* **视觉语义SLAM关键技术研究**

单位：浙江大学

类型：硕士学位论文

年份：2021

**摘要：**

本文主要研究内容：通过利用深度网络提取特征点和语义类别信息，并融入到传统视觉SLAM中。1. 深度学习提取特征点和描述子；2. 基于语义类别信息的SLAM系统。

稀疏地图构建仍是主流；在复杂环境中传统视觉SLAM的鲁棒性不理想，深度学习擅长提取图像的深层特征和语义信息，与视觉SLAM相结有望获得更好的鲁棒性。

**关键词：**SLAM，特征点法，深度学习，语义分割，BA优化

* **融入语义信息的VSLAM研究综述**

类型：期刊文章/贵州大学学报

日期：2022.09

**摘要：**

在传统VSLAM框架基础上引入深度学习获取环境语义信息，构建3D点云语义地图。根据传统VSLAM在视觉里程计、回环检测、建图上的不足，通过语义信息来改善VSLAM系统的性能。总结与展望。

**关键词：**SLAM，深度学习，语义分割，3D语义地图，高层次交互任务

**前言：**

1. 传统VSLAM：视觉传感器、前端（视觉里程计）、后端（非线性优化）、回环检测、地图构建
2. VSLAM相关方案：LSD-SLAM，RGBD-SLAMV2，ORB-SLAM
3. 神经网络方案：SSD，YOLOv3，SegNet，MaskRCNN
4. **语义与视觉里程计**

**视觉里程计**，从传感器获取数据进行处理，用于估计位姿和短期的局部地图构建。位姿估计的准确性很大程度上取决于帧与帧之间的数据关联和帧中关键点的提取。

**不足一**，在一些弱纹理和光照强烈变化的环境中，往往不能提取到足够的关键点用于位姿估计；**不足二**，环境中出现的动态物体会造成错误的数据关联，给位姿估计造成较大误差。

**优化思路一，**直接从图像中提取语义信息用于位姿估计；**优化思路二，**将语义信息作为约束条件，协助提高位姿估计精度。思路二中更倾向于多种约束条件共同作用，来解决数据关联中的误关联问题。

1. **语义与回环**

**回环检测，**校准位姿估计中的累积误差，对同一位置进行有效识别，对位姿轨迹实现有效闭环。传统方法是词袋模型和相似性度量。

**不足一，**为了达到有效区分，词典树往往较为庞大，在一些移动应用上受到限制；**不足二，**相似性度量计算量庞大。

**优化思路一，**语义信息协助传统方法提高相似特征搜索的速度和精度；**优化思路二，**卷积神经网络提取语义特征，使用特征点进行相似性度量，判断回环。

1. **语义与建图**

传统VSLAM可以构建出度量地图和拓扑地图，缺少环境内容信息，实用性不高。

**不足，**机器人虽然能感知环境中的空间占据信息，却无法深入理解物体的具体信息。

**优化思路一，**将2D语义信息与VSLAM构建出来的地图按照一定的方式融合，使地图具有语义信息；

**优化思路二，**基于对象模型，首先分割场景中的语义对象，其次由不同的对象构成环境地图；

**优化思路三，**采用多视角方式，将不同视角的语义信息相结合，构建信息更丰富法环境地图。

1. **总结和展望**

**总结：**

前端（视觉里程计）主流选择：与其他动点检测方法形成约束解决数据关联问题。

语义分割网络在视觉里程计和回环检测中的研究还不足。

构建地图：语义信息得到了充分利用和拓展。

**趋势：**

一方面，语义信息的利用呈现多约束条件共同作用的趋势。例如语义信息与视觉惯导的融合、不同语义信息相互约束、多SLAM系统相互约束。

另一方面，深入各行各业。例如稀疏语义地图增强、面向交互的局部稠密化技术。（可应用于AR空间打造、服务机器人、无人机低空自主飞行）

此外，面向边缘计算的语义VSLAM实现架构与算法部署是上述应用得以实时实施的关键。

* **基于环境语义信息的同步定位与地图构建方法综述**

类型：期刊文章/工程科学学报

日期：2021.06

**摘要：**

传统VSLAM技术在实时性上达到了较高水平，但在精度和鲁棒性上有较大缺陷，交互性能不足。本文综述了环境语义信息运用在同步定位与地图构建领域在最新研究进展和突出研究成果。并将传统VSLAM和语义SLAM做了对比。展望未来。

**关键词：**SLAM，深度学习，系统定位，地图构建，语义同步定位与地图构建技术

**前言：**

视觉SLAM系统大多将估计相机位姿作为主要任务，通过多视几何理论构建三维地图。

1. **语义SLAM系统概述**

借助语义信息一方面提高定位精度和鲁棒性，另一方面借助语义信息将数据关联从传统的像素级别提升到物体级别，帮助机器人进行自主环境理解和人机交互。

1. **语义与SLAM系统定位**

移动机器人对系统定位的要求：定位实时性、定位精度、定位鲁棒性

定位实时性以达到较高水平，定位精度和鲁棒性有待提高

* 1. **语义与定位精度**

1. **面向单目初始化**

**问题：**尺度模糊、随时间漂移

**研究成果：**

一、基于混合多尺度可变模型的目标检测系统检测物体，并与SLAM系统中的关 键帧进行数据关联，然后对物体尺寸进行估计，生 成地图地标，最后地标与相机位姿共同参与局部 优化，以解决尺度模糊和漂移的问题，保证了全局 地图的一致性。

二、贝叶斯框架内， 通过基于深度学习的目标检测算法观测物体，确 定其先验高度，然后将局部地图中3D点投影到目 标检测的2D图像上，确定目标检测区域的边界点 并投影回3D空间，进而确定物体的实际高度，最 后将先验高度与实际高度之比设置为尺度因子，用于单目SLAM系统中的尺度矫正，从而准确估 计移动机器人位姿。

