



审查

视觉 SLAM 概述：从传统到语义

陈伟峰^{1,2}
李振雄²尚光涛¹, 吉爱红³, 周成军², 王希阳², 徐崇辉²、胡凯^{2,*}¹ 泉州信息工程学院机械与电子工程学院, 泉州 362000; 002021@nuist.edu.cn² 南京信息工程大学自动化学院, 南京 210044; 20201222014@nuist.edu.cn (GS); 20211257010@nuist.edu.cn (捷克); 20211267006@nuist.edu.cn (XW); 20211249101@nuist.edu.cn (CX); 20211257005@nuist.edu.cn (ZL)³ 南京航空航天大学机电工程学院运动仿生与智能机器人实验室, 南京 210016; meeahji@nuaa.edu.cn

* 通讯地址: 001600@nuist.edu.cn

抽象的: 视觉SLAM (VSLAM) 由于其传感器成本低廉、易于与其他传感器融合、环境信息更丰富等优点而得到迅速发展。传统的基于视觉的SLAM研究已经取得了许多成果，但在充满挑战的环境中可能无法达到预期的结果。深度学习推动了计算机视觉的发展，深度学习与SLAM的结合越来越受到人们的关注。语义信息作为高层环境信息，可以使机器人更好地理解周围环境。本文从传统VSLAM和结合深度学习的语义VSLAM两个方面介绍了VSLAM技术的发展。对于传统的VSLAM，我们详细总结了间接和直接方法的优缺点，并给出了一些经典的VSLAM开源算法。此外，我们重点发展基于深度学习的语义VSLAM。我们从典型的神经网络CNN和RNN入手，详细总结了神经网络对于VSLAM系统的改进。稍后我们重点讨论目标检测和语义分割对于VSLAM语义信息引入的帮助。我们相信，未来智能时代的发展离不开语义技术的帮助。将深度学习引入VSLAM系统中提供语义信息，可以帮助机器人更好地感知周围环境，为人们提供更高层次的帮助。



引用: 陈, W. 尚, G. 吉, A.; 周, C.;
王X; 徐, C.; 李, Z. Hu, K. 视觉 SLAM
概述：从传统到语义。遥感。2022年, 14
, 3010.10.3390/rs14133010

<https://doi.org/>

学术编辑: Fabio Remondino、Radosław Zimroz、Denis Guilhot 和 Vittorio Cannas

收稿日期: 2022-05-29 接

受日期: 2022-06-17 发布

日期: 2022-06-23

出版商注: MDPI 对于已发布地图和机构
隶属关系中的管辖权主张保持中立。



版权: © 2022 作者版权所有。被许可方
MDPI, 瑞士巴塞尔。本文是根据知识共
享署名 (CC BY) 许可证 (<https://creativecommons.org/licenses/by/4.0/>) 的条款和条件分发的开放获取文
章。

关键词: SLAM; 深度学习; 神经网络; 计算机视觉; 语义; 智能时代

一、简介

人们需要移动机器人自己执行一些任务，这就需要机器人能够适应陌生的环境。因此，SLAM^[1]（同步定位和建图）能够在陌生的环境中进行定位和建图，已经成为自主移动机器人的必备能力。自1986年首次提出以来，SLAM引起了众多研究人员的广泛关注，并在机器人、虚拟现实等领域迅速发展。SLAM是指基于位置和地图的自我定位，以及基于自我定位构建增量地图。主要用于解决机器人在未知环境中移动时的定位和地图构建问题^[2]。SLAM作为一项基础技术，早期已应用于移动机器人定位与导航。随着计算机技术（硬件）和人工智能（软件）的发展，机器人研究受到越来越多的关注和投入。许多研究人员致力于让机器人变得更加智能。SLAM被认为是推动移动机器人真正自主的关键^[3]。

有学者根据采用的传感器不同将SLAM分为激光SLAM和视觉SLAM (VSLAM) ^[4]。与VSLAM相比，由于起步较早，激光

国外SLAM研究相对成熟，长期以来一直被认为是移动机器人的首选解决方案。与人眼类似，VSLAM主要利用图像作为环境感知的信息源，比激光SLAM更符合人类的理解，信息量也更多。近年来，基于相机的VSLAM研究引起了研究人员的广泛关注。由于价格便宜、安装方便、环境信息丰富、易于与其他传感器融合等优点，许多基于视觉的SLAM算法应运而生[5]。VSLAM的优点是环境信息更丰富，被认为能够赋予移动机器人更强的感知能力，并在一些特定场景中得到应用。因此，本文重点关注VSLAM并梳理其衍生的算法。基于各类激光雷达的SLAM不在本文讨论范围内。有兴趣的读者可以参考[6–8]和文献中的其他来源。

作为自主机器人导航的解决方案之一，传统的VSLAM本质上是基于图像几何特征的简单环境理解[9]。由于传统的VSLAM仅利用环境的几何特征，如点、线等，来面对这种低级的几何信息，因此可以达到较高的实时性。面对光照、纹理和动态物体的变化普遍存在，这就表现出明显的不足，在位置精度和鲁棒性方面存在缺陷[10]。虽然传统视觉SLAM构建的地图包含了环境中的重要信息，一定程度上满足了机器人的定位需求。对机器人自主导航和避障任务的支持不足。而且无法满足智能机器人与环境、人类的交互需求。[11]。

人们对智能移动机器人的需求日益增加，这对机器人的自主能力和人机交互能力提出了很高的要求。[12]。传统的VSLAM算法可以满足机器人的基本定位和导航需求，但无法完成“帮我关卧室门”、“去厨房给我拿一个苹果”等更高级别的任务。为了实现这样的目标，机器人需要识别场景中物体的信息，找出它们的位置并构建语义地图。借助语义信息，数据关联从传统的像素级升级到对象级。此外，感知几何环境信息被分配语义标签以获得高级语义图。它可以帮助机器人理解自主环境和人机交互[13]。我们相信深度学习的快速发展为将语义信息引入VSLAM提供了一座桥梁。尤其是在语义地图构建方面，将其与VLAM相结合可以使机器人获得对场景的高层次感知和理解。显着提高机器人与环境的交互能力[14]。

2016年，Cadena等人。[15]首先提出将SLAM的发展分为三个阶段。在他们的描述中，我们处于一个稳健感知的阶段，如图1。他们从经典、算法、鲁棒性三个方面描述了不同时代SLAM的侧重点和贡献。参考号[16]总结了2010年到2016年基于视觉的SLAM算法的发展，并提供了帮助初学者的工具包。优素福等人。[17号]讨论了VSLAM的基本框架，并总结了几个数学问题，以帮助读者做出最佳选择。巴夫勒等人。[18]总结了机器人SLAM技术并指出了机器人场景理解的发展趋势。Servieres等人从视觉与视觉惯性的融合出发。[19]回顾和比较了SLAM中重要的方法并总结了新兴的优秀算法。阿扎姆等人。[20]对基于特征的方法进行了全面的研究。他们根据在环境中观察到的视觉特征对审查的方法进行分类。此外，他们还提出了SLAM未来发展可能出现的问题和解决方案。参考号[21]详细介绍了基于单目、双目、RGB-D、视觉惯性融合的SLAM方法，并给出了存在的问题和未来的方向。参考号[22]描述了VSLAM的机遇和挑战。

几何到深度学习，并预测了VSLAM在未来语义时代的发展前景。

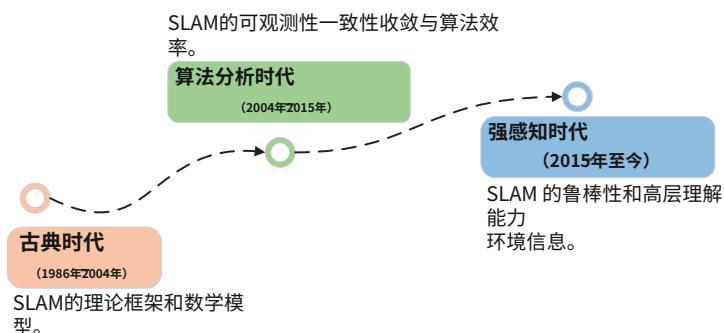


图1。 SLAM发展时代概述。SLAM的发展经历了理论框架、算法分析和高级鲁棒感知三个主要阶段。时间点没有严格限制，而是代表了SLAM在某一阶段的发展以及人们感兴趣的热点问题。

正如你所看到的，有一些基于视觉的SLAM技术的调查和总结。但大多只关注VSLAM的某一方面，而没有对VSLAM的发展进行更全面的总结。此外，上述综述更多地关注传统的视觉SLAM算法，而没有详细介绍与深度学习相结合的语义SLAM。因此，有必要对基于视觉的SLAM算法进行全面的回顾，以帮助研究人员和学生开展视觉SLAM技术的研究，以了解这个大领域的概况。

为了让读者对SLAM领域有更深入、更全面的了解，我们回顾了通用SLAM算法从诞生到现在的历史。此外，我们还总结了推动SLAM解决方案技术演进的关键解决方案。从点问题的形成到最常用的状态方法描述了SLAM的工作。我们不只关注某一方面，

显示所带来的研究之间的联系的关键主要方法

SLAM接近其当前状态。此外，我们回顾了SLAM从传统到语义的演变，这个视角涵盖了历史上主要的、有趣的和领先的设计方法。在此基础上，我们对深度学习SLAM算法进行了全面的总结。还对语义VSLAM进行了详细的讲解，帮助读者更好地理解语义VSLAM的特点。我们认为我们的工作可以帮助读者更好地理解机器人环境感知。我们在语义VSLAM方面的工作可以为读者提供更好的想法，并为未来SLAM研究甚至机器人自主感知提供有益的参考。因此，本文对基于视觉的SLAM技术的发展进行了全面的补充和更新。此外，

本文将基于视觉的SLAM的发展分为两个阶段：传统VSLAM和融合深度学习的语义VSLAM。使读者能够更好地了解VSLAM的研究热点，把握VSLAM的发展方向。我们认为传统的阶段SLAM问题主要解决的是框架问题

算法。在语义时代，SLAM注重与深度学习相结合的高级态势感知和系统鲁棒性。

我们的审查对现有技术做出了以下贡献：

- 我们比较全面地回顾了基于视觉的SLAM的发展，回顾了基于环境语义信息的同步定位和地图构建领域的最新研究进展。
- 我们从卷积神经网络（CNN）和循环神经网络（RNN）开始，详细描述深度学习在VSLAM中的应用。据我们所知，这是第一篇从神经网络角度介绍VSLAM的综述。

- 我们详细描述了语义信息与VSLAM的结合，并指出了语义时代VSLAM的发展方向。主要介绍和总结了语义信息与传统视觉SLAM结合在系统定位和地图构建方面的优秀研究成果，并对传统视觉SLAM和语义SLAM进行了深入的比较。最后提出了语义SLAM的未来研究方向。

具体来说，在第1本文详细介绍了传统VSLAM的特点，包括直接法和基于前端视觉里程计的间接法，并对基于深度相机的VSLAM和集成IMU的经典VSLAM进行了比较。在部分2，本文分为两部分。我们首先从CNN和RNN这两个神经网络介绍深度学习和VSLAM的结合。我们认为，将深度学习引入语义VSLAM是语义VSLAM发展的前提。此外，这个阶段也可以被视为语义VSLAM的开始。然后，本文从目标检测和语义分割方面描述了深度学习引领语义VSLAM进入高级阶段的过程。所以本文总结了发展

从定位、建图、消除三个方面探讨语义VSLAM方向的动态对象。在部分3，本文介绍了一些主流的SLAM数据集，以及该领域的一些优秀实验室。最后，我们总结了当前的研究并指出了VSLAM未来的研究方向。的截面表为

本文内容如图2。

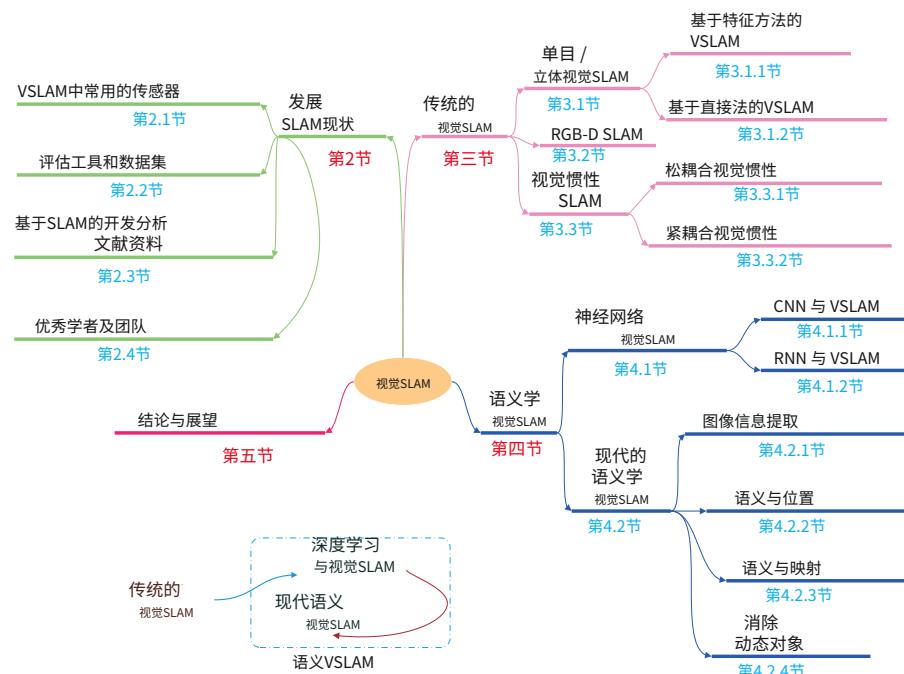


图 2.本文其余部分的结构图。本文重点研究第二章语义VSLAM。我们认为神经网络的引入是语义VSLAM的开始。我们从深度神经网络入手，描述其与VSLAM的结合，然后从对象检测和语义分割方面详细解释现代语义VSLAM

基于深度学习，并做出总结和展望。

2.SLAM的发展现状

2.1. VSLAM中常用的传感器

VSLAM中使用的传感器通常包括单目相机、立体相机和RGB-D相机。单目摄像头和立体摄像头原理相似，可用于广泛的室内外环境。RGB-D相机作为相机的一种特殊形式，主要通过主动发射来直接获取图像深度

红外结构光或计算飞行时间 (TOF)。使用方便，但对光线敏感，大多数情况下只能在室内使用[23]。赛事相机作为近年来出现的一种新型相机传感器，拍摄出的照片有别于传统相机。事件相机就是“事件”，可以简单到“像素亮度变化”。事件相机输出的变化是像素亮度，基于事件相机的SLAM算法还只处于初步研究阶段[24]。此外，作为基于视觉的经典SLAM系统，视觉惯性融合在很多方面都取得了优异的成果。图中3，我们比较了不同相机的主要特点。

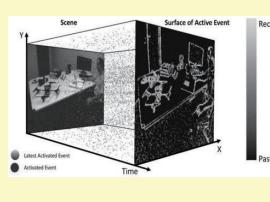
相机	优势	坏处
 单目	结构简单，成本低，可 室内和室外使用。	仅靠图像不能 确定这个真实的比例。
 立体声	可测量的距离越远；可以用 室内和室外。	视差计算非常消耗资 源。
 RGB-D	可以提供更丰富的信息， 并且也不需要那么耗时或需 要双眼 深度计算。	测量范围窄、噪声大、 视场小 视力，易受影响 阳光干扰。
 事件	事件相机具有低延迟、高动态 范围 (HDR)、无运动模糊、 极低功耗和低数据量等优点 带宽。	单一事件的有效信 息很少 和稀疏和 数据不完整。

图 3. 不同相机之间的比较。事件摄像机并不是特定类型的摄像机，而是能够获取“事件信息”的摄像机。“传统相机”工作频率恒定，在拍摄高速物体时存在滞后、模糊、曝光过度等天然缺陷。然而，事件相机是一种类似于人眼的基于神经的信息处理方法，不存在这些问题。

2.2. 评估工具和数据集

SLAM 问题已经存在了几十年。在过去的几十年里，涌现了许多优秀的算法，尽管各自的侧重点不同，但它们都不同程度地促进了SLAM技术的快速发展。每个算法都必须公平地进行比较。一般来说，我们可以从耗时、复杂度、准确度等多个角度来评价一个SLAM算法。不过，最重要的一点是我们最注重它的准确性。ATE（绝对轨迹误差）和RPE（相对位姿误差）是用于评估SLAM精度的两个最重要的指标。相对位姿误差用于计算相同两个时间戳内位姿变化的差异，适合估计系统漂移。绝对轨迹误差直接计算相机位姿的真实值与SLAM系统的估计值之间的差异。ATE和RPE的基本原理如下。

假设：给定的姿态估计是 Δ 。下标表示时间t（或帧），其中假设估计姿势和真实姿势的每一帧的时间对齐，并且总帧数相同。

ATE：绝对轨迹误差是估计位姿与真实位姿的直接差异，可以直接反映算法的准确性和全局轨迹一致性。需要注意的是，估计的位姿和groundtruth通常不在同一个坐标系中，所以我们需要先将它们配对：对于立体SLAM和RGB-D SLAM，尺度是统一的，所以我们需要计算一个变换矩阵通过最小二乘法从估计位姿到真实位姿 $S\epsilon_{\text{东南欧}}(3)$ 。对于尺度不确定的单目相机，我们需要计算类似的变换矩阵 $S\epsilon_{\text{辛}}(3)$ 从估计位姿到真实位姿。因此第i帧的ATE定义如下：

$$F_{\text{我}} = \sqrt{\sum_{i=1}^n \| \text{反式}(\text{我}) - \text{反式}(S\epsilon_i) \|^2} \quad (1)$$

与RPE类似，均方根误差建议用于ATE统计。

$$\text{均方根误差 } (F_1: n, \Delta) := \sqrt{\frac{1}{n} \sum_{i=1}^n \| \text{反式}(\text{我}) - \text{反式}(S\epsilon_i) \|^2} \quad (2)$$

RPE：相对位姿误差主要描述相隔固定时间差的两帧的准确度（与真实位姿相比） Δ ，相当于里程表直接测量的误差。因此帧1的RPE定义如下：

$$E_{\text{我}} = (\sqrt{\sum_{i=1}^{n-\Delta} \| \text{反式}(\text{我}) - \text{反式}(\text{我}+i) \|^2})^{-1} \quad (3)$$

给定总数n和间隔 Δ ，我们可以得到 $(E_{\text{我}} = n - \Delta)RPE$ 。然后我们可以使用均方根误差均方根误差计算这个误差并获得一个population化值：

$$\text{均方根误差 } (E_1: n, \Delta) = \sqrt{\frac{1}{n} \sum_{i=1}^{n-\Delta} \| \text{反式}(\text{我}) - \text{反式}(\text{我}+i) \|^2} \quad (4)$$

反式(我)表示相对位姿误差的平移部分。我们可以从数据集的大小来评估算法的性能均方根误差价值。但在实践中我们发现，选择的选择有很多。 Δ 。为了综合衡量算法的性能，我们可以计算平均值均方根误差遍历所有 Δ ：

$$\text{均方根误差} = (E_1: n = \frac{1}{n} \sum_{\Delta=1}^n \text{均方根误差 } (E_1: n, \Delta)) \quad (5)$$

埃沃[25]是Python版SLAM系统的评估工具，可以使用多种数据集。除了可以获得ATE和RPE数据外，还可以绘制测试算法和真实轨迹的对比图。是一款非常方便的评估套件。SLAMBench2 [26]是一个公开可用的软件框架，可通过可扩展的数据集列表评估当前和未来的SLAM系统。它包括开放和封闭源代码，同时使用可比较和指定的性能指标列表。它支持多种现有的SLAM算法和数据集，例如ElasticFusion [27]，ORB-SLAM2 [28]，和OKVIS [29]，并且集成新的SLAM算法和数据集非常简单。

此外，我们还需要使用数据集来测试算法的具体可视化。用于测试SLAM性能各个方面常用数据集如表所示1。TUM数据集主要包括多视图数据集、3D物体识别与分割、场景识别、3D模型匹配、VSALM等各个方向的数据。根据应用方向可分为TUM RGB-D [30]，TUM MonoVO [31]，和慕尼黑工业大学六 [32]。其中，TUM RGB-D数据集主要包含具有真实地面轨迹的室内图像。此外，它还提供了两种评估轨道局部精度和全局一致性的度量，即相对姿态误差和绝对轨迹误差。TUM MonoVO用于评估单眼系统，

包含室内和室外图像。由于场景的多样性，地面真实性不可用，而是执行具有相同起始位置的大型序列，从而允许评估循环漂移。采用TUM VI对视觉惯性里程计进行评估。基蒂 [33] 数据集是由卡尔斯鲁厄理工学院和丰田美国理工学院联合创建的著名户外环境数据集。是全球最大的自动驾驶场景下计算机视觉算法评测数据集，包括单目视觉、双目视觉、Velodyne Lidar、POS轨迹等，是户外应用最广泛的数据集。欧洲洛克[34] 数据集 苏黎世联邦理工学院开发的视觉惯性数据集。城市景观[35]是一个与自动驾驶相关的数据集，专注于像素级场景分割和实例标注。此外，许多数据集用于各种场景，例如 ICL-NUIM [36], 纽约大学 RGB-D [37]，可可女士[38]， ETC。

表格1。 SLAM 的常见开源数据集。

数据集	传感器	环境	地面-真相	可用性	发展
城市景观	立体声	户外的	全球定位系统	[35]	戴姆勒股份公司研发部、马克斯·普朗克信息学研究所、达姆施塔特工业大学 Visaul 推理小组
基蒂	立体/3D 激光扫描仪	户外的	全球定位系统/惯导系统	[33]	卡尔斯鲁厄理工学院和丰田美国理工学院
TUM RGB-D	RGB-D	室内的	运动捕获	[30]	
慕尼黑工业大学 单一VO	单目	室内/室外循环漂移		[31]	慕尼黑工业大学
慕尼黑第六大学	立体声/IMU	室内室外	运动捕获	[32]	
欧洲洛克	立体声/IMU	室内的	车站/运动捕获	[34]	苏黎世联邦理工学院
ICL-NUIM	RGB-D	室内的	3D表面模型SLAM估计	[36]	伦敦帝国理工学院

2.3. 基于文献数据的SLAM发展分析

SLAM自问世以来，在机器人领域得到了广泛的应用。如图4，本文选取了近二十年来约1000篇与移动机器人相关的热门文章，制作了这张关键词热图。圆圈越大，关键词出现的频率越高。圆圈层从内到外展示了从过去到现在的时间，颜色越红越有吸引力。连接线表示不同关键字之间存在联系（数据来自 Web of Science 核心合集）。如图5视觉SLAM和语义SLAM相关论文的引用次数正在迅速增加。尤其是在2017年左右，视觉SLAM和语义SLAM的引用量猛增。传统的VSLAM研究已经取得了许多进展。为了使机器人能够从更高的层面感知周围环境，语义VSLAM的研究受到了广泛的关注。近年来，语义SLAM受到越来越多的关注。此外，如图所示6，本文从Web of Science核心合集中选取了约5000篇文章。从发表的有关SLAM的期刊标题来看，SLAM是机器人学界感兴趣的话题。

从以上数据可以看出，SLAM研究一直是热门话题。随着深度学习的快速发展，计算机视觉领域取得了前所未有的进步。因此，VSLAM也迎来了快速发展时期。将语义信息与VSLAM相结合将是很长一段时间的热门话题。语义VSLAM的发展可以让机器人真正实现自主。

