

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-17

论文引用格式: Lan Tiansheng, Zhang Ling, Liu Libo, Zhang Ruonan. Wave-particle duality-driven feature representation paradigm for point cloud place recognition[J/OL]. Journal of Image and Graphics, XXXX: 1-17. DOI: 10.11834/jig.250631. (兰天升, 张玲, 刘立波, 张若楠. 波粒二象性驱动的点云场景识别特征表达范式[J/OL]. 中国图象图形学报, XXXX: 1-17. DOI: 10.11834/jig.250631. [DOI: 10.11834/jig.250631])

波粒二象性驱动的点云场景识别特征表达范式

兰天升¹, 张玲¹, 刘立波², 张若楠^{1*}

1. 宁夏大学前沿交叉学院, 宁夏中卫市 755000; 2. 宁夏大学信息工程学院, 宁夏银川市 750021

摘要: 目的 针对现有有点云场景识别方法中传统实数域卷积运算表达能力有限, 难以充分挖掘点云场景中丰富结构及语义信息的问题, 本文探索一种融合跨学科理论的新型点云特征表达范式, 以提升点云场景识别的性能与可解释性。方法 受物理学波粒二象性理论启发, 本文将点云中的点抽象建模为基本态粒子, 创新地构建兼具“波动性”与“粒子性”的点云特征表达范式, 并提出由三大核心模块组成的场景识别方法: 1) 点云波动性表达模块: 基于单粒子的波动属性, 对点云中各点的语义信息进行完备性建模, 增强其局部上下文感知能力; 2) 点云粒子性信息交互模块: 模拟多粒子相干作用下的“粒子性”行为, 刻画点特征间的动态依赖与高阶交互关系; 3) 点云特征紧凑化编码模块: 将融合波粒二象性的点特征高效聚合为紧凑且具有代表性的全局表示, 用于最终的识别任务。在实现上, 本文为每个模块提出两种实例进行展示。结果 在四个数据集上的综合评估显示, 所提范式下的不同实例均能有效提升多种点云识别基准模型的性能; 其中“波函数直接建模与广义平均池化”的组合计算高效, 以 6.12MB 参数量使模型收敛速度提升近 1 倍, 达到 8ms。结论 所提波粒二象性特征表达范式具有良好的通用性和即插即用特性, 验证了跨学科理论应用于点云场景特征表达的有效性, 为点云场景识别提供了新的研究方向。

关键词: 点云; 场景识别; 波粒二象性; 特征表达; 深度学习

Wave-particle duality-driven feature representation paradigm for point cloud place recognition

Lan Tiansheng¹, Zhang Ling¹, Liu Libo², Zhang Ruonan^{1*}

1. School of Advanced Interdisciplinary Studies, Ningxia University, Zhongwei 755000, Ningxia Hui Autonomous Region, China; 2. School of Information Engineering, Ningxia University, Yinchuan 750021, Ningxia Hui Autonomous Region, China

Abstract: Objective As a fundamental data modality for autonomous driving, robotic navigation, and large-scale environmental perception, 3D point clouds pose stringent demands on place recognition technologies. Accurate place recognition relies on extracting robust environmental descriptors; however, despite recent advancements, mainstream deep learning methods are fundamentally constrained by three prominent limitations. First, conventional feature representations are strictly confined to the real-number domain, which inherently restricts models to capturing only geometric coordinates and local structural information. This paradigm discards crucial spatial attributes such as the point distribution phase, the rela-

收稿日期: 2025-12-16; 修回日期: 2026-05-06

* 通信作者: 张若楠, 1992年生, 博士, 准聘副教授. E-mail: zhangm@nxu.edu.cn

基金项目: 宁夏自然科学基金(No. 2026AAC050038); 国家自然科学基金(No. 62506179); 银川市科技项目(No. 2025RC09); 宁夏科技创新领军人才计划项目(No. 2022GKLRXLX03)

Supported by: Natural Science Foundation of Ningxia (No. 2026AAC050038); National Natural Science Foundation of China (No. 62506179); Science and Technology Project of Yinchuan City (No. 2025RC09); Ningxia Science and Technology Innovation Leading Talent Program Project (No. 2022GKLRXLX03)

tive positions of non-adjacent points, and complex global topological relations. Consequently, this leads to low recognition accuracy for places exhibiting similar local structures but possessing distinct global configurations. Second, existing feature interaction mechanisms struggle to align with the "unordered, sparse, and irregular" intrinsic properties of raw point clouds. Traditional grid-based convolutions, which rely on fixed local receptive fields, often suffer from "local information over-focus," severely neglecting long-range semantic associations necessary for holistic place understanding. Finally, resolving the pervasive trade-off between recognition performance and computational efficiency remains elusive. High-accuracy baseline models typically demand excessive network depth and parameter scales—often reaching tens or hundreds of megabytes—yielding inference latencies that far exceed the 10ms real-time threshold, thereby failing to meet the strict practical application requirements of autonomous systems. Aiming to systematically address these limitations of traditional real-number domain operations, this study explores a novel point cloud feature representation paradigm deeply integrated with interdisciplinary theories, striving to significantly enhance both the performance and interpretability of large-scale point cloud place recognition. **Methods** Inspired by the foundational wave-particle duality theory in physics, this study abstracts each individual point in the 3D point cloud as an elementary particle with fundamental physical states, innovatively constructing a comprehensive point cloud feature representation framework that embodies both "wave nature" and "particle nature". Correspondingly, an end-to-end place recognition method consisting of three synergistic core modules is proposed, with two specific implementation instances designed for each module to demonstrate the paradigm's remarkable flexibility. 1) Point Cloud Wave Property Expression Module: Leveraging the wave attribute of single particles, this initial module maps raw 3D point clouds from the constrained real-number space into a high-dimensional Hilbert space, representing each point as a complex-valued wave function. The amplitude component encodes fundamental geometric positions and local structural details, whereas the phase component intricately captures relative spatial relationships and global distribution trends. This dual encoding achieves comprehensive modeling of both geometric and semantic point information, vastly enhancing local context perception. 2) Point Cloud Particle Property Information Interaction Module: Emulating the "particle nature" behavior observed under multi-particle coherence, this phase-driven interaction mechanism constructs a dynamic spatial graph based on the wave function similarities between points, treating individual points as graph nodes and interaction intensities as adaptive edge weights. By mathematically simulating particle energy transfer and interactive information exchange, it facilitates the adaptive fusion of high-dimensional feature information across both immediate adjacent points and semantically correlated non-adjacent distant points. This efficiently characterizes the dynamic dependencies and high-order semantic interactions that conventional real-number convolutions miss. 3) Point Cloud Feature Compact Encoding Module: To address the inherent high dimensionality and computational burden of complex-valued features, this final module leverages parameter-sharing complex-valued convolutions for deep feature extraction. It then applies efficient aggregation strategies—specifically proposing a Generalized Mean (GeM) pooling approach or a structurally refined Soft-NetVLAD instance—to compress these high-dimensional interactions into a highly compact, representative global feature vector tailored for the final retrieval and recognition task. Results Comprehensive and rigorous evaluations were conducted across four foundational benchmark datasets (Oxford, U. S. , R. A. , B. D.) and two complex generalization datasets (KITTI Sequence 00, RobotCar Season). The empirical findings confirm that all instances of the proposed interdisciplinary paradigm effectively and consistently enhance the performance of established benchmark models—including PointNet++, MinkLoc3D, and PCAN—in terms of both top-1% average recall and retrieval precision. Notably, the specific combination of direct wave function modeling and GeM pooling achieves an unprecedented balance between performance and deployment efficiency. Requiring merely 6.12MB of learnable parameters, this streamlined instance nearly doubles the model convergence speed during training and reduces the inference time to a mere 8ms, strictly satisfying the critical 10ms real-time processing requirement dictated by autonomous driving and robotic navigation systems. Furthermore, the wave-particle paradigm exhibits exceptional robustness; in rigorous perturbation tests, it delivered a 7.1% average performance gain under severe rotational variances and secured a 5.8% average performance enhancement in demanding cross-domain environmental tests. Extensive ablation studies unequivocally corroborate the structural necessity of modeling wave-particle coherence, demonstrating that the removal of any core module directly leads to immediate and notable performance degradation. **Conclusion** The proposed wave-particle duality paradigm exhibits excellent architectural versatility and plug-and-play charac-

teristics, enabling significant performance enhancements of existing baseline models without necessitating extensive structural modifications. It empirically verifies the profound effectiveness of interdisciplinary theoretical fusion in successfully breaking through traditional point cloud feature representation bottlenecks, resolving issues including real-number domain spatial limitations, mismatched feature interaction mechanisms, and rigid performance-efficiency trade-offs. This work provides a robust, interpretable, and highly efficient new research direction for point cloud place recognition, paving the way for future optimizations in complex-valued computation and expanding scenario-specific applications.

Key words: point cloud; place recognition; wave-particle duality; feature representation; deep learning

Lan Tiansheng, Zhang Ling, Liu Libo, Zhang Ruonan. Wave-particle duality driven feature representation paradigm for point cloud place recognition [J/O]. 中国图象图形学报. DOI: 10.11834/jig.250631. (兰天升, 张玲, 刘立波, 张若楠. 波粒二象性驱动的点云场景识别特征表达范式[J/O]. 中国图象图形学报. DOI: 10.11834/jig.250631.)

