



## 基于双曲乘积量化的无监督持续学习

朱广泽, 刘兴波, 聂秀山, 王少华

引用本文:

朱广泽, 刘兴波, 聂秀山, 等. 基于双曲乘积量化的无监督持续学习[J]. 智能系统学报, 2026, 21(5): 1237-1248.  
ZHU Guangze, LIU Xingbo, NIE Xiushan, et al. Unsupervised continual learning based hyperbolic product quantization[J]. CAAI transactions on intelligent systems, 2026, 21(5): 1237-1248.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202507031>

## 您可能感兴趣的其他文章

### 对抗样本三元组约束的度量学习算法

Metric learning algorithm with adversarial sample triples constraints  
智能系统学报. 2021, 16(1): 30-37 <https://dx.doi.org/10.11992/tis.202009050>

### 半监督类保持局部线性嵌入方法

Semi-supervised class preserving locally linear embedding  
智能系统学报. 2021, 16(1): 98-107 <https://dx.doi.org/10.11992/tis.202003007>

### 面对类别不平衡的增量在线序列极限学习机

Incremental online sequential extreme learning machine for imbalanced data  
智能系统学报. 2020, 15(3): 520-527 <https://dx.doi.org/10.11992/tis.201904040>

### 加权PageRank改进地标表示的自编码谱聚类算法

An autoencoder spectral clustering algorithm for improving landmark representation by weighted PageRank  
智能系统学报. 2020, 15(2): 302-309 <https://dx.doi.org/10.11992/tis.201904021>

### 结合谱聚类的标记分布学习

Label distribution learning based on spectral clustering  
智能系统学报. 2019, 14(5): 966-973 <https://dx.doi.org/10.11992/tis.201809019>

### 一种加入类间因素的曲线聚类算法

Curve clustering algorithms by adding the differences among clusters  
智能系统学报. 2019, 14(2): 362-368 <https://dx.doi.org/10.11992/tis.201709029>

DOI: 10.11992/tis.202507031

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20260706.1631.002>

# 基于双曲乘积量化的无监督持续学习

朱广泽, 刘兴波, 聂秀山, 王少华

(山东建筑大学 计算机与人工智能学院, 山东 济南 250101)

**摘要:** 无监督持续学习 (unsupervised continual learning, UCL) 是一种能够在无需人工标注的前提下, 从动态变化的任务流中持续获取知识的学习范式, 对于构建具备长期持续学习能力的智能系统具有重要意义。现有的无监督持续学习方法普遍存在早期任务性能不佳和灾难性遗忘的问题, 主要原因在于特征信息的多样性不足和结构对齐能力薄弱。为了有效捕获数据的层次结构语义信息并增强特征信息的多样性, 提出了一种新颖的无监督持续学习方法——基于双曲乘积量化的无监督持续学习方法 (unsupervised continual learning based hyperbolic product quantization, HPQUCL), 首次在 UCL 范式中引入双曲空间中的乘积量化机制。该方法利用双曲几何的层次结构特性增强了特征表示的多样性, 有效缓解了模型早期任务性能不佳的问题。此外, 设计了一种基于双曲距离的混合重放策略, 基于洛伦兹距离选择结构上具有代表性的样本进行重放, 有效减缓了灾难性遗忘问题。为了进一步提升模型的判别能力, 提出了一种在双曲空间中原始表示与量化表示之间的对比学习损失。在多个基准数据集上实验证明, HPQUCL 在知识保持和早期任务性能方面显著优于主流的无监督持续学习方法。结果表明, 引入双曲量化与结构感知重放机制为提升无监督持续学习的性能提供了有效途径。

**关键词:** 持续学习; 无监督学习; 乘积量化; 码本; 双曲空间; 洛伦兹距离; 对比学习; 灾难性遗忘

**中图分类号:** TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2026)05-1237-12

中文引用格式: 朱广泽, 刘兴波, 聂秀山, 等. 基于双曲乘积量化的无监督持续学习 [J]. 智能系统学报, 2026, 21(5): 1237-1248.

英文引用格式: ZHU Guangze, LIU Xingbo, NIE Xiushan, et al. Unsupervised continual learning based hyperbolic product quantization[J]. CAAI transactions on intelligent systems, 2026, 21(5): 1237-1248.

## Unsupervised continual learning based hyperbolic product quantization

ZHU Guangze, LIU Xingbo, NIE Xiushan, WANG Shaohua

(School of Computer and Artificial Intelligence, Shandong Jianzhu University, Jinan 250101, China)

**Abstract:** Unsupervised continual learning (UCL) is a paradigm for continuously acquiring knowledge from dynamically changing task streams without manual annotations, critical for building intelligent systems with long-term continuous learning capabilities. Existing UCL methods typically suffer from poor early task performance and catastrophic forgetting, stemming from insufficient feature diversity and weak structural alignment. To capture data's hierarchical semantics and enhance feature diversity, we propose a novel method—hyperbolic product quantization-based unsupervised continual learning (HPQUCL)—the first to introduce hyperbolic space product quantization into UCL. It leverages hyperbolic geometry's hierarchical properties to boost feature diversity, alleviating poor early task performance. Additionally, a hybrid replay strategy based on hyperbolic distance is designed, selecting structurally representative samples via Lorentz distance for replay to mitigate catastrophic forgetting. A contrastive loss between original and quantized representations in hyperbolic space is also proposed, which further enhances discriminative ability. Experiments on multiple benchmarks show HPQUCL significantly outperforms mainstream UCL methods in knowledge retention and early task performance, demonstrating hyperbolic quantization and structure-aware replay as effective for improving UCL.

**Keywords:** continual learning; unsupervised learning; product quantization; codebook; hyperbolic space; Lorentzian distances; contrastive learning; catastrophic forgetting

收稿日期: 2025-07-31. 网络出版日期: 2026-07-07.

基金项目: 国家自然科学基金青年基金项目 (62206160); 山东省自然科学基金青年基金项目 (ZR2022QF082).

通信作者: 刘兴波. E-mail: [sc1xb@mail.sdu.edu.cn](mailto:sc1xb@mail.sdu.edu.cn).

©《智能系统学报》编辑部版权所有

随着人工智能系统在实际应用中面临不断变化和动态扩展的数据环境, 实现模型的持续学习和适应能力成为推动智能系统长期发展的关键。特别是在开放环境中, 模型常常需要在不访问所

有历史数据的前提下持续适应新任务或新类别。为此,持续学习 (continual learning, CL)<sup>[1-2]</sup> 被提出以模拟人类大脑在长期记忆中不断积累知识的能力。现有的一些研究方法大多基于监督学习范式<sup>[3-11]</sup>。但是,现实世界中高质量、带标签的数据样本是稀缺的,使得依赖精确监督信号的方法在实际部署中受到严重限制。相比之下,无监督持续学习 (unsupervised continual learning, UCL) 作为一种在缺乏标签信息的情形下实现长期知识积累的学习范式,可以很好地解决这一问题。然而,在 UCL 范式中,模型在应对灾难性遗忘和提升早期任务表现方面仍存在明显不足,仍需持续探索更有效的解决方案。

由于没有标签,不能依赖监督信号,现有的无监督持续学习方法主要有以下几类: 1) 使用变分自编码器 (variational autoencoder, VAE) 和生成式重放机制 (generative replay) 试图重建和重放历史知识的方法<sup>[12-16]</sup>, 但这类方法在处理大规模数据集时常面临可扩展性不足和计算开销过大的挑战。2) 基于知识蒸馏的方法<sup>[17-19]</sup>, 通过提取历史模型中的关键信息以实现持续学习, 但这类方法通常需要依赖预先已知的任务切换点。3) 基于自监督学习 (如对比学习) 的方法<sup>[14,20-21]</sup>, 此类方法主要将有监督持续学习扩展到无监督表征学习框架下, 结合多个主流的无监督表征学习方法进行比较和评估。在现有的基于无监督表征学习的持续学习的工作中, CURL (continual unsupervised representation learning)<sup>[14]</sup> 是最早尝试在 UCL 范式下进行表征学习的方法之一, 但其在处理任务间语义差异较大的复杂情境时表现出明显的局限性。随后, 有研究<sup>[20]</sup> 将多种现有的有监督持续学习方法扩展到无监督设定下, 系统性评估了现有的无监督表征学习框架 (例如 Simsiam<sup>[22]</sup> 与 Barlow-Twins<sup>[23]</sup>) 在 UCL 范式下的表现, 但其仅进行了扩展和评估, 尚未提出专门面向 UCL 挑战的有效策略。鉴于 UCL 范式中亟需创新性方法的研究需求, Chen 等<sup>[21]</sup> 提出了 CUCL (Codebook for unsupervised continual learning) 方法。该方法作为一种即插即用的方法, 不仅将有监督持续学习方法扩展到无监督设定下, 还与最先进的无监督表征学习方法进行了比较, 包括 Simsiam<sup>[22]</sup>、Barlow-twins<sup>[23]</sup>、BYOL (exploring simple siamese representation learning)<sup>[17]</sup>、SimCLR (bootstrap your own latent a new approach to self-supervised learning)<sup>[18]</sup> 和 SwAV (unsupervised learning of visual features by contrasting cluster assignments)<sup>[19]</sup>。此外, CUCL 还首次提出