1. **面向数据关联**

**VSLAM中关联数据分为：**短期关联（特征匹配）、长期关联（回环检测）

**新思路：**基于中期数据的关联机制

**研究成果：**

一、概率数据关联机制

二、全新视觉语义里程计框架（VSO）。实现中期连续点跟踪，可以直接融入直接法和间接法视觉里程计当中。改善了漂移现象。

1. **面向位姿优化**

**常规优化算法：**基于粒子滤波的优化算法、基于非线性优化的算法

**优化思路：**

在基于粒子滤波的优化算法中，可利用环境语义信息更新粒子状态，进而更精准的进行位姿估计。

在非线性优化的算法中，可以借助目标检测算法，将物体作为路标，融合到一个优化算法中。

对象级的定位与建图，与基于特征点的SLAM系统相比，对象级的SLAM系统可以提供更多的几何约束和尺度一致性，从而得到更好的精度。

1. **面向重定位与回环检测**

**重定位：**跟踪丢失时重新找回当前姿态，恢复相机位姿估计。

**回环检测：**识别机器人曾经到过的场景，得到几何一致的映射，从而解决漂移问题。

**问题：**传统的SLAM算法大都依赖于低级别的几何问题，使得重定位和回环检测通常依赖与相机的视角，在特征不明显或重复性的纹理环境中容易检测失败。

**优化思路：**语义SLAM系统利用目标识别推测标志物的种类和大小，产生易于识别的标志物，进而提高系统的定位精度。

* 1. **语义与定位鲁棒性**

鲁棒性是指在复杂环境中容易出现信息丢失的情况。

1. **面向弱纹理环境**

**问题：**特征可区别性较弱，缺少显著特征。

**优化思想：**利用深度学习提取环境中的立体特征并用于优化位姿估计。

1. **面向光照变化**

**问题：**传统VSLAM系统中，特征描述子对光照变化敏感，光照的变化容易导致目标表面颜色等特征的变化，对系统的定位鲁棒性造成干扰。

**优化思路：**

一、环境中的物体类别与光照和时间变化无关，可以利用环境语义信息构造新的特征描述子，从而优化鲁棒性。

二、通过优化特征选择策略，结合语义信息进行过滤，使得所关注目标对于光照更加鲁棒。

1. **面向动态场景**

**传统VSLAM存在问题：**传统VSLAM大多基于静态环境，处理动态环境的定位问题时，容易出现错误匹配。

**优化思路：**引入环境语义信息有效的对静态和动态特征点进行划分，移除动态特征点，从而减少干扰，提高鲁棒性。

运动一致性校验

多视几何和深度学习方法实现移动物体的检测

**语义SLAM存在问题：**目前大多语义信息用于动态场景是将所有潜在运动物体直接剔除，但如果动态目标物体并没有发生运动，且占据相机视角的大部分，那么将目标物体上的特征点都剔除，会对位姿估计产生严重误差。

**优化思路：**

一、引入对于地图点是否为静态点的概率估计，实现地图点在动静态之间平滑过渡。

二、(SOFSLAM). 该系统采用一种新的语义光流动态特征 检测方法，使用SegNet[49]产生像素级的语义分割 结果，获取运动先验（静态、潜在动态的和动态）， 并将运动先验作为掩膜去除动态和潜在动态特征 之间的匹配，然后利用语义静态特征的匹配计算 基本矩阵，最后利用基本矩阵和对极约束寻找并 剔除真正的动态特征，在跟踪和优化模块中保留 剩余的静态特征，实现动态环境下相机位姿的精 确估计。

* 1. **小结**

**对比：**相较于传统VSLAM，语义SLAM可以直接提取图像特征，无需人工特征提取和匹配。

**目前存在问题：**

目前环境语义信息主要依赖深度学习的方法，不同算法间网络架构差异大；

对训练数据库有较强的依赖性；

系统性能严重依赖于环境目标识别和语义分割结果的准确性。

**发展现状和趋势：**

当前环境语义信息多与SLAM局部的子模块相结合，例如位姿估计和回环检测等。

也有部分学者开始尝试搭建端到端的SLAM系统。（这个创新方向太新，不要尝试）

1. **语义与SLAM地图构建**

**稀疏**点云地图用于定位；

**稠密**点云地图、八叉树地图用于重建、避障、导航；

**语义**地图用于交互。

语义信息与SLAM地图构建融合的**研究方向：**

面向场景的语义地图构建：侧重于对环境的感知；

面向对象的语义地图构建：侧重于机器人与实体进行交互。

* 1. **面向场景的语义地图构建**

**（1）面向场景的语义地图：**是指使用语义分割算法对2D图像进行像素级分割，提取图片中的语义信息，并将其与3D点云进行融合，以构建全场景的语义地图。

**（2）早期构建方法：**

例：SemanticFusion算法：构建像素级语义地图的典型代表，一个完整的语义SLAM系统，实现了室内环境语义地图构建。

早期方法缺点：虽然是完整的语义地图构建框架，但是计算资源消耗大，难以在机器人上实际部署使用。

**（3）轻量化构建方法**

例：SRM算法：提高了语义地图构建实时性；

Kimera算法：精度方面达到较高水平，实时性较差。

**（4）**同时，有人倾向于通过控制计算资源的调用解决面向场景的语义地图构建算法存在的实时性和精度问题。

**（5）面向对象的语义地图构建引出**

**研究方法：**面向场景的语义地图大多采用深度学习方法，将2D语义信息映射到3D点云中。目前的方法主要围绕语义分割和语义融合。

**问题：**机器人无法与环境中的个体进行交互；像素级语义分割计算量较大，实时性不理想。

**改进：**基于目标检测算法和点云分割算法提出了面向对象的语义地图构建算法。

* 1. **面向对象的语义地图**

**面向对象的语义地图**是指语义地图中仅包含 部分实例的语义信息，语义信息以聚类的方式独 立于地图. 因此，可以允许机器人对地图中每一个 实体的语义信息进行操作和维护。

**早期算法：**大多只关注了地图构建问题，而忽略了利用动态物体属性优化系统性能。普遍实时性较差。

**问题：**面向对象的语义地图构建算法将工作重心置于实体本身，包括环境中每一个实体的类别信息和位置信息，在这种地图中，实体独立于环境。**但是，**目前面向对象的语义地图构建方法大多需要同时处理实例分割和语义分割两项任务，系统的实时性比较差. **另外，**在处理动态物体时，大多数算法直接剔除动态物体，这对于地图构建来说信息损失较多，因此，如何表示动态环境中的动态物体是一个亟需解决的问题。

* 1. **小结**

上述两种语义地图占用空间大，无 法应用于大规模场景工作的移动机器人.

在度量地图基础上构建的语义地图虽然在细节上对场景的表现更加丰富，但是通常无法表达某一场景的语义类别，因此，拓扑地图作为传统SLAM研究中不可或缺的一部 分，因其可表示环境的连通关系，并显著降低地图的存储空间

这种拓扑‒度量融合的混合地图形式一直是机器人导航领域的研究重点

综上所述，面向场景的语义地图和面向对象的语义地图均需要大量的计算资源来处理地图构建中的语义获取和数据融合，而拓扑语义地图难以满足机器人与环境交互的需要.因此，部分学者开始研究提取环境中的标志物、文字等信息替代 传统意义的语义目标，从而简化语义地图的构建

1. **语义SLAM与传统SLAM对比**

**传统VSLAM：**

关键性能：精度、鲁棒性、实时性

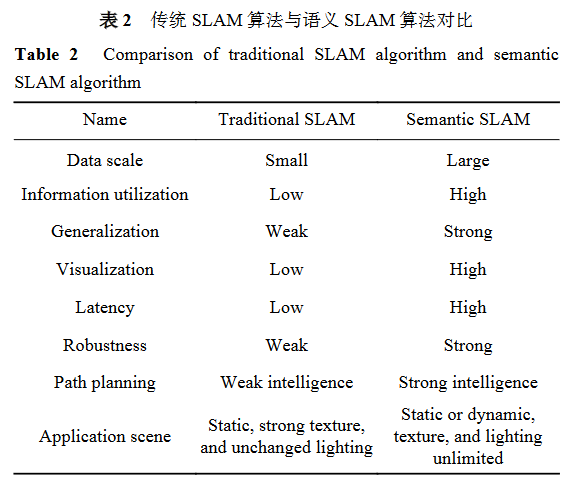
构建地图时提取的是环境中的几何特征信息，常见的有SIFT和SURF特征。几何特征不能满足导航、避障、交互。

**早期语义地图构建：**

一般采用模型匹配库的方式。局限性大。

**现状：**

机器学习与传统VSLAM相结合，能够直接学到环境中高层次特征。



**总结：**（1）传统SLAM方法一般基于静态环 境假设,而语义SLAM可以感知物体（人、汽车、动 物等）的可移动属性；（2）语义SLAM中的物体知 识描述可以共享，通过维护共享知识库可提高 SLAM系统的可扩展性和存储效率；（3）语义SLAM 可提高机器人导航的智能化程度，如机器人在搬 动物体的任务中，利用语义信息优化路径.

1. **未来展望**
   1. **SLAM对语义的优化与联合推理**
   2. **端到端SLAM系统**

机器人在未知环境中根据视觉传感器获取的图像视频信息，直接生成一系列的任务决策，这一般被称为端到端的视觉SLAM系统.

端到端的SLAM系统可以解决同时进行学习环境特征、位姿估计和任务决策等任务。

重点研究方向：鲁棒性差；训练难度大；样本获取困难。

* 1. **主动SLAM系统**

机器人自主探索环境

拓扑语义SLAM

重点研究方向：如何利用环境中的语义对象更加高效生成探索目标是提升机器人探索效率的重要方向.