0 引言

点云作为自动驾驶、机器人导航等领域的核心数据形式,其场景识别技术因为能够为环境感知提供关键位置信息而成为研究热点(Qi等,2017; Komorowski等,2021)。随着深度学习技术的发展,点云场景识别从传统手工特征提取进入端到端学习阶段——国际上,PointNet/PointNet++(Qi等,2017)解决了处理无序点云的问题,PointNetVLAD算法(PointNetVLAD,PN-VLAD)(Uy等,2018)、明科夫斯基基定位3D算法(Minkowski-Based Localization 3D, MinkLoc3D)(Komorowski等,2021)通过优化特征聚合策略提升了检索性能与效率,Transformer与图卷积技术(Hui等,2021;Shu等,2024)进一步强化了点间特征交互能力;国内研究则在国际算法框架的基础上,更聚焦技术的工程化落地与场景适配,形成了贴合实际应用需求的研究方向。Zhang等(2019)提出的PCAN、Sun等(2020)提出的DAGC,分别通过注意力机制与双路径图卷积优化,强化了点云局部与全局特征的关联,有效提升了复杂场景下特征的判别能力。针对端侧设备的算力限制,Liu等(2024)设计的LPS-Net通过参数共享机制实现了模型轻量化,Li等(2025)提出的双路径Transformer则兼顾了多尺度特征感知与推理效率,二者很好地平衡了识别精度与部署成本。此外,Pan等(2025)还将相关技术拓展至航空点云场景,通过多特征融合的点

Transformer实现了大尺度遥感点云的精准分类,进一步拓宽了技术的应用边界。

尽管现有研究已取得显著进展,但三大局限仍制约着技术的进一步发展:其一,实数域表达的固有缺陷,主流方法仅能捕捉几何位置与局部结构信息,丢失点分布相位、非相邻点相对位置及全局拓扑等关键属性,导致局部相似但全局不同场景的识别精度偏低(Li等,2025);其二,特征交互机制与点云特性不匹配,传统卷积依赖固定网格与局部感受野(Komorowski等,2021),与点云“无序、稀疏、不规则”的属性矛盾,易出现“局部信息过聚焦”,忽略全局语义关联(Sun等,2020);其三,性能与效率难以平衡,高性能方法多通过增加模型深度与参数规模以提升精度,如PPT系列方法(Hui等,2021;Hou等,2022),但推理速度远超10ms实时阈值,无法满足自动驾驶等场景需求(Komorowski等,2021;Qiu等,2024)。

深入研究发现,当前点云场景识别研究面临的瓶颈,主要在于其特征表达域存在局限。这种受限的表达空间不仅制约了局部特征的代表能力,也在很大程度上限制了后续的特征间交互机制的建模空间,从而导致整体识别性能受限。为突破实值域表达的不完备性所带来的约束,本文从量子力学的波粒二象性理论中获得启发。该理论构建于希尔伯特空间,相较传统实值欧氏空间,具备更强的数学完备性与表达能力,从而为点云中点特征的建模提供坚实的理论支撑。

进一步而言,波粒二象性揭示了微观粒子同时具备波动性与粒子性。这一物理现象又为点云特征的高阶建模提供新的视角:可将点云中的每个点抽象为一个“兼具波动性与粒子性的基本态粒子”,从而在点级别上实现信息表达的完备性与交互过程的统一。

基于此,本研究创新地提出了一种波粒二象性
©中国图象图形学报版权所有

驱动的点云特征表达范式。该范式首先将点云映射至复数域,并构建对应的“波动性”函数,以突破实数域信息表达的局限性;其次,引入粒子相干机制,模拟多粒子系统中粒子的动态交互,以增强特征间的高阶语义交互能力;最后,设计高效的特征紧凑化编码模块,以实现融合波粒二象性信息后的特征精简。本方法旨在为点云场景识别技术提供一种兼具高识别精度、结构轻量化与强泛化能力的研究思路,助力点云场景识别技术在实际场景中的高效转化。

1 相关工作

深度学习技术推动点云场景识别从无序点云处理向高精度特征表达发展,现有识别方法按特征载体的不同,可分为基于点的特征表达方法与基于体素的特征表达方法两类,详细分析如下:

1.1 基于点的特征表达方法

此类方法直接对原始点云进行特征学习,核心逻辑是通过“局部特征提取—全局聚合”构建场景描述子,适配点云稀疏、不规则的特性。其中,PN-VLAD利用NetVLAD编码PointNet提取的局部特征,生成具有强判别力的全局描述子,以显著提升大规模场景检索性能,但VLAD聚合过程易丢失精细几何细节;PCAN结合PointNet++与注意力机制,强化局部特征的判别力,DAGC通过双向注意力与动态图卷积网络(Dynamic Graph Convolutional Neural Network, DGCNN)建模特征动态关联,但两者均因结构复杂度增加导致计算开销增加;邻居点特征学习网络(Learning Features Between Neighboring Points, LBNP)(Wang等,2025)通过学习邻居点间的相互关系,以捕捉更精细的局部几何结构;胶囊基定位3D网络(Capsule-based Localization 3D Network, CapsLoc3D)(Zhang等,2024)采用胶囊网络实现特征紧凑化表达,以减少冗余信息;双路径Transformer(Li等,2025)解耦位置信息与几何关系,实现宏观分布与微观结构的同步感知;PPT系列方法(Hui等,2021;Shu等,2024)通过层次化架构以深度挖掘多尺度特征,但其多金字塔层级与图卷积模块加剧了模型复杂度;(Pan等2025)针对航空点云场景,将Point Transformer与多特征融合策略结合,有效解决了地面物体区分难题,拓展了技术的应用边界。

此外,近年来在点云特征表达的完备性与轻量

化设计方面出现了一批创新研究。在局部特征捕捉与表示方面,王昌硕等(2023)针对现有网络难以有效捕捉准确局部形状信息的问题,提出一种基于局部区域动态覆盖的3D点云分类方法,通过构建各向异性的空间几何体覆盖局部特征空间,显著增强了特征表达对稀疏点云的鲁棒性;程浩喆等(2025)针对大规模无标注点云特征表达泛化性不足的问题,提出基于双向拟合掩码重建的多模态自监督点云表示学习方法,有效挖掘了点云的内在结构特征。同时,为兼顾性能与效率,周新志等(2024)利用灵活的图卷积操作以自适应捕捉动态结构,在保证轻量化参数的同时提升了特征提取效率。此类现有方法均在实数域内进行特征优化,无法有效表征点分布的相位特性与全局拓扑结构;部分方法为兼顾全局与局部信息,牺牲了模型的轻量化与通用性,如大规模点云描述子(Large-Scale Point Cloud Descriptor, LPD)(Liu等,2019)引入先验知识提升性能,却限制了其在不同场景的适配能力。

1.2 基于体素的特征表达方法

此类方法通过将点云离散化为较为规则的体素网格,利用稀疏3D卷积以提取特征,其核心优势是能够复用成熟的2D卷积技术,平衡特征表达能力与计算效率。其中,MinkLoc3D(Komorowski等,2021)率先采用明科夫斯基卷积处理体素化点云,并用广义平均池化(Generalized Mean Pooling, GeM)替代NetVLAD,在简化特征表达的同时提升了推理效率;选择性特征融合定位(Selective Feature Fusion Localization, SelfLoc)(Qiu等,2024)设计选择性特征融合机制,进一步增强特征判别力;体素曼巴网络(Voxel Mamba Network, VMN)(Zhang等,2024)基于状态空间模型构建体素骨干网络,通过群组无关策略提升计算效率,缓解空间邻近性损失;体素一点云特征融合(Voxel-Point Feature Fusion, VPF)(Geiger等,2012)采用混合体素一点云表示,融合两类特征的优势,实现更高效的3D特征提取。

虽然体素化方法在特征提取效率上具有优势,且被广泛拓展应用于各种3D视觉任务中,但体素化过程会不可避免地丢失点云的精细几何结构。例如,周昊等(2024)指出在基于体素的3D目标检测中,对点云进行体素化会损失部分数据信息和结构信息,由此提出一种融合点云深度信息的补偿方法。然而,在面对大规模场景识别时,此类稀疏卷积的计

算开销会随体素分辨率提升而显著增加,仍难以同时满足高精度识别与轻量化部署的需求。

综上可知,现有方法普遍未能突破实数域表达的局限,且性能与效率的平衡多以牺牲特征丰富性或通用性为代价。本研究提出的波粒二象性驱动的点云特征表达范式,从跨学科理论融合视角出发,为解决上述问题提供了新的技术路径。