特征信息缺乏多样性是导致模型在前几个任务上的次优问题的主要原因, 并在 UCL 的范式下将基于对比学习的乘积量化<sup>[24]</sup> 应用到持续学习中, 很大程度上缓解了某些无监督学习方法中存在的次优问题。然而, CUCL 中基于欧氏空间的乘积量化方法主要侧重于挖掘同一图像在不同视图下的相似性特征, 因此量化过程中存在一定的局限性。例如, 1) 语义结构建模的缺失: 当数据本身呈现明显的层次结构时, 量化过程中的码字空间不仅需要具备结构上的紧凑性, 还应确保对类间语义距离拟合的合理性。而此类方法<sup>[25-27]</sup> 的码字学习机制往往独立于语义结构, 仅作为压缩或聚类工具, 忽视了图像数据中普遍存在的层次化语义结构。2) 欧氏空间的距离失真问题: 欧氏空间作为平直空间, 难以有效表示具有层次结构或语义嵌套关系的数据, 导致在层次语义建模中出现距离失真, 无法准确拉开或保持类间真实语义距离, 从而限制了码字对类别区分特征的表达能力。3) 特征表达能力受限影响下游性能: 由于语义结构与量化过程的不一致, 导致生成的量化表示缺乏多样性和判别性, 不仅难以捕捉关于其他类别的语义信息, 也会削弱模型在后续分类和任务迁移中的性能, 亟需引入结构感知的量化机制以提升表达质量。

为了解决以上问题, 本文以 CUCL<sup>[14]</sup> 为基础, 进一步探索了 UCL 范式下更具有创新性的方法。本文希望将传统的基于欧氏空间的乘积量化方法转换到非欧氏空间中, 即将量化学习过程整合到双曲几何空间<sup>[28]</sup> 中以增强图像的结构化语义的捕获<sup>[29-31]</sup>, 以此来为特征表示注入关于其他类别的信息以增强特征信息的多样性, 从而解决 UCL 范式下模型的次优问题。在双曲空间中, 将骨干网络提取的特征基于码本 (codebook, cb) 进行软量化, 从而得到具有语义区分性和多样性的量化表示。其中, 码本的构建是基于包含类别多样性信息的码字。此外, 为了有效缓解持续学习过程中的灾难性遗忘问题, 本文还设计了一种基于双曲空间的混合重放样本选择策略。该策略充分利用了双曲空间在建模语义层次结构方面的优势, 通过将历史任务样本嵌入至双曲空间中, 从而更准确地保持其原有的语义结构和类间关系。具体而言, 首先对每个历史样本的双曲特征表示进行量化, 并选取与其特征子向量双曲距离最近的码字作为其代理码字, 以此作为重放判断的几何基准。在此基础上, 本文依据候选样本到其代理码字之间的双曲距离, 将所有样本划分为困难

样本、中等样本和简单样本。最终, 从上述 3 类样本区域中按照设定比例分层采样, 以构建一个在几何分布和难度层次上均具有良好多样性的重放样本集合。

综上, 在 UCL 范式下, 本文提出了一个新的方法: 基于双曲乘积量化的无监督持续学习 (unsupervised continual learning based hyperbolic product quantization, HPQUCL), 本文的主要贡献有以下 3 个方面:

1) 本文首次将双曲空间中的乘积量化引入到无监督持续学习任务中。与传统欧氏空间下的量化方法不同, 在双曲空间进行量化不仅可以增加特征表示的多样性信息, 还可以增强无监督特征的结构表达能力, 以缓解模型在前几个任务上的次优问题。

2) 本文设计了一种基于双曲空间的混合重放样本选择策略。该策略利用了双曲空间在建模语义层次结构方面的优势, 根据双曲距离混合选取历史任务样本, 缓解了持续学习中的灾难性遗忘问题。

3) 本文在多个基准数据集上进行了广泛的实验评估。实验结果表明, 该方法在缓解灾难性遗忘和提升模型性能方面具有显著优势, 也有效改善了模型在前几个任务上次优的问题。

## 1 相关工作

在本节中, 将从持续学习和乘积量化 2 个方面, 简单回顾 2 个领域的相关工作。

### 1.1 持续学习

模型在面对持续学习任务时暴露出的灾难性遗忘问题引发了广泛关注。为缓解该问题, 持续学习作为应对长期学习中知识保持与更新之间矛盾的重要方法, 已发展出多种应对策略。现有方法大致可归为 3 类: 基于参数隔离的方法<sup>[4,10,32-34]</sup>、基于正则化的方法<sup>[3,11,35]</sup> 和基于记忆的方法<sup>[7,36-37]</sup>。

基于参数隔离的方法通常为不同的任务分配不同的模型参数, 以防止后续任务干扰先前学习的知识。例如, 通过迭代修剪将多个任务添加到单个网络方法 (adding multiple tasks to a single network by iterative pruning, PackNet)<sup>[32]</sup> 利用剪枝和零化技术, 动态地将不同任务的网络子空间分配给不同的任务。此外, 通过硬注意力机制 (overcoming catastrophic forgetting with hard attention to the task, HAT)<sup>[33]</sup>, 在神经网络单元级别屏蔽与旧任务相关的参数, 防止它们被当前任务更新, 从而实现任务间的隔离, 减少遗忘。通过平均梯度情节记忆进行高效的终身学习 A-GEM (efficient lifelong

learning with A-GEM)<sup>[4]</sup> 提出了一种基于梯度的约束方法, 利用小的经验回放缓冲区来隔离新旧任务的梯度。使用模块化网络与任务驱动先验的方法 (efficient continual learning with modular networks and task-driven priors, MNTDP)<sup>[10]</sup> 通过引入模块化网络结构和任务驱动的先验信息, 在任务增量学习过程中动态选择和重用特定模块, 实现高效的知識保留与迁移。

基于正则化的方法通过在损失函数中加入额外的正则项, 来限制模型对旧任务重要参数的修改以减轻灾难性遗忘。例如, 克服神经网络中的灾难性遗忘的方法 (rabinowitz overcoming catastrophic forgetting in neural networks, EWC)<sup>[35]</sup> 将模型参数的更新限制在多个任务的共同低损区域, 而不是只属于新任务的低损区域, 从而有助于缓解灾难性的遗忘问题。相对于 EWC 需要计算费时的费舍信息矩阵, 通过突触智能进行持续学习的方法 (continual learning through synaptic intelligence, SI)<sup>[11]</sup> 在训练过程中动态地估计这些参数重要性, 从而效率更高。记忆感知突触方法 (memory aware synapses: learning what (not) to forget, MAS)<sup>[3]</sup> 可以在模型训练的同时持续更新参数的重要性估计, 而不是等任务训练完再去计算。

基于记忆的方法通过存储部分旧任务的数据或特征信息, 在训练新任务时回顾或重放这些记忆。例如, 一种利用增量分类器和表示学习的方法 (incremental classifier and representation learning, iCARL)<sup>[7]</sup> 采用最近均值分类策略, 将输入图像的标签分配给与其在潜在特征空间中距离最近的类别原型。用于通用持续学习的暗经验方法 (dark experience for general continual learning: a strong, simple baseline, DER)<sup>[36]</sup> 是一种基于经验重放并结合软标签蒸馏的简单而强大的连续学习方法, 通过重放数据和知识共同缓解灾难性遗忘。面向持续学习的图像编码与再生学习 (learning to encode and regenerate images for continual learning, EEC)<sup>[37]</sup> 方法通过编码器将图像压缩存储, 再用解码器重建图像进行经验重放, 有效减少存储并保留旧任务知识。终身无监督混合方法 (representational continuity for unsupervised continual learning, LUMP)<sup>[20]</sup> 则通过数据增强和新样本与缓冲区中旧样本的插值来减缓遗忘。

