DOI:10.19651/j.cnki.emt.2518227

小样本类增量提示的细粒度车辆识别*

冉烨军 金良琼 罗树霞 李琼忆 陶 永
(贵州民族大学数据科学与信息工程学院 贵阳 550025)

摘 要:在细粒度车辆识别领域,深度学习面临一个挑战:各种新车型源源不断推出,然而我收集并标注数据的能力有限,这会导致"小样本类增量学习问题"问题。针对上述挑战,本文提出了一种新方法,基于提示的小样本类增量学习,旨在使模型在少量新车辆类别样本下既能识别原有类别又能学习新增类别,而无需重新训练或依赖大量原始数据。这种方法结合了提示机制和预训练的视觉转换器(ViT)模型的优势。我们设计了两种提示——域提示和 FSCIL 提示,以解决 FSCIL 中的挑战。在类增量学习中,Stanford Cars 和 CompCars 这两个数据集的平均精度到达了 70.47%和 73.56%,优于目前现有的方法。

关键词:小样本学习;提示学习;视觉转换;原型分类器
中图分类号:TN014 **文献标识码**:A **国家标准学科分类代码**:510.4050

Fine grained vehicle recognition with small sample class incremental hints

Ran Yejun Jin Liangqiong Luo Shuxia Li Qiongyi Tao Yong
(Guizhou University for Nationalities, School of Data Science and Information Engineering, Guiyang 550025, China)

Abstract: In the field of fine-grained vehicle recognition, deep learning faces a challenge: various new car models are constantly being introduced, but my ability to collect and annotate data is limited, which can lead to the problem of "small sample class incremental learning". In response to the above challenges, this article proposes a new method based on prompt based small sample class incremental learning, aiming to enable the model to recognize existing categories and learn new categories with a small number of new vehicle category samples, without the need for retraining or relying on a large amount of raw data. This method combines the advantages of prompt mechanisms and pre trained visual transformer (ViT) models. We have designed two types of prompts-domain prompts and FSCIL prompts-to address the challenges in FSCIL. In class incremental learning, the average accuracy of Stanford Cars and CompCars datasets reached 70.47% and 73.56%, respectively, which is superior to current existing methods.

Keywords: small sample learning; prompt learning; visual transformation; prototype classifier

0 引 言

细粒度车辆分类识别技术作为智能交通管理[1]的核心算法之一。通过精确识别车辆的型号、制造商和生产年份等信息,ITS 能够更有效地管理交通流量,提高道路安全性,并优化交通规划。在执法和安全领域,细粒度汽车分类识别有助于追踪和识别涉案车辆。例如,在犯罪调查中,警方可以通过识别特定车型来缩小嫌疑范围,或者通过车牌伪造检测[2]来打击犯罪行为。

细粒度车辆识别是计算机视觉中的一个复杂任务,它比普通图像分类更难,因为不同车辆类别间相似度高,而同一类别内的车辆又差异大。近年来,随着深度学习技术的不断进步[3-4]以及大规模车辆数据集的涌现[5],细粒度车辆识别领域取得了显著的进展[6]。通过在大规模数据集上全面训练深度神经网络(DNN),基于深度学习在大规模数据集上训练 DNN 能挖掘数据潜在信息,但 ResNet 等模型因容量有限,有时不能满足新需求。ViT 等[7]预训练大型模型的出现,因其可扩展性和可调整性强,在很多方面超过了 ResNet。

近年来,在类增量学习中,由于新类别数据稀少,传统 ViT 模型[7]训练面临挑战。为解决此问题,我们引入了基于提示学习的新方法,借鉴 NLP 中的策略,通过可学习的

收稿日期:2025-03-02
* 基金项目:国家自然科学基金(62062024)、贵州省高等学校大数据分析与智能计算重点实验室(黔教技[2023]012 号)、贵州省高等学校智能算法与智能软件协同创新团队(黔教技[2023]061 号)项目资助

指令让模型利用已有知识，而非重新学习。这样，模型能在不遗忘旧类别的情况下，以较少的计算快速掌握新类别。

现有的小样本类增量学习方法在处理细粒度车辆识别时面临诸多挑战，尤其是灾难性遗忘和计算复杂度高的问题。传统的微调方法需要重新训练整个模型，计算成本高昂；重放策略虽然能够缓解遗忘问题，但需要存储大量旧类样本，增加了存储负担；元学习方法虽然能够快速适应新任务，但在新类别加入时表现不稳定。针对这些问题，本文提出了基于提示的小样本类增量学习方法（PT-FSCIL），通过引入域提示（D-Prompt）和 FSCIL 提示（F-Prompt），在不改变预训练模型参数的情况下，引导模型适应新任务，同时通过原型分类器简化分类过程，显著降低了计算复杂度。

在此背景下，提出了结合提示来完成细粒度小样本车辆识别。主要贡献总结如下：

1）本文提出的 D-Prompt 提示与 F-Prompt 提示，提高了模型的泛化能力和判别能力。

2）为了提高 D-Prompt 提示和 F-Prompt 提示之间的正交性，我们引入了提示规则策略。该策略利用 Frobenius 规范来量化正交性，从而使模型的泛化能力更高。

3）在两个公共的细粒度车辆数据集上进行了实验，验证了所提出的 PT-FSCIL 方法的有效性。

1　相关工作

1.1　细粒度车辆识别

车辆制造与型号识别（VMMR）是细粒度图像分类的难题，因类内多样、类间相似。常见方法是先定位关键兴趣区域（ROI），再用深度卷积神经网络（DCNN）提取特征，以减少尺寸和背景干扰。

具体而言，Yu 等[8]和 Liu 等[9]采纳了区域提议网络（RPN）框架来精准定位 ROI，并进一步利用深度卷积神经网络（CNN）的强大能力来抽取这些区域的特征。Fang 等[10]则另辟蹊径，他们依据 CNN 特征映射的响应模式，智能地识别出图像中的关键部分，并整合这些部分在不同层次上的连接特征，以此为基础训练支持向量机（SVM）分类器。