2 波粒二象性的点云场景特征表达

为更充分地挖掘点云场景中的完备信息以提升点云场景识别性能,本研究提出一种波粒二象性驱动的大规模点云场景识别范式,如图1所示,该范式包括三个核心模块:点云波性表达模块、点云粒子性信息交互模块与点云特征紧凑化编码模块,定义为:

$$F(X) = E(P(W(X))) \quad (1)$$

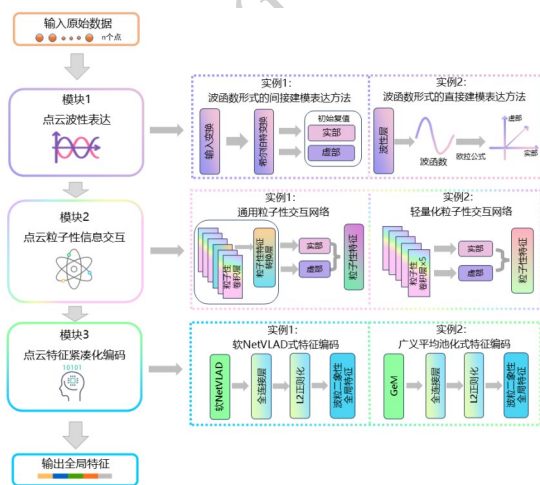


图1 波粒二象性的点云场景识别特征表达范式

Fig. 1 Wave-Particle Duality-Based Feature Representation Paradigm for Point Cloud Scene Recognition

式中 $X = [x_1, x_2, \dots, x_n]$ 为一个具有 n 个点的点云场景。点云波性表达模块 $W(\cdot)$ 利用波粒二象性中的“波动性”以完备地表示点云,可有效解决点云场景特征表达完备性受限的问题;点云粒子性信息交互模块 $P(\cdot)$ 依据波粒二象性中“粒子性”有效地反映了基本粒子间的融合机理,可有效解决点云场景中信息交互欠佳的问题;点云特征紧凑化编码模块 $E(\cdot)$ 则对点云场景特征进一步地精简,输出紧凑且适配的点云特征表达,以实现高效的点云场景识别,并为下游点云任务提供特征表达的提取思路。

以下将具体介绍每个模块。

2.1 模块1:点云波性表达

点云波性表达模块将欧氏空间三维坐标表示的点云转换为具有波性的粒子,即“波性”粒子。这区别于传统点云表示方法受限于欧氏空间实数域的情景,解决了传统方法难以充分挖掘点云表达的丰富性信息的问题。该模块通过将点云中离散的点映射到连续的波性表示,为后续在复数域或更高维度的操作提供基础,突破传统实数域表达的限制,进而更完备地表示点云信息。本质上,“波动性”式建模模拟物理学中单个粒子在希尔伯特空间具有波动性的现象,较欧氏空间更完备地表现粒子属性,并为探索点云更丰富的信息表达提供途径。

2.2 模块2:点云粒子性信息交互

点云粒子性信息交互模块是实现“波性”粒子交互的过程,即波粒二象性中“粒子性”的体现形式,主要通过对多个单独的波进行混合操作实现。具体而言,本研究引入粒子相干方法来探索点间的相关性与深层特征,捕捉整体关联关系。相干过程将点的“波性”信息进行有效融汇,形成能够反映点云场景特征整体分布、结构及交互关系的信息交互。本质上,点云粒子性信息交互模块模拟了物理学中大量具有“波性”的粒子相互作用涌现出“粒子性”的现象,实现了从单个粒子独立性信息向多个粒子聚合性信息的有效转换,并为获得紧凑且信息丰富的全局点云场景特征表达奠定基础。

2.3 模块3:点云特征紧凑化编码

点云特征紧凑化编码模块负责对特征进行精简并输出。鉴于粒子性信息交互模块已尽可能全面地探索了点云场景特征表达的可能性,导致其特征维度较高,因此应用于点云场景识别任务时,需该模块对粒子性特征进一步精简,输出紧凑的全局点云场景特征,从而形成最终的波粒二象性融合表达。

3 波粒二象性的场景特征表达实例

为验证第2节所述“波粒二象性特征表达范式”的有效性与灵活性,本章将针对范式的点云波性表达模块、点云粒子性信息交互模块以及点云特征紧凑化编码模块,分别设计并展示两种具体实例(如图1所示),旨在为点云场景识别方法提供多样化的特征表达途径。

为清晰展现所提范式的实现逻辑与各核心模块间的协同演进关系,本节将上述模块的执行流程总结为算法1。该算法不仅涵盖从原始点云输入到全局特征向量输出的完整数据流向,还详细展示了各模块在不同实例配置(如直接/间接建模、通用/轻量化交互及多样化编码策略)下的分支路径,为后续小节性能分析与实验验证提供了明确的计算逻辑支撑。

3.1 点云波性表达模块

为建模完备的点云场景特征表达范式,本研究

Algorithm 1: 波粒二象性驱动的点云场景识别特征算法

Input: 原始点云数据 $X = \{x_1, x_2, \dots, x_n\}$

Output: 全局特征向量 F_{out}

1. 对原始点云数据 X 进行预处理(详见 4.1.1 节)

// 模块1:点云波性表达

2. **if** 波函数间接建模:

3. 令 X_{reg} 为通过 T-Net 正则化后的 X ;

4. 利用希尔伯特变换计算初始复值表达 $c(x)$ (公式3)

5. **else if** 波函数直接建模:

6. 计算点云 X 中第 j 个点的幅值 f_j^a 和相位 θ_j (公式4)

7. 根据欧拉公式重构复值波函数 \hat{f}_j (公式6)

// 模块2:点云粒子性信息交互

8. **if** 通用粒子性交互网络:

9. 通过5层 Particle-Conv 与1层 Transform layer 提取特征;

10. 执行复值卷积并取模融合特征,得到 X' (公式11)

11. **else if** 轻量化粒子性交互网络:

12. 移除特征转换层

13. 计算点间动态相干交互特征 $\{c_j\}, j \in [1, n]$ (公式10)

// 模块3:点云特征紧凑化编码模块

14. **if** 软 NetVLAD 式特征编码:

15. 基于软聚类中心 s 修正权重并聚合局部特征(公式13)

16. 经全连接层与 L_2 归一化输出 F_{out} (公式14)

17. **else if** 广义平均池化(GeM)式编码:

18. 设定广义平均池化的指数参数 $i = 3$

19. 对交互特征进行全局聚合以计算 F_{out} (公式15)

20. **return** F_{out}

从波粒二象性的波动性角度出发,通过将点云中的每个点类比为微观粒子,并利用其波动性特征,将点构建为能够完备描述其空间信息和潜在属性的波函数形式。其定义如下:

$$\hat{f}_j = |f_j^a| \cdot e^{i\theta_j}, j = 1, 2, \dots, n \quad (2)$$

式中 $|\cdot|$ 被视为幅值的模数, f_j^a 对应于第 j 个点的振幅信息, i 为虚数部分的表示,满足 $i^2 = -1$, θ_j 为 j 点在一个波周期里的相位,表示第 j 个点的当前位置, $e^{i\theta_j}$ 则表示相位的周期函数。使用该波函数,每个中间特征都可以用复值来进行表示,而复值的获得则需要相位和振幅的辅助。基于式(2),本研究提出了两种实例方法,分别为波函数形式的直接建模表达方法与波函数形式的间接建模表达方法。

3.1.1 波函数形式的间接建模表达方法

间接建模表达方法将点云特征间接建模为复数域中的表达。该方法利用了复数能够同时承载振幅和相位信息的特性,与式(2)通过欧拉公式展开后的复值形式在数学上是等价的。首先,通过一个变换网络(Transform Network, T-Net)正则化点云 P ,随后利用式(1)将实数域点云转换为初始复值表达。在训练过程中,该复值表达的初始虚部由希尔伯特变换生成,其最终虚部则通过网络学习得到。为详细阐述这一转换过程,本研究将点云中的每个点 x 定义为复值表达,其对应的转换过程如式(3)所示:

$$c(x) = r(x) + i\hat{r}(x), \hat{r}(x) = \text{Hilbert}[r(x)] \quad (3)$$

在上述公式中, $\text{Hilbert}[\cdot]$ 表示希尔伯特变换公式,其定义为:

$$\text{Hilbert}[X] = X * \frac{1 - (-1)^t}{\pi t} \quad (4)$$

式中, t 是时间变量,为非线性的积分过程。因此,在训练过程中,可采用一个卷积操作进行逼近。其中 $c(x)$, $r(x)$, $\hat{r}(x)$ 分别代表点云的复数域表达、实部和虚部,这里的虚部可以由实部经过希尔伯特变换得到。

3.1.2 波函数形式的直接建模表达方法

直接建模表达方法是直接通过式(2)将点云建模为波函数形式。首先定义点的波性层为:

$$f_j^a = W^a \cdot f_j^{in}, j \in [1, n] \quad (4)$$

式中, f_j^{in} 为表示波性层的输入。不同于式(2),该式中去除了幅值的模值计算,由于 $f_j^a e^{i\theta_j} = f_j^a e^{i\theta_j + \pi}$,即可

通过调节相位来实现去除模值的计算。对于点云波表示中的相位,可用波性层来进行建模。不同之处在于,此处需使用激活函数约束相位范围,使其满足在区间 $m \in [0, 2\pi]$ 内,定义为:

$$\theta_j = 2\pi \cdot \text{Sigmoid}(W^{\theta} \cdot f_j^{\text{in}}), j \in [1, n] \quad (5)$$