### 1.2 乘积量化

乘积量化 (product quantization, PQ) 作为大规模检索系统中广泛采用的技术, 能够实现高效的向量压缩与近似最近邻搜索。其核心思想是将高

维向量空间分解为多个低维子空间,并在每个子空间中独立地进行矢量量化,从而显著降低存储成本并加速向量匹配过程。在 PQ 的基础上,后续研究提出了多种改进策略。例如,优化子空间划分方式的方法 (optimized product quantization, OPQ)<sup>[25]</sup> 引入旋转矩阵对原始向量进行线性变换,以最小化量化误差。传统 PQ 方法通常独立于特征学习过程,缺乏对数据分布与语义结构的建模能力。为克服这一局限,近年来的研究逐渐转向深层乘积量化范式<sup>[38]</sup>,即将 PQ 融入深度神经网络架构中<sup>[39-41]</sup>,以实现特征学习与量化的端到端联合优化。为了应对大规模数据中标注成本高昂、人工标签稀缺的问题,一些无监督量化方法被提出,其中,基于自动编码器的方法,例如,用于压缩域相似性搜索的无监督神经量化方法 (unsupervised neural quantization for compressed-domain similarity search, UNQCDSS)<sup>[42]</sup> 利用自编码器学习特征的低维表示,并通过解码器重构原始特征,确保压缩后的特征保留足够的信息。基于对比学习的方法,例如,基于码记忆的对比量化用于无监督图像检索 (contrastive quantization with code memory for unsupervised image retrieval, CQCM)<sup>[43]</sup> 引入去偏机制和码字记忆模块,有效缓解了传统对比学习在量化场景下的假负样本问题与训练退化问题。基于可微量化的方法,例如,用于端到端嵌入压缩的可微积量化 (beyond product quantization: deep progressive quantization for image retrieval, DPQ)<sup>[44]</sup> 通过设计可微的软量化过程,使

得离散编码可以通过反向传播进行优化,实现了嵌入层的端到端压缩。可微优化的乘积量化及其扩展 (differentiable optimized product quantization and beyond, DOPQ)<sup>[45]</sup> 通过引入可微性,使得模型可以在训练过程中同时优化量化参数和任务目标,在保持高压缩率的同时,显著提升了近似最近邻搜索的准确性。

## 2 基于双曲量化的无监督持续学习

在本章,首先介绍本文中涉及的一些符号的定义。随后,本章将详细阐述本文所提出的 HPQUCL 算法的量化过程、双曲量化对比学习以及双曲混合样本选择策略。

### 2.1 符号定义

持续学习的目标是在来自一系列任务的数据流上训练机器学习模型,在无监督持续学习的框架下,将任务序列表示为  $t \in \{1, 2, \dots, T\}$ , 其中每个任务对应的数据集表示为  $\mathcal{D}_t = \{\mathbf{X}_{t,i}\}_{i=1}^{N_t}$ , 每个任务有  $N_t$  个来自同一分布的样本。HPQUCL 的目标是在这种任务增量设置的前提下,通过训练一个持续学习的模型  $F = \{\theta_F\}$ , 使其能够不断积累知识,同时尽可能避免遗忘已经学到的内容,  $\theta_F$  表示模型参数。

### 2.2 双曲量化机制

本文设计了一种应用在无监督持续学习中的多子空间双曲嵌入可训练量化的框架。图 1 给出了所提出的 HPQUCL 流程的整体概览,其中包括量化过程和双曲混合样本选择策略。

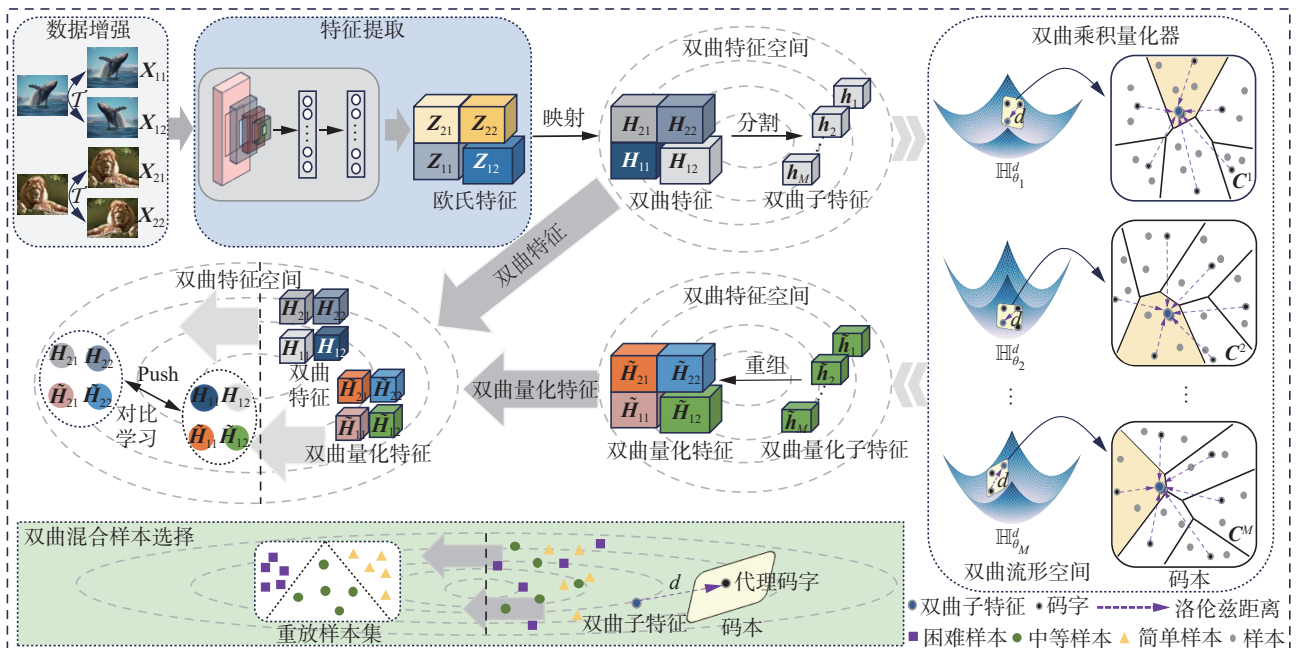


图 1 HPQUCL 框架

Fig. 1 Framework of HPQUCL

受近期在双曲空间量化方面的进展<sup>[46-48]</sup>的启发, 本文采用洛伦兹模型来描述双曲空间, 同时引入一种非欧氏几何构造策略。具体而言, 将向量空间 $\mathbb{V}$ 构建为 $M$ 个具有负常曲率的 $d$ 维的双曲空间的笛卡尔积:

$$\mathbb{V} = \mathbb{H}_{\theta_1}^d \times \mathbb{H}_{\theta_2}^d \times \cdots \times \mathbb{H}_{\theta_M}^d$$

式中 $\mathbb{H}_{\theta_m}^d$ 是曲率为 $\theta_m < 0$ 的洛伦兹双曲流形空间<sup>[49]</sup>, 在 $M$ 个双曲空间中选定参考点 $q = (q_1, q_2, \dots, q_M) \in \mathbb{V}$ , 其在流形上的切空间 $\Gamma_q \mathbb{V}$ 同样可视为各子流形切空间的笛卡尔积, 经过参考点的任意切向量 $\mathbf{v} \in \Gamma_q \mathbb{V}$ 都满足 $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M)$ , 其中 $\mathbf{v}_m \in \Gamma_{q_m} \mathbb{H}_{\theta_m}^d$ 。

该方法旨在对一对经过数据增强变换的图像视图进行特征提取和一致性建模。具体来说, 对于一个包含 $N_i$ 个样本的数据集 $\mathcal{D}_i = \{\mathbf{X}_{i,j}\}_{j=1}^{N_i}$ , 从 $\mathcal{D}_i$ 中选取的每个批次经过数据增强 $\mathcal{T}$ 独立生成两组不同的视图, 将这两组视图输入到特征提取网络得到两组特征表示 $\mathbf{Z}_a$ 和 $\mathbf{Z}_b$ , 其中 $\mathbf{Z}_a \in \mathbb{R}^D$ ,  $\mathbf{Z}_b \in \mathbb{R}^D$ , 每个视图通过特征提取网络获得欧氏空间下的特征之后, 将其分割成 $M$ 个相同维度的子特征 $\{\mathbf{z}_m\}_{m=1}^M$ 。接下来需要将欧氏空间中的特征 $\mathbf{Z}_a$ 和 $\mathbf{Z}_b$ 映射到双曲空间中, 得到双曲空间中的双曲特征 $\mathbf{H}_a$ 和 $\mathbf{H}_b$ , 以及其对应的双曲子特征 $\{\mathbf{h}_m\}_{m=1}^M$ 。具体做法如下: 首先, 通过正交投影将欧氏空间中的子特征 $\mathbf{z}_m$ 映射到参考点 $q$ 所在的切空间 $\Gamma_{q_m}$ , 得到切向量 $\mathbf{v}_m = \text{proj}_{q_m}(\mathbf{z}_m)$ 。然后将所有切向量 $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M)$ 通过指数映射转移到双曲空间乘积流形 $\mathbb{V}$ 上, 得到最终的双曲子特征 $\mathbf{h}$ :

$$\mathbf{h} = \exp_q(\mathbf{v}) = (\exp_{q_1}^{\theta_1}(\mathbf{v}_1), \exp_{q_2}^{\theta_2}(\mathbf{v}_2), \dots, \exp_{q_M}^{\theta_M}(\mathbf{v}_M))$$