此外，还有研究团队选择直接以全局图像作为输入，例如，Hu 等[11]提出了一种创新的池化策略，该策略融入了可学习的空间权重掩码，以精细指导神经网络的特征学习过程。而 Xiang 等[12]的研究则深入到了组件间的拓扑关系层面，他们不仅检测并整合了相关组件的特征，还通过引入拓扑约束，精确估算了这些组件之间复杂关系的概率分布。

1.2　类增量学习

少样本类别增量学习（few-shot class incremental learning，FSCIL）[13]是一个活跃的研究领域如图 1 所示，它涵盖了多种方法，包括传统机器学习、元学习、特征空间处理、重放策略以及动态网络结构设计等。在传统机器学习框架下，FSCIL 的研究往往从基础策略出发，如监督学习算法和统计分布分析。例如，文献[14]将半监督学习引入 FSCIL，通过在每个增量学习阶段引入少量未标记数据点，来可以来缓解新类别样本稀缺而导致过拟合的问题。然而，半监督学习在 FSCIL 中面临挑战，如未标记数据质量与标注数据的兼容问题、算法复杂度增加导致的计算成本上升，以及未标记数据可能引入的分布偏移降低伪标签可信度。同时，FSCIL 还需适应数据分布变化，并克服新类样本有限的难题，常通过数据增强技术来提升模型性能。另外，文献[15]采用 Polya-伽玛分布来模拟数据的不确定性，通过高斯过程对数据进行建模，从而提高了模型对数据分布的适应性。但统计分布方法也存在局限性，比如对于极端不平衡的数据分布或高度复杂的分布形态，模型可能难以准确估计数据的真实分布，进而影响学习效果。

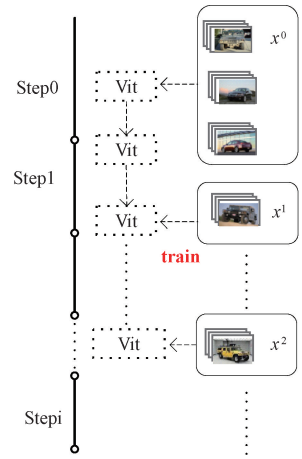


图 1　增量学习步骤模型图

Fig. 1　Incremental learning step model diagram

元学习方法则借鉴了少样本学习的思想，并侧重于原型学习策略的应用。这类方法通过构建原型来表示每个类别的特征，从而提高了新类别样本的识别能力。例如，DeepSLDA 方法[16]在每个增量学习阶段选择性地更新模型参数，以防止过拟合，并通过最小化新旧类别原型之间的余弦相似性来最大化它们之间的分离度。然而，元学习方法也存在一些局限性。如原型选择依赖初始样本，样本不佳则影响模型；类别增多时，原型空间复杂度和计算量激增，影响实时性和可扩展性。

在特征和特征空间处理方面，一些研究从特征解耦和子空间表示的角度对 FSCIL 问题进行建模。例如，LUCIR 方法[17]通过特征解耦技术，将特征分解为与类别相关的部分和与类别无关的部分，从而提高了模型的泛化能力。然而，在特征解耦的过程中，有可能会丢失一些原本对分类有益的信息。这些信息可能因解耦操作的近似性或简化处理而被遗漏，进而对模型的分类能力造成负面影响。

动态网络结构设计是 FSCIL 领域的另一个重要方向。例如，有研究提出了使用神经气体网络（NG）来学习知识表

示的拓扑结构。这种方法通过保持 NG 的拓扑稳定性来防止对旧类别的遗忘,同时利用 NG 的动态增长来提高新类别样本的表示能力。这种方法通过动态调整网络结构来适应不断变化的类别分布,从而提高了模型的适应性。然而,动态网络结构设计虽具优势,但也存在挑战,算法设计不当可能增加模型复杂度和计算量,影响训练与推理效率。

重放策略是 FSCIL 中常用的另一种方法。例如,直接重放方法 iCaRL[18] 通过存储来自旧类别的样本并在训练过程中使用它们来进行回放。这种方法通过保留旧类别的信息来防止遗忘。但重放策略也有局限性,直接重放需大量存储空间且可能因样本选择不均导致偏见;生成重放则依赖生成模型质量,模型不佳时重放样本可能不准确,影响模型性能。

近年来,基于预训练视觉与语言 Transformer 的少样本增量学习方法通过多模态提示机制实现了跨模态知识迁移。然而,此类方法依赖文本-图像对齐标注,难以应用于缺乏语言描述的细粒度车辆识别任务。相比之下,PT-FSCIL 专注于单一视觉模态,提出域提示(D-Prompt)与任务特定提示(F-Prompt)的双重机制,直接优化视觉特征空间。此外,PT-FSCIL 通过原型分类器与正交性约束,显著降低了模型复杂度,避免了多模态融合的计算开销。

2 方 法

2.1 问题描述

在细粒度车辆识别中,面临少样本类增量学习(FSCIL)的挑战,包含初始训练及后续的连续增量学习步骤,如图 2 所示。此框架针对车辆识别,要求模型准确预测每张仅含一辆车的图像的细粒度类别。

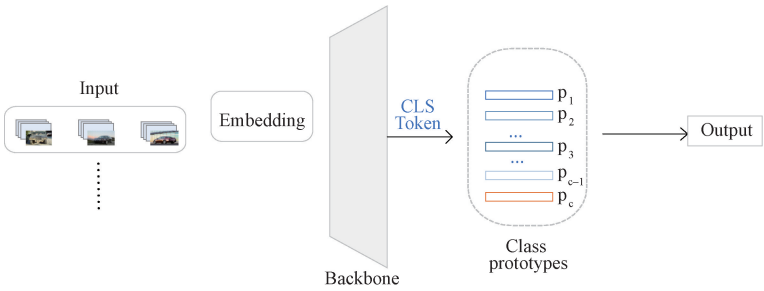


图 2 小样本类增量图像识别示意图

Fig. 2 Schematic diagram of incremental image recognition for small sample classes