* 1. **多机器人系统**

多机器人系统一般采用两种方式构建环境地图：一种是每个机器人单独构建局部地图，再将所有的地图进行合并；另一种是利用子母式的多机器人系统，子机器人负责感知环境，并将环境信息发送到母机器人，母机器人负责地图构建。

* 1. **语义评价标准**

**SLAM系统指标：**绝对轨迹误差（ATE）；相对位姿误差（RPE）

以上两个标准针对位姿估计结果进行评定，对地图构建暂时没有公认的评价标准。

**语义SLAM系统标准：**语义信息获取的准确性；如何评价语义地图构建效果

1. **结论**

目前语义SLAM在 工业界尚未受到广泛关注，主要限制在于精确的 语义分析需要大量计算资源，实时性受到制约. 此 外，语义SLAM系统的精度和鲁棒性依赖于环境 语义信息提取的准确性.

* **机器人室内语义建图中的场所感知方法综述**

类型：期刊文章\自动化学报

日期：2017.4

**摘要：**

根据实现场所感知目标所采用的线索对已有方法进行分类介绍. 主要分成 3 个大类: 基于环境布局几何信息的方法、基于环 境布局视觉信息的方法、基于用户指导信息的方法，此外还将一些特殊方法进行单独分类，并阐述特点和局限性。总结与展望。

**关键词：**场所分类；语义建图；室内语义地图；场所感知；语义场所标注

1. **概念辨析**

场所可以从事特定活动，涉及到特定物品和行为，有对应的符号描述或者代指的词汇标签类。

1. **利用环境布局的几何信息实现场所感知**

**2.1基于2D几何信息的场所感知**

* **动态环境下的语义地图构建**

类型：期刊文章/计算机科学（开源代码中的论文）

日期：2018.04

**摘要：**

现SLAM存在的问题：如何通过剔除动态物体来获取较高的定位精度；理解周围场景中存在的物体及其位置。

本文：动态环境下构建语义地图，在ORB-SLAM2上进行改进。

跟踪线程中加入动静点检测，用来剔除动点特征点；光流法；

目标检测线程对关键图像进行目标检测；

地图构建线程构建Octo-Map地图，构建3D目标数据库；

关键词：动点检测，目标检测，语义地图构建，视觉SLAM，光流法

* 1. **引言**

SSD目标检测框架：保证检测精度的同时提高了检测速度，相比YOLO框架提高了检测精度，相比RCNN框架提高了检测速度。

实验室环境背景

结合基于RGBD相机的视觉SLAM系统和基于回归预测的多尺度预测的深度卷积神经网络SSD框架来设计算法，以剔除动态物体点、融合位姿信息和语义信息以及构建语义目标数据库，实时构建高层次语义地图。

MobileNet小型化网络：可以再嵌入式平台上运行，运算量降低。

ORB-SLAM2：基于特征点的视觉SLAM算法。

* 1. **动态信息检测和3D语义信息获取**

提高定位精度：光流法动态检测

语义地图构建：结合2D目标检测和3D点云信息

目标检测实时性：Mobile Net网络和SSD检测框架

3D定位精度：点云聚类分割

* 1. 动态信息检测

基于光流

首先设置一个光流幅值较大的阈值，然后利用相邻两帧特征点求解光流场，由于匹配点对数较多，需要用RANSAC方法提高鲁棒性。

* 1. 2D目标检测

用Mobile Net网络作为SSD目标检测网络的前端提取网络。

* 1. 3D语义信息获取
* **Ongoing evolution of visual SLAM from geometry to deep learning : challenges and opportunities**

Date: 2018.12

**Abstract:**

基于几何模型的视觉SLAM的发展进程，引出深度学习是一项非常可行的方法，结合语义分割阐述语义信息在导航等场景中的优势。

1. **Introduction**

基于模型的视觉SLAM可分为基于特征点法和直接法两类。基于特征的方法是在图像中提取稀疏特征信息，用于特征匹配和运动估计；基于直接法是在光度一致的假想下直接用稠密或者半稠密像素进行运动估计。

系统建图鲁棒性和高层次语义感知是基于模型的VSLAM面临的挑战。

深度学习的方式能够自动的从端到端中学习特征信息，不需要人为的设计特征读取方法，提取到的特征也更加稳定有效。

在视觉SLAM中融合深度学习方面的技术主要体现在三个方向：单目图像场景深度估计；视觉里程计估计；语义地图生成。

1. **Model-based SLAM methods**
2. **Feature-based visual SLAM methods**

ORB-SLAM 特征法中的代表

RGB-D SLAM能够生成稠密且精确的三维地图。

1. **Direct visual SLAM methods**

DTAM系统，每一帧能够生成一个稠密的三维地图。

SVO 可以在CPU上运行的半稠密视觉里程计；

LSD-SLAM 可以在CPU上运行的大规模直接单目视觉；

DSO 直接稀疏里程计。直接法中的代表。

NID-SLAM 单目相机直接方法，在外观变化的环境中表现稳定。

1. **summary**

ORB-SLAM 是特征点法的代表，DSO 是直接法的代表，深度学习可以提高视觉SLAM 系统的鲁棒性。

1. **Deep neural networks for vision SLAM**
2. Convolutional neural network (CNN)

卷积神经网络：

1. Deep recurrent neural network (RNN)

循环神经网络：主要用于捕捉视频片段的时间动态性。LSTM 被广泛应用来解决简单RNN网络的梯度消失问题。

RNN与CNN通常一起使用，形成RCNN，在视频片段中捕捉特征。适用于在视频片段中的位姿估计问题。

1. Auto-encoder

自编码器，自CNN演变而来，一种特殊的深度神经网络。自从提出全连接网络( FCN ) 以来，自动编码器被广泛用于深度估计和语义分割。

1. Dataset

TUM/NYU：深度相机采集的室内彩色深度图像。NYU还提供了一些用于语义分割的标注图像。

PASCAL VOC/synthia dataset/COCO dataset：针对图像分割问题

ADE20K：包含20000个像素级的语义标注图像

1. **Depth estimation with deep learning**

**深度估计**

Data-drive方法为深度估计提供了一个思路，基于深度学习的深度估计可分为有监督和无监督方法。

1. **Supervised methods (有监督学习)**

CNN-SLAM：仅在深度估计方面使用CNN，其他环节仍然使用特征点法。

有监督的方法需要大量的标签数据来训练网络，由于收集标记数据集的成本太高，限制了其应用。

1. **Unsupervised methods（无监督学习）**

最近出现了使用无监督深度学习的深度估计方法。其主要思想来源于自编码器的表示能力。编码器是一个CNN，为左输入图像预测深度图，解码器是一个wrap函数，从右输入图像和预测深度图合成一个重建的左图像。重构误差作为代价函数来训练CNN。

无监督学习在位姿估计、重定位和深度估计等方面具有较大潜力。

1. **Summary**

无监督学习不需要标签数据，适用于自主学习。无监督学习的深度估计在构建稠密地图时也很重要。

1. **Pose estimation with deep learning**
2. **Relocalization with deep learning**