除此之外,该相位表达还可以使用其他简化形式,例如:简化为无区间限制。实际操作中,式(4)与(5)需要构建指数级表达,操作并不便利。因此可采用欧拉公式将其进行展开,具体如下:

$$\hat{f}_j = |f_j^a| \cdot \cos \theta_j + i |f_j^a| \cdot \sin \theta_j, j = 1, 2, \dots, n \quad (6)$$

式中,第 j 个点的中间特征 \hat{f}_j 可以分解为两个实值部分,分为实部和虚部。相比波函数间接建模表达方法而言,虽然两种方法最终都表示为复值,但该方法精度更高。因为间接建模表达方法忽略波函数到复值的转换过程,而是直接利用希尔伯特变换模拟并生成复值表达,本质是通过一层卷积层近似。

3.2 点云粒子性交互动模块

在点的波性建模之上,为实现更有效地建模点云特征之间的动态交互与融合,本研究提出点云粒子性交互动模块。该模块通过将点云特征视为可相互作用的粒子,使得特征间的关系可以被显式地表达和操作。基于式(2),根据点云中两个点之间的相互作用,绘制对应波,如图2所示。给定 $\hat{f} = \hat{f}_1 + \hat{f}_2$ 作为相互作用后的表示。 \hat{f} 的相位和振幅可以通过以下方式计算:

$$\theta_j = \theta_1 + \psi(|f_2| \cdot \sin(\theta_2 - \theta_1)) \quad (7)$$

$$|f| = |f_1| + |f_2| \cos(\theta_2 - \theta_1) \quad (8)$$

$$|f|^2 = |f_1|^2 + |f_2|^2 + 2|f_1||f_2|\cos(\theta_2 - \theta_1) \quad (9)$$

在上述方式中, ψ 表示两个输入的反正切函数。振幅受相位差的影响,存在三种情况,分别对应图2中的(a)、(b)和(c)所示的组合效果。在这些情况下,两个波相互作用较为复杂,具体情况取决于相位差。而现有大多数点云处理方法仅使用实数域引导操作,是该公式所表示的一个特例,忽略了相位的重要性,直接将其赋值为 π 的整数倍,为 $\theta_j = \pi m$,其中 m 是一个整数,对应波表示当中, $e^{i\theta} = -1$ 也因此忽略了相位所带来的波相互作用之间的自由度,因而表示特征时其完备性受限。

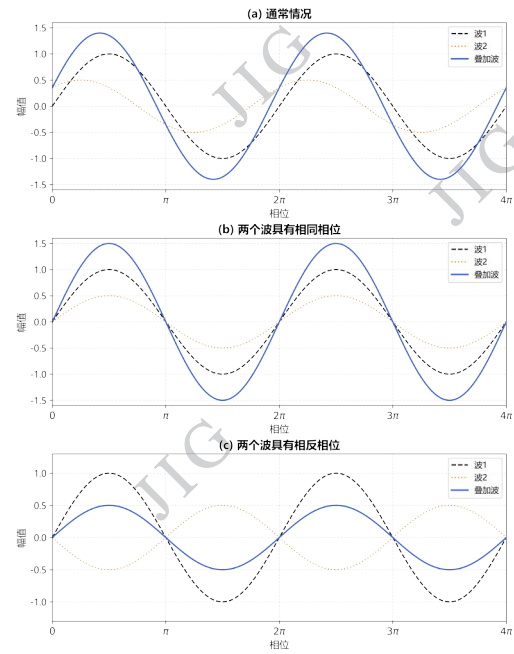


图2 点云的点波相互作用

Fig. 2 Point-Wave Interaction of Point Clouds

通过对点云的波进行表示,可明显得出波表示下的点在表示信息层面更具完备性,通过在点交互过程中互动信息,进而为后续的性能提升提供途径。因此,整个交互过程可表示为: $c_j = W_j^R f_j^a \cdot \cos \theta_j + W_j^I f_j^a \cdot \sin \theta_j, j \in [1, n]$ (10)

式中, W_j^R 和 W_j^I 分别为第 j 个点的实部和虚部的权重值, θ_j 表示第 j 个点的相位角值。注意到, θ_j 相位角动态地参与了整个交互过程,而这是现有的大部分基于点学习的点云场景识别方法中所忽略的。基于此,本研究进一步设计了两种实例网络,分别是通用粒子性交互动网络和轻量化粒子交互网络。

3.2.1 通用粒子性交互动网络

该网络顺承4.1节获取到的点云复值表达,进行复值卷积操作(如图3),整体经过5层粒子性卷积层和1个粒子性特征转换层及最后的归一化操作完成。最后对每个点都输出1024维的特征。每个粒子性卷积层根据式(10),可简化为:

$$F * X = (A * R - B * I) + i(B * R + A * I) \quad (11)$$

式中, F 与 X 分别表示复值滤波与复值点云, $F = A + iB$, $X = R + iI$ 。最后的求模过程则表示为 $X' = \sqrt{R^2 + I^2}$ 。其中 R 与 I 分别表示最后一层的实部与虚部的部分。该网络通过复值卷积操作,实现了粒子性交互动过程,有助于增强特征的代表性,并促进最终性能的提升。

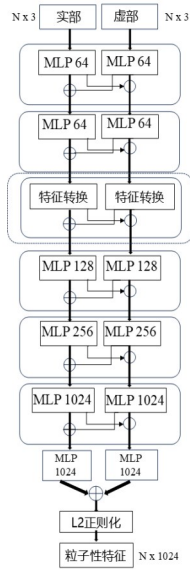


图3 点云粒子性交互模块网络实例

Fig. 3 Network Instance of the Point Cloud Particle Interaction Module

3.2.2 轻量化粒子性交互网络

与上一小节所不同,本小节致力于更轻量化的粒子性交互过程,在实际验证过程中,可考虑去除图3中虚线框部分的粒子性特征转换层,但前提是

配合使用直接表达方法使用。这是由于直接表达方法从起始处对点云中的点进行建模,可表达更加完备的场景信息,故在此无需使用粒子性特征变换层以挖掘更多信息,会存在冗余性。

3.3 点云特征紧凑化编码模块

为进一步精简点云特征表达,本研究提出两种实例方法,分别为软NetVLAD式特征编码与广义平均池化式特征编码。

3.3.1 软NetVLAD式特征编码

在实现计算全局特征时,采用NetVLAD^[4]改进版本,即软NetVLAD式特征编码方法,实现对所提取的复值特征进行紧凑编码,最终获取到一个全局的紧凑的特征表达。为此,本节引入软聚类中心 s ,使其参与聚类中心的权值分配,但并不参与最后的计算。这样可削弱不重要聚类中心的权重占比,该过程可视为在编码融合过程中动态调节和去冗余的因子,使得特征间可以进行信息传递。当不使用软聚类时,该模块退化为NetVLAD。具体定义如下:定义 $K+T$ 个聚类中心,仅 K 个聚类中心作用于最终结果,而 T 个聚类中心则作用于 K 个聚类中心的权

重修正。定义场景视觉词的聚类中心集 $c = \{c_1, c_2, \dots, c_{K+T}\}$, $c_k \in \mathbb{R}^D$,最终特征表示为:

$$S(X) = \{S_1(X'), \dots, S_k(X')\}, S_k(X') \in \mathbb{R}^D \quad (12)$$

$$S_k(X') = \sum_{i=1}^n \left[\frac{(e^{w_k^T x'_i + b_k})}{\sum_{k'} e^{w_{k'}^T x'_i + b_{k'}}} \right] (x'_i - c_k) \quad (13)$$

式中,权重项为 $\{w_k\}$,偏置项为 $\{b_k\}$,式(13)表示每个特征向量 x_i 对于特征词 $S_k(X')$ 的贡献。之后,使用一个全连接层 $\sigma(\cdot)$ 配合 L_2 范式得到最终的全局特征,表示为:

$$F_{out} = \|\sigma(S(X'))\|_2 \quad (14)$$

3.3.2 广义平均池化式特征编码

软NetVLAD固好,但针对其聚类参数的学习耗时费力,故另提供一种较为轻量级的广义平均池化方法,定义为:

$$F_{out} = \phi \left(\left(\frac{i}{n} \sum_{j=1}^n c_j^i \right)^{\frac{1}{i}} \right) \quad (15)$$

式中, i 为广义平均池化的指数参数。 c_j^i 为式(10)的输出交互特征。实际中设置 $i=3$ 。最后,再添加一个全连接层,记为 ϕ ,以获得最终的紧凑的具有代表性的特征 F_{out} 。

