



局部关系泛化表征的小样本增量学习

赵一凡¹, 李甲^{2*}, 田永鸿¹

1. 北京大学计算机学院, 北京 100871

2. 北京航空航天大学计算机学院虚拟现实技术与系统国家重点实验室, 北京 100191

* 通信作者. E-mail: jiali@buaa.edu.cn

收稿日期: 2022-06-10; 接受日期: 2022-08-28; 网络出版日期: 2023-05-30

广东省重点领域研发计划项目 (批准号: 2020B0101380001)、国家自然科学基金 (批准号: 62132002, 61825101, 61922006) 和博士后面上基金 (批准号: 2022M710212) 资助

摘要 小样本学习作为计算机视觉的新兴研究热点, 旨在通过少量的已知样本实现对全新概念的快速识别和理解. 现有的小样本学习技术只针对给定的新类别物体进行分类和识别, 忽略了旧有基础类别的表示和区分. 针对小样本增量学习问题中的类间可区分性差、增量类别难泛化两大研究难点, 本文引入局部关系学习的思想, 提出局部关系泛化表征的模型学习方法. 为确保增量过程中的类别可区分性, 本文首先使用局部空间关联约束增量表征能力, 在此基础上, 使用特征重映射方法将查询向量与支撑集映射在同一度量空间, 有效确保了微小差异的可区分性. 同时, 为缓解增量过程中样本不足导致的归纳偏置问题, 本文提出空间泛化式原型生成算法, 利用分布特性快速生成虚拟原型, 促进样本的有效表征. 在此基础上, 本文利用元学习训练方法, 提出了局部泛化联合的增量学习框架, 通过联合基础类别的局部表征和增量类别的快速泛化约束, 有效缓解了灾难性遗忘和表征不足的困难, 实验证明本方法在经典小样本增量任务中取得了当前最好的性能表现.

关键词 小样本学习, 增量学习, 局部关系, 泛化表征, 元学习

1 引言

相比于传统的机器视觉, 人类视觉由于其特殊的记忆和推理功能, 往往能在观察少量样本的基础上, 形成对某一类别物体的认知和识别能力. 这一优势往往与经验记忆的积累以及对其他物体的认知概念密不可分. 受这种认知推理模式的启发, 现有工作倾向于通过对机器视觉系统进行训练和改造, 使其获得对小样本的学习能力. 小样本学习 (few-shot learning, FSL) ^[1~4] 通常旨在利用基础类别的物体语义进行训练, 从而使得机器学习系统具有在少量样本的新类别上快速泛化识别的能力. 为赋予机器学习系统这一能力, 前沿的小样本学习工作通常使用基于元学习 (meta learning) 的模型泛化表

引用格式: 赵一凡, 李甲, 田永鸿. 局部关系泛化表征的小样本增量学习. 中国科学: 信息科学, 2023, 53: 1132–1146, doi: 10.1360/SSI-2022-0231
Zhao Y F, Li J, Tian Y H. Generalized representation of local relationships for few-shot incremental learning (in Chinese). Sci Sin Inform, 2023, 53: 1132–1146, doi: 10.1360/SSI-2022-0231

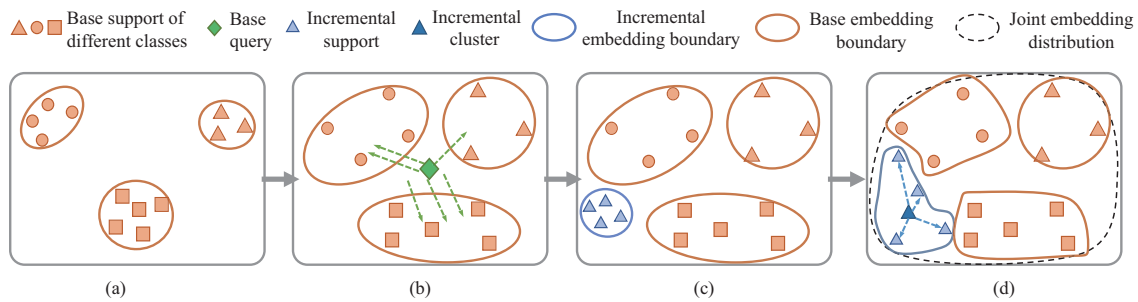


图 1 (网络版彩图) 局部关系泛化表征原理示意. (a) 初始基础类别原型^[8]分布; (b) 局部关系映射的元学习; (c) 增量类别原型初始化; (d) 快速泛化后的增量类别原型特征

Figure 1 (Color online) Motivation of generalized representation of local relationships. (a) Vanilla base embedding by prototypes^[8]; (b) meta learning with local-aware reprojection; (c) incremental prototype initialization; (d) incremental prototype after fast generalization

示^[5~7]和基于度量矩阵的表征学习两大类方法^[8~12],从而使得模型在新类别上形成快速表征.然而,与人类视觉系统类似,机器学习模型,例如,卷积神经网络(convolutional neural networks, CNNs),会快速遗忘之前学到的已知知识,从而在联合推理多种物体、认知复杂世界时,存在巨大的挑战和困境.

基于上述观察和挑战,一个实际可用的理想模型需要具有以下两个要素:(1)对先验知识或基础类别知识不遗忘;(2)对新类别物体具有快速泛化表征能力.近期少数工作提出了全新的小样本增量学习任务^[13](class-incremental few-shot learning, CIFSL),其目的在于研究在不断增加新类别的过程中,原有视觉系统能否对已知基础类别和新增类别形成共性的泛化表征.这一问题广泛存在于视觉任务的各种场景,例如,在生物领域对新物种进行发现并识别^[14],在深空探测^[15]、异常检测^[16]领域,需要仅依赖于几张图像,使视觉模型可以快速区分已有类别和新增类别.

然而,小样本增量学习作为一个全新的基础任务,目前对其探索的方法较为有限.最新工作^[13,17,18]主要借鉴增量学习(incremental learning)中的相关研究,用以克服或平衡严重的基础类别遗忘问题.早期典型研究,例如,Rebuffi等^[17]提出在新类别学习的基础上,对旧的基础知识进行缓存更新.在此基础上,Dong等^[18]和Cheraghian等^[19]分别提出基于关系保持和语义感知过程进行蒸馏的策略,使模型在新类别学习的过程中,对原有的特征先验进行保留学习.然而,这类方法在增加新类别的过程中,仅使用少量数据进行迭代训练,会导致基础类别的特征分布严重被破坏,而使用基础类别进行训练,则带来庞大的计算开销,也非增量学习的研究侧重点.为解决这一问题,现有先进方法^[20~24]通过原型网络^[8]的方式,构建平均特征进行度量学习,从而避免了反复梯度下降过程导致的原有特征损失.例如,Chen等^[21]提出使用量化表征的方式构建特征向量,同时学习到精要的基础嵌入特征. Shi等^[23]提出缓解灾难性遗忘的重点在于基础表征而非增量过程.具体而言,应该在训练基础类别的过程中,构建更加平缓的泛化表征.在此基础上,Zhou等^[22]提出在原型网络的基础上,构建虚拟的基础类别原型,使图像实现在特征空间的平滑过渡.该类型方法尽管取得了一定的性能优势,但其主要研究侧重于环境基础类别的灾难性遗忘(catastrophic forgetting),对新增类别仍然表现较差.而增量学习的另一目的是要求模型对于新增类别可以实现快速自适应的表征,本文将在后序中对这一点进行具体讨论.除此之外,现有方法在基础类别的表征过程中,均忽略了局部空间的关联关系,从而极大地影响了物体细节的表达能力.