重定位

1. **Ego-motion estimation with deep learning**

自运动，两个时刻的相对运动

1. **Sensor fusion with deep learning**

传感器融合

1. **Semantic mapping with deep learning**

将语义分割和视觉SLAM相结合，可以同时估计语义三维地图和相机运动。

1. **Semantic segmentation**

介绍了几种采用VGG作为基础网络架构的神经网络用于语义分割，

ResNet残差网络：已广泛应用于语义分割任务。或许浅层网络效果比深层网络效果更好。

金字塔解析网络PSPNet：

SegNet：编码器-解码器架构

1. **Semantic mapping**
2. **Summary**

Weekly supervised learning（弱监督学习） 用于视觉SLAM成为可能

1. **Open challenges and future opportunities**

鲁棒性的提高、语义信息的融合以及学习能力的融入称为视觉SLAM下一步发展的核心。

1. **ImageNet-Scale dataset for learning based visual SLAM**

无监督学习通过增加未标记的数据集数量，可以不断提高性能，无监督学习有望利用大规模的数据，提升视觉SLAM在鲁棒性和语义理解方面的能力。

1. **Semantic SLAM with high-level understanding**

对象级语义信息具有提高位姿估计精度和鲁棒性的潜力，二位姿估计也可用于语义分割。

对环境的高层次理解也是语义SLAM的发展方向

1. **Adaptive SLAM methods for diffirent sensing modalities**

针对不同传感器模态的自适应SLAM

深度学习框架下的多传感器融合和管理，可以容纳不同的感知模式来代替标定过程。

1. **Integration of model based methods with deep learning**

基于模型的SLAM过度依赖于特征的成功检测与匹配，利用深度学习可以准确提取复杂环境下的特征，有监督的深度学习来增强鲁棒性和提取匹配的特征已经实现，无监督的深度学习来增强鲁棒性和特征匹配是另一种思路。

如何利用深度学习保持地图的全局一致性。

1. **Conclusions**

趋势：更高层次的认知和感知，具有学习能力和自适应能力的视觉SLAM系统，在估计结果中增加更有意义是结果，比如姿态、深度、三维点云、语义地图。

大规模数据集的可用性是深度学习方法广泛应用的关键。采用无监督学习的尝试有望进一步巩固深度学习对视觉SLAM的贡献。

* **An overview on visual SLAM: from tradition to semantic**

Date：2022.06

**Abstract：**

传统VSLAM和结合深度学习VSLAM，传统从间接法、直接法介绍，结合深度学习的从CNN\RNN入手，重点研究了目标检测和语义分割对VSLAM的影响。

1. **Introduction**
2. **Development status of SLAM**
   1. **sensors commonly used in VSLAM**
   2. **assessment tools and dataset**
   3. **SLAM development analysis based on literature data**
   4. **Outstanding scholars and teams**
3. **Traditional VSLAM**
   1. **monocular/stereo VSLAM**
      1. **VSLAM based on the feature-based method**
      2. **VSLAM based on direct method**
   2. **RGB-D SLAM**
   3. **Visual-inertial SLAM**

**视觉惯性SLAM**

* + 1. **loosely coupled visual-inertial**
    2. **tightly coupled visual-inertial**

1. **semantic VSLAM**

深度学习在VSLAM中的应用降低了对特征点的依赖性，提高了系统的鲁棒性。深度学习可以代替传统视觉SLAM中的一些环节，例如回环检测、位姿估计，来改进视觉SLAM系统。

* 1. **neural networks with VSLAM**

**CNN：**从图像中捕获空间特征，识别物体间的关系。优势在于提取特征然后进行分类、识别、预测、决策。

**RNN：**可以处理图像或数值型数据，本身具有记忆功能，可以帮助处理上下文特征。优势在于建立相邻帧之间的关系，高层特征具有更好的区分度，可以更好的完成数据关联。

* + 1. **CNN with VSLAM**

CNN通过训练数据进行学习，避免了特征提取。

CNN可以与VSLAM的特征提取与匹配、深度估计、位姿估计等多个环节相结合，并可以取得较好的成果。

**应用：**

与单目相机相结合，视觉深度估计弥补了单目相机无法进行深度估计的问题；

也改善了传统VSLAM中位姿估计的缺陷；

复杂环境下跟踪丢失的问题；

在闭环检测中使用图像中更高层次的语义信息进行匹配，能克服复杂环境，匹配特征耕鲁棒；

随着更复杂更优的深度学习模型的引入，在提高VSLAM系统各项性能的同时，如何保持系统的轻量化也是一个问题。

* + 1. **RNN with VSLAM**

**RNN**具有记忆和参数共享的特点，在学习非线性特征方面具有一定优势。在VSLAM中的适用范围较小，但在建立相邻帧之间的一致性具有很大的优势。RNN是深度学习中数据驱动时序建模的常用方法。IMU输出的数据对时序有严重依赖型，特别适用于RNN模型。RNN在融合视觉-惯性数据时更有效

**LSTM**（long short-term memory networks）：最常见的循环网络之一。可以记住和合理忘记一定数量的先前训练的数据，解决了反向传播的梯度消失问题。

**GRU：**易于训练，可提高训练效率

深度学习学习特征：将旋转和平移分开训练，比传统方法具有更好的适应性。

**DROID-SLAM：**核心是一个更新算子，该算法用于端到端神经网络架构，在复杂环境中具有优势

在位姿估计中，引入端到端的深度学习方法，不需要特征匹配等复杂的几何方法，求解图像帧之间的相对位置只需要比对相邻帧的参数即可。

现有方法有很多是CNN和RNN相结合的方式来提高VSLAM整体性能，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对时序模型(运动模型)和数据关联模型(图像序列)进行隐式建模。

* 1. **modern semantic VSLAM**

根据目标检测方法的不同，语义SLAM可以分为两种：一种是传统方法检测目标，使用大量二进制单词和数据库提供实时检测。另一种是使用深度学习的方法，利用语义信息可以将数据关联从传统的像素级提升到对象级，提高复杂环境的准确性。

* + 1. **image information extraction**

**图像信息提取**

目标检测和图像语义分割是从图像中提取语义信息的方法。图像的语义分割是在像素级别上理解图像，获取图像中的深层信息。

**目标检测：**目标检测只获取图像的目标信息和空间信息。目标检测只用候选框框出物体类别，因此比语义分割更快，但语义分割具有更高的精度，但速度要慢得多。YOLO算法是目前应用最广泛的目标检测算法之一。

**语义分割：**语义分割可以预测不同物体的位置和类别，还可以准确描绘不同类别目标之间的边界。但对于语义分割，普通的卷积网络无法获取足够的信息。基于语义分割，提出的算法有：

FCN：全卷积神经网络。将局部信息与全局信息相结合。

SegNet：提出了编码-解码结构，将两个独立的网络结合，提高了分割精度，但牺牲了运行时间。

PSPNet：和金字塔结合，将每个层次融合输出。

Mask R-CNN：不仅具有像素级的分类（语义分割），还具有对象的位置信息（目标检测）。主要思想是在Fast R-CNN基础上增加一个分支进行语义分割。

**summary：**在精度方面以达到一定高度，但在实时性方面仍需提高。

* + 1. **semantic with location**

**语义和定位**

引入语义信息可以有效改善视觉SLAM中尺度不确定性和累计漂移，从而提高定位精度。通常引入语义信息可以避免光照等复杂环境的影响，但同时也会减缓系统的运行速度。

* + 1. **semantic with mapping**

**语义和建图**

在获取环境中特征点等几何信息的同时还获取语义对象的位置、属性、类别等语义信息。利用目标检测、语义分割等算法获取语义信息，进而集成到环境地图中构建语义地图。构建语义地图的研究方向主要分为面向场景和面向对象的语义地图构建两个方向。

**面向场景：**大多基于深度学习的方法将二维语义信息映射到三维点云。可以帮助机器人更好的理解环境。此类算法需要对场景中的物体进行像素级语义分割，导致计算量大，实时性较低。

Kimera：一种成熟的面向场景的语义SLAM算法。

**面向对象：**指只包含部分实例语义信息，且语义信息以聚类的方式独立存在的地图。这种地图允许机器人操作并维护地图上每个实体的语义信息，更利于交互。目前大多数语义地图构建方法需要同时处理实例分割和语义分割，导致系统实时性差。此外，在处理动态对象时，大多算法通过消除动态对象来获取鲁棒性，但同时会丢失很对有用信息，因此处理动态场景是个待解决的问题。

