动态综述

DOI：10.16818/j.issn1001－5868.2023112202

动态环境下基于深度学习的视觉SLAM研究综述

罗　元，沈吉祥*，李方宇
（重庆邮电大学 光电工程学院，重庆 400065）

摘　要： 目前的同时定位与地图构建(Simultaneous Localization and Mapping，SLAM)研究大多是基于静态场景的假设，而实际生活中动态物体是不可避免的。在视觉SLAM系统中加入深度学习，可以协同剔除场景中的动态物体，有效提升视觉SLAM在动态环境下的鲁棒性。文章首先介绍了动态环境下基于深度学习的视觉SLAM分类，然后详细介绍了基于目标检测、基于语义分割和基于实例分割的视觉SLAM，并对它们进行了分析比较。最后，结合近年来视觉SLAM的发展趋势，通过对动态环境下基于深度学习的视觉SLAM存在的主要问题进行分析，总结了未来可能的发展方向。

关键词： 视觉同时定位与地图构建；深度学习；动态环境；语义信息
中图分类号： TP242.6　**文章编号：** 1001－5868(2024)01－0001－10

Review of Visual SLAM Research Based on Deep Learning in Dynamic Environments

LUO Yuan，SHEN Jixiang，LI Fangyu
（College of Optoelectronic Engineering，Chongqing University of Posts and Telecommunications，Chongqing 400065，CHN）

Abstract： The current research on simultaneous localization and mapping (SLAM) in academia mostly assumes static scenes，but dynamic objects are inevitable in real-life scenarios. Integrating deep learning into visual SLAM systems can collaboratively eliminate dynamic objects from the scene，effectively enhancing the robustness of visual SLAM in dynamic environments. This paper first introduces a classification of deep learning-based visual SLAM in dynamic environments and then provides a detailed overview of visual SLAM systems based on object detection，semantic segmentation，and instance segmentation. A comparative analysis of these approaches is also presented. Finally，considering the recent trends in the development of visual SLAM，the paper analyzes the main challenges of deep learning-based visual SLAM in dynamic environments and summarizes potential future directions.

Key words： visual SLAM；deep learning；dynamic environment；semantic information

0　引言

同时定位与地图构建(Simultaneous Localization and Mapping，SLAM)是指移动机器人在没有任何先验信息的情况下同时完成自身的定位与周围环境地图的构建[1-3]。根据所用传感器的不同，可以把SLAM分为激光SLAM和视觉SLAM两大类。激光SLAM发展较早，其激光传感器计算量较小，测距准确，不易受光照影响，但是在纹理信息丰富的区域不能很好地提取特征信息[4]。视觉SLAM是把相机作为传感器来获取外部环境信息从而进行定位与建图。按照视觉SLAM工作方式的不同，可以把相机分为深度相机(RGB-D)、单目相机(Monocular)、双目相机(Stereo)三大类[5-6]。

收稿日期：2023－11－22.
基金项目：国家自然科学基金青年基金项目(52205183).
* **通信作者：**沈吉祥

虽然视觉 SLAM 相比激光 SLAM 精度较差，但是因为视觉 SLAM 可以获取更多外部环境信息而得到了广泛的应用。

视觉 SLAM 技术经过长期发展，已经出现了许多优秀成果，但仍存在一些问题。现有的视觉 SLAM 算法很多都将外部环境假设为静态场景，忽视了动态物体对视觉 SLAM 系统的影响[7-8]。比如移动机器人在复杂的动态环境中工作时，高动态目标会使相邻的两帧图像之间产生巨大的不一致，从而严重影响视觉 SLAM 系统的准确性和鲁棒性。另外，构建的地图也会由于每帧图像中动态目标的位置不同而产生大量模糊区域，导致后期无法供导航使用。基于静态场景的假设极大地限制了视觉 SLAM 的发展以及视觉 SLAM 在实际生活中的应用。在视觉 SLAM 系统中加入深度学习网络，可以获取环境的语义信息。利用语义信息来剔除场景中的动态物体，生成具有语义信息的语义地图，可以丰富移动载体对环境的理解，获得高层次的感知，提升动态环境下视觉 SLAM 系统的准确性和鲁棒性。

本文对近年来动态环境下基于深度学习的视觉 SLAM 进行了梳理和分类，重点介绍了基于目标检测、语义分割和实例分割的视觉 SLAM，最后对动态环境下基于深度学习的视觉 SLAM 进行展望和总结。