针对上述问题,如图1(a)所示,增量学习在经过基础支撑类别(base support classes)的训练之后,会形成针对特定任务和类别的原始表征模式,引发较为严重的过拟合现象,从而严重影响对增量类别

(incremental classes) 的继续学习和表示. 在嵌入待识别的查询图像 (query) 后, 原有的基础类别难以对查询样本进行有效表征. 因此本文提出图 1(b) 中的局部关系保持的特征重映射方法, 通过利用样本的局部空间特性, 将支撑类别 (support classes) 重新投影到与待识别样本相同的度量空间中. 并利用局部特征两两之间的相似度作为度量, 通过元学习的泛化表征, 学习到具有空间分辨率和自适应映射的支撑原型 (prototypes). 在此基础上, 在增量任务过程中 (图 1(c)), 继续增加增量任务的原型仍然会导致其在流形空间中的密集表示, 难以形成稳定的泛化性表征. 因此, 不同于现有方法, 本文提出一种空间泛化式原型生成算法, 通过在多个样本原型的基础上, 增加自适应的泛化分布假设, 使用虚拟的泛化原型作为监督信号进行快速学习, 有效减缓了增量类别由于样本数量少导致的过拟合现象, 如图 1(d) 所示. 在此基础上, 本文进一步提出一种多阶段的局部 - 泛化表征联合的增量学习框架, 通过 (1) 在基础类别训练任务中, 利用元学习方式构建局部映射关系引导的泛化性表征; (2) 在增量学习任务中, 利用快速泛化虚拟原型表征, 实现快速增量式原型修正, 在实现增量类别的识别基础上有效缓解灾难性遗忘现象. 基于该增量学习框架, 本文通过在基于常规分类数据集 miniImageNet、细粒度数据集 CUB-200-2011 上的实验进行定性和定量比较, 证明了本文方法的有效性.

本文的具体贡献主要有以下 3 个方面: (1) 将局部关系学习的思想引入小样本增量学习, 并提出了局部关系保持的特征重映射算法, 有效促进了小样本增量学习中基础任务的泛化表征; (2) 提出了一种空间泛化式原型生成算法, 通过利用分布特性快速生成虚拟原型, 有效缓解样本不足引起增量类别表征不足的问题, 并与现有方法展开分析讨论; (3) 利用局部泛化联合的增量学习框架, 通过基础类别的局部表征和增量类别的快速泛化两部分约束, 有效缓解了灾难性遗忘和表征不足的缺陷, 实验证明了本方法相比于现有最好方法取得了明显的性能优势.

本文的其余部分组织安排如下: 第 2 节介绍了论文研究的相关工作, 第 3 节介绍了本文提出的局部泛化映射表征的小样本增量学习方法, 第 4 节通过定性和定量实验验证了本文所提算法的有效性. 第 5 节对本文结论进行总结概述.

2 相关工作

本文研究的小样本增量学习任务受小样本任务和增量学习任务定义的启发, 早期研究也尝试使用增量学习方法和小样本学习方法解决这一全新的任务. 除上述两方面前期研究之外, 近年来小样本增量学习任务也逐渐成为领域研究的热点问题. 下面将对这三部分进行进一步介绍.

2.1 增量学习

增量学习是在模型训练稳定后继续对模型进行新知识的补充学习的过程, 其主要研究思路分为三类. (1) 基于回放 (RePlay) 的方法: 该类型方法, 例如, iCaRL^[17], CLEAR^[25], A-GEM^[26] 等, 通过对旧数据的样本进行选择性的保留, 在更新新数据类别时, 对旧有数据进行回放训练. iCaRL^[17] 使用知识蒸馏损失函数更新模型参数, 同时保留了新旧知识间的表征能力, 以缓解在旧有知识的灾难性遗忘问题. (2) 基于正则化的学习方法: 例如, LwF^[27] 通过对新学习的增量任务增加正则化约束, 保证旧有的知识不被新学到的记忆覆盖. 以及 Liu 等^[28] 进一步通过 Fisher 信息矩阵构建弹性的重量固定化方法, 有效约束了序列化的增量任务的学习和表征能力. (3) 基于参数隔离的方法: 典型代表方法, 例如, PackNet^[29] 使用迭代剪枝和网络重训练操作, 在增加少量自适应参数的基础上, 兼顾了基础任务和后序增量任务的表征模式能力. 增量学习的经典方法通过在灾难性遗忘与新增类别学习过程中取得兼顾平衡, 在保持基础任务表征能力的基础上, 兼顾下游任务的学习表示. 然而, 该类型方法仍然依赖于大

量的训练样本,在实际的小样本增量任务中,由于增量类别样本数量极少,仍然难以获得有效的表征.

2.2 小样本学习

小样本学习 (few-shot learning) 是近年来领域研究的热点问题,重点研究如何通过一个或几个少量的可学习样本,实现对物体的认知和理解能力. 主流的小样本学习研究方法主要分为两大类:基于元学习的方法^[5~7,30~32]和基于度量表征的方法^[8~12,33,34]. 元学习方法旨在通过学习泛化表征的元学习器 (meta-learner), 通过片段式学习 (episodic training) 的方式得到与任务无关的先验知识. 例如, Rezende 等^[30] 提出使用非条件化的采样策略,并通过给定的概念学习基础类别的新样例,并将这种概念推广到新类别的全新样本. Laenen 和 Bertinetto^[31] 提出使用片段式训练的方式,通过模拟测试环境中的少量样本情况,使网络学习到任务无关的通用知识,从而对新概念的学习更加鲁棒. 与之相对,另一类方法通过构建原型表征 (prototype) 或使用关联关系对样本之间的距离进行度量学习. 例如, Snell 等^[8] 提出使用样本均值作为原型以替代原有的全连接学习模式,有效避免了梯度下降导致的网络特征更新和偏移现象. 同时文献^[33] 提出使用神经网络学习查询集合与支撑集集合的关联映射关系,从而利用这种度量关系形成快速的表征能力. 然而,小样本学习任务旨在研究网络特征能否在新概念的少数类别中形成泛化,对于基础类别的表征和嵌入缺乏考虑,对基础类别会形成严重的灾难性遗忘现象.

2.3 小样本增量学习

小样本增量学习的任务是指仅利用增量类别的少量可训练图像,从而形成对基础类别和新增类别的联合表征. 为了解决增量学习和小样本学习中的经典问题,少数前沿研究提出对小样本增量学习进行研究. 例如, Tao 等^[13] 提出使用神经气体的表征方式,使得新增类别和原有样本在流形空间保持合理的拓扑约束,防止空间坍塌或严重的类别不平衡现象. 而如前所述 Dong 等^[18] 和 Cheraghian 等^[19] 借鉴增量学习中的概念模式,通过增加额外的旧有知识数据联合新样本进行训练. 但该类型方法使用了额外的旧有样本,在增量任务中往往是不可获得的,同时该类型方法在基础类别的表示能力会获得较大的损害,导致最终性能的进一步下降. 为缓解增量学习导致的灾难性遗忘现象,现有主流方式^[20~23] 通过利用基础类别取得更好的泛化表征,而对增量类别使用原型网络进行度量. 例如, Shi 等^[23] 提出缓解增量任务性能下降的关键在于寻找平缓的泛化区间,避免网络的局部过拟合现象. Zhou 等^[22] 提出使用混合多个样本空间的方式,对基础类别的训练数据进行增强,用以获得更加鲁棒的特征表示. Zhang 等^[20] 提出在下游训练过程中采用参数隔离的方式对上游任务的特征进行保持,而对于下游任务,使用图神经网络进行微调. 然而,上述方法具有两个明显的缺点: (1) 仅仅基于基础类别训练的方法^[22,23] 严重忽略了增量任务的表征特性,对增量类别缺乏泛化能力; (2) 这些类别在基础类别表示过程中,忽略了局部特征间的关联关系,从而导致在使用相同的特征编码器编码特征后,在嵌入空间难以区分. 综合上述两点,本文提出了局部特征与泛化联合表示的小样本增量学习方法,在增强嵌入空间自适应表征的同时,兼顾下游任务的快速泛化学习.