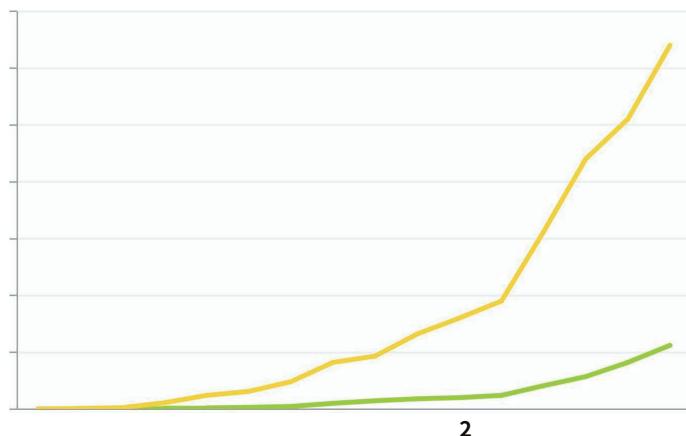
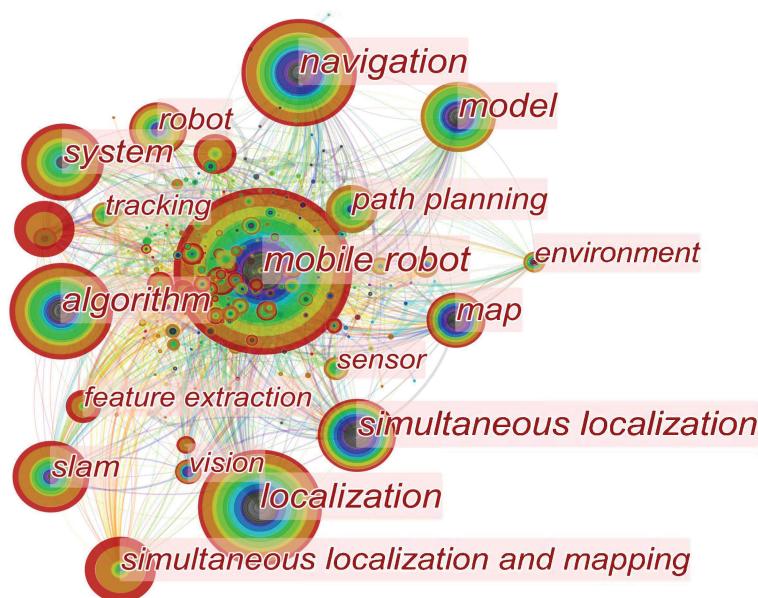


图 5. 近年来 Web of Science 关于视觉 SLAM 和语义 SLAM 的文章被引用情况（数据截至 2021 年 12 月）。

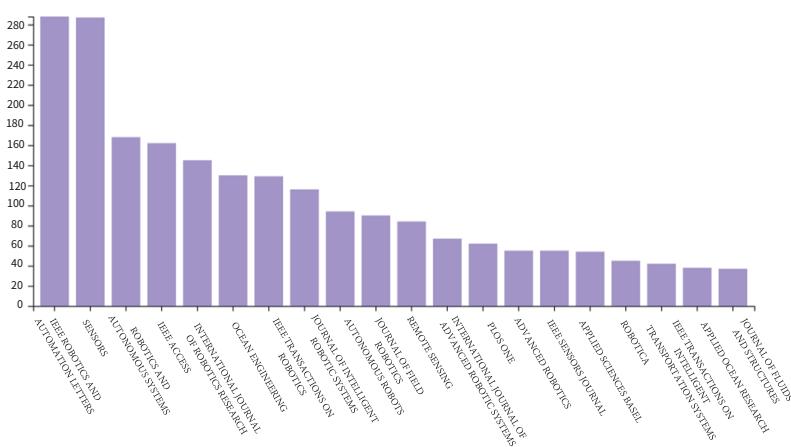


图 6. Web of Science 上有关 SLAM 的出版物标题。

2.4. 优秀学者及团队

此外，许多学者和团队为SLAM的研究做出了不可磨灭的贡献。如图7，我们分析了2000年至2022年约4000篇文章（数据来自Web of Science网站）。字体越大表明该作者受到的关注最多，反之亦然。他们所属的国家如图所示⁸。德国慕尼黑工业大学的计算机视觉小组是该领域的领导者。团队发布了DSO等多种经典视觉SLAM解决方案[39]和LSD-SLAM[40]，提高了视觉SLAM各方面的性能。瑞士苏黎世大学的机器人与感知小组也通过开发SVO和VO/VIO轨迹评估工具，为SLAM技术的快速发展做出了贡献。此外，苏黎世联邦理工学院计算机视觉与集成实验室也在该领域做出了大量努力。此外，他们在大规模户外测绘视觉语义定位领域取得了很多突破性进展。西班牙萨拉戈萨大学的机器人、传感和实时组SLAM实验室是SLAM发展的最大贡献者之一。实验室推出的ORB-SLAM系列是视觉SLAM领域的标志性方案，对SLAM的研究产生了深远的影响。此外，众多学者和团队的努力推动了视觉语义SLAM的快速发展，为未来解决各种问题奠定了基础。桌子2展示了优秀团队的作品和团队网站供大家参考，您可以通过团队名称后面的数字查看团队的网站。

一些学者在语义VSLAM研究方面做出了突出贡献。尼科·森德豪夫[41]例如，他们的团队在机器人场景理解和语义VSLAM方面取得了许多进展。该团队致力于让机器人理解它所看到的东西是最令人着迷的目标之一。为此，他们通过将对象检测与同步定位和建图(SLAM)技术相结合，开发了语义建图和语义SLAM的新方法。团队[42]的研究人员是澳大利亚机器人视觉中心的一部分，总部位于澳大利亚布里斯班昆士兰科技大学。他们致力于研究SLAM(同时定位和建图)的新方法，通过结合几何和语义信息来创建具有语义意义的地图。我们相信，这种语义丰富的地图将帮助机器人理解我们复杂的世界，并最终增加机器人在家庭和工业部署场景中交互的范围和复杂性。

表2.一些伟大的团队和他们的贡献。

团队	作品
帝国理工学院戴森机器人实验室[43]	代码-SLAM[44], ElasticFusion[27], 融合++[45], 语义融合[46]
慕尼黑工业大学信息学系计算机视觉组 慕尼黑工业大学[47]	D3VO[48], DM-VIO[49], LSD-SLAM[40], LDSO[50], DSO[39]
弗莱堡大学自主智能系统[51]	G映射[52], RGB-D SLAMv2[53]
科大空中机械人组[54]	VINS-单声道[55], VINS-融合[56], 基于事件的立体视觉 里程计[57]
华盛顿大学机器人和状态估计实验室[58]	飞镖[59], DA-RNN[60], RGB-D映射[61]
机器人、感知和实时小组 UNIVERSIDAD DE 萨拉戈萨[62]	ORB-SLAM2[28]、实时单目物体猛击[63]

图 7. 视觉SLAM领域杰出学者。

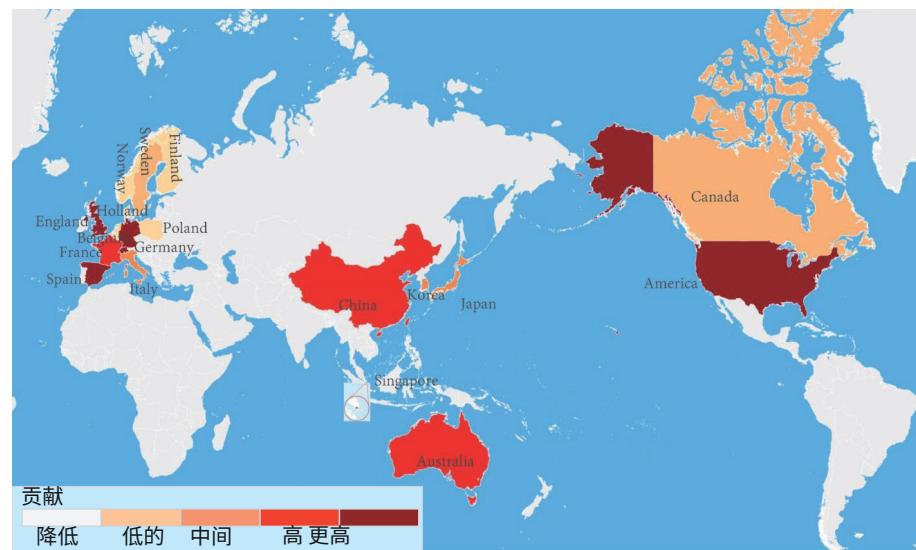


图 8. 不同国家在SLAM领域的贡献（颜色从浅到深表示贡献从低到高）。

3. 传统VSLAM

卡德纳等人。[\[15\]](#) 提出了经典的VSLAM框架，主要由前端和后端两部分组成，如图9。前端提供实时相机姿态估计，后端提供地图更新和优化。具体来说，成熟的视觉SLAM系统包括传感器数据采集、前端视觉里程计、后端优化、闭环检测、地图构建等模块。[\[64\]](#)。

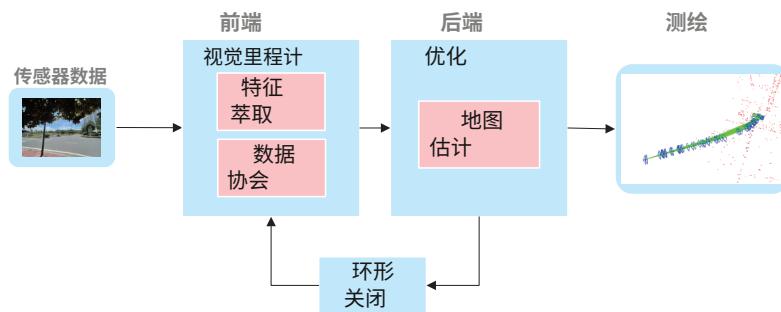


图 9.典型的视觉SLAM系统框架。

3.1. 单目/立体 VSLAM

本节我们将详细阐述基于单目或立体相机的VSLAM算法。对于VSLAM系统来说，视觉里程计作为SLAM的前端，是不可或缺的一部分[65]。参考号[20]指出，根据前端视觉里程计采集的图像信息不同，VSLAM可分为直接法和间接法。间接方法需要从采集的图像中选择一定数量的代表点，称为关键点，并在后续图像中检测和匹配它们以获得相机位姿。它不仅保存了图像的关键信息，而且减少了计算量，因此得到广泛应用。直接法利用图像的全部信息，无需预处理，直接对像素强度进行操作，在纹理稀疏的环境下具有更高的鲁棒性[66]。间接法和直接法都受到了广泛的关注并得到了不同程度的发展。

3.1.1. 基于特征方法的VSLAM

间接VSLAM的核心是检测、提取和匹配几何特征（点、线或面），估计相机位姿，并在保留重要信息的同时构建环境地图，可以有效减少计算量，因此得到了广泛的应用。^[67]。基于点特征的VSLAM方法由于其简单性和实用性长期以来被认为是间接VSLAM的主流方法^[68]。

特征提取早期多采用角点提取方法，如Harris[69]，快速地 [70]，GFTT [71]然而，在许多场景下，简单的角点无法提供可靠的特征，这促使研究人员寻求更稳定的局部图像特征。目前，典型的基于点特征的VSLAM方法首先使用特征检测算法，例如SIFT[72]、冲浪[73] 和 ORB [74]，提取图像中的关键点进行匹配。然后在最小化重投影误差后获得姿态。图像中的特征点和相应的描述符用于数据关联。此外，初始化时的数据关联是通过特征描述符的匹配完成的[75]。表中3中，我们列出了常见的传统特征提取算法并比较了它们的主要性能，以帮助读者有更全面的了解。

表3。常用特征提取算法对比表。

方法	年	类型	速度	回转不变性	规模不变性	照明不变性	反对-不变性
ORB [74]	2011年	观点	高的	是的	是的	是的	更强
冲浪[73]	2008年	观点	中间	是的	是的	不	星期
快速地 [70]	2006年	观点	高的	不	是的	不	星期
筛选变换[72]	2004年	观点	低的	是的	是的	是的	强的
希托马西 [71] 哈里斯[69]	1994年	康纳	中间	是的	不	是的	星期
迷幻剂[76]	1988年	康纳	低的	是的	不	是的	星期
	2010年	线	中间	是的	是的	是的	更强

戴维森等人。[\[77\]](#) 在2007年提出了MonoSLAM，该算法被认为是第一个实时单目VSLAM算法，可以实现实时漂移自由运动结构恢复。前端跟踪稀疏特征点shi-Tomasi角点进行特征点匹配，后端使用扩展卡尔曼滤波器（EKF）[\[78\]](#) 进行优化，可以实时在线构建稀疏环境地图。该算法在SLAM研究中具有里程碑意义，但EKF方法导致存储量和状态量呈平方增长，因此不适合大规模场景。同年，PTAM问世[\[79\]](#) 改善了MonoSLAM无法在多种环境下长时间稳定工作的问题。PTAM作为第一个在后端使用非线性优化的SLAM算法，解决了基于滤波器的方法中数据快速增长的问题。此外，它首次将跟踪和映射分离到两个不同的线程中。前端使用FAST角点检测利用图像特征提取和估计相机运动，后端负责非线性优化和映射。不仅保证了SLAM在相机位姿计算中的实时性，也保证了整个SLAM系统的准确性。但由于没有环回检测模块，长时间运行时会累积错误。

2015年，MUR-Artal等人。提出了ORB-SLAM[\[80\]](#)。该算法被认为是PTAM的优秀继承者，并且在PTAM的基础上增加了闭环检测模块，有效降低了累积误差。作为一个使用ORB特征匹配的实时单目视觉SLAM系统，整个过程都是围绕ORB特征进行的。如图10创新地使用了跟踪、局部建图、闭环检测三个线程。另外，闭环检测线程使用词袋模型DBoW[\[81\]](#) 用于循环闭合。基于BoW模型的闭环方法可以通过检测图像相似度来快速检测闭环。此外，在处理速度和地图构建精度方面均取得了良好的成绩。后来几年，团队推出了ORB-SLAM2[\[28\]](#) 和ORB-SLAM3[\[82\]](#)。ORB-SLAM系列因其实时CPU性能和鲁棒性而成为最广泛使用的视觉SLAM解决方案之一。然而，ORB-SLAM系列严重依赖环境特征，因此在没有纹理特征的环境中可能很难获得足够的特征点。

基于点特征的SLAM系统过于依赖点特征的质量和数量。在走廊、窗户、白墙等弱纹理场景中很难检测到足够的特征点，影响系统的鲁棒性和准确性，甚至导致跟踪失败。另外，由于相机的快速移动、光照变化等原因，点特征的匹配数量和质量都会严重下降。为了改进基于特征的SLAM算法，线特征在SLAM系统中的应用引起了越来越多的关注[\[83\]](#)。常用的线条特征提取算法是LSD[\[76\]](#)。

近年来，随着计算机计算能力的提高，基于VSLAM的在线特征也得到了快速发展。史密斯等人。[\[84\]](#)在2006年提出了基于单目VSLAM算法的在线特征提取。线条由两个端点表示，在SLAM系统中利用线条特征来检测和跟踪小场景中线条的两个端点。该系统可以单独使用线特征，也可以与点线特征结合使用，这在VSLAM研究中具有开创性意义。2014年，Perdices等人。提出了LineSLAM，一种基于线的单目SLAM算法[\[85\]](#)。对于线条提取，本方案采用[中的线条提取方案[\[86\]](#)]。每次获取关键帧时它都会检测线条。然后使用无迹卡尔曼滤波器（UKF）来预测当前相机状态和地线的矢量概率分布。然后，将线路预测结果与检测到的线路进行匹配。由于该方案没有闭环，而且线段不是有限长度而是无限长度，因此在实际中使用起来比较困难。

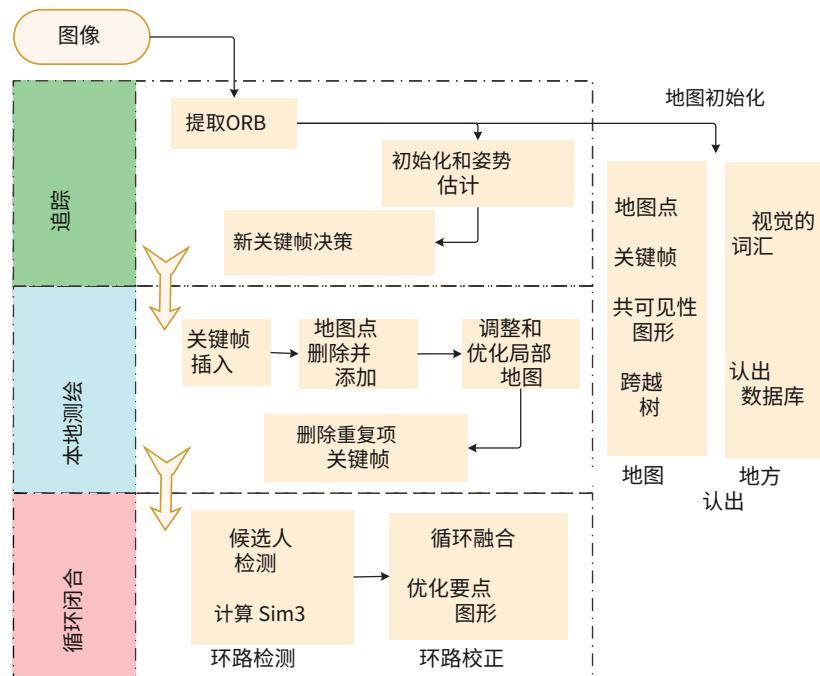


图 10. ORB-SLAM 流程图。

如图11，与单独的点特征或线特征相比，组合点特征和线特征的结合增加了特征观测和数据关联的数量。此外，线特征对光线变化的敏感度低于点特征，这提高了原系统的定位精度和鲁棒性。[\[76\]](#)。

2016 年，克莱因等人。[\[87\]](#) 采用点线融合的方法来改善跟踪。由于相机快速移动导致图像模糊，导致 SLAM 系统失效。2017 年，普马罗拉等人。[\[88\]](#) 发表了单目 PL-SLAM，以及 Gomez-Ojeda 等人。[\[89\]](#) 同年发表立体PL-SLAM。两种算法在ORB-SLAM的基础上，利用LSD检测算法检测线特征，然后将SLAM各环节中的点线特征进行组合。即使大多数点特征消失，它也可以工作。此外，它提高了SLAM系统的准确性、鲁棒性和稳定性，但实时性不佳。

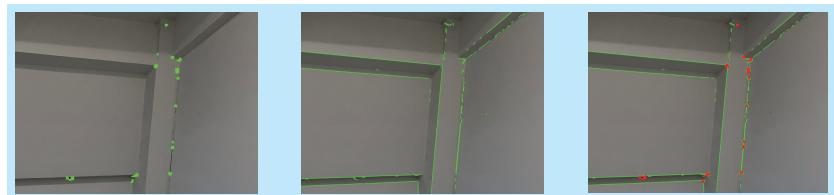


图 11. 弱纹理环境下点和线特征提取的比较。从左到右依次为ORB点特征提取、LSD线特征提取、点线组合特征提取。

另外，在某些环境下，还存在一些明显的表面特征，引起了一些研究者的极大兴趣。参考号 [\[90\]](#) 提出了平面与直线相结合的地图构建方法。通过将表面特征引入实时 VSLAM 系统，通过结合低层特征，减少误差，提高系统鲁棒性。2017 年，Li 等人。[\[91\]](#) 针对人工环境提出了一种基于点、线、面融合的VSLAM算法。点特征用于机器人当前位姿的初始估计。直线和平面用于描述环境。然而，大多数飞机仅存在于人工环境中，在自然环境中很少能找到合适的飞机。这些都限制了它的应用范围。

与仅依赖点特征的方法相比，仅依赖线或面的SLAM系统大多数情况下只能在人工环境中稳定工作。结合点、线、面特征的VSLAM方法提高了弱纹理场景、光照变化、相机快速移动等情况下的定位精度和鲁棒性。然而线面特征的引入增加了特征提取和匹配的时间消耗，从而降低了SLAM系统的效率。因此，基于点特征的VSLAM算法仍然占据主流地位[92]。桌子4显示几何特征的比较。

表 4. 几何特征比较表。

特征	好处	不利之处
观点	是最流行、最常用的功能，易于存储和匹配，而且速度普遍较快。	在强光和相机快速旋转的环境下很难提取足够的特征。
线	它具有自然光照和视角不变性，同时更先进的功能还提高了跟踪的鲁棒性和准确性。特别是在某些人工场景（室内、走廊）中，可以克服无纹理或不可靠纹理的干扰。	线段的检测和匹配时间比特征点长。后端也没有标准的、通用的SLAM优化和环回模块。线特征匹配也比较困难，例如线段容易断裂、没有很强的几何约束（如极线几何约束）、对纹理缺失的地方识别能力不强等。
飞机	在人工环境中具有更稳定的效果。	范围较小，只能在某些人工环境下操作。

3.1.2. 基于直接法的VSLAM

与基于特征的方法不同，直接方法直接对像素强度进行操作，并且可以保留有关图像的所有信息。此外，直接法取消了特征提取和匹配的过程，因此计算效率优于间接法。此外，它对复杂纹理的环境具有良好的适应性。在缺少特征的环境下依然能保持良好的效果 [93]。直接法与光流法类似，它们都有一个很强的假设：灰度不变性，其原理如图12。

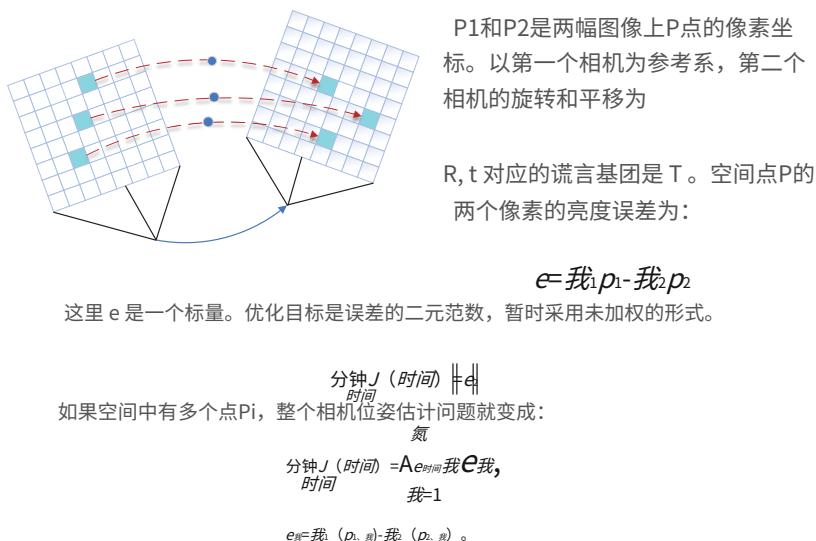


图 12. 直接法示意图。

2011 年, 纽科姆等人。[94] 提出了 DTAM 算法, 该算法被认为是第一个实用的 VSLAM 直接方法。DTAM 允许通过将输入图像与重建地图创建的图像进行比较来进行跟踪。该算法对环境进行精确而详细的重建。然而, 它影响了存储和处理数据的计算成本, 因此只能在 GPU 上实时运行。LSD-SLAM [40] 忽略无纹理区域以提高运行效率, 并且可以在 CPU 上实时运行。LSD-SLAM 是另一种主要的间接方法, 将无特征提取与半密集重建相结合, 其核心是使用半密集重建的视觉里程计。该算法由三个步骤组成: 跟踪、深度估计和地图优化。首先, 最小化光度误差以估计传感器位姿。其次, 选择关键帧进行深度估计。最后, 在地图优化步骤中, 将新的关键帧合并到地图中, 并使用姿态优化算法进行优化。2014 年, 福斯特等人。[95] 提出了半直接视觉 SLAM 算法 SVO。由于该算法不需要为每一帧提取特征, 因此可以以高帧速率运行, 这使其能够在低成本嵌入式系统中运行[80]。SVO 结合了特征点法和直接法的优点。该算法分为两个主线程: 运动估计和建图。运动估计是通过特征点匹配进行的, 而映射是通过直接方法进行的。SVO 有很好的效果, 但作为一种纯粹的视觉方法, 它只进行短期的数据关联, 这限制了它的准确性[82]。2018 年, 恩格尔等人。[39] 提议的 DSO。DSO 可以有效地利用任何图像像素, 这使得它即使在无特征区域也具有鲁棒性, 并且可以获得比 SVO 更准确的结果。DSO 可以在特征点检测器性能较差的情况下计算准确的相机姿态, 从而提高低纹理区域或模糊图像的鲁棒性。此外, DSO 使用几何和光度相机校准结果来进行高精度估计。但 DSO 仅考虑局部几何一致性, 因此不可避免地产生累积误差。此外, 它不是一个完整的 SLAM, 因为它不包括闭环、地图重用等。