3.4 训练策略

为优化模型训练并提升点云场景识别性能,本节详细探讨并比较了四种关键训练策略:三元组损失 $\ell_T(T_3)$ 、最大三元组损失 $\ell_{MaxT}(T_3)$ 、四元组损失 $\ell_Q(T_4)$ 以及最大四元组损失 $\ell_{MaxQ}(T_4)$ 。这些策略在负样本优化上有所侧重:基于最大优化的方法优先处理难负样本,而基于求和的方法则倾向于优化平均难度的负样本。具体而言,三元组损失通过引入边距,旨在确保负样本与锚点的距离始终大于正样本与锚点的距离,从而有效提升负样本的区分度。在此基础上,四元组损失进一步引入错样本,以约束三元组中 $\ell_T(T_3)$ 负样本组之间的差异性,增强模型的判别能力。以上数学表达式为:

$$T_3 = (X_a, X_p, \{X_n\}) \quad (16)$$

$$T_4 = (X_a, X_p, \{X_n\}, X_n^-) \quad (17)$$

$$\ell_T(T_3) = \sum_i H[\rho_p - \rho_{n_i} + \alpha] \quad (18)$$

$$\ell_{MaxT}(T_3) = \max_i H[\rho_p - \rho_{n_i} + \alpha] \quad (19)$$

$$\ell_Q(T_4) = \ell_T(T_3) + \sum_i H[\rho_p - \rho_{n_i} + \beta] \quad (20)$$

$$\ell_{MaxQ}(T_4) = \ell_{MaxQ}(T_3) + \max_i H[\rho_p - \rho_{n_i} + \beta] \quad (21)$$

式中, T_3 与 T_4 分别为三元组、四元组。 X_a, X_p, X_n 和 X_{n_i} 分别表示为锚点、正样本、负样本集与错样本。 $H[\cdot]$ 表示最大边缘计算中使用的 Hinge 损失。 样本间差异性通过欧氏距离 $d(\cdot)$ 进行度量, 涉及三类距离计算:

$$\rho_p = d(G(X_a), G(X_p)), \rho_{n_i} = d(G(X_a), G(X_{n_i}))$$

$$\rho_{n_i} = d(G(X_{n_j}), G(X_n))$$

分别为正样本与锚点之间的距离、负样本与锚点之间的距离、错样本与负样本的距离。式(18)和式(19)中 α 为边距。式(20)和式(21)中 β 为另一组正样本与错样本之间的边距。

4 实验

本节旨在通过多维度的实验以全面验证所提波粒二象性特征表达范式在点云场景识别任务中的有效性、泛化性及轻量化优势。实验方案主要从四个方面展开: 首先, 在多个大规模公开数据集上评估不同实例的识别精度, 并与现有主流基准模型进行对比; 然后, 针对模型参数规模与推理耗时进行量化评估, 验证其在实际部署中的实时性表现; 其次, 通过跨数据集实验及鲁棒性测试, 验证模型在不同复杂环境下的泛化能力; 最后, 进行消融实验以深入分析各核心模块及超参数设置的合理性与必要性。

4.1 数据集及实验设置

4.1.1 数据集

该部分涉及两类数据集, 第一类是基础训练测试数据集, 而第二类则是泛化性测试数据集。

第一类数据集构建于使用激光雷达设备采集得到的四个数据集上, 包括 Oxford、University Sector (U. S.)、Residential Area (R. A.) 与 Business District (B. D.)。在训练前, 需要对这些数据集进行预处理操作, 包括去除地面、切分子图及对切分后的子图进行归一化处理。随后将预处理后的子图作为网络输入进行训练, 每个子图关联 1 个 UTM 坐标以指示其位置。针对训练元组构建, 距离在 10 米内视为正样本, 距离大于等于 50 米视为负样本。

第二类数据集包括: 1) KITTI Sequence 00 (Geiger A 等, 2012) 数据集, 该数据集的前 170 个场景用于构建数据库, 其余部分用作查询序列。将每个场景中的前 4096 个点应用于测试。2) RobotCar Season (Maddern W 等, 2016) 数据集, 每张图像均对应一张

点云。本实验涉及六个类别, 即夏季、雪、黄昏、太阳、夜雨和参考。参考类别用作数据库, 而其他类别则用作查询序列。

4.1.2 评估环境及指标

本实验使用 Tensorflow 与 Pytorch 深度学习框架实现, 网络的训练是在 Nvidia T4 与 V100 图形处理器 (Graphics Processing Unit, GPU) 环境下进行。点云场景识别任务通常使用的衡量方式为平均召回率, 具体为 1% 的平均召回率 (Average Recall@1%, @1%) 与 1 的平均召回率 (Average Recall@1, @1)。如果检索到的场景是队列场景的真实邻居, 则认为其检索成功。对于 RobotCar Season 数据集, 为测试每个类别中匹配的准确性, 估计位姿和真实位姿之间的距离误差在 5 米范围内且角度误差在 10 度内视为正确匹配。

4.1.3 实验细节

该部分对本研究所提的波粒二象性驱动的点云场景识别特征表达范式下多个实例配置进行描述, 如表 1 所示。所有方法的输入点云均包含 4096 个点。每个点云依次通过点云波性表达、点云粒子性信息交互及点云特征紧凑化编码三个模块。

在点云特征紧凑化编码模块中, 经实验验证, 针对软 NetVLAD 实例, 确定聚类大小 K 为 32, 且引入软聚类中心 s , 最终输出的全局特征维度为 256, 具体参见 4.3.1 小节。针对 GeM 实例, GeM 的系数 i 设置为 3。当作为泛化方法使用时, 可将间接-NetVLAD-256D 与直接-GeM-128D 这两种实例配置迁移至当前点云场景识别方法中以提升其特征表达性能。

4.2 性能比较及分析

本小节将进行深入细致的定量性分析, 首先给出本文所提波粒二象性驱动的点云场景识别特征表达范式下多个实例方法的比对结果, 随后展现所提 2 类实例与主流方法的对比实验与分析。具体包括泛化性分析、复杂度分析、性能提升分析与鲁棒性分析。

4.2.1 多种实例方法分析

针对不同实例的实验验证结果如表 2 所示。相较于其它实验, 其中有 2 种实验配置较为典型, 分别是: 间接-NetVLAD-256D (简称间接方法) 与直接-GeM-128D (简称直接方法)。因此, 后续实验均基于这两种配置来进行效果比对。

表1 不同方法实例的实验配置总结

Table1 Summary of Experimental Configurations for Different Method Instances

配置名称	点云波性表达模块		点云粒子性交互模块		点云特征紧凑化编码模块		输出维度	
	实例1: 间接建模	实例2:直接建模	实例1: 通用	实例2: 轻量化	实例1: 软NetVLAD	实例2:GeM	128	256
间接-NetVLAD-128D	√		√		√		√	
间接-NetVLAD-256D	√		√		√			√
间接-GeM-128D	√		√			√	√	
间接-GeM-256D	√		√			√		√
直接-NetVLAD-128D		√		√	√		√	
直接-NetVLAD-256D		√		√	√			√
直接-GeM-128D		√		√		√	√	
直接-GeM-256D		√		√		√		√

表2 不同实例实验结果

Table2 Experimental Results of Different Instances

Methods	Oxford		U.S.		R.A.		B.D.	
	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1	AR@1%	AR@1
间接+NetVLAD+256D	85.9	70.9	74.8	61.6	71.3	59.0	66.6	59.3
间接+NetVLAD+128D	84.3	69.1	79.1	64.4	71.0	57.7	65.5	58.1
间接+GeM+256D	85.0	71.6	93.8	69.4	77.6	66.8	72.8	65.6
间接+GeM+128D	86.5	73.5	83.4	71.8	77.9	65.4	72.6	65.9
直接+GeM+128D	84.5	72.0	87.6	75.8	77.6	70.2	76.5	70.2
直接+GeM+256D	82.6	69.3	87.0	73.8	78.0	68.6	74.1	66.7
直接+NetVLAD+128D	82.2	67.1	78.4	64.2	69.7	58.7	66.8	59.9
直接+NetVLAD+256D	84.4	70.0	79.7	65.8	73.7	60.5	67.9	60.5

为验证所提方法的泛化性和有效性,本研究选取该领域经典且高引用的方法作为基准进行比较。通过在经典方法上应用本文所提方法框架,验证所提方法对传统框架的普适性,并展示其在原始方法上可提升的性能效果,故本次实验主要涉及五种经典比对方法:PN-VLAD(Uy M A 等, 2018)、PCAN(Zhang W 等, 2019)、LPD(Liu Z 等, 2019)、PPT-Net(Liu C 等, 2024)、MinkLoc3D(Komorowski J 等, 2021)。

在比对本研究所提出的间接方法时,由于除MinkLoc3D方法外,其余方法均采用NetVLAD编码。为实现公平比较,针对MinkLoc3D比对方法,采用与其他比对方法一致的NetVLAD编码方式。

在比对本研究所提出的直接方法时,由于原始

LPD方法的聚合过程较为复杂,所需训练时长较长,且占用内存过大。为实现轻量级比较,因此,在对比该方法时,本节去除了LPD中邻居聚合的部分进行比对,标识为LPD*。所比对的五类方法验证结果如表3所示,在大多数情况下,使用本研究所提的两种实例方法后,性能均得到了提升。与此同时,所需的训练迭代次数更少。这表明本研究所提方法具有好的泛化性,且更易训练。尽管在间接方法的比对中,MinkLoc3D在Oxford基准测试中取得了最优结果,但由于其过度依赖Oxford模式,使得在其他三个基准测试的泛化性能并不佳。对于直接方法,当嵌入到现有方法中时,性能都大幅提升,尤其是在三类室内数据集上。这里可知直接方法更佳,因为间接方法所采用的希尔伯特变换是种简化得到复值表示的

