

基于 NeRF 的 SLAM 研究综述^①

喻伟东¹, 鲁 静¹, 程晗蕾^{1,2}

¹(远光软件股份有限公司 区块链及元宇宙实验室, 珠海 519085)

²(澳门科技大学 创新工程学院, 澳门 999078)

通信作者: 鲁 静, E-mail: lujing8102@126.com



摘 要: 随着神经辐射场 (NeRF) 的提出, 其基于神经隐式表示场景的方法在生成高保真地图方面具有显著优势, 将 NeRF 应用于同时定位与地图构建 (SLAM) 中, 即基于 NeRF 的 SLAM 方法, 能够在实现高精度的定位的同时进行连续的 3D 建模, 通过渲染新视角并预测未知区域, 提高场景重建的质量和细节. 为了跟踪该领域的最新研究成果, 对近年来基于 NeRF 的 SLAM 的关键算法进行了回顾和综述. 首先介绍了 NeRF 技术的核心原理并全面概述了基于 NeRF 的 SLAM 方法的框架, 其次重点探讨了基于 NeRF 的 SLAM 的改进和优化, 包括提高神经隐式表征效率、解决大尺度场景建图问题、增加回环和全局优化实现全局一致性和解决动态干扰问题, 最后对基于 NeRF 的 SLAM 方法进行了展望, 为相关研究人员提供有价值的参考, 以促进更多创新研究.

关键词: 同时定位与地图构建; 神经辐射场; 神经隐式表示; 三维重建; 移动机器人

引用格式: 喻伟东, 鲁静, 程晗蕾. 基于 NeRF 的 SLAM 研究综述. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9865.html>

Review of NeRF-based SLAM Research

YU Wei-Dong¹, LU Jing¹, CHENG Han-Lei^{1,2}

¹(Blockchain and Metaverse Laboratory, YGSOFT Inc., Zhuhai 519085, China)

²(Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China)

Abstract: The neural radiation field (NeRF) has significant advantages in generating high-fidelity maps thanks to its neural implicit representation-based scene. The application of NeRF in simultaneous localization and mapping (SLAM), namely the NeRF-based SLAM method, enables continuous 3D modeling while achieving high-precision localization to enhance the quality and detail of the scene reconstruction by rendering new perspectives and predicting unknown regions. To track the latest research results in this field, this study reviews and summarizes the key algorithms of NeRF-based SLAM in recent years. Firstly, the core principle of NeRF technology is introduced and a comprehensive overview of the framework of NeRF-based SLAM methods is given, followed by focusing on the improvements and optimizations of NeRF-based SLAM, including improving the efficiency of neural implicit representation, solving the large-scale scene building problem, adding loopback and global optimization to achieve global consistency and solving the dynamic interference problem. Finally, an outlook on the NeRF-based SLAM method is presented to provide valuable references for related researchers to promote more innovative research.

Key words: simultaneous localization and mapping (SLAM); neural radiance field (NeRF); implicit neural representation; 3D reconstruction; mobile robot

① 收稿时间: 2024-11-15; 修改时间: 2024-12-09; 采用时间: 2024-12-24; csa 在线出版时间: 2025-03-04

同时定位与地图构建 (simultaneous localization and mapping, SLAM) 技术是移动机器人在未知环境移动过程中构建增量式地图的同时, 根据构建的地图计算自身位置, 实现自主定位和导航的过程^[1,2], SLAM 技术目前在机器人、自动驾驶和增强现实 (augmented reality, AR) 等多个领域有着广泛的应用. 传统基于几何模型的 SLAM 在缺乏特征的环境^[3]和动态环境^[4,5]中的表现常常受限, 这降低了 SLAM 系统的准确性和鲁棒性. 近年来, 将深度学习与传统几何方法融合用于提升 SLAM 在复杂环境下的精度和环境适应性成为 SLAM 技术发展的趋势, 深度学习在 SLAM 系统视觉里程计^[6,7]、闭环检测^[8]和全局优化^[9,10]等多个方面已经实现对 SLAM 系统模块的单点替代.

神经辐射场 (neural radiance field, NeRF)^[11]作为一种新颖的 3D 隐式神经表示方法, 通过神经网络直接学习场景的连续体积密度和颜色, 为 3D 场景理解和新视角合成提供了一种全新的途径, 将 NeRF 与 SLAM 技术结合可同时进行相机位姿估计和环境建模, 端到端替代传统 SLAM. 基于 NeRF 的 SLAM 技术结合了深度学习强大的场景表征能力和 SLAM 的实时定位与建图需求, 能够进行连续的 3D 建模, 可以渲染新视角并预测位置区域, 提高场景重建的质量和细节, 此外, 通过反向传播算法计算损失函数来优化相机位姿和 NeRF 地图, 对稠密地图进行非常细致的优化, 为解决传统 SLAM 面临的挑战提供了新的思路.

针对基于 NeRF 的 SLAM 技术, 有相关综述已经对其进行了归纳总结. 例如, Yang 等^[12]基于对 30 篇相关文章的研究, 从跟踪和建图两个方面展开详细阐述了 NeRF 在 SLAM 系统中的应用; Tosi 等^[13]全面梳理了现有基于 NeRF 的 SLAM 算法, 将基于 NeRF 的 SLAM 方法分为不同的类别, 并提供了一个详细的分类体系. 然而前述相关综述没有详细描述针对早期基于 NeRF 的 SLAM 方法改进和优化的演化过程以及相关方法的优缺点对比, 基于此, 本研究从基于 NeRF 的 SLAM 方法的改进和优化角度出发, 详细阐述了基于 NeRF 的 SLAM 技术改进和优化方法的优缺点对比以及相关方法的适用场景, 为相关研究人员选择或改进算法提供参考.

为了进行基于 NeRF 的 SLAM 研究, 本研究采用了系统的文献收集方法, 通过使用谷歌学术 (Google Scholar) 进行相关文献的搜索和收集, 谷歌学术提供期

刊论文、会议论文、预印本等更为丰富和全面的文献, 便于了解该领域的最新研究进展. 通过使用关键词“NeRF+SLAM”和“Neural implicit slam”进行搜索, 然后对文献标题、摘要、关键词等进行相关性筛选, 最终得到 49 篇基于 NeRF 的 SLAM 方法相关论文, 收集的最终文献集截至 2024 年 11 月, 搜索和收集相关文献的方法如图 1 所示.

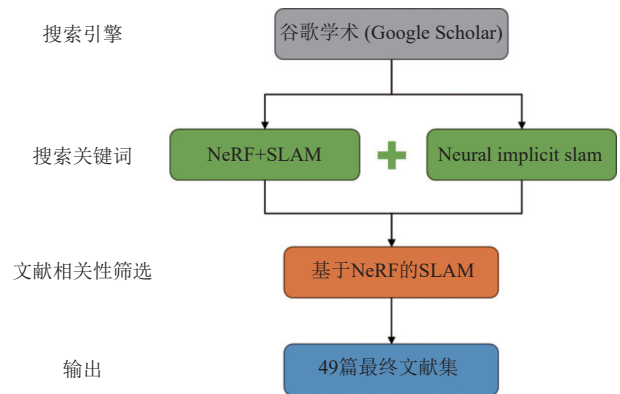


图 1 搜索和收集相关文献的方法

本研究将探讨基于 NeRF 的 SLAM 技术的最新研究进展, 首先全面阐述了基于 NeRF 的 SLAM 方法的框架, 其次重点阐述了改进现有方法的现有策略和研究现状, 最后对基于 NeRF 的 SLAM 技术的发展趋势进行了展望. 本研究通过系统地研究基于 NeRF 的 SLAM 方法, 旨在为相关研究人员提供有价值的参考, 以促进更多创新研究.

1 基于 NeRF 的 SLAM 方法

本节首先从 NeRF 技术的核心原理出发, 然后基于 NeRF 技术与 SLAM 技术的结合, 全面梳理出基于 NeRF 的 SLAM 方法的框架.

1.1 NeRF 技术的核心原理

NeRF^[11]的核心原理在于使用一个全连接神经网络, 通常是一个多层感知机 (multilayer perceptron, MLP)^[14], 来隐式地表示三维场景的连续体积场, 同时, 通过优化一个连续的体积场函数进行复杂场景图像的新视角合成^[15]. 如图 2 所示, NeRF 技术主要通过沿相机射线采样 5D 坐标、采样点颜色和密度计算、体渲染合成 2D 图像以及场景表示优化等 4 个步骤来实现.

(1) 沿相机射线采样 5D 坐标

NeRF 模型会沿着相机视点到场景中的某一点的

射线进行采样, 每条射线由一系列的采样点组成, 这些采样点代表了相机到场景中特定点的视线路径, 采样点分布在射线上, 位置 $r(t)$ 的采样点可基于式(1)表示.

$$r(t) = o + t \times d \quad (1)$$

式中, $r(t)$ 为采样点的位置, o 为相机的位置; t 为射线方向上的距离; d 为视角方向, 可用方位角 θ 和俯仰角 φ 来表示, 即 $d = (\theta, \varphi)$. 采样点 $r(t)$ 的空间位置可以用三维坐标表示为 $x = (X, Y, Z)$, 射线上的采样点 $r(t)$ 可以基于 $(X, Y, Z, \theta, \varphi)$ 这5个变量组成的5D坐标来表示.