到目前为止, VSLAM 在直接和间接方法上都取得了很多成果。桌子 5 比较了直接法和间接法的优缺点, 以帮助读者更好地理解。

表 5. 直接法与间接法的比较。

方法	好处	缺点
间接	(1) 特征点本身对光线、运动、旋转不敏感, 因此相对稳定。 (2) 相机移动速度较快 (相对直接法) 并且能够成功跟踪, 具有较好的鲁棒性。 (3) 研究时间长, 方案成熟。	(1) 关键点的提取、描述和匹配需要较长的时间。 (2) 不能使用特征点丢失场景。 (3) 只能构建稀疏映射。
直接的	(1) 速度快, 可以节省计算特征点、描述符的时间。 (2) 可用于特征缺失的情况 (如白墙), 这种情况下特征点法会迅速恶化。 (3) 可以构建半稠密甚至稠密映射。	(1) 由于假设灰度级不变, 因此容易受到光照变化的影响。 (2) 需要缓慢的相机移动或较高的采样频率 (可以通过图像金字塔来改进)。 (3) 单像素或像素块的区分度不强, 采取重数量而不重质量的策略。

3.2. RGB-D SLAM

RGB-D 相机是近年来推出的视觉传感器。它可以同时采集环境彩色图像和深度图像, 主要通过主动发射红外结构光或计算飞行时间 (TOF) 直接获得深度图。[96]。RGB-D 相机作为一种特殊相机, 可以获取空间的三维信息。

更方便。因此三维重建受到了广泛的关注和发展[97]。

KinectFusion [98] 是第一个基于 RGB-D 相机的实时 3D 重建系统。它使用深度创建的点云通过 ICP (迭代最近点) 来估计相机位姿。然后根据相机位姿拼接多帧点云集合，并通过 TSDF (截断符号距离函数) 模型表达重建结果。3D 模型可以通过 GPU 加速实时构建。然而，该系统尚未通过闭环进行优化。而且，长期运行会出现明显误差，RGB-D 相机的 RGB 信息没有得到充分利用。相比之下，ElasticFusion [27] 充分利用了 RGB-D 相机的颜色和深度信息。它通过 RGB 颜色一致性来估计相机位姿，并通过 ICP 来估计相机位姿。然后通过不断优化和重建地图来提高相机位姿的估计精度。最后采用面元模型进行地图表示，但只能在室内小场景中重建。连续的[99] 在 KinectFusion 的基础上增加了闭环，首次利用变形图进行 3d 刚体重建的非刚体变换。因此使得两环闭包重建的结果重叠，在室内环境中取得了良好的效果。与上述算法相比，RGB-D SLAMv2 [53] 是一个非常优秀且全面的系统。包含图像特征检测、优化、闭环等模块，适合初学者进行二次开发。

虽然 RGB-D 相机使用起来比较方便，但是 RGB-D 相机对光线极其敏感。再者，存在着空间狭窄、噪音大、视野小等诸多问题，所以大部分情况只在室内使用。此外，现有的算法必须使用 GPU 来实现。所以主流的传统 VSLAM 系统仍然不采用 RGB-D 相机作为主要传感器。然而，在室内三维重建中，RGB-D 相机被广泛使用。此外，由于能够构建密集的环境地图，语义 VSLAM 方向，RGB-D 相机被广泛使用。桌子 6 展示了基于 RGB-D 相机的经典 SLAM 算法。

表 6. 适用于带有 RGB-D 相机的传感器的一些 SLAM 算法。

方法	年	摄像机追踪	循环闭合	代码资源
KinectFusion [98]	2011 年	直接的	不	[100]
连续的 [99] RGB-D	2012 年	直接的	是的	[101]
SLAMv2 [53] 弹性融合	2013 年	间接	是的	[102]
[27] DVO-SLAM [104]	2016 年	直接的	是的	[103]
BundleFusion [106]	2017 年	直接的	是的	[105]
RGBDTAM [108]	2017 年	杂交种	是的	[107]
	2017 年	直接的	是的	[109]

3.3. 视觉惯性 SLAM

纯视觉 SLAM 算法已经取得了很多成果。然而，仅使用相机作为单一传感器仍难以解决相机快速移动和光照不良造成的图像模糊的影响。[110]。IMU 被认为是与相机最互补的传感器之一。可以在短时间内获得高频的准确估计，并减少动态物体对相机的影响。另外，相机数据可以有效修正 IMU 的累积漂移[111]。同时，由于相机和 IMU 的小型化和成本降低，视觉惯性融合也获得了快速发展。此外，它成为传感器融合的首选方法，受到许多研究人员的青睐[112]。目前，视觉惯性融合根据状态向量中是否添加图像特征信息可以分为松耦合和紧耦合两种。[113]。松耦合意味着 IMU 和相机分别估计它们的运动，然后融合它们的姿态估计。紧耦合是指 IMU 的状态和 IMU 的状态的组合。

相机共同构建运动和观测方程，然后进行状态估计[114]。

3.3.1. 松耦合视觉惯性

松耦合核心是融合视觉传感器和IMU分别计算的位置和姿态。融合对两者得到的结果没有影响。一般情况下，融合是通过EKF进行的。史蒂芬·韦斯 [115] 在他们的博士论文中提供了开创性的见解。参考号 [116] 提出了一种高效的松耦合方法，通过使用RGB-D相机和IMU获得了良好的实验结果。松耦合实现比较简单，但融合结果容易出错，这方面的研究还很少。

3.3.2. 视觉惯性紧耦合

紧耦合的核心是通过优化的滤波器将视觉传感器和IMU的状态结合起来。需要将图像特征添加到特征向量中共同构造运动方程和观测方程。然后进行状态估计以获得位姿信息。紧耦合需要充分利用视觉和惯性测量信息，方法实现复杂，但可以获得较高的位姿估计精度。因此，它也是主流的方法，并且在这方面已经取得了很多突破。

2007年，Mourikis等人。[117] 提出MSCKF。MSCKF的核心是以紧耦合的方式融合EKF下的IMU和视觉信息。与单独的VO算法相比，MSCKF可以适应更剧烈的运动和纹理损失，具有更高的鲁棒性。速度和准确度也能达到很高的水平。MSCKF已广泛应用于机器人、无人机、AR/VR领域。但由于后端采用卡尔曼滤波方法，无法利用全局信息进行优化，且没有环回检测会导致误差累积。参考号 [29] 提出了基于双目视觉和IMU融合的OKVIS。但它只输出六自由度位姿，没有环回检测和地图，所以它不是严格意义上的完整SLAM。虽然精度不错，但是跑的时间长了，姿势就会松动。虽然这两种算法都取得了不错的效果，但并没有得到广泛推广。缺少闭环模块不可避免地会导致长时间运行时累积错误。

VINS-Mono的出现[55] 打破了这种局面。2018年，香港科技大学（HKUST）的团队推出了单目惯性紧耦合VINS-Mono算法。此后又发布了扩展版本Vins-Fusion，支持多传感器集成，包括单目+IMU、立体+IMU，甚至仅立体，还提供带GPS的版本。VINS-mono 是视觉和 IMU 的经典融合。其定位精度与OKVIS相当，并且比OKVIS拥有更完整、更稳健的初始化和闭环检测流程。同时，VINS-Mono为视觉SLAM的研究和应用树立了标准，更多的是单目+IMU。在机器人导航，特别是无人机自主导航中，单目相机不受RGB-D相机（易受光照影响且深度信息有限）和立体相机（占用较大空间）的限制。可适应室内、室外及不同光照环境，适应性好。

惯性传感器作为相机的补充，可以有效解决单一相机无法应对的问题。视觉惯性融合必将成为SLAM研究的长期热点方向。然而多个传感器的引入会导致数据量的增加，这对计算能力提出了很高的要求[118]。因此，我们认为视觉惯性融合的下一个热点问题将体现在传感器融合数据的高效处理上。如何更好地利用不同传感器的数据将是一个长期有吸引力的热点问题。由于视觉传感器信息获取丰富、使用方便且价格低廉，构建的环境地图更接近人类识别的真实环境。经过几十年的发展，基于视觉的SLAM技术已经取得了许多优秀的成果。桌子7总结了一些

最好的基于视觉的SLAM算法，比较它们在关键领域的性能，并提供开源地址以帮助读者做出更好的选择。

表 7. 最佳基于视觉的 SLAM 算法。

	方法	传感器	前端	后端	循环闭合	测绘	代码资源
视觉的	单SLAM [77]	中号	磷	F	不	疏	[119]
	PTAM [79]	中号	磷	氧	不	疏	[120]
	ORB-SLAM2 [28]	中号/中号/右号	磷	氧	是的	疏	[121]
	PL-SVO [122]	中号	PL	氧	不	疏	[123]
	PL-SLAM [88]	多发性硬化症	PL	氧	是的	疏	[124]
	DTAM [94]	中号	D	氧	不	稠密	[125]
	SVO [95]	中号	H	氧	不	疏	[126]
	LSD-SLAM [40]	多发性硬化症	D	氧	是的	半密	[127]
	数字SO[39]	中号	D	氧	不	疏	[128]
	方法	传感器	耦合	后端	循环闭合	测绘	代码资源
视觉惯性	MSCKF [117]	米+我	时间	F	不	疏	[129]
	奥克维斯 [29]	S+我	时间	氧	不	疏	[130]
	罗维奥 [131]	米+我	时间	F	不	疏	[132]
	VINS-单声道 [55]	米+我	时间	氧	是的	疏	[133]

传感器：M代表单目摄像头；S代表立体相机；R代表RGB-D相机，I代表IMU。前端：P代表Point；PL代表点-线；D代表直接；H代表混合动力。后端：F代表Filtering；O代表优化。耦合：T代表Tightly。

本章我们对传统的基于视觉的SLAM算法进行了总结，并总结了一些优秀的算法供大家参考，希望能让读者有更全面的了解。接下来，我们将介绍具有语义信息融合的VSLAM，旨在更深入地探索SLAM领域。

4. 语义VSLAM

语义SLAM是指不仅能够获取未知环境的几何信息和机器人运动信息，而且能够检测和识别场景中的目标的SLAM系统。它可以获取自身的功能属性、与周围物体的关系等语义信息，甚至了解整个环境的内容[134]。传统的VSLAM以点云等形式来表示环境，这对我们来说是一堆无意义的点。为了从几何和内容两个层面感知世界，为人类提供更好的服务，机器人需要进一步抽象这些点的特征并理解它们。[135]。随着深度学习的发展，研究人员逐渐意识到它对SLAM问题可能的帮助[136]。语义信息可以帮助SLAM在更高层次上理解地图。此外，它减少了SLAM系统对特征点的依赖，提高了系统的鲁棒性[137]。

现代语义VSLAM系统离不开深度学习的帮助，通过学习获得的特征属性和关联关系可以用于不同的任务中[138]。深度学习作为机器学习的一个重要分支，在图像识别方面取得了令人瞩目的成果[139]、语义理解[140]，图像匹配[141]、3D重建[142]，以及其他任务。深度学习在计算机视觉中的应用可以大大缓解传统方法遇到的问题[143]。传统的VSLAM系统在许多方面取得了值得称赞的成果，但仍然存在许多具有挑战性的问题有待解决[144]。参考号 [145] 详细总结了基于深度学习的VSLAM，并指出了传统VSLAM存在的问题。这些作品[146–149]建议用深度学习来替代传统SLAM的一些模块，如闭环、位姿估计等，以改进传统方法。

机器学习是人工智能的一个子集，它使用统计技术提供从计算机“学习”数据的能力，而无需复杂的编程。与特定任务的算法不同，深度学习是基于学习数据的机器学习的子集。它的灵感来自于所谓的人工神经网络的功能和结构。深度学习通过学习将世界显示为更简单的概念和层次结构，并根据不太抽象的概念计算更抽象的表示，从而获得了巨大的灵活性和力量。传统机器学习和深度学习最重要的区别是数据扩展的性能。当数据非常小时，深度学习算法不能很好地工作，因为它们需要大数据才能完美地识别和理解它。机器学习算法的性能取决于识别和提取的特征的准确性。另一方面，深度学习算法从数据中识别这些高级特征，从而减少了为每个问题开发全新特征提取器的工作量。深度学习是机器学习的一个子集，与传统的机器学习算法相比，它已被证明是一个更强大、更有前途的行业分支。它以其分层的特点实现了传统机器学习无法实现的许多功能。SLAM系统需要采集环境中的大量信息，因此需要计算的数据量巨大，而深度学习模型正好适合解决这个问题。

本文认为语义VSLAM是一个不断演化的过程。早期，一些研究人员试图通过使用CNN等神经网络提取环境中的语义信息来提高VSLAM的性能。现代阶段，目标检测、语义分割等深度学习方法是推动语义VSLAM发展的有力工具。因此，在本章中，我们将首先描述典型神经网络在VSLAM中的应用。我们认为这是现代语义VSLAM发展的前提。神经网络在VSLAM中的应用为现代语义VSLAM提供了模型。本文认为神经网络是将语义信息引入现代语义VSLAM系统的桥梁并获得快速发展。

4.1. 具有 VSLAM 的神经网络

数字13展示了CNN和RNN的典型框架。CNN可以从图像中捕获空间特征，这有助于我们准确地识别图像中的对象及其与其他对象的关系[150]。RNN的特点是可以处理图像或数值数据。由于网络本身的记忆能力，它可以学习具有上下文相关性的数据类型[151]。此外，其他类型的神经网络如DNN（深度神经网络）也有一些尝试性的工作，但都处于起步阶段。本文指出，CNN的优点是用一定的模型提取事物的特征，然后根据特征进行分类、识别、预测或决策。对VSLAM的不同模块都有帮助。另外，本文认为RNN在帮助建立邻近帧之间的一致性方面具有很大的优势。此外，高层特征具有更好的区分性，可以帮助机器人更好地完成数据关联。

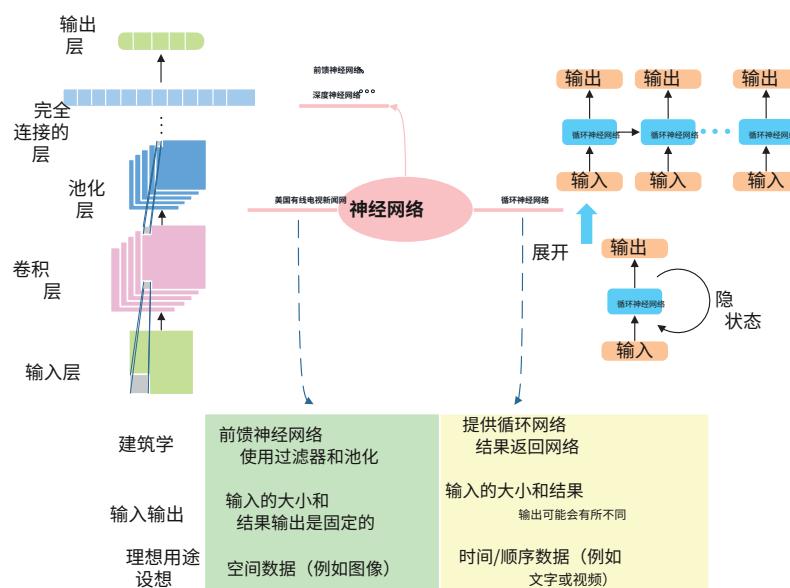


图 13.CNN 和 RNN 的结构框图。CNN 适合从层次或空间数据中提取未标记的特征。RNN 适用于时态数据和其他类型的顺序数据。

4.1.1. CNN 与 VSLAM

传统的帧间估计方法采用基于特征的方法或直接通过多视图几何识别相机位姿的方法[152]。基于特征方法需要复杂的特征提取和匹配。直接方法依赖于像素强度值，这使得传统方法很难在强光照或稀疏纹理等环境下获得期望的结果[153]。相比之下，方法

基于深度学习的更加直观、简洁。那是因为它们不需要提取环境特征、特征匹配和复杂的几何运算[154]。

由于 CNN 的特征检测层通过训练数据进行学习，因此避免了特征检测在显示中提取并在使用过程中从训练数据中隐式学习。参考文献。[155,156] 等作品都做了详细的总结。

CNN 在图像处理方面的优势已经得到充分验证。例如，视觉深度估计改善了单目相机无法获得可靠的深度估计的问题深度信息[157]。2017 年，Tateno 等人。[158] 提出了实时 SLAM 系统“CNN-SLAM”基于 LSD-SLAM 框架下的 CNN。如图 14，算法通过训练深度估计网络得到可靠的深度图模型。采用 CNN 进行深度预测，输入到传统姿态估计等后续模块中，以提高定位和建图精度。此外，框架中还添加了 CNN 语义分割模块，为 VSLAM 系统的高级信息感知提供帮助。使用网络进行类似的工作

估计深度信息包括 Code-SLAM [42] 和 DVS0 [159] 基于立体声相机。同年，戈达尔等人。[160] 提出了一种无监督图像深度估计方案。利用立体数据集改进无监督学习，然后使用单帧进行姿态估计，与其他方案相比有很大的改进。

CNN 不仅解决了传统方法无法利用单目相机获得可靠深度数据的问题，而且改善了传统方法在相机位姿估计方面的缺陷。2020 年，杨等人。[48] 提出 D3VO。该方法从深度估计、位姿估计和不确定性估计三个方面使用了深度学习。将预测深度、位姿和不确定性紧密结合起来直接视觉里程计，同时提高前端跟踪和后端非线性优化的性能。然而，自监督方法很难适应所有环境。此外，秦等人。[161] 提出了一种基于语义特征的

2020年出现的定位方法，有效解决了传统视觉SLAM方法容易出现跟踪丢失的问题。其原理是利用CNN在地下停车场狭窄拥挤、缺乏GPS信号、光线昏暗、纹理稀疏的环境中检测语义特征。然后使用U-Net[162]进行语义分割，将地面上的停车线、减速带等指标分开，然后利用里程表信息。将语义特征映射到全局坐标系以构建停车场地图。然后将语义特征与之前构建的地图进行匹配以定位车辆。最后利用EKF将视觉定位结果与里程表信息进行融合，保证系统在地下停车环境下能够获得连续稳定的定位结果。朱等人。[163]通过使用CNN来关注光流输入的不同象限来学习旋转和平移。然而，端到端替代视觉里程计的方法简单粗暴，缺乏理论支持和泛化能力。

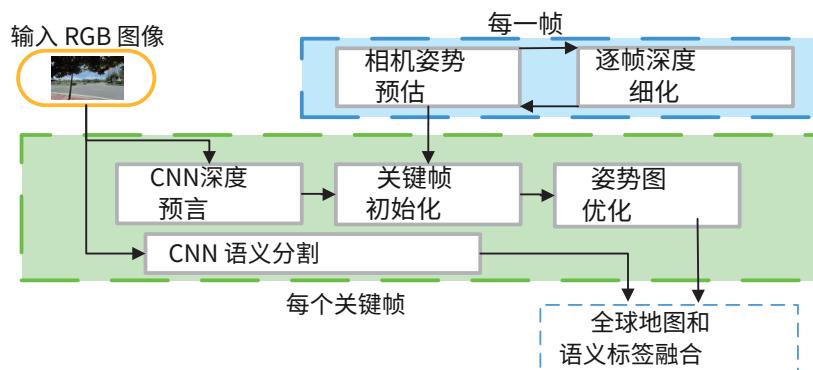


图 14.CNN-SLAM的结构。

闭环检测可以消除累积的轨迹误差和地图误差，决定整个系统的精度，本质上是一个场景识别问题[164]。传统方法是通过人工设计的稀疏特征或像素级密集特征来匹配。深度学习可以通过神经网络学习图像中的高级特征。此外，利用深度学习强大的识别能力提取图像的更高水平的鲁棒特征，其识别率可以达到更高的水平。这样，系统可以对图像变化有更强的适应能力，例如

透视和光照提高闭环图像识别能力[165]。

因此，基于深度学习的场景识别可以提高闭环检测的准确性，CNN对于闭环检测也获得了许多可靠的效果。梅蒙等人。[166]提出了一种基于字典的深度学习方法，与传统的Bow字典不同，采用了更高层次、更抽象的深度学习特征。该方法不需要创建词汇，内存效率更高，运行速度比同类方法更快。然而，该论文仅基于相似度得分检测周期，因此不具有广泛的代表性。李等人。[167]提出了一种基于学习特征的视觉SLAM系统DXSLAM，解决了上述方法的局限性。使用 CNN 从每一帧中提取局部和全局特征，

然后，这些特征被输入到现代 SLAM 管道中，用于姿势跟踪、局部映射和重新定位。与传统的基于BOW的方法相比，它实现了更高的效率和更低的计算成本。此外，秦等人。[168] 使用 CNN提取环境语义信息，并将视觉场景建模为语义

子图。利用语义信息可以有效提高环回检测的效率。参考文献。[169,170] 等人详细描述了深度学习在多个方面取得的成就。然而，随着更复杂和更好的模型的引入，

如何保证模型计算的实时性？如何在资源受限的平台上更好的设置闭环检测模型，以及轻量级的
模型也是一个大问题[171]。

CNN在替代传统VSLAM算法的一些模块，如深度估计、回环检测等方面取得了良好的效果。其稳定性仍然不如传统的VSLAM算法[172]。相比之下，CNN系统的语义信息提取带来了更好的效果。利用CNN对传统VSLAM的流程进行优化，提取具有更高层次特征的环境语义信息，使得传统VSLAM取得更好的效果。使用神经网络提取语义信息并将其与VSLAM相结合将是一个令人非常感兴趣的领域。借助语义信息，数据关联从传统的像素级升级到对象级。感知几何环境信息被分配语义标签以获得高级语义图。它可以帮助机器人理解自主环境和人机交互。桌子8展示了CNN网络在VSLAM中的一些主要应用环节。有的涉及很多方面，这里只列出主要贡献。

表 8.CNN 用于 VSLAM。

部分	方法	贡献
图像 预估	深度 CNN-SLAM [158]	仅对关键帧进行深度估计，提高了计算效率。
	UnDeepVo [173]	以无监督的方式实现了真实比例的单目视觉里程计。
	代码-SLAM [44]	实现了实时单目 SLAM 系统，可以同时优化相机运动和地图。
	DVSO [159]	设计一种新颖的深度网络，通过两阶段过程细化单个图像的预测深度。
姿态估计	德通等人。[174]	它仅使用点的位置，而不使用局部点的描述符。
	维网 [175]	自然而巧妙地结合特定区域的信息的能力可以显着减少漂移。
	D3VO [48]	所提出的单目视觉里程计框架在三个层面上利用深度学习网络。
	朱等人。[163]	提出一种新颖的四分支网络，利用卷积神经网络（CNN）关注光流输入的不同象限来学习旋转和平移。
环闭合	梅蒙等人。[166]	两个深度神经网络一起使用可以加速闭环检测并忽略移动物体对闭环检测的影响。
	李等人。[167]	使用词袋（BoW）方法训练局部特征的视觉词汇。基于局部特征、全局特征和词汇，构建了一种高可靠的闭环检测方法。
	秦等人。[168]	通过仅保留来自对象检测的语义和几何信息，将视觉场景建模为语义子图。
语义学 信息	CNN-SLAM [158]	通过整合几何和语义信息，生成带有语义信息的地图。
	纳西尔等人。[176]	实现实时语义分割并保持良好的区分效率。CNN 多视图的语义预测可以概率性地集成到地图中。
	语义融合[46]	
	秦等人。[161] 鲍曼 等人。[177]	提出了一种用于视觉 SLAM 框架的新颖语义特征。 提出了传感器状态和语义地标定位的优化问题。