方式,会损失部分精度,而直接方法中的最终复值是 通过波函数推导得出,对复值的表示更为精准。

表3 在基准数据集上所训练的模型泛化性测试性能
Table3 Generalization Test Performance of Models Trained on Benchmark Datasets

方法	Oxford		University Sector (U.S.)		Residential Area (R.A)		Business District (B.D.)	
	@1%	@1	@1%	@1	@1%	@1	@1%	@1
PN-MAX	79.6	61.3	67.8	53.5	60.2	44.9	54.4	46.1
PN-VLAD	80.3	62.6	72.7	63.2	60.8	56.1	65.3	57.2
+间接方法	83.7	68.3	80.6	67.3	72.2	58.1	69.2	62.6
+直接方法	84.5	72.0	87.6	75.8	77.6	70.2	76.5	70.2
PAN	81.9	65.4	78.5	62.6	69.7	57.8	63.9	57.2
PCAN	83.8	69.1	79.1	62.5	71.2	56.9	66.8	58.1
+间接方法	86.0	72.2	74.9	59.7	69.6	55.6	65.3	57.0
+直接方法	84.5	72.0	87.6	75.8	77.6	70.2	76.5	70.2
LPD	94.4	85.4	92.7	79.4	87.4	77.8	83.3	75.8
LPD*	84.0	69.8	77.8	62.9	73.7	60.2	66.7	59.6
+间接方法	94.4	85.9	93.7	83.5	88.4	80.4	86.3	81.4
+直接方法	82.6	68.0	81.5	67.7	76.2	63.2	69.9	62.6
PPT	97.2	92.1	96.5	88.3	89.1	79.8	85.6	79.4
+间接方法	97.9	93.2	97.9	92.3	92.9	85.7	89.0	82.9
+直接方法	98.1	93.5	97.6	91.8	92.9	85.7	90.8	85.6
MinkLoc3D(NetVLAD)	95.2	86.8	79.5	71.3	79.8	66.1	88.7	73.9
+间接方法	92.9	82.4	79.7	72.4	83.9	70.1	89.2	75.3
MinkLoc3D (GeM)	92.5	83.1	86.0	76.5	83.5	76.8	91.0	78.5
+直接方法	97.1	91.3	94.6	83.8	90.6	80.8	85.6	78.4

4.2.2 复杂度分析

表4详列了学习参数、每个点云所需的执行时间和运算开销(Floating Point Operations, FLOPs)的分析。其中,LPD*计入了手工特征的处理耗时。与PN-VLAD、PCAN和LPD三类方法相比,本研究所提间接方法展现出较为明显的优势,但其参数量仅为14.4MB,且运行时间最快,为11ms。尽管MinkLoc3D凭借3D稀疏卷积架构在参数量上具有竞争力,因此MinkLoc3D需学习的参数最少,但其运行时间约为间接方法的两倍。

进一步地,本研究所提直接方法采用更轻量化的GeM编码策略,避免了NetVLAD复杂的聚类学习过程,实验结果显示,直接方法的参数量仅为6.12MB,执行时间缩短至8ms,显著优于现有主流方法。此外,图4展示了使用本研究所提直接方法

后的参数量变化。得益于更完备的信息表示能力和更高效的编码模块,直接方法故在训练过程中得以更快收敛,节约超过一半的收敛时间,并且所需的参数量大幅减少。

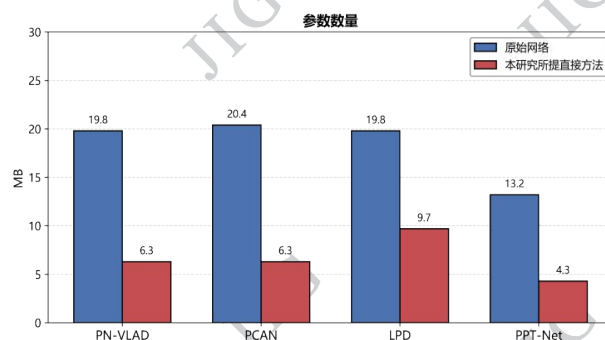


图4 所提直接表达法与现有方法参数量对比
Fig. 4 Parameter Quantity Comparison Between the Proposed Direct Representation Method and Existing Methods

表4 不同模型的复杂度分析

Table 4 Complexity Analysis of Different Models

模型 / 方法	参数量	执行时间	FLOPs
PN-VLAD	19.8MB	15ms	4.2GB
PCAN	20.4MB	55ms	7.7GB
LPD	19.8MB	26ms	7.8GB
LPD*	19.8MB	917ms	7.8GB
PPT-Net	13.2MB	22ms	3.2GB
MinkLoc3D	1.1MB	21ms	3.5GB
间接方法	14.4MB	11ms	10.7GB
直接方法	6.12MB	8ms	10.7GB

4.2.3 性能提升分析

为证明本研究所提方法的优越性,本节在基准线上额外比较三种不同模式,其中A表示仅采用最大池化操作,B表示使用NetVLAD编码,而C则表示使用NetVLAD编码加上下文注意力机制。

表5展示了训练基线和增强数据集后的性能效果,训练基线用于测试泛化性能的增益,而训练增强数据集则用于评估增加输入数据后性能的增益。通过实验数据分析可知,本研究所提间接方法在未使

用任何增强模块,如上下文注意力机制,其在两类数据集上依旧表现良好,且该方法使用较小的聚类数32。整体而言,相较于基于实数系列的方法(A、B和C),该方法在这两个训练数据集上都实现了性能提升,进一步验证其有效性。

对本研究所提直接方法而言,其在基准数据集上泛化性能增益最高,提升增益超过间接方法,特别是在U. S.、R. A.与B. D.数据集上;其在增强数据集上,在除B. D.数据集的其他三个数据集上比间接方法表现更好,而在B. D.数据集略微逊色,仅超过了A方法。究其原因,不同于基准数据集,训练增强数据集的数据中除了Oxford数据集之外,还加入了其他三个数据集。由于本研究所提间接方法采用软NetVLAD策略,能够更精细地捕捉局部信息,从而使性能得以提高。然而直接方法采用GeM编码配合输出更低维度的特征表达,虽然实现了特征的精简,但在面对训练数据增多的情况下,这种编码策略则显得过于简单而导致部分局部细节被忽略。因此,在处理数据量较小的B. D.数据集时,明显比不上本研究所提间接方法的编码能力,进而影响性能表现。

表5 在基准数据集与增强数据集上的测试结果性能表

Table 5 Performance Comparison of Test Results on Benchmark Datasets and Augmented Datasets

数据集	方法	Oxford		University Sector (U.S.)		Residential Area (R.A)		Business District (B.D.)	
		@1%	@1	@1%	@1	@1%	@1	@1%	@1
		基准数据集	A	79.6	61.3	67.8	53.5	60.2	44.9
B	80.3		62.6	72.7	63.2	60.8	56.1	65.3	57.2
C	83.8		69.1	79.1	62.5	71.2	56.9	66.8	58.1
间接	83.7		69.3	80.6	67.3	72.2	58.1	69.2	62.6
直接	84.5		72.0	87.6	75.8	77.6	70.2	76.5	70.2
增强数据集	A	73.9	54.2	79.3	62.2	75.1	60.2	69.5	59.0
	B	80.1	63.3	90.1	86.1	93.1	82.7	86.5	80.1
	C	86.4	70.7	94.1	83.7	92.3	82.3	87.0	80.3
	间接	84.7	69.3	95.8	87.2	92.5	85.6	87.6	82.6
	直接	87.8	75.0	96.5	87.3	94.0	85.9	85.3	79.5

4.2.4 旋转鲁棒性

为了检验噪声鲁棒性,表6选用随机抖动点进行噪声模拟实验。两种噪声参数设置为 $\sigma=0.005$

和 $\sigma=0.05$ 。通过表中数据可知,本研究所提间接方法在大部分情况下表现最佳。虽然在第一个噪声条件下,C方法在Oxford数据集上略好,但其在剩余数

数据集上的性能表现一般,这凸显出C方法在基准数据集上对噪声的鲁棒性并不均衡。而本文所提直接方法在四个数据集上均展现出最佳性能,尤其是在U. S.、R. A. 及 B. D. 这三个数据集上,这体现了通

过波函数获取直接“波性”相比采用简化版希尔伯特函数获取间接“波性”的方法而言,在建模点云特征时更为精准,从而促进性能提升。

表6 在基准数据集上训练的模型,针对输入噪声测试的结果性能表

Table 6 Performance Results of Models Trained on Benchmark Datasets for Input Noise Tests