* + 1. **elimination of dynamic objects**

**消除动态物体**

1. **conclusions and prospect**

对于语义建图应用场景的调查

动态场景：空中机器人

自动驾驶：

* **Efficient construction of topological semantic map with 3D information**

基于三维信息的拓扑语义图的高效构建

Date：2018.10

**Abstract：**

拓扑语义地图框架，我们通过将卷积神经网络识别的三维场景语义信息投影到二维平面上来构建二维语义地图。通过RGB - D SLAM三维空间建图算法实现环境的三维重建。离线识别二维地图中的交叉点，利用拓扑地图中交叉点的语义标注构建完整的基于对象的语义地图。

* **SDF-SLAM: Semantic Depth Filter SLAM for Dynamic Environments**

SDF-SLAM：用于动态环境的语义深度过滤器 SLAM

Date：2020

**Abstract：**

面向动态环境的视觉语义SLAM系统SDF - SLAM：Semantic Depth Filter SLAM，利用深度滤波技术直接判断三维地图点是否为动态。

首先，通过语义光流法将语义信息融入到原始的纯几何SLAM系统中，进行可靠的地图初始化。其次，设计满足高斯均匀混合分布的语义深度滤波器来描述每个地图点的逆深度。第三，在贝叶斯估计框架中更新三维地图点的逆深度，并将三维地图点分为活跃的和不活跃的。最后，仅利用主动地图点实现鲁棒的相机位姿跟踪。

* **SOF-SLAM: A Semantic Visual SLAM for Dynamic Environments**

SOF-SLAM：动态环境的语义视觉SLAM

Date：2019

**Abstract：**

SOF - SLAM：Semantic Optical Flow SLAM，一种面向动态环境的视觉语义SLAM系统，基于ORB - SLAM2的RGB - D模式构建。

去除动态特征

提出了一种新的动态特征检测方法- -语义光流法，有效合理地去除动态特征。SegNet生成的像素级语义分割结果作为所提出的语义光流中的掩膜，得到一个可靠的基础矩阵，然后用于筛选出真正的动态特征。跟踪优化模块只保留剩余的静态特征，实现动态环境下相机位姿的准确估计。

* **Dynamic Scene Semantics SLAM Based on Semantic Segmentation**

基于语义分割的动态场景语义SLAM

Date：2020

**Abstract：**

本文结合视觉SLAM系统ORB - SLAM2和PSPNet语义分割网络，提出了PSPNet - SLAM系统，利用光流和语义分割检测并剔除动态点，实现动态场景语义SLAM。

关键词：光流法；室内动态场景；剔除动态点

* **An FPGA Based Energy Efficient DS-SLAM Accelerator for Mobile Robots in Dynamic Environment**

基于FPGA的高能效DS-SLAM加速器，用于动态环境中的移动机器人

Date：2021

**Abstract：**

DS- SLAM (面向动态环境的语义SLAM)是视觉语义SLAM的代表性工作之一。

* **Object Detection-based Semantic Map Building for A Semantic Visual SLAM System**

基于对象检测的语义视觉SLAM系统语义图构建

Date：2020

**Abstract：**

基于使用RGBD相机的视觉SLAM系统创建，基于语义对象的地标被连续保存到语义地图中，不仅可以描述上下文信息，还可以在动态环境中更新语义地图，对象可以被被动移动。

* **DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments**

DS-SLAM：面向动态环境的语义视觉SLAM

Date：2018

**Abstract：**

一种面向动态环境的鲁棒语义视觉SLAM，即DS - SLAM。DS - SLAM中五个线程并行运行：跟踪、语义分割、局部建图、闭环和稠密语义地图创建。DS - SLAM将语义分割网络与移动一致性检测方法相结合，减少了动态物体的影响，在动态环境下定位精度得到了很大的提高。同时，生成稠密的语义八叉树图，可用于高层任务。它是高动态环境下最先进的SLAM系统之一。

DS-SLAM采用SegNet实时提供基于caffe的像素级语义分割。

1. **introduction**

本文将语义分割网络与光流法相结合，重点研究了在基于视觉的SLAM中减少动态物体的影响，同时提供八叉树地图的语义表示，可用于机器人的高层任务。DS - SLAM概述如图所示。

主要贡献：动态场景中精度和鲁棒性优于ORB-SLAM2；集成了ROS系统；过滤掉动态部分；创建一个单独的线程来构建稠密八叉树地图

1. **Related works**
   1. semantic SLAM

语义地图通常分为几何部分和语义部分

为了准确地进行语义分割，我们使用语义分割网络来遵循最新的工作路线。

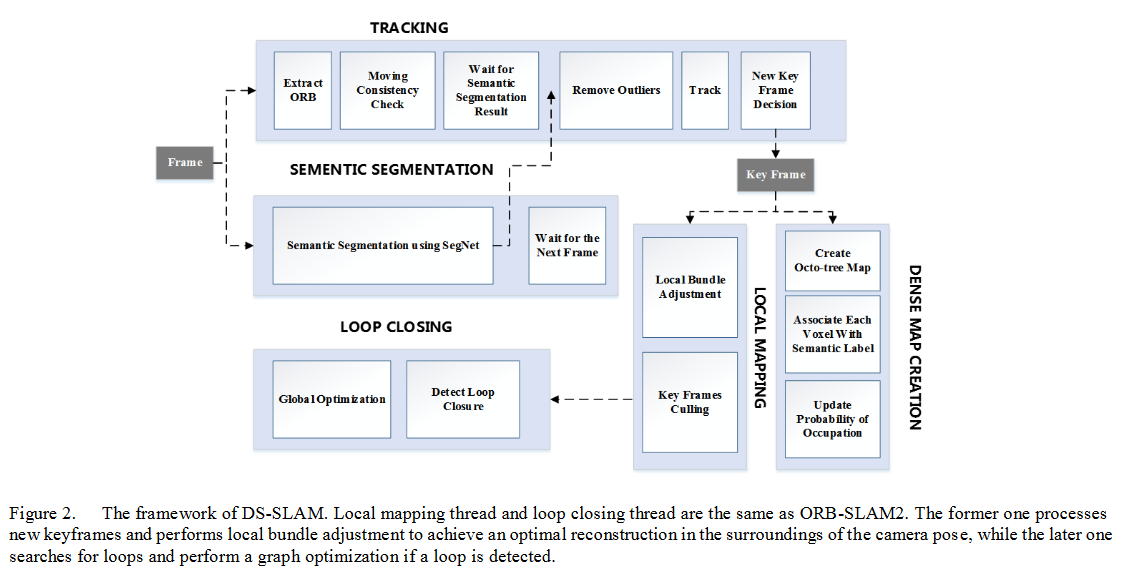
语义信息用于八叉树地图构建和剔除异常值。

* 1. SLAM in dynamic environments

静态环境中识别出动态对象然后丢弃，通过光流的不一致性来区分静态背景和动态目标。

1. **System introduction**

本部分将详细介绍DS - SLAM。本部分包含五个方面的内容。首先，给出了DS - SLAM框架的结构图。其次，简要介绍了DS - SLAM中采用的实时语义分割方法。然后介绍了用于检查特征点移动一致性的算法。随后，展示了异常值剔除方法，该方法结合语义分割和移动一致性检查来过滤动态对象。最后，我们给出了构建语义八叉树地图的方法。



* 1. framework of DS-SLAM

tracking, semantic segmentation, local mapping, loop closing and dense map creation

跟踪、语义分割、局部建图、闭环和稠密地图创建

Kinect2采集的原始RGB图像在跟踪线程和语义分割线程中同时进行处理。跟踪线程首先提取ORB特征点，然后粗略检查特征点的运动一致性并保存潜在的异常值。然后跟踪线程等待语义分割线程预测的具有像素级语义标签的图像。在分割结果到达后，基于该结果和之前检测到的潜在异常值，将位于运动物体中的ORB特征点异常值丢弃。然后，通过匹配剩余的稳定特征点计算变换矩阵。