1 动态环境下基于深度学习的视觉 SLAM 分类

近年来随着深度学习的飞速发展，出现了许多优秀的深度学习网络，如 YOLO[9]，SegNet[10]，Mask R-CNN[11]，PSPNet[12]，ICNet[13] 等。为弥补传统方法的不足，研究者们尝试将视觉 SLAM 与深度学习相结合，把获得的语义信息作为判断动态目标的先验信息，来解决视觉 SLAM 在动态环境中面临的问题。目前，研究者们已经提出了大量优秀的语义视觉 SLAM 系统，如 DS-SLAM[14]，DynaSLAM[15]，PSPNet-SLAM[16]，SG-SLAM[17]，GSL-VO[18] 等。这些视觉 SLAM 系统的精度相比于经典的 ORB-SLAM2[19]，ORB-SLAM3[20] 在动态环境下有大幅度的提高，有效提升了传统 SLAM 在动态环境下的鲁棒性。

深度学习的加入虽然提高了视觉 SLAM 系统在动态环境中的定位精度，但是还存在需要改进的地方。首先基于深度学习的视觉 SLAM 系统严重依赖于硬件的性能，实时性不能得到保证，并且所采用的深度学习网络精度是有限的，在某些复杂的场景中，不能精确地分割目标。其次在剔除动态物体时，会将动态物体所在的区域整个剔除，造成动态区域中静态特征点的剔除，但这些静态点对后续相机的位姿估计和定位是有帮助的。因此，怎样构建动态环境下高精度的实时视觉 SLAM 系统成为了当下的研究热点[21]。

目前根据深度学习网络的不同，动态环境下基于深度学习的视觉 SLAM 主要分为三类：基于目标检测的视觉 SLAM、基于语义分割的视觉 SLAM 以及基于实例分割的视觉 SLAM。这三种视觉 SLAM 的提出，极大地推动了动态环境下视觉 SLAM 的发展。

2 基于目标检测的视觉 SLAM

如何从图像中提取语义信息是现代视觉 SLAM 中的一个热点研究问题，其本质是从场景中提取物体特征信息[22]。虽然卷积神经网络(Convolutional Neural Networks，CNN)等也有助于语义信息的提取，但现代视觉 SLAM 更依赖于目标检测等语义提取模块[23]。目标检测可以获取图像的目标信息和空间信息，并且可以通过绘制物体的候选框来识别每个物体的类别[24]。

2.1 目标检测算法

基于深度学习的目标检测算法根据实现步骤可以分为："双阶段"目标检测算法和"单阶段"目标检测算法。早期的目标检测采用双阶段架构，在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。常见的算法有：R-CNN[25]，Fast R-CNN[26]，Faster R-CNN[27] 等。随后 YOLO 创造性地提出了单阶段架构，它直接一步完成物体的分类和定位，直接输出回归得到的候选框及其类别。单阶段算法减少了目标检测算法的步骤，不需要创建候选框，直接将目标框定位问题转化为回归问题，在速度上具有优越性。常见的算法有：YOLO 系列和 SSD[28]。近年来，基于目标检测的视觉 SLAM 系统也大多采用"单阶段"算法，如 Detect-SLAM[29]，AHY-SLAM[30]，YOLO-SLAM[31] 以及 Wang 等[32] 提出的面向动态场景的高效语义 SLAM 系统等。

2.2　典型的基于目标检测的视觉 SLAM 系统

2.2.1　Detect-SLAM

Detect-SLAM 是一种新颖的机器人视觉系统，它将 SLAM 与基于深度神经网络的目标检测器集成在一起，使这两种功能相互受益，有助于机器人在未知动态环境中可靠高效地完成任务。Detect-SLAM 首次将 SLAM 和基于深度神经网络（Deep Neural Networks，DNN）的检测器结合起来，并同时完成三个任务：提高 SLAM 在动态环境中的鲁棒性、提高目标检测性能、构建语义地图。它的框架如图 1 所示。