PT-FSCIL 问题可以被假设成为一个包含(i+1)个阶段的学习任务,每个阶段对应一个任务。这里,我们定义了从阶段 0 到阶段 i 的一系列训练集,初始训练集和标签集分别表示为 $X^{(0)}$ 和 $Y^{(0)}$,在这里,对于第 t 个增量学习步骤($t=1,2,\cdots,N$)的训练集和标签集分别表示为 $\{(\boldsymbol{x}_1^{(t)},y_1^{(t)}),(\boldsymbol{x}_2^{(t)},y_2^{(t)}),\cdots,(\boldsymbol{x}_{mk}^{(t)},y_{mk}^{(t)})\}$,其中 $y_i^{(t)}\in Y(t)$,$X^{(t)}$ 是一个 m-way k-shot 训练集,意味着这一步有 m 个新的车辆类,每个类的训练样本数为 k。标签集 $Y(t)$ 可以表示为 $\{y_1^{(t)},y_2^{(t)},\cdots,y_m^{(t)}\}$,其中 m 是新类的数量,模型的更新过程可以表示为 $M_t=T(M_{t-1},\boldsymbol{X}^t,\boldsymbol{Y}^t)$,其中 T 表示训练过程。更新后的模型 M_t 通过 $X^{(t)}$ 和 $Y^{(t)}$ 进行训练,同时保持对之前学习的类 $Y^{(0)}\cup Y^{(1)}\cdots\cup Y^{(t-1)}$ 的识别能力。在每个增量学习步骤之后,使用测试集对模型 M_t 进行评估。测试集包含所有之前学习的类 $Y^{(0)}\cup Y^{(1)}\cdots\cup Y^{(t-1)}$ 的样本,以检验其对当前为止遇到的所有类别,从而全面评估模型在类别增量学习过程中的整体性能。

2.2 提示池框架

本文借鉴 NLP 领域的提示学习策略,通过可学习的指令模板引导模型复用预训练表征[19],而非重构特征空间。该方法在基类特征迁移过程中,利用动态优化的提示向量实现跨领域知识迁移,在保持基类识别能力的同时,将新类样本的领域适应过程转化为低维提示空间优化问题,显著降低了模型更新的计算复杂度。具体而言,由于大型预训练模型所学习的知识和捕获的模式广泛,加入了提示学习这种机制后,使得模型在处理零样本和稀缺热点数据方面展现出较好的性能。在本讨论中,聚焦于两种关键提示设计策略:D-Prompt 和 F-Prompt(如图 3 所示),在图 2 的 Backbone 的基础上嵌入了 Domain-Prompt、FSCIL-Prompt,最后使用原型网络分类器分类。

D-Prompt 作为一种域特定的提示机制,旨在通过将其融入预训练模型中,使模型的特征表示能力能够针对当前数据集的特定域进行适配。这种机制通过精细地调整模型,使其更好地捕捉域内数据的独特特征。然而,F-Prompt 则侧重于任务特定性。它通过向模型中添加与少量样本任务紧密相关的提示,不仅传递了域信息,还额外附加了任务层面的信息。该设计能让模型更灵活的适应新类别,同时在新任务中保持对旧类别的识别,为实现两种机制的有效嵌入,我们采用前缀调整法,将它们嵌入到 Transformer 模型的适当 self-attention layer 中,既可以确保信息传递又不改变模型结构。此外,我们还对传统的 softmax 分类器进行了替换,引入了原型分类器。原型分类器无需依赖梯度反向传播进行优化,而是根据样本特征输出直接计算原型进行分类决策。

首先给出一些相关的符号和定义,由于 ViT 模型是由

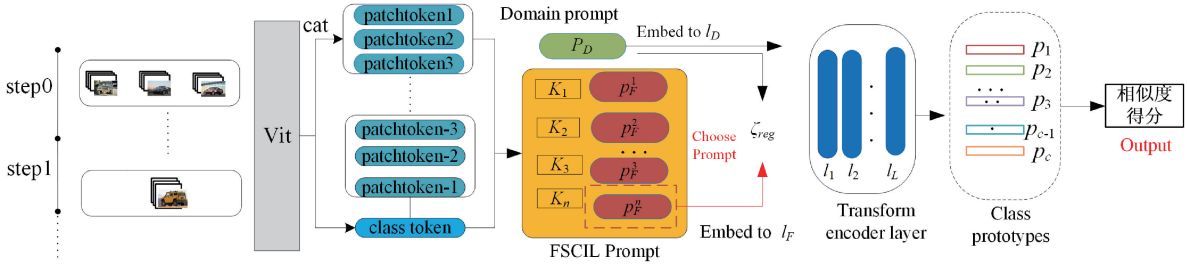


图 3　基于 Vision Prompt 的小样本细粒度车辆识别模型整体架构图

Fig. 3　Shows the overall architecture diagram of a small sample fine-grained vehicle recognition model based on Vision Prompt

N 个连续的多头注意力机制(multi-head self-attention)组成。每一层都有相同的函数我们用 f^l 表示。函数的输入和输出序列用 $X^{(l)}, Y^{(l)} \in \boldsymbol{R}^{N\times D}$ 表示。其中 N 代表是分块的总数，D 代表的是嵌入的维度。通过把 D-Prompt 和 F-Prompt 两种提示嵌入到输入 $X^{(l)}$ 上，其中两种提示的维度需要与 $X^{(l)}$ 一致。

1) D-Prompt 提示

D-Prompt 该部分提示拥有对整个细粒度汽车数据集信息。用张量 $\boldsymbol{d} \in \boldsymbol{R}^{L_d\times D}$ 表示，其中 L_d 代表序列长度，D 代表嵌入的维度，假设想把 D-Prompt 附加到第 l 层上，我们定义了以下函数。

$$h_d^l = f(\boldsymbol{d}, \boldsymbol{h}^{(l)}) \tag{1}$$

其中，f 定义了如何将提示附加到隐藏层中的方法，我们用 $h^{(l)}$ 表示第 l 个 MSA 层的输入嵌入特征。

2) F-Prompt 提示

F-Prompt 是对任务特定的，这意味着它们需要整合并反映随着学习阶段不断累积的知识。为了明确这一点，定义为 $F=\{f_t\}_1^T$ 是一组任务相关的参数，其中 $f_t \in \boldsymbol{R}^{L_f\times D}$，$L_f$ 代表的是序列 F-Prompt 长度，D 代表嵌入维度。T 代表任务总数。在训练期间，每个任务的 ID 是可识别的，我们设计了掩码隐藏与当前不相关的会话提示。同时，f_t 与特定的键 $k_t \in \boldsymbol{R}^D$ 相关联。

