90	82.93	64.42	81.82
DN4		45.41	63.51	59.84	88.65	46.84	74.92
TALDS-Net		—	—	—	—	81.73	93.24
DLDA		49.44	69.35	60.86	89.5	55.12	74.46
LDCA	Conv-4	55.72	80.76	56.8	91.91	75.45	91.63
BiFRN		61.39	78.86	76.22	90.66	—	—
C2-Net		66.42	81.23	81.29	91.08	78.66	89.43
BSFA		—	—	49.98	67.52	68.16	82.41
HelixFormer		59.81	73.76	75.48	89.93	78.52	90.87
MFRNet(本文)		67.05	81.53	80.33	92.62	80.32	94.32

注: 加粗数据表示性能最优。

从表 2 中实验结果可以看出, 本文提出的 MFRNet 方法在多个细粒度分类数据集 (Stanford Dogs、Stanford Cars、CUB-200-2011) 上均取得了最优性能, 全面超越现有基于全局特征 (如 PN)、局部特征 (如 DN4、TALDS-Net)、融合特征 (如 Bi-FRN、C2-Net)、基于 Transformer 的方法 (如 QS-Former、HelixFormer 和 ViTFSL-Baseline) 以及基于自监督学习的方法 (如 bdc+var+invar)。在使用 ResNet-12 作为特征提取网络时, MFRNet 在所有数据集的 1-shot 和 5-shot 设置中均实现当前最佳性能, 尤其在 CUB 数据集上的提升最为显著, 分别比 BSFA 提高 7.07% 和 4.76%, 体现出其在细粒度视觉分类任务中的强大判别能力。而在轻量级的 Conv-4 网络结构下, MFRNet 在 5-shot 设置中依然表现突出, 验证了其良好的结构适应性与鲁棒性。相比使用 Transformer 等复杂结构的优化方法 (如 BSFA、QSFormer、HelixFormer 和 ViTFSL-Baseline), MFRNet 在采用更传统主干网络的条件下仍取得领先, 说明所提出的特征剥离与细化机制能更有效地提取判别信息, 并通过全局与局部信息融合进一步增强了模型的泛化能力。与基于自监督学习的方法 bdc+var+invar 进行对比时, MFRNet 同样表现出色。bdc+var+invar 在 CUB-200-2011 上的 5-shot 准确率为 94.37%, 而 MFRNet 在相同设置下达到了 95.52%, 超出了 bdc+var+invar 的表现。这表明, MFRNet 在通过多尺度特征细化和自适应注意力机制提高特征表示的判别性方面, 比自监督学习的方法更具优势。

综上, 本文方法在不同网络结构、多个数据集以及多种设置下均表现出一致性和先进性, 充分验证了其在小样本分类任务中的有效性和广泛

适应性。

此外, 为了全面评估方法性能, 本文不仅从分类精度, 还从参数量、计算量和推理延迟对比了本文方法与较新的 TALDS-Net、DLDA 和 LDCA 方法, 结果如表 3 所示。

表 3 各项指标对比结果
Table 3 Comparison results of various indicators

模型	参数量/10 ⁶	FLOPs/10 ⁹	推理延迟/ms
TALDS-Net	13.14	1.82	3.14
DLDC	0.11	0.36	0.85
LDCA	1.78	0.78	3.09
本文	5.29	0.42	1.35

实验结果表明, MFRNet 在模型效率与分类性能间实现了显著平衡。相较于 TALDS-Net, 其参数量降低 59.7%、计算复杂度减少 76.9%, 推理速度提升 57.0%。与轻量化模型 DLDC 相比, 虽然参数量增加约 46 倍, 但计算复杂度仅提升 16.7%, 推理延迟维持在可接受范围。相较于集成上下文增强的 LDCA, MFRNet 在参数量增加 197% 的代价下, 计算复杂度降低 46.2%, 推理速度提升 56.3%。综上所述, 本文方法通过多重冗余抑制与轻量化注意力机制协同优化, 在分类精度显著提升的同时, 实现参数量、计算效率、模型体积与推理速度的多维度均衡, 为资源受限场景下的高效小样本分类提供了创新性解决方案。