3 局部关系泛化表征的小样本增量学习

3.1 问题定义与概述

问题定义. 小样本增量学习旨在基础类别学习的基础上,通过少量新给定的增量类别的样本,实现对基础和增量类别的联合快速识别. 与经典增量学习任务定义类似,小样本增量学习要求模型在

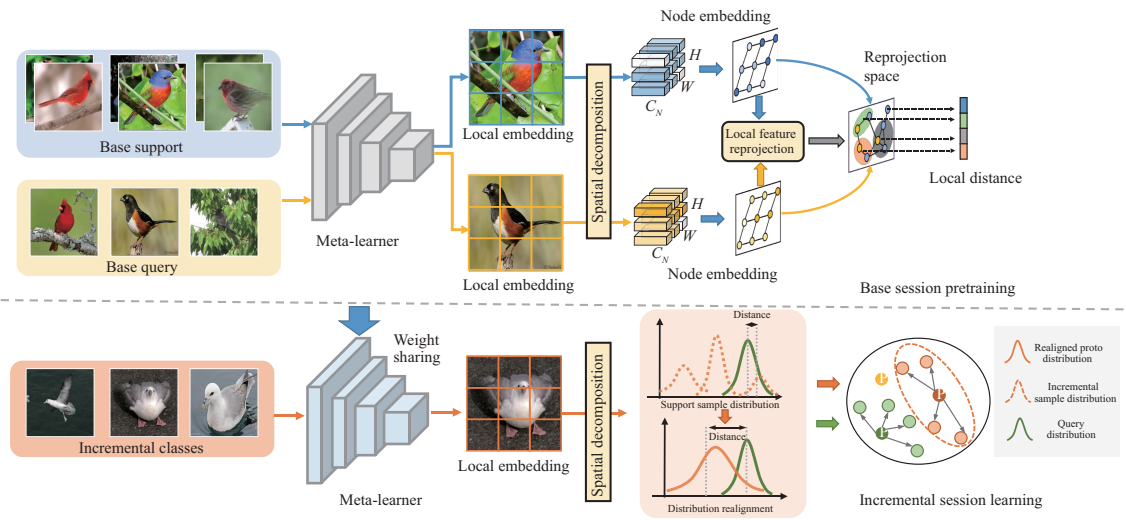


图 2 (网络版彩图) 局部泛化关系表征的小样本增量学习流程图。(1) 在基础会话训练阶段, 本文使用局部关系保持的特征重映射将支撑集和查询集映射在同一空间; (2) 在增量训练阶段, 使用空间泛化式原型生成策略对齐增量类别的分布

Figure 2 (Color online) Pipeline of generalized representation of local relationships for few-shot incremental learning. Our approach is composed of two essential parts. (1) In the base training session, we propose a local relationship-aware reprojection module with meta learning strategy to align the support and query features in the same space; (2) in the incremental session, we propose a spatial-aware generalized prototype generation strategy to align the incremental distribution with others

连续的 t 个训练会话 (sessions) 序列中, 即 $\{\mathcal{D}^1, \dots, \mathcal{D}^t\}$. 其中 $\mathcal{D}^t = \{(x_k^t, y_k^t)\} \in \mathcal{X}^t \times \mathcal{C}^t$. $\mathcal{X}^t, \mathcal{C}^t$ 分别代表第 t 次会话的输入样本空间与标签空间. 在小样本增量学习任务中, 基础会话 \mathcal{D}^1 通常包含较多的训练样本和类别, 而后序 $2 \sim t$ 次会话被设置为一个 N -way K -shot 的小样本问题, 即包含 N 个新增未知类别, 每个类别包含 K 个可训练样本, $K \ll |\mathcal{D}^1|$. 与经典增量学习任务的设置相同, 在训练时仅当前第 t 个会话序列的样本 \mathcal{D}^t 可见, 且不同会话之间的样本空间不存在交叉, 即 $\forall i, j \in \{1, \dots, t\}$ and $i \neq j, \mathcal{C}^i \cap \mathcal{C}^j = \emptyset$. 小样本增量学习的最终目标是在经过全部 t 个训练会话之后, 对全部可见类别 $\mathcal{C} = \cup_{i=1}^t \mathcal{C}^i$ 集合进行评价, 计算基础类别和增量会话类别的平均准确率.

问题分析与流程概述. 如第 1 和 2 节所述, 现存小样本增量学习方法的两大问题是 (1) 如何使用大量基础类别样本形成代表性特征, 确保流形空间中增量类别可区分能力; (2) 如何在确保基础类别不产生灾难性遗忘的同时, 仅利用少量样本对增量类别形成泛化性的表征? 针对上述两部分问题, 本文提出一种局部泛化映射表征的小样本增量学习框架, 具体如图 2 所示. 相比于前述小样本增量方法, 本文 (1) 首先将局部关联关系引入特征表示, 并使用原型自适应映射方法确保不同原型之间的可区分性, 具体见 3.2 小节; (2) 在确保基础特征映射不变的情况下, 利用少量样本的分布约束, 形成空间泛化式的增量原型表征, 见 3.3 小节. 在此基础上, 第 3.4 小节进一步简述提出局部泛化联合的增量学习框架, 在不增加额外参数的基础上实现小样本增量学习表征.

3.2 局部关系保持的特征重映射

小样本增量学习的一大挑战是如何通过基础类别的充足训练数据, 形成对下游会话表征能力较强、同时特征区分清晰的预训练特征提取器. 为解决上述问题, 如图 2 所示, 本小节在预训练强监督的基础上, 使用元学习器 (meta-learner) Φ 对下游增量训练任务进行模拟, 通过片段式 N -way K -shot

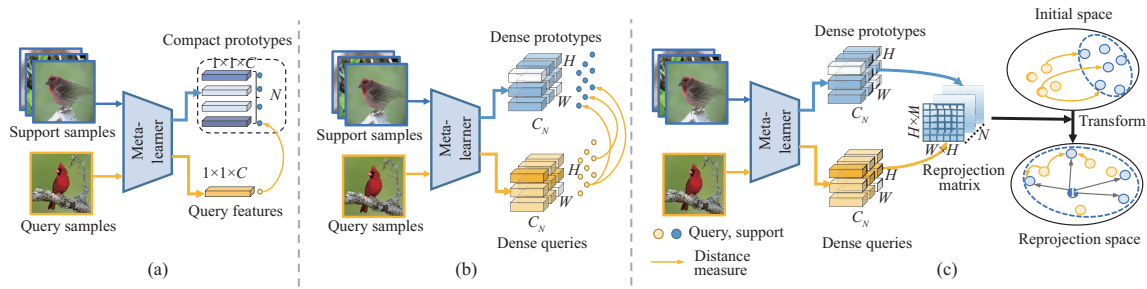


图 3 (网络版彩图) 三类典型距离度量策略对比. (a) 无空间感知的原型度量 [8, 9]; (b) 使用空间分辨率的局部度量 [35, 36]; (c) 本文提出的局部关系特征重映射方法

Figure 3 (Color online) Comparisons of three typical metric learning strategies. (a) Vanilla prototypical network without spatial awareness [8, 9]; (b) local-aware prototypical network with spatial resolution [35, 36]; (c) our local reprojection prototypical network