为了确保匹配到最适合的 F-Prompt，建立一个查询函数 $q(\cdot)$ 旨在减小在训练期间与键 k_t 之间的距离。

$$\mathcal{L}_{\text{dist}}(\boldsymbol{x}, \boldsymbol{k}_t) = g(q(\boldsymbol{x}), \boldsymbol{k}_t) \tag{2}$$

其中，x 是输入图像，$q(\cdot)$ 可以是各种计算相似度距离的函数，为了简化处理，本文决定采用一个已经预训练完成且参数保持冻结的视觉(ViT)模型作为 $q(\cdot)$，该模型的输出基于其头部的(CLS_Token)向量，能够有效地捕捉图像的高层特征。

在测试期间，为了选择合适的 f_t，本文通过计算输入样本与 $\boldsymbol{K}_t$ 之间的距离。并将特定任务的知识加入到模型中。

$$f_t = \underset{t=0,\cdots,T}{\operatorname{argming}}(q(\boldsymbol{x}), \boldsymbol{k}_t) \tag{3}$$

其中，f_t 代表的是与输入样本最接近的提示，$q(\boldsymbol{x})$ 是查询函数，另外，为了提高训练速度，在开始前使用 f_{t-1} 初始化 f_t，然后继续训练。

3) 正则化机制模块

传统方法例如基于梯度投影通常依赖于隐含的正交性假设，这种假设可能不够明确，正交性控制不够直接，导致提示之间的干扰或知识覆盖。而基于 Frobenius 范数的正则化项(公式(4))相比传统方法中隐含的正交性假设，这种显式定义能更精准地控制提示空间的独立性，避免模型训练后出现灾难性遗忘，提供更精准的控制能力。而且对 F-Prompt 提示的更新不会覆盖或干扰 D-Prompt 提示所表示的全面知识。正则化损失函数可以定义如下：

$$\mathcal{L}_{reg}(i) = \| \boldsymbol{d} \times \boldsymbol{f}_t^{\mathrm{T}} \| \tag{4}$$

其中，$\|\cdot\|$代表 Frobenius norm，$\mathcal{L}_{reg}(i)$ 来衡量 $\boldsymbol{d}$ 和 $\boldsymbol{f}_t$ 之间的差异。如果 $\mathcal{L}_{reg}(i)$ 的值为零的话，则代表是正交的，如果值不为零的话，代表 $\boldsymbol{d} \times \boldsymbol{f}_t^{\mathrm{T}}$ 包含了更多相似提示的知识。对于任意阶段 i，总损失函数定义如下：

$$\min_{d, f_t} \mathcal{L}(f(\boldsymbol{x}), y) + \lambda\, \mathcal{L}_{dist}(\boldsymbol{x}, \boldsymbol{k}_t) + a\, \mathcal{L}_{reg}(i) \tag{5}$$

其中，$\boldsymbol{x} \in D_{test}^{i}$，$f(\boldsymbol{x})$ 代表模型的输出，总损失函数包括 $\mathcal{L}_{dist}$，以及正则化损失 $\mathcal{L}_{reg}$，a 和 λ 代表对总损失的权重。

4) ViT 中嵌入提示

如图 2 所示，在没有加入提示符作为一种外部信号引入时，模型无法针对特定域或功能进行自适应调整，可能无法充分捕捉特定域或功能的特征。相比而言，如图 3 嵌入了提示符旨在调控 Transformer 模型内部的注意力机制，特征表示能够更专注于特定域或功能的关键特征。接下来，我们将细致探讨如何在 ViT 模型中整合提示信息。

具体而言，在 ViT 的每一层 l 中，$x \in \boldsymbol{R}^{N\times D}$ 表示 $\boldsymbol{N}$ 个 $\boldsymbol{F}$ 维的特征向量序列，ViT 是一个函数 $Z: \boldsymbol{R}^{N\times D} \rightarrow \boldsymbol{R}^{N\times D}$，由 $Z_1, Z_2, \cdots, Z_l$ 总共 l 层组成，如式(6)所示。

$$Z_l(x) = f_l(A_l(\boldsymbol{x})) \tag{6}$$

其中，函数 $f_l(\cdot)$ 是独立于其他的变换特征函数，$A_l(\cdot)$ 是注意力机制，会关注到与自己相关的重要信息。输入会首先经历线性变换，被拆解成 3 个关键组成部分：查询 $\boldsymbol{Q}^{(l)}$，键 $\boldsymbol{K}^{(l)}$ 和值 $\boldsymbol{V}^{(l)}$。这些组件随后参与注意力权重的计算以及输出特征的生成。在原始的 ViT 架构中，自注意力机制的实现过程可以概括为上述步骤，公式定义如式(7)所示。

$$\text{Attention}(\boldsymbol{Q}^{(l)},\boldsymbol{K}^{(l)},\boldsymbol{V}^{(l)}) = \text{softmax}((\frac{\boldsymbol{Q}^{(l)}(\boldsymbol{K}^{(l)})^{\mathrm{T}}}{\sqrt{d_k}})\boldsymbol{V}^{l}) \tag{7}$$

为有效融合提示到 ViT 模型,本文采用前缀调优技术,在输入序列前加特定任务的虚拟令牌(Prefix),训练时仅优化这些令牌参数,保持 ViT 其他参数不变。对于 D-Prompt 和 F-Prompt,本文将它们嵌入到键(K)和值(V)中。为了确保提示符能够与键和值张量的维度完美匹配,本文对其进行了适当的调整,形成了组合提示符 $P = concat(\boldsymbol{P}_k,\boldsymbol{P}_v)$。随后,我们修改了自注意力运算流程,以便能够充分利用这一新的组合提示符。

$$\boldsymbol{K}' = [\boldsymbol{P}_k;\boldsymbol{K}^{(l)}],\boldsymbol{V}' = [\boldsymbol{P}_v;\boldsymbol{V}^{(l)}] \tag{8}$$

$$Attention(\boldsymbol{Q}^{(l)},\boldsymbol{K}^{(')},\boldsymbol{V}^{(')}) = \text{softmax}((\frac{\boldsymbol{Q}^{(l)}(\boldsymbol{K}^{(')})^{T}}{\sqrt{d_k}})\boldsymbol{V}') \tag{9}$$

