

Research on SLAM Algorithms from Traditional to Deep Learning Based

Jingtian Xu Ting Yang*

Xuzhou Medical University, Xuzhou, Jiangsu, 221000, China

Abstract

This paper systematically reviews the research progress of SLAM algorithms from traditional methods to those based on deep learning. It first introduces the basic concepts and system structure of SLAM algorithms, including core modules such as front-end processing, back-end optimization, and loop closure detection. Then, it analyzes the characteristics of traditional SLAM systems, such as visual SLAM and laser SLAM, and points out their low adaptability in dynamic environments and low-texture areas. On this basis, it focuses on discussing the integrated application of deep learning techniques, such as convolutional neural networks (CNN), Transformer models, and recurrent neural networks (RNN), in feature extraction, pose estimation, and semantic understanding in SLAM technology. It summarizes the advantages and disadvantages of representative algorithms such as LIFT-SLAM, DeepVo, DLD-SLAM, and NeRF-SLAM. Finally, it points out that deep learning technology can significantly enhance the adaptability of SLAM systems in complex environments, providing a reference for the further development of SLAM technology.

Keywords

SLAM; deep learning; deep neural networks

从传统到基于深度学习的 SLAM 算法研究

徐靖添 杨婷*

徐州医科大学, 中国·江苏 徐州 221000

摘要

本文系统梳理了从传统SLAM到基于深度学习的SLAM算法的研究进展。首先介绍了SLAM算法的基本概念与系统结构,包括前端处理、后端优化、回环检测等核心模块。随后分析了传统SLAM系统如视觉SLAM和激光SLAM的特点并指出其在动态环境和低纹理的低适应性。在此基础上,重点探讨了深度学习技术如卷积神经网络(CNN)、Transformer模型、循环神经网络(RNN)等在SLAM技术中特征提取、位姿估计及语义理解等方面的融合应用。总结了LIFT-SLAM、DeepVo、DLD-SLAM、NeRF-SLAM等代表性算法的优势与不足。最后指出深度学习技术可显著提升SLAM系统在复杂环境下的适应性,为SLAM技术的进一步发展提供参考。

关键词

SLAM; 深度学习; 深度神经网络

1 引言

SLAM最早由伦纳德和杜兰特·怀特于二十世纪90年代提出,早期SLAM研究集中在扩展卡尔曼滤波和图优化方法,初步构成SLAM的理论框架,但这些传统SLAM技术往往适用于纹理丰富清晰的环境,应用的范围相对有限。

在二十一世纪,以相机为基础传感器的视觉SLAM取得极大进展,其代表算法ORB-SLAM和PL-SVO通过特征匹配可实现较高鲁棒性和实时性。激光雷达SLAM的出现

则进一步丰富了SLAM应用。

而在二十一世纪的第二个十年,随着深度学习的发展,深度学习技术如卷积神经网络(convolutional neural network,CNN)和循环神经网络(recurrent neural network,RNN)已经成功应用于SLAM系统中特征提取、位姿估计、深度估计和图像处理等各个环节,提高SLAM系统在纹理稀疏和复杂,动态环境中的鲁棒性。

同步定位和地图构建(Simultaneous Localization and Mapping, SLAM),作为核心技术广泛应用于机器人,自动驾驶,AR/VR,旨在让设备在未知环境中实时确定更新自身位置和构建地图。SLAM的核心包含两部分。定位(Localization),通过跟踪传感器确定设备在环境中当前的位置和姿态(如坐标,旋转角度)。构建地图(Mapping),根据传感器捕获数据构建当前环境地图,传统SLAM多采

【作者简介】徐靖添(2005-),男,中国江苏苏州人,本科,从事传统到基于深度学习的SLAM研究。

【通讯作者】杨婷(1985-),女,中国山东日照人,博士,讲师,从事路径规划算法研究。

用 2D 栅格地图和 3D 点云地图。

2 SLAM 系统结构

SLAM 总体可以分为五个部分：数据输入，前端处理（Front-End），后端优化（Back-End），闭环检测（LCD）和地图构建（Mapping）。SLAM 系统运行的第一个步骤，便是传感器采集信息后传输给前端里程计进行对数据初步分析处理，进行特征点提取以计算相邻帧间的相对位姿并生成一个图表，这一环节中所得的数据相对粗糙，需在后端优