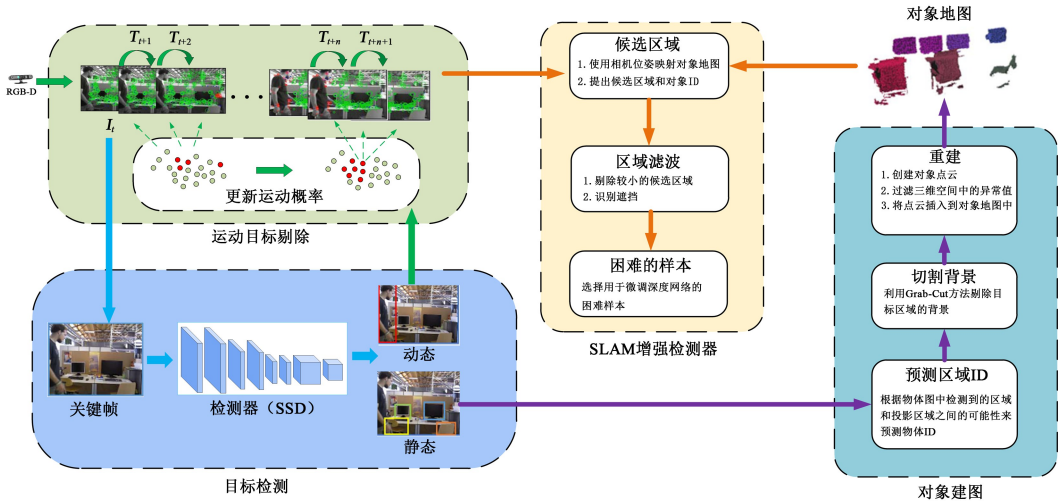


图 1　Detect-SLAM 的框架

在 Detect-SLAM 中，研究者们将一个基于 DNN 的目标检测器融入到 ORB-SLAM2 系统中。ORB-SLAM2 主要有三个并行线程：跟踪线程、局部建图线程、闭环检测线程。与它相比，Detect-SLAM 加入了 3 个新的线程：运动目标剔除、对象建图、SLAM 增强检测器。运动目标剔除模块可以过滤掉与移动对象相关联的特征，对象建图模块可以重建关键帧中检测到的静态对象，生成具有密集点云的对象地图。SLAM 增强检测器则可以利用目标地图作为先验信息，提高在具有挑战性环境中的检测性能。

为了评估 Detect-SLAM 系统在高动态环境下的鲁棒性，研究者们将该系统与在静态环境和低动态环境下产生良好结果的 ORB-SLAM2 系统进行了对比。在 TUM RGB-D 数据集的 7 个高动态序列、2 个低动态序列和 1 个静态序列上分别进行了实验。实验结果表明，在高动态场景下 Detect-SLAM 系统估计的轨迹更接近于真实轨迹，但在静态场景下与 ORB-SLAM2 相比，无法得到理想的结果。这是因为在静态场景中，Detect-SLAM 系统过滤掉了一些对相机位姿估计和后续建图有益的静态信息，从而影响了整个系统的定位精度。

2.2.2　AHY-SLAM

AHY-SLAM 是一种基于自适应阈值均匀化特征提取和 YOLOv5 目标检测的动态场景视觉 SLAM 算法。该算法在 ORB-SLAM2 的基础上增加了关键帧选择模块、阈值计算模块和目标检测模块，并且在跟踪线程中增加了均匀化 ORB 特征点提取和动态特征点剔除步骤。AHY-SLAM 利用光流法减少了关键帧选择中描述符的计算量，提取了更适合于动态场景的均匀特征点，具有处理高动态场景的能力。AHY-SLAM 的系统架构如图 2 所示，蓝色部分是 AHY-SLAM 相对于 ORB-SLAM2 的创新之处。

图像帧进入 AHY-SLAM 系统后，需要先经过关键帧选择和目标检测两个模块的处理。在关键帧选择模块中，AHY-SLAM 提取图像的 FAST 角点，利用 LK 光流进行跟踪，并根据选择策略判定是否为关键帧。若不是关键帧，计算相机位姿并继续判断下一帧。若是关键帧，则输出到阈值计算模块，将图像分成几个小区域，根据每个区域的灰度值分别计算阈值。所有阈值都导出到跟踪线程中的均匀化 ORB 特征点提取步骤。在目标检测模块中，AHY-SLAM 使用 YOLOv5 对目标进行检测，并在被检