采样的方式多次模拟训练过程, 构建泛化表征的嵌入空间 $F = \Phi(x; \theta)$, 其中 θ 为模型参数. 在此基础上, 现有的小样本增量方法包括小样本方法 [8, 9], 通常使用池化操作后的样本平均特征作为原型向量: 对于第 $c \in C$ 个类别的特征, 构建图 3(a) 所示的原型特征 P^c :

$$P^c = \frac{1}{|\mathcal{D}_c^t|} \frac{1}{WH} \sum_{i=1}^{|\mathcal{D}_c^t|} \sum_{j=1}^{WH} F_{i,j}^c \in \mathbb{R}^{1 \times 1 \times C_N}, \quad (1)$$

其中 W, H, C_N 分别代表特征的宽、高以及通道维数. 然而这种方法不可避免地忽略了样本空间特征上的关联关系, 因此部分研究 [35, 36] 提出原型特征的保留空间尺度的局部关联关系, 如图 3(b) 所示. 然而在增量学习任务中, 由于无法使用大量的样本对元学习 $\Phi(x; \theta)$ 进行重新训练, 仅仅使用局部特征作为区分将会导致拥有大量样本的支撑集合和未知的查询样本无法处于相同的分布区域. 在增量会话 $\mathcal{D}^t, t > 1$ 中, 由于无法获得足够的训练样本, 这一现象将会更加严重. 因此, 受文献 [37, 38] 启发, 本小节使用 Ridge 回归作为特征映射, 将输入的查询 (query) 样本与支撑原型 (support prototypes) 利用局部特征关联关系, 将其映射在同一嵌入空间, 从而实现稳定的度量关系. 具体地, 如图 3(c) 所示, 本小节提出使用可学习的空间关联矩阵 $S \in \mathbb{R}^{|\mathcal{C}| \times WH \times WH}$, 其中 $|\mathcal{C}|$ 代表样本空间的类别数量. 利用该局部关系矩阵 S 对原型特征 P 进行重映射, 可以得到如下约束过程:

$$\hat{S} = \arg \min_S \|S \cdot P - Z\| + \lambda \|S\|_2^2 \in \mathbb{R}^{|\mathcal{C}| \times WH \times C_N}, \quad (2)$$

其中 Z 代表查询特征经过元学习器编码后的特征. Ridge 回归可以保证重映射向量的稀疏性 [37, 38], 避免投影过程陷入局部最优解, 确保嵌入特征的泛化性. 上述过程可以借鉴字典学习等求解方法进行获得, 在这里本文使用简单的闭式解对原型特征进行重新投影. 则式 (2) 可以得到如下形式:

$$\hat{S} = ZP(P P^T + \lambda I)^{-1} \in \mathbb{R}^{|\mathcal{C}| \times WH \times WH}. \quad (3)$$

根据原型网络的度量方式并使用蒸馏温度 τ [12, 38, 39] 作为熵值分布的平滑, 使用局部关联度量的相似度可以表征为

$$\text{Sim}(P^c, Z^c) = \frac{\exp(-\tau \|Z^c P^c (P^c P^{cT} + \lambda I)^{-1} P^c - Z^c\|^2)}{\sum_{k=1}^{|\mathcal{C}^t|} \exp(-\tau \|Z^k P^k (P^k P^{kT} + \lambda I)^{-1} P^k - Z^k\|^2)}. \quad (4)$$

根据该相似性度量函数 $\text{Sim}(P^c, Z^c)$, 可以得到第 c 个增量类别的语义原型度量. 为简化求解过程, 由于上述式 (4) 可以求得稳定的闭式解, 在保证矩阵可逆的前提下, 将对可学习矩阵 \hat{S} 的求解问题转化

为对原型 \mathbf{P}^c 的直接约束. 通过上述局部关联关系的表示, 可以将支撑向量 \mathbf{P} 和查询样本特征 \mathbf{Z} 映射在相同的高维语义空间, 在保证特征差异性的同时确保了度量空间的稳定性. 在上述增量学习过程中, 有效缓解了由不同会话、训练测试集合等导致的样本分布差异问题.

3.3 空间泛化式原型生成

经典的强监督学习过程将大量的样本作为监督约束, 可以生成特征充分泛化的原型向量. 而对于小样本增量任务, 由于增量会话中的样本个数较少, 单纯使用其作为监督会导致网络特征产生严重的过拟合现象, 这一现象在使用局部空间保持的原型表征中表现尤为显著. 同时, 文献 [40] 表明, 在样本数量充足的条件下, 训练得到的特征分布应符合高斯或正态分布. 因此, 本文提出解决小样本增量任务训练的一大关键是在增量阶段利用样本特性, 模拟在充足条件下的样本分布. 因此, 针对会话 \mathcal{D}_c^t 的样本对 $\{(\mathbf{x}_k^t, \mathbf{y}_k^t)\}$, 根据其统计特性, 计算每个类别 $c \in \mathcal{C}$ 的正态分布的均值 μ 和方差 σ^2 :

$$\begin{aligned} \mu_t(c) &= \frac{1}{|\mathcal{D}_c^t|} \sum_{\mathbf{x} \sim \mathcal{X}_c^t} \Phi(\mathbf{x}) \in \mathbb{R}^{W \times H \times C_N}, \\ \sigma_t^2(c) &= \frac{1}{|\mathcal{D}_c^t|} \sum_{\mathbf{x} \sim \mathcal{X}_c^t} (\Phi(\mathbf{x}) - \mu_t(c))^2. \end{aligned} \tag{5}$$

根据式 (5), 每个类别的特征分布可以表示为 $\mathcal{N}(\mu, \sigma^2)$, 则可以形成如下的泛化式原型空间:

$$\mathbb{S}_c = \{(\mathbf{Z}, c) | x \sim \mathcal{N}(\mu(c), \sigma^2(c))\}. \tag{6}$$

然而, 仅适用上述的泛化原型作为指导约束, 将会导致对样本局部特征的忽略, 形成统一的全局最优过程. 为使得这一方法具有充分的可扩展性并充分发挥样本的表征能力, 受流形空间 (manifold) 混合等思想的启发, 变换式 (4), 可以构建如下的损失约束过程:

$$\arg \min_{\mathbf{P}} \mathcal{L}_{\text{inc}}(\Sigma_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}^t \cup (\mathbf{Z}, c) \sim \mathbb{S}_c, c=y} \text{Sim}(\mathbf{P}, (1 - \alpha) \cdot \mathbf{Z} + \alpha \cdot \Phi(\mathbf{x}))), \tag{7}$$

其中超参数 α 用于平衡泛化流形与样本特异性之间的权重, 该值越大, 泛化程度越低. 则在该空间中, 本小节通过使用流形特征混合的方式, 对原型 $\mathbf{P} \in \mathbb{R}^{|\mathcal{C}^t| \times W \times H \times C_N}$ 使用具有随机性的泛化度量进行优化求解. 从而在保持空间维度向量的同时, 在保持空间维度向量的同时, 产生尽可能多的泛化样本, 有效防止了由于增量会话中样本数量过少导致的过拟合问题.

3.4 局部泛化联合表征的小样本增量学习框架

为实现在基础会话过程中的细粒度局部表征, 同时获得在增量任务中的快速泛化能力, 本小节提出一种局部泛化联合的增量学习框架, 具体细节如算法 1 所示. 该算法流程主要包含三部分: (1) 基础类别强监督预训练; (2) 基础类别元学习过程; (3) 增量类别快速泛化学习. 具体而言, (1) 基础类别会话 \mathcal{D}^1 的强监督预训练过程主要在经典的交叉熵损失的基础上, 使用局部关系保持和式 (4) 中的特征重投影算法, 优化如下的损失函数:

$$\mathcal{L}_{\text{CE}}(\mathcal{D}^1; \theta) = \Sigma_{(\mathbf{x}, \mathbf{y}) \in \mathcal{D}^1} \mathbf{y} \log(\text{Sim}(\mathbf{P}^{\mathbf{y}}, \Phi(\mathbf{x}^{\mathbf{y}}; \theta))), \tag{8}$$

其中 θ 是特征提取器 Φ 的可学习参数. $\mathbf{P}^{\mathbf{y}} \triangleq \{\mathbf{P}^c, \mathbf{y}_c = 1\}$ 为概率类别的简化表示. 通过上述强监督过程, 模型可以学习到空间关联表示的嵌入特征.