噪声	方法	Oxford		University Sector (U.S.)		Residential Area (R.A)		Business District (B.D.)	
		@1%	@1	@1%	@1	@1%	@1	@1%	@1
0.005	A	79.6	61.2	68.5	53.2	61.1	46.0	54.6	46.7
	B	81.1	62.7	77.5	62.9	70.0	56.2	65.3	67.4
	C	83.7	69.1	78.6	61.9	70.8	57.4	66.4	57.8
	间接	83.4	68.2	80.2	67.0	71.6	58.7	68.2	60.9
	直接	84.4	71.9	87.6	75.7	77.5	70.1	76.4	70.1
0.05	A	74.7	56.4	67.0	50.0	63.9	49.0	55.2	47.5
	B	80.9	62.4	77.4	63.3	69.2	56.1	65.4	67.3
	C	80.2	63.3	76.0	60.6	69.6	56.5	66.9	58.9
	间接	80.6	65.2	77.0	65.1	70.8	58.9	67.1	59.8
	直接	85.5	70.3	84.6	72.6	76.9	68.0	75.2	68.3

为了检验旋转鲁棒性,表7测试了不同旋转度下各类方法的性能结果。本文所提的间接方法和直接方法带来了最高的性能增益,分别平均提升6.2%和7.1%,远高于A、B、C三种方法。这一显著表现

进一步彰显基于波粒二象性的点云场景特征表达的独特优势,即通过将点云建模为粒子,不仅能获得更加完备的特征表达,还能增强模型的稳定性。

表7 在KITTI与RobotCar Season数据集上的鲁棒性实验

Table 7 Robustness Experiments on the KITTI and RobotCar Season Datasets

方法	30度	60度	90度	120度	150度	180度	平均
A	29.3	18.4	14.1	17.7	15.6	14.2	18.2
B	34.2	24.4	14.7	16.9	14.6	13.5	19.7(+1.5)
C	39.3	23.1	16.1	19.6	16.8	15.8	21.8(+3.6)
间接	42.6	26.9	18.4	22.4	18.2	17.7	24.4(+6.2)
直接	45.0	29.4	18.3	22.5	19.1	17.2	25.3(+7.1)

为了进一步检验跨域鲁棒性,表8采用基线数据集上预训练好的模型,和PN-VLAD方法进行对比,测试在KITTI和RobotCar Season数据集上的跨域性能。可以看出,相较于PD-VLAD,本研究所提间接方法在KITTI数据集上性能提升了2.4%,

本研究所提两种方法在RobotCar Season数据集上分别平均提升4.0%和5.8%。整体而言,本研究所提两种方法较PN-VLAD有更好的跨域鲁棒性。

而直接方法在KITTI数据集上表现不佳,其中一个原因是其场景差异性较大,且测试集规模较小,容易导致结果产生一定偏差。

4.3 消融实验

本节核心讨论所提波粒二象性点云场景特征范式在基准数据集的前1%的平均召回率性能下的消融实验,并深入分析各模块设置的合理性与必要性。

表8 KITTI与RobotCar Season数据集上的跨域鲁棒性实验

Table 8 Cross-Domain Robustness Experiments on the KITTI and RobotCar Season Datasets

方法	KITTI@1%	RobotCar Season					
		Summer	Snow	Dusk	Sun	Night-rain	Average
PN-VLAD	52.4	85	84.9	78.1	78.0	8.5	66.9
间接方法	54.8	90.5	85.8	85.6	78.5	14.0	70.9(+4.0)
直接方法	50.0	90.0	87.4	86.6	86.1	13.4	72.7(+5.8)

4.3.1 各模块有效性分析

本部分旨在测试每个模块对性能的影响,如表9所示,涉及三种情况:1)无模块1,即不使用“波性”表达,而使用PointNet提取单特征表示;2)无模块2,即不使用“粒子性”交互,而使用串联层和全连接层

来输出相同维度的特征;3)无模块3,即只采用最大池化与全连接层输出相同的特征维度。从结果可分析得出,缺失任一模块都会导致性能下降,这也揭示了本研究所提范式中的每一模块都具有其作用。

表9 不同模块对最终性能的影响

Table 9 Impact of Different Modules on the Final Performance

实例	模块	最大差值(Oxford)	Oxford	U.S.	R.A.	B.D
直接	无模块1	-3.1	80.5	75.5	65.3	64.1
	无模块2	-16.8	66.8	77.3	71.4	65.1
	无模块3	-2.5	81.1	69.8	60.1	55.1
	全包含	0	83.6	78.4	72.2	69.2
间接	无模块1	-3.92	81.8	82.5	75.3	71.0
	无模块2	-55.48	33.1	24.4	23.2	23.7
	无模块3	-7.99	79.7	80.9	68.8	64.9
	全包含	0	84.5	87.6	77.6	76.5

输出特征维度如表10所示。对于间接方法而言,维度为256时效果最佳,维度的继续递增并无法带来全部数据集的性能增益,故选取 $D=256$ 。对于直接方法而言,当维度为128时增益最佳,维度的递增会导致部分性能下降,故选取 $D=128$ 。

间接方法软NetVLAD聚类个数分析如表11所示,输出特征维度设置为256,软NetVLAD适配的聚类数量为 K 。随着聚类个数的逐步增加,性能在大部分数据集上都得到提升,当聚类个数高于32后,性能则呈现下降趋势。因此本研究在最终实验将聚类数设置为32,相比于其他方法将聚类数设置为64,本方法能够有效节省一半的聚类学习时间。

直接方法的编码策略有效性如表12所示,本研究比对了最大池化、NetVLAD和GeM对应的表现性能,它们的复杂度分别为 $O(n)$ 、 $O(\ln)$ 和 $O(n^2)$,由于

表10 最终特征输出维度大小对平均召回率前1%性能的影响

Table 10 Impact of the Final Feature Output Dimension Size on the Top-1% Average Recall Performance

实例	特征维度	Oxford	U.S.	R.A.	B.D
间接	D64	81.0	77.4	65.6	60.9
	D128	82.4	78.2	70.4	65.1
	D256	83.6	78.4	72.2	69.2
	D512	82.8	80.1	74.7	69.0
直接	D64	81.8	82.1	76.2	71.7
	D128	84.5	87.6	77.6	76.5
	D256	82.6	87.0	78.0	74.1
	D512	84.2	86.8	80.1	75.3

在实验中 GeM 的系数为 3, 可知 $O(n)=O(\ln)<O(n^2)$ 。其中最大池化是较为简单且粗略的方法, 用于调整输出组合大小, 而不具备保留信息的能力。而 NetVLAD 需学习聚类超参, 成本较高。GeM 则是一个轻量级无参数学习的编码方法, 简单高效, 表现出众。可以发现, 最大池化在不同数据集上波动很大, 因为最大池化虽然简单, 但其效果不佳。而 NetVLAD 是三种策略中最复杂的, 但性能有时不如最大池化, 而 GeM 性能表现出众。因此, 在考虑轻量化兼顾效果时, 优先选择 GeM。

表 11 聚类大小对平均召回率前 1% 性能的影响

Table 11 Impact of Cluster Size on the Top-1% Average Recall Performance

聚类个数 K	Oxford	U.S.	R.A.	B.D
16	82.7	77.2	72.7	67.9
32	83.6	78.4	72.2	69.2
64	82.5	78.5	71.0	65.9
128	83.7	76.0	72.0	67.6

表 12 不同编码策略下的表现性能

Table 12 Performance Under Different Encoding Strategies

策略	Oxford	U.S.	R.A.	B.D
Max-Pooling	79.7	80.9	68.8	64.9
NetVLAD	82.2	78.4	69.7	66.8
GeM	84.5	87.6	77.6	76.5

4.3.2 训练策略分析

如表 13 所示, 实验结果表明, 针对本研究所提两种方法, 在大规模数据集上, 最大四元组损失策略表现出优越的性能。针对直接方法, 对于中小型数据集(如 U. S. 与 R. A. 数据集), 三元组损失则展现出更佳的效果, 此时四元组损失可能因其复杂性而略显笨重。鉴于本研究所提方法在大规模 Oxford 数据集上训练, 并在所有数据集上进行测试, 因此训练策略最终选取最大四元组损失。

5 结论

本研究针对现有三维点云场景识别方法在实数域表达受限、高计算开销及模型复杂度高局限性的

表 13 选用不同训练策略对平均召回率前 1% 性能的影响

Table 13 Impact of Different Training Strategies on the Top-1% Average Recall Performance

实例	训练策略	Oxford	U.S.	R.A.	B.D
间接	三元组损失	79.8	72.9	66.3	59.8
	四元组损失	88.0	83.3	81.7	81.6
	最大三元组损失	89.4	87.7	84.7	84.1
	最大四元组损失	90.7	86.2	83.0	84.1
直接	三元组损失	82.4	83.6	76.2	73.6
	四元组损失	82.4	81.8	74.9	72.2
	最大三元组损失	82.4	84.7	77.4	73.1
	最大四元组损失	84.45	87.6	77.6	76.5

受物理学波粒二象性理论启发, 提出了一种新的点云场景识别特征表达范式。该范式由点云波性表达、点云粒子性信息交互及点云特征紧凑化编码三个核心模块组成, 并针对每个模块设计了两种具体实现方案, 通过将点云信息从实数域映射至高维希尔伯特空间, 本方法利用波函数的幅值与相位实现了对场景结构与语义信息的完备建模, 并模拟粒子相干作用刻画了特征间的动态交互。