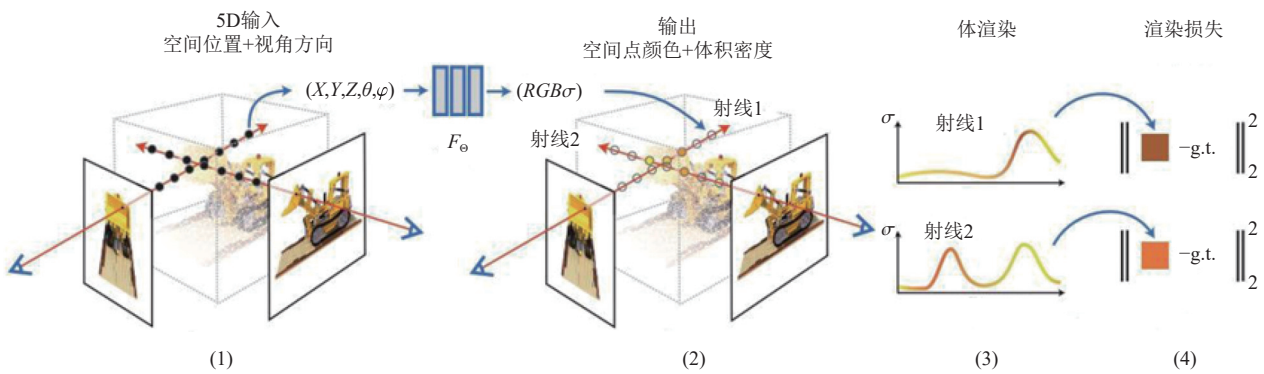


图2 NeRF 技术核心原理^[11]

(2) 采样点颜色和体积密度预测

NeRF 通过使用一个 MLP 网络来预测采样点的颜色和体密度, MLP 网络的输入为采样点的 5D 坐标, 输出为空间点的颜色和体密度, 计算公式如下^[11]:

$$F_{\Theta} : (x, d) \rightarrow (c, \sigma) \quad (2)$$

其中, F_{Θ} 表示 MLP 网络, Θ 表示 MLP 网络的权重, 输入为空间点的位置 $x = (X, Y, Z)$ 和视角方向 $d = (\theta, \varphi)$, 输出为空间点的颜色 $c = (r, g, b)$ 和体密度 σ . 对于每个采样点, 经过 MLP 网络训练后可预测该采样点 $r(t)$ 的颜色值 $c(r(t))$ 和密度值 $\sigma(r(t))$.

(3) 体渲染合成 2D 图像

NeRF 使用渲染方程进行体渲染 (volume rendering)^[16]从而生成 2D 图像, 将沿射线的的所有采样点的颜色和密度信息累计起来, 以计算出新视角下最终 2D 图像上每个像素的颜色值, 渲染方程公式如下所示:

$$C(r(t)) = \int_0^T c(r(t)) \cdot \sigma(r(t)) \cdot T(t) dt \quad (3)$$

其中, $C(r(t))$ 为最终合成的 2D 图像颜色, $c(r(t))$ 和 $\sigma(r(t))$ 分别表示 MLP 网络预测的采样点的颜色值和密度值, T 为射线长度, $T(t)$ 为累计透射率, 即从相机到该采样点的光线被吸收的比例, 其计算公式如下所示:

$$T(t) = \exp\left(-\int_0^t \sigma(r(u)) du\right) \quad (4)$$

(4) 场景表示优化

NeRF 模型的体渲染过程是可微渲染^[17,18], 可通过反向传播算法计算损失函数来优化 MLP 网络的权重, 使用梯度下降的方法以最小化渲染图像和真实图像之间的差异来优化场景表示, 优化过程可以使得 NeRF 以任意新视角渲染出高质量的图像, 损失函数通常使用均方误差 (MSE) 表示, 如下式所示:

$$\mathcal{L} = \min_{\sigma, c} \frac{1}{n} \sum_{i=1}^n \|\hat{I}_i(\sigma, c) - I_i\|_2^2 \quad (5)$$

其中, \mathcal{L} 为损失函数, $\hat{I}_i(\sigma, c)$ 和 I_i 分别代表 NeRF 渲染图像和实际观测图像.

1.2 基于 NeRF 的 SLAM 方法概述

SLAM 技术是机器人利用传感器 (激光雷达、视觉相机等) 在没有先验地图信息的情况下实现机器人的定位和周边环境建图, 早期的 SLAM 方法主要是激光 SLAM^[19-21], 激光 SLAM 具有定位精度和可靠性高等优势, 但价格昂贵且不适合大范围的室外场景; 而视觉 SLAM^[22-24]以视觉相机作为传感器, 具有成本低和在室内外环境中均可开展工作等优势, 随着计算机视觉的发展, 视觉 SLAM 因其能够感知环境丰富的颜色和纹理信息而成为研究的热点. 传统的视觉 SLAM 技术主要关注于定位问题, 而生成的地图, 无论是稀疏的还是稠密的, 都仅作为辅助定位的工具, 这种方法未能

充分地记录和精确地表达环境的细节. 随着 NeRF 的提出, NeRF 在生成高保真地图方面具有显著优势^[25], 不仅能够提供精细的场景细节, 还能模拟复杂的光照和环境变化^[26], 能充分捕捉并精确表达周围环境的全貌, 同时, SLAM 能为 NeRF 训练提供初始位姿并通过反向传播算法来优化位姿, 从而实现对环境的高保真度重建和相机位姿的精确估计, 相关研究尝试将 NeRF 与 SLAM 相结合, 即基于 NeRF 的 SLAM 方法.

基于 NeRF 的 SLAM 方法的核心在于使用 MLP 网络来隐式表示复杂场景, 并基于体渲染可微分特性通过反向传播算法最小化损失函数来优化位姿和 MLP 网络参数, 从而实现高精度的定位和高保真的隐式场

景表示. 基于 NeRF 的 SLAM 系统一般包括两个并行运行的线程: 跟踪线程和建图线程^[12], 跟踪线程采用视觉里程计 (visual odometry, VO) 方式或者逆渲染^[18]方式, 视觉里程计方式直接采用传统 SLAM 的前端跟踪模块, 如 Orbeez-SLAM^[27]、Go-SLAM^[28]等, 通过前端跟踪和后端 NeRF 建图解耦达到良好的跟踪性能; 逆渲染方式则通过逆渲染反向传播算法最小化损失函数来更新相机位姿, 如 iMAP^[29]、NICE-SLAM^[30]等, 由于基于视觉里程计方式 (VO) 的前端跟踪只是简单的采用传统 SLAM 的前端跟踪模块, 下面以逆渲染方式的前端跟踪为例大致介绍基于 NeRF 的 SLAM 系统的工作流程, 如图 3 所示.

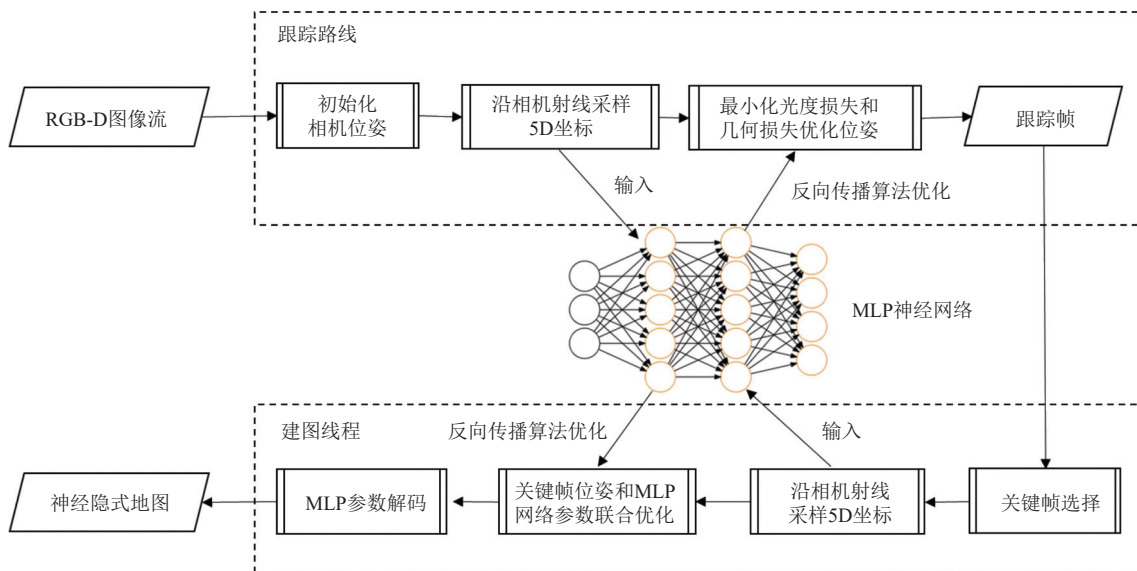


图 3 基于 NeRF 的 SLAM 系统大致的工作流程

跟踪线程: 在这一线程中, 系统接收到输入的 RGB-D 图像流后, 初始化相机位姿, 然后通过沿相机射线采样 5D 坐标, 基于 NeRF 模型的反向传播算法最小化光度损失 (即渲染图像颜色与实际图像颜色之间的差异) 和几何损失 (即渲染深度图与实际深度测量之间的差异) 来优化每帧相机位姿 $\{R, t\}$, 即旋转矩阵 R 和平移向量 t , 这些参数定义了相机在世界坐标系中的位置和方向.

建图线程: 在这一线程中, 系统会从跟踪线程接收关键帧, 首先需要决定哪些关键帧被选为关键帧并用于优化, 关键帧的选取通常基于某些标准, 例如逆渲染深度与 RGBD 获取的深度值进行比较, 超过一定阈值则认为关键帧; 其次, 处理选取的关键帧, 通过沿相

机射线采样 5D 坐标, 并通过反向传播算法最小化光度损失和几何损失来联合优化 NeRF 模型参数 (即 MLP 网络参数 θ) 和关键帧位姿 $\{R, t\}$, 通过相机位姿与 NeRF 模型参数的联合优化; 最后通过 MLP 解码得到渲染所需的颜色和体密度参数, 并通过体渲染得到隐式场景地图.