因此,  $M$ 个子特征就映射成 $M$ 个双曲子特征 $\{\mathbf{h}_m\}_{m=1}^M$ , 在双曲量化模型中将量化码本 $\mathbf{C}$ 表示为 $\mathbf{C} = \mathbf{C}^1 \times \mathbf{C}^2 \times \cdots \times \mathbf{C}^M$ , 其中第 $m$ 个码本 $\mathbf{C}^m = \{\mathbf{c}_{m1}, \mathbf{c}_{m2}, \dots, \mathbf{c}_{mK}\} \in \mathbb{H}_{\theta_m}^d$ 包含了 $K$ 个码字。对于双曲子特征 $\mathbf{h}_m$ , 本文希望通过在码本 $\mathbf{C}^m$ 内计算加权聚合, 获得其软量化近似 $\tilde{\mathbf{h}}_m \in \mathbb{H}_{\theta_m}^d$ , 软量化后的向量 $\tilde{\mathbf{h}}_m$ 应当与码本 $\mathbf{C}^m$ 中的所有码字之间的加权期望平方距离最小化, 因此本文通过以下优化定义软量化向量:

$$\tilde{\mathbf{h}}_m \triangleq \underset{\mu \in \mathbb{H}_{\theta_m}^d}{\text{argmin}} w_{mk} d_{\mathcal{L}_m}^2(\mathbf{c}_{mk}, \mu)$$

式中权重项 $w_{mk}$ 表示输入子特征与第 $m$ 个码本中第 $k$ 个码字之间的相对相似度, 其值由洛伦兹距离度量。然后使用双曲乘积量化器量化这些双曲子特征, 在量化过程中, 计算双曲子特征与码字之间的平方洛伦兹距离<sup>[50]</sup>:

$$d_{\mathcal{L}_m}^2(\mathbf{h}_m, \mathbf{c}_{mk}) = \|\mathbf{h}_m - \mathbf{c}_{mk}\|_{\mathcal{L}}^2 = -2\theta_m - 2\langle \mathbf{h}_m, \mathbf{c}_{mk} \rangle_{\mathcal{L}}$$

式中 $\langle \mathbf{h}_m, \mathbf{c}_{mk} \rangle_{\mathcal{L}}$ 表示双曲子特征与码字的洛伦兹内积。对于双曲子特征 $\mathbf{h}_m \in \mathbb{H}_{\theta_m}^d$ , 该方法的目标是获

得其双曲量子特征 $\tilde{\mathbf{h}}_m$ , 量化过程为

$$\tilde{\mathbf{h}}_m = \frac{\sum_{k=1}^K w_{mk} \mathbf{c}_{mk}}{\sqrt{-1/\theta_m} \left\| \sum_{k=1}^K w_{mk} \mathbf{c}_{mk} \right\|_{\mathcal{L}}}$$

式中,  $\|\mathbf{a}\|_{\mathcal{L}} = \sqrt{\|\mathbf{a}\|_{\mathcal{L}}^2}$ , 表示向量 $\mathbf{a}$ 的洛伦兹范数模长。权重项 $w_{mk}$ 通过温度调控的 softmax 函数计算得到, 定义为

$$w_{mk} = \frac{\exp(-d_{\mathcal{L}_m}^2(\mathbf{h}_m, \mathbf{c}_{mk})/\tau)}{\sum_j \exp(-d_{\mathcal{L}_m}^2(\mathbf{h}_m, \mathbf{c}_{mj})/\tau)}$$

式中 $\tau$ 表示量化温度缩放参数。该权重在量化过程中, 一方面, 它实现了对各码字的加权融合, 利用码字与子特征向量之间的距离来度量每个码字的贡献, 从而生成连续可微的近似表示。另一方面, 它保留了子向量与多个码字之间的结构关系, 有助于增强表示的表达能力与判别性。对每个双曲嵌入子特征量化之后, 通过组合这些量子特征 $\tilde{\mathbf{h}}_m$ 获得两组视图 $\mathbf{Z}_a$ 和 $\mathbf{Z}_b$ 在双曲空间中的量化特征 $\tilde{\mathbf{H}}_a$ 和 $\tilde{\mathbf{H}}_b$ 。

### 2.3 双曲量化对比学习

借鉴传统对比学习的思想, 本文引入了一种基于洛伦兹相似度的交叉对比学习方法, 用于比较不同视图下的特征表示 $\mathbf{H}$ 和 $\tilde{\mathbf{H}}$ 。与传统对比学习相同, 如果 $\mathbf{H}$ 和 $\tilde{\mathbf{H}}$ 来源于同一张图像, 则将其视为正样本对。如果它们来自不同的图像, 则视为负样本对。因此, 可以根据正样本对 $(i, j)$ 在 $\mathbf{H}$ 和 $\tilde{\mathbf{H}}$ 之间应用对比损失函数:

$$\mathcal{L}_{\text{HPQUCL}} = -\log \frac{\exp(-d_{\mathcal{L}_m}^2(\mathbf{H}_i, \tilde{\mathbf{H}}_j)/\tau_l)}{\sum_{n=1}^{N_B} \mathbf{1}_{[n \neq j]} \exp(-d_{\mathcal{L}_m}^2(\mathbf{H}_i, \tilde{\mathbf{H}}_n)/\tau_l)}$$

式中:  $d_{\mathcal{L}_m}^2$ 代表洛伦兹距离, 这里用于表示正样本对之间的相似度;  $\tau_l$ 是一个非负温度参数;  $\mathbf{1}_{[n \neq j]}$ 是一个指示函数, 当 $n \neq j$ 时, 指示函数的值为 1。

HPQUCL 作为一种即插即用的方法, 在一些传统的无监督表征学习方法结构下进行评估, 包括 SimSiam<sup>[22]</sup>、Barlow-twins<sup>[23]</sup>、BYOL<sup>[17]</sup>、SimCLR<sup>[18]</sup>和 SwAV<sup>[19]</sup>。对于不同的 UCL 范式, 传统的无监督表征学习的损失不同, 为了方便评估, 这里将传统的无监督表征学习的损失统一记为 $\mathcal{L}_{\text{UNSUP}}$ 。因此, 整体损失结合了 HPQUCL 方法的损失和传统的无监督表征学习的损失, 具体形式为

$$\mathcal{L} = \mathcal{L}_{\text{HPQUCL}} + \mathcal{L}_{\text{UNSUP}}$$

### 2.4 双曲混合样本选择

为了缓解灾难性遗忘问题并保留先前任务的

知识, 本文提出了一种在双曲空间中的混合重放样本选择策略。具体而言, 首先选择码本中距离双曲嵌入子向量  $\mathbf{h}_m$  最近的那个码字  $\mathbf{c}_{mk}$  作为代理码字, 然后将候选缓冲区样本划分为 3 类区域: 1) 具有较大距离和高不确定性的困难样本, 2) 具有较小距离和高置信度的简单样本, 3) 距离居中、代表性较强的中等样本。从每个区域中按比例选取样本, 以构建一个多样且信息丰富的重放样本集合, 从而在保留难度的同时兼顾样本分布的多样性。为了更准确地度量每个子向量  $\mathbf{h}_m$  与其代理码字  $\mathbf{c}_{mk}$  之间的距离, 将距离度量从欧氏空间迁移至双曲空间, 并采用洛伦兹距离计算所有子向量与其对应的代理码字之间的距离之和:

$$\text{dis}(\mathbf{Z}) = \sum_m^M d_{\mathcal{L}_m}^2(\mathbf{h}_m, \mathbf{c}_{mk})$$

因此, 通过洛伦兹距离选择旧任务对应的样本特征进行存储, 不仅能够回顾之前任务所学的知识, 还能够通过再次训练这些样本来进一步学习当前任务的更多信息。

### 3 实验及分析

为了验证所提方法的有效性, 本文在 3 个广泛使用的数据集 Split CIFAR100、Split TinyImageNet 和 Split MiniImageNet 进行了实验。本节将从数据集、实验设置和实验结果及分析 3 个方面具体介绍实验内容。