化中进行进一步处理，通过全局优化算法消除在前端里程中产生的累计误差，得到精确的位姿，进而进行构建全局一致的地图，而闭环检测系统则负责在以上数个步骤过程中识别已经访问过的节点，修正累计误差使机器人获得识别相同场景的能力。在此将 SLAM 大体分为两个部分进行粗略介绍。

2.1 前端处理

核心任务是传感器输入数据预处理、提取环境特征和计算相对位姿。

表 1-1 前端处理各环节

关键环节	主要内容	常用方法
数据预处理	图像（点云）去畸变，点云去噪等	迭代最近点（ICP）
特征提取和匹配	从预处理数据中提取特征并进行匹配	提取：ORB, CNN 匹配：暴力匹配，迭代最近点（ICP）
相对位姿估计	基于特征匹配结果计算相邻时刻位姿	视觉里程计（VO），激光里程计（LO）

2.2 后端处理

核心任务为对前端处理获取的输出位姿和地图信息进行优化，相称累积误差从而构建全局一致性地图。

V-SLAM 方案和基于直接法的 V-SLAM 方案。

3 传统 SLAM

3.1 视觉 SLAM

视觉 SLAM：相较于其他 SLAM，V-SLAM 直接使用原始图像作为传感器输入数据。大致可分为基于特征的

3.2 激光 SLAM

激光 SLAM 主要分为 2D 激光 SLAM 和 3D 激光 SLAM。激光雷达根据激光线数，可分为单线激光雷达和多线激光雷达两种类型。在视觉 SLAM 难以正常工作的情况下。激光雷达可直接测量距离，获取更为精确的环境信息。进而在构建高精度地图的同时获得精确位姿，并在长时间运行情况下有较强鲁棒性。如图 1-1。

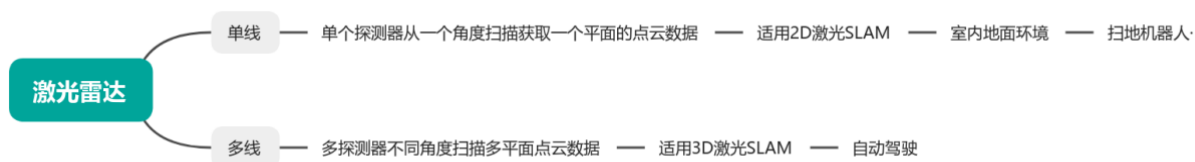


图 1-1 激光雷达分类

3.3 语义 SLAM

语义 SLAM 指在传统 SLAM 基础上，为其增加理解语义信息的技术。除 SLAM 的基础能力外，还拥有对当前环境中物体进行分类的能力，使机器人或者智能设备对环境有更深入的了解。

方法结合的关键。目前广泛应用于医学，机器人，自动驾驶等领域，CNN 是大多数神经网络的主要核心。当前深度学习与 SLAM 结合发展的主要发展趋势是深度学习与传统 SLAM 方法相结合。借助深度学习神经网络在特征提取和匹配、图像感知、位姿估计等方面的优越性与 SLAM 系统的同步定位和地图构建任务相结合，利用 SLAM 系统各环节算法构建神经网络，显著提高其精度、效率和鲁棒性，在一定程度上实现自主同步定位和地图构建。

4 基于深度学习 SLAM

4.1 深度学习概念

深度学习（DL）是机器学习的一个重要分支，核心是通过模拟人脑神经网络结构的算法模型，从数据中自动学习特征和规律，实现对复杂任务的处理（如图像识别、自然语言理解等）。

4.3 深度学习与视觉 SLAM

4.2 深度学习结合 SLAM 方法

深度学习的核心部分是神经网络如卷积神经网络（CNN）和循环神经网络（RNN）也是深度学习与 SLAM

4.3.1 前端处理

文献^[1]中 LIFT-SLAM 即一种将深度学习用于 V-SLAM 特征提取的基于 LIFT 网络混合 V-SLAM 算法，在 ORB-SLAM 系统后端进行特征提取。LIFT-SLAM 以监督端到端的方法实现局部特征检测、方向估计等。其适用于图像补丁，优点在于传统 V-SLAM 不利的条件下（弱光照，无特征环境）