2 基于 NeRF 的 SLAM 的改进和优化

基于 NeRF 的 SLAM 方法结合了 NeRF 的连续体积表示和 SLAM 实时定位与建图的功能, 可以对相机位姿和地图进行非常细致的优化, 实现精准定位的同时建立稠密细腻的三维场景. 2021 年提出的 iMAP^[29] 是神经隐式表示在 SLAM 中的首次应用, 是第 1 个真

正意义上的基于 NeRF 的 SLAM 系统, 开创性提出了在 SLAM 中使用 MLP 来隐式表征场景地图, 但由于 iMAP 基于经典 NeRF, SLAM 过程需要不断用历史帧训练网络, 并且 iMAP 采用单个 MLP 网络, 受限于网络的大小, MLP 网络会逐渐遗忘掉此前构建的地图, 无法应用于大场景空间; 同时, 早期的基于 NeRF 的 SLAM 方法的跟踪线程缺乏回环检测的支持, 导致定位精度不如经典视觉 SLAM 方法高; 最后, 由于运用了多层 MLP, 导致运算量大、NeRF 训练时间长, 很难做到实时。后续的相关研究在提高神经隐式表征效率、解决大尺度场景建图灾难性遗忘问题、增加回环检测和全局优化实现全局一致性以及解决动态干扰等多个方面进行了卓有成效的研究。

2.1 提高神经隐式表征效率

2.1.1 显式表征与隐式表征相结合提升表征效率

早期的基于 NeRF 的 SLAM 系统采用基于坐标的 MLP 隐式表示方法, 如 iMAP^[29]、NeRF-SLAM^[31]、

vMAP^[32]等, 通过基于坐标的 MLP 来连续隐式地表征全局场景, 能够实现高质量的地图和纹理重建, 但由于不允许局部场景更新以及受网络容量大小的限制, 导致不适合在线场景重建。为了提高神经隐式表征效率来提高重建性能, 研究者们考虑将传统的三维显式表征(如网格、点云等)与基于 MLP 的神经隐式表征相结合来优化神经隐式表征效率, 主要分为基于网格的神经隐式表征方法、基于三平面的神经隐式表征方法、基于点的隐式表征方法和基于二次曲面的神经隐式表征方法等, 相关方法对比如表 1 所示。

(1) 基于网格的神经隐式表征方法

基于网格的神经隐式表征方法结合了神经网络的表征能力和网格结构的特性, 网格的优点是邻域查找和上下文聚合是快速和直接的^[47], 但需要预先指定网格的分辨率, 基于网格的神经隐式表征方法可以采用分层特征网格、八叉树网格、哈希网格等来压缩内存。

表 1 显式表征与隐式表征结合进行位置编码方法对比

场景编码方式	代表算法	优势	局限性	适用场景
基于坐标的 MLP 表征	iMAP ^[29] 、NeRF-SLAM ^[31] 、vMAP ^[32] 、MLM-SLAM ^[33]	可以在任意空间分辨率进行采样; 连续表征进行表面空洞填充。	不允许局部场景更新; 重建过度平滑; 难以实时。	适用于需要高分辨率的重建和渲染场景。
分层特征网格	NICE-SLAM ^[30] 、NICER-SLAM ^[34] 、SNI-SLAM ^[35] 、NeSLAM ^[36] 、LONER ^[37]	不同尺度上表示场景, 有助于同时捕捉场景的粗略结构和精细细节; 允许局部更新并可扩展到大型室内场景; 保持细节的同时减少内存和计算需求。	多层次的特征表示可能增加计算复杂性, 表现出较慢的运行性能。	适合于大型室内场景的精细重建。
基于网格的隐式表征	Vox-Fusion ^[38] 、SHINE-Mapping ^[39] 、NGEL-SLAM ^[40]	减少内存需求, 更高效地处理大规模数据集; 可变分辨率允许在密集区域更细致的划分。	随着八叉树深度增加, 需要更为复杂的数据结构来表示; 重建未观察区域困难。	适合需要对隐式地图进行增量扩展场景。
哈希网格	Co-SLAM ^[41] 、LONER ^[37] 、NEWTON ^[42] 、MoD-SLAM ^[43]	能够显著减少训练 NeRF 所需时间, 更适合实时应用。	对内存要求更高, 尤其是在大规模场景中; 在没有观测的区域, 可能无法提供准确的空间信息。	适合需要实时应用场景。
基于三平面的隐式表征	ESLAM ^[44] 、S3-SLAM ^[45] 、PLG-SLAM ^[46]	能够更有效地学习场景高频细节; MLP 解码器变得更小, 在计算效率和三维重建质量之间取得平衡。	视角依赖性缺失, 无法捕捉和模拟视角变化对外观的影响; 额外 MLP 解码器可能引入额外的计算负担;	适合需要在参数数量和重建质量之间取得平衡的场景。
基于点的隐式表征	Point-SLAM ^[47] 、Loopy-SLAM ^[48] 、PIN-SLAM ^[49] 、GLORIE-SLAM ^[50] 、CP-SLAM ^[51]	相比基于网格的方法无需提前指定分辨率, 具有自适应性, 能够根据场景动态调整点云密度; 不需要维护固定的网格结构, 可以有效进行地图校正。	点云稀疏、无结构且可能包含噪声, 使得分割、识别和重建等处理任务更加复杂。	可有效进行地图校正, 适合回环检测和全局优化场景。
基于二次曲面的隐式表征	Q-SLAM ^[52]	比点云更紧凑, 提高存储效率并解决大体积点云最近邻搜索效率低下问题。	在特定环境下(如室外或光线变化大的场景)的性能尚未得到充分验证, 可能存在局限性。	适合室内需要在精确的几何建模场景。

分层特征网格 (hierarchical feature grids) 使用多个不同分辨率的网格层次来表示场景, 每个层次的网格可以捕捉不同尺度的场景特征, 从而允许系统在保持细节的同时减少内存和计算需求. NICE-SLAM^[30] 利用分层特征网格在不同尺度上捕捉场景细节, 分层特征网格不仅融合了局部和全局信息, 还通过预训练的几何先验知识, 增强了对场景结构的理解, 此外, 通过对特征网格进行动态局部更新, 减少计算负担, 避免全局更新导致信息丢失, 从而提高地图的长期一致性和准确性. NICER-SLAM^[34] 采用分层特征网格捕捉和表示场景几何细节, 通过将分层特征网格分为粗略级别 (coarse level) 和精细级别 (fine level), 并采用相应的 MLP 解码器用于场景的几何表示, 使得能够有效处理不同大小的物体和场景结构.

八叉树网格 (octree grid) 通过将三维空间划分为多个立方体 (体素), 并采用八叉树的方式存储, 这种数据结构可以有效地处理稀疏体积数据, 通过仅存储非空体素来减少内存, 在处理大规模 3D 数据时特别有用, 因为它可以显著减少存储的计算资源的需求并且支持快速查询和操作. Vox-Fusion^[38] 通过稀疏体素八叉树有效地组织和存储 3D 空间数据, 只保留非空体素信息, 极大地降低了存储开销, 并且支持快速查询和操作, 这种数据结构允许以稀疏方式存储 3D 场景, 从而提高了神经隐式表示的效率. SHINE-Mapping^[39] 采用了八叉树这种稀疏数据结构来高效地存储局部特征向量, 这些特征向量通过一个共享的浅层多层感知机 (MLP) 作为解码器, 被转换成符号距离场 (SDF), 通过在局部特征之间进行平滑的插值处理, 该方法能够支持在任意分辨率下查询几何信息, 此外, 通过引入增量八叉树结构, 该方法还能够实现地图构建的增量更新.

哈希网格 (hash grid) 是一种用于高效表示和渲染大规模 3D 场景的数据结构, 通过将空间划分为不同分辨率的网格单元, 并使用哈希表来存储每个单元的特征, 从而实现对大规模场景的紧凑表示, 主要优点在于能够通过多分辨率哈希编码快速训练和渲染 NeRF 模型. Instant-NGP^[53] 利用哈希网格加速 NeRF 的训练和渲染, 通过多分辨率哈希编码和优化的网格结构, 可以将 NeRF 的训练速度以小时计的时间开销缩短到秒级, 方法的核心在于提高了采样效率和网络的查询速度, 从而显著减少训练时间. 但在处理未观测区域时可

能留下空洞, 因哈希网格通常依赖于观测数据来构建空间的数据表示, 在没有观测的区域, 哈希网格可能无法提供准确的空间信息, 从而导致重建的表面出现不连续或缺失的部分, 基于此, Co-SLAM^[41] 结合了坐标编码 (one-blob 编码) 和稀疏参数编码 (哈希网格) 以实现快速收敛和表面空洞填充, 通过 one-blob 编码在未观测区域实现平滑的洞填补, 并将场景表示为多分辨率哈希网格, 利用了哈希网格的高收敛速度以及其表示高频局部特征的能力, 同时保持了坐标编码的平滑一致性先验.

(2) 基于三平面的隐式表征方法

三平面编码是一种参数化的场景表示方法, 通过将显式特征投影到 3 个轴正交的特征平面上, 可以有效地存储和处理三维数据, 这种表达方式通过将笛卡尔坐标映射到更高维度的空间, 使得网络能够更有效地学习场景高频细节, 同时, 三平面表达方式使得 MLP 解码器变得更小, 由于减少了网络复杂度, 将大部分的表达能力转移给显式特征, 能够保持参数的稀疏性以提高渲染效率. ESLAM^[44] 通过结合三平面编码和 MLP 来提高神经隐式表示的效率, 在 ESLAM 中, 三平面编码被用来表示场景的局部高频特征, 而全局低频特征则通过多层感知器 (MLP) 网络来表示, 这种结合了三平面编码和 MLP 的方法, 既利用了三平面编码在高频细节方面的优势, 又利用了 MLP 在捕捉全局特征和保持一致性方面能力, 从而实现了在未见区域的准确、平滑和填孔的重建, 提高场景重建的质量和鲁棒性, 同时保持了实时性能和参数效率. S3-SLAM^[45] 基于稀疏三平面编码, 这种编码方式通过使用哈希网格对正交平面特征进行稀疏化, 仅使用 2%–4% 的常用三平面参数 (从 100 MB 减少到 2–4 MB), 就能在高达 512 分辨率下高效地实现场景重构, 可以显著减少了模型的内存消耗, 在保持较低参数数量的同时, 实现高分辨率和高质量的场景重建.

(3) 基于点的神经隐式表征方法

基于点的隐式神经表示方法通过将三维空间中的点与 MLP 网络相结合, 来实现对场景的高效和高精度表征, 方法的核心在于使用 MLP 来学习和编码每个点的位置信息以及局部场景信息. 与基于网格的方法不同, 基于点的方法不需要存储整个用于建图的输入帧历史, 可以更加高效地使用内存资源, 同时, 由于不需要维护固定的网格结构, 使得可以有效进行地图校正,

在回环检测和全局优化方面更有优势. Point-SLAM^[47]利用点云对场景特征进行编码,实现了细粒度的几何和颜色重建,同时采用动态点密度策略,允许系统根据输入信息的密度自适应地调整锚点的密度,这种策略在细节较少的区域减少运行时间和内存使用,而在需要解析精细细节的区域则增加点密度,从而提高计算和存储效率. Loopy-SLAM^[48]将神经场景表示锚定在点上,允许在进行地图更新时无需从头计算密集地图,这种灵活性提高了重建效率,并减少了计算负担. PIN-SLAM^[49]的隐式地图基于稀疏的可优化的神经点,利用体素哈希技术高效地索引神经点,并通过快速隐式地图注册算法,无需最近点关联,可在适用的 CPU 上实现实时性能.