测对象周围绘制检测框。对具有不同动态水平的对象，AHY-SLAM 对其进行分类，最后将检测的结果输出到跟踪线程中进行动态特征点剔除。

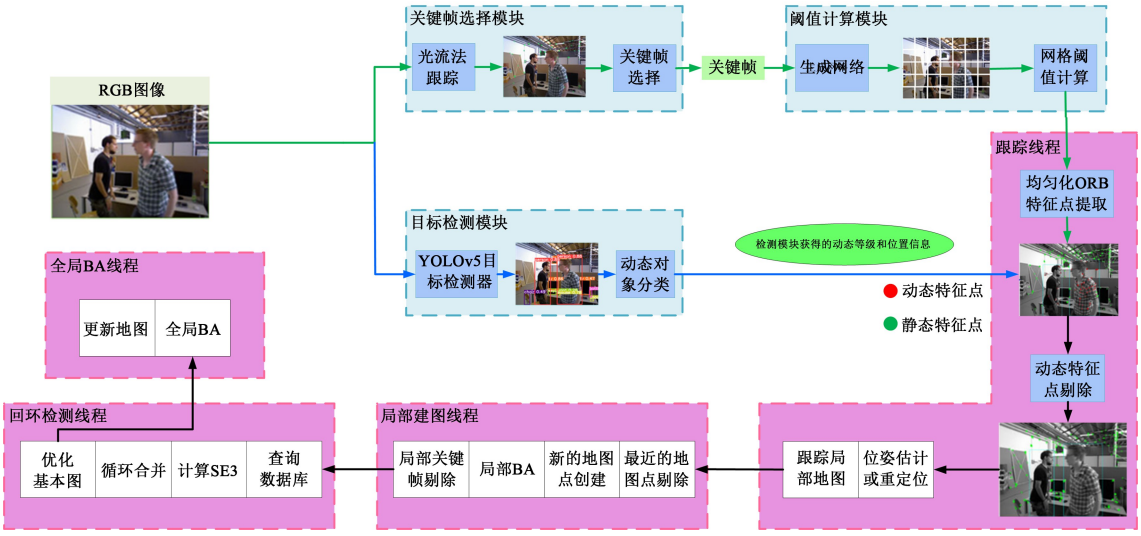


图 2 AHY-SLAM 的系统架构

与 ORB-SLAM2 相比，AHY-SLAM 显著提高了 TUM 开放数据集中多个动态场景序列的姿态估计精度，绝对姿态估计精度最高可提高 97%。为了进一步验证算法的可行性，研究者们还将 AHY-SLAM 与其他动态环境下优秀的视觉 SLAM 进行了比较，如：Detect-SLAM，DS-SLAM，DynaSLAM，RDS-SLAM[33]。实验结果表明，在动态目标运动幅度较大的序列中，AHY-SLAM 的效果要优于 DS-SLAM 和 Detect-SLAM，但在运动目标移动较少的情况下，DS-SLAM 的效果更好。这是因为 AHY-SLAM 没有考虑到动态对象上也可能存在静态特征点，而是直接消除所有可能的特征点，导致剩余的静态点较少，影响了跟踪定位的精度。在实时性方面，AHY-SLAM 比其他动态场景的 SLAM 方法表现更好，但由于增加了目标检测过程，相比于 ORB-SLAM2 耗时更多。

2.2.3 小结

在动态环境下基于目标检测的视觉 SLAM 相比于使用其他深度学习方法的视觉 SLAM 在速度方面具有优越性。近年来，研究者们也提出了许多优秀的基于目标检测的视觉 SLAM 系统，表 1 中列举了一些代表。

虽然基于目标检测的视觉 SLAM 在实时性方面表现良好，但在精度方面还需要进一步加强，例如：在剔除动态目标时往往会剔除整个区域，导致静态特征点的浪费，影响整个系统的定位精度。因此，在保证实时性的同时，又能保证动态环境下视觉 SLAM 的高精度，是未来重要的发展方向。

表 1 基于目标检测的视觉 SLAM 系统

系统	年份	传感器	场景	动态检测
Detect-SLAM	2018	RGB-D	室内	SSD
Wang Zemin 等提出的系统	2019	RGB-D	室内	YOLOv3
YOLO-SLAM	2022	RGB-D	室内	Darknet19-YOLOv3
YPD-SLAM[34]	2022	RGB-D	室内	YOLOFastestv2
AHY-SLAM	2023	RGB-D	室内	YOLOv5

3 基于语义分割的视觉 SLAM

3.1 语义分割

图像的语义分割是在像素级别上理解图像，从而获得图像中的深层次信息，包括空间、类别和边缘等。在深度学习应用到计算机视觉领域之前，纹理基元森林和随机森林等语义分割方法用得比较多。基于深度神经网络的语义分割技术突破了传统语义分割的瓶颈[35]，极大地推动了语义分割的发展。与目标检测相比，基于深度神经网络的语义分割方法具有更高的准确性。

目前语义分割技术已应用到视觉 SLAM 当中，并涌现了一大批优秀的视觉 SLAM 系统，如 DS-SLAM，STDyn-SLAM[36]，KSF-SLAM[37] 和 PSPNet-SLAM 等。应用语义分割技术可以有效提高视觉 SLAM 系统在动态环境下的精度，实现移动机器人在复杂动态环境下的准确位姿估计。

3.2　典型的基于语义分割的视觉 SLAM 系统

3.2.1　DS-SLAM

DS-SLAM 是一种面向动态环境的鲁棒语义视觉 SLAM，它是高动态环境下最先进的 SLAM 系统之一。DS-SLAM 将语义分割网络 SegNet 与移动一致性检测方法相结合，减少了动态物体的影响，同时生成稠密的语义八叉树地图，可用于完成高层次任务。它的系统框图如图 3 所示。

DS-SLAM 中含有五个并行的线程：跟踪线程、语义分割线程、局部建图线程、闭环检测线程和稠密语义地图构建线程。原始的 RGB 图像先同时在跟踪线程和语义分割线程中进行处理。在跟踪线程中，先提取 ORB 特征点，然后粗略地检查特征点的移动一致性，并保存潜在的异常点。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离群点，将位于运动物体中的 ORB 离群特征点剔除。然后通过匹配剩余的稳定特征点计算变换矩阵，从而提升系统在动态场景下的鲁棒性和准确性。