算法 1 局部泛化联合表征的小样本增量学习算法

Input: 会话序列 $\{\mathcal{D}^1, \dots, \mathcal{D}^t\}$, $\mathcal{D}^t = \{(\mathbf{x}_k^t, \mathbf{y}_k^t)\} \in \mathcal{X}^t \times \mathcal{C}^t$;
Output: 特征提取器 Φ_{θ^1} , 基础和增量原型联合分布 $\{\hat{\mathbf{P}}^c | c \in \mathcal{C}\}$;

- 1: 随机初始化基础类别原型: $\mathbf{P} \in \mathbb{R}^{C^1 \times WH \times C}$;
 { /* 基础类别强监督预训练 */ }
- 2: **for** $(\mathbf{x}, \mathbf{y}) \in \mathcal{D}^1$ **do**
- 3: 根据式 (3) 和 (4) 计算相似度 $\text{Sim}(\mathbf{P}, \Phi_{\theta^0}(\mathbf{x}))$;
- 4: 使用梯度下降更新网络参数, 学习率为 γ_{ce} : $\theta^0 \leftarrow \theta^0 - \gamma_{ce} \nabla \mathcal{L}_{CE}(\text{Sim}, \mathbf{y}; \theta^0)$;
- 5: **end for**
 { /* 基础类别元学习过程 */ }
- 6: **for** 采样任务 $\mathcal{T} \sim \mathcal{D}^1$ **do**
- 7: 划分 $\mathcal{T} = \{(\mathbf{x}, \mathbf{y})\}_{i=1}^{K \times N}$ 为伪支撑集 $\{(\mathbf{x}^{qry}, \mathbf{y}^{qry})\}$ 和伪查询集 $\{(\mathbf{x}^{sp}, \mathbf{y}^{sp})\}$;
- 8: 根据式 (1) 计算查询集合的原型 $\mathbf{P} = \text{Proto}(\mathbf{x}^{qry}, \mathbf{y}^{qry})$;
- 9: 根据式 (3) 和 (4) 计算相似度 $\text{Sim}(\mathbf{P}, \Phi_{\theta^1}(\mathbf{x}))$;
- 10: 使用伪标签 \mathbf{k} 更新网络参数, 学习率为 γ_m : $\theta^1 \leftarrow \theta^1 - \gamma_m \nabla \mathcal{L}_{meta}(\text{Sim}, \mathbf{k}; \theta^1)$;
- 11: **end for**
 { /* 增量类别快速泛化学习 */ }
- 12: **for** $t = 2, \dots, T$ **do**
- 13: 利用元学习器参数 Φ_{θ^1} 提取嵌入特征, 根据式 (5) 计算增量类别 c 的样本特征分布 $\mathbb{S}_c = \mathcal{N}(\mu(c), \sigma^2(c))$;
- 14: **for** $(\mathbf{x}, \mathbf{y}) \in \mathcal{D}^t$ **do**
- 15: 根据式 (6) 采样泛化原型 $(\mathbf{Z}, c) \sim \mathbb{S}_c$;
- 16: 根据式 (7) 流形混合度量 $\text{Sim}(\mathbf{P}, (1 - \alpha) \cdot \mathbf{Z} + \alpha \cdot \Phi(\mathbf{x}))$;
- 17: 更新原型: $\hat{\mathbf{P}} \leftarrow \hat{\mathbf{P}} - \gamma_{inc} \nabla \mathcal{L}_{inc}(\text{Sim}, \mathbf{y}; \hat{\mathbf{P}})$;
- 18: **end for**
- 19: **end for**
- 20: **return** $\Phi_{\theta^1}, \hat{\mathbf{P}}$.

(2) 然而这种强监督方式容易带来一定的过拟合现象, 为增强特征提取器对下游增量任务的泛化性表示, 本小节使用片段式训练方式, 通过在基础会话中采样 N -way K -shot 的训练任务 $\mathcal{T} = \{\mathbf{x}, \mathbf{k}\} \sim \mathcal{D}^1$, 其类别标签 \mathbf{k} 被重新设定为 $1, \dots, N$, 构建平均后的临时原型 \mathbf{H} , 该元学习训练过程损失约束与上式类似:

$$\mathcal{L}_{meta}(\mathcal{D}^1; \theta) = \sum_{\mathcal{T} \sim \mathcal{D}^1} \sum_{(\mathbf{x}, \mathbf{k}) \in \mathcal{T}} \mathbf{k} \log(\text{Sim}(\mathbf{H}^{\mathbf{k}}, \Phi_{\theta^0}(\mathbf{x}^{\mathbf{k}}; \theta))). \quad (9)$$

上述元学习优化过程在第一阶段参数为 Φ_{θ^0} , 经过元学习训练后得到网络参数 θ^1 . 并根据该训练参数, 根据式 (1) 重新计算基础类别的原型 $\{\hat{\mathbf{P}}^1, \dots, \hat{\mathbf{P}}^{|\mathcal{C}^1|}\}$.

(3) 在本框架的增量学习阶段, 本小节提出使用空间泛化式原型生成算法, 用于约束新类别的 $|\mathcal{C}^t|$ 个原型权重. 使用参数隔离算法固定元学习器 Φ_{θ^1} , 在泛化特征表示的基础上, 修正下游任务的权重学习过程. 对于第 t 次增量会话, 本框架利用式 (7) 的约束条件, 计算优化后的原型 $\hat{\mathbf{P}} = \arg \min_{\mathbf{P}} \mathcal{L}_{inc}(\mathcal{D}^t; \mathbf{P})$. 通过联合上述三部分渐进式的损失约束, 本框架可以逐步得到空间关联自适应表示、特征表达能力强、下游增量快速泛化的最终参数模型 Φ_{θ^1} , 以及优化后的增量原型表征 $\{\hat{\mathbf{P}}^c | c \in \mathcal{C}\}$, 使用式 (4) 进行度量即可得到各个类别的联合分布概率.

4 实验结果与分析

4.1 实验设置

本小节沿用主流小样本增量的任务设定^[13,20,21,41], 使用常规数据集 miniImageNet^[42] 以及经典细粒度评测基准 Caltech UCSD Birds-200-2011 (CUB200)^[43] 进行评测. 本小节将对实验数据集、评价指标、实现细节分别进行介绍.

4.1.1 实验数据集和评价指标

miniImageNet. miniImageNet^[42] 是由大规模挑战赛 ImageNet^[44] 选取代表性子集进行构建的, 共包含 100 个不同的语义类别, 每个语义类别包含 600 张分辨率为 84×84 的图像. 本文根据经典方法^[13,20] 的设定, 将 100 个类别分别划分为 60 个基础类别 (即会话 \mathcal{D}^1) 和 40 个增量类别. 其中增量类别被均匀地划分为 8 个会话 (即会话 $\mathcal{D}^2, \dots, \mathcal{D}^9$), 其中每个会话包含 5 个类别, 每个类别中有 5 张图像用于训练, 形成 5-way 5-shot 的典型小样本问题.