(4) 基于二次曲面的隐式表征方法

二次曲面将 3D 场景分解为一组稀疏的二次曲面,通过将给定点云分割成多个面片,并将每个面片拟合到一个二次隐式函数中,每个函数随后与其他几何描述符(中心位置和协方差矩阵)结合,共同描述一个

3D 场景,二次曲面表征比传统点云更为紧凑,提高存储效率的同时可以解决点云扫描子图匹配和全局地图表示方法受到大体积点云中最近邻搜索效率低下的问题^[54]. Q-SLAM^[52]利用二次曲面进行隐式场景表示来提高效率和精度,通过将三维场景中的点云转换为二次曲面的集合,显著减少全局地图存储所需的内存,而且基于在二次曲面周围采样而不是整个 3D 空间,显著降低了复杂性.

2.1.2 优化几何表征提升表征效率

经典 NeRF^[11]和早期的基于 NeRF 的 SLAM 方法(如 iMAP^[29])采用体密度来进行场景几何表征,NeRF 模型的 MLP 网络解码输出的是体密度和颜色值.然而,体密度的预测效果容易出现噪点和低保真的几何形状.为了优化几何表征提升表征效率,相关研究在基于 NeRF 的 SLAM 系统中采用占用表示和符号距离函数(signed distance function, SDF),即在 NeRF 模型的 MLP 网络解码输出为占用表示或 SDF 和颜色值.相关方法对比如表 2 所示.

表 2 优化几何表征方法对比

几何表征	代表算法	优势	局限性	适用场景
体密度	iMAP ^[29] 、MeSLAM ^[55] 、NEWTON ^[42] 、LONER ^[37] 、DIM-SLAM ^[56]	提供连续场景表征,可以用于平滑地表示物体表面和内部结构.	容易出现噪点、低保真的几何形状.	适合需要考虑物体内部属性的应用.
占用表示	NICE-SLAM ^[30] 、Point-SLAM ^[47] 、Loopy-SLAM ^[48] 、Uncle-SLAM ^[57] 、vMAP ^[32] 、KN-SLAM ^[58]	实现简单;可高效表示空间占用情况.	以体素形式表示,可能导致在表示细节时受分辨率限制;不提供物体表面形状详细信息.	适合快速判断空间占用状态的应用.
SDF/TSDF	ESLAM ^[44] 、NICER-SLAM ^[34] 、SHINE Mapping ^[39] 、Co-SLAM ^[41] 、Go-SLAM ^[28]	提供连续的场景表征,能够提供精确的几何形状;SDF 可以支持如追踪等任务,对于重建表面非常有帮助.	在需要精确几何重建的场景中计算比较复杂;网络学习起来不如占用直观.	适合需要精确表面信息的应用.

● 占用表示通过一个二值函数来表示空间中的每个点是否被占用,这种表示方法简单并可高效表示空间占用情况,占用表示通过离散化的体素网格和二值化占用状态简化了空间表示和处理,使得系统能够快速进行空间查询和更新,这对于实时 SLAM 系统至关重要,此外,占用表示局部更新特性和几何一致性也更适合大规模场景的重建和维护,相比之下,体密度需要对整个空间进行连续的密度估计,这在计算上更为复杂和资源密集,且在处理噪声和不确定性时可能不够鲁棒. NICE-SLAM^[30]、Point-SLAM^[47]、Loopy-SLAM^[48]等采用占用表示来优化几何表征提高表征效率.

● SDF 是一个定义在空间中的函数,可表示为

$f(x,y,z) = d$, 它返回一个点到物体表面的最短距离,表示点相对于表面的内部或外部的的位置,通过使用 MLP 来拟合 SDF 函数,可以实现对复杂形状的隐式表示,这种表示方法对于捕捉场景的几何细节非常有效,尤其是在复杂的室内场景中. iDF-SLAM^[59]采用 SDF 来进行几何表征,可以更准确地进行深度估计,从而有助于更清晰地定义几何表面和改善场景重建任务. ESLAM^[44]采用隐式的 TSDF 来表示几何形状,TSDF 是 SDF 的一种特殊形式,通过引入截断机制来限制考虑的距离范围,实验结果表明 TSDF 比常见的基于渲染的表示(体积密度和占用率)收敛得更快,从而提高神经隐式的效率和重建质量.然而,使用 SDF 代替体密

度也会牺牲由渲染函数带来的填补空洞的性能^[12], 总的来说, 体密度更加适合渲染, 而 SDF 更侧重于重建, 因此如何找到 SDF 值和体密度之间的权衡是一个需要解决的问题. 有研究通过控制 SDF 到密度的转换来兼顾渲染和重建质量, NICER-SLAM^[34]作为 NICE-SLAM^[30]的继承者, SDF 被用来替代在 NICE-SLAM 中使用的占用表示, 通过局部自适应的方法, 即在体积渲染方程中引入一个参数 β 来控制从 SDF 到体密度的转换, 参数 β 模拟了物体表面附近的平滑程度, 将 SDF 转换为体密度, 以便在体积渲染方程中使用, 可以动态调整不同区域 SDF 值的平滑程度从而更好地适应地图的几何复杂性, 以此达到 SDF 和体密度重建和渲染之间的平衡.

2.2 解决大尺度场景建图灾难性遗忘问题

在 SLAM 过程中需要不断拓宽地图, 而早期基于 NeRF 的 SLAM 系统受限于 MLP 的容量, 导致大尺度范围建图存在灾难性遗忘问题, 同时在大规模室内场景中因累计误差而影响精度和鲁棒性, 如 iMAP^[29]使用两步迭代的方法分别对相机位姿和场景表示参数的联合优化进行跟踪和映射, 然而由于只使用单个 MLP, 受到神经网络泛化能力的限制, 建图范围仅局限于室内小场景, 难以预测复杂的几何或者大型场景. 现有扩大场景重建规模的研究主要包括基于局部子图表示、基于分层特征网格和基于八叉树的方法等 3 种, 相关方法的对比如表 3 所示, 也可以将几种方法结合使用以提高大尺度场景重建的效果和效率^[60].

表 3 大尺度场景三维重建方法对比

分类	代表算法	优势	局限性
局部子图	MeSLAM ^[55] 、PLG-SLAM ^[46] 、NISB-map ^[61] 、CP-SLAM ^[51] 、NEWTON ^[42] 、NGEL-SLAM ^[40] 、Loopy-SLAM ^[48] 、Vox-Fusion++ ^[60]	精度高, 可以更精细地捕捉场景细节; 更灵活, 通过调整子图大小和数量来优化性能.	内存消耗大; 计算成本高; 在地图融合中容易产生伪影.
分层特征网格	NICE-SLAM ^[30] 、NICER-SLAM ^[34] 、SNI-SLAM ^[35]	分层场景表示允许局部更新场景从而能够扩展到大型室内场景的详细重建; 内存效率高;	计算复杂性高, 可能表现出较慢的运行性能; 分层场景需要预训练几何先验.
八叉树	Vox-Fusion ^[38] 、SHINE-Mapping ^[39] 、NeRF-LOAM ^[62] 、Vox-Fusion++ ^[60]	可对未知场景进行增量地图构建; 递归地将空间划分为更小的区域提高存储效率; 计算效率高, 适合实时应用.	精度不高; 灵活性受到限制.

● 基于分层特征网格结构解决灾难性遗忘问题. 分层特征网格的设计允许对场景的局部区域进行更新, 这是大规模场景重建的关键. 通过仅更新可视范围内的特征网格, 可以有效地处理新的观测数据, 而不需要重新优化整个场景. NICE-SLAM^[30]通过基于分层特征网格的分层场景表示方法来解决传统神经隐式方法在大场景重建中的局限性, 使用 3 个层级的

● 基于局部子图表示方法实现大尺度场景重建. 通过将大型、复杂场景划分为多个小的、可管理的子图, 并基于多个 MLP 网络实现高效、准确和鲁棒的重建, 从而允许对每个局部区域进行独立的处理和优化, 进而解决灾难性遗忘问题, 然而, 这种方法内存消耗大且在地图融合中容易产生伪影. 在 iMAP^[29]工作的基础上, MeSLAM^[55]将大场景划分为数个子区域, 使用多个 MLP 网络来表示场景的不同部分, 每个单一的 MLP 网络仅代表场景的一部分, 最后在重建时进行拼接实现大场景的精确建图. PLGSLAM^[46]使用渐进式场景表示方法来处理大型室内场景, 即当相机移动到局部场景表示的边界时动态初始化新的局部场景表示, 整个场景被划分为多个局部场景呈现, 在局部场景表示中使用三平面来捕获局部高频特征, 同时结合 MLP 网络来处理低频特征, 从而实现在未观测区域的平滑和场景完成, 可以显著提高大型室内场景的场景表示能力. NEWTON^[42]通过局部神经场的集合来捕捉和更新场景的表示, 根据运行时的观察动态分配新的神经场, 从而允许模型在需要时扩展其表示, 而不受限于预先定义的全局结构; 同时为了有效表示无边界场景, 提出了一种球面参数化方法, 允许将 3D 空间中的点映射到单位球面上, 不仅能够高效地分配特征, 还不受视锥边界的限制, 不会出现由于视锥边界造成的伪影, 球面参数化能够充分利用训练视图的信息, 而不受视场限制, 从而有效地表示和处理无界场景.

特征网格 (粗糙、中等、精细) 和相应的 MLP 解码器用于场景的几何表示, 同时使用 1 个体素网格用于场景颜色表示, 这种分层结构允许地图局部更新, 而不是全局更新, 从而使得系统能够有效地处理大型室内场景.

● 基于八叉树结构对隐式地图进行增量扩展. 这类方法采用八叉树的结构来划分场景并支持动态扩展,

允许在不知道环境的情况下进行动态扩展,这对于处理大型和未知场景尤为重要. Vox-Fusion^[38]采用八叉树的场景划分结构并动态分配新的体素,通过 Morton 编码对体素坐标进行编码以提高检索速度,从而对隐式地图进行增量扩展,完成室外大场景的建图. SHINE-Mapping^[39]采用哈希表来存储八叉树结构中每层的特征数据,并通过特有的 Morton 编码将高维信息映射到单一维度,这种设计允许在不预先分配内存的情况下轻松扩展地图范围. NeRF-LOAM^[62]使用基于八叉树并结合神经隐式嵌入实现大尺度建图,为了解决未知的大规模室外环境下的增量里程计和地图构建问题,设计了动态体素嵌入策略,通过扩展查找表并基于最大 Morton 编码更新,以存储体素的访问信息,未访问的体素将被分配初始化的嵌入,使得能够轻松地扩展地图而无需任何预分配或耗时的循环.