其中,$\boldsymbol{P}_k$ 与 $\boldsymbol{P}_v$ 代表提示的键和值分量,$concat$ 表示拼接操作,通过把这些额外的键和值与原始的输入的键和值拼接就可以影响到注意力权重的和输出特征的计算。从而能使模型更具有适应性和泛化能力,提示的算法如算法 1 所示。

Algorithm 1: Prompt training and model optimization for FSCIL

Input: Training dataset $D = \{(x_i,y_i)\}_{i=1}^{N}$, Pretrained ViT model M(L layers), Domain prompt d(embed at layer l_d), FSCIL prompt f(embed at layer l_f), IL sessions S, Feature transformation function F

Output: Optimized d and f

Initialize d, f and the corresponding prompt key K
for each session i in S do
  for each $D_j = (x_j,y_j)$ in D do
    $Z^{(0)} = x_i$
    for each layer l in L do
      if l in l_d then
        $Z^{(l)} = F^{(l)}(Z^{(l-1)};d^{(i)})$
      else if l in l_f then
        $Z^{(l)} = F^{(l)}(Z^{(l-1)};f^{(i)})$
      else
        $Z^{(l)} = F^{(l)}(Z^{(l-1)})$
      end if
    end for
Calculate $\mathcal{L}_{dist}$ using Eq. 2
    Update the parameters of two prompts with $\mathcal{L}_{reg}$ using Eq. 4
  end for
end for
return d, f

2.3 原型分类器

原型网络是一种数据分类算法,它通过样本与代表数据特征的原型向量的距离来进行分类。这里使用预训练模型 Vit 的 CLS_Token 加权平均值作为类别原型。用公式定义如下:

$$\boldsymbol{p}_i = \frac{\sum_{j=1}^{n}\boldsymbol{w}_j \cdot f(\boldsymbol{x}_j)}{\sum_{j=1}^{n}\boldsymbol{w}_j} \tag{10}$$

其中,$\boldsymbol{p}_i$ 是类别 i 的原型向量,$\sum_{j=1}^{n}\boldsymbol{w}_j \cdot f(\boldsymbol{x}_j)$ 表示的是所有样本特征向量 $f(\boldsymbol{x}_j)$ 与其权重 $\boldsymbol{w}_j$ 的点积之和。$\sum_{j=1}^{n}\boldsymbol{w}_j$ 表示所有样本权重之和,用作归一化因子,以确保原型向量的长度不会因权重的大小而改变。这里的权重我们根据预测的概率进行相应设定,概率值越高对原型贡献率越大,反之亦然。我们这里采用的欧氏距离分类,将样本分配到离它最近的原型,公式定义为:

$$\hat{y} = \underset{i=1,\cdots,k}{\text{argmin}} \| f(\boldsymbol{x}) - p_i \|_2 \tag{11}$$

其中,x 为输入样本,k 为总类别数。

3 实验与分析

在本节中,通过广泛的实验验证了 PT-FSCIL 方法的有效性,3.1 节介绍了数据集,3.2 节介绍了实验的设置,3.3 节通过消融实验验证了方法的有效性。

3.1 数据集

如表 1 所示,为了验证 PT-FSCIL 方法嵌入模型的有效性,本文采用了两个实验数据集,Stanford Cars[20],CompCars[21]这两个公开数据集涵盖了不同天气下和光照条件下的图像(图 4)。根据类增量学习问题,将数据集划分为两个交集为空的集合 Y_0,Y_{new} 分别代表初始阶段训练集和增量学习集,然后将 Y_{new} 分成 Y_1,Y_2,Y_3,…,Y_N 每个类别包含 m 辆车,用于增量学习步骤。

表 1 两个数据集的详细统计数据

Table 1 Detailed statistical data of three datasets

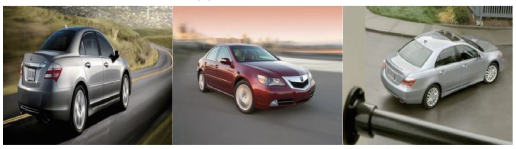
类别	Stanford Cars	Comp Cars
Total classes	196	431
Total training samples	8 144	16 016
Total testing samples	8 041	14 939
Initial classes	100	311
Incremental steps	12	12
New classes per step	8	10
Training samples per class	5	5

3.2 实验设置

1) 实验细节

本文的实验是在 pytorch 2.0.0,具有一张 RTX 4090D

(a) Stanford Cars



(b) CompCars

图 4　两个不同类型的汽车数据集

Fig. 4　Two different types of automotive datasets

(24 GB)×1 的计算机平台上训练，预训练模型本文使用 ViT Base/16，输入的图片通过剪裁后为 224 pixel×224 pixel×3 pixel，即 $H=224$，$W=224$，$C=3$。D-Prompt 与 F-Prompt 的大小设置为 10，其中类原型的维数与每个数据集的类数量保持一致，学习率设置为 1×10^{-5}，其次我们利用交叉熵损失进行误差计算，并使用 Adam 优化器来最小化损失函数。

2) 评价指标

此任务可以视为细粒度车辆识别小样本分类任务，以 AA(average accuracy)和 PD(performance dropping)作为评价指标。在每一次增量学习阶段后我们记录前一次的准确率和所有增量学习阶段的平均准确率，这两个指标的计算方式如下：

$$\begin{cases} AA = \dfrac{1}{n}\sum_{i=1}^{n} A_i \\ PD = A_0 - A_N \end{cases} \tag{12}$$

其中，n 代表增量步骤有几次，A_i 代表第 i 次增量学习步骤的精度，A_0 代表第一次增量学习的分类精度，A_N 代表最后一次分类的精度。

3.3　消融实验分析

1) 对比实验

为了全面验证提出的 PT-FSCIL 方法的性能。本文选取一些最先进和具有代表性的方法如 Imprint[22]、CEC[23]、iCaRL[18]、LUCIR[17]、DeeSIL[24]、Deep SLDA[16]、M-FCIL[25]。其中 CEC、Imprint 和 PL-FSCIL 是专为 FSCIL 设计的能验证 PT-FSCIL 在基础性能上的提升，iCaRL[13]和 LUCIR[21]是经典的增量学习标准对比基准。为了使实验结果公平，本文还额外引入了 M-FCIL 方法和与最新的 CLIP 模型进行基准测试。