3.4 消融实验

为了证明双流注意力融合模块 (DSAF)、可分离通道注意力 (SCA) 和自适应特征细化网络 (AFRN) 的有效性, 本文在 MiniImageNet、TieredImageNet 和 CUB 数据集上进行消融实验研究, 实验结果如表 4 所示。

表 4 消融实验结果
Table 4 Ablation experiment results

Backbone	DSAF	SCA	AFRN	MiniImageNet		TieredImageNet		CUB-200-2011	
				1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
√				62.67	77.88	68.48	83.46	71.98	86.64
√	√			65.49	84.65	71.55	85.79	72.31	87.67
√	√	√		67.86	86.55	73.03	87.41	74.32	87.76
√	√	√	√	70.49	88.96	75.42	89.85	89.34	95.52
√	√		√	68.56	87.05	73.56	88.12	74.88	88.32
√	√	√		67.86	86.55	73.03	87.41	74.32	87.76
√		√	√	69.52	88.45	74.85	88.36	76.25	90.27
√	√	√	√	70.49	88.96	75.42	89.85	89.34	95.52
√	跳跃连接	√	√	69.15	88.24	74.35	88.36	88.59	94.58
√	√	√	√	70.49	88.96	75.42	89.85	89.34	95.52

注: 加粗数据表示性能最优。

本文消融实验结果表明,各模块以递进方式显著提升模型性能。基准模型(仅 Backbone)在 MiniImageNet 1-shot 任务中准确率为 62.67%。引入 DSAF 模块实现多尺度特征融合后,准确率提升至 65.49%,表明跨层级特征交互可有效增强语义表征。叠加 SCA 模块进一步抑制空间与通道冗余,准确率增至 67.86%,印证了解耦式注意力机制在特征净化中的有效性。

除此之外还进行了模块之间两两匹配的消融实验。结果如表 4,移除 DSAF 模块会导致模型在 1-shot 和 5-shot 设置下的平均准确率分别下降 0.54% 和 0.33%,证明了跨层次特征融合的重要性;移除 SCA 模块后,模型性能下降更为显著,1-shot 和 5-shot 准确率平均降低了 1.75% 和 1.64%,凸显了通道-空间解耦在抑制冗余信息中的关键作用;而移除 AFRN 模块的影响最大,导致 1-shot 和 5-shot 准确率分别下降 2.35% 和 2.53%,充分验证了层级化特征细化机制对于提升判别性特征表达的贡献。这些实验结果一致表明,本文提出的 3 个核心模块通过协同作用,共同提升了模型在小样本分类任务中的性能。

为验证 DSAF 模块移除跳跃连接的必要性,在 ResNet-12 主干网络、SCA 与 AFRN 启用的控制条件下,对比其在 3 个数据集上的性能差异。结果表明,移除跳跃连接后,所有数据集及 1-shot/5-shot 设置的精度均稳定提升:MiniImageNet(1.34%/0.72%)、TieredImageNet(1.07%/1.49%)、CUB-200-2011(0.75%/0.94%),粗粒度数据集平均提升 1.15%,高于细粒度 0.84%。从机制看,保留跳跃连接会传递浅层噪声干扰多层池化筛选,移除后可切断噪声路径,提升特征纯净度;同时适配小样本场景,降低过拟合风险,并与主干残差形成

“梯度传递-特征筛选”协同闭环。综上,移除跳跃连接是 DSAF 模块实现冗余抑制、提升性能的必要设计,为 MFRNet 整体优势提供支撑。

进一步观察结果,在 CUB-200-2011 细粒度分类任务中,DSAF、SCA 与 AFRN 任一模块的缺失均导致性能显著下降,该现象源于细粒度分类对判别性局部特征高度依赖,而对类内冗余极度敏感的特性。CUB-200-2011 数据集中不同鸟类之间的区分往往依赖于羽毛纹理、翅膀细节等局部特征,而背景复杂、噪声较多。传统方法容易将浅层冗余信息传播至深层,掩盖这些关键差异。本文所构建的模块链恰针对此核心挑战。DSAF 实现跨尺度关键区域定位,SCA 负责抑制特征空间中的无关响应,AFRN 则完成判别性特征的终极筛选与强化。三者依次协作,构成从特征定位、净化至提炼的完整通路,缺一不可。因此,各模块移除所引发的性能衰减非但不是模型缺陷,反而证实了 MFRNet 架构设计与细粒度分类内在需求的高度契合,凸显了其系统性抑制冗余、增强判别性特征提取的有效性。