2.3 增加回环检测和全局优化实现全局一致性

早期的基于 NeRF 的 SLAM,如 iMAP^[29]和 NICE-

SLAM^[30],在稠密建图方面更具优势,但由于没有回环检测和全局优化来消除累计误差,在定位精度方面不如基于特征的传统 SLAM,如 ORB-SLAM2^[23],相关研究在增加回环检测和全局优化实现全局一致性进行了尝试,通过减少累计误差,提高定位准确性.

2.3.1 增加回环检测

传统 SLAM 系统具有低延迟和高精度的追踪性能,并通过回环检测确保全局一致性.基于 NeRF 的 SLAM 系统能够实现逼真的场景重建,但是早期基于 NeRF 的 SLAM 系统由于缺乏回环检测,即便将回环检测集成到系统中,更新位姿非常耗时,映射网络收敛速度较慢,在迅速更新 MLP 表示以适应回环方面仍面临挑战,从而存在较高的延迟,导致缺乏全局一致性而导致定位表现不佳,相关研究通过在基于 NeRF 的 SLAM 系统中加入回环检测来提高定位精度和鲁棒性,现有的方法可分为基于解耦的方法和基于耦合的方法两种,如表 4 所示.

表 4 增加回环检测方法对比

方法	代表算法	优势	局限性
基于解耦	NEWTON ^[42] 、Orbeez-SLAM ^[27] 、MoD-SLAM ^[43] 、Go-SLAM ^[28] 、HI-SLAM ^[63] 、NEGL-SLAM ^[40] 、GLORIE-SLAM ^[50]	(1) 独立优化,可能在某些情况下带来更好的性能; (2) 更为灵活,可以更容易地替换或更新SLAM系统的某个部分.	(1) 跟踪和建图独立进行,可能导致数据冗余和不一致性; (2) 无法有效利用建图信息来辅助跟踪,导致跟踪误差累积.
基于耦合	基于局部子图的方法: MIPS-Fusion ^[64] 、NF-Atlas ^[65] 、CP-SLAM ^[51] 、Vox-Fusion++ ^[60] 基于点的方法: PIN-SLAM ^[49] 、Loopy-SLAM ^[48]	(1) 共享同一场景表示来保持跟踪和建图之间数据一致性,有助于提高系统鲁棒性; (2) 跟踪和建图是紧密集成的,更有效地利用建图信息来辅助跟踪,减少误差传播; (3) 减少数据传输和处理的延迟,更适合实时应用.	(1) 不如解耦方法灵活,更新或替换系统的某个部分可能需要更多的工作; (2) 需要更高的计算复杂性,因为它们需要同时处理跟踪和建图任务.

(1) 基于解耦的方法

在基于 NeRF 的 SLAM 系统中,基于解耦的方法直接采用传统 SLAM 的视觉里程计作为前端,如 ORB-SLAM2^[23]、ORB-SLAM3^[24]、DROID-SLAM^[66]等,前端通过与后端 NeRF 建图进行解耦,回环检测方法直接继承传统 SLAM,基于解耦的方法没有在前端位姿估计中使用 MLP 进行优化,同时大多采用多分辨率哈希网格以加速后端 MLP 收敛速度,有助于在回环检测过程中保持系统的实时性和效率.由于跟踪模块只专注于相机位姿估计,而不增加同时更新地图表示的复杂性,基于解耦的方法具有较好的跟踪性能^[13],但是,由于跟踪过程独立于建图线程,导致数据冗余和不一致性. NEWTON^[42]和 Orbeez-SLAM^[27]的回环检测是利用 ORB-SLAM2^[23]的闭环处理来实现的,通过使用 ORB-

SLAM2 的跟踪系统,继承 ORB-SLAM2 的回环检测过程,使用词袋模型和关键帧数据库来检测新的关键帧和数据库中的帧是否相似区域,并通过特征点匹配、求解和优化 Sim(3) 来确定是否检测到回环. Go-SLAM^[28]通过将 DROID-SLAM^[66]的框架扩展至实时闭环检测场景,并利用 Instant-NGP^[53]算法实现与地图的集成,通过检测当前关键帧与历史关键帧之间的共视性来确定闭环以解决尺度漂移问题,但由于 RGB 输入缺乏场景的原始深度信息,导致场景重建误差较大, HI-SLAM^[63]在此基础上通过搜索潜在的回环闭合并执行基于 Sim(3) 的位姿图束调整 (PGBA) 和基于 Hessian 矩阵的先验深度估计来克服 Go-SLAM 误差较大的问题. NGEL-SLAM^[40]使用 ORB-SLAM3^[24]的传统特征跟踪模块来实现回环检测,这个模块负责跟踪相机的姿态,并在检

测到回环时进行闭环融合,同时,通过多个神经隐式场来表示场景,每个隐式场对应一个局部区域,这种表示方法允许系统快速调整闭环,从而维护全局一致性。

(2) 基于耦合的方法

基于耦合的方法将回环检测与位姿估计和地图构建紧密集成,使得回环检测的结果可以直接反馈和优

化整个系统的位姿和地图表示,基于耦合方法虽然可以提供更全局的一致性,但可能会增加系统的计算复杂性和处理开销.现有基于耦合的回环检测方法可以分为基于局部子图的方法和基于点的方法^[49],也有相关研究将两种方法结合以更好地提升全局一致性^[48].相关方法对比如表 5 所示。

表 5 基于耦合的回环检测方法对比

方法	优势	局限性	适用场景
基于子图	可减少长程建图时的累积误差,定位和建图精度更高;建图的全局一致性和准确性更高.	子图之间存在数据冗余;在进行子图间的回环检测时,可能需要较大的计算量.	适用于全局一致性和精度上要求更高的场景.
基于点	在全局姿态调整中具有弹性和可变性,有助于修正回环;处理的数据量相对较小,计算复杂度较低;处理速度上通常更快,适合于需要快速响应的场景.	可能需要复杂的邻域查询和插值计算,增加了计算复杂度;可能需要有效的特征提取机制来保证点的代表性和区分性.	适用于内存效率和实时性要求高的场景.

基于局部子图实现回环检测.将环境划分为多个局部子图,并在这些子图之间检测和闭环回环从而实现全局一致性.MIPS-Fusion^[64]通过其独特的子图管理策略实现回环检测,基于刚性配准多个隐式地图子图来完成闭环检测过程,系统还引入了一种透视性阈值机制来识别回环,但目前的设计仅限于校正小规模漂移,并不支持对大规模漂移的校正.NF-Atlas^[65]通过神经特征体积(子地图)与位姿图优化相结合实现回环检测,将神经特征体积(子地图)视为位姿图节点,将体积间的相对位姿视为位姿图边,实现了整个神经特征场局部的刚性和全局的弹性,这意味着在全局层面上,避免了增量映射时的灾难性遗忘,当回环闭合时,由于基于弹性位姿图的表示,只需更新神经体积的原点而无需重新建图,简化了回环检测的处理。

基于点的神经隐式表示进行回环检测.基于点的神经隐式表示特别适合执行地图校正,不需要存储用于映射的输入帧的整个历史,以执行地图校正,PIN-SLAM^[49]利用神经隐式表示中提取的神经点特征进行回环检测,这些特征用于识别之前访问过的环境位置,

通过体素哈希技术高效索引神经点,有助于快速识别和匹配回环中的关键特征点,神经点闭环时具有固有的弹性和可变形性,能够随着全局姿态调整而变形使得地图在调整全局姿态时能保持弹性并避免结构重复,增强了回环检测的能力.为了纠正大规模漂移,Loopy-SLAM^[48]使用全局位置识别来在线触发循环闭合,即将每个全局关键帧添加到一个 DBoW3 数据库,用于全局位置识别,并直接在场景表示上通过简单有效的刚性校正来全局对齐轨迹和子图,由于是基于点的神经表示,使得地图校正变得高效,避免了在重叠区域产生可见的接缝,并在轨迹捕获结束时对颜色和几何特征进行特征细化,允许校正大规模漂移。

2.3.2 全局优化

基于 NeRF 的 SLAM 系统通过后端建图中引入全局优化,即在全局范围内约束全局轨迹的漂移,以保证整个系统的一致性,通过减少累计误差,提高跟踪精度,基于 NeRF 的 SLAM 系统与传统 SLAM 一样主要采用图优化的方式进行全局优化,具体可分为全局束调整和位姿图优化.相关方法对比如表 6 所示。

表 6 基于 NeRF 的 SLAM 全局优化方法对比

方法	代表算法	优势	局限性	适用场景
全局束调整 (GBA)	Go-SLAM ^[28] 、Co-SLAM ^[41] 、S3-SLAM ^[45] 、EC-SLAM ^[67] 、NGEL-SLAM ^[40] 、MUTE-SLAM ^[68] 、GLORIE-SLAM ^[50]	优化所有或采样相机位姿和场景表示来最小化重投影误差,通常能够获得较高的精度;全局一致性更高.	计算成本高,需要处理大量数据和复杂的优化问题;不适合实时.	适合对精度要求更高的小型或中型静态场景.
位姿图优化 (PGO)	NF-Atlas ^[65] 、CP-SLAM ^[51] 、HI-SLAM ^[63]	通过优化位姿图中的边(即相对位姿)来调整相机位姿,计算效率更高;适合实时场景.	精度不如全局束调整;全局一致性不如GBA.	对实时性要求高的场景.

全局束调整(global bundle adjustment, GBA)同时优化所有关键帧或采样关键帧的位姿和地图点,

通过非线性最小化二乘方法最小化重投影误差,以提升整体地图的精度,方法更侧重于系统位姿和场景表

示的联合优化,与基于 NeRF 的 SLAM 后端优化更契合,但需要处理大量数据和复杂的优化问题,更适合对精度要求更高的场景. Go-SLAM^[28]通过一个单独的后端跟踪线程来运行全局 BA,从而对所有历史关键帧进行优化,以保持 3D 几何的全局一致性,生成全局一致的位姿和深度预测,全局 BA 的引入减缓了 Go-SLAM,但显著改善了全局位姿估计. Co-SLAM^[41]在地图和相机位姿联合优化中使用全局 BA,在所有关键帧执行全局 BA,但不存储完整的关键帧,只选择存储代表每个关键帧的像素子集(约 5%)来进行相机位姿和场景表示联合优化,从而能更频繁地插入新的关键帧,并维护更大的关键帧数据库. S3-SLAM^[45]采用分层光束法平差(HBA)用于优化局部外观并确保全局几何一致性,在每次迭代中从所有关键帧采样 10% 的光线以保持全局一致性,同时,建立局部滑动窗口,并从中采样更多(大于 10%)的光线优化局部相机位姿和重建局部场景,局部样本射线的数量根据每个关键帧的最佳损失进行加权的. EC-SLAM^[67]提出了一种全局约束的 BA 策略,通过计算所有关键帧与当前帧的相机之间的光心距离和视角来动态确定滑动窗口中的关键帧数据和顺序,以消除随机性的影响,确保全局 BA 的有效性.