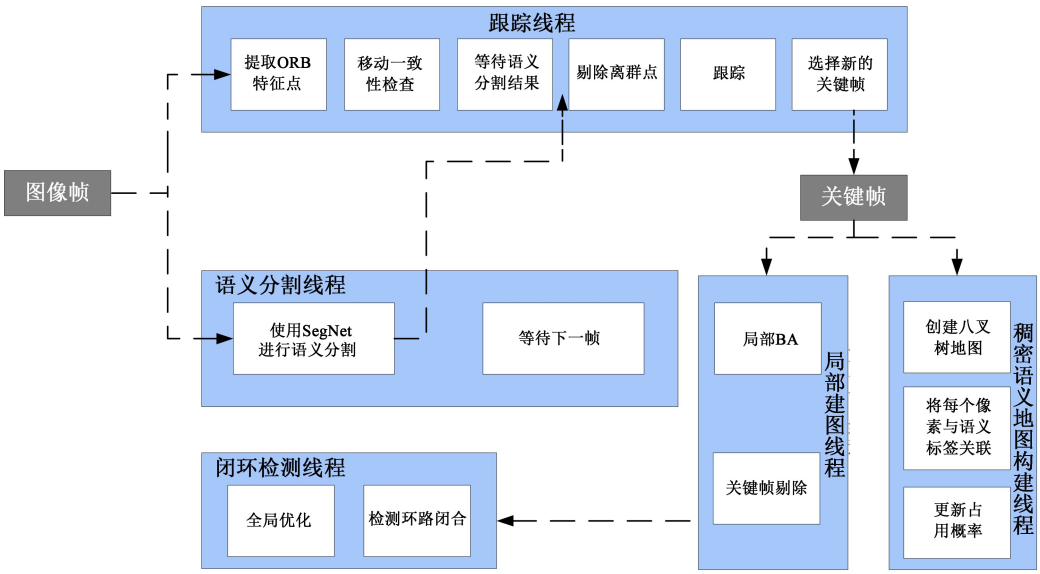


图 3　DS-SLAM 的系统框图

研究者们在 TUM RGB-D 数据集和真实环境中进行了实验。实验结果表明，与 ORB-SLAM2 相比，DS-SLAM 的绝对轨迹精度可以提高一个数量级，并且系统的鲁棒性明显优于 ORB-SLAM2。但是，DS-SLAM 还有需要加强的地方，例如：语义分割网络中可识别的对象类型受到限制，导致其应用范围有限。另外，当检测到回环时，八叉树地图也需要重新构建，导致系统的耗时增加。

3.2.2　PSPNet-SLAM

PSPNet-SLAM 是一种通过金字塔场景解析网络检测动态对象的语义 SLAM。该系统基于 ORB-SLAM2 构建，并引入 PSPNet 作为并行语义线程。在语义线程中，研究者们使用 PSPNet 进行逐像素的语义分割。由于其金字塔的网络结构，可以获得更多的上下文信息以及它们之间的关系，故在动态特征点的检测和剔除上比其他结构的全卷积神经网络更有效。在几何线程中，研究者们首先提出最优误差补偿单应矩阵来补偿前一帧特征点的投影变换误差引起的特征点偏移。其次，提出了一种反向蚁群策略，利用动态点社区分布的特性来选择检测到的特征点，从而提升几何线程动态特征点检测的鲁棒性和实时性。PSPNet-SLAM 的系统架构如图 4 所示。

为了验证系统的性能，研究者们将 PSPNet-SLAM 在 TUM 数据集上与其他优秀的 SLAM 系统进行了对比实验。实验结果表明，与 ORB-SLAM2 相比，该系统实现了许多改进，包括在动态环境下的定位精度和鲁棒性。与性能优异的 DynaSLAM 相比，其几何线程的实时性也得到了保证，但整个系统的实时性，仍然是需要面对的问题。