最后,为深入探究可分离通道注意力(SCA)与自适应特征细化网络(ARFN)在特征冗余抑制中的协同作用机制,本研究基于 Grad-CAM 技术对 3 个小样本基准数据集进行了可视化分析。如图 6 所示的实验结果表明,通过在双流注意力融合(DSAF)后依次引入 SCA 模块和 ARFN 模块构建的级联优化架构,能够显著提升模型的特征选择能力。该架构有效抑制了背景噪声和冗余特征的干扰响应,同时使核心前景目标的特征激活区域聚焦度提升。这种动态特征优选机制使得模型在有限样本条件下,能够建立更具判别性的特征表征空间。

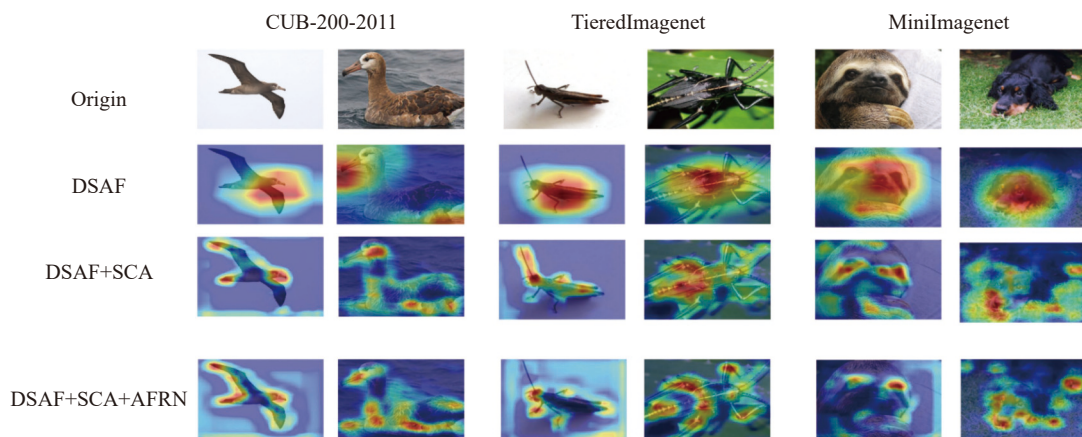


图 6 Gard-CAM 激活图
Fig. 6 Grad-CAM activation map

4 结束语

根据对小样本图像分类方法中融合全局和局部特征时容易产生冗余信息的现状研究, 采用系统性冗余抑制策略, 提出了多尺度特征融合细化网络 (MFRNet)。

针对小样本分类任务, 本文提出一种融合-细化两阶段的冗余抑制网络。首先通过双流注意力融合模块实现多尺度特征的协同表征, 随后采用可分离通道注意力机制解耦通道与空间特征以抑制信息冗余, 最终基于自适应特征细化网络完成判别性特征的优化选择。在多个标准数据集上的实验结果表明, 该方法显著提升了小样本分类性能, 各模块的有效性得到了充分验证。

尽管 MFRNet 在小样本图像分类任务中存在一定的优势, 但还有一些局限性, 超参数仍需人工调整。因此, 未来工作将聚焦于研究自适应超参数学习方法。

参考文献:

- [1] LU Jiang, GONG Pinghua, Ye Jieping, et al. A survey on machine learning from few samples[J]. *Pattern recognition*, 2023, 139: 109480.
- [2] SONG Yisheng, WANG Ting, CAI Puyu, et al. A comprehensive survey of few-shot learning: evolution, applications, challenges, and opportunities[J]. *ACM computing surveys*, 2023, 55(13): 1–40.
- [3] WANG Y, YAO Q, KWOK J T, et al. Generalizing from a few examples: a survey on few-shot learning[J]. *ACM computing surveys (CSUR)*, 2020, 53(3): 1–34.
- [4] 周伯俊, 陈峙宇. 基于深度元学习的小样本图像分类研究综述[J]. *计算机工程与应用*, 2024, 60(8): 1–15.
ZHOU Bojun, CHEN Zhiyu. Survey of few-shot image classification based on deep meta-learning[J]. *Computer engineering and applications*, 2024, 60(8): 1–15.
- [5] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning[C]//International Conference on Learning Representations. Toulon: ICLR, 2017.
- [6] MISHRA N, ROHANINEJAD M, CHEN X, et al. A simple neural attentive meta-learner[EB/OL]. (2017–07–01) [2025–01–01]. <https://arxiv.org/abs/1707.03141>.
- [7] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//International Conference on Machine Learning. Sydney: PMLR, 2017: 1126–1135.
- [8] NICHOL A, ACHIAM J, SCHULMAN J. On first-order meta-learning algorithms[EB/OL]. (2018–03–08) [2025–01–01]. <https://arxiv.org/abs/1803.02999>.
- [9] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[C]//Advances in Neural Information Processing Systems. Barcelona: NIPS, 2016, 29.
- [10] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[C]//Advances in Neural Information Processing Systems. Long Beach: NIPS, 2017, 30.
- [11] 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: CVPR, 2018: 1199–1208.
- [12] SATORRAS V G, ESTRACH J B. Few-shot learning with graph neural networks[C]//International Conference on Learning Representations. Vancouver: ICLR, 2018.
- [13] BOIMAN O, SHECHTMAN E, IRANI M. In defense of nearest-neighbor based image classification[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage: IEEE, 2008: 1–8.
- [14] LI Wenbin, WANG Lei, XU Jinglin, 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: CVPR, 2019: 7260–7268.
- [15] LIU Yang, ZHENG Tu, SONG Jie, et al. Dmn4: Few-shot learning via discriminative mutual nearest neighbor neural network[C]//Proceedings of the AAAI Conference on Artificial Intelligence. online: AAAI, 2022: 1828–1836.
- [16] 文浪, 苟光磊, 白瑞峰, 等. 基于邻域融合和特征增强的小样本细粒度图像分类[J]. *计算机工程*, 2026, 52(2): 158–166.
WEN Lang, GOU Guanglei, BAI Ruifeng, et al. Few-shot fine-grained image classification based on neighborhood fusion and feature enhancement[J]. *Computer engineering*, 2026, 52(2): 158–166.
- [17] 李小菲, 苟光磊, 韩岩奇, 等. 多尺度知识引导局部增强的小样本细粒度图像分类方法[J]. *智能系统学报*, 2024, 19(5): 1157–1167.
LI Xiaofei, GOU Guanglei, HAN Yanqi, et al. Multiscale knowledge-guided local enhancement method for few-shot fine-grained image classification[J]. *CAAI transactions on intelligent systems*, 2024, 19(5): 1157–1167.
- [18] ZHANG Xuchao, GAO Yifeng, LIN J, et al. Tapnet: mul-