位姿图优化 (pose graph optimization, PGO) 将相机的位姿和约束表示为图,通过优化位姿图中节点(即关键帧位姿)和边(即位姿之间的相对变换),能够有效消除累计误差,方法更侧重于位姿优化,精度和全局一致性不如全局束调整,但计算效率更高,适合实时场景. NF-Altas^[65]通过将大地图表示为多个通过位姿图链接

的子图(神经特征体积),神经特征体积视为位姿图的节点,不同神经特征体积视为位姿图的边,这些边连接不同的节点,系统被设计为局部刚性和全局弹性,即子地图被允许在大尺度上进行位姿图优化求解以适应全局的一致性而局部区域的姿态和映射是不变的,当闭环发生后,只需对子图的位姿进行更新和优化,而无需重新映射整个区域. 针对多机协同场景, CP-SLAM^[51]采用位姿图优化进行全局优化,在中央服务器对子地图融合之后,建立以每帧姿态为节点和相对位姿和环路相对位姿为边的全局姿态图模型,采用 Levenberg-Marquardt 算法解决非线性全局姿态图优化问题,旨在减少建图和跟踪中的累积误差,由于采用基于点的神经隐式表示,可以根据优化后的位姿来优化点云的 3D 位置.

2.4 解决动态干扰问题

目前为止,大多数基于 NeRF 的 SLAM 方法都是以不移动物体的静态环境为基本假设,而在有动态干扰的场景中,会导致位姿估计中错误的关联和累计漂移放大,同时产生伪影或灾难性遗忘而影响重建质量,虽然 NICE-SLAM^[30]基于跟踪线程中的异常损失值进行基于深度阈值的像素滤波,但并不能完全消除动态物体的影响,因此,部分改进工作将基于 NeRF 的 SLAM 方法扩展到动态场景中,然而动态环境中部署基于 NeRF 的 SLAM 系统尚未得到充分研究,目前已有的解决基于 NeRF 的 SLAM 系统的动态干扰问题的方法包括基于语义分割和结合事件相机等,且以基于语义分割方法为主,如表 7 所示.

表 7 基于 NeRF 的 SLAM 动态环境算法对比

方法	动态环境算法	关键方法	语义先验	优势和局限性对比
基于语义分割	DN-SLAM ^[69]	利用ORB特征进行对象跟踪,并采用语义分割、光流和分段任意模型(SAM),有效地识别和隔离场景中的动态对象.	SAM ^[70]	优势: 提供丰富的环境信息,包括物体的类别和实例,有助于理解场景的语义内容. 局限性: 语义分割网络可能会增加计算负担,影响实时性;依赖于语义分割的准确性,对于未预先定义的动态物体无法有效识别和处理.
	NID-SLAM ^[71]	通过语义分割和掩膜修正,精确地从RGB-D图像中移除动态物体,并彻底恢复被遮挡的背景.	YOLO	
	DDN-SLAM ^[72]	结合语义特征和混合高斯分布模型的特征点分割方法,用于区分动态、静态和潜在静态物体.	YOLOv9	
	RoDyn-SLAM ^[73]	通过一种运动掩码生成方法,用于过滤无效的采样光线,有效地融合了光流掩码和语义掩码,以增强运动掩码的精度,从而识别并区分动态和静态区域.	OneFormer ^[74]	
	DynaMoN ^[75]	利用语义分割和通用运动掩码来处理动态内容,有助于在初始相机位姿估计和静态聚焦的光线采样中处理动态场景.	DeepLabV3 ^[76]	
结合事件相机	EN-SLAM ^[77]	结合事件相机和RGBD相机的数据,利用事件相机高动态范围和高分辨率的特性生成事件流创建动态物体掩码.	—	优势: 适用于高动态范围场景和高速运动环境和需要快速响应的实时应用. 局限性: 事件相机成本更高,可能限制商业化应用的普及.

基于语义分割的方法. 基于语义分割的方法是基于 NeRF 的 SLAM 系统解决动态干扰问题采用的常用方法, 使用深度学习进行语义分割, 剔除动态物体, 减少其对位姿估计的影响, 提升基于 NeRF 的 SLAM 在动态环境下的鲁棒性, 然而语义分割网络可能会增加计算负担, 导致系统运行速度降低, 影响实时性, 而实时性对于基于 NeRF 的 SLAM 系统至关重要, 此外, 动态物体处理依赖于语义分割的准确性, 对于未进行语义先验的动态物体可能无法有效识别和处理.

DN-SLAM^[69]基于 ORB 特征跟踪动态物体, 利用语义分割获得潜在的移动物体, 并结合光流和分段任意模型 (SAM) 进行精细分割, 通过剔除动态物体来减少冗余, 从而提高系统在动态环境中的性能. NID-SLAM^[71]通过语义分割来识别图像中的动态物体, 并生成语义掩码, 通过提出的深度引导的语义掩码增强方法来消除边缘区域的不准确性, 将修正后的深度信息与准确地语义掩码结合来检测和去除动态物体, 并将静态地图投射到当前帧以修复被这些物体遮挡的背景. DDN-SLAM^[72]通过结合语义分割和高斯混合模型对动态物体进行特征点分割以限制动态物体的干扰, 利用光流和重投影误差检查进行长期数据关联和背景恢复, 以及应用体渲染技术, 有效地从动态环境中去除动态物体, 同时保留潜在的动态对象, 并实现高质量的场景重建. RoDyn-SLAM^[73]通过创新性地融合光流掩码和语义掩码生成运动掩码, 精确区分和过滤动态物体, 同时利用边缘 warp 损失和分割优化策略, 增强了位姿估计的几何一致性, 有效处理动态环境中的动态物体.

结合事件相机的方法. 基于事件相机捕捉场景中的动态变化, 以适应快速变化的环境, 事件相机具有低时延和高动态范围的特点, 适合高速运动和复杂光照场景, 并且对数据存储和计算资源的需求较小, 在实时性方面具有明显优势, 结合事件相机的基于 NeRF 的 SLAM 系统目前仍处于研究的早期阶段. EN-SLAM^[77]结合事件相机和 RGBD 相机的数据, 利用事件相机的高动态范围和高时间分辨率优势来补偿在快速运动中丢失的信息, 通过提出一种可微分的相机响应函数 (CRF) 渲染技术来共享辐射场将统一表示映射到 RGB 和事件相机数据, 以解决事件和 RGB 相机之间显著的区别, 此外, 通过将事件生成模型整合到优化过程中, 能够区分动态和静态物体, 并生成动态物体掩码, 以便在跟踪和映射过程中排除动态物体的影响, 然而, EN-SLAM

侧重于室内场景, 在无限长的轨迹中可能面临挑战.

3 未来研究方向展望

3.1 实时性能的提升

基于 NeRF 的 SLAM 方法能达到不错的建图效果, 但是基于神经隐式的场景表示方法渲染速度慢、占用内存资源多, 实时性能面临挑战, 而 SLAM 需要快速甚至实时的后端建图, 尽管采用哈希网格的神经隐式表示可以提高计算速度, 但需要更高的内存需求, 研究如何改进网络结构、采用更高效的采样策略、优化算法提升和硬件加速实时性能的同时保持高精度全局一致性仍然是未来的一个重要的研究方向.

3.2 提升动态场景下的鲁棒性

传统的 SLAM 方法和 NeRF 方法主要关注静态场景, 而现实生活场景通常涉及物体运动或拓扑变化, 如移动的车辆和行人, 虽然一些基于 NeRF 的 SLAM 方法试图解决这个问题, 但目前仍处于研究的初期, 仍有很多的改进空间, 特别是在高度动态的环境中, 现有的动态场景下基于 NeRF 的 SLAM 方法主要是基于深度学习的语义分割的方法, 实时性目前受限于分割网络的速度同时依赖于语义先验, 研究如何平衡分割质量和分割网络速度是未来的一个研究方向.

3.3 提升全局一致性

通过增加回环检测和全局优化实现全局一致性需要消耗大量的计算资源, 在实时应用中容易出现性能瓶颈, 基于解耦的方法虽然具有更好的跟踪性能, 但跟踪与建图独立进行, 无法有效利用建图信息来辅助跟踪, 导致跟踪误差累积, 基于耦合的方法跟踪和建图紧密集成的, 更有效地利用建图信息来辅助跟踪, 减少误差传播, 但资源消耗大, 影响实时性. 研究如何通过优化算法或引入额外的约束来减少累积误差, 进一步提高全局一致性是一个重要研究方向.