4.1.2. RNN 与 VSLAM

RNN（循环神经网络）的研究始于20世纪80年代和90年代，并在21世纪初发展成为经典的深度学习算法之一。长短期记忆网络（LSTM）是最常见的循环神经网络之一[178]。LSTM 是 RNN 的变体，它会记住可控量的先前训练数据或者更恰当地忘记它[179]。如图15，给出了 LSTM 的结构及其不同模块的状态方程。具有特殊隐式单元的 LSTM 可以长时间保存输入。LSTM 继承了 RNN 模型的大部分特点，解决了梯度逐渐减少带来的梯度消失问题。

梯度回传过程。作为RNN的另一种变体，GRU（Gated Recurrent Unit）更容易训练并且可以提高训练效率[180]。RNN由于其记忆性和参数共享性，在学习序列的非线性特征方面具有一定优势。引入卷积神经网络CNN构建的RNN可以处理涉及序列输入的计算机视觉问题[181]。

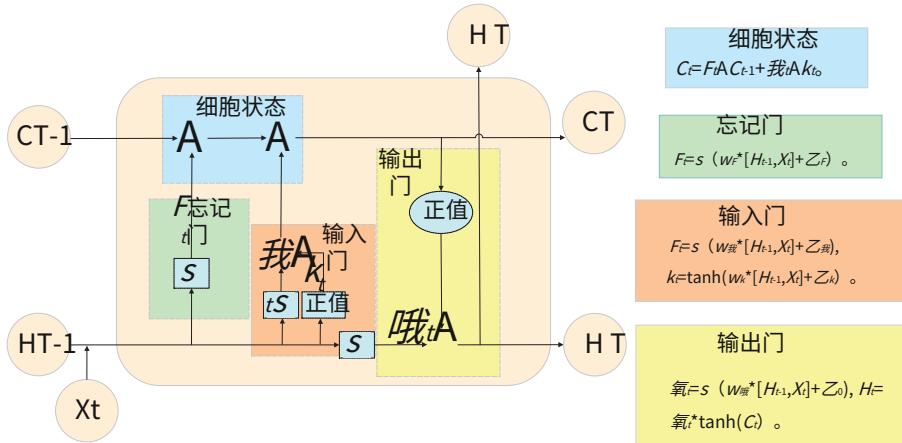


图 15.LSTM的基本框架。

在姿态估计中，引入端到端的深度学习方法来求解视觉图像帧之间的姿态参数，无需特征匹配和复杂的几何运算。通过直接输入附近帧可以快速获取帧之间的相对位姿参数[182]。薛等人。[183] 利用深度学习学习特征选择的过程，实现基于RNN的姿态估计。位姿估计中，旋转和位移分别训练，相比之下具有更好的适应性

用传统方法。2021 年，Teed 等人。[184] 介绍了DROID-SLAM，其核心是可学习的更新算子。如图16，更新运算符是 $3 \times$ 具有 H 隐藏状态的 3 个卷积 GRU。更新算子的迭代应用创建了一系列收敛到反映真实重建的固定点的态度和深度。该算法是一种用于视觉SLAM的端到端神经网络架构，在具有挑战性的环境中比以前的工作具有很大的优势。

现有方法大多采用CNN与RNN相结合来提高VSLAM的整体性能。CNN 和 RNN 可以使用单独的层进行组合，将 CNN 的输出作为 RNN 的输入。一方面，它可以通过CNN自动学习VO问题的有效特征表示。另一方面，它可以通过RNN隐式建模时序模型（运动模型）和数据关联模型（图像序列）[185]。2017 年，Yu 等人。[60] 结合RNN和KinectFusion对RGB-D采集的图像进行语义标注，重建3D语义图。他们在RNN中引入了一种新的闭环单元来解决GPU计算资源消耗的问题。该方法充分利用RNN的优势来实现语义信息的标注。高层次的特征具有更好的辨别力，帮助机器人更好地完成数据关联。由于使用RGB-D相机，它们只能在室内环境中操作。深度序列SLAM [186] 很好的解决了这个问题。在该方案中，使用可训练的 CNN+RNN 架构来从单个单目图像序列中联合学习视觉和位置表示。使用 RNN

整合短图像序列的时间信息。同时，利用这些网络的动态信息处理功能，首次实现了端到端位姿和序列位置学习。此外，能够从大型驾驶数据集的单个图像序列中学习有意义的时间关系。

在运行时间、精度和计算需求方面，基于序列的方法显着优于
优于传统方法，可在室外环境下稳定运行。

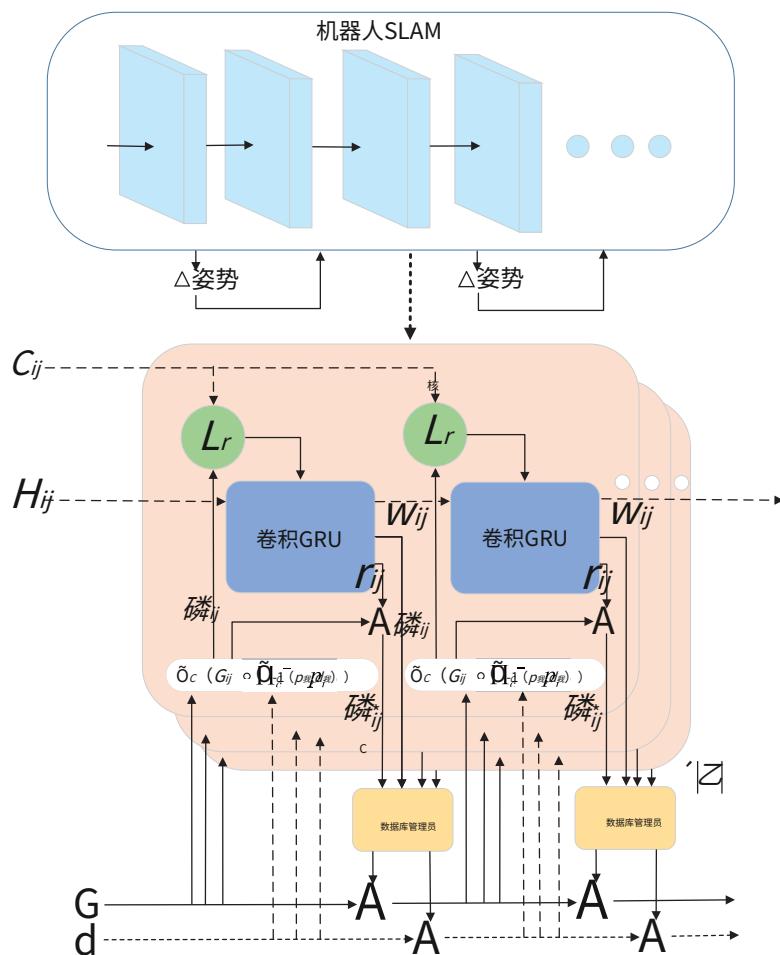


图 16.DROID-SLAM的核心框架。

CNN可以与VLSAM的很多环节结合起来，比如特征提取与匹配、深度估计、位姿估计等，并且在这些方面都取得了不错的效果。相比之下，RNN的适用范围较小，但在帮助建立邻近帧之间的一致性方面有很大的优势。RNN是深度学习中数据驱动时序建模的常用方法。IMU输出的高帧率角速度、加速度等惯性数据对时序有严格的依赖性，特别适合RNN模型。基于此，克拉克等人。[\[175\]](#)提出使用常规的小型LSTM网络来处理IMU的原始数据，获得IMU数据下的运动特征。最后，他们将视觉运动特征与IMU运动特征相结合，并将其发送到核心LSTM网络中进行特征融合和姿态

估计。其原理如图所示[17号](#)。

与姿态估计相比，我们认为RNN因其贡献而更具吸引力。视觉惯性数据融合的关键。该方法可以有效地融合视觉-惯性数据，并且比传统方法更加方便。类似的工作，例如[\[187,188\]](#)，证明融合策略的有效性，与直接融合相比，它提供了更好的性能。本文在表中给出了RNN对部分VSLAM的贡献[9](#)。

本节从经典神经网络CNN和RNN开始介绍深度学习与传统VSLAM的结合。桌子[10](#)展示了—些将神经网络与VSLAM相结合的优秀算法。

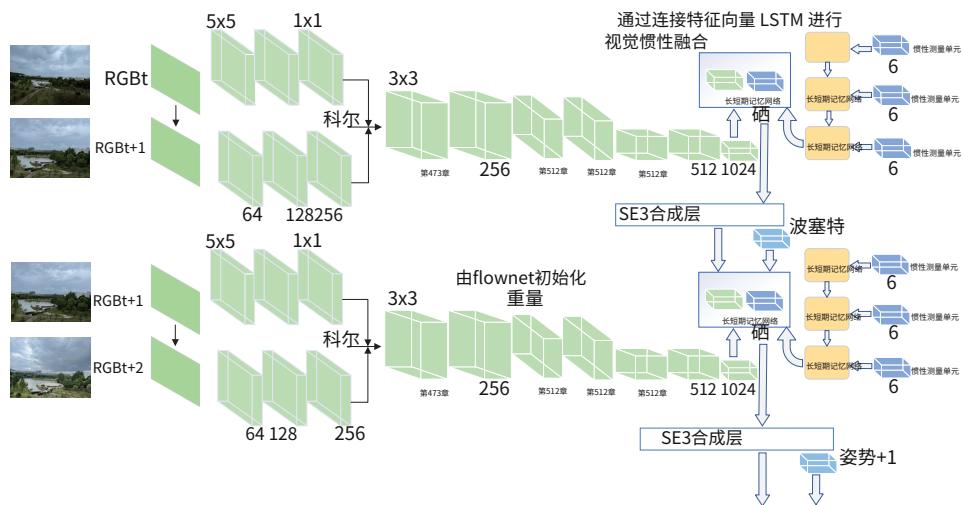


图 17. 克拉克等人。提出了使用 LSTM 进行视觉惯性融合的框架。

表 9.RNN 用于 VSLAM。

部分	方法	贡献
薛等人。[183]		提出一种双分支循环网络，利用当前的 CNN 进行特征表示和 RNN 进行图像序列推理，分别学习旋转和平移。
视氧浓度德等人。[184]		它包括通过密集束调整层对相机姿势和像素深度进行循环迭代更新。
DA-RNN [60]		用于联合 3D 场景映射和语义标记的新颖框架。
深度序列SLAM [186]		一种可训练的 CNN+RNN 架构，用于从路线的单个单目图像序列中联合学习视觉和位置表示。
克拉克等人。[175]		它是第一个用于视觉惯性里程计的端到端可训练方法，它在中间特征表示级别上执行数据融合。
维奥陈等人。[188] 亚辛等人。[189] 黄等人。[190]	深VIO [187]	它减少了相机-IMU 校准不准确以及数据不同步和丢失的影响。它提出了一种新颖的单目 VIO 端到端选择性传感器融合框架。
	D3VO [48]	使用对抗性训练和自适应视觉惯性传感器融合。
	深度序列SLAM [186]	首次提出视觉惯性+深度数据集的融合方法，进一步增强视觉与惯性传感器的优势互补。

表 10. 神经网络与VSLAM相结合的优秀算法。

	方法	年	传感器	神经网络	监督
视氧浓度	CNN-SLAM [158]	2017年	单目	美国有线电视新闻网	监督
	深沃[191]	2017年	单目	R-CNN	监督
	代码-SLAM [44]	2018年	单目	优网	监督
	DVSO [159]	2018年	立体声	调度网	半监督
	UnDeepVo [173]	2018年	单目	VGG编码器-解码器	无监督
	CNN-SVO [192]	2019年	单目	美国有线电视新闻网	杂交种
	甘沃[193]	2019年	单目	生成网络	无监督
	李等人。[194]	2019年	单目	美国有线电视新闻网	监督
	D3VO [48]	2020年	单目	美国有线电视新闻网	杂交种
	深度序列SLAM [186]	2020年	单目	CNN+RNN	监督
维奥	深度SLAM [145] 提升-SLAM [195]	2021年	单目	R-CNN	无监督
	张等人。[196]	2021年	单目	深度神经网络	监督
	维网 [175]	2017年	立体声	U-Net编码器-解码器	无监督
	VIO学习者 [197]	2020年	单目+IMU	CNN + LSTM	监督
	DeepVIO [187]	2019年	立体声+IMU	美国有线电视新闻网	无监督

4.2. 现代语义VSLAM

深度学习在姿态估计、深度估计、回环检测等方面取得了很多成果。然而，在VSLAM中，深度学习目前还无法撼动传统方法的统治地位。然而，将深度学习应用于语义VSLAM研究可以获得更有价值的发现，从而能够快速推动语义VSLAM的发展。参考文献。[\[60,158,168\]](#) 使用CNN或RNN来提取环境中的语义信息，以提高传统VSLAM中不同模块的性能。语义信息用于姿态估计和环回检测。它显着提高了传统方法的性能，并证明了语义信息对于VSLAM系统的有效性。本文认为，这为现代语义VSLAM的发展提供了技术支撑，是现代语义VSLAM的开始。利用目标检测、语义分割等深度学习方法创建语义地图，是语义SLAM发展的重要代表时期。参考文献。[\[135,200\]](#) 指出语义SLAM根据目标检测方法的不同可以分为两种类型。一是使用传统方法检测目标。实时单目物体SLAM是最常见的一种，使用大量的二进制词和物体模型数据库来提供实时检测。然而，它非常有限，因为语义类（例如“汽车”）的3D对象实体有很多种类。SLAM的另一种方法是使用深度学习方法进行对象识别，例如[\[46\]](#)。

语义和SLAM看似是独立的模块，但其实不然。在许多应用中，两者齐头并进。一方面，语义信息可以帮助SLAM提高建图和定位的准确性，特别是对于复杂的动态场景[\[201\]](#)。传统SLAM的建图和定位大多基于像素级的几何匹配。有了语义信息，我们可以将数据关联从传统的像素级升级到物体级，提高复杂场景的准确率[\[第202章\]](#)。另一方面，通过利用SLAM技术计算物体之间的位置约束，可以将一致性约束应用于同一物体在不同角度、不同时间的识别结果，从而提高语义理解的准确性。语义和SLAM的融合不仅对两者精度的提高有很大贡献，而且促进了SLAM在机器人领域的应用，例如机器人路径规划和导航、根据人类指令搬运物体、做家务、陪伴人类运动等，ETC。

例如，我们想要一个机器人从卧室走到厨房去拿一个苹果。这是如何运作的？依靠传统的SLAM，机器人计算自己的位置（自动）和苹果的位置（手动），然后进行路径规划和导航。如果苹果在冰箱里，还需要手动设置冰箱和苹果的关系。然而，现在有了我们的语义SLAM技术，人类可以更自然地向机器人发送“请去厨房给我拿一个苹果”，然后机器人会自动完成剩下的工作。如果机器人在作业过程中前方有污染的地面，传统的路径规划算法需要手动标记污染区域，以便机器人可以绕过它[\[203\]](#)。

语义信息可以帮助机器人更好地理解周围环境。将语义信息集成到VSLAM中是一个不断发展的领域，近年来受到越来越多的关注。本节将从定位、建图和动态物体去除两个方面来阐述我们对语义VSLAM的理解。我们认为深度学习对于VSLAM的最大贡献是语义信息的引入。它可以不同程度地提高传统方法不同模块的性能。尤其是在语义图的构建上，推动了整个智能机器人领域的创新。

4.2.1. 图像信息提取

现代语义VSLAM与传统VSLAM的核心区别在于目标检测模块的集成。它可以获取环境中物体的属性和语义信息[204]。语义VSLAM的第一步是从相机获得的图像中提取语义信息。此外，通过对图像信息进行分类，可以获得基于图像信息的语义信息[205]。传统的目标检测依赖于可解释的机器学习分类器，例如决策树和SVM，来分类和实现目标特征。但检测过程缓慢、准确率低、泛化能力弱[206]。基于深度学习的图像分类可以分为对象检测、语义分割和实例分割，如图18。



图 18.从左到右分别是YOLOv5、Deeplabv3、Mask R-CNN的测试渲染器。

如何更好地从图像中提取语义信息是计算机视觉领域的一个研究热点问题，其本质是从场景中提取物体特征信息[207]。

我们认为，虽然CNN等神经网络也有助于语义信息提取，但现代语义VSLAM更多地依赖于目标检测等语义提取模块。目标检测和图像语义分割都是从图像中提取语义信息的方法。图像语义分割是在像素级别上理解图像，以获得图像中的深层信息，包括空间、类别和边缘。基于深度语义分割技术

神经网络突破传统语义分割的瓶颈[208]。与语义分割相比，目标检测仅获取图像的物体信息和空间信息。此外，它通过绘制对象的候选框来识别每个对象的类别，因此目标检测比语义分割更快[209]。与物体检测相比，语义分割技术的精度更高，但速度却低很多[210]。

目标检测分为一级和两级结构[211]。早期的目标检测算法使用两阶段架构。创建一系列候选框作为样本后，通过卷积神经网络进行样本分类。常见的算法包括R-CNN [212], Fast R-CNN [213]，更快的 R-CNN [214]，等等。后来，YOLO[215]创造性地提出了一级结构。它直接一步完成了两阶段的两步，一步完成了物体的分类和定位，直接输出回归得到的候选框及其类别。一级减少了目标检测算法的步骤，直接将目标框定位问题转化为回归问题理论，无需创建候选框，速度优越。常见的算法有YOLO和SSD[216]。

2014年，R-CNN的出现颠覆了传统的物体检测方案，提高了检测精度，推动了物体检测技术的快速发展。其核心是提取候选区域，然后通过Alexnet获取特征向量，最后使用SVM分类和帧校正。然而，由于R-CNN采用串行特征提取方法，特征提取的速度受到限制。Ross在2015年提出Fast R-CNN就很好地解决了这个问题。Fast R-CNN中使用感兴趣区域池化（ROI Pooling）操作来提高特征效率

提取，并使用区域生成网络（RPN）进行坐标校正。RPN 中设置了许多候选框（锚点）。然后判断 anchor 与背景的依赖关系，计算出 anchor 的覆盖范围，判断目标是否被覆盖。此外，YOLO 提高了预测的准确性，加快了处理速度并增加了识别对象的类型，并提出了目标分类和检测的联合训练方法。YOLO 是使用最广泛的目标检测算法之一，提供实时检测以及此后的一系列改进版本。

与目标检测不同，语义分割不仅可以预测图像中目标的位置和类别，还可以准确描述不同种类目标之间的边界。然而，在语义分割技术中，普通的卷积神经网络无法获得足够的信息。为了解决这个问题，Long 等人提出了全卷积神经网络 FCN [217]。与 CNN 相比，FCN 没有全连接层。新的 FCN 获取特征图的空间位置，并将不同深度层的输出与层次结构融合。该方法将局部信息与全局信息相结合，提高了语义分割的准确性。在 Badriarayansn 等人提出的 Segnet 网络中 [218]，提出了编码器-解码器结构，结合两个独立的网络来提高分割的准确性。然而，两个独立网络的组合严重降低了检测速度。赵等人提议的 PSPNet [219] 以及金字塔模块，融合各个层级的特征，如金字塔，最后融合输出，进一步提高分割效果。

近年来，计算机性能的不断提高促进了视觉实例分割的快速发展。实例分割不仅有像素级别的分类（语义分割），还有不同物体的位置信息（目标检测），甚至可以检测到同一物体。2017 年，He 等人提出了 Mask R-CNN [220]。该算法是实例分割的创举。如图 19，其主要思想是在 Faster R-CNN 的基础上增加一个用于语义分割的分支。

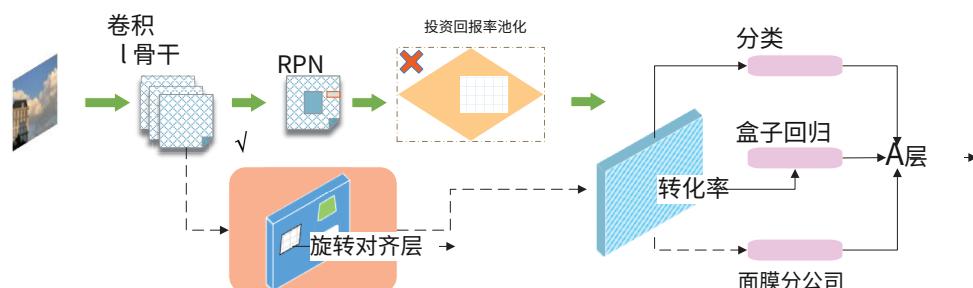


图 19. MASK-RCNN 的框架。

虽然基于神经网络的目标检测和分割技术已经很完善，但它需要依靠强大的计算能力来实现实时处理。VSLAM 对实时操作的要求很高，因此如何有效地将所需对象及其语义信息与环境分离将是一项长期而艰巨的任务。作为语义 VSLAM 的基础，在处理完语义分割后，我们将 p

请注意，我们将详细介绍语义信息对不同事物的影响
VSLAM 的各个方面。我们绍三个对象检测和定位、映射和语义分割的各个方面都是 s 的一种手段。桌子 11 显示了一些语义分割的贡献。
将对象命名为“移除”。
然而，语义
提取语义信息
图像信息
分割精度更好。实例分割融合了目标检测和语义分割，在精度上有突出的表现，但不能保证运行速度。对于一些无法提供原论文的方案，我们提供了开源代码，例如 YOLOv5。
对象检测
etection 比 s 快

表 11. 经典图像检测算法的一部分。

场地	模型	年	贡献
物体检测	R-CNN [212]	2014年	第一个成功地将深度学习应用于目标检测的算法。图像特征提取仅执行一次。
	快速 R-CNN [213]	2015年	
	更快的 R-CNN [214]	2017年	融入网络后，综合性能得到了很大的提升。
	固态硬盘[216]	2016年	SSD 是单相模型的早期化身。
	优洛[215]	2016年	将检测视为回归问题，使用网络输出位置和类别。
语义学分割	YOLOv5 [221]	2020年	环境配置简单，模型训练速度非常快。
	FCN [217]	2015年	它开启了卷积神经网络在语义分割中的首次应用。
	分段网络[218]	2017年	采用完全对称的结构。空洞卷积。
	DeepLabv1 [222]	2014年	
	DeepLabv3+ [223]	2018年	大大减少参数数量。
实例分割	PSP网[219]	2017年	金字塔池模块可以聚合来自不同区域的上下文信息。
	Mask R-CNN [220]	2017年	它不仅可以检测图像中的目标，还可以对每个目标给出高质量的分割结果。
	优拉特[224]	2019年	基于一阶段目标检测算法，整体架构设计非常轻量，在速度和效果上都取得了不错的效果。

4.2.2. 语义与位置

位置精度是SLAM系统中最基本的考核标准之一，是移动机器人执行多项任务的前提条件[225]。引入环境语义信息可以有效改善视觉SLAM定位中的尺度不确定性和累积漂移，从而不同程度地提高定位精度。226]。

鲍曼等人。[177]提出了传感器状态估计和语义地标位置优化问题，该问题集成了度量信息、语义信息和数据关联。从目标检测中获得语义信息后，他们引入了期望最大化（EM），并根据语义分类的结果计算数据关联的概率。他们成功地将语义SLAM转化为概率问题，提高了SLAM系统的定位精度。然而，本文中有许多强有力的假设。比如物体三维中心的投影要接近检测网络的中心，这在实际中不容易满足。

2020 年，赵等人。[227]西安交通大学提出了一种针对大规模室外环境的标志性视觉语义 SLAM 系统。其核心是将ORB-SLAM中的3D点云与卷积神经网络模型PSPNET-101中的语义分割信息相结合。它可以构建大规模环境的 3D 语义图。他们提出了一种将真实地标与点云地图关联起来的方法。它将建筑地标与语义点云关联起来，并将从 Google 地图获得的地标与用于城市区域导航的语义 3D 地图关联起来。借助语义点云，系统在无需GPS信息的大范围室外环境下实现基于地标的重定位。其流程如图20。2018年，苏黎世联邦理工学院提出VSO[228] 基于语义信息的自动驾驶场景。该方案解决了室外光照变化环境下的视觉SLAM定位问题。它在语义信息与图像之间建立约束，利用语义信息不受视角、尺度、光照影响的优点。同样，斯滕堡等人。[229] 还针对此类问题提出了解决方案。