#### 3.1 数据集

Split CIFAR100<sup>[51]</sup> 将 CIFAR-100 随机分为 10 个任务, 每个任务含 10 个不同类别, 每类有 500 个训练样本和 100 个测试样本。Split TinyImageNet 将 TinyImageNet<sup>[52]</sup> 的 100 个类别划分为 10 个任务, 每个任务 10 类, 图像尺寸为 64×64。Split MiniImageNet<sup>[53]</sup> 从 ImageNet 选取 100 类, 分为 10 个任务, 每类 500 训练样本和 100 测试样本, 图像尺寸为 84×84。

#### 3.2 实验设置

**对比方法** 首先, 本文将 HPQUCL 插入多种主流的持续学习方法中, 包括 SI<sup>[11]</sup>、A-GEM<sup>[4]</sup>、DER<sup>[36]</sup>、LUMP<sup>[20]</sup>, 得到其无监督形式并与 CUCL<sup>[21]</sup> 进行比较。其次, 在一些主流的无监督学习方法设定下, 包括 Simsiam<sup>[22]</sup>、Barlow-twins<sup>[23]</sup>、BYOL<sup>[17]</sup>、SimCLR<sup>[18]</sup> 和 SwAV<sup>[19]</sup>, 也将 HPQUCL 与 CUCL 进行了比较。最后, 将 HPQUCL 在 Simsiam 框架下与最新的无监督/弱监督基线进

行比较, 包括 EDSR(Effective data selection and replay for unsupervised continual learning)<sup>[54]</sup>、EVOLVE(Evolve: Enhancing unsupervised continual learning with multiple experts)<sup>[28]</sup>、iUL-LRCP(Prototype-Based continual learning with label-free replay buffer and cluster preservation loss)<sup>[55]</sup>、DSGD(Dynamic sub-graph distillation for robust semi-supervised continual learning)<sup>[56]</sup>。

**参数设置** 本文使用与 CUCL 相同的超参数。在无监督学习中通过 KNN<sup>[57]</sup> 分类器进行评估。所有实验均使用 ResNet18 作为主干网络在无监督设定下, 每个任务训练 200 个 epoch, batch size 为 256。在有监督设定下, 量化部分的温度参数设置为 5, 损失函数温度为 0.5。HPQUCL 使用 24 位量化器: 包含 8 个码本, 每个码本有 8 个码字 (codeword), 维度为 16。在所有实验中, 保持 0.03 的统一学习率。对于 HPQUCL 中的混合重放样本选择方法, 将 buffer 设置为 20 用于保存旧任务的重放样本。

**评价指标** 本文采用以下 3 种指标评估持续学习模型的性能: 平均准确度 (average accuracy, ACC)、反向迁移 (backward transfer, BWT) 和平均平均准确度 (mean average accuracy, MAA)。其中, ACC 表示模型在训练完所有任务后在旧任务上的表现。MAA 表示模型在训练每个新任务后, 对所有已学任务的平均准确率。BWT 表示模型在学习新任务后, 在旧任务上的表现。

#### 3.3 实验结果及分析

本节将从所提方法与对比方法的性能比较、消融实验和参数敏感性实验 3 个方面展示实验结果并进行分析。

##### 3.3.1 实验结果对比与分析

**与有监督方法及其无监督变体的比较** 为了验证 HPQUCL 方法在缓解灾难性遗忘和改进次优解方面的有效性, 本文将 HPQUCL 融入现有的多种变体方法中并与 CUCL 进行了综合对比评估, 如表 1 所示。从表中可以看出, HPQUCL 几乎在所有的变体方法上都优于 CUCL。一个可能的原因是, HPQUCL 有效地保留了样本之间的结构信息, 增强了类别间的判别性。其中, 在 MAA 方面, 基于 Simsiam 架构下算法改进的方法, 3 个数据集上可以观察到 SI、DER 和 LUMP 的 w/HPQUCL 方法比 w/CUCL 方法分别平均改进了 0.59、1.29 和 1.91 百分点。在基于 Barlow-twins 架构下算法改进的方法, 将 HPQUCL 与 SI、DER 和 LUMP 融合时, 可以观察到 w/HPQUCL 方法比

w/CUCL 方法分别平均改进了 0.89、1.08 和 1.12 百分点。这表明本文提出的方法在改进模型的次优问题上有着良好的效果。在 BWT 方面,

HPQUCL 方法也呈现出显著的优势 (如表 1 加粗数值), 证明该方法在一定程度上也克服了模型的灾难性遗忘现象。

表 1 不同数据集下, HPQUCL 和 CUCL 与有监督基准方法及其无监督变体的实验结果对比  
Table 1 Comparison of experimental results between HPQUCL, CUCL, and supervised baseline methods as well as their unsupervised variants on different datasets %

| 设置           | 方法                   | Split CIFAR100         |       |              | Split TinyImageNet     |       |              | Split MiniImageNet     |       |              |
|--------------|----------------------|------------------------|-------|--------------|------------------------|-------|--------------|------------------------|-------|--------------|
|              |                      | MAA                    | ACC   | BWT          | MAA                    | ACC   | BWT          | MAA                    | ACC   | BWT          |
| Supervised   | SI <sup>[11]</sup>   | 59.17                  | 5.77  | -32.00       | 52.12                  | 44.42 | -35.27       | 50.12                  | 44.97 | -34.52       |
|              | A-GEM <sup>[4]</sup> | 58.10                  | 54.77 | -31.24       | 52.00                  | 48.74 | -29.67       | 50.84                  | 48.09 | -30.32       |
|              | DER <sup>[36]</sup>  | 72.48                  | 67.97 | -20.09       | 60.83                  | 56.74 | -21.40       | 58.52                  | 54.49 | -23.40       |
| Simsiam      | SI                   | 69.75                  | 74.32 | -2.66        | 51.37                  | 58.80 | -0.98        | 54.79                  | 63.24 | -0.30        |
|              | w/CUCL               | 75.79                  | 76.68 | -5.12        | 62.45                  | 64.48 | -4.49        | 64.91                  | 67.05 | -4.44        |
|              | w/HPQUCL             | 75.93 <sub>+0.14</sub> | 75.66 | <b>-2.38</b> | 63.24 <sub>+0.79</sub> | 65.88 | <b>-0.29</b> | 65.74 <sub>+0.83</sub> | 68.29 | <b>-3.68</b> |
|              | DER                  | 73.05                  | 75.31 | -4.04        | 51.01                  | 58.66 | -0.82        | 56.78                  | 63.72 | -0.61        |
|              | w/CUCL               | 76.32                  | 77.24 | -4.39        | 62.41                  | 65.10 | -4.04        | 65.03                  | 67.15 | -4.42        |
|              | w/HPQUCL             | 77.13 <sub>+0.81</sub> | 75.83 | -4.61        | 63.57 <sub>+1.16</sub> | 64.97 | <b>-0.02</b> | 66.92 <sub>+1.89</sub> | 68.24 | <b>-3.21</b> |
|              | LUMP <sup>[20]</sup> | 63.95                  | 69.83 | -3.60        | 53.60                  | 61.12 | -5.93        | 52.36                  | 59.54 | -4.52        |
|              | w/CUCL               | 72.19                  | 74.78 | -0.82        | 58.51                  | 59.42 | -2.80        | 58.67                  | 60.25 | -2.36        |
|              | w/HPQUCL             | 74.12 <sub>+1.93</sub> | 75.46 | -2.13        | 60.22 <sub>+1.17</sub> | 58.33 | <b>-0.25</b> | 60.77 <sub>+2.10</sub> | 61.42 | <b>-0.23</b> |
| Barlow-twins | SI                   | 73.13                  | 73.52 | -8.16        | 60.39                  | 59.72 | -8.73        | 61.53                  | 61.97 | -7.91        |
|              | w/CUCL               | 74.40                  | 74.93 | -7.68        | 61.59                  | 62.44 | -6.11        | 62.67                  | 64.50 | -6.61        |
|              | w/HPQUCL             | 74.77 <sub>+0.37</sub> | 75.22 | -7.70        | 62.33 <sub>+0.74</sub> | 65.12 | <b>-5.37</b> | 64.23 <sub>+1.56</sub> | 65.78 | <b>-5.54</b> |
|              | DER                  | 73.15                  | 73.46 | -7.63        | 60.40                  | 60.04 | -8.82        | 61.79                  | 62.47 | -7.93        |
|              | w/CUCL               | 74.62                  | 75.47 | -7.16        | 61.57                  | 60.88 | -8.82        | 63.10                  | 63.82 | -7.64        |
|              | w/HPQUCL             | 75.60 <sub>+0.98</sub> | 74.49 | <b>-5.11</b> | 62.74 <sub>+1.17</sub> | 61.06 | <b>-7.62</b> | 64.18 <sub>+1.08</sub> | 64.03 | <b>-7.11</b> |
|              | LUMP                 | 67.18                  | 66.73 | -6.33        | 55.26                  | 57.78 | -2.62        | 54.91                  | 57.66 | -2.63        |
|              | w/CUCL               | 72.16                  | 73.66 | -3.88        | 59.43                  | 61.58 | -3.04        | 58.60                  | 61.80 | -1.97        |
|              | w/HPQUCL             | 73.07 <sub>+0.91</sub> | 69.64 | <b>-2.71</b> | 60.34 <sub>+0.91</sub> | 60.16 | <b>-2.80</b> | 60.11 <sub>+1.51</sub> | 63.25 | -3.22        |