* 1. semantic segmentation

segnet在实时性上优于像素级语义分割的caffe。

* 1. moving consistency check

语义分割结果可以从另一个线程获得，所以只需要判断分割结果中的关键点是否运动，如果一个分割的对象中一些点是运动的，那么可以认为这是一个动态对象。

本文识别动态点：第一步是计算光流金字塔得到当前帧中匹配的特征点。第二步如果匹配对距离图像边缘过近或者匹配对中心的3 × 3图像块像素差过大，则舍弃匹配对。第三步，利用内点最多的RANSAC寻找基础矩阵。然后利用基础矩阵计算当前帧中的核线。最后，判断匹配点到其对应核线的距离是否小于某一阈值。如果距离大于阈值，则判定匹配点在移动。

* 1. outliers rejection

剔除异常值

仅用运动一致性提取动态物体轮廓困难且耗时，采用语义分割网络可以很容易得到完整轮廓。将两者结合以完成两级语义知识库额建立：动态物体和静态物体。如果运动一致性检验产生一定数量的动态点落在分割物体的轮廓中，则确定该物体是运动的，并剔除位于目标轮廓中所有的特征点。

在DS - SLAM中，由于使用了语义分割网络，可以很容易地获得物体的完整轮廓。思路是结合语义信息和移动一致性检查结果，完成物体移动或不移动的两级语义知识库的建立。如果有一定数量的运动一致性检验产生的动态点落在分割对象的轮廓上，则判定该对象是运动的。如果分割后的物体被判定为运动，则去除位于物体轮廓上的所有特征点。

理论上，DS - SLAM适用于任意多个被识别和分割的动态物体。包括人但不局限于人。

在语义分割结果出来后，如果没有人被检测到，那么所有的ORB特征将直接与最后一帧进行匹配来预测位姿。否则，利用移动一致性检查结果判断人员是否移动。如果确定人员为静止状态，则直接进行位姿预测，否则在匹配前将落在人员轮廓内的所有ORB特征点剔除。

* 1. dense semantic 3D octo-tree map building

语义八叉树映射线程从跟踪线程得到新的关键帧，从语义分割线程得到分割结果。利用关键帧的变换矩阵和深度图像生成局部点云，然后将局部点云转换并保存在局部八叉树地图中。

八叉树地图灵活、紧凑、可更新、存储高效、易于导航。

建模和语义融合过程都是以概率的方式完成，因此可以方便更新体素属性。

DS - SLAM面向处理动态环境，因此地图中不应存在动态物体。

为了解决复杂情况下，语义分割不完整或错误的情况，提出一种评分方法，来最小化动态物体的影响。被观察到多次被占用的体素认为是稳定被占用的体素。

1. **Experimental results**
   1. TUM

高动态环境：行走的人

低动态环境：坐着的人

DS-SLAM能够有效地处理高动态环境，精度和鲁棒性明显提高，但对于低动态环境，精度和鲁棒性提升不明显，原因是ORB-SLAM2能够有效处理低动态环境，提升空间有限。

实时性也能达到要求。

* 1. real environment

由于对数比值比评分法用于滤除地图中不稳定的体素，因此重建的结果不受移动人员的影响。此外，二维代价地图由八叉树地图生成，可用于导航。

1. **Conclusion**

通过剔除动态对象来提升在动态环境中的精度和鲁棒性。

实时性符合要求但没有明显提升。

稠密的八叉树地图可以用来完成高级任务。

与ROS系统集成，可以进行真实环境的实验。

1. **总结：**

极线几何检测外点

物体含动态点过多就认为是动态，剔除

八叉树地图

缺点：

沿极线方向的外点无法准确检测出来

利用特征点中外点比例判断动态物体，特征点数量受物体纹理影响大

Segnet分割效果有提升空间

* **A Mobile Robot Visual SLAM System With Enhanced Semantics Segmentation**

增强语义分割的移动机器人视觉SLAM系统

Date：2020

**Abstract：**

基于两轮移动机器人，开发了一个更高效的语义SLAM系统，利用语义分割识别每个关键帧中的人、椅子等物体。在对环境有初步了解的情况下，融合RGB - D相机和编码器信息，进行定位并创建无动态物体的稠密彩色八叉树地图。增强的关键帧语义分割极大地提高了系统的效率。此外，融合不同的传感器可以大大提高定位精度。并与DRE - SLAM、DS - SLAM进行比较，

* **MISD-SLAM: Multimodal Semantic SLAM for Dynamic Environments**

Misd-slam：面向动态环境的多模态语义slam

Date：2022

**Abstract：**

多模态语义SLAM系统( MISDSLAM )，去除环境中的动态物体，利用语义信息重建静态背景。MISD - SLAM构建3个主要过程：实例分割、动态像素去除和语义三维地图构建。结合多视图几何约束和K - means聚类算法，去除未定义但运动的像素。同时，重建了具有语义信息的三维稠密点云地图，在不破坏动态物体的前提下恢复了静态背景。

* **RS-SLAM: A Robust Semantic SLAM in Dynamic Environments Based on RGB-D Sensor**

Rs-slam：基于Rgb - D传感器的动态环境下的鲁棒性语义SLAM

Date：2021

**Abstract：**

提出了一种鲁棒的语义RGB - D SLAM系统RS - SLAM。RS - SLAM能够同时检测运动物体和可移动物体，利用基于贝叶斯更新的上下文信息对分割结果进行细化，提供了更准确的感兴趣区域提取。在此基础上，提出了一种运动目标检测方法。引入映射过程中的更新策略，实时构建仅包含静态世界的3D语义OctoMap。

* **SOLO-SLAM: A Parallel Semantic SLAM Algorithm for Dynamic Scenes**

Solo-slam：一种面向动态场景的并行语义slam算法

Date：2022

**Abstract：**

SOLO - SLAM并行完成语义线程和SLAM线程的任务，从而有效提高SLAM系统的实时性。

* **MaskFusion: Real-Time Recognition, Tracking and Reconstruction of Multiple Moving Objects**

掩模融合：多个运动目标的实时识别、跟踪和重建

Date：2018

**Abstract：**

MaskFusion，一个实时的，对象感知，语义和动态的RGB - D SLAM系统，基于实例级语义分割创建对象级语义地图。MaskFusion不需要知道能识别的对象模型，充分展示了实例级语义分割的优势。

* **Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping**

Kimera：一个用于实时度量语义定位和映射的开源库

Date：2020

**Abstract：**

开源的C + +库用于实时的度量语义视觉惯性SLAM，实现了三维场景下的网格重建和语义标注。Kimera采用模块化设计，具有四个关键组件：用于快速准确状态估计的视觉惯性里程计( VIO )模块、用于全局轨迹估计的鲁棒位姿图优化器、用于快速网格重建的轻量级3D Mesher模块和稠密的3D度量语义重建模块。四个模块能单独运行。

* **DynaSLAM: Tracking, Mapping and Inpainting in Dynamic Scenes**

Dynaslam：动态场景中的跟踪、映射和修复

Date：2018

**Abstract：**

DynaSLAM视觉SLAM系统，该系统基于ORBSLAM2 ，增加了动态目标检测和背景修复功能。剔除动态物体：结合多视图几何和深度学习。DynaSLAM在单目、立体和RGB - D配置的动态场景中具有鲁棒性。

先用CNN对先验动态对象进行分割，然后对移动的静态对象进行标注。

Mask RCNN和多视图几何的融合方法主要针对本身是静态物体但发生移动的情况。

1. **Introduction**

本文提出一种实时算法来处理动态物体。通过在ORB-SLAM2系统中增加一个前段阶段来实现。在单目和立体情况下，建议使用卷积神经网络，在RGB-D情况下，提出了结合多视图几何模型和基于深度学习的算法来检测动态物体。