- tivariate time series classification with attentional prototypical network[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 6845–6852.
- [19] QIAO Qian, XIE Yu, ZENG Ziyin, et al. Talds-net: task-aware adaptive local descriptors selection for few-shot image classification[C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Seoul: IEEE, 2024: 3750–3754.
- [20] YANG Yajie, FENG Yuxuan, ZHU Li, et al. Feature fusion network based on few-shot fine-grained classification[J]. *Frontiers in neurorobotics*, 2023, 17: 1301192.
- [21] KIM J, CHI M. SAFFNet: self-attention-based feature fusion network for remote sensing few-shot scene classification[J]. *Remote sensing*, 2021, 13(13): 2532.
- [22] HOU Ruibing, CHANG Hong, MA Bingpeng, et al. Cross attention network for few-shot classification[EB/OL]. (2019–10–17)[2025–01–01]. <https://arxiv.org/abs/1910.07677>.
- [23] 王雪松, 吕理想, 程玉虎, 等. 注意力集合表示的多尺度度量小样本图像分类[J]. *中国图象图形学报*, 2024, 29(11): 3371–3382.
- WANG Xuesong, LYU Lixiang, CHENG Yuhu, et al. Attention set representation for multiscale measurement of few-shot image classification[J]. *Journal of image and graphics*, 2024, 29(11): 3371–3382.
- [24] 周彬, 鲜浩, 秦艺嘉, 等. 融合混合注意力机制的局部原型小样本分类模型[J]. *计算机科学与探索*, 2026, 20(2): 584–595.
- ZHOU Bin, XIAN Hao, QIN Jiayi, et al. A local prototypes-based few-shot classification model embedded hybrid attention[J]. *Journal of frontiers of computer science and technology*, 2026, 20(2): 584–595.
- [25] 曹增辉, 陈浩, 曹雅慧. 基于类注意力的原型网络改进方法[J]. *自动化与信息工程*, 2025, 46(1): 59–65.
- CAO Zenghui¹, CHEN Hao¹, CAO Yahui. Improvement method of prototype network based on class attention[J]. *Automation & Information Engineering*, 2025, 46(1): 59–65.
- [26] CHEN Zixuan, HE Zewei, LU Zheming. DEA-net: single image dehazing based on detail-enhanced convolution and content-guided attention[J]. *IEEE transactions on image processing*, 2024, 33: 1002–1015.
- [27] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: CVPR, 2018: 7132–7141.
- [28] LIU Guilin, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[C]//Proceedings of the European Conference on Computer Vision. Munich: ECCV, 2018: 85–100.
- [29] WAH C, BRANSON S, WELINDER P, et al. The caltech-ucsd birds-200–2011 dataset: CNS-TR-2011–001[R]. [S. l. : s. n.], 2011.
- [30] DATASET E. Novel datasets for fine-grained image categorization[C]//First Workshop on Fine Grained Visual Categorization. Colorado Spring: CVPR, 2011: 2.
- [31] KRAUSE J, STARK M, DENG J, et al. 3D object representations for fine-grained categorization[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Sydney: ICCV, 2013: 554–561.
- [32] ZHANG Chi, CAI Yujun, LIN Guosheng, et al. DeepEMD: few-shot image classification with differentiable earth mover's distance and structured classifiers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. online: CVPR, 2020: 12203–12213.
- [33] WANG Xixi, Wang Xiao, JIANG Bo, 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.
- [34] LAI Jinxiang, YANG Siqian, LIU Wenlong, et al. tSF: transformer-based semantic filter for few-shot learning[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 1–19.
- [35] HU Ling, WU Wei. Enhancing few-shot image classification with a multi-faceted self-supervised and contrastive learning approach[J]. *IEEE access*, 2024, 12: 164844–164861.
- [36] WU Jijie, CHANG Dongliang, SAIN A, et al. Bi-directional feature reconstruction network for fine-grained few-shot image classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Washington, DC: AAAI, 2023: 2821–2829.
- [37] MA Zhenxiang, CHEN Zhenduo, ZHAO Lijun, et al. Cross-layer and cross-sample feature optimization network for few-shot fine-grained image classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2024: 4136–4144.
- [38] WANG Maofa, YAN Bingchen. LDCA: local descriptors with contextual augmentation for few-shot learning[EB/OL].

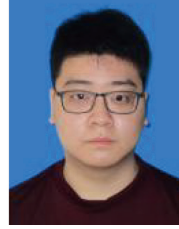
(2024-01-24)[2025-01-01]. <https://arxiv.org/abs/2401.13499>.

- [39] SONG Qijun, ZHOU Siyun, XU Liwei. Learning more discriminative local descriptors for few-shot learning [EB/OL]. (2023-05-15)[2025-01-01]. <https://arxiv.org/abs/2305.08721>.
- [40] ZHA Zican, TANG Hao, SUN Yunlian, et al. Boosting few-shot fine-grained recognition with background suppression and foreground alignment[J]. *IEEE transactions on circuits and systems for video technology*, 2023, 33(8): 3947-3961.
- [41] WANG Guangpeng, WANG Yongxiong, PAN Zhiqun, et al. Vitfsl-baseline: a simple baseline of vision transformer network for few-shot image classification[J]. *IEEE access*, 2024, 12: 11836-11849.
- [42] ZHANG Bo, YUAN Jiakang, LI Baopu, 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: ACM, 2022: 2135-2144.

作者简介:



吕伏, 副教授, 博士, 主要研究方向为智能数据处理、图像与视觉信息计算、大数据与云计算。发表学术论文 33 篇。E-mail: 38458786@qq.com。



张旭, 硕士研究生, 主要研究方向为小样本学习、图像分类。E-mail: 2696357169@qq.com。



张紫扬, 硕士研究生, 主要研究方向为目标检测。E-mail: 734759802@qq.com。