**与无监督方法的比较** 为了更好地与 CUCL 方法进行比较, 本文将 HPQUCL 方法融入到不同的无监督学习方法中, 如表 2 所示。表 2 给出了一些主流的无监督学习方法设定下 HPQUCL 方法和 CUCL 方法的对比结果。HPQUCL 方法整体上带来了显著的提升, MAA 方面, w/HPQUCL 在 Simsiam、Barlow-twins、BYOL、SimCL 和 SwAV 上分别平均提升了 2.16、1.15、1.12、1.23 和 1.09 百分点, 这表明该方法在解决模型的次优问题上具有良好的通用性和有效性。主要的原因在于, 在双曲空间中的码字与子向量的距离拟合更加真实有效, 从而提高了码字的学习能力, 生成更具有多样性信息的量化特征表示。在 BWT 方面, HPQUCL 方法也呈现出良好的表现 (如表 2 加粗

数值)。比起 CUCL 中仅挑选困难边界样本的方法, HPQUCL 的双曲混合选择样本的策略在几何分布和难度层次上均具有良好多样性。因此, HPQUCL 在缓解模型的灾难性遗忘问题上更具有优势。

**与最新方法的比较** 为了更好地说明 HPQUCL 方法的先进性, 本文将 HPQUCL 方法与最新的 4 个无监督/弱监督基线作对比, 如表 3 所示。从实验结果来看, 在同样的实验设定下, HPQUCL 在 Split MiniImageNet 数据集上所有度量表现最佳。在 Split CIFAR100 和 Split TinyImageNet 数据集上, HPQUCL 始终能够将 BWT 的衰减幅度控制在较低水平。HPQUCL 有效减弱了新任务数据分布对旧知识边界的挤压, 说明 HPQUCL 在缓解模型的灾难性遗忘问题上更具有先进性。

表 2 不同数据集下, HPQUCL 和 CUCL 与无监督基准方法的实验结果对比

Table 2 Comparison of experimental results between HPQUCL, CUCL, and unsupervised baseline methods across different datasets %

| 方法         | Split CIFAR100         |       |              | Split TinyImageNet     |       |              | Split MiniImageNet     |       |              |
|------------|------------------------|-------|--------------|------------------------|-------|--------------|------------------------|-------|--------------|
|            | MAA                    | ACC   | BWT          | MAA                    | ACC   | BWT          | MAA                    | ACC   | BWT          |
| Simsiam    | 67.48                  | 73.87 | -1.20        | 50.47                  | 57.18 | -2.62        | 57.33                  | 64.36 | -0.92        |
| w/CUCL     | 76.07                  | 75.81 | -6.26        | 63.23                  | 64.42 | -5.18        | 64.61                  | 66.17 | -5.30        |
| w/HPQUCL   | 77.73 <sub>+1.66</sub> | 75.66 | <b>-3.34</b> | 65.23 <sub>+2.00</sub> | 66.74 | <b>-3.45</b> | 67.42 <sub>+2.81</sub> | 67.15 | <b>-3.42</b> |
| BYOL       | 75.45                  | 77.80 | -1.40        | 56.90                  | 63.98 | 1.16         | 57.75                  | 66.56 | 3.96         |
| w/CUCL     | 76.78                  | 77.67 | -4.19        | 63.90                  | 65.66 | -4.58        | 64.50                  | 66.99 | -3.74        |
| w/HPQUCL   | 77.25 <sub>+0.47</sub> | 76.48 | <b>-1.14</b> | 65.97 <sub>+2.07</sub> | 66.45 | <b>-3.21</b> | 65.42 <sub>+0.92</sub> | 67.23 | <b>-0.42</b> |
| Barlowtins | 72.91                  | 72.96 | -8.47        | 59.90                  | 59.76 | -8.22        | 61.26                  | 63.32 | -6.92        |
| w/CUCL     | 75.29                  | 75.18 | -8.82        | 63.12                  | 63.58 | -7.27        | 64.31                  | 66.16 | -6.56        |
| w/HPQUCL   | 76.78 <sub>+1.49</sub> | 77.56 | -9.34        | 63.33 <sub>+0.21</sub> | 63.97 | <b>-6.29</b> | 66.23 <sub>+1.92</sub> | 63.24 | <b>-5.12</b> |
| SimCRL     | 74.39                  | 74.14 | -6.58        | 63.49                  | 63.24 | -6.56        | 63.15                  | 65.30 | -4.71        |
| w/CUCL     | 75.38                  | 75.88 | -5.23        | 63.81                  | 64.14 | -5.36        | 63.72                  | 64.77 | -5.09        |
| w/HPQUCL   | 76.23 <sub>+0.85</sub> | 77.48 | <b>-4.22</b> | 66.23 <sub>+2.42</sub> | 65.96 | <b>-3.34</b> | 64.13 <sub>+0.41</sub> | 67.56 | -6.22        |
| SwAV       | 73.18                  | 74.44 | -4.20        | 58.52                  | 62.94 | -0.76        | 62.03                  | 62.66 | -4.67        |
| w/CUCL     | 74.77                  | 74.86 | -6.39        | 61.95                  | 62.84 | -5.40        | 63.22                  | 64.29 | -5.63        |
| w/HPQUCL   | 75.66 <sub>+0.89</sub> | 75.23 | <b>-5.36</b> | 62.23 <sub>+0.28</sub> | 61.66 | <b>-4.34</b> | 65.33 <sub>+2.11</sub> | 65.89 | <b>-3.36</b> |

表 3 不同数据集下, Simsiam w/HPQUCL 和最新方法实验结果对比

Table 3 Performance comparison of SimSiam w/ HPQUCL against recent state-of-the-art methods across different datasets %

| 方法                       | Split CIFAR100 |              |              | Split TinyImageNet |              |              | Split MiniImageNet |              |              |
|--------------------------|----------------|--------------|--------------|--------------------|--------------|--------------|--------------------|--------------|--------------|
|                          | MAA            | ACC          | BWT          | MAA                | ACC          | BWT          | MAA                | ACC          | BWT          |
| Simsiam                  | 67.48          | 73.87        | -1.20        | 50.47              | 57.18        | -2.62        | 57.33              | 64.36        | -0.92        |
| w/HPQUCL                 | 77.73          | 75.66        | <b>-3.34</b> | 65.23              | 66.74        | <b>-3.45</b> | <b>67.42</b>       | <b>67.15</b> | <b>-3.42</b> |
| EDSR <sup>[54]</sup>     | 76.45          | 79.83        | -5.43        | 67.95              | 75.98        | -5.16        | 66.75              | 66.56        | 5.96         |
| EVOLVE <sup>[28]</sup>   | 77.08          | 79.48        | -4.14        | 69.97              | 74.45        | -4.21        | 67.13              | 66.02        | -5.42        |
| iUL-LRCP <sup>[55]</sup> | <b>78.91</b>   | <b>79.96</b> | -3.47        | <b>72.90</b>       | <b>77.76</b> | -3.89        | 67.26              | 66.32        | -6.92        |
| DSGD <sup>[56]</sup>     | 76.22          | 75.43        | -3.57        | 67.44              | 70.78        | -4.44        | 66.72              | 65.43        | -4.43        |

### 3.3.2 消融实验

为了深入探究 HPQUCL 框架中关键组件的实际贡献, 本文开展了详尽的消融实验。根据

HPQUCL 的设计方案, 本文从不同码本数量和不同重放样本选择方法 2 个角度验证所提方法的有效性。表 4 和表 5 展示了消融实验的结果。

表 4 SimSiam w/HPQUCL 对应不同码本数量的实验结果

Table 4 Experimental results of SimSiam w/HPQUCL with different numbers of codebooks %

| 码本数量 | Split CIFAR100 |       |       | Split MiniImageNet |       |       |
|------|----------------|-------|-------|--------------------|-------|-------|
|      | MAA            | ACC   | BWT   | MAA                | ACC   | BWT   |
| 0    | 68.54          | 72.67 | -1.23 | 57.33              | 67.36 | -1.92 |
| 2    | 73.23          | 74.44 | -7.56 | 62.97              | 64.54 | -5.77 |
| 4    | 75.99          | 74.68 | -5.24 | 65.37              | 65.33 | -4.56 |
| 8    | <b>77.73</b>   | 75.66 | -5.34 | <b>67.42</b>       | 67.15 | -3.42 |