4 结语

基于 NeRF 的 SLAM 方法能够进行连续的 3D 建模, 可以渲染新视角并预测位置区域, 提高场景重建的质量和细节, 此外, 通过反向传播算法计算损失函数来优化相机位姿和 NeRF 地图, 然而, 基于 NeRF 的 SLAM 方法目前仍然存在很多需要解决的问题和挑战, 如实时性能提升、动态场景处理、计算资源优化和提升全局一致性等, 需要后续研究者们进一步解决, 通过本研

究的梳理和展望, 希望可以为研究者们进一步改进基于 NeRF 的 SLAM 技术, 为基于 NeRF 的 SLAM 在机器人、自动驾驶和增强现实领域的应用和创新提供帮助。

参考文献

- 1 Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: Part I. *IEEE Robotics & Automation Magazine*, 2006, 13(2): 99–110. [doi: [10.1109/MRA.2006.1638022](https://doi.org/10.1109/MRA.2006.1638022)]
- 2 Bailey T, Durrant-Whyte H. Simultaneous localization and mapping (SLAM): Part II. *IEEE Robotics & Automation Magazine*, 2006, 13(3): 108–117. [doi: [10.1109/MRA.2006.1678144](https://doi.org/10.1109/MRA.2006.1678144)]
- 3 Li RH, Wang S, Gu DB. Ongoing evolution of visual SLAM from geometry to deep learning: Challenges and opportunities. *Cognitive Computation*, 2018, 10(6): 875–889. [doi: [10.1007/s12559-018-9591-8](https://doi.org/10.1007/s12559-018-9591-8)]
- 4 Wen SH, Li PJ, Zhao YJ, *et al.* Semantic visual SLAM in dynamic environment. *Autonomous Robots*, 2021, 45(4): 493–504. [doi: [10.1007/s10514-021-09979-4](https://doi.org/10.1007/s10514-021-09979-4)]
- 5 Su P, Luo SY, Huang XC. Real-time dynamic SLAM algorithm based on deep learning. *IEEE Access*, 2022, 10: 87754–87766. [doi: [10.1109/ACCESS.2022.3199350](https://doi.org/10.1109/ACCESS.2022.3199350)]
- 6 DeTone D, Malisiewicz T, Rabinovich A. SuperPoint: Self-supervised interest point detection and description. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Salt Lake City: IEEE, 2018. 337–33712. [doi: [10.1109/CVPRW.2018.00060](https://doi.org/10.1109/CVPRW.2018.00060)]
- 7 Sarlin PE, DeTone D, Malisiewicz T, *et al.* SuperGlue: Learning feature matching with graph neural networks. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2020. 4937–4946. [doi: [10.1109/CVPR42600.2020.00499](https://doi.org/10.1109/CVPR42600.2020.00499)]
- 8 Arandjelovic R, Gronat P, Torii A, *et al.* NetVLAD: CNN architecture for weakly supervised place recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016. 5297–5307. [doi: [10.1109/CVPR.2016.572](https://doi.org/10.1109/CVPR.2016.572)]
- 9 Tateno K, Tombari F, Laina I, *et al.* CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017. 6565–6574. [doi: [10.1109/CVPR.2017.695](https://doi.org/10.1109/CVPR.2017.695)]
- 10 Zhou HZ, Ummenhofer B, Brox T. DeepTAM: Deep tracking and mapping with convolutional neural networks. *International Journal of Computer Vision*, 2020, 128(3): 756–769. [doi: [10.1007/s11263-019-01221-0](https://doi.org/10.1007/s11263-019-01221-0)]
- 11 Mildenhall B, Srinivasan PP, Tancik M, *et al.* NeRF: Representing scenes as neural radiance fields for view synthesis. *Proceedings of the 16th European Conference on Computer Vision (ECCV)*. Glasgow: Springer, 2020. 405–421. [doi: [10.1007/978-3-030-58452-8_24](https://doi.org/10.1007/978-3-030-58452-8_24)]
- 12 Yang KY, Cheng YQ, Chen ZH, *et al.* SLAM meets NeRF: A survey of implicit SLAM methods. *World Electric Vehicle Journal*, 2024, 15(3): 85. [doi: [10.3390/wevj15030085](https://doi.org/10.3390/wevj15030085)]
- 13 Tosi F, Zhang Y, Gong Z, *et al.* How NeRFs and 3D Gaussian splatting are reshaping SLAM: A survey. *arXiv:2402.13255v2*, 2024.
- 14 Taud H, Mas JF. Multilayer perceptron (MLP). Camacho Olmedo MTC, Paegelow M, Mas JF, *et al.* *Geomatic Approaches for Modeling Land Change Scenarios*. Cham: Springer, 2018. 451–455. [doi: [10.1007/978-3-319-60801-3_27](https://doi.org/10.1007/978-3-319-60801-3_27)]
- 15 Sun SH, Huh M, Liao YH, *et al.* Multi-view to novel view: Synthesizing novel views with self-learned confidence. *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018. 162–178. [doi: [10.1007/978-3-030-01219-9_10](https://doi.org/10.1007/978-3-030-01219-9_10)]
- 16 Novák J, Georgiev I, Hanika J, *et al.* Monte Carlo methods for volumetric light transport simulation. *Computer Graphics Forum*, 2018, 37(2): 551–576. [doi: [10.1111/cgf.13383](https://doi.org/10.1111/cgf.13383)]
- 17 Kato H, Beker D, Morariu M, *et al.* Differentiable rendering: A survey. *arXiv:2006.12057*, 2020.
- 18 Gao RC, Qi Y. A brief review on differentiable rendering: Recent advances and challenges. *Electronics*, 2024, 13(17): 3546. [doi: [10.3390/electronics13173546](https://doi.org/10.3390/electronics13173546)]
- 19 Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with Rao-blackwellized particle filters. *IEEE Transactions on Robotics*, 2007, 23(1): 34–46. [doi: [10.1109/tro.2006.889486](https://doi.org/10.1109/tro.2006.889486)]
- 20 Hess W, Kohler D, Rapp H, *et al.* Real-time loop closure in 2D LIDAR SLAM. *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm: IEEE, 2016. 1271–1278. [doi: [10.1109/icra.2016.7487258](https://doi.org/10.1109/icra.2016.7487258)]
- 21 Shan TX, Englot B. LeGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Madrid: IEEE, 2018. 4758–4765. [doi: [10.1109/iros.2018.8594299](https://doi.org/10.1109/iros.2018.8594299)]
- 22 Davison AJ. Real-time simultaneous localisation and

- mapping with a single camera. Proceedings of the 9th IEEE International Conference on Computer Vision. Nice: IEEE, 2003. 1403–1410. [doi: [10.1109/iccv.2003.1238654](https://doi.org/10.1109/iccv.2003.1238654)]
- 23 Mur-Artal R, Tardos JD. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras. IEEE Transactions on Robotics, 2017, 33(5): 1255–1262. [doi: [10.1109/tro.2017.2705103](https://doi.org/10.1109/tro.2017.2705103)]
- 24 Campos C, Elvira R, Rodríguez JGG, *et al.* ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. IEEE Transactions on Robotics, 2021, 37(6): 1874–1890. [doi: [10.1109/tro.2021.3075644](https://doi.org/10.1109/tro.2021.3075644)]
- 25 Remondino F, Karami A, Yan ZY, *et al.* A critical analysis of NeRF-based 3D reconstruction. Remote Sensing, 2023, 15(14): 3585. [doi: [10.3390/rs15143585](https://doi.org/10.3390/rs15143585)]
- 26 Gao K, Gao YN, He HJ, *et al.* NeRF: Neural radiance field in 3D vision, a comprehensive review. arXiv:2210.00379, 2023.
- 27 Chung CM, Tseng YC, Hsu YC, *et al.* Orbeez-SLAM: A real-time monocular visual SLAM with ORB features and NeRF-realized mapping. Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA). London: IEEE, 2023. 9400–9406. [doi: [10.1109/icra48891.2023.10160950](https://doi.org/10.1109/icra48891.2023.10160950)]
- 28 Zhang YM, Tosi F, Mattoccia S, *et al.* GO-SLAM: Global optimization for consistent 3D instant reconstruction. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023. 3704–3714. [doi: [10.1109/iccv51070.2023.00345](https://doi.org/10.1109/iccv51070.2023.00345)]
- 29 Sucar E, Liu SK, Ortiz J, *et al.* iMAP: Implicit mapping and positioning in real-time. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021. 6209–6218. [doi: [10.1109/iccv48922.2021.00617](https://doi.org/10.1109/iccv48922.2021.00617)]
- 30 Zhu ZH, Peng SY, Larsson V, *et al.* NICE-SLAM: Neural implicit scalable encoding for SLAM. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022. 12776–12786. [doi: [10.1109/cvpr52688.2022.01245](https://doi.org/10.1109/cvpr52688.2022.01245)]
- 31 Rosinol A, Leonard JJ, Carlone L. NeRF-SLAM: Real-time dense monocular SLAM with neural radiance fields. Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Detroit: IEEE, 2023. 3437–3444. [doi: [10.1109/iros55552.2023.10341922](https://doi.org/10.1109/iros55552.2023.10341922)]
- 32 Kong X, Liu SK, Taher M, *et al.* vMAP: Vectorised object mapping for neural field SLAM. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023. 952–961. [doi: [10.1109/cvpr52729.2023.00098](https://doi.org/10.1109/cvpr52729.2023.00098)]
- 33 Li MR, He JM, Wang YY, *et al.* End-to-end RGB-D SLAM with multi-MLPs dense neural implicit representations. IEEE Robotics and Automation Letters, 2023, 8(11): 7138–7145. [doi: [10.1109/lra.2023.3311365](https://doi.org/10.1109/lra.2023.3311365)]
- 34 Zhu ZH, Peng SY, Larsson V, *et al.* NICER-SLAM: Neural implicit scene encoding for RGB SLAM. Proceedings of the 2024 International Conference on 3D Vision (3DV). Davos: IEEE, 2024. 42–52. [doi: [10.1109/3dv62453.2024.00096](https://doi.org/10.1109/3dv62453.2024.00096)]
- 35 Zhu ST, Wang GM, Blum H, *et al.* SNI-SLAM: Semantic neural implicit SLAM. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024. 21167–21177. [doi: [10.1109/cvpr52733.2024.02000](https://doi.org/10.1109/cvpr52733.2024.02000)]
- 36 Deng TC, Wang YB, Xie HL, *et al.* NeSLAM: Neural implicit mapping and self-supervised feature tracking with depth completion and denoising. IEEE Transactions on Automation Science and Engineering, 2024. [doi: [10.1109/TASE.2025.3541064](https://doi.org/10.1109/TASE.2025.3541064)]
- 37 Isaacson S, Kung PC, Ramanagopal M, *et al.* LONER: LiDAR only neural representations for real-time SLAM. IEEE Robotics and Automation Letters, 2023, 8(12): 8042–8049. [doi: [10.1109/LRA.2023.3324521](https://doi.org/10.1109/LRA.2023.3324521)]
- 38 Yang XR, Li H, Zhai HJ, *et al.* Vox-Fusion: Dense tracking and mapping with voxel-based neural implicit representation. Proceedings of the 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Singapore: IEEE, 2022. 499–507. [doi: [10.1109/ismar55827.2022.00066](https://doi.org/10.1109/ismar55827.2022.00066)]
- 39 Zhong XG, Pan Y, Behley J, *et al.* SHINE-Mapping: Large-scale 3D mapping using sparse hierarchical implicit neural representations. Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA). London: IEEE, 2023. 8371–8377. [doi: [10.1109/icra48891.2023.10160907](https://doi.org/10.1109/icra48891.2023.10160907)]
- 40 Mao YX, Yu X, Zhang ZQ, *et al.* NGEL-SLAM: Neural implicit representation-based global consistent low-latency SLAM system. Proceedings of the 2024 IEEE International Conference on Robotics and Automation (ICRA). Yokohama: IEEE, 2024. 6952–6958. [doi: [10.1109/icra57147.2024.10611269](https://doi.org/10.1109/icra57147.2024.10611269)]
- 41 Wang HY, Wang JW, Agapito L. Co-SLAM: Joint coordinate and sparse parametric encodings for neural real-time SLAM. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023. 13293–13302. [doi: [10.1109/cvpr52729.2023.01277](https://doi.org/10.1109/cvpr52729.2023.01277)]