CUB-200-2011. CUB 数据集^[43] 是最经典的常用细粒度分类数据集, 广泛用于常规模型的性能评测. 该数据集包含 200 个不同的鸟类种属, 共计 11788 张分辨率为 224×224 的图像. 根据经典方法和数据集的划分要求, 本文将 200 个类别划分为 100 个基础类别 (即会话 \mathcal{D}^1) 和 100 个增量类别. 其中 100 个增量类别被均匀地划分为 10 次增量会话, 即 10-way 5-shot 问题.

评价指标. 与前述代表性工作^[13,20,21,41] 相同, 本文使用在每次会话 $\mathcal{D}^1, \dots, \mathcal{D}^t$ 后的 Top-1 准确率作为评价标准. 在对模型进行评价时, 一般根据最终会话 \mathcal{D}^t 后的准确率进行判断, 同时使用 PD (performance dropping) 计算增量会话导致的性能下降, 即基础会话 \mathcal{D}^1 的准确率与最后一次增量会话 \mathcal{D}^t 准确率的差值.

4.1.2 实现细节

为与经典方法实现公平对比, 本文采用 ResNet18^[45] 作为基础特征提取器 $\Phi(\cdot)$, 沿用经典方法^[13,20,21,41] 的对比方式, 在计算细粒度数据集 CUB 时, 采用 ImageNet^[44] 预训练后的模型. 在计算 miniImageNet 数据集时, 网络以随机初始化开始训练. 在基础会话 \mathcal{D}^1 中使用 SGD 优化器训练 100 轮次, 学习率分别设置为 0.002 (CUB) 和 0.001 (miniImageNet), 在增量会话中使用 Adam 作为优化器训练 10 轮次. 元学习阶段的采样使用 20-way 10-shot 模式, 模拟其泛化性. 可学习温度 τ 设置为 $\frac{16}{WH}$. 数据增强方法和 batchsize 大小与经典方法^[20] 一致.

4.2 与现有算法的对比

CUB-200-2011. 为与经典方法实现公平对比, 本文采用相同的设置, 在不增加任何额外训练参数的前提下, 选取了 9 种代表性的小样本增量学习方法. 其中典型增量方法 iCaRL^[17] 与 Rebalance^[46] 为当前最先进的增量学习方法, 本文将其引入增量小样本学习任务当中进行对比. 本文基线为基础会话与增量会话的参数隔离方法, 其中增量部分使用小样本数据进行原型初始化操作. 上述参数隔离基线方法可以明显地缓解灾难性遗忘问题, 并超越了部分早期方法. 由表 1^[47] 可以看出, 典型增量方法在样本数据量较少的情况下, 出现了严重的性能下降, 其 PD 超过 45%, 无法很好地解决灾难性遗忘问题. 同时, 表 1 的第一行数值为使用增量样本进行微调的结果, 可以看出在对模型骨干进行微调之后, 模型仅具有对当前输入样本的拟合能力, 其他基础类别和之前增量会话类别的表征能力大幅降低. 在此基础上, TOPIC^[13], SPPR^[41] 等方法尽管实现了增量类别的快速表示, 但是产生了严重的灾难性

表 1 CUB 数据集上的 10-way 5-shot 增量学习准确率 (%) 对比
 Table 1 Classification accuracy (%) on CUB dataset for 10-way 5-shot incremental learning^{a)}

Method	Pub. year	Accuracy in each session \uparrow											PD \downarrow
		1	2	3	4	5	6	7	8	9	10	11	
Finetune ^[13]		68.68	43.70	25.05	17.72	18.08	16.95	15.10	10.06	8.93	8.93	8.47	60.21
iCaRL ^{[17]*}	CVPR 17	68.68	52.65	48.61	44.16	36.62	29.52	27.83	26.26	24.01	23.89	21.16	47.52
Rebalance ^{[46]*}	CVPR 19	80.94	70.32	62.96	57.19	51.06	46.70	44.03	40.15	36.75	34.88	32.09	48.85
TOPIC ^[13]	CVPR 20	68.68	62.49	54.81	49.99	45.25	41.40	38.35	35.36	32.22	28.31	26.28	42.40
FSSL ^[47]	CVPR 20	72.77	69.33	65.51	62.66	61.10	58.65	57.78	57.26	55.59	55.39	54.21	18.56
FSSL+SS ^[47]	CVPR 20	75.63	71.81	68.16	64.32	62.61	60.10	58.82	58.70	56.45	56.41	55.82	19.81
IDLVQ-C ^[21]	ICLR 20	77.37	74.72	70.28	67.13	65.34	63.52	62.10	61.54	59.04	58.68	57.81	19.56
SPPR ^[41]	CVPR 21	68.68	61.85	57.43	52.68	50.19	46.88	44.65	43.07	40.17	39.63	37.33	31.35
CEC ^[20]	CVPR 21	75.85	71.94	68.50	63.50	62.43	58.27	57.73	55.81	54.83	53.52	52.28	23.57
FACT ^[22]	CVPR 22	75.90	73.23	70.84	66.13	65.56	62.15	61.74	59.83	58.41	57.89	56.94	18.96
Baseline	–	76.27	71.14	66.89	61.82	60.30	56.66	54.93	52.93	51.36	50.37	48.52	27.75
Ours	–	78.11	75.65	72.73	68.43	68.46	66.48	65.75	64.39	63.21	62.57	61.81	16.30

a) *: performances reported by [22]. PD \downarrow : performance dropping compared to the base session \mathcal{D}^1 .

表 2 miniImageNet 数据集上的 5-way 5-shot 增量学习准确率 (%) 对比
 Table 2 Classification accuracy (%) on miniImageNet dataset for 5-way 5-shot incremental learning^{a)}

Method	Pub. year	Accuracy in each session \uparrow									PD \downarrow
		1	2	3	4	5	6	7	8	9	
Finetune ^[13]	–	61.31	27.22	16.37	6.08	2.54	1.56	1.93	2.60	1.40	59.91
iCaRL ^{[17]*}	CVPR 17	61.31	46.32	42.94	37.63	30.49	24.00	20.89	18.80	17.21	44.10
Rebalance ^{[46]*}	CVPR 19	61.31	47.80	39.31	31.91	25.68	21.35	18.67	17.24	14.17	47.14
TOPIC ^[13]	CVPR 20	61.31	50.09	45.17	41.16	37.48	35.52	32.19	29.46	24.42	36.89
FSSL ^[47]	CVPR 20	66.48	61.75	58.16	54.16	51.10	48.53	46.54	44.20	42.28	24.20
FSSL+SS ^[47]	CVPR 20	68.85	63.14	59.24	55.23	52.24	49.65	47.74	45.23	43.92	22.93
IDLVQ-C ^[21]	ICLR 20	64.77	59.87	55.93	52.62	49.88	47.55	44.83	43.14	41.84	22.93
CEC ^[20]	CVPR 21	72.00	66.83	62.97	59.43	56.70	53.73	51.19	49.24	47.63	24.37
F2M ^[23]	NeurIPS 21	67.28	63.80	60.38	57.06	54.08	51.39	48.82	46.58	44.65	22.63
Baseline	–	70.86	65.71	61.66	58.51	55.49	52.68	50.07	48.08	46.64	24.22
Ours	–	71.57	66.94	63.06	60.20	57.41	54.38	51.71	50.13	49.02	22.55

a) *: performances reported by [23]. PD \downarrow : performance dropping compared to the base session \mathcal{D}^1 .