2) Stanford Cars 数据集结果展示

如表 2 中展示了训练后的实验结果，从数据来看，PT-FSCIL 在平均准确率(AA 70.47%)上显著优于其他方法(如 Imprint 67.87%、M-FSCIL 69.21%)，尤其在中期学习阶段(session 5～10)表现更稳定(如 session 5 达 73.25%，远高于同期 iCaRL 的 63.4%、CEC 的 55.57%)。虽然其最终性能下降(PD 25.58%)略高于部分方法(如 Imprint 22.7%)，但其长期抗遗忘能力在(session 6～10)期间仍保持较高水平(72.4%～67.86%)，表明其通过渐进式训练有效平衡了新任务学习与旧知识保留，综合性能领先。

3) CompCars 数据集结果展示

如图 5 的实验结果显示，PT-FSCIL 方法在初始训练阶段达到了 85.09%的准确率领先，随着增量学习的增加，优势持续扩大。相对于其他方法，iCaRL 和 LUCIR 因模型的网络权重改变导致旧知识的遗忘严重，性能下降较快。而 Deep SLDA 和 DeeSIL 虽然考虑到解耦模型，但他们对细粒度提取的特征不足，Imprint 和 CEC 方法则受限于弱判别性特征带来的分类误差。基于此，PT-FSCIL 则通过两种提示有效区分了相似度高的车辆类别，维持了新旧知识平衡，在少样本情况任保持稳定学习。

表 2　Stanford Cars 数据集每次增量学习阶段的准确率和平均准确率以及下降率

Table 2　Accuracy, average accuracy, and decline rate of each incremental learning stage in the Stanford Cars dataset

%

Methods	Accuracy in each session ↑													AA ↑	PD ↓
	0	1	2	3	4	5	6	7	8	9	10	11	12		
Imprint[22]	81.07	78.16	75.57	72.89	70.86	68.17	67.01	65.26	63.36	61.76	60.26	59.66	58.37	67.87	22.70
iCaRL[18]	77.40	72.70	70.60	67.20	65.90	63.40	62.90	61.90	60.50	60.60	60.10	59.69	59.22	64.77	**18.18**
CEC[23]	73.07	68.88	65.26	61.19	58.09	55.57	53.22	51.34	49.14	47.45	44.32	42.25	40.37	54.62	32.70
Deep SLDA[16]	80.32	76.20	75.33	74.40	73.42	73.12	71.16	70.12	66.99	64.10	63.21	59.12	56.88	69.56	23.44
DeeSIL[24]	78.45	74.98	71.30	69.28	67.17	66.16	65.43	64.25	63.66	61.01	59.44	58.22	56.12	65.80	22.33
LUCIR[17]	80.45	75.98	72.30	70.28	68.17	65.16	64.43	63.25	60.66	60.01	59.44	58.32	57.12	65.81	23.33
M-FSCIL[25]	**82.10**	**81.42**	**77.52**	**75.30**	**74.22**	**73.90**	65.87	64.02	63.90	61.49	60.20	59.99	59.87	69.21	22.23
PT-FSCIL	81.89	75.40	74.75	72.22	73.22	73.25	**72.40**	**70.28**	**67.98**	**68.33**	**67.86**	**62.22**	**56.31**	**70.47**	25.58

4) 提示学习的实验结果

表 3 可以直观展示 3 个方法对精确率和下降率方面的

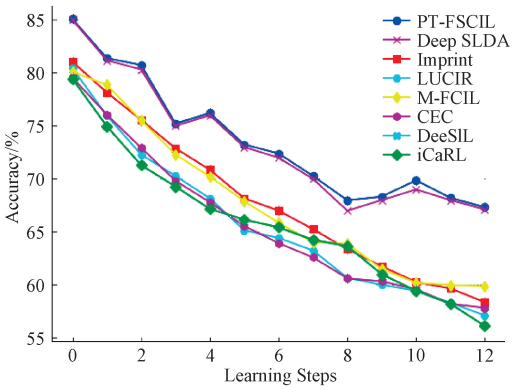


图 5 CompCars 数据集与不同方法随训练阶段次数增加的准确率

Fig. 5 Accuracy of the CompCars dataset and different methods as the number of training stages increases

贡献,表中的记号代表的是在实验部分包含该模块,从表格可以看出,单独使用原型分类器效果低下,加入所提出的 DP 模块,性能提升了 1.33%,说明该模块是有效的,假设只用 FP 模块和 PC 模块性能有明显提升,相比只用原型分类器提升了 2.57%,3 个模块同时使用时,能使这两个提升使模型到达最佳性能。

表 3 在 Standfor Cars 数据集上进行消融实验,三个不同组件 D-Prompt,F-Prompt, Prototype Classifier 对性能的影响

Table 3 Conduct ablation experiments on the Standfor Cars dataset to investigate the impact of three different components, D-Prompt, F-Prompt, and Prototype Classifier, on performance

FP	DP	PC	AA/% ↑	PD/% ↓
√	√	√	70.47	25.58
√		√	70.12	16.46
	√	√	70.19	17.22
		√	68.86	20.88

如表 4 实验结果显示,VPT 在 3 个数据集上取得了最高分类精度,但其计算代价与参数显著高于其他模型,表明其性能依赖复杂运算支撑。ResNet18 在 CIFAR-10 上的精度 91.68%低于 VPT 5.12%,说明固定参数模型难以适配多任务需求。D-Prompt 在 CIFAR-10 和 Flowers-102 的参数量仅为 23.05 K 和 93.80 K,降幅达 95%以上,而计算量(17.26 G)比 VPT 降低 2.5%。但在 STL-10 与 Flowers-102 上的精度损失仅 2.6%和 2.0%,反映其在精度与效率间实现更优平衡。