拥有强鲁棒性,高精度特征提取和匹配及增量图优化效率。不足之处则在于大型环境中性能取决于闭环检测,如无法识别循环,则误差无限增加,因为漂移累积不能通过位姿图优化进行校正,及速度与其他先进算法相比较慢。DeepVo 是最引人注目的端到端方法,文献^[2]中,DeepVo 将循环神经网络(RNN)和卷积神经网络(CNN)相结合提取图像特征。RNN 通过 CNN 的学习特征估计相对位姿,同时使用两个堆叠的长短期记忆层(LSTM)估计相机位姿变化,全程由神经网络自主完成。优点在于自主神经网络所给予的优秀位姿估计能力以及对动态环境的适应性和无需手工设计特征便利性,缺点则是新环境的不适应性。文献^[3]中提出轻量级卷积神经网络和循环神经网络结合的一种使用深度特征提取器、深度特征匹配器和轻量级深度闭环检测器的 V-SLAM。主要思想是将深度特征提取器、深度特征匹配器和深度闭环检测器与传统 SLAM 相结合。其中深度特征提取器提供稳定特征点,深度特征匹配器则提供位姿估计所需的连续正确匹配点对。

4.3.2 闭环检测

闭环检测(LCD)可帮助 SLAM 在长期运行中消除累积误差,早期应用词袋模型手工提取特征点,而嵌入神经网络的 LCD 可表现出较传统 SLAM 方法鲁棒性更强且无需手工标注。因此在文献^[4]中提出的将深度学习与视觉 SLAM 中的闭环检测部分相结合提出的基于 MobileNet V2(ECMobileNet)的新型轻量级神经网络,可有效减少深度神经网络复杂性。原理是将高效通道注意力(ECA)插入压缩 ECMobileNet 中,通过减少作数的方法降低神经网络复杂性,较现有神经网络方法更加高效和轻量。ECMobileNet 可在确保 LCD 长期实时运行的同时减少神经网络复杂度。在 GPU 上 TensorFlow 实验已验证 ECMobileNet 优于现有神经网络方法,但在 CPU 等微处理器上效果仍有待验证。

4.3.3 后端处理

文献^[5]中提出 DLD-SLAM 以解决动态环境下的同步定位问题。DLD-SLAM 基于 RGB-D 相机的传统 ORB-SLAM 算法,额外增添一语义分割区域来解决动态区域问题,提高动态环境下 SLAM 系统的精度和运行效率。创新性采用采用 GCNv2-tiny 深度神经网络取代传统的 ORB-SLAM 的特征提取和匹配方法可稳定提取分布均匀的特征点。同时,额外增加的语义分割线程中结合 GSConv 卷积增强 YOLOv5s 目标检测网络,减少深度神经网络的复杂度,达到高效率和高精度及轻量化。DLD-SLAM 较传统 ORB-SLAM 在动态环境下的精度,效率和鲁棒性都显著提高。文献^[6]中,将实例分割算法 MASK R-CNN 与经典目标跟踪算法 Deepsort 相结合以提取检测目标 2D 语义信息并构建图像联系。后使用 RGB-D 版本开源 SLAM 系统 ORB-SLAM2 进行稀疏跟踪,重建目标语义图、目标对象及优化位姿。提出了一种构建面向对象的语义图方法,将跟踪目标与实例分开。TUM

数据集测试验证该方法较传统 SLAM 方法有较高精度,高定位性能,同时高精度分割环境中物体的点云模型。缺点是难满足实时性要求,忽略利用数据关联性提高语义信息的准确性。

4.4 深度学习与激光雷达 SLAM

4.4.1 点云特征提取及配准

Hu 等人^[7]提出基于学习的特征提取方法 SpinNet,由空间点转换器及三维卷积神经网络的特征提取器组成,具有旋转不变性。Li 等人^[8]提出基于深度学习网络的端到端的点云配准方法 Leopard,引入 Transformer 注意力机制,缺点则是缺少精度。

4.4.2 回环检测

在文献^[9]中提出的 OverlapNet,利用深度神经网络,凭借激光雷达扫描生成信息提供提供成对三维扫描间重叠及相对偏航角估计,适用于激光雷达 SLAM 的闭环检测环节。无需通过相邻帧间位姿估计,而是通过成对 3D LiDar 间图像重叠及偏航角估计实现闭环检测要求。该方法使用的激光雷达扫描获得信息包括深度,强度,法线及缓解值,引入语义信息。经 KITTI 数据集测试该方法具有较高的精度及鲁棒性。缺点则是不同环境下精度难以保证。

