* **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

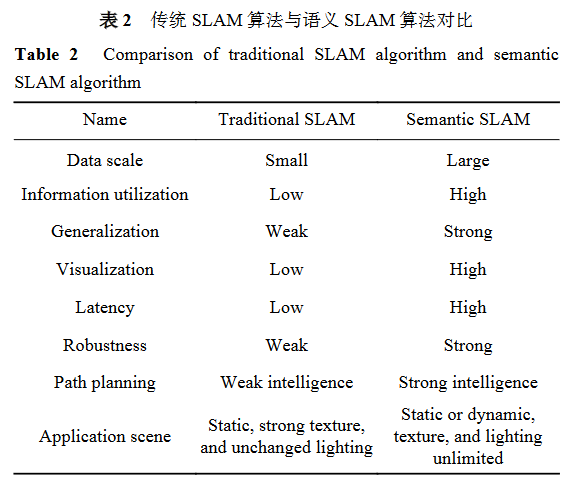
构建地图时提取的是环境中的几何特征信息，常见的有SIFT和SURF特征。几何特征不能满足导航、避障、交互。

**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。

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。

* **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：**

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**