本文主要工作，识别检测动态物体并剔除，重建地图中被动态物体遮挡的部分。

在单目和双目中，利用CNN进行像素级语义分割，在SLAM中不再对他们进行特征提取；在RGBD中，结合多视图几何算法和深度学习识别动态物体。

1. **Related work**

常用的异常值剔除算法：RANSAC和鲁棒代价函数。

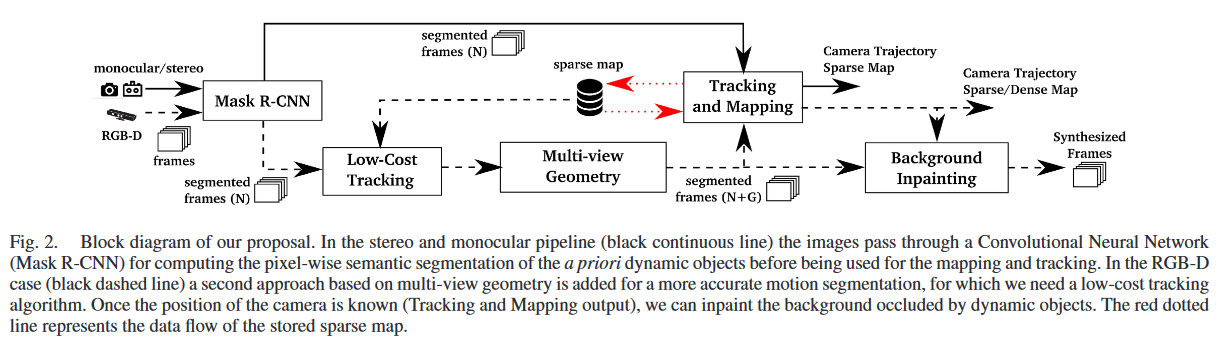
其他学者提出的方法存在的问题：1. 无法检测到动态物体保持静止；2. 可以检测先验动态物体，但无法检测静态物体发生移动。

本文将针对以上两个问题剔除解决办法。

1. **System description**

在图2中，单目和双目直接经过Mask R-CNN进行像素级语义分割，RGBD相机图像需要经过多视图几何算法分为若干图像在进行分割。然后将大部分时间都是静止的动态对象定位新动态对象。分割后的帧用于估计相机位姿和构建场景地图。

用先前帧的静态信息重建当前帧中被动态物体挡住的背景。动态物体的特征点不会被跟踪也不被映射。



* 1. segmentation of potentially dynamic content using a CNN

Mask R-CNN可以同时提供像素级语义分割和实例标签。

Mask R-CNN的输入是RGB原始图像。

输出是包含所有动态物体分割的一幅图。

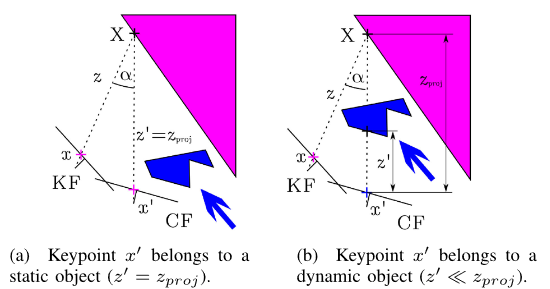
* 1. Low-Cost Tracking

低成本的跟踪线程：将动态物体分割后，利用静态物体特征跟踪相机位姿。

* 1. Segmentation of dynamic content using Mask R-CNN and Multi-View Geometry

Mask R-CNN对于动态物体可以检测出来，静止的人或者车等，他是通过学习将这一类的物体进行语义分割，但是对于发生移动的静态物体无法检测，例如运动的人拿着一本书。

重叠帧？？



在判断静态物体是否运动时，通过估计两个关键帧之间投影的夹角，大于30度就认为被遮挡，可以忽略。

当深度差大于某个阈值的时候则可以认为是动态物体。

一些动态关键点会出现在边界上，可能会造成一些问题，？？对所有动态关键点进行分类，多视图一致性解决此问题。

当深度学习和多视图几何同时检测到目标，则掩膜区域是几何方法的结果。

掩膜区域也会包含只有一种方法检测出来的区域。

* 1. Tracking and mapping

Input：RGB原图和掩膜结果

落在边缘高梯度区域的关键点剔除

* 1. Background inpainting

有些没有足够信息的场景不会被重建

在动态场景中合成的地图在静态SLAM中也能用。

1. **Experimental results**
   1. TUM dataset

包含低运动（坐着手部运动）和高运动（行走）

相机四种运动模式：1.半球；2.x-y-z轴；3.横摇，俯仰，偏航；4.静止。

ORB特征提取在真实区域和重建区域都能工作，都能与先前的关键帧中的特征点进行匹配。

对于深度学习的改进：动态对象的分割和动态片段的细化

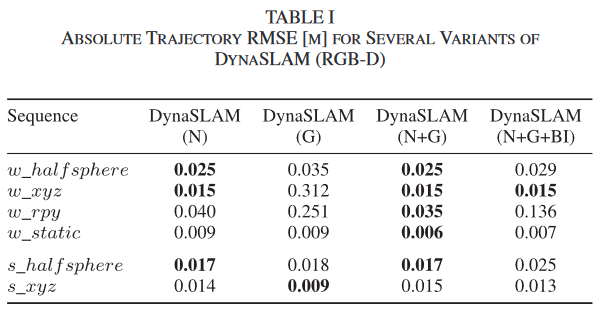
多视图几何的误差来源：动态物体的分割需要一段时间才精确，这段时间会产生误差。

相机定位前加入背景修复会导致跟踪精度降低：背景重建与相机位姿强相关。

背景修复应放在跟踪之后。

在高动态环境中，Dyna的精度要更高

在低运动环境中，精度方面没有明显优势。



DynaSLAM（N+G+BI）中由于使用了重建之后的影像进行跟踪，二相机姿态会对重建影像有较大影响，因此这种效果比较差。

* 1. KITTI dataset

交通环境

在具有先验动态性的物体上提取特征，跟踪精度比较高，当车辆处于停止状态时，误差会大一些，因为将这些物体目标去掉后检测到的特征点会处于较远或者弱纹理区域，从而导致跟踪精度下降。

但是在回环检测和重定位中结果较为稳健，因为构建的地图中仅仅包含结构性物体对象，没有动态对象。

待解决问题：在动态物体没有运动时能够被用来提取特征点用于跟踪，而该对象不会在最终的地图中出现。

* 1. Timing analysis

增加多视图几何模块会降低实时性

没有进行工程优化，运行时间比较久

1. **Conclusion**

Dyna在大多数情况下精度最高，可以在长期应用中使用。

在K数据集中精度较低

后期可改进的地方：实时性，更真实的修复技术。

1. **总结**

多视图几何判断外点

多视图掩膜区域和语义分割区域都去掉

背景修复

缺点：

只要物体上存在一个动点，就将该物体判断为动态，然后剔除；

* **DynaSLAM II: Tightly-Coupled Multi-Object Tracking and SLAM**

Dynaslam II：紧耦合多目标跟踪与SLAM

Date：2021

**Abstract：**

DynaSLAM II针对立体和RGB - D相机配置的视觉SLAM系统，紧密集成了多目标跟踪能力。DynaSLAM II利用实例语义分割和ORB特征来跟踪动态对象。

* **A survey: which features are required for dynamic visual simultaneous localization and mapping?**

综述：动态视觉同时定位与建图需要哪些特征?