遗忘现象. 同时, 除本方法外的最好方法, 例如 FACT^[22], IDLVQ-C^[21] 等均侧重于对基础类别的约束表征, 对于增量类别往往表征能力较差. 为此, 本文方法使用与文献 [20] 相同的基线方法, 并利用局部表征与泛化映射联合的方式, 取得了在最终会话轮次 13% 的增量性能提升. 在使用本文提出的联合学习框架的基础上, 本文方法取得了在最后会话增量后, 全部数据集 61.81% 的准确率, 领先现有方法近 4%, 并取得了最低的性能下降值 PD.

miniImageNet. 为验证在常规图像上的小样本增量学习性能, 如表 2 所示, 第一行数值为使用微调方法的增量学习性能. 在常规数据集上使用微调等方式会导致更为严重的灾难性遗忘现象, 综合

表 3 模块消融实验分析

Table 3 Ablation studies of incremental learning on CUB benchmark^{a)}

$\mathcal{M}_{\text{Proj}}$	$\mathcal{M}_{\text{Local}}$	$\mathcal{M}_{\text{Meta}}$	$\mathcal{M}_{\text{GenP}}$	Acc (\mathcal{D}^1) (%)	Acc (\mathcal{D}^t) (%)	Δ (Rel. Acc) (%)	PD \downarrow (%)
Baseline ^[20]	–	–	Finetune	76.27	48.52	(+0.0)	27.75
✓	–	✓	Mix	75.68	58.46	(+9.94)	17.22
✓	✓	–	Finetune	78.30	59.08	(+10.56)	19.22
✓	✓	✓	Finetune	77.99	61.13	(+12.61)	16.86
✓	✓	✓	Mix	78.11	61.81	(+13.29)	16.30

a) $\mathcal{M}_{\text{Proj}}$, $\mathcal{M}_{\text{Local}}$, $\mathcal{M}_{\text{Meta}}$, and $\mathcal{M}_{\text{GenP}}$ denote feature reprojection, local embedding, meta learning constraints, and generalized prototypes, respectively.

准确率降低为 1.40%，即由于增量类别图像过少，无法实现很好的表征，同时在基础类别造成严重的遗忘现象。与 CUB 数据集类似，单纯的增量学习方法 (iCaRL^[17] 与 Rebalance^[46]) 未兼顾到增量会话中样本数量少的特性，虽在一定程度上缓解了遗忘现象，但性能仍然存在较大的瓶颈。本文的基线方法基于文献 [20]，并在此基础上取得了会话 $\mathcal{D}^1 \sim \mathcal{D}^9$ 全面的性能领先。本文方法大幅超越最新方法 CEC^[20], F2M^[23] 等 8 种最新方法，同时实现了最小的性能下降。

4.3 消融实验与性能分析

模块消融分析. 为分别验证各个模块的作用，本文将所提出方法的各个部分进行拆分，即 3.2 小节中的特征重投影约束 $\mathcal{M}_{\text{Proj}}$ 与局部空间嵌入 $\mathcal{M}_{\text{Local}}$ ，式 (9) 所构建的元学习约束 $\mathcal{M}_{\text{Meta}}$ ，以及 3.3 小节提出的泛化原型约束。如表 3 所示，本小节通过实验对所提方法的各个模块进行消融，其中第一行所示为使用参数解耦的原型基线方法，取得了 48.52% 的准确率。在增加本文提出的局部关系引导的特征重映射方法后，其基础会话准确率和增量会话准确率获得了明显的提升，在会话 \mathcal{D}^t 准确率提升 10.56%。同时，为验证本文提出的混合原型训练策略、元学习约束、局部特征保持等设计方法，分别在二、三、四行在完整模型的基础上对其进行消融，均导致了 1%~4% 的性能下降，验证了本文所提方法的有效性。观察最后一列数据 PD 可以得知，本文方法取得了最少的性能下降，确保了基础类别和增量类别同时有效表征。

局部关系影响分析. 为验证本文提出的局部关系表征对性能的影响，本小节进一步计算在不同空间分辨率下的模型表征能力。如图 4(a) 所示，紫色为消除本文所提的 4 个模块的基线方法，即表 3 第一行数据。红色与蓝色分别表示原型特征 \mathbf{P} 分辨率为 5×5 与分辨率为 1×1 时的度量结果。经典小样本方法通常使用 1×1 的原型代替全连接作为最终的概率度量。在本文提出的特征重映射模块中，当仅降低分辨率而保留其他模块时，其最终准确率从 61.81% 下降为 58.45%。综上所述可以看出，使用局部关系表征在促进增量类别表征的前提下，可有效增强基础会话类别的表示能力。

融合因子影响分析. 为验证不同融合因子 (式 (7) 中 α) 对增量学习表征的影响，如图 4(b) 所示，其中红色线条表示使用混合原型表征与增量特征后的 10 个 session 准确率，其中 α 被设置为 0~1 之间的随机值。其中蓝色结果为不使用泛化原型混合后的表征性能，可以发现最终会话出现了近 1% 的性能下降。而仅使用泛化原型结果进行约束，即紫色结果，会导致网络学习到更多的随机扰动，严重影响性能分析结果。结合泛化结果，同时兼顾特征自身表征，可以在原型泛化的基础上快速获得增量类别的自适应能力。

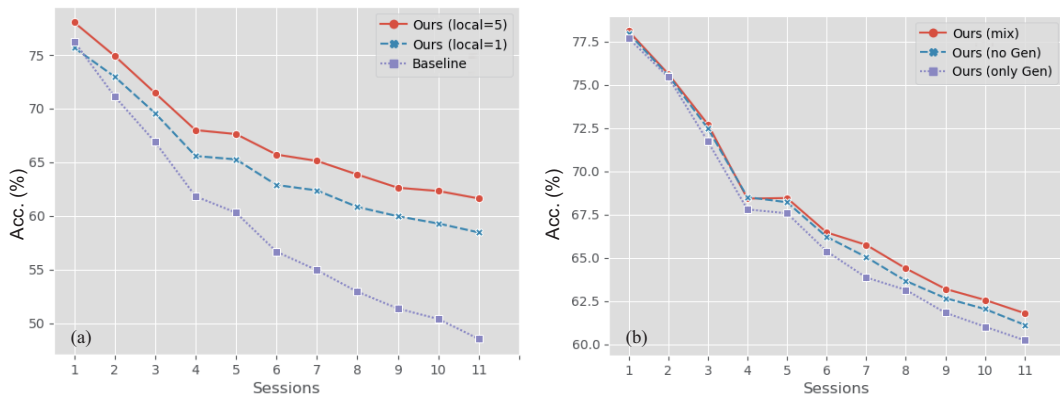


图 4 (网络版彩图) 影响因素分析. (a) 局部关系影响分析; (b) 融合因子影响分析

Figure 4 (Color online) Effects of different components. (a) Analyses on local feature resolutions; (b) analyses on generalized prototypes

5 结论

本文提出了一种局部泛化表征的小样本增量学习方法,用于解决现有方法在学习过程中类间可区分性差、增量类别难泛化两大难题.面对上述问题,本文通过引入局部关系学习思想,同时使用特征重映射方法将查询样本和支撑样本映射在相同的流形空间,有效避免了微小差异难表达,空间尺度表示不充分等研究困难.同时由于增量过程中样本严重不足,本文提出一种空间泛化式原型生成算法,使用分布特性快速模拟虚拟原型,使模型学习到近似的泛化结果.在上述两部分算法的基础上,本文提出了一种基于元学习训练的局部泛化联合表示的增量学习框架,快速缓解小样本增量学习中的灾难性遗忘和模型表征能力不足等困境.最后,本文通过实验证明所提方法超越了 9 种先进方法,不仅可以在常规数据集上有效实现快速泛化,同时可以实现对细粒度特征的高效建模和表示.