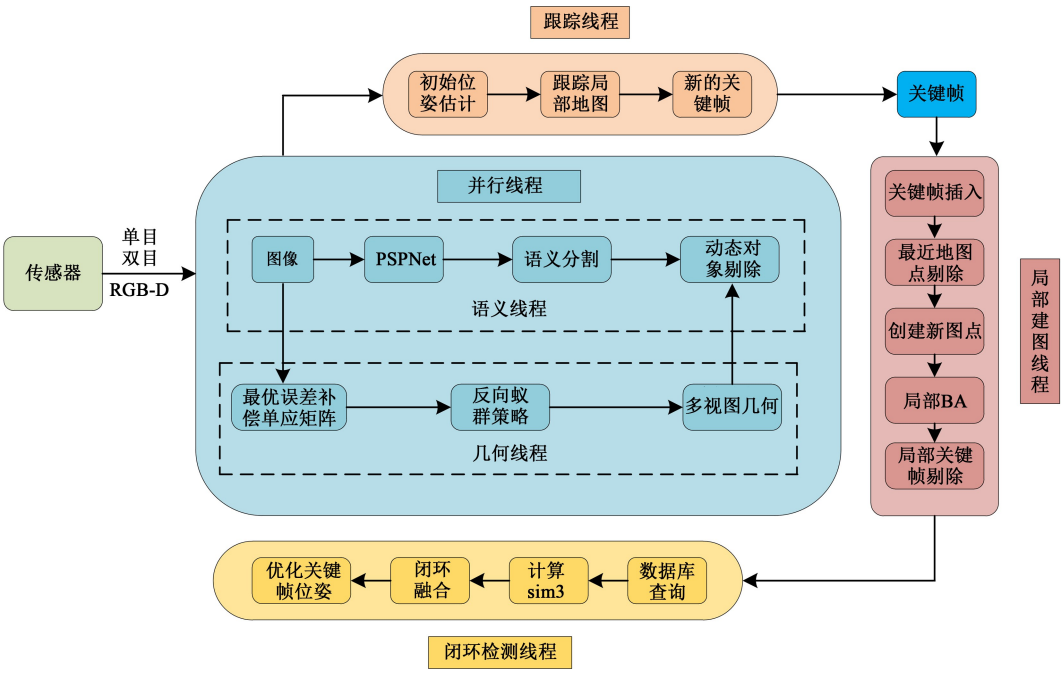


图 4 PSPNet-SLAM 的系统架构

3.2.3 小结

在视觉 SLAM 中加入语义分割技术，可以有效提高动态环境下系统的定位精度。目前研究者们已经提出了许多基于语义分割的视觉 SLAM 系统，表 2 中列举了一些性能表现良好的代表。

相比于基于目标检测的视觉 SLAM，基于语义分割的视觉 SLAM 在精度方面更加准确，但在实时性方面，仍有待提高。视觉 SLAM 是需要应用到实际生活中的，实时性也是不可忽视的参数之一。因此，怎样进一步减少基于语义分割的视觉 SLAM 系统的耗时成为了研究者们新的研究目标。

表 2 基于语义分割的视觉 SLAM 系统

系统	年份	传感器	场景	动态检测
Li Xuanpeng 等提出的系统[38]	2016	RGB-D	室内/室外	Deeplabv2
DS-SLAM	2018	RGB-D	室内	SegNet
Han Shuangquan 等提出的系统[39]	2020	RGB-D	室内	PSPNet
PSPNet-SLAM	2020	RGB-D	室内	PSPNet
STDyn-SLAM	2022	Stereo	室外	SegNet

4 基于实例分割的视觉 SLAM

4.1 实例分割

近年来，计算机性能的不断提高促进了计算机视觉中实例分割的快速发展。实例分割不仅具有像素级的分类，还具有不同对象的位置信息，甚至可以检测到相同的对象。举例来说，如果一张图片中有多个人，对于语义分割来说，可以将所有人的像素都归为一类。但是实例分割可以将不同人的像素归为不同的类，也就是说实例分割比语义分割的分割精度更高。

2017 年，He 等提出的 Mask R-CNN 是实例分割的开创性工作。它的提出为提升动态环境下视觉 SLAM 的鲁棒性提供了新的思路，随后研究者们基于实例分割提出了大量优秀的视觉 SLAM 系统，如：DynaSLAM，SLAMANTIC[40]，YO-SLAM[41]，DRSO-SLAM[42]，RDMO-SLAM[43] 等。实例分割既可以在像素层面上进行分类，又可以定位出不同实例，可以实现在复杂动态环境下的精细化分割，极大地推动了动态环境下基于深度学习的视觉 SLAM 的发展。

4.2 典型的基于实例分割的视觉 SLAM 系统

4.2.1 DynaSLAM

DynaSLAM 在 ORB-SLAM[44] 的基础上增加了运动分割方法，使得其在动态环境下使用单目、双目和 RGB-D 相机进行位姿估计和定位时都具有鲁棒性。该系统准确地跟踪了相机，并创建了一个静态的、可重用的场景地图。图 5 展示了该系统的架构。黑色实线为单目和双目通道，黑色虚线为 RGB-D 通道，红色虚线表示存储的稀疏地图数据流。在 RGB-D 情况下，研究者们使用多视图几何从两个方面改进动态内容分割。首先，对之前通过