实验结果表明, 该范式在保持高度轻量化的同时, 显著提升了多种主流基准模型的识别性能。特别是, 波函数直接建模表达方法结合广义平均池化的轻量化方案效果显著, 且计算高效, 仅用 6.12MB 参数量即可将模型收敛速度提升近 1 倍, 达到 8ms, 满足自动驾驶、机器人导航的实时性需求。

本研究验证了跨学科理论融合在点云场景特征表达领域的有效性, 所提范式具备良好的通用性和即插即用特性, 为点云分析提供了新的思路和高效的解决方案。未来工作将进一步探索该范式在点云分割、目标检测等更广泛三维视觉任务中的应用, 并结合边缘计算开展轻量化研究, 以提升其在真实场景中的高效部署能力。

参考文献 (References)

- Arandjelović R, Gronat P, Torii A, Pajdla T and Sivic J. 2016. NetVLAD: CNN architecture for weakly supervised place recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 4393-4402 [DOI: 10.1109/CVPR.2016.513]

- Cheng H Z, Feng Y, Zheng J F and Wan B. 2025. Multimodal self-supervised point cloud representation learning based on bi-directional fitting mask reconstruction. *Journal of Software*, 36(5): 2092-2115 (程浩喆, 冯勇, 郑江峰, 万博. 2025. 基于双向拟合掩码重建的多模态自监督点云表示学习. *软件学报*, 36(5): 2092-2115) [DOI: 10.13328/j.cnki.jos.007134]
- Geiger A, Lenz P and Urtasun R. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE: 3354-3361 [DOI: 10.1109/CVPR.2012.6248074]
- Hui L, Yang H, Cheng M, Liu Y and Zhang S. 2021. Pyramid point cloud transformer for large-scale place recognition [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 6098-6107 [DOI: 10.1109/ICCV48922.2021.00608]
- Hou Z, Yan Y, Xu C, Li J and Wang H. 2022. Hitpr: Hierarchical transformer for place recognition in point cloud [C]//Proceedings of the 2022 International Conference on Robotics and Automation. Philadelphia, USA: IEEE: 2612-2618 [DOI: 10.1109/ICRA46639.2022.9811770]
- Hao W, Zhang W, Jin H, Liu S and Chen L. 2024. SAGE-Net: Employing spatial attention and geometric encoding for point cloud based place recognition [J]. *IEEE Robotics and Automation Letters*, 9(6): 4958-4965 [DOI: 10.1109/LRA.2024.3389456]
- Huang Y, Zhou S, Zhang J, Li M and Chen W. 2024. Voxel or Pillar: Exploring efficient point cloud representation for 3D object detection [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 2426-2435 [DOI: 10.1609/aaai.v38i5.28845]
- Jégou H, Douze M, Schmid C and Pérez P. 2010. Aggregating local descriptors into a compact image representation [C]//Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE: 3304-3311 [DOI: 10.1109/CVPR.2010.5539970]
- Komorowski J, Tyszkiewicz J, Mielczarek B and Cielniak G. 2021. MinKLoc3D: Point cloud based large-scale place recognition [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 1789-1798 [DOI: 10.1109/WACV48630.2021.00184]
- Liu C, Chen G, Song R, Wang Y and Zhang X. 2024. LPS-Net: Lightweight parameter-shared network for point cloud-based place recognition [C]//Proceedings of the 2024 IEEE International Conference on Robotics and Automation. Yokohama, Japan: IEEE: 448-454 [DOI: 10.1109/ICRA57147.2024.10610595]
- Li X, Wang Q, Qiu B, Chen Y and Zhao J. 2025. Dual-path geometric relation-aware transformer for point cloud classification and segmentation [J]. *Applied Soft Computing*, 174: 112801 [DOI: 10.1016/j.asoc.2024.112801]
- Liu Z, Zhou S, Suo C, Li J and Wang Y. 2019. LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE: 2831-2840 [DOI: 10.1109/ICCV.2019.00292]
- Maddern W, Pascoe G, Linegar C and Newman P. 2016. 1 year, 1000 km: The Oxford RobotCar dataset [J]. *The International Journal of Robotics Research*, 36(1): 3-15 [DOI: 10.1177/0278364916679441]
- Pan J, Cao J, Xing S, Li W and Zhao Y. 2025. An aerial point cloud classification using point transformer via multi-feature fusion [J]. *Scientific Reports*, 15: 22924 [DOI: 10.1038/s41598-025-92924-x]
- Qi C R, Su H, Mo K and Guibas L J. 2017. PointNet: Deep learning on point sets for 3D classification and segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 625-660 [DOI: 10.1109/CVPR.2017.169]
- Qi C R, Yi L, Su H and Guibas L J. 2017. PointNet++: Deep hierarchical feature learning on point sets in a metric space [C]//Proceedings of the Advances in Neural Information Processing Systems 30. Long Beach, USA: Curran Associates: 5105-5114 [DOI: 10.5555/3295222.3295263]
- Qiu Q, Gao H, Wang W, Chen L and Zhang X. 2024. SelfLoc: Selective feature fusion for large-scale point cloud-based place recognition [J]. *Knowledge-Based Systems*, 295: 111794 [DOI: 10.1016/j.knosys.2024.111794]
- Sun Q, Liu H, He J, Zhang C and Wang L. 2020. DAGC: Employing dual attention and graph convolution for point cloud based place recognition [C]//Proceedings of the 2020 International Conference on Multimedia Retrieval. Dublin, Ireland: ACM: 224-232 [DOI: 10.1145/3372278.3390680]
- Shu D W and Kwon J. 2024. Hierarchical bidirected graph convolutions for large-scale 3-D point cloud based place recognition [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 35(7): 9651-9662 [DOI: 10.1109/TNNLS.2023.3289095]
- Uy M A and Lee G H. 2018. PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4393-4402 [DOI: 10.1109/CVPR.2018.00450]
- Wang C S, Wang H, Ning X, Tian S W and Li W J. 2023. 3D point cloud classification method based on dynamic coverage of local regions. *Journal of Software*, 34(4): 1962-1976 (王昌硕, 王含, 宁欣, 田生伟, 李卫军. 2023. 基于局部区域动态覆盖的3D点云分类方法. *软件学报*, 34(4): 1962-1976) [DOI: 10.13328/j.cnki.jos.006764]
- Wang L, Huang M, Yang Z, Li J and Zhang H. 2025. LBNP: Learning

features between neighboring points for point cloud classification [J]. PLOS ONE, 20(1): e0314086 [DOI: 10.1371/journal.pone.0314086]

Wang Y, Sun Y, Liu Z, Sarma S E, Bronstein M M and Solomon J M. 2019. Dynamic graph CNN for learning on point clouds [J]. ACM Transactions on Graphics, 38(5): 1-12 [DOI: 10.1145/3326362]

Zhang G, Fan L, He C, Liu Y and Zhao J. 2024. Voxel Mamba: Group-free state space models for point cloud based 3D object detection [C]//Proceedings of the 38th International Conference on Neural Information Processing Systems. Red Hook, USA: Curran Associates: 81489-81509 [DOI: 10.48550/arXiv.2405.14852]

Zhang W and Xiao C. 2019. PCAN: 3D Attention Map Learning Using Contextual Information for Point Cloud Based Retrieval [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4230-4239 [DOI: 10.1109/CVPR.2019.00433]

Zhang J, Zhang Y, Liao M, Wang Z and Li C. 2024. CapsLoc3D: Point cloud retrieval for large-scale place recognition based on 3D capsule networks [J]. IEEE Transactions on Intelligent Transportation Systems, 25(7): 6811-6823 [DOI: 10.1109/ITITS.2024.3367891]

Zhou H, Qi H G, Deng Y Q, Li J J, Liang H and Miao J. 2024. 3D object detection and classification combined with point cloud depth

information. Journal of Image and Graphics, 29(8): 2399-2412 (周昊, 齐洪钢, 邓永强, 李娟娟, 梁浩, 苗军. 2024. 融合点云深度信息的3D目标检测与分类. 中国图象图形学报, 29(8): 2399-2412) [DOI: 10.11834/jig.230568]

Zhou X Z, Jiang L, Wei M and Zhong J R. 2024. Lightweight point cloud classification based on deformable 3D graph convolution. Computer Engineering, 50(9): 255-265 (周新志, 蒋林, 魏敏, 钟佳容. 2024. 基于可变形三维图卷积的轻量级点云分类研究. 计算机工程, 50(9): 255-265) [DOI: 10.19678/j.issn.1000-3428.0068471]

作者简介

兰天升:男,工学学士在读,主要研究领域为点云智能处理与分析。E-mail: qumei2023@163.com.

张玲:女,工学学士在读,主要研究领域为计算机视觉。E-mail: zling202511@163.com.

刘立波:女,博士,教授,博士生导师,主要研究领域为智能信息处理、计算机视觉。E-mail: liulib@163.com.

张若楠(通讯作者):女,博士,准聘副教授,硕士生导师,主要研究领域为量子启发式多源异构信息感知、面向人机感知共友好的多媒体处理与分析。E-mail: zhangrn@nxu.edu.cn.