- 42 Matsuki H, Tateno K, Niemeyer M, *et al.* NEWTON: Neural view-centric mapping for on-the-fly large-scale SLAM. *IEEE Robotics and Automation Letters*, 2024, 9(4): 3704–3711. [doi: [10.1109/lra.2024.3367277](https://doi.org/10.1109/lra.2024.3367277)]
- 43 Zhou H, Guo ZT, Liu SH, *et al.* MoD-SLAM: Monocular dense mapping for unbounded 3D scene reconstruction. *arXiv:2402.03762*, 2024.
- 44 Johari MM, Carta C, Fleuret F. ESLAM: Efficient dense SLAM system based on hybrid representation of signed distance fields. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver: IEEE, 2023. 17408–17419. [doi: [10.1109/cvpr52729.2023.01670](https://doi.org/10.1109/cvpr52729.2023.01670)]
- 45 Zhang ZY, Zhang YZ, Wu YM, *et al.* S3-SLAM: Sparse tri-plane encoding for neural implicit SLAM. *arXiv:2404.18284*, 2024.
- 46 Deng TC, Shen GL, Qin T, *et al.* PLGSLAM: Progressive neural scene representation with local to global bundle adjustment. *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2024. 19657–19666. [doi: [10.1109/cvpr52733.2024.01859](https://doi.org/10.1109/cvpr52733.2024.01859)]
- 47 Sandström E, Li Y, Van Gool L, *et al.* Point-SLAM: Dense neural point cloud-based SLAM. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2023. 18387–18398. [doi: [10.1109/iccv51070.2023.01690](https://doi.org/10.1109/iccv51070.2023.01690)]
- 48 Liso L, Sandstrom E, Yugay V, *et al.* Loopy-SLAM: Dense neural SLAM with loop closures. *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2024. 20363–20373. [doi: [10.1109/cvpr52733.2024.01925](https://doi.org/10.1109/cvpr52733.2024.01925)]
- 49 Pan Y, Zhong XG, Wiesmann L, *et al.* PIN-SLAM: LiDAR SLAM using a point-based implicit neural representation for achieving global map consistency. *IEEE Transactions on Robotics*, 2024, 40: 4045–4064. [doi: [10.1109/tro.2024.3422055](https://doi.org/10.1109/tro.2024.3422055)]
- 50 Zhang GL, Sandström E, Zhang YM, *et al.* GIORIE-SLAM: Globally optimized RGB-only implicit encoding point cloud SLAM. *arXiv:2403.19549*, 2024.
- 51 Hu JR, Mao M, Zhang GJ, *et al.* CP-SLAM: Collaborative neural point-based SLAM. *Proceedings of the 37th International Conference on Neural Information Processing Systems*. New Orleans: ACM, 2023. 1712.
- 52 Peng CS, Xu CF, Wang Y, *et al.* Q-SLAM: Quadric representations for monocular SLAM. *arXiv:2403.08125*, 2024.
- 53 Müller T, Evans A, Schied C, *et al.* Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 2022, 41(4): 102. [doi: [10.1145/3528223.3530127](https://doi.org/10.1145/3528223.3530127)]
- 54 Xia C, Xu CF, Rim P, *et al.* Quadric representations for LiDAR odometry, mapping and localization. *IEEE Robotics and Automation Letters*, 2023, 8(8): 5023–5030. [doi: [10.1109/lra.2023.3290510](https://doi.org/10.1109/lra.2023.3290510)]
- 55 Kruzhkov E, Savinykh A, Karpyshev P, *et al.* MeSLAM: Memory efficient SLAM based on neural fields. *Proceedings of the 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Prague: IEEE, 2022. 430–435. [doi: [10.1109/smc53654.2022.9945381](https://doi.org/10.1109/smc53654.2022.9945381)]
- 56 Li H, Gu XD, Yuan WH, *et al.* Dense RGB slam with neural implicit maps. *Proceedings of the 11th International Conference on Learning Representations (ICLR)*. Kigali: OpenReview.net, 2023.
- 57 Sandström E, Ta K, Van Gool L, *et al.* UncLe-SLAM: Uncertainty learning for dense neural SLAM. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. Paris: IEEE, 2023. 4539–4550. [doi: [10.1109/iccvw60793.2023.00488](https://doi.org/10.1109/iccvw60793.2023.00488)]
- 58 Wu XM, Liu ZM, Tian YX, *et al.* KN-SLAM: Keypoints and neural implicit encoding SLAM. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 2512712. [doi: [10.1109/tim.2024.3378264](https://doi.org/10.1109/tim.2024.3378264)]
- 59 Ming YH, Ye WC, Calway A. iDF-SLAM: End-to-end RGB-D SLAM with neural implicit mapping and deep feature tracking. *arXiv:2209.07919*, 2022.
- 60 Zhai HJ, Li H, Yang XR, *et al.* Vox-Fusion++: Voxel-based neural implicit dense tracking and mapping with multi-maps. *arXiv:2403.12536*, 2024.
- 61 Xiang BC, Sun YX, Xie ZQ, *et al.* NISB-map: Scalable mapping with neural implicit spatial block. *IEEE Robotics and Automation Letters*, 2023, 8(8): 4761–4768. [doi: [10.1109/lra.2023.3289152](https://doi.org/10.1109/lra.2023.3289152)]
- 62 Deng JY, Wu Q, Chen XYL, *et al.* NeRF-LOAM: Neural implicit representation for large-scale incremental LiDAR odometry and mapping. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2023. 8184–8193. [doi: [10.1109/iccv51070.2023.00755](https://doi.org/10.1109/iccv51070.2023.00755)]
- 63 Zhang W, Sun TC, Wang S, *et al.* HI-SLAM: Monocular real-time dense mapping with hybrid implicit fields. *IEEE Robotics and Automation Letters*, 2024, 9(2): 1548–1555.

- [doi: [10.1109/lra.2023.3347131](https://doi.org/10.1109/lra.2023.3347131)]
- 64 Tang YJ, Zhang JZ, Yu ZN, *et al.* MIPS-Fusion: Multi-implicit-submaps for scalable and robust online neural RGB-D reconstruction. *ACM Transactions on Graphics (TOG)*, 2023, 42(6): 246. [doi: [10.1145/3618363](https://doi.org/10.1145/3618363)]
- 65 Yu X, Liu YL, Mao ST, *et al.* NF-Atlas: Multi-volume neural feature fields for large scale LiDAR mapping. *IEEE Robotics and Automation Letters*, 2023, 8(9): 5870–5877. [doi: [10.1109/lra.2023.3300281](https://doi.org/10.1109/lra.2023.3300281)]
- 66 Teed Z, Deng J. DROID-SLAM: Deep visual SLAM for monocular, stereo, and RGB-D cameras. *Proceedings of the 35th International Conference on Neural Information Processing Systems*. ACM, 2021. 1266.
- 67 Li GH, Chen Q, Yan YX, *et al.* EC-SLAM: Real-time dense neural RGB-D SLAM system with effectively constrained global bundle adjustment. *arXiv:2404.13346*, 2024.
- 68 Yan YF, He RM, Liu ZH. MUTE-SLAM: Real-time neural SLAM with multiple tri-plane hash representations. *arXiv:2403.17765*, 2024.
- 69 Ruan CY, Zang QY, Zhang KH, *et al.* DN-SLAM: A visual SLAM with ORB features and NeRF mapping in dynamic environments. *IEEE Sensors Journal*, 2024, 24(4): 5279–5287. [doi: [10.1109/jsen.2023.3345877](https://doi.org/10.1109/jsen.2023.3345877)]
- 70 Kirillov A, Mintun E, Ravi N, *et al.* Segment anything. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2023. 3992–4003. [doi: [10.1109/ICCV51070.2023.00371](https://doi.org/10.1109/ICCV51070.2023.00371)]
- 71 Xu ZH, Niu JW, Li QF, *et al.* NID-SLAM: Neural implicit representation-based RGB-D SLAM in dynamic environments. *Proceedings of the 2024 IEEE International Conference on Multimedia and Expo (ICME)*. Niagara Falls: IEEE, 2024. 1–6. [doi: [10.1109/ICME57554.2024.10687512](https://doi.org/10.1109/ICME57554.2024.10687512)]
- 72 Li MR, Zhou YM, Jiang GA, *et al.* DDN-SLAM: Real-time dense dynamic neural implicit SLAM. *arXiv:2401.01545*, 2024.
- 73 Jiang HC, Xu YM, Li KJ, *et al.* RoDyn-SLAM: Robust dynamic dense RGB-D SLAM with neural radiance fields. *IEEE Robotics and Automation Letters*, 2024, 9(9): 7509–7516. [doi: [10.1109/lra.2024.3427554](https://doi.org/10.1109/lra.2024.3427554)]
- 74 Jain J, Li JC, Chiu M, *et al.* OneFormer: One Transformer to rule universal image segmentation. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver: IEEE, 2023. 2989–2998. [doi: [10.1109/cvpr52729.2023.00292](https://doi.org/10.1109/cvpr52729.2023.00292)]
- 75 Schischka N, Schieber H, Karaoglu MA, *et al.* DynaMoN: Motion-aware fast and robust camera localization for dynamic neural radiance fields. *IEEE Robotics and Automation Letters*, 2025, 10(1): 548–555. [doi: [10.1109/LRA.2024.3511399](https://doi.org/10.1109/LRA.2024.3511399)]
- 76 Chen LC, Papandreou G, Schroff F, *et al.* Rethinking atrous convolution for semantic image segmentation. *arXiv:1706.05587*, 2017.
- 77 Qu DL, Yan C, Wang D, *et al.* Implicit event-RGBD neural SLAM. *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2024. 19584–19594. [doi: [10.1109/cvpr52733.2024.01852](https://doi.org/10.1109/cvpr52733.2024.01852)]
- (校对责编: 张重毅)