表 5 BYOL w/HPQUC 对应不同重放样本选择方法的实验结果  
Table 5 Experimental results of BYOL w/HPQUCL with different replay sample selection methods %

| 方法          | Split CIFAR100 |       |       | Split MiniImageNet |       |       |
|-------------|----------------|-------|-------|--------------------|-------|-------|
|             | MAA            | ACC   | BWT   | MAA                | ACC   | BWT   |
| easy-sample | 59.48          | 66.24 | -8.55 | 55.54              | 67.42 | -7.89 |
| hard-sample | 77.11          | 76.93 | -2.76 | 65.23              | 66.78 | -2.64 |
| mix-sample  | 77.25          | 76.48 | -1.14 | 65.42              | 67.23 | -0.42 |

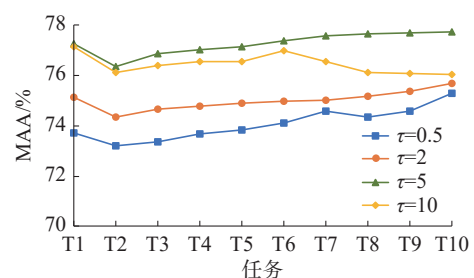
1) 为了验证特征表示的信息多样性是导致模型在前几个任务上的次优问题的主要原因, 本文选择 SimSiam 作为无监督基础架构, 在双曲量化器中分别使用不同的码本数量进行量化。具体而言, 保持其他设置不变, 码本数量分别设置为 0、2、4、8 个, 在 Split CIFAR100 和 Split MiniImageNet 数据集上的实验结果如表 4 所示。结果表明, 不使用共享码本时, 特征无法通过共享码本注入其他类别特征的信息, 模型性能较低。随着码本数量的增加, 特征被分割成多个子向量, 不同的码字可以为不同子向量的量化表示提供互补的多样性信息, 生成更具有多样性信息的量化特征表示, 从而带来更优的模型性能。

2) 为了验证双曲混合样本选择策略优势, 本文选择 BYOL 作为无监督基础架构。样本重放数量固定为 20, 其他参数保持不变, 将 20 个样本按照以下方式选择: 全部选择为简单样本 (easy-sample)、全部选择为困难样本 (hard-sample) 和按照双曲混合样本选择方式进行选择 (mix-sample)。在 Split CIFAR100 和 Split MiniImageNet 数据集上的实验结果如表 5 所示。结果表明, 全部使用简单样本作为重放样本在解决灾难性遗忘问题上的效果最差, 可能的原因是模型丢失了对边界样本或复杂样本的记忆。全部使用困难样本作为重放样本, 一定程度上增强了模型的记忆能力。使用双曲混合样本选择 (mix-sample) 的方案在 BWT 上表现最好, 此方案兼顾了简单和困难样本, 同时也重放了中等样本, 在几何分布和难度层次上均具有良好多样性。由此验证了本文所提的双曲混合样本选择方案的有效性。

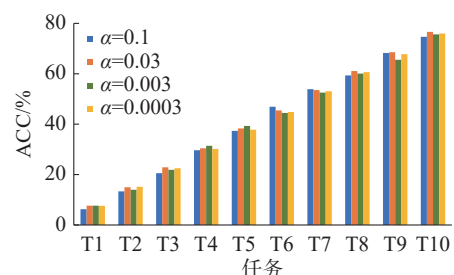
### 3.3.3 参数实验

为了进一步验证模型的稳定性, 本文以 SimSiam w/HPQUCL 为例, 对模型中的量化温度  $\tau$ 、样本重放内存大小 buffer 和整体学习率  $\alpha$  分别在 Split CIFAR100 数据集上做敏感性测试。图 2(a) 给出了量化温度  $\tau$  对模型 MAA 的影响, 图 2(b) 给

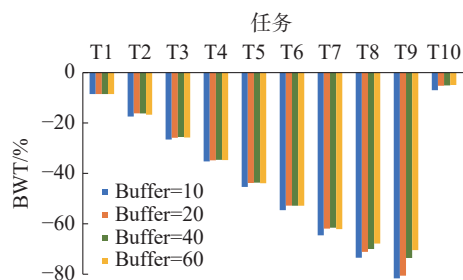
出了整体学习率  $\alpha$  对模型 ACC 的影响, 图 2(c) 给出了样本重放内存大小 buffer 对模型 BWT 的影响。从图 2(a) 可以看出,  $\tau = 5$  时, 在大部分任务上的 MAA 均优于其他对比组。在任务前期 (T1~T5), 模型在前几个任务上的次优问题也得到了有效的缓解, 在中后期任务 (T5~T10) 中, 模型的表现尤为稳定和出色。从图 2(b) 可以看出, 整体学习率  $\alpha$  的选择对 ACC 的影响非常微弱。从图 2(c) 可以看出, 在训练至第 10 个任务 (T10) 后, 所有设置的缓冲区容量 (buffer) 均表现出不同程度的 BWT 回升趋势, 表明模型在学习新任务的同时对旧任务的遗忘有所缓解。尤其值得注意的是, 在 buffer 为 60 的设定下, 最终的 BWT 值最接近 0, 优于较小缓冲区容量的对比。这一结果说明, 更大的样本重放缓冲区能够更充分地保留各历史信息, 从而有效抑制灾难性遗忘现象。但是, 随着缓冲区大小的增加, 模型需要访问和处理更多的历史样本, 从而导致更高的计算开销与时间成本。因此, 在实际应用中, 应根据可用资源在性能与成本之间进行合理权衡。



(a) 量化温度  $\tau$  对模型 MAA 的影响



(b) 学习率  $\alpha$  对模型 ACC 的影响



(c) 样本重放内存大小buffer对模型BWT的影响

图 2 参数  $\tau$ 、buffer 和  $\alpha$  分别对模型 MAA、ACC 和 BWT 的影响

Fig. 2 Effects of parameters,  $\tau$ , buffer, and  $\alpha$  on the model's MAA, ACC, and BWT, respectively

## 4 结束语

本文提出了一种基于双曲乘积量化的无监督持续学习方法。该方法为解决模型次优问题和灾难性遗忘问题提供了新的思路,即将传统的欧氏特征空间转换为更具有表现力的双曲特征空间,不仅增强了特征的表达能力与多样性,也通过双曲空间的几何性质提升了旧任务知识的保留效果。在此基础上,本文也引入了双曲量化对比学习机制,以增强量化特征的判别性与一致性,引导模型捕捉具有判别性的特征。大量的实验证明了所提方法的有效性。然而,本文所提方法对具有层次结构的数据较为敏感,适用于特定场景,具有一定局限性。未来可在语义层次更丰富的数据集上进一步验证其有效性,并探索更具普适性与灵活性的持续学习模型。

## 参考文献:

- [1] MCCLOSKEY M, COHEN N J. Catastrophic interference in connectionist networks: the sequential learning problem[M]//Psychology of Learning and Motivation. Amsterdam: Elsevier, 1989: 109–165.
- [2] RING M B. CHILD: a first step towards continual learning[J]. *Machine learning*, 1997, 28(1): 77–104.
- [3] ALJUNDI R, BABILONI F, ELHOSEINY M, et al. Memory aware synapses: learning what (not) to forget[C]//Computer Vision – ECCV 2018. Cham: Springer, 2018: 144–161.
- [4] CHAUDHRY A, RANZATO M, ROHRBACH M, et al. Efficient lifelong learning with A-GEM[C]//International Conference on Learning Representations. New Orleans: OpenReview.net, 2019: 1–20.
- [5] CHEN Cheng, ZHANG Ji, SONG Jingkuan, et al. Class gradient projection for continual learning[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa: ACM, 2022: 5575–5583.
- [6] LI Zhizhong, HOIEM D. Learning without forgetting[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2018, 40(12): 2935–2947.
- [7] 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. Honolulu: IEEE, 2017: 5533–5542.
- [8] RUSU A A, RABINOWITZ N C, DESJARDINS G, et al. Progressive neural networks[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016: 1–5.
- [9] SPRECHMANN P, JAYAKUMAR S M, RAE J W, et al. Memory-based parameter adaptation[C]//International Conference on Learning Representations. Vancouver: OpenReview.net, 2018: 1–16.
- [10] VENIAT T, DENOYER L, RANZATO M. Efficient continual learning with modular networks and task-driven priors[C]//International Conference on Machine Learning. Vienna: PMLR, 2021: 1–30.
- [11] ZENKE F, POOLE B, GANGULI S. Continual learning through synaptic intelligence[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney: PMLR, 2017: 3987–3995.
- [12] ACHILLE A, ECCLES T, MATTHEY L, et al. Life-long disentangled representation learning with cross-domain latent homologies[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: ACM, 2018: 9895–9905.
- [13] RAMAPURAM J, GREGOROVA M, KALOUSIS A. Lifelong generative modeling[J]. *Neurocomputing*, 2020, 404: 381–400.
- [14] RAO D, VISIN F, RUSU A A, et al. Continual unsupervised representation learning[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019: 7645–7655.
- [15] WU Chenshen, HERRANZ L, LIU Xialei, et al. Memory replay GANs: learning to generate new categories without forgetting[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal: Curran Associates Inc., 2018: 1–10.
- [16] YE Fei, BORS A G. Learning latent representations across multiple data domains using lifelong VAEGAN[C]//Computer Vision–ECCV 2020. Cham: Springer, 2020: 777–795.
- [17] GRILL J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent: a new approach to self-supervised learning[C]//Proceedings of the 34th Conference on Neural Information Processing Systems. Vancouver: Curran Associates