Mask R-CNN 得到的动态对象进行细化分割，得到精度更高的分割结果，避免造成静态特征点的浪费。其次，对在 Mask R-CNN 阶段未被检测到的动态对象进行实例标记，进一步避免动态特征点的残留。在单目和双目情况下，直接通过 Mask R-CNN 对图像进行分割，使得属于先验动态物体的关键点既不被跟踪也不被映射。

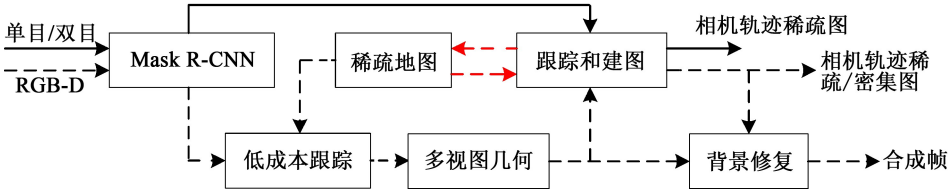


图 5　DynaSLAM 的系统架构

研究者们将 DynaSLAM 与 ORB-SLAM，ORB-SLAM2 以及动态环境下其他优秀的视觉 SLAM 系统在 TUM 数据集和 KITTI 数据集上进行了对比实验。在 TUM 数据集中，DynaSLAM 在单目情况下的精度与 ORB-SLAM 相似，但在较早初始化的情况下获得了场景的静态地图。在 KITTI 数据集中，除了动态物体代表场景重要部分的情况，DynaSLAM 的精度都略低于单目和双目 ORB-SLAM，但该系统估计的地图只包含结构对象，因此可以在长期应用中重用。未来研究者们还将对 DynaSLAM 进行扩展，除了加强实时性外，还将使用更精细的背景修复技术，进一步提高系统的精度。

4.2.2　DRSO-SLAM

DRSO-SLAM 是一种基于语义信息和光流的动态 RGB-D SLAM 算法，该算法利用基于语义信息的像素级分割网络进行系统的粗略自运动估计，并结合光流和语义信息提取动态场景中的运动物体，消除系统中的动态特征点，然后仅利用静态特征点进行 SLAM 系统的位姿估计。DRSO-SLAM 的系统框图如图 6 所示。

DRSO-SLAM 系统基于 ORB-SLAM2 算法框架，分为动态特征点检测线程、跟踪线程、局部建图线程和闭环检测线程。在动态特征点检测线程中，首先基于 Mask R-CNN 估计潜在的运动物体，并使用静态特征点进行粗略的自运动估计，然后使用金字塔光流跟踪图像中的特征点，最后使用对极几何剔除图像帧中缺失和虚假的动态特征点，仅使用静态特征点估计系统的初始位姿。DRSO-SLAM 算法在 TUM 数据集上得到了验证，在低动态场景下具有较高的相机位姿精度。在高动态场景下与 ORB-SLAM2 相比，该系统的均方根误差改善值平均为 95.02%。未来研究者们还将在不影响精度的前提下，进一步改进语义信息提取线程，提高实时性，实现实时运行实例分割算法。

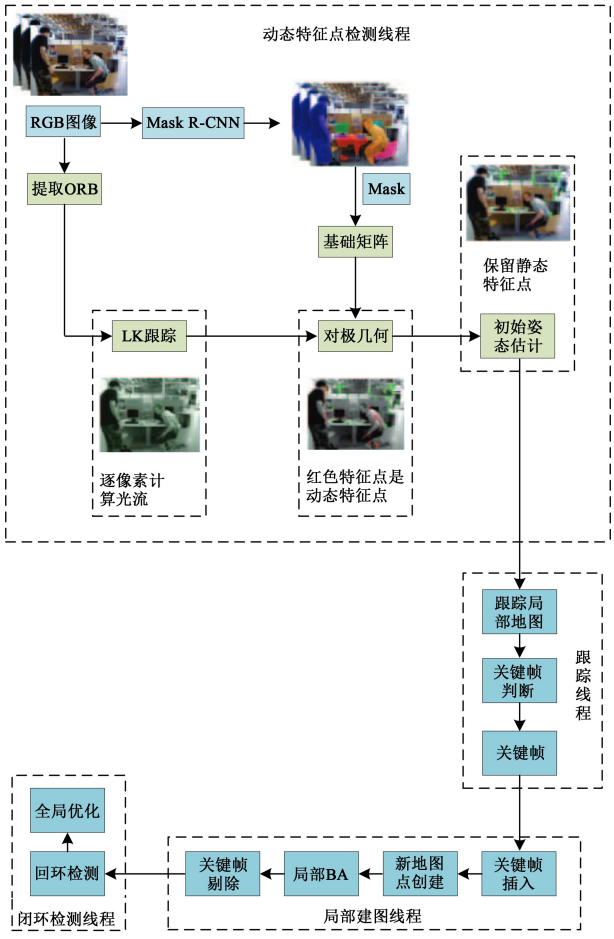


图 6　DRSO-SLAM 的系统框图

4.2.3　小结

实例分割是目标检测和语义分割的结合，它可以将输入图像中的目标检测出来，并且对目标的每个像素分配类别标签。实例分割是上述三种图像分割方法中最难的，也是分割最精细的，可以实现复杂动态环境下的精确分割。近年来出现了许多优秀的基于实例分割的视觉 SLAM 系统，表 3 中列出了一部分代表，它们在动态环境下都表现出了良好的鲁