在轨迹估计方面，几何特征只能为相机位姿提供短期约束，会在大范围内产生较大偏差

的环境。相比之下，物体作为更高层次的特征，当光照强度、观察距离和角度变化时，可以保持其语义信息不变。例如，桌子在任何光线和角度下仍然是桌子，其更稳定的性能可以为相机姿势提供长期约束。此外，语义SLAM可以有效解决传统视觉SLAM对光照变化敏感、干扰系统定位鲁棒性的问题。我们认为 VSLAM 定位本质上是相机姿态估计。语义信息可以提高传统 VSLAM 系统在强光照和高相机旋转下的定位精度。然而在实际应用中，语义信息的引入不可避免地会减慢整个系统的运行速度，这是VSLAM亟待解决的问题。我们认为，在大多数情况下，传统的VSLAM在定位精度方面仍然表现良好。然而，语义帮助VSLAM系统提高定位精度也值得研究。桌子12比较了 VSLAM 定位的传统方法和语义方法之间的差异。

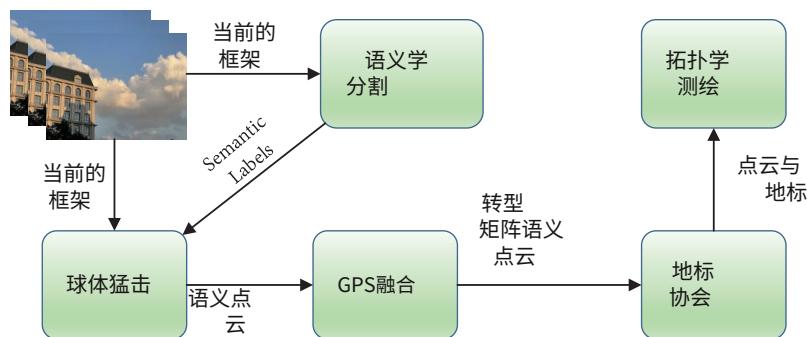


图 20.赵等人。提出了一种使用语义信息的大规模室外定位过程。

表 12.VSLAM 定位的传统方法与语义方法的比较。

方法	特征
传统的	对极几何，透视-n 点，迭代最近点，光流……
语义学	语义标签、数据关联

几何特征只能提供相机姿势的短期约束
在强光和快速运动的环境中可能会失败。
当光强、观察距离、角度变化时，语义信息可以保持不变。

4.2.3. 语义与映射

VSLAM和深度学习的另一个关键节点是SLAM的语义图构建，大多数语义VSLAM系统都是基于这个思想[230]。对于机器人来说，要像人类一样了解环境并从一个地方到另一个地方执行不同的任务，需要比几何地图所能提供的技能不同的技能。第231章]。机器人应该有能力以人类为中心理解其环境。未来需要区分房间和走廊，或者厨房和客厅的不同功能[第232章]。因此，涉及人类的语义属性

概念（例如房间类型、物体及其空间布局），被认为是未来机器人的必要属性[233]。近年来，随着深度学习的快速发展，包含语义信息的语义图逐渐进入人们的视野[234]。语义SLAM系统中的语义地图使机器人能够获取环境特征点等几何信息。此外，它还识别环境中的对象并获取位置、属性和类别等语义信息。与传统VSLAM构建的地图相比，机器人可以具备感知能力。对于机器人来说，处理这些问题具有重要意义

复杂的环境和完整的人机交互[235]。语义地图构建是SLAM研究的热点之一[236]。2005年, Galindo 等人。[第237章]提出了语义图的概念。如图21, 它由两个并行的层表示: 空间表示和语义表示。它为机器人提供了类似于人类对环境的推理能力(例如, 卧室是包含床的房间)。后来, 瓦苏代万等人。[238]进一步加强了人们对语义图的理解。

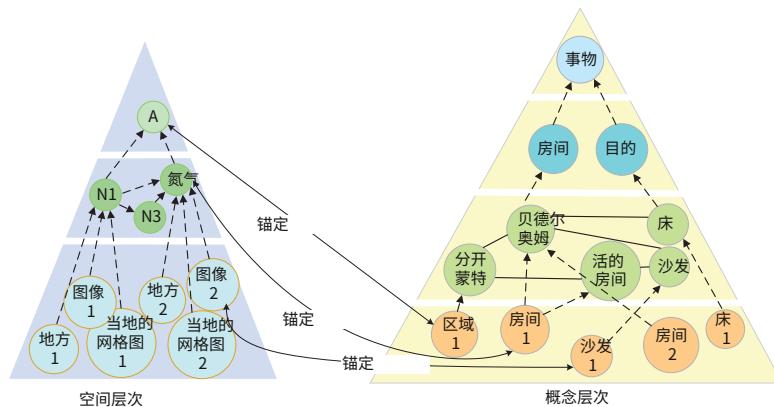


图 21.Galindo 文章中提到的语义图概念。

近年来, 深度学习技术发展迅速。越来越多的研究人员将深度学习与SLAM技术结合起来。他们使用目标检测、语义分割和其他算法来获取有关环境的语义信息。此外, 将其集成到环境地图中以构建环境

环境语义图[239]。如图22语义图构建的研究主要分为两个方向: 面向场景的语义图构建和面向对象的语义图构建。



图 22.不同类型的语义地图。

大多数面向场景的语义地图都是基于深度学习方法, 将2D语义信息映射到3D点云。面向场景的语义地图可以帮助机器人更好地理解其环境[240]。2020年, MIT提出Kimera[第241章]。这是一种成熟的面向场景的语义SLAM算法。参考号 [第242章] 提出了一种面向场景的语义图构建算法。基于RTABMAP [243], YOLO用于目标检测。在粗略估计出物体的位置后, 他们使用Canny算子检测深度图像中目标物体的边缘。然后他们通过基于区域的边缘处理实现了物体的精确分割

增长算法。他们通过非深度学习语义分割算法, 解决了传统语义图构建中计算资源消耗大的问题, 实时构建了面向场景的语义图。面向场景的语义图将帮助机器人更好地理解环境, 构建更具表现力的机器人

环境图。然而，这种方法无法为机器人认识环境提供更多的帮助，阻碍了机器人与个体的环境进行交互，在一定程度上限制了机器人的智能化程度。[\[244\]](#)。另外，此类算法需要对场景中的物体进行像素级的语义分割，导致系统计算量大、实时性低。因此，一些学者转向面向对象的语义图构建算法[\[245\]](#)。

面向对象的语义图是指只包含部分实例语义信息的图，语义信息以聚类的方法独立存在[\[246\]](#)。这种类型的地图允许机器人操作和维护地图上每个实体的语义信息。这样更有利于机器人理解环境并与环境中的实体进行交互，提高地图的实用性[\[第247章\]](#)。参考[\[45\]](#)提出了一种基于mask-RCNN和KinectFusion算法的基于体素的语义视觉SLAM系统。通过Mask-RCNN算法进行目标检测后，将目标检测结果与基于体素前景理论的TSDF模型融合，构建面向对象的语义图。虽然保证了检测的准确性，但仍然无法解决算法实时性差的问题。参考号[\[248\]](#)提出了一种轻量级的面向对象的SLAM系统，有效解决了数据关联和姿态估计的问题，解决了上述方法实时性较差的问题。核心框架基于ORB-SLAM2开发，并使用YOLOv3作为对象检测器来融合语义线程。在追踪器线程中，融合边界框、语义标签和点云信息，构建面向对象的半稠密语义图。实验结果表明，与ORB-SLAM2相比，该方案能够处理复杂环境中不同尺度和方向的多类物体，能够更好地表达环境。然而，对于一些大型物体，精确的姿态估计是不可能的。同样，伦敦大学学院提出了DSP-SLAM[\[249\]](#)。

目前大多数语义图构建方法需要同时处理实例分割和语义分割，这导致系统的实时性较差[\[250\]](#)。桌子13列出了一些语义图构建工作。另外，在处理动态对象时，大多数算法都是通过消除动态对象来实现系统的鲁棒性，这会使系统丢失很多有用的信息。因此，面向动态场景的SLAM是一个亟待解决的问题[\[第251章\]](#)。

表 13.优秀语义映射算法的一部分。

参考	年	传感器	语义标签	地图	贡献
维内特等人。 [第252章]	2015年	S	随机森林	体素	第一个系统可以实时执行密集、大规模、户外场景的语义重建。使用时间信息和高阶团来强制每个图像标记结果的标记一致性。
赵等人。 [第253章]	2016年	D	支持向量机	体素	
李等人。 [第254章]	2016年	D	深度实验室v2	体素	不需要为序列中的每个帧获得语义分割。
语义融合 [46]	2016年	D	CNN 与 CRF	瑟菲尔	允许 CNN 从多个视点进行的语义预测概率性地融合成密集的语义注释图。
杨等人。 [255]	2017年	S	CNN 与 CRF	网格	此外，通过新颖的 CRF 模型优化 3D 网格标签。
全景视融合 [256]	2020年	D	带有 CRF Mask 的 PSPNET 带有 CRF 的 R-CNN	体素	一种新颖的事物层面的在线体积语义映射系统。
基梅拉 [第241章]	2020年	S+我	逐像素	网	它是模块化的，允许替换每个模块或单独执行它们。
AVP-SLAM [161]	2020年	中+我+E	优网	体素	自动停车。
路线图 [第257章]	2021年	R+M+I+E	美国有线电视新闻网	体素	车载测绘、云端维护、用户端定位的框架。

传感器：S代表立体摄像头；M代表单目相机；I代表IMU；E代表编码器；R代表RTK-GPS，D代表RGB-D相机。

4.2.4. 消除动态对象

传统的VSLAM算法假设环境中的物体是静态或低运动的，这影响了VSLAM系统在实际场景中的适用性[258]。当环境中存在动态物体（如人、车辆、宠物）时，它们会给系统带来错误的观测数据，降低系统的准确性和鲁棒性[第259章]。传统方法通过RANSAC算法解决一些异常值对系统的影响。然而，如果动态物体占据了大部分图像区域或者运动物体速度很快，仍然无法获得可靠的观测数据[260]。如图23，由于动态物体，相机无法准确捕捉数据。因此如何解决动态物体对SLAM系统的影响成为许多研究人员的目标。

目前，针对动态物体给SLAM系统带来扰动问题的解决方案是一致的。即在视觉里程计之前，采用目标检测和图像分割算法来过滤掉图像中的动态区域。然后使用静态环境点计算相机附近的位置并构建包含语义信息的地图[261]。数字24显示出典型的结构。虽然不能完全解决动态物体的影响，但系统的鲁棒性大大提高。

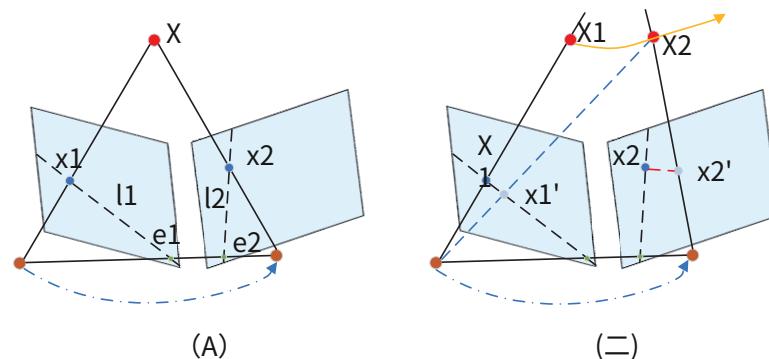


图 23. 传统方法利用几何约束来判断物体是否运动。例如，在（A），X是空间中的静止点，因此可以顺利地得到空间变换关系。在（B），空间点X1运动到X2后会带来系统误差。

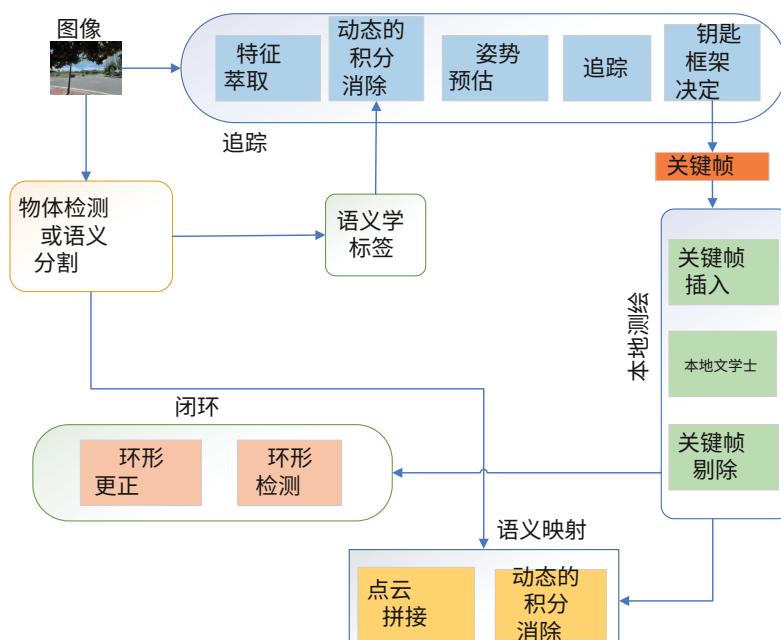


图 24. 典型的动态对象移除框架。

2018 年, Bescos 等人。[\[262\]](#) 基于ORB-SLAM2提出了动态场景视觉SLAM的DynaSLAM 算法。该系统提供单目、立体和 RGB-D 相机的接口。对于单目和立体相机, 使用MASK-RCNN 对每帧中的动态物体进行分割, 以避免SLAM系统中动态物体的特征提取。如果使用RGB-D相机, 则使用多视图几何的方法进行更精确的运动分割。动态片段将从当前帧和地图中删除。但是, 此方法选择删除所有可能移动的物体, 例如停放的汽车。这可能会导致剩余的静止特征点太少并影响相机姿态估计。同年, 清华大学团队提出了完整的SLAM系统DS-SLAM[\[263\]](#) 基于 ORB-SLAM2。其核心是在ORB-SLAM2中添加了语义网络分割, 并作为单独的线程实时运行。它可以去除场景中的物体动态分割并创建单独的线程来构建密集的语义八叉树图, 以帮助机器人实现更高级别的任务。

一些方法使用语义信息来隐藏被认为是动态的对象。虽然此类方法在一定程度上改善了动态对象对系统的影响, 但一刀切的做法可能会导致系统丢失许多有用的特征点。例如, 停在路边的一辆车可以被视为一个动态物体, 其携带的所有特征点都被过滤掉[\[264\]](#)。然而, 停在路边的汽车可以用作系统中的可靠特征点。然而, 它甚至可以成为高质量特征点的主要来源。参考 [\[265\]](#)提出将语义信息集成到传统的VSLAM方法中。该方法不需要运动检测。置信度的引入, 赋予每个物体不同的可能运动概率, 来判断物体是否在运动。此外, 语义标签分布与地图点观测一致性相结合, 以估计每个 3D 点测量的可靠性。然后将其用于位姿估计和优化步骤的地图中。此方法可以处理被视为动态但静止的对象, 例如停在路边的汽车。参考 [\[266\]](#)基于光流方法去除动态物体。其核心思想基于ORB-SLAM2。在其前端, 四个 CNN 神经网络用于同时预测每帧的深度、姿态、光流和语义掩模。通过计算深度和姿态合成的刚性光流并与估计光流进行比较, 得到初始运动区域。该算法能够将运动物体与当前场景区分开来, 并保留静态物体的特征点。避免仅根据类别属性去除运动物体, 导致SLAM系统跟踪失败。文章[\[第267章\]](#)提出了一种基于 ORB-SLAM2 构建的视觉 SLAM 系统, 借助 Mask-RCNN 获得的语义信息和 RGB-D 相机提供的深度信息, 通过丢弃运动特征点, 在动态环境中鲁棒且准确地执行。该方法试图通过找出从可移动物体中提取的静态特征点来利用更可靠的特征点进行相机位姿估计, 这在静态物体无法在场景中提供足够的特征点时受益匪浅。

语义信息可以更好地帮助系统解决动态对象带来的计算资源消耗较高的干扰。但现有方案普遍实时性不够, 无法广泛推广到实用机器人, 应用场景受到很大限制。[\[268\]](#)。此外, 语义信息可能无法以相机帧速率获得, 或者可能并不总是提供准确的数据[\[269\]](#)。将图像区域分配给错误的语义类可能会不必要地将其排除在姿势估计之外, 这在稀疏纹理的环境中可能至关重要[\[270\]](#)。目前该问题的解决方案主要集中在使用光流等方法来检测场景中移动的物体[\[第271章\]](#)。现有算法虽然在数据集上取得了不错的效果, 但在实际工程中并没有取得非常可靠的结果。桌子[14](#)展示了近年来利用深度神经网络改善动态环境的VSLAM算法。

表 14. 近年来一些优秀的针对动态场景的VSLAM算法。

模型	年	传感器	场景	动态检测	数据集	代码资源
雷迪等人。[第272章]	2016年	立体声	户外的	[273]	基蒂	0
DynaSLAM [262]	2018年	单目/立体/ RGB-D	室外/室内	Mask R-CNN	KITTI/TUM RGB-D	[274]
DS-SLAM [263]	2018年	RGB-D	室内的	段网	TUM RGB-D	[275]
检测-SLAM [276] 王 等人。[278] 斯拉曼 蒂克 [265] 动态 SLAM [280]	2018年 2019年 2019年 2018年	RGB-D RGB-D 单目/立体 立体声	室内的 室内的 户外的 户外的	固态硬盘 YOLOv3 Mask R-CNN 级联[第281章]	TUM RGB-D 纽约大学深度数据集 V2	[第277章] 0
STDyn-SLAM [第283章]	2022年	立体声	户外的	网段网络	基蒂	[第284章]
姿势融合 [第285章] RDS- SLAM [第287章] YO- SLAM [第289章]	2018年 2021年 2021年	RGB-D RGB-D RGB-D	室内的 室内的 室内的	开放姿势 SegNet/Mask R-CNN 酸奶	弗莱堡 RGB-D SLAM TUM RGB-D TUM RGB-D	[第286章] [288] 0
张等人。[290] DOE- SLAM [第292章]	2021年	全景 单目	数据 室内的	酸奶 自发*	[第291章] TUM RGB-D	0 0
DRSO-SLAM [293]	2021年	RGB-D	室内的	Mask R-CNN	TUM RGB-D	0
DDL-SLAM [294]	2020年	RGB-D	室内的	DUNet	TUM RGB-D	0
RDMO-SLAM [295]	2021年	RGB-D	室内的	Mask R-CNN	TUM RGB-D	0

代码资源：0表示无代码资源。动态检测：[第242章] 代表请参考本文；self-initiated * 代表参考本文提出的方法。

五、结论与展望

同时定位和建图是机器人领域的一个主要研究问题，该领域已经投入了大量的精力来开发新方法以最大限度地提高其鲁棒性和可靠性。基于视觉的SLAM技术经历了多年的发展，涌现出许多优秀的算法，并成功应用于机器人、无人机等各个领域。深度学习的快速发展推动了计算机领域的创新，两者的结合成为一个活跃的研究领域。因此，VS-LAM的研究受到越来越多的关注。此外，随着智能时代的到来，对移动机器人的自主性提出了更高的要求。为了实现机器人的高级环境感知，语义VSLAM被提出并迅速发展。传统的VSLAM在构建环境地图时仅恢复环境的几何特征，无法满足机器人导航、人机交互、自主探索等应用的需求。但早期的语义图构建方法普遍采用模型库匹配方式，需要提前构建对象模型库，局限性很大，不利于推广应用。随着计算机性能的提高和深度学习技术的快速发展，VSLAM技术与深度学习技术相结合，填补了传统VSLAM系统的不足。近年来，深度学习技术作为最具前景和优势的计算机视觉处理方法，受到SLAM研究者的广泛关注。在语义SLAM系统中，可以通过深度学习技术从预训练图像集和实时感知图像集直接学习环境语义信息。它还可以用来更好地利用大型数据集，赋予系统更大的泛化能力。语义SLAM系统在构建语义地图时，可以利用深度学习方法对环境中的物体进行检测和分类，构建信息更丰富的地图，具有更好的实用性。

在本文中，我们研究了大多数最先进的视觉 SLAM 解决方案，这些解决方案使用功能来定位机器人并绘制其周围环境的地图。我们根据基于特征的视觉SLAM方法所依赖的特征类型对它们进行分类；传统的VSLAM和与深度学习相结合的VSLAM。每个类别的优点和缺点都经过彻底调查，并且在适用的情况下，每个解决方案克服的挑战

被突出显示。这项工作证明了使用视觉作为解决 SLAM 问题的唯一外部感知传感器的重要性。这主要是因为相机是一种理想的传感器，因为它重量轻、无源、低功耗，并且能够捕获有关场景的丰富而独特的信息。然而，由于移动的人或物体、无特征区域的幻影、白天和黑夜之间的过渡或任何其他不可预见的情况，视觉的使用需要在可变照明条件下具有良好性能和一致性的可靠算法。因此，使用视觉作为唯一传感器的 SLAM 系统仍然是一个具有挑战性和前景的研究领域。图像匹配和数据关联仍然分别是计算机视觉和机器人视觉领域的开放研究领域。探测器和描述符的选择直接影响系统跟踪显着特征、识别先前看到的区域、构建一致的环境模型和实时工作的性能。尽管数据库不断增长且环境不断变化且复杂，但数据关联尤其需要长期导航。接受不良关联将导致整个 SLAM 系统出现严重错误，这意味着位置计算和地图构建将不一致。

此外，我们还重点介绍了融合语义信息的 VSLAM 的发展。结合语义信息的VSLAM系统在鲁棒性、精度和高水平感知方面取得了更好的效果。语义VLSAM的研究将受到更多关注。语义VSLAM将从根本上提高机器人的自主交互能力。

结合其他研究，我们对VSLAM的未来发展做出如下展望：

(1) 工程应用。经过几十年的发展，VSLAM已经得到了广泛的应用应用于机器人等多个领域。然而SLAM对环境光照、高速运动、运动干扰等问题比较敏感，因此如何提高系统的鲁棒性、长时间构建大规模地图都是值得挑战的。SLAM使用的两个主要场景是基于智能手机或无人机等嵌入式平台，以及3D重建、场景理解和深度学习。如何平衡实时性和准确性是一个重要的悬而未决的问题。针对动态、非结构化、复杂、不确定和大规模环境的解决方案仍有待探索。

(2) 理论支撑。通过深度学习学到的信息特征仍缺乏直观的意义和明确的理论指导。目前深度学习主要应用于SLAM的局部子模块，如深度估计、闭环检测等。然而，如何将深度学习应用到整个SLAM系统中仍然是一个很大的挑战。传统的VSLAM在定位导航方面仍然有优势。虽然传统方法的一些模块通过深度学习进行了改进，但深度学习的范围普遍不广，在某些数据集上可能取得不错的效果，但在另一个场景下可能会不稳定。定位建图过程涉及大量数学公式，而深度学习在处理数学问题上存在弊端，同时利用深度学习进行相关训练的数据较少，这种方法较为传统。SLAM框架并没有呈现出显着的优势，目前还没有可用。SLAM技术的主要算法。未来，SLAM将逐步吸收深度学习方法并完善训练数据集，以提高定位建图的准确性和鲁棒性。

(3) 高层次环境信息感知与人机交互行动。随着深度学习的进一步发展，语义VSLAM的研究和应用将有巨大的发展空间。未来智能时代，人们对 智能自主移动机器人的需求将快速增长。如何利用语义VSLAM技术更好地提高机器人的自主能力将是一项长期而艰巨的任务。尽管近年来已经取得了一些优秀的成果，但与经典的VSLAM算法相比，语义VSLAM仍处于发展阶段。目前，语义SLAM的开源解决方案并不多，语义SLAM的应用还处于初级阶段，主要是因为准确的语义地图的构建需要大量的计算资源。这