4.4.3 深度神经网络与 SLAM

神经辐射场(NeRF),一种新兴三维隐性神经表达方法,可以通过神经网络学习场景连续体积、密度及颜色。文献^[10]中,将 NeRF 与传统 SLAM 结合,通过深度学习的学习能力满足 SLAM 系统对场景构建及实时定位需求,实现三维连续场景构图,同时利用反向传播算法对位姿估计及稠密地图构建进行进一步优化,缺点则是实时性难以保证,仅在静态环境下有较高精度,另需要提高全局一致性。

4.4.4 深度学习与语义 SLAM

语义 SLAM 即将语义信息与传统 SLAM 相结合,其核心目的是对对象的识别和检测。在文献^[11]中 Yang 等人提出的 Cube SLAM 是一种联合估计位姿及动态轨迹 SLAM 方法,核心思路是将静态与动态物体区分估计,将静态物体提取特征与 2D 物体相关联,动态物体则是通过三角化测量,直接使用系数灌注算法跟踪像素得到三维位置特征点。重要意义在于首次证明语义对象检测可以和几何 SLAM 在同一框架内相结合受益,该方法在 SUNRGBD 数据集检测三维目标及 KITTI 数据集上测试相机位姿估计得出最佳精度。

文献^[12]提出的 QuadricSLAM 结合了最先进的物体检测的技术和 SLAM 方法,其核心思路是通过将二次参数化为闭合曲线表示物体,推导出基于因子图 SLAM 公式。经过 TUM RGB 数据集测试得出,较视觉里程计,轨迹质量提升,较 ORB-SLAM2,性能略有不足。

5 结语

本文着重探讨了深度学习与传统 SLAM 的相融合,为

SLAM带来了突破性的进展,从SLAM的特征提取匹配到回环检测,地图构建。深度学习方法凭借其强悍的学习能力有效弥补了传统SLAM在如动态复杂环境,稀疏纹理环境,低光度环境下低适应性。SLAM技术从传统到基于深度学习的演进,不仅是对算法框架及技术的革新,更是将人类新兴技术与传统相结合的思路及方法的重新塑造。

参考文献

- [1] Bruno H M S,Colombini E L.LIFT-SLAM:a deep-learning feature-based monocular visual SLAM method[J]. Neurocomputing,2021,455:97-100.
- [2] Wang S,Clark R,Wen H,et al.DeepVO:Towards End-to-End Visual Odometry with Deep Recurrent Convolutional Neural Networks[J]2017 IEEE International Conference on Robotics and Automation(ICRA),Singapore,2017,2043-2050.
- [3] Zhu B,Yu A,Hou B,et al.A Novel Visual SLAM Based on Multiple Deep Neural Networks[J].Applied Sciences,2023,13(17):9630
- [4] Zhou D,Luo Y,Zhang Q,et al.A Lightweight Neural Network for Loop Closure Detection in Indoor Visual SLAM[J].International Journal of Computational Intelligence Systems,2023,16(1):49.
- [5] Yu H,Wang Q,Yan C,et al.DLD-SLAM:RGB-D Visual Simultaneous Localisation and Mapping in Indoor Dynamic Environments Based on Deep Learning[J].Remote Sensing,2024,16(2):246.
- [6] Sun Y,Hu J,Yun J,et al.Multi-Objective Location and Mapping Based on Deep Learning and Visual Slam[J]. Sensors,2022,22(19):7576.
- [7] Ao S,Hu Q,Yang B,et al.SpinNet:Learning a General Surface Descriptor for 3D Point Cloud Registration[Z].//2021IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Nashville,TN,USA,2021:11748-11757.
- [8] Li Y,Harada T.Lepard:Learning partial point cloud matching in rigid and deformable scenes[Z].//2022 IEEE/CVF Conference onComputerVision and PatternRecognition(CVPR),NewOrleans,L A,USA,2022:5544-5554.
- [9] Chen X,Läbe T,Milioto A,et al.OverlapNet:A siamese network for computing LiDAR scan similarity with applications to loop closing and Localization[J].Autonomous Robots,2021,46(1):61-81.
- [10] 喻伟东,鲁静,程哈蕾.基于NeRF的SLAM研究综述[J].计算机系统应用,2025,34(04):18-33.
- [11] Yang S,Scherer S.CubeSLAM:Monocular 3-D Object SLAM[J]. IEEE Transactions on Robotics,2019,35(4):925-938.
- [12] Nicholson L,Milford M,Sünderhauf N.QuadricSLAM: Dual Quadrics from Object Detections as Landmarks in Object-oriented SLAM[Z].IEEE Robotics and Automation Letters,2019,4(1):1-8.