表 4 对比不同骨干网络上不同数据集的 Top1 精确率(Acc),可训练参数(Params)和复杂度(GFLOPs)

Table 4 Compares the Top1 accuracy (Acc), trainable parameters (Params), and complexity (GFLOPs) of different datasets on different backbone networks

Models	CIFAR-10		STL-10		Flowers-102		GFLOPs
	Acc/% ↑	Params ↓	Acc/% ↑	Params ↓	Acc/% ↑	Params ↓	(Mac) ↓
ResNet18[26]	91.68	11.18 M	87.83	11.18 M	84.65	11.23 M	**1.82** G
VPT[27]	**96.89**	99.85 K	**98.98**	99.85 K	**97.38**	308.84 K	17.71 G
D-Prompt	91.82	**23.05** K	96.35	**92.08** K	95.35	**93.80** K	17.26 G

图 6 提供了消融实验更详细的可视化视图,展示每次模型训练阶段的准确性。很明显,在整个学习过程中,DP AND FP AND PC 的组合在大多数学习步骤中保持了最高的准确率,尽管它的准确率也有所下降,DP AND PC 的组合紧随其后,表现出较好的准确率维持能力。FP AND PC 的组合在中间位置,其准确率下降趋势与 DP AND PC 相似,但整体准确率略低。单独使用 PC 方法的准确率在所有步骤中都是最低的,尤其是在学习步骤 8 之后,准确率下降得更为明显。因此,只有 3 个组件的共同使用在准确率和下降率方面才更有竞争力。

5)正则化系数的影响

消融研究揭示了快速正则化系数 α 在平衡领域知识与任务 特定知识方面的重要作用。如表 5 所示,当 α 设置为适度的值时,模型的平均准确率(AA)有所提升:对于 Standfor Cars 数据集,$\alpha=0.002$ 时达到最优;而对于 CompCars 数据集,则在 $\alpha=0.020$ 时表现最佳。这表明,

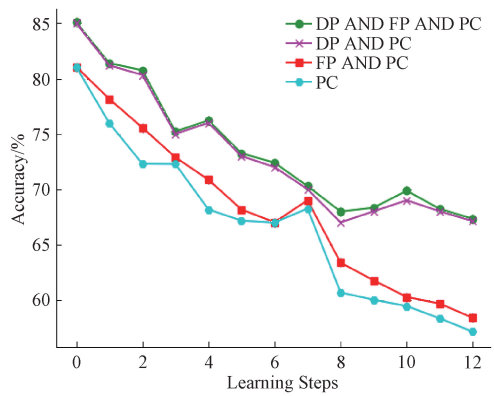


图 6 不同组件在 Standfor Cars 数据集上的影响

Fig. 6 Impact of different components on the Standfor Cars dataset

适量的快速正则化有助于提升学习精度。然而,当正则化过度(例如 $\alpha=0.050$)时,AA 会略有下降,这提示存在一

个临界值，超过此值后，知识的区分能力反而会下降，这些发现验证了快速正则化机制在促进知识正交性方面的有效性，从而增强了模型在区分一般知识与特定任务知识方面的能力。

6）现实场景模拟

为了验证模型在真实场景下的性能，本文通过在 Stanford Cars 数据集上添加不同类型的噪声进行模拟类增量学习实验（如图 7 所示）。实验分为 3 个阶段：无噪声、标签噪声、遮挡＋高斯噪声。结果显示，随着训练增加，所有方法的准确率都下降。M-FSCIL 在无噪声阶段表现最好；PT-FSCIL 在标签噪声阶段鲁棒性较强；面对复合噪声，PT-FSCIL 初始准确率最高，但所有方法最终都跌破 40%，其中 CEC 受影响最大。PT-FSCIL 因元学习驱动的原型扩展策略在后期阶段衰减斜率低，保持了较好的可持续学习。

表 5　正则化系数的取值对 Standfor Cars 和 CompCars 数据集准确率和下降率的影响

Table 5　Impact of regularization coefficient valueson the accuracy and descent rate of Standfor Cars and CompCars datasets　%

α	Stand for Cars		CompCars	
	AA ↑	PD ↓	AA ↑	PD ↓
0.001	74.40	15.55	73.03	24.45
0.002	**75.36**	14.89	**74.35**	24.39
0.010	75.06	15.65	73.89	**24.08**
0.050	73.80	15.17	73.55	24.16
0.100	73.46	**14.83**	73.39	24.33

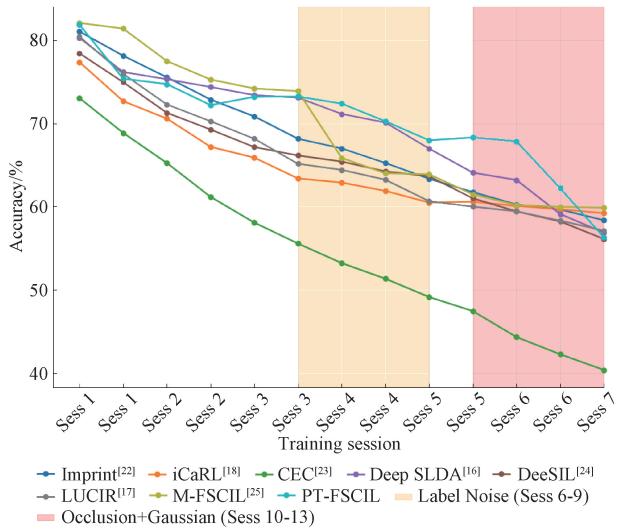


图 7　噪声渐进场景下的性能

Fig. 7　Performance in a noisy asymptotic scene

4　结　　论

本文提出了一种名为 PT-FSCIL 的新方法，该方法利用预训练的视觉 Transformer（ViT）中的提示策略，有效应对细粒度车辆识别领域的少样本类增量学习挑战。PT-FSCIL 模型擅长在数据稀缺的新任务和领域中学习并做出适应。在标准数据集上的测试结果显示，PT-FSCIL 模型表现优异。通过消融实验，本文验证了 PT-FSCIL 方法中各个组成部分的必要性。此外，D-Prompt 和 F-Prompt 的引入显著提升了模型对新数据和任务的特征提取效能。在难以获取充足标签数据的情境下，提示学习显得尤为重要。然而，我们的方法也存在一个潜在的限制：在处理数据分布复杂的场景时，其性能可能会受到影响，且当前的原型分类器可能因过于简单而不足以应对。未来的研究方向将聚焦于优化原型分类器，并探索更高效的提示整合策略，以期进一步强化学习过程。

参考文献

[1]　翟永杰，刘璇，王新颖，等. 基于全局与局部注意力的车辆方位场景识别[J]. 电子测量技术，2024，47(14)：96-107.
ZHAI Y J，LIU X，WANG X Y，et al. Vehicle orientation scene recognition based on global and local attention[J]. Electronic Measurement Technology，2024，47(14)：96-107.