- Inc., 2020: 21271–21284.
- [18] CHEN Ting, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the 37th International Conference on Machine Learning. Virtual: PMLR, 2020: 1597–1607.
- [19] CARON M, MISRA I, MAIRAL J, et al. Unsupervised learning of visual features by contrasting cluster assignments[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020: 9912–9924.
- [20] MADAAN D, YOON J, LI Yuanchun, et al. Representational continuity for unsupervised continual learning[C]//10th International Conference on Learning Representations. Baltimore: OpenReview.net, 2022: 1–18.
- [21] CHEN Cheng, SONG Jingkuan, ZHU Xiaosu, et al. CUCL: codebook for unsupervised continual learning[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM, 2023: 1729–1737.
- [22] CHEN Xinlei, HE Kaiming. Exploring simple Siamese representation learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 15745–15753.
- [23] ZBONTAR J, JING Li, MISRA I, et al. Barlow twins: self-supervised learning via redundancy reduction[C]//International Conference on Machine Learning. Virtual: PMLR, 2021: 12310–12320.
- [24] JANG Y K, CHO N I. Self-supervised product quantization for deep unsupervised image retrieval[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 12065–12074.
- [25] GE Tiezheng, HE Kaiming, KE Qifa, et al. Optimized product quantization[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2014, 36(4): 744–755.
- [26] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90.
- [27] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [28] YU Xiaofan, ROSING T, GUO Yunhui. Evolve: enhancing unsupervised continual learning with multiple experts[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2024: 2355–2366.
- [29] CHEN Jingzhou, XIONG Fengchao, QIAN Yuntao, et al. Hierarchical contrastive learning for multigranularity ship classification with learnable class queries[J]. *IEEE transactions on geoscience and remote sensing*, 2025, 63: 5622916.
- [30] LIN Qinghong, CHEN Xiaojun, ZHANG Qin, et al. Deep unsupervised hashing with latent semantic components[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual: AAAI, 2022: 7488–7496.
- [31] MA Yueen, SONG Zixing, HU Xuming, et al. Graph component contrastive learning for concept relatedness estimation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Washington D. C. : AAAI, 2023: 13362–13370.
- [32] MALLYA A, LAZEBNIK S. PackNet: adding multiple tasks to a single network by iterative pruning[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7765–7773.
- [33] SERRÀ J, SURIS D, MIRON M, et al. Overcoming catastrophic forgetting with hard attention to the task[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018: 4548–4557.
- [34] WANG Mingyang, ADEL H, LANGE L, et al. Rehearsal-free modular and compositional continual learning for language models[C]//Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Mexico City: Association for Computational Linguistics, 2024: 469–480.
- [35] KIRKPATRICK J, PASCANU R, RABINOWITZ N, et al. Overcoming catastrophic forgetting in neural networks[J]. *Proceedings of the national academy of sciences of the United States of America*, 2017, 114(13): 3521–3526.
- [36] BUZZEGA P, BOSCHINI M, PORRELLO A, et al. Dark experience for general continual learning: a strong, simple baseline[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc. , 2020: 15920–15930.
- [37] AYUB A, WAGNER A R. EEC: learning to encode and regenerate images for continual learning[C]//International Conference on Learning Representations. Virtual: OpenReview.net, 2021: 1–12.
- [38] ZHAN Jingtao, MAO Jiaxin, LIU Yiqun, et al. Jointly optimizing query encoder and product quantization to improve retrieval performance[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Virtual: ACM, 2021: 2487–2496.
- [39] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2015: 1–14.
- [40] XIE Saining, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern

- Recognition. Honolulu: IEEE, 2017: 5987–5995.
- [41] CUI Hui, ZHU Lei, LI Jingjing, et al. Scalable deep hashing for large-scale social image retrieval[J]. *IEEE transactions on image processing*, 2020, 29: 1271–1284.
- [42] MOROZOV S, BABENKO A. Unsupervised neural quantization for compressed-domain similarity search[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 3036–3045.
- [43] WANG Jinpeng, ZENG Ziyun, CHEN Bin, et al. Contrastive quantization with code memory for unsupervised image retrieval[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual: AAAI, 2022: 2468–2476.
- [44] GAO Lianli, ZHU Xiaosu, SONG Jingkuan, et al. Beyond product quantization: deep progressive quantization for image retrieval[C]//Proceedings of the International Joint Conference on Artificial Intelligence. Macau: IJCAI, 2019: 723–729.
- [45] LU Zepu, LIAN Defu, ZHANG Jin, et al. Differentiable optimized product quantization and beyond[C]//Proceedings of the ACM Web Conference. Austin: ACM, 2023: 3353–3363.
- [46] GAO Zhi, WU Yuwei, HARANDI M, et al. Curvature-adaptive meta-learning for fast adaptation to manifold data[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2023, 45(2): 1545–1562.
- [47] GU A, SALA F, GUNEL B, et al. Learning mixed-curvature representations in products of model spaces[C]//International Conference on Learning Representations. Vancouver: OpenReview.net, 2018: 1–13.
- [48] QIU Zexuan, LIU Jiahong, CHEN Yankai, et al. HiHPQ: hierarchical hyperbolic product quantization for unsupervised image retrieval[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Washington D. C. : AAAI, 2024: 4614–4622.
- [49] 徐蓉, 姜峰, 姚鸿勋. 流形学习概述[J]. *智能系统学报*, 2006, 1(1): 44–51.
- XU Rong, JIANG Feng, YAO Hongxun. Overview of manifold learning[J]. *CAAI transactions on intelligent systems*, 2006, 1(1): 44–51.
- [50] LAW M T, LIAO Renjie, SNELL J, et al. Lorentzian distance learning for hyperbolic representations[C]//Proceedings of the International Conference on Machine Learning. Long Beach: PMLR, 2019: 3672–3681.
- [51] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[EB/OL]. (2009–04–08) [2025–07–31]. <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>.
- [52] DENG Jia, DONG Wei, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 248–255.
- [53] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016: 3630–3638.
- [54] LIU Hanmo, DI Shimin, LI Haoyang, et al. Effective data selection and replay for unsupervised continual learning [C]//2024 IEEE 40th International Conference on Data Engineering. Utrecht: IEEE, 2024: 1449–1463.
- [55] AGHASANLI A, LI Yi, ANGELOV P. Prototype-based continual learning with label-free replay buffer and cluster preservation loss[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville: IEEE, 2025: 6547–6556.
- [56] FAN Yan, WANG Yu, ZHU Pengfei, et al. Dynamic sub-graph distillation for robust semi-supervised continual learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2024: 11927–11935.
- [57] WU Zhirong, XIONG Yuanjun, YU S X, et al. Unsupervised feature learning via non-parametric instance discrimination[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3733–3742.

#### 作者简介:



朱广泽, 硕士研究生, 主要研究方向为无监督持续学习、联邦学习和跨模态检索。E-mail: [228239765@qq.com](mailto:228239765@qq.com)。



刘兴波, 教授, 博士生导师, CCF AI 专委会执行委员、CCML 专委会通讯委员。主要研究方向为多媒体检索、计算机视觉。获山东省科技进步二等奖 1 项、山东省人工智能自然科学一等奖 1 项。以第一或通信作者发表学术论文 17 篇。E-mail: [sclxb@mail.sdu.edu.cn](mailto:sclxb@mail.sdu.edu.cn)。



聂秀山, 教授, 博士生导师, 山东建筑大学计算机和人工智能学院院长, CCF AI 专委会秘书、CCML 专委会执行委员。主要研究方向为多媒体检索、计算机视觉、机器学习、数据挖掘。获山东省科技进步二等奖 2 项、山东省人工智能自然科学一等奖 1 项。以第一或通信作者发表学术论文 30 余篇。E-mail: [nixsh@hotmail.com](mailto:nixsh@hotmail.com)。