严重影响了SLAM的实时性。未来随着硬件水平的不断提高，SLAM系统实时性差的问题可能会得到很大改善。

(四) 建立健全评价体系。语义VSLAM技术已发展

近年来迅速发展。但与传统的VSLAM相比，暂时还没有完善的评价标准。在SLAM系统研究中，一般采用ATE或RPE来评估系统性能。然而，这两种评价标准都是基于SLAM系统的位姿估计结果，对于建图效果尚无普遍认可的可靠评价标准。对于语义SLAM系统来说，如何评价语义信息获取的准确性以及如何评价语义地图构建的效果，是语义SLAM系统评价标准中应该考虑的问题。而且，仅靠主观指标来评价也不是长久之计。未来，如何建立系统的语义VSLAM评价指标将是一个热点话题。

作者贡献：概念化、WC、KH和GS；方法论、WC、KH和GS；软件，CZ和XW；形式分析、WC、KH和GS；调查、WC、KH和AJ；写作——初稿准备，GS；写作——复习WC、KH和GS；编辑、WC、KH和GS；可视化、AJ、XW、CX和ZL；监督、WC、KH和AJ；项目管理，

WC、KH和AJ；资金收购，AJ所有作者均已阅读并同意稿件的出版版本。

资金：该工作得到了国家重点研发计划（批准号：2019YFB1309600）、国家自然科学基金（批准号：51875281和51861135306）的支持。这些资金来自AJ

数据可用性声明：不适用。

致谢：本文的研究得到了国家重点研发计划（批准号：2019YFB1309600）、国家自然科学基金（批准号：51875281和51861135306）的支持。作者谨向对本文提出宝贵修改的审稿人和编辑表示衷心的感谢。

利益冲突：作者宣称没有利益冲突。

缩写

本手稿中使用了以下缩写：

SLAM	刺激定位与映射 视觉刺激定位与映射 卷积神经网络
视觉SLAM	
美国有线电视新闻网	
循环神经网络	循环神经网络惯性测量单元
惯性测量单元	
埃沃	用于评估里程计和SLAM绝对轨迹误差的Python包
吃	
视网膜色素上皮	相对位姿误差
慕尼黑工业大学	慕尼黑工业大学飞行时间
到F	
中央处理器	中央处理单元 图形处理单元 二进制字
图形处理器	袋 无味卡尔曼滤波器 迭代最近点 截断
弓	符号距离函数 视觉里程计
英国联合基金会	
ICP	
TSDF	
视氧浓度	
维奥	视觉惯性里程计深度神经网络
深度神经网络	
长短期记忆网络	长短期记忆网络门控循环单元
格鲁乌	

3D 三维
 EM 期望最大化
 麻省理工学院
 无人机

参考

1. 史密斯, RC; Cheeseman, P. 关于空间不确定性的表示和估计。[交叉引用] 国际。J.机器人。资源。1986年,5, 56–68。
2. 邓, G.; 李, J. 李, W. Wang, H. SLAM: 用于地图绘制的深度图像信息和用于定位的惯性导航系统。2016 年亚太智能机器人系统会议 (ACIRS) 会议记录, 日本东京, 2016 年 7 月 20-22 日; 第 187–191 页。
3. 崔L.; Ma, C. SOF-SLAM: 动态环境的语义视觉 SLAM。IEEE 接入2019年,7, 166528–166539 [交叉引用]
4. 布列松, G.; 阿尔赛德, Z.; 于L.; Glaser, S. 同步定位和地图绘制: 自动驾驶当前趋势调查。IEEE 传输。英特尔。呃。2017年,2, 194–220。[交叉引用]
5. 卡尔森, N.; 贝尔纳多, 艾德; 奥斯特罗斯基, J.; 贡萨尔维斯, L.; 皮尔贾尼安, P.; 慕尼黑, 缅因州 用于稳健定位和地图绘制的 vSLAM 算法。2005 年 IEEE 国际机器人与自动化会议论文集, 西班牙巴塞罗那, 2005 年 4 月 18-22 日; 第 24-29 页。
6. 赫斯, W.; 科勒, D.; 拉普, H.; Andor, D. 2D LiDAR SLAM 中的实时闭环。2016 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 瑞典斯德哥尔摩, 2016 年 5 月 16-21 日; 第 1271–1278 页。
7. 格里塞蒂, G.; 斯塔尼斯, C.; Burgard, W. 通过自适应提议和选择性重采样, 使用 Rao-Blackwellized 粒子滤波器改进基于网格的 SLAM。2005 年 IEEE 国际机器人与自动化会议论文集, 西班牙巴塞罗那, 2005 年 4 月 18-22 日; 第 2432–2437 页。
8. 可汗, 穆; Zaidi, SAA; 伊什蒂亚克, A.; 布哈里, 南苏丹; 萨默, S.; Farman, A. LiDAR-SLAM 和基于 LiDAR 的传感器技术的比较调查。2021 年穆罕默德·阿里·真纳大学国际计算机会议 (MAJICC) 会议记录, 巴基斯坦卡拉奇, 2021 年 7 月 15 日至 17 日; 第 1-8 页。
9. 古普塔, A.; Fernando, X. 无人机中的同步定位与建图 (SLAM) 和数据融合: 最新进展和挑战。无人机2022年,6, 85. [交叉引用]
10. 阿伦, A.; Nirmaladevi, P. 对当前用于提高 CBIR 性能的语义层算法的调查。材料科学与工程会议系列论文集, 印度金奈, 2021 年 2 月 1 日; p. 012118。
11. 布尔格拉, A.; 博宁-丰特, F.; 字体, EG; Torres, AM 将深度学习与稳健估计相结合, 实现离群值弹性水下视觉图 SLAM。J. Mar. Sci. 工程师。2022年,10, 511. [交叉引用]
12. 阿拉蒂斯, MB; Hancke, GP 对自主移动机器人和传感器融合方法的挑战的回顾。IEEE 接入2020年,8, 39830–39846。[交叉引用]
13. 王P.; 程, J.; Feng, W. 利用 3D 信息高效构建拓扑语义图。J · 英特尔。模糊系统2018年,35, 3011–3020。[交叉引用]
14. 王S.; 克拉克, R. 温, H. Trigoni, N. 通过深度神经网络进行端到端、序列到序列的概率视觉里程计。国际。J.机器人。资源。2017年, 37, 513–542。[交叉引用]
15. 卡德纳, C.; 卡隆, L.; 卡里略, H.; 拉蒂夫, Y.; 斯卡拉穆扎, D.; 内拉, J.; 里德, I. Leonard, JJ 《同时定位和绘图的过去、现在和未来: 迈向稳健感知时代》。IEEE 传输。机器人。2016年,32, 1309–1332。[交叉引用]
16. 竹富, T.; 内山, H.; Ikeda, S. Visual SLAM 算法: 2010 年至 2016 年的一项调查。IPSJ 翻译。计算。维斯。应用。2017年,9, 16. [交叉引用]
17. 优素福, K.; 巴布-哈迪亚沙尔, A.; Hoseinnezhad, R. 视觉里程计和视觉 SLAM 概述: 移动机器人的应用。英特尔。工业系统2015年,1, 289–311。[交叉引用]
18. 巴夫勒, H.; 桑切斯-洛佩斯, JL; 施密特, EF; Voos, H. 从 SLAM 到态势感知: 挑战与调查。arXiv2021年, arXiv: 2110.00273。
19. Servières, M.; 雷诺丁, V.; 杜普伊斯, A.; Antigny, N. 视觉和视觉惯性 SLAM: 最先进的技术、分类和实验基准。J. Sens. 2021年, 2021年, 2054828。[交叉引用]
20. 阿扎姆, R.; 塔哈, T.; 黄, S. Zweiri, Y. 基于特征的视觉同步定位和建图: 一项调查。SN 应用程序。科学。2020年,2, 224。[交叉引用]
21. 马卡里奥·巴罗斯, A.; 米歇尔, M.; 莫林, Y.; 科雷, G.; Carrel, F. 视觉 SLAM 算法的综合综述。机器人技术2022年,11, 24. [交叉引用]
22. 李, R.; 王, S. Gu, D. 视觉 SLAM 从几何到深度学习的持续演变: 挑战和机遇。认知。计算。2018年,10, 875–889。[交叉引用]
23. Medeiros Esper, I.; 斯莫尔金, O.; 曼科, M.; 波波夫, A.; 来自 PJ; Mason, A. 用于 3D 重建的 RGB-D 多相机姿势估计的评估。应用。科学。2022年,12, 4134。[交叉引用]
24. 左, Y.; 杨, J. 陈, J. 王X; 王, Y. Kneip, L. DEVO: 具有挑战性的条件下的深度事件相机视觉里程计。arXiv2022年, arXiv: 2202.02556。
25. EVO。Python。在线提供: <https://github.com/MichaelGrupp/evo> (2022 年 4 月 25 日访问)。

26. 博丹, B.; 瓦格斯塔夫, H.; 萨埃迪, S.; 纳尔迪, L.; 韦斯帕, E.; 莫尔, J.; 尼斯贝特, A.; 卢扬, M.; 弗伯, S.; 戴维森, AJ; 等人。SLAMBenchmark2: 视觉 SLAM 的多目标头对头基准测试。2018 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 澳大利亚布里斯班, 2018 年 5 月 21-25 日; 第 3637–3644 页。
27. 惠兰, T.; 萨拉斯·莫雷诺, RF; 格洛克尔, B.; 戴维森, AJ; Leutenegger, S. ElasticFusion: 实时密集 SLAM 和光源估计。国际。J. 机器人。资源。2016年,35, 1697–1716。[交叉引用]
28. Mur-Artal, R.; Tardós, JD ORB-SLAM2: 用于单目、立体和 RGB-D 相机的开源 SLAM 系统。IEEE 传输。机器人。2017年,33, 1255–1262。[交叉引用]
29. 洛特内格, S.; 林恩, S.; 博塞, M.; 西格沃特, R.; Furgale, P. 使用非线性优化的基于关键帧的视觉惯性里程计。国际。J. 机器人。资源。2014年,34, 314–334。[交叉引用]
30. TUM RGB-D。在线提供: <https://vision.in.tum.de/data/datasets/rbgd-dataset> (2022 年 4 月 25 日访问)。
31. TUM MonoVo。在线提供: <http://vision.in.tum.de/mono-dataset> (2022 年 4 月 25 日访问)。
32. 慕尼黑工业大学六世。在线提供: <https://vision.in.tum.de/data/datasets/visual-inertial-dataset> (2022 年 4 月 25 日访问)。
33. 基蒂。在线提供: <http://www.cvlibs.net/datasets/kitti/> (2022 年 5 月 22 日访问)。
34. EuRoc。在线提供: <https://projects.asl.ethz.ch/datasets/doku.php?id=kmavvisualinertialdatasets> (2022 年 4 月 25 日访问)。
35. 城市景观。在线提供: <https://www.cityscapes-dataset.com/> (2022 年 4 月 25 日访问)。
36. ICL-NUIM。在线提供: <https://www.doc.ic.ac.uk/ahanda/VaFRIC/iclnuim.html> (2022 年 4 月 25 日访问)。
37. 纽约大学RGB-D。在线提供: <https://cs.nyu.edu/silberman/datasets/> (2022 年 4 月 25 日访问)。
38. 可可女士。在线提供: <https://paperswithcode.com/dataset/coco> (2022 年 4 月 25 日访问)。
39. 恩格尔, J.; 科尔通, V.; Cremers, D. 直接稀疏里程计。IEEE 传输。模式肛门。马赫。英特尔。2018年,40, 611–625。[交叉引用] [考研]
40. 恩格尔, J.; 舍普斯, T.; Cremers, D. LSD-SLAM: 大规模直接单目 SLAM。计算机视觉论文集 – ECCV 2014, 瑞士苏黎世, 2014 年 9 月 6–12 日; 第 834–849 页。
41. 尼科·森德豪夫。在线提供: <https://nikosuenderhauf.github.io/projects/sceneunderstanding/> (2022 年 6 月 12 日访问)。
42. 语义SLAM.ai。在线提供: <http://www.semantic slam.ai/> (2022 年 6 月 12 日访问)。
43. 帝国理工学院戴森机器人实验室。在线提供: <http://www.imperial.ac.uk/dyson-robotics-lab> (2022 年 4 月 25 日访问)。
44. 布洛施, M.; 查诺夫斯基, J.; 克拉克, R. 洛伊特内格, S.; Davison, AJ CodeSLAM - 学习密集视觉 SLAM 的紧凑、可优化表示。在 2018 年 IEEE/CVF 计算机视觉和模式识别会议论文集中, 美国犹他州盐湖城, 2018 年 6 月 18 日至 23 日; 第 2560–2568 页。
45. 麦科马克, J.; 克拉克, R. 布洛施, M.; 戴维森, A.; Leutenegger, S. Fusion++: 体积对象级 SLAM。2018 年 3D 视觉国际会议 (3DV) 会议记录, 意大利维罗纳, 2018 年 9 月 5–8 日; 第 32–41 页。
46. 麦考马克, J.; 汉达, A.; 戴维森, A.; Leutenegger, S. SemanticFusion: 使用卷积神经网络的密集 3D 语义映射。2017 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 新加坡, 2017 年 5 月 29 日至 6 月 3 日; 第 4628–4635 页。
47. 慕尼黑工业大学信息学系计算机视觉小组。在线提供: <https://vision.in.tum.de/研究> (2022 年 4 月 25 日访问)。
48. 杨, N.; 斯坦伯格, Lv; 王, R. Cremers, D. D3VO: 单目视觉里程计的深度、深度姿势和深度不确定性。2020 年 IEEE/CVF 计算机视觉和模式识别会议 (CVPR) 会议记录, 美国华盛顿州西雅图, 2020 年 6 月 13 日至 19 日; 第 1278–1289 页。
49. Stumberg, Lv; Cremers, D. DM-VIO: 延迟边缘化视觉惯性里程计。IEEE 机器人。自动。莱特。2022年,7, 1408–1415。[交叉引用]
50. 高X.; 王, R. 德梅尔, N.; Cremers, D. LDSO: 带环路闭合的直接稀疏里程计。2018 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 西班牙马德里, 2018 年 10 月 1–5 日; 第 2198–2204 页。
51. 弗莱堡自主智能系统大学。在线提供: http://ais.informatik.uni-freiburg.de/index_en.php (2022 年 5 月 23 日访问)。
52. 格里塞蒂, G.; 斯塔尼斯, C.; Burgard, W. 使用 Rao-Blackwellized 粒子滤波器改进网格映射技术。IEEE 传输。机器人。2007年,23, 34–46。[交叉引用]
53. 恩德雷斯, F.; 赫斯, J.; 斯特姆, J.; 克雷默斯, D.; Burgard, W. 使用 RGB-D 相机进行 3D 映射。IEEE 传输。机器人。2013年,30, 177–187。[交叉引用]
54. 香港科技大学空中机器人小组。在线提供: <https://uav.hkust.edu.hk/> (2022 年 4 月 25 日访问)。
55. 秦T.; 李, P. Shen, S. VINS-Mono: 一种稳健且多功能的单目视觉惯性状态估计器。IEEE 传输。机器人。2018年,34, 1004–1020。[交叉引用]
56. 秦T.; 潘, J. 曹S.; Shen, S. 基于优化的通用多传感器局部里程计估计框架。arXiv2019年, arXiv: 1901.03638。
57. 周, Y.; 加列戈, G.; Shen, S. 基于事件的立体视觉里程计。IEEE 传输。机器人。2021年,37, 1433–1450。[交叉引用]
58. 华盛顿大学机器人和状态估计实验室。在线提供: <http://rse-lab.cs.washington.edu/projects/> (2022 年 4 月 25 日访问)。
59. 施密特, T.; 纽科姆, R.; Fox, D. DART: 使用消费者深度相机进行密集铰接式实时跟踪。奥顿。机器人。2015年,39, 239–258。[交叉引用]
60. 向, Y.; Fox, D. DA-RNN: 与数据相关的循环神经网络的语义映射。arXiv2017年, arXiv: 1703.03098。

61. 亨利, P.; 克莱宁, M.; 赫布斯特, E.; 任, X. Fox, D. RGB-D 映射: 使用深度相机对室内环境进行密集 3D 建模。在实验机器人: 第十二届国际实验机器人研讨会, Khatib, O.、Kumar, V.、Sukhatme, G., 编辑; 施普林格: 德国柏林/海德堡, 2014 年; 第 477–491 页。
62. 萨拉戈萨大学机器人、感知和实时小组。在线提供: <http://robots.unizar.es/slamlab/> (2022 年 4 月 25 日访问)。
63. 加尔韦斯-洛佩斯, D.; 萨拉斯, M.; 塔多斯, JD; Montiel, JMM 实时单目物体 SLAM。机器人。奥顿。系统。2016年,75, 435–449。[交叉引用]
64. 塔赫里, H.; 夏, ZC SLAM; 定义和演变。工程师。应用。阿蒂夫。英特尔。2021年,97, 104032。[交叉引用]
65. 林L.; 王, W. 罗, W.; 宋L.; Zhou, W. 具有解耦相机姿态估计的无监督单目视觉里程计。数字。信号处理。2021年,114, 103052。[交叉引用]
66. 朱K.; 蒋X; 方, Z.; 高, Y. 藤田, H.; 黄, J.-N. 用于直接视觉里程计的光度传输。基于知识的系统2021年,213, 106671。[交叉引用]
67. 古克卢, O.; Can, AB k-SLAM: 一种适用于大型室内环境的快速 RGB-D SLAM 方法。计算。维斯。图像理解。2019年,184, 31–44。[交叉引用]
68. 蔡L.; 叶, Y. 高X.; 李, Z. 张, C. 基于仿射变换的 ORB 特征提取的改进视觉 SLAM。光学2021年,227, 165421。[交叉引用]
69. 哈里斯, C.; Stephens, M. 组合角点和边缘检测器。阿尔维愿景会议记录, 英国曼彻斯特, 1988 年 8 月 31 日至 9 月 2 日。
70. 罗斯滕, E.; Drummond, T. 用于高速角点检测的机器学习。计算机视觉会议录 - ECCV 2006, 奥地利格拉茨, 2006 年 5 月 7-13 日; 第 430–443 页。
71. 建波, S.; 托马西. 良好的跟踪功能。1994 年 IEEE 计算机视觉和模式识别会议论文集, 美国华盛顿州西雅图, 1994 年 6 月 21-23 日; 第 593–600 页。
72. Lowe, DG 来自尺度不变关键点的独特图像特征。国际。J. 计算机。维斯。2004年,60, 91–110。[交叉引用]
73. 贝, H.; 埃斯, A.; 图伊特拉尔斯, T.; Van Gool, L. 加速稳健特征 (SURF)。计算。维斯。图像理解。2008年,110, 346–359。[交叉引用]
74. 卢布, E.; 拉博, V.; 科诺利格, K.; Bradski, G. ORB: SIFT 或 SURF 的有效替代方案。2011 年计算机视觉国际会议论文集, 西班牙巴塞罗那, 2011 年 11 月 6-13 日; 第 2564–2571 页。
75. 阿里, AM; Nordin, 基于 MJ SIFT 的单目 SLAM, 具有用于室内导航的多云功能。TENCON 2010—2010 IEEE Region 10 会议记录, 日本福冈, 2010 年 11 月 21 日至 24 日; 第 2326–2331 页。
76. 吉奥伊, RGv; 雅库博维奇, J.; 莫雷尔, JM; Randall, G. LSD: 具有错误检测控制的快速线段检测器。IEEE 传输。模式肛门。马赫。英特尔。2010年,32, 722–732。[交叉引用]
77. 戴维森, AJ; 爱达荷州里德; 北达科他州莫尔顿; Stasse, O. MonoSLAM: 实时单相机 SLAM。IEEE 传输。模式肛门。马赫。英特尔。2007年,29, 1052–1067。[交叉引用]
78. 胡K.; 吴, J. 翁L.; 张, Y. 郑, F. 庞, Z. Xia, M. 一种基于联邦卡尔曼滤波器置信度的新型联邦学习方法。国际。J. 马赫. 学习。赛博恩。2021年,12, 3607–3627。[交叉引用]
79. 克莱因, G.; Murray, D. 小型 AR 工作空间的并行跟踪和映射。2007 年第六届 IEEE 和 ACM 国际混合与增强现实研讨会论文集, 日本奈良, 2007 年 11 月 13-16 日; 第 225–234 页。
80. 穆尔-阿塔尔, R.; 蒙蒂尔, JMM; Tardós, JD ORB-SLAM: 多功能且精确的单目 SLAM 系统。IEEE 传输。机器人。2015年,31, 1147–1163。[交叉引用]
81. 加尔韦斯·洛佩斯, D.; Tardos, JD 用于图像序列中快速位置识别的二进制词袋。IEEE 传输。机器人。2012年,28, 1188–1197。[交叉引用]
82. 坎波斯, C.; 埃尔维拉, R.; 罗德里格斯, JJG; 蒙蒂尔, JMM; Tardós, JD ORB-SLAM3: 用于视觉、视觉惯性和多地图 SLAM 的精确开源库。IEEE 传输。机器人。2021年,37, 1874年至1890年。[交叉引用]
83. 瓦希托夫, A.; 芬克, J.; Moreno-Noguer, F. 从点和线进行精确和线性的时间姿态估计。在计算机视觉—ECCV 2016; 施普林格: Cham, 瑞士, 2016; 第 583–599 页。
84. 史密斯, P.; 爱达荷州里德; Davison, AJ 直线实时单目 SLAM。BMVC2006年,6, 17–26。
85. 佩尔迪斯, E.; 洛佩斯, LM; Cañas, JM LineSLAM: 使用线条和 UKF 进行视觉实时定位。在 ROBOT2013: 第一届伊比利亚机器人会议: 机器人技术进展, Armada, MA、Sanfeliu, A.、Ferre, M., 编辑; 施普林格国际出版社: Cham, 瑞士, 2014 年; 第一卷, 第 663–678 页。
86. 蒙特罗, AS; 纳亚克, A.; 斯托伊梅诺维奇, M.; Zagaria, N. 基于图像轮廓上重复片段方向的鲁棒线提取。2009 年 IEEE 安全与国防应用计算智能研讨会论文集, 加拿大安大略省渥太华, 2009 年 7 月 8-10 日; 第 1-7 页。
87. 克莱因, G.; Murray, D. 提高基于关键帧的 SLAM 的敏捷性。计算机视觉会议录 - ECCV 2008, 法国马赛, 2008 年 10 月 12-18 日; 第 802–815 页。
88. 普马罗拉, A.; 瓦希托夫, A.; 阿古多, A.; 桑菲柳, A.; Moreno-Noguer, F. PL-SLAM: 带有点和线的实时单目视觉 SLAM。2017 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 新加坡, 2017 年 5 月 29 日至 6 月 3 日; 第 4503–4508 页。
89. 戈麦斯-奥赫达, R.; 莫雷诺, 足协; 祖尼加-诺埃尔, D.; 斯卡拉穆扎, D.; Gonzalez-Jimenez, J. PL-SLAM: 通过点和线段组合的立体 SLAM 系统。IEEE 传输。机器人。2019年,35, 734–746。[交叉引用]