Date：2021

**Abstract：**

动态SLAM将特征分为动态和静态两类。动态SLAM为智能机器人在复杂环境下的作业提供更加鲁棒的定位。动态SLAM可以跟多目标跟踪相结合。

本文综述讨论不同层次特征的优缺点。

本文对基于点特征和基于点线特征的SLAM进行研究

1. **Introduction**

**相机类型：**传统相机（单目、立体、RGB-D）、事件相机

单目：无法直接估计全局尺度，必须增加额外的传感器或先验来克服尺度模糊。

双目：根据左右相机之间基线的长度直接估计深度和全局尺度。

RGB-D：可以直接获取深度信息和全局尺度。但对光照极为敏感。

事件相机：独立像素的亮度随时间变化。记录更连续。更低的延时、更好的动态范围，适合动态场景。

**特征匹配：**

低级特征：像素块、点、线等纹理、几何信息

高级特征：信息融入语义标签

**优化：**

Filter-based：使用概率分布获得信息，基于滤波的方法对处理统计信息方面有优势，深度估计，传感器融合

Frame-based：选取少量过去帧使用光束平差法，对视觉SLAM更有效。

**动态场景鲁棒性问题解决方法：**

检测+剔除：将动态特征剔除，利用静态特征构建地图等

检测+跟踪：利用动态特征来估计相机位姿

**Existing surveys on dynamic SLAM or its components**

**Article organization**

**鲁棒SLAM：**

动态特征鲁棒的判断

处理遮挡

保证地图长期一致性

剔除动态对象导致特征变少

**多目标跟踪SLAM：**

缺失对遮挡物的数据处理

单目系统相对尺度问题

存在噪声的测量的概率数据关联

1. **Low-level-feature-based dynamic SLAM**

**2.1 概述：**

**低级特征：**

点特征：可以更好的表现纹理；点特征可以获得更好的鲁棒性；

线特征：更好的表现几何结构，适用于人造环境；完全基于线特征的SLAM系统不如基于点特征的SLAM系统；

**基于点特征的匹配方法：**

基于像素：对光照敏感；限制其在长期数据关联上的应用；

基于描述子：对光照更加鲁棒；通常利用运动信息和特征的三维位姿进行辅助匹配；没有传感器的情况下很难预测目标运动；不能再高动态环境下获得足够信息关联，可以利用光流法解决；

数据驱动：利用深度学习获得特征提取器；易于结合检测和数据驱动，使描述子方法摆脱了运动估计的负担。

**基于线特征的匹配方法：**

方法：线段检测；二进制描述子；深度学习

比基于点特征的更耗时，动态目标跟踪效果比较差

**2.2 鲁棒SLAM**

区分动态和静态特征，即运动分割

**运动分割：**

1. 光流法：特征在二维投影图像空间中投影的运动学；

点特征：

线特征：

1. 几何法：
2. 基于运动的方法：

**讨论：**

1. 方法总结：

光流法不需要额外先验信息，但是对光线敏感；

几何方法不需要先验信息，但是单独使用几何特征会导致运动退化；

基于运动的方法可以轻易解决运动退化的问题；

推荐多个方法结合。

1. 存在问题：

需要通过历史连续帧检测动态，过多计算量大，过少不鲁棒，构建概率模型检测动态特征；

遮挡问题，剔除遮挡信息会导致信息太少，保存遮挡信息来提高鲁棒性；

把动态特征剔除后，特征太少，缺少数据关联，解决方案：多目标跟踪SLAM

* **VDO-SLAM: A Visual Dynamic Object-aware SLAM System**

Vdo-slam：一个可视化的动态对象感知SLAM系统

Date：2021

**Abstract：**

利用语义信息，在没有物体形状和几何模型先验知识的情况下，实现对动态刚性物体的准确运动估计和跟踪。

光流+语义分割

基于双目/rgb-d

对动态目标的位姿变化进行精确估计并估算出场景中目标移动速度

能够处理由于语义分割失败而导致的间接性遮挡问题

**第一个完整的能够实现运动分割，动态目标追踪，估算跟随着静态和动态结构相机的位姿，估算场景中的每一个刚体的完全SE(3)位姿变化，提取速度信息和可以在现实户外场景中被演示证明的动态SLAM系统。**

* **Co-Fusion: Real-time Segmentation, Tracking and Fusion of Multiple Objects**

协同融合：多目标的实时分割、跟踪和融合

Date：2017

**Abstract：**

利用多模态拟合方法，每个对象都可以独立地从背景中移动，可以有效的被跟踪，保持了每个分割对象的3D模型，并通过融合来改善。

可以在对象级别上保持场景描述。

* **ReFusion: 3D Reconstruction in Dynamic Environments for RGB-D Cameras Exploiting Residuals**

融合：Rgb - D相机利用残差进行动态环境下的三维重建

Date：2019

**Abstract：**

没有利用神经网络的动态语义SLAM

RGB-D传感器，纯几何方法分割动态区域，具体的：在KinectFusion稠密slam系统的基础上，计算每个像素点的残差，通过自适应阈值分割得到大致动态区域，形态学处理得到最终动态区域，与此同时，可得到静态背景的TSDF地图。

* **SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation**

Segnet：一种用于图像分割的深度卷积编码器-解码器架构

Date：2017

**Abstract：**

语义像素级分割的深度全卷积神经网络，编码解码结构

内存和精度的平衡

1. **Architecture**

Segnet由一个编码器网络、一个解码器网络和一个逐像素分类层组成。编码层和解码层是对称结构，每个编码层都有对应的解码层，

编码器网络由VGG16的前13层卷积层组成，

池化层采用最大池化，不但保存了最大权值还保存了最大权值的相对位置，在反池化操作中仍将最大权值放到原有位置，其他位置用0填充。

Soft-max层逐像素进行分类，将每个像素 分到概率最大的那个类。

* **An improved SegNet network model for accurate detection and segmentation of car body welding slags**

改进的SegNet网络模型用于车体焊渣的精确检测与分割

Date：2021

**Abstract：**

在segnet网络中加入上下文提取器，增强感知全局和局部信息的能力。

利用残差层和BN层优化segnet网络，利用dropout层避免过拟合。

* **Brain Tumor Segmentation from MRI Images using Hybrid Convolutional Neural Networks**

使用混合卷积神经网络从MRI图像中分割脑肿瘤

Date：2020

**Abstract：**

Segnet和U-net和resnet18相结合创造新的网络模型。主要用到编码解码结构和跳跃连接。

* **Multi-scale feature fusion network for pixel-level pavement distress detection**

面向像素级路面破损检测的多尺度特征融合网络

Date：2022

**Abstract：**

在编码解码结构中串联了带有故障分类特征的故障位置信息。训练时间稍长。

* **Robust optimization of SegNet hyperparameters for skin lesion segmentation**

面向皮损分割的SegNet超参数鲁棒优化

Date：2021

**Abstract：**

贝叶斯优化的segnet网络

* **Mask R-CNN**

Date：2018

**Abstract：**

* **Optimized segmentation with image inpainting for semantic mapping in dynamic scenes**

结合图像修复的优化分割用于动态场景下的语义映射

Date：2023

**Abstract：**

为了解决静态物体被动态物体遮挡而出现建图缺失或者错误的情况，提出了一种图像修复技术，，在原有的语义分割网络上串联了一个修复网络和一个相同的语义分割网络，从而得到完整的语义分割结果。最终与SLAM相结合够建立更加完整稳定的语义八叉树地图。

* **Fully Convolutional Networks for Automatically Generating Image Masks to Train Mask R-CNN**

自动生成图像掩膜的全卷积网络训练Mask R – CNN

Date：2023

**Abstract：**

基于Mask R-CNN的自动生成图像掩模的方法

* **A Dense Semantic Mapping System based on CRF-RNN Network**

基于CRF - RNN网络的稠密语义映射系统

Date：2017

**Abstract：**

带有语义信息标注的三维可视化地图，CRFRNN进行语义分割，使用RGB-D数据作为系统的输入，用自己生成的RGB-D数据集测试语义映射系统，能够可靠的标注三维点云地图中的语义信息。

* **DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs**

Deeplab：深度卷积网络、空洞卷积和全连接Crfs的语义图像分割

Date：2017

**Abstract：**

将深度学习用于语义图像分割，