参考文献

- 1 Miller E G, Matsakis N E, Viola P A. Learning from one example through shared densities on transforms. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2000. 464–471
- 2 Li F-F, Fergus R, Perona P. One-shot learning of object categories. *IEEE Trans Pattern Anal Machine Intell*, 2006, 28: 594–611
- 3 Lake B M, Salakhutdinov R, Tenenbaum J B. Human-level concept learning through probabilistic program induction. *Science*, 2015, 350: 1332–1338
- 4 Jiang S, Zhu Y, Liu C, et al. Dataset bias in few-shot image recognition. *IEEE Trans Pattern Anal Mach Intell*, 2022. doi: 10.1109/TPAMI.2022.3153611
- 5 Ravi S, Larochelle H. Optimization as a model for few-shot learning. In: Proceedings of International Conference on Learning Representations (ICLR), 2017
- 6 Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of International Conference on Machine Learning (ICML), 2017. 1126–1135
- 7 Yu Z, Chen L, Cheng Z, et al. Transmatch: a transfer learning scheme for semi-supervised few-shot learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020. 12856–12864
- 8 Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS), 2017. 30
- 9 Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS), 2016. 3637–3645

- 10 Oreshkin B N, Rodriguez P, Lacoste A. Tadam: task dependent adaptive metric for improved few-shot learning. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS), 2018. 719–729
- 11 Liu B, Cao Y, Lin Y, et al. Negative margin matters: understanding margin in few-shot classification. In: Proceedings of European Conference on Computer Vision (ECCV), 2020. 438–455
- 12 Dhillon G S, Chaudhari P, Ravichandran A, et al. A baseline for few-shot image classification. In: Proceedings of International Conference on Learning Representations (ICLR), 2019
- 13 Tao X, Hong X, Chang X, et al. Few-shot class-incremental learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020. 12183–12192
- 14 Li Y, Yang J. Few-shot cotton pest recognition and terminal realization. *Comput Electron Agr*, 2020, 169: 105240
- 15 Yang X, Nan X, Song B. D2N4: a discriminative deep nearest neighbor neural network for few-shot space target recognition. *IEEE Trans Geosci Remote Sens*, 2020, 58: 3667–3676
- 16 Lu Y, Yu F, Reddy M K, et al. Few-shot scene-adaptive anomaly detection. In: Proceedings of European Conference on Computer Vision (ECCV), 2020. 125–141
- 17 Rebuffi S A, Kolesnikov A, Sperl G, et al. iCaRL: incremental classifier and representation learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 2001–2010
- 18 Dong S, Hong X, Tao X, et al. Few-shot class-incremental learning via relation knowledge distillation. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 1255–1263
- 19 Cheraghian A, Rahman S, Fang P, et al. Semantic-aware knowledge distillation for few shot class-incremental learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021. 2534–2543
- 20 Zhang C, Song N, Lin G, et al. Few-shot incremental learning with continually evolved classifiers. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021
- 21 Chen K, Lee C G. Incremental few-shot learning via vector quantization in deep embedded space. In: Proceedings of International Conference on Learning Representations (ICLR), 2020
- 22 Zhou D W, Wang F Y, Ye H J, et al. Forward compatible few-shot class-incremental learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022. 9046–9056
- 23 Shi G, Chen J, Zhang W, et al. Overcoming catastrophic forgetting in incremental few-shot learning by finding flat minima. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS), 2021. 34
- 24 Cheraghian A, Rahman S, Ramasinghe S, et al. Synthesized feature based few-shot class incremental learning on a mixture of subspaces. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021. 8661–8670
- 25 Rolnick D, Ahuja A, Schwarz J, et al. Experience replay for continual learning. In: Proceedings of Advances in Neural Information Processing Systems, 2019. 32
- 26 Chaudhry A, Ranzato M A, Rohrbach M, et al. Efficient lifelong learning with A-GEM. In: Proceedings of International Conference on Learning Representations (ICLR), 2019
- 27 Li Z, Hoiem D. Learning without forgetting. *IEEE Trans Pattern Anal Mach Intell*, 2018, 40: 2935–2947
- 28 Liu X, Masana M, Herranz L, et al. Rotate your networks: better weight consolidation and less catastrophic forgetting. In: Proceedings of the 24th International Conference on Pattern Recognition (ICPR), 2018. 2262–2268
- 29 Mallya A, Lazebnik S. PackNet: adding multiple tasks to a single network by iterative pruning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. 7765–7773
- 30 Rezende D, Danihelka I, Gregor K, et al. One shot generalization in deep generative models. In: Proceedings of International Conference on Machine Learning (ICML), 2016. 1521–1529
- 31 Laenen S, Bertinetto L. On episodes, prototypical networks, and few-shot learning. In: Proceedings of Advances in Neural Information Processing Systems (NeurIPS), 2021. 34
- 32 Ren M, Triantafillou E, Ravi S, et al. Meta-learning for semi-supervised few-shot classification. In: Proceedings of International Conference on Learning Representations (ICLR), 2018
- 33 Sung F, Yang Y, Zhang L, et al. Learning to compare: relation network for few shot learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. 1199–1208
- 34 Tang S, Chen D, Bai L, et al. Mutual CRF-GNN for few-shot learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021. 2329–2339
- 35 Lifchitz Y, Avrithis Y, Picard S, et al. Dense classification and implanting for few-shot learning. In: Proceedings of

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019. 9258–9267
- 36 Li W, Wang L, Xu J, et al. Revisiting local descriptor based image-to-class measure for few-shot learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019. 7260–7268
- 37 Bertinetto L, Henriques J F, Torr P, et al. Meta learning with differentiable closed-form solvers. In: Proceedings of International Conference on Learning Representations (ICLR), 2018
- 38 Wertheimer D, Tang L, Hariharan B. Few-shot classification with feature map reconstruction networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021. 8012–8021
- 39 Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. 2015. ArXiv:1503.02531
- 40 Yang S, Liu L, Xu M. Free lunch for few-shot learning: distribution calibration. 2021. ArXiv:2101.06395
- 41 Zhu K, Cao Y, Zhai W, et al. Self-promoted prototype refinement for few-shot class-incremental learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021. 6801–6810
- 42 Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*, 2015, 115: 211–252
- 43 Wah C, Branson S, Welinder P, et al. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011
- 44 Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009. 248–255
- 45 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 770–778
- 46 Hou S, Pan X, Loy C C, et al. Learning a unified classifier incrementally via rebalancing. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019. 831–839
- 47 Mazumder P, Singh P, Rai P. Few-shot lifelong learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2021. 35: 2337–2345

Generalized representation of local relationships for few-shot incremental learning

Yifan ZHAO¹, Jia LI^{2*} & Yonghong TIAN¹

1. *School of Computer Science, Peking University, Beijing 100871, China;*

2. *State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China*

* Corresponding author. E-mail: jjiali@buaa.edu.cn

Abstract Few-shot learning is a booming research area in computer vision that aims to recognize and grasp novel concepts by learning from a limited number of known samples. The existing studies on few-shot learning focus only on recognizing novel categories while neglecting the understanding of base knowledge. This paper introduces the concept of local relationship learning and proposes a generalized representation method for local relationships to address the two major research problems in few-shot incremental learning tasks, i.e., inferior inter-class distinguishability and the difficult generalization of incremental categories. To enhance the distinguishability, this paper first adopts the local spatial relationship to regularize the incremental representation ability. To alleviate inductive biases caused by the lack of data in the incremental process, this paper proposes a spatial generalization prototype generation algorithm, which uses distribution characteristics to quickly generate virtual prototypes and promote the effective representation of samples. Benefiting from the meta-learning training mechanism, this paper proposes a joint locality and generalization awareness incremental learning framework, which effectively alleviates catastrophic forgetting and distinguishability difficulties by combining the local representation of the base category and the fast generalization constraint of the incremental category. Our experimental results demonstrated that the proposed method achieves state-of-the-art results on few-shot incremental learning tasks.

Keywords few-shot learning, incremental learning, local relationship, generalized representation, meta learning