[2]　过鑫炎，朱硕，孙佳豪，等. 基于注意力机制融合特征的车辆目标检测方法[J]. 电子测量技术，2024，47(9)：52-60.
GUO X Y，ZHU SH，SUN J H，et al. Vehicle object detection method based on attention mechanism fused features[J]. Electronic Measurement Technology，2024，47(9)：52-60.

[3]　王子鹏，孙鹏，程耀瑜，等. 基于深度学习增强的散射介质成像重建[J]. 国外电子测量技术，2024，43(12)：

1-7.
WANG Z P, SUN P, CHENG Y Y, et al. Deep learning-enhanced scattering medium imaging reconstruction[J]. Foreign Electronic Measurement Technology,2024,43(12): 1-7.
[4] 刘宇鹏,雷少波,樊浩研,等.基于深度强化学习的无线多址接入方法研究[J].国外电子测量技术,2024,43(8):10-16.
LIU Y P, LEI SH B, FAN H Y, et al. Research on wireless multi-access method based on deep reinforcement learning [J]. Foreign Electronic Measurement Technology,2024,43(8): 10-16.
[5] 张武,刘秀清.基于改进YOLOv5的SAR图像飞机目标细粒度识别[J].国外电子测量技术,2024,43(6):143-151.
ZHANG W, LIU X Q. Fine-grained recognition of aircraft targets in SAR images based on improved YOLOv5 [J]. Foreign Electronic Measurement Technology,2024,43(6):143-151.
[6] 李冰锋,冀得魁,杨艺.基于改进MMAL的细粒度图像分类研究[J].电子测量技术,2024,47(17):172-179.
LI B F, JI D K, YANG Y, et al. Research on fine-grained image classification based on improved MMAL [J]. Electronic Measurement Technology, 2024,47(17): 172-179.
[7] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [C]. International Conference on Learning Representations,2021.
[8] YU S, WU Y, LI W, et al. A model for fine-grained vehicle classification based on deep learning [J]. Neurocomputing, 2017, 257: 97-103.
[9] LIU M, YU C, LING H, et al. Hierarchical joint CNN-based models for fine-grained cars recognition [C]. Springer International Publishing, 2016: 337-347.
[10] FANG J, ZHOU Y, YU Y, et al. Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 18(7): 1782-1792.
[11] HU Q, WANG H, LI T, et al. Deep CNNs with spatially weighted pooling for fine-grained car recognition [J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(11): 3147-3156.
[12] XIANG Y, FU Y, HUANG H. Global topology constraint network for fine-grained vehicle recognition [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(7): 2918-2929.
[13] TAO X, HONG X, CHANG X, et al. Few-shot class-incremental learning[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12183-12192.
[14] REN M, TRIANTAFILLOU E, RAVI S, et al. Meta-learning for semi-supervised few-shot classification [J]. ArXiv preprint arXiv: 1803.00676, 2018.
[15] ACHITUVE I, NAVON A, YEMINI Y, et al. Gp-tree: A gaussian process classifier for few-shot incremental learning[C]. International Conference on Machine Learning. PMLR, 2021: 54-65.
[16] HAYES T L, KANAN C. Lifelong machine learning with deep streaming linear discriminant analysis[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 220-221.
[17] HOU S, PAN X, LOY C C, et al. Learning a unified classifier incrementally via rebalancing [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 831-839.
[18] REBUFFI S A, KOLESNIKOV A, SPERL G, et al. iCaRL: Incremental classifier and representation learning [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE,2017.
[19] LIU P, YUAN W, FU J, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing [J]. ACM Computing Surveys, 2023, 55(9):1-35.
[20] 赵勋,王家宝,李阳,等.细粒度图像分类的互补注意力方法[J].中国图象图形学报,2021,26(12):2860-2869.
ZHAO X, WANG J B, LI Y, et al. Complementary attention method for fine-grained image classification[J]. Chinese Journal of Image and Graphics, 2021, 26(12): 2860-2869.
[21] YANG L, LUO P, CHANGE LOY C, et al. A large-scale car dataset for fine-grained categorization and verification[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3973-3981.
[22] QI H, BROWN M, LOWE D G. Low-shot learning with imprinted weights [J]. IEEE, 2017, DOI: 10.1109/CVPR.2018.00610.
[23] ZHANG C, SONG N, LIN G, et al. Few-shot incremental learning with continually evolved classifiers[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 12455-12464.
[24] BELOUADAH E, POPESCU A. DeeSIL: Deep-shallow incremental learning[C]. European Conference on Computer Vision(ECCV) Workshops, 2018: 151-157.
[25] LI J, BAI Y, LOU Y, et al. Memory-based label-text tuning for few-shot class-incremental learning [J]. ArXiv preprint arXiv:2207.01036, 2022.
[26] HE K M, ZHANG X, REN SH Q, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
[27] JIA M, TANG L, CHEN B C, et al. Visual prompt tuning[C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 709-727.

作者简介

冉烨军,硕士研究生,主要研究方向为小样本图像分类。
E-mail:1569398787@qq.com

罗树霞,硕士研究生,主要研究方向为统计模型。
E-mail:2269122837@qq.com

李琼忆,硕士研究生,主要研究方向为统计模型。
E-mail:2738137306@qq.com

陶永,硕士研究生,主要研究方向为统计模型。
E-mail:1050989109@qq.com

金良琼(通信作者),副教授,硕士生导师,主要研究方向为统计模型与统计计算。
E-mail:1969549665@qq.com