- 90.哎呀, 美联社; 契赫洛夫, D.; 卡尔威, A.; Mayol-Cuevas, W. 发现视觉 SLAM 中的更高层次结构。*IEEE 传输。机器人。* 2008年, 24, 980–990。[交叉引用]
- 91.李H.; 胡, Z. Chen, X. PLP-SLAM: 一种基于点线面特征融合的视觉SLAM方法。*机器人*2017年, 39, 214–220。
- 92.张N.; 赵, Y. 使用点和线的快速稳健的单目视觉惯性里程计。*传感器*2019年, 19, 4545。[交叉引用]
- 93.他, X.; 高, W. 盛, C.; 张, Z. 潘, S. 段, L.; 张, H. 卢欣。基于优化视觉点线特征的激光雷达视觉惯性里程计。*遥感。* 2022年, 14, 622。[交叉引用]
- 94.纽科姆, RA; 洛夫格罗夫, SJ; Davison, AJ DTAM: 实时密集跟踪和地图绘制。2011 年计算机视觉国际会议论文集, 西班牙巴塞罗那, 2011 年 11 月 6-13 日; 第 2320–2327 页。
- 95.福斯特, C.; 皮佐利, M.; Scaramuzza, D. SVO: 快速半直接单目视觉里程计。2014 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 中国香港, 2014 年 5 月 31 日至 6 月 7 日; 第 15-22 页。
- 96.张杰; 加尼什, P.; 沃勒, K.; 威利斯, A.; Brink, K. 低带宽和计算限制 RGB-D 平面语义 SLAM。*传感器*2021年, 21, 5400。[交叉引用]
[考研]
- 97.菲拉托夫, A.; 扎斯拉夫斯基, M.; Krinkin, K. 多无人机 3D 建筑重建方法。*数学*2021年, 9, 3033。[交叉引用]
- 98.纽科姆, RA; 伊扎迪, S.; 希利格斯, O.; 莫利诺, D.; 金, D. 戴维森, AJ; 科希, P.; 肖顿, J.; 霍奇斯, S.; Fitzgibbon, A. KinectFusion: 实时密集表面映射和跟踪。2011 年第 10 届 IEEE 国际混合与增强现实研讨会论文集, 瑞士巴塞尔, 2011 年 10 月 26-29 日; 第 127–136 页。
- 99.凯斯, M.; 法伦, M.; 约翰逊, H.; Leonard, J. Kintinous: 空间扩展的 kinectfusion。摘自 RSS 研讨会 RGB-D: 使用深度相机进行高级推理, 澳大利亚悉尼, 2012 年 7 月 9-10 日; p. 9.
- 100.Kinect融合。在线提供: <https://github.com/ParikaGoel/KinectFusion> (2022 年 4 月 21 日访问)。
- 101.连续的。在线提供: <https://github.com/mp3guy/Kintinously> (2022 年 4 月 21 日访问)。
- 102.RGB-DSLAMv2。在线提供: https://github.com/felixendres/rbgdslam_v2 (2022 年 4 月 21 日访问)。
- 103.ElasticFusion。在线提供: <https://github.com/mp3guy/ElasticFusion> (2022 年 4 月 21 日访问)。
- 104.严Z.; 叶, M. Ren, L. 带有概率 Surfel 地图的密集视觉 SLAM。*IEEE 传输。维斯。计算。图形。* 2017年, 23, 2389–2398。[交叉引用]
- 105.DVO-SLAM。在线提供: https://github.com/tum-vision/dvo_slam (2022 年 4 月 21 日访问)。
- 106.戴, A.; 尼埃纳, M.; 佐尔霍夫, M.; 伊扎迪, S.; Theobalt, C. BundleFusion: 使用动态表面重新集成进行实时全局一致 3D 重建。
ACM 翻译。图形。 2017年, 36, 1. [交叉引用]
- 107.捆绑融合。在线提供: <https://github.com/niessner/BundleFusion> (2022 年 4 月 21 日访问)。
- 108.Concha, A.; Civera, J. RGBDTAM: 一种经济高效且准确的 RGB-D 跟踪和绘图系统。2017 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 英国, 科罗拉多州, 加拿大, 2017 年 9 月 24-28 日; 第 6756–6763 页。
- 109.RGBDTAM。在线提供: <https://github.com/alejocb/rbgdtam> (2022 年 4 月 21 日访问)。
- 110.刘, Y.; 赵, C. Ren, M. 一种用于室内移动机器人的增强型混合视觉惯性里程计系统。*传感器*2022年, 22, 2930。[交叉引用]
- 111.谢H.; 陈, W. Wang, J. 基于分层森林的低延迟一致视觉惯性 SLAM 快速在线闭环。*机器人。奥顿。系统。* 2022年, 151, 104035。[交叉引用]
- 112.李, W.; 埃肯霍夫, K.; 杨, Y. 日内瓦, P.; Huang, G. 具有在线校准功能的视觉惯性轮里程计。2020 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 2020 年 10 月 24 日至 1 月 24 日, 美国内华达州拉斯维加斯; 第 4559–4566 页。
- 113.程J.; 张L. Chen, Q. 一种改进的单目视觉惯性 SLAM 初始方法。*电子产品*2021年, 10, 3063。[交叉引用]
- 114.荣格, JH; 查, J. 钟JY; 金, TI; Seo, MH; 公园, SY; 杨, JY; Park, CG 在城市地区使用低级 IMU 的单目视觉惯性轮里程计。
IEEE 传输。英特尔。运输。系统。 2022年, 23, 925–938。[交叉引用]
115. Weiss, S. 微型直升机基于视觉的导航。博士论文, 苏黎世联邦理工学院, 瑞士苏黎世, 2012 年。
- 116.法尔克斯, JM; 卡斯帕, M.; Sibley, G. 用于鲁棒直接视觉里程计的惯性辅助密集和半密集方法。2016 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 韩国大田, 2016 年 10 月 9-14 日; 第 3601–3607 页。
- 117.Mourikis, 人工智能; Roumeliotis, SI 用于视觉辅助惯性导航的多状态约束卡尔曼滤波器。2007 年 IEEE 国际机器人与自动化会议论文集, 意大利罗马, 2007 年 4 月 10-14 日; 第 3565-3572 页。
- 118.维思, D.; 卡穆里, M.; 达斯, S. Fallon, M. 紧耦合激光雷达视觉惯性里程计的统一多模式地标跟踪。*IEEE 机器人。自动。莱特。* 2021年, 6, 1004–1011。[交叉引用]
119. 单SLAM。在线提供: <https://github.com/rrg-polito/mono-slam> (2022 年 4 月 22 日访问)。
120. PTAM。在线提供: <https://github.com/Oxford-PTAM/PTAM-GPL> (2022 年 4 月 22 日访问)。
- 121.ORB-SLAM2。在线提供: https://github.com/raulmur/ORB_SLAM2 (2022 年 4 月 22 日访问)。
- 122.戈麦斯-奥赫达, R.; 布里亚莱斯, J.; Gonzalez-Jimenez, J. PL-SVO: 通过结合点和线段的半直接单目视觉里程计。2016 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 韩国大田, 2016 年 10 月 9-14 日; 第 4211–4216 页。
- 123.PL-SVO。在线提供: <https://github.com/rubengooj/pl-svo> (2022 年 4 月 22 日访问)。
- 124.PL-SLAM。在线提供: <https://github.com/rubengooj/pl-slam> (2022 年 4 月 22 日访问)。
- 125.DTAM。在线提供: <https://github.com/anuranbaka/OpenDTAM> (2022 年 4 月 22 日访问)。

126. SVO。在线提供：https://github.com/uzh-rpg/rpg_svo（2022年4月22日访问）。
127. LSD-SLAM。在线提供：https://github.com/tum-vision/lsts_lam（2022年4月21日访问）。
128. DSO。在线提供：<https://github.com/JakobEngel/dso>（2022年4月22日访问）。
129. MSCKF-单声道。在线提供：https://github.com/daniilidis-group/msckf_mono（2022年4月22日访问）。
130. 奥克维斯。在线提供：<https://github.com/ethz-asl/okvis>（2022年4月21日访问）。
131. 布洛施, M.; 奥马里, S.; 哈特, M.; Siegwart, R. 使用直接基于 EKF 的方法的鲁棒视觉惯性里程计。2015 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 德国汉堡, 2015 年 9 月 28 日至 10 月 2 日; 第 298–304 页。
132. 罗维奥。在线提供：<https://github.com/ethz-asl/rovio>（2022年4月22日访问）。
133. VINS-单声道。在线提供：<https://github.com/HKUST-Aerial-Robotics/VINS-Mono>（2022年4月22日访问）。
134. 苏阿莱, M.; 金, G.-W. 基于 3D MODT 的语义感知动态 SLAM。传感器**2021**,**21**, 6355。[交叉引用]
135. 王S.; 郭, G.; 隋, H.; 周, Y. 张, H. Li, J. CDSFusion: 使用 CPU 计算的室内环境密集语义 SLAM。遥感。**2022**,**14**, 979。[交叉引用]
136. 维什尼亞科夫, B.; 斯吉布涅夫, I.; 谢维尔丁, V.; 索罗金, A.; 马萨洛夫, P.; 哈扎赫梅多夫, K.; Arseev, S. 具有基于 dcnn 的特征点检测、匹配和密集点云聚合的实时语义 slam。国际。拱。摄影图。遥感空间信息科学。**2021**,**XLIII-B2-2021**, 399–404。[交叉引用]
137. 李, P.; 张, G. 周, J. 姚, R. 张X; 周健。基于动态场景目标检测的Slam算法研究。2019 年先进机电系统国际会议 (ICAMechS) 论文集, 日本滋贺, 2019 年 8 月 26 日至 28 日; 第 363–367 页。
138. 徐D.; 维达尔迪, A.; Henriques, JF Moving SLAM: 非刚性场景中的完全无监督深度学习。2021 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 捷克共和国布拉格, 2021 年 9 月 27 日至 10 月 1 日; 第 4611–4617 页。
139. 胡, K.; 吴, J. 李, Y. 卢, M. 翁L.; Xia, M. FedGCN: 用于非欧几里得空间数据的基于联邦学习的图卷积网络。数学。**2022**,**10**, 1000。[交叉引用]
140. 胡K.; 翁, C.; 张, Y. 金, J. Xia, Q. 水下视觉增强概述: 从传统方法到最新的深度学习。J.Mar.Sci. 工程师。**2022**,**10**, 241。[交叉引用]
141. 胡K.; 李, M. 夏, M. Lin, H. 用于水域分割的多尺度特征聚合网络。传感器**2022**,**14**, 206。[交叉引用]
142. Lechelek, L.; 霍纳, S.; 佐鲁尔, R.; 诺丁, M.; Guillemin, C. MR 图像 3D 重建的混合方法。J. 影像学**2022**,**8**, 103。[交叉引用]
143. 胡K.; 丁, Y. 金, J. 翁L.; Xia, M. 基于多尺度深度时空特征的骨骼运动识别。应用。科学。**2022**,**12**, 1028。[交叉引用]
144. 迈克尔, E.; 萨默斯, TH; 伍德, 助教; 曼齐, C.; Shames, I. 大规模语义 SLAM 的概率数据关联。arXiv**2022**, arXiv: 2202.12802。
145. 李, R.; 王, S. Gu, D. DeepSLAM: 具有无监督深度学习的鲁棒单目 SLAM 系统。IEEE 传输。工业电子。**2021**,**68**, 3577–3587。[交叉引用]
146. Garg, R.; Bg、VK; 卡内罗, G.; Reid, I. 用于单视图深度估计的无监督 CNN: 几何救援。计算机视觉会议录 - ECCV 2016, 荷兰阿姆斯特丹, 2016 年 10 月 11-14 日; 第 740–756 页。
147. 穆克吉, A.; 查克拉博蒂, S.; Saha, SK 学习 SLAM 中位置识别的深度表示。模式识别和机器智能会议录, 印度加尔各答, 2017 年 12 月 5-8 日; 第 557–564 页。
148. 高X.; 张, T. 视觉 SLAM 系统中使用深度神经网络检测循环的无监督学习。奥顿。机器人。**2017**,**41**, 1-18。[交叉引用]
149. 哲, J.; Eoh, G. 用于视觉位置识别的条件不变特征提取的变分贝叶斯方法。应用。科学。**2021**,**11**, 8976。[交叉引用]
150. 穆穆尼, A.; Mumuni, F. 计算机视觉中几何变换不变特征表示的 CNN 架构: 综述。SN 计算。科学。**2021**,**2**, 340。[交叉引用]
151. 马, R.; 王, R. 张, Y. 皮泽尔, S.; 麦吉尔, SK; 罗森曼, J.; 弗拉姆, J.-M. RNN SLAM: 重建 3D 结肠以在结肠镜检查期间可视化缺失区域。医学。图像肛门。**2021**,**72**, 102100。[交叉引用] [考研]
152. 王K.; 马, S. 陈, J. 任, F.; Lu, J. 深度视觉里程计的方法、挑战和应用: 面向复杂和新兴领域。IEEE 传输。认知。开发。系统。**2022**,**14**, 35–49。[交叉引用]
153. 洪, S.; 班贡哈卡纳, A.; 帕克, J.-M.; 崔, M. 辛, H.-S. 基于视觉 SLAM 的行星构建机器人测绘方法。传感器**2021**,**21**, 7715。[交叉引用]
154. 段, C.; 容金格, S.; 黄, J. 金, K. Thurrow, K. 运输机器人中视觉 SLAM 的深度学习: 综述。运输。萨夫。环境。**2020**,**1**, 177–184。[交叉引用]
155. 卢, SY; 沙克里, M.; 唐SH; 马什霍尔, S.; 张H. 深度预测和视觉 SLAM 的在线相互适应。arXiv**2021**, arXiv: 2111.04096。
156. 金, JJY; 乌尔施勒, M.; 里德尔, PJ; Wicker, JS SymbioLCD: 使用 CNN 提取的对象和视觉词袋进行基于集成的循环闭合检测。2021 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 捷克共和国布拉格, 2021 年 9 月 27 日至 10 月 1 日; p. 5425。

157. 斯廷贝克, A.; Nex, F. 基于 CNN 的密集单目视觉 SLAM, 用于紧急情况下的实时无人机探索。*无人机*2022年,6, 79. [交叉引用]
158. 塔特诺, K.; 通巴里, F.; 莱娜, I.; Navab, N. CNN-SLAM: 具有学习深度预测的实时密集单目 SLAM。2017 年 IEEE 计算机视觉和模式识别会议 (CVPR) 会议记录, 美国夏威夷州檀香山, 2017 年 7 月 21-26 日; 第 6565-6574 页。
159. 杨, N.; 王, R. 施图克勒, J.; Cremers, D. 深度虚拟立体里程计: 利用单目直接稀疏里程计的深度预测。计算机视觉论文集 — ECCV 2018, 德国慕尼黑, 2018 年 9 月 8 日至 14 日; 第 835-852 页。
160. 戈达尔, C.; 奥达, OM; Brostow, GJ 具有左右一致性的无监督单目深度估计。2017 年 IEEE 计算机视觉和模式识别会议 (CVPR) 会议记录, 美国夏威夷州檀香山, 2017 年 7 月 21-26 日; 第 6602-6611 页。
161. 秦T.; 陈, T. 陈, Y. Su, Q. AVP-SLAM: 停车场自动驾驶车辆的语义视觉映射和定位。IEEE/RSJ 国际智能机器人系统会议论文集, 美国内华达州拉斯维加斯, 2020 年 10 月 24 日至 2020 年 1 月 24 日; 第 5939-5945 页。
162. Ronneberger, O.; 费舍尔, P.; Brox, T. U-Net: 用于生物医学图像分割的卷积网络。医学图像计算和计算机辅助干预论文集 — MICCAI 2015, 德国慕尼黑, 2015 年 10 月 5-9 日; 第 234-241 页。
163. 朱R.; 杨, M. 刘, W.; 宋, R. 严, B.; Shaw, Z. DeepAVO: 通过深度视觉里程计的特征提取进行高效姿势精炼。神经计算2022年,第 467章, 22-35. [交叉引用]
164. 罗, Y.; 肖, Y. 张, Y. Zeng, N. 视觉 SLAM 中循环闭合的检测: 一种基于堆叠分类自动编码器的方法。光电. 莱特。2021年,17 号, 354-360. [交叉引用]
165. 温S.; 赵, Y. 袁X.; 王, Z. 张, D. Manfredi, L. 未知环境下基于深度强化学习的主动 SLAM 路径规划。英特尔。服务。机器人。2020 年,13, 263-272. [交叉引用]
166. 梅蒙, 阿肯色州; 王, H. Hussain, A. 使用监督和无监督深度神经网络进行单目 SLAM 系统的闭环检测。机器人。奥顿。系统。2020 年,126, 103470. [交叉引用]
167. 李D.; 施X.; 长, Q. 刘, S. 杨, W. 王, F. 魏, Q. Qiao, F. DXSLAM: 具有深度特征的稳健高效的视觉 SLAM 系统。在 2020 年 IEEE/RSJ 国际智能机器人与系统会议 (IROS) 会议记录中, 美国内华达州拉斯维加斯, 2020 年 10 月 24 日至 2020 年 1 月 24 日; 第 4958-4965 页。
168. 秦, C.; 张, Y. 刘, Y. Lv, G. 多目标场景中基于图匹配的语义闭环检测。J. 维斯。交流。图像代表。2021年,76, 103072. [交叉引用]
169. 陈, C.; 王, B.; 卢CX; 特里戈尼, A.; Markham, A. 定位和绘图深度学习调查: 迈向空间机器智能时代。arXiv2020年, arXiv: 2006.12567.
170. 叶X.; 吉, X. 孙, B.; 陈, S. 王, Z. Li, H. DRM-SLAM: 通过场景深度融合实现单目 SLAM 的密集重建。神经计算2020年,第396章, 76-91. [交叉引用]
171. 曹J.; 曾, B.; 刘, J. 赵, Z. 苏宇。一种基于深度学习算法的同步定位建图重定位新方法。计算。电。工程师。2017年,63, 79-90. [交叉引用]
172. 阿尔沙德, S.; 金, G.-W. 深度学习在视觉和激光雷达 SLAM 环路闭合检测中的作用: 一项调查。传感器2021年,21, 1243. [交叉引用]
173. 李, R.; 王, S. 朗, Z.; Gu, D. UnDeepVO: 通过无监督深度学习的单目视觉里程计。2018 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 澳大利亚布里斯班, 2018 年 5 月 21-25 日; 第 7286-7291 页。
174. 德通, D.; 马里谢维奇, T.; Rabinovich, 《AJA 走向几何深度 SLAM》。arXiv2017年, arXiv: 1707.07410.
175. 克拉克, R.; 王, S. 温, H. 马卡姆, A.; Trigoni, A. VINet: 视觉惯性里程计作为序列到序列学习问题。AAAI 会议记录, 加拿大魁北克市, 2017 年 10 月 23-26 日。
176. 纳西尔, T.; 奥利维拉, GL; 布洛克斯, T.; Burgard, W. 在具有挑战性的感知条件下语义感知的视觉定位。2017 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 新加坡, 2017 年 5 月 29 日至 6 月 3 日; 第 2614-2620 页。
177. 鲍曼, SL; 阿塔纳索夫, N.; 丹尼利迪斯, K.; Pappas, GJ 语义 SLAM 的概率数据关联。2017 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 新加坡, 2017 年 5 月 29 日至 6 月 3 日; 第 1722-1729 页。
178. 胡K.; 郑, F. 翁L.; 丁, Y. Jin, J. 基于特征增强的时空差分 LSTM 动作识别算法。应用。科学。2021年,11, 7876. [交叉引用]
179. 陈, W.; 郑, F. 高, S. Hu, K. 具有差分结构的 LSTM 及其在动作识别中的应用。数学。问题。工程师。2022年,2022年, 7316396. [交叉引用]
180. 曹B.; 李, C. 宋, Y. 秦, Y. Chen, C. 基于 CNN 和 GRU 的网络入侵检测模型。应用。科学。2022年,12, 4184. [交叉引用]
181. 陈, EZ; 王, P. 陈X.; 陈, T. Sun, S. 用于 MRI 图像重建的金字塔卷积 RNN。IEEE 传输。医学。影像学2022年。[交叉引用] [考研]
182. 桑, H.; 江, R. 王, Z. 周, Y. He, B. 一种用于未知环境中自主视觉导航的新型神经多存储记忆网络。IEEE 机器人。自动。莱特。2022年,7, 2039-2046. [交叉引用]

- 183.薛F.; 王Q. 王X; 董, W. 王, J. Zha, H. 深度视觉里程计的引导特征选择。计算机视觉论文集—ACCV 2018, 澳大利亚珀斯, 2018 年 12 月 2-6 日; 第 293–308 页。
- 184.特德, Z.; Deng, J. DROID-SLAM: 用于单目、立体和 RGB-D 相机的深度视觉 SLAM。arXiv2021年, arXiv: 2108.10869。
- 185.图兰, M.; 阿尔马里奥格鲁, Y.; 阿劳霍, H.; 科努科格鲁, E.; Sitti, M. Deep EndoVO: 一种用于内窥镜胶囊机器人的基于循环卷积神经网络 (RCNN) 的视觉里程计方法。神经计算2018年,275, 1861年至1870年。[交叉引用]
- 186.钱坎, M.; Milford, M. DeepSeqSLAM: 用于联合全局描述和基于序列的位置识别的可训练 CNN+RNN。arXiv2020年, arXiv: 2011.08518。
- 187.汉, L.; 林, Y. 杜, G. Lian, S. DeepVIO: 使用 3D 几何约束的单目视觉惯性里程计的自监督深度学习。2019 年 IEEE/RSJ 国际智能机器人与系统会议 (IROS) 论文集, 中国澳门威尼斯人, 2019 年 11 月 3-8 日; 第 6906–6913 页。
- 188.陈, C.; 罗莎, S.; 苗, Y. 卢CX; 吴, W. 马卡姆, A.; Trigoni, N. 用于神经视觉惯性里程计的选择性传感器融合。2019 年 IEEE/CVF 计算机视觉和模式识别会议 (CVPR) 会议记录, 美国加利福尼亚州长滩, 2019 年 6 月 15 日至 20 日; 第 10534–10543 页。
- 189.阿尔马里奥格鲁, Y.; 图兰, M.; 萨普特拉, MRU; 德古斯芒, PPB; 马卡姆, A.; Trigoni, N. SelfVIO: 自监督深度单目视觉惯性里程计和深度估计。神经网络。2022年,150, 119–136。[交叉引用]
- 190.黄, A.; 费, X. 崔, S. Soatto, S. 视觉惯性里程计的无监督深度完成。IEEE 机器人。自动。莱特。2020年,5, 1899-1906。[交叉引用]
- 191.王S.; 克拉克, R. 温, H. Trigoni, N. DeepVO: 利用深度循环卷积神经网络实现端到端视觉里程计。2017 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 新加坡, 2017 年 5 月 29 日至 6 月 2 日; 第 2043–2050 页。
- 192.卢, SY; 阿米里, AJ; 马什霍尔, S.; 唐SH; 张, H. CNN-SVO: 使用单图像深度预测改进半直接视觉里程计中的映射。2019 年机器人与自动化国际会议 (ICRA) 会议记录, 加拿大魁北克省蒙特利尔, 2019 年 5 月 20-24 日; 第 5218–5223 页。
- 193.阿尔马里奥格鲁, Y.; 萨普特拉, MRU; 古斯芒, PPBd; 马卡姆, A.; Trigoni, N. GANVO: 无监督深度单目视觉里程计和生成对抗网络的深度估计。2019 年机器人与自动化国际会议 (ICRA) 会议记录, 加拿大魁北克省蒙特利尔, 2019 年 5 月 20-24 日; 第 5474–5480 页。
- 194.李, Y.; 牛久, Y.; Harada, T. 无监督单目视觉里程计的位姿图优化。2019 年机器人与自动化国际会议 (ICRA) 会议记录, 加拿大魁北克省蒙特利尔, 2019 年 5 月 20-24 日; 第 5439-5445 页。
- 195.布鲁诺, 英国皇家海军; Colombini, EL LIFT-SLAM: 一种基于深度学习特征的单目视觉 SLAM 方法。神经计算2021年,第455章, 97–110。[交叉引用]
- 196.张S.; 卢, S. 她。; Bao, Z. 通过无监督深度学习进行立体视觉里程计姿势校正。传感器2021年, 21, 4735。[交叉引用]
- 197.沙姆韦尔, EJ; 林格伦, K.; 梁, S. Nothwang, WD 无监督深度视觉惯性里程计, 具有 RGB-D 图像在线纠错功能。IEEE 传输。模式识别。马赫。英特尔。2020年,42, 2478–2493。[交叉引用][考研]
- 198.金, Y.; 尹, S. 金, S. Kim, A. 用于视觉惯性传感器融合的无监督平衡协方差学习。IEEE 机器人。自动。莱特。2021年,6, 819–826。[交叉引用]
- 199.古尔特克, M.; 尤塞菲, A.; 阿斯兰, MF; 索伊坎, M.; 杜尔杜, A.; Masiero, A. YTU 数据集和基于循环神经网络的视觉惯性里程计。测量2021年,184, 109878。[交叉引用]
- 200.关, P.; 曹, Z. 陈, E. 梁, S. 谭, M. Yu, J. 一种基于点和对象的实时语义视觉 SLAM 方法。国际。J.Adv. 机器人。系统。2020年, 17 号, 1729881420905443。[交叉引用]
- 201.亨佩尔, T.; Al-Hamadi, A. 用于扩展和增强视觉 SLAM 的在线语义映射系统。工程师。应用。阿蒂夫。英特尔。2022年,111, 104830。[交叉引用]
- 202.钱Z.; 帕塔斯, K.; 傅, J. 具有自主对象级数据关联的语义 SLAM。2021 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 中国西安, 2021 年 5 月 30 日至 6 月 5 日; 第 11203–11209 页。
- 203.巴勃罗·F·阿尔坎塔里拉博士。在线提供: <https://blog.slamcore.com/age-of-perception> (2022 年 6 月 12 日访问)。
- 204.张L.; 魏, L. 沉, P. 魏, W.; 朱, G.; Song, J. 基于目标检测和改进 Octomap 的语义 SLAM。IEEE 接入2018年,6, 75545–75559。[交叉引用]
- 205.胡K.; 张, D. Xia, M. CDUNet: 遥感图像云检测 UNet。遥感。2021年,13, 4533。[交叉引用]
- 206.胡K.; 张, Y. 翁, C.; 王, P. 邓, Z. 刘燕。一种基于生成对抗网络和自然图像质量评价指标的水下图像增强算法。J.Mar.Sci. 工程师。2021年,9, 691。[交叉引用]
- 207.帕扎尼, AAJ; Vasanthanayaki, C. 通过与增强型 ROI 池化 (FrRNet-ERol) 框架相结合的更快 R-CNN 进行卫星图像中的对象检测。地球科学。通知。2022年,15, 553–561。[交叉引用]
- 208.Hoang, TM; 周, J. Fan, Y. 使用编码器-解码器匹配语义分割进行图像压缩。2020 年 IEEE/CVF 计算机视觉和模式识别研讨会 (CVPRW) 会议记录, 美国华盛顿州西雅图, 2020 年 6 月 14 日至 19 日; 第 619–623 页。
- 209.胡, K.; 卢, F. 卢, M. 邓, Z. 刘Y. 一种基于SSD和特征增强的海洋目标检测算法。复杂2020年,2020年, 5476142。[交叉引用]
- 210.邵, F.; 陈L.; 邵, J. 吉, W. 肖, S. 叶, L. 庄, Y. 肖, J. 弱监督目标检测和定位的深度学习: 一项调查。神经计算2022年,第496章, 192-207。[交叉引用]

211. 梁W.; 徐, P. 郭L.; 白, H.; 周, Y. Chen, F. 3D 对象检测调查。多媒体。工具应用**2021年**,*80*, 29617–29641。[交叉引用]
212. 吉尔希克, R.; 多纳休, J.; 达雷尔, T.; Malik, J. 用于精确对象检测和语义分割的丰富特征层次结构。2014 年 IEEE 计算机视觉和模式识别会议论文集, 美国俄亥俄州哥伦布, 2014 年 6 月 23–28 日; 第 580–587 页。
213. Girshick, R. Fast R-CNN。2015 年 IEEE 国际计算机视觉会议 (ICCV) 会议记录, 智利圣地亚哥, 2015 年 12 月 7–13 日; 第 1440–1448 页。
214. 任, S.; 他, K. 吉尔希克, R.; Sun, J. Faster R-CNN: 利用区域提议网络实现实时目标检测。IEEE 传输。模式肛门。马赫。英特尔。**2017年**,*39*, 1137–1149。[交叉引用] [考研]
215. 雷德蒙, J.; 迪夫瓦拉, S.; 吉尔希克, R.; Farhadi, A. 你只看一次: 统一的实时对象检测。2016 年 IEEE 计算机视觉和模式识别会议 (CVPR) 会议记录, 美国内华达州拉斯维加斯, 2016 年 6 月 27–30 日; 第 779–788 页。
216. 刘W.; 安格洛夫, D.; 埃尔汗, D.; 塞格迪, C.; 里德, S.; 傅, C.-Y.; Berg, AC SSD: 单次多盒探测器。计算机视觉会议录 - ECCV 2016, 荷兰阿姆斯特丹, 2016 年 10 月 11–14 日; 第 21–37 页。
217. 龙, J.; 谢尔哈默, E.; Darrell, T. 用于语义分割的全卷积网络。2015 年 IEEE 计算机视觉和模式识别会议 (CVPR) 会议记录, 美国马萨诸塞州波士顿, 2015 年 6 月 7–12 日; 第 3431–3440 页。
218. 巴德里纳拉亚南, V.; 肯德尔, A.; Cipolla, R. SegNet: 用于图像分割的深度卷积编码器-解码器架构。IEEE 传输。模式肛门。马赫。英特尔。**2017年**,*39*, 2481–2495。[交叉引用] [考研]
219. 赵H.; 石, J.; 齐, X. 王X.; Jia, J. 金字塔场景解析网络。2017 年 IEEE 计算机视觉和模式识别会议 (CVPR) 会议记录, 美国夏威夷州檀香山, 2017 年 7 月 21–26 日; 第 6230–6239 页。
220. 他, K.; 吉奥萨里, G.; 多拉尔, P.; Girshick, RB Mask R-CNN。IEEE 国际计算机视觉会议论文集, 意大利威尼斯, 2017 年 10 月 22–29 日; 第 2980–2988 页。
221. YOLOv5。在线提供: <https://github.com/ultralytics/yolov5> (2022 年 4 月 23 日访问)。
222. 陈, L.-C.; 帕潘德里欧, G.; 科基诺斯, I.; 墨菲, K.; Yuille, AL 使用深度卷积网络和全连接 crf 进行语义图像分割。arXiv**2014年**, arXiv: 1412.7062。
223. DeepLabv3+。在线提供: <https://github.com/Tramac/awesome-semantic-segmentation-pytorch> (2022 年 4 月 23 日访问)。
224. 博利亚, D.; 周, C.; 肖, F. Lee, YJ YOLACT: 实时实例分割。2019 年 IEEE/CVF 国际计算机视觉会议 (ICCV) 会议记录, 韩国首尔, 2019 年 10 月 27 日至 11 月 2 日; 第 9156–9165 页。
225. Dang, X.; 荣, Z. 梁 X. 基于传感器融合的消除动态环境中 SLAM 移动物体的方法。传感器**2021年**,*21*, 230。[交叉引用]
226. 乔普, F.; 涅托, JI; 西格沃特, RY; Cadena, C. 基于优化的语义 SLAM 的超二次对象表示。arXiv**2021年**, arXiv: 2109.09627。
227. 赵Z.; 毛, Y. 丁, Y. 任, P. 郑 N. 基于视觉的大规模室外环境地标语义 SLAM。2019 年第二届中国认知计算与混合智能研讨会 (CCHI) 论文集, 中国西安, 2019 年 9 月 21–22 日; 第 149–154 页。
228. 利亚诺斯, K.-N.; 舍恩伯格, JL; 波尔菲斯, M.; Sattler, T. VSO: 视觉语义里程计。计算机视觉论文集 – ECCV 2018, 德国慕尼黑, 2018 年 9 月 8 日至 14 日; 第 246–263 页。
229. 斯滕堡, E.; 托夫特, C.; Hammarstrand, L. 使用语义分割图像的长期视觉定位。2018 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 澳大利亚布里斯班, 2018 年 5 月 21–25 日; 第 6484–6490 页。
230. 戴, W.; 张, Y. 李, P. 方, Z.; Scherer, S. 在动态环境中使用点相关性的 RGB-D SLAM。IEEE 传输。模式肛门。马赫。英特尔。**2022年**,*44*, 373–389。[交叉引用]
231. 韩X.; 李, S. 王X.; Zhou, W. 室内场景中移动机器人的语义映射: 一项调查。信息**2021年**,*12*, 92. [交叉引用]
232. 廖Z.; 胡, Y. 张, J. 齐, X. 张X.; Wang, W. SO-SLAM: 具有比例比例和对称纹理约束的语义对象 SLAM。IEEE 机器人。自动。莱特。**2022年**,*7*, 4008–4015。[交叉引用]
233. Ran, T.; 袁L.; 张, J. 他, L.; 黄, R. Mei, J. 不仅看, 而且推断: 语义 SLAM 数据关联推理的多重假设聚类。IEEE 传输。仪器。测量。**2021年**,*70*, 1–9. [交叉引用]
234. 殷, R.; 程, Y. 吴, H. 宋, Y. 于, B.; Niu, R. FusionLane: 使用深度神经网络进行车道标记语义分割的多传感器融合。IEEE 传输。英特尔。运输。系统。**2022年**,*23*, 1543–1553。[交叉引用]
235. 汉, B.; Xu, L. 具有掩模环闭合的单目 SLAM 系统。2020 年中国控制与决策会议 (CCDC) 论文集, 中国合肥, 2020 年 8 月 22–24 日; 第 4762–4768 页。
236. 杨S.; 范, G.; 白, L.; 赵, C. Li, D. SGC-VSLAM: 动态室内环境的语义和几何约束 VSLAM。传感器**2020年**,*20*, 2432. [交叉引用] [考研]
237. 加林多, C.; 萨菲奥蒂, A.; 科拉德斯基, S.; 布施卡, P.; 费尔南德斯-马德里加尔, JA; Gonzalez, J. 移动机器人的多层次语义图。2005 年 IEEE/RSJ 国际智能机器人和系统会议论文集, 加拿大艾伯塔省埃德蒙顿, 2005 年 8 月 2–6 日; 第 2278–2283 页。
238. 瓦苏代万, S.; 加希特, S.; 阮, V.; Siegwart, R. 移动机器人的认知地图——一种基于对象的方法。机器人。奥顿。系统。**2007年**,*55*, 359–371。[交叉引用]
239. 岳, Y.; 温, M. 赵, C. 王, Y. Wang, D. COSEM: 自主机器人协作语义地图匹配框架。IEEE 传输。工业电子。**2022年**,*69*, 3843–3853。[交叉引用]

240. 阿舒尔, R.; 阿卜杜勒卡德尔, M.; 迪亚斯, J.; 北卡罗来纳州阿尔莫萨; Taha, T. 机器人探索期间的语义危险标签和风险评估映射。*IEEE 接入*2022年,10, 16337–16349。[交叉引用]
241. 松香, A.; 阿巴特, M.; 张, Y. Carbone, L. Kimera: 用于实时度量语义定位和映射的开源库。2020 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 虚拟会议, 2020 年 5 月 31 日至 8 月 31 日; 第 1689–1696 页。
242. 明远, M.; 和伟, Z.; 斯蒙, L.; Baochang, Z. SEMANTIC-RTAB-MAP (SRM): 一种在深度图像上使用 CNN 的语义 SLAM 系统。数学。成立。计算。2019年,2, 29–41。
243. 拉贝, M.; Michaud, F. 用于在线大规模和长期运行的基于外观的环路闭合检测。*IEEE 传输。机器人*。2013年,29, 734–745。[交叉引用]
244. 梅尼尼, D.; 库马尔, S.; 奥斯瓦尔德先生; 桑德斯特伦, E.; 斯明奇塞斯库, C.; Gool, LV 用于室内场景联合 3D 重建和语义分割的实时在线学习框架。*IEEE 机器人。自动。莱特*。2022年,7, 1332–1339。[交叉引用]
245. 庄, C.; 王, Z. 赵, H. Ding, H. 基于 3D 物体姿态估计的语义部分分割方法, 使用 RGB-D 图像进行分箱拾取。*机器人。计算机集成制造商*。2021年,68, 102086。[交叉引用]
246. 苏萨, YC; Bassani, 通过深度视觉特征的整合进行 HDF 拓扑语义映射。*IEEE 机器人。自动。莱特*。2022年,7, 4110–4117。[交叉引用]
247. 王F.; 张, C.; 张, W. 方, C.; 夏, Y. 刘, Y. Dong, H. 基于对象的移动机器人可靠视觉导航。*传感器*2022年,22, 2387。[交叉引用] [考研]
248. 吴, Y.; 张, Y. 朱, D. 冯, Y. 科尔曼, S.; Kerr, D. EAO-SLAM: 基于集成数据关联的单目半密集对象 SLAM。在 2020 年 IEEE/RSJ 国际智能机器人与系统会议 (IROS) 会议记录中, 美国内华达州拉斯维加斯, 2020 年 10 月 24 日至 1 月 24 日; 第 4966–4973 页。
249. 王J.; 伦茨, M.; Agapito, L. DSP-SLAM: 具有深度形状先验的面向对象 SLAM。2021 年 3D 视觉国际会议 (3DV) 会议记录, 英国伦敦, 2021 年 12 月 1-3 日; 第 1362–1371 页。
250. Fu, J.; 黄Q. 多尔蒂, K.; 王, Y. Leonard, JJ 在基于对象的 SLAM 中提出模糊性的多假设方法。2021 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 捷克共和国布拉格, 2021 年 9 月 27 日至 10 月 1 日; 第 7639–7646 页。
251. Zhai, R.; 袁勇。复杂城市环境中利用语义信息的视觉辅助 GNSS 定位方法。*遥感*。2022年,14, 869。[交叉引用]
252. 维尼特, V.; 米克西克, O.; 利德加德, M.; 尼埃纳, M.; 戈洛代茨, S.; 弗吉尼亚州普里萨卡留; 卡勒, O.; 默里, DW; 伊扎迪, S.; 佩雷斯, P.; 等人。用于大规模语义场景重建的增量密集语义立体融合。2015 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 美国华盛顿州西雅图, 2015 年 5 月 26-30 日; 第 75–82 页。
253. 赵Z.; Chen, X. 使用 RGB-D 相机为移动机器人构建 3D 语义地图。*英特尔。服务。机器人*。2016年,9, 297–309。[交叉引用]
254. 李X.; Belaroussi, R. 单目 SLAM 的半密集 3D 语义映射。*arXiv*2016年, arXiv: 1611.04144。
255. 杨S.; 黄, Y. Scherer, S. 通过高效的高阶 CRF 进行语义 3D 占用映射。2017 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 加拿大不列颠哥伦比亚省温哥华, 2017 年 9 月 24-28 日; 第 590–597 页。
256. 成田G.; 塞诺, T.; 石川, T.; Kaji, Y. PanopticFusion: 事物层面的在线体积语义映射。2019 年 IEEE/RSJ 国际智能机器人与系统会议 (IROS) 论文集, 中国澳门威尼斯人, 2019 年 11 月 3-8 日; 第 4205–4212 页。
257. 秦T.; 郑, Y. 陈, T. 陈, Y. Su, Q. 用于自动驾驶视觉定位的轻量级语义图。2021 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 中国西安, 2021 年 5 月 30 日至 6 月 5 日; 第 11248–11254 页。
258. 严L.; 胡X.; 赵L.; 陈, Y. 魏, P. Xie, H. DGS-SLAM: 动态环境中结合几何和语义信息的快速鲁棒 RGBD SLAM。*遥感*。2022年,14, 795。[交叉引用]
259. 胡, Z.; 赵, J. 罗, Y. Ou, J. 动态场景中基于改进 DeepLabv3+ 的语义 SLAM。*IEEE 接入*2022年,10, 21160–21168。[交叉引用]
260. 刘X.; 纳尔达里, GV; 奥赫达, 足球俱乐部; 陶, Y. 周, A.; 唐纳利, T.; 曲, C.; 陈西南; 罗梅罗, 英国皇家空军; 泰勒, CJ; 等人。茂密林冠下实时语义 SLAM 大规模自主飞行。*IEEE 机器人。自动。莱特*。2022年,7, 5512–5519。[交叉引用]
261. 陈, B.; 彭, G.; 他, D. 周, C.; Hu, B. 基于动态目标检测的视觉 SLAM。2021 年第 33 届中国控制与决策会议 (CCDC) 论文集, 中国昆明, 2021 年 5 月 22-24 日; 第 5966–5971 页。
262. 贝斯科斯, B.; 法西尔, JM; 西维拉, J.; Neira, J. DynaSLAM: 动态场景中的跟踪、映射和修复。*IEEE 机器人。自动。莱特*。2018年, 3, 4076–4083。[交叉引用]
263. Yu, C.; 刘, Z.; 刘XJ; 谢F.; 杨, Y. 魏, Q. Fei, Q. DS-SLAM: 针对动态环境的语义视觉 SLAM。2018 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 西班牙马德里, 2018 年 10 月 1-5 日; 第 1168–1174 页。
264. 金子, M.; 石见, K.; 小川, T.; 山崎, T.; Aizawa, K. Mask-SLAM: 使用语义分割进行掩蔽的基于鲁棒特征的单目 SLAM。2018 年 IEEE/CVF 计算机视觉和模式识别研讨会 (CVPRW) 会议记录, 美国犹他州盐湖城, 2018 年 6 月 18 日至 22 日; 第 371–3718 页。

- 265.Schörghuber, M.; 斯坦宁格, D.; 卡邦, Y.; 胡门伯格, M.; Gelautz, M. SLAMANTIC—利用语义改进动态环境中的 VSLAM。2019 年 IEEE/CVF 国际计算机视觉研讨会 (ICCVW) 会议记录, 韩国首尔, 2019 年 10 月 27-28 日; 第 3759–3768 页。
- 266.Lv, X.; 王, B.; 叶, D. Wang, SJA 用于鲁棒映射的语义流引导运动消除方法。arXiv2020年, arXiv: 2010.06876。
- 267.袁X.; Chen, S. SaD-SLAM: 基于语义和深度信息的视觉 SLAM。2020 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 美国内华达州拉斯维加斯, 2020 年 10 月 24 日至 1 月 24 日; 第 4930–4935 页。
- 268.温S.; 李, P. 赵, Y. 张, H. 孙F.; Wang, Z. 动态环境中的语义视觉 SLAM。奥顿。机器人。2021年, 45, 493–504。[交叉引用]
- 269.吴, J.; 熊, J. 提高动态场景中 VIO 线条特征的鲁棒性。测量。科学。技术。2022年, 33, 065204。[交叉引用]
- 270.王, M.; 王, H. 王, Z. Li, Y. 一种基于物联网的利用语义对象特征的无人机视觉重定位方法。无线通信移动计算会议记录, 克罗地亚杜布罗夫尼克, 2022 年 5 月 30 日至 6 月 3 日。
- 271.陆Z.; 胡, Z. Uchimura, K. 动态室外环境中的 SLAM 估计: 综述。智能机器人与应用会议录, 新加坡, 2009 年 12 月 16-18 日; 第 255–267 页。
- 272.雷迪, ND; 阿巴斯内贾德, I.; 雷迪, S.; 阿拉斯加蒙达尔; Devalla, V. 使用立体相机进行轨迹优化的增量实时多体 VSLAM。2016 年 IEEE/RSJ 国际智能机器人和系统会议 (IROS) 会议记录, 韩国大田, 2016 年 10 月 9-14 日; 第 4505–4510 页。
- 273.伦茨, P.; 齐格勒, J.; 盖革, A.; Roser, M. 用于城市环境中移动物体检测的稀疏场景流分割。2011 年 6 月 5-9 日在德国巴登巴登举行的 2011 年 IEEE 智能汽车研讨会 (IV) 会议记录; 第 926–932 页。
- 274.动态SLAM。在线提供: <https://github.com/BertaBescos/DynaSLAM> (2022 年 4 月 24 日访问)。
- 275.DS-SLAM。在线提供: <https://github.com/ivipsourcecode/DS-SLAM> (2022 年 4 月 24 日访问)。
- 276.钟F.; 王, S. 张, Z. 陈, C.; Wang, Y. Detect-SLAM: 使目标检测和 SLAM 互惠互利。2018 年 IEEE 计算机视觉应用冬季会议 (WACV) 会议记录, 美国内华达州太浩湖, 2018 年 3 月 12-15 日; 第 1001–1010 页。
- 277.检测SLAM。在线提供: <https://github.com/liadbiz/detect-slam> (2022 年 4 月 24 日访问)。
- 278.王Z.; 张, Q. 李, J. 张, S. Liu, J. 用于动态场景的计算高效语义 SLAM 解决方案。遥感。2019年, 11, 1363。[交叉引用]
- 279.斯拉曼蒂克。在线提供: <https://github.com/mthz/slamanitic> (2022 年 4 月 25 日访问)。
- 280.爱荷华州巴桑; 刘, P. 波尔菲斯, M.; Geiger, A. 大规模动态环境的鲁棒密集映射。2018 年 IEEE 国际机器人与自动化会议 (ICRA) 会议记录, 澳大利亚布里斯班, 2018 年 5 月 21-25 日; 第 7510–7517 页。
- 281.戴, J.; 他, K. Sun, J. 通过多任务网络级联实现实例感知语义分割。2016 年 IEEE 计算机视觉和模式识别会议 (CVPR) 会议记录, 美国内华达州拉斯维加斯, 2016 年 6 月 27-30 日; 第 3150–3158 页。
- 282.动态SLAM。在线提供: <https://github.com/AndreiBarsan/DynSLAM> (2022 年 4 月 25 日访问)。
- 283.埃斯帕扎, D.; Flores, G. STDyn-SLAM: 动态室外环境中 VSLAM 的立体视觉和语义分割方法。IEEE接入2022年, 10, 18201–18209。[交叉引用]
- 284.STDyn-SLAM。在线提供: <https://github.com/DanielaEsparza/STDyn-SLAM> (2022 年 4 月 25 日访问)。
- 285.张T.; Nakamura, Y. Posefusion: 动态人类环境中的密集 rgb-d slam。国际实验机器人研讨会论文集, 阿根廷布宜诺斯艾利斯, 2018 年 11 月 5-8 日; 第 772–780 页。
- 286.姿势融合。在线提供: <https://github.com/conix-center/posefusion> (2022 年 4 月 25 日访问)。
- 287.刘, Y.; Miura, J. RDS-SLAM: 使用语义分割方法的实时动态 SLAM。IEEE接入2021年, 9, 23772–23785。[交叉引用]
- 288.RDS-SLAM。在线提供: <https://github.com/yubaoliu/RDS-SLAM> (2022 年 4 月 25 日访问)。
- 289.赖, D.; 李, C. He, B. YO-SLAM: 针对动态环境的鲁棒视觉 SLAM。2021 年通信、信息系统和计算机工程国际会议 (CISCE) 论文集, 中国北京, 2021 年 5 月 14 日至 16 日; 第 720–725 页。
- 290.张, Y.; 徐, X. 张, N. Lv, Y. 用于动态环境中折反射全景相机的语义 SLAM 系统。传感器2021年, 21, 5889。[交叉引用] [考研]
- 291.舍恩拜因, M.; Geiger, A. 增强曼哈顿世界中的全方位 3D 重建。2014 年 IEEE/RSJ 国际智能机器人和系统会议论文集, 美国伊利诺伊州芝加哥, 2014 年 9 月 14-18 日; 第 716–723 页。
- 292.胡X.; Lang, J. DOE-SLAM: 动态对象增强视觉 SLAM。传感器2021年, 21, 3091。[交叉引用]
- 293.余, N.; 甘, M. 于, H.; Yang, K. DRSO-SLAM: 用于室内动态场景的动态 RGB-D SLAM 算法。2021 年第 33 届中国控制与决策会议 (CCDC) 论文集, 中国昆明, 2021 年 5 月 22-24 日; 第 1052–1058 页。
- 294.艾, Y.; 瑞, T. 卢, M. 傅L.; 刘, S. Wang, S. DDL-SLAM: 动态环境中与深度学习相结合的鲁棒 RGB-D SLAM。IEEE接入2020年, 8, 162335–162342。[交叉引用]
- 295.刘, Y.; Miura, J. RDMO-SLAM: 使用光流语义标签预测的动态环境实时视觉 SLAM。IEEE接入2021年, 9, 106981–106997。[交叉引用]