棒性。但实例分割也存在不足，与语义分割一样实时性不能得到保证，另外，在处理遮挡问题时，会造成目标信息的丢失，这些问题也是当前的研究热点。

表 3 基于实例分割的视觉 SLAM 系统

系统	年份	传感器	场景	动态检测
DynaSLAM	2018	Monocular/Stereo/RGB-D	室内/室外	Mask R-CNN
SLAMANTIC	2019	Monocular/Stereo	室外	Mask R-CNN
YO-SLAM	2021	RGB-D	室内	Yolact
DRSO-SLAM	2021	RGB-D	室内	Mask R-CNN
RDMO-SLAM	2021	RGB-D	室内	Mask R-CNN

5 展望与总结

通过对基于目标检测、基于语义分割和基于实例分割的视觉 SLAM 系统的介绍和分析可以发现，三类视觉 SLAM 系统各有优势并都存在不足。结合近年来的发展趋势，本文对动态环境下基于深度学习的视觉 SLAM 方向做出如下展望：

1)实时性的提升。视觉 SLAM 系统是需要应用到实际生活中的，实时性是必须要考虑的因素。引入深度学习的方法虽然可以有效提升视觉 SLAM 系统在动态环境下的鲁棒性和准确性，但深度学习的加入会使得系统的实时性降低。因此，减少系统耗时，提升实时性，是动态环境下基于深度学习的视觉 SLAM 重要发展方向之一。

2)更精确的分割。细小物体的分割一直是图像分割的难点。在面对复杂动态环境中的小物体，基于深度学习的方法往往会丢失细节，无法做到精确分割，造成动态特征点的残留，影响位姿估计的精度。如何提高深度学习网络的分割精度，以应对更加复杂的动态环境，是未来必须要面对的问题。

3)保留更多的静态特征点。深度学习网络在剔除动态特征点时，往往会将动态特征点所在的区域整个剔除，造成动态区域中静态特征点的剔除，但这些静态点所包含的静态信息对后续相机的跟踪和地图的构建是有益的。利用背景修复等技术减少动态区域中静态特征点的浪费，进一步提高系统的精度，也是未来的重要研究方向。

4)硬件的升级。基于深度学习的视觉 SLAM 严重依赖于计算机的硬件性能，硬件的提升可以大大提高深度学习网络处理图像的速度，有效提升视觉 SLAM 系统的实时性。

综上所述，本文首先介绍了动态环境下基于深度学习的视觉 SLAM 分类，然后详细介绍了基于目标检测、基于语义分割和基于实例分割的视觉 SLAM，并对它们进行了分析比较。最后，结合近年来视觉 SLAM 的发展趋势，展望了未来的发展方向，为后续研究提供了参考。

参考文献：

[1] Bresson G, Alsayed Z, Li Y, et al. Simultaneous localization and mapping: A survey of current trends in autonomous driving[J]. IEEE Trans. on Intelligent Vehicles, 2017, 2(3): 194-220.

[2] Li R H, Wang S, Gu D B. Ongoing evolution of visual SLAM from geometry to deep learning: challenges and opportunities [J]. Cognitive Computation, 2018, 10(6): 875-889.

[3] 胡章芳，张 杰，程 亮. 融合特征法与直接法的 RGB-D 稠密视觉 SLAM 算法[J]. 半导体光电，2020，41(4)：548-554，559.

Hu Zhangfang, Zhang Jie, Cheng Liang. RGB-D dense visual SLAM algorithm combining feature method and direct method [J]. Semiconductor Optoelectronics, 2020, 41(4): 548-554, 559.

[4] 田 野，陈宏巍，王法胜，等. 室内移动机器人的 SLAM 算法综述[J]. 计算机科学，2021，48(9)：1-26.

Tian Ye, Chen Hongwei, Wang Fasheng, et al. Overview of SLAM algorithms for indoor mobile robots [J]. Computer Science, 2021, 48(9): 1-26.

[5] Teed Z, Deng J. Droid-SLAM: Deep visual slam for monocular, stereo, and RGB-D cameras [J]. Advances in Neural Information Processing Systems, 2021, 34: 16558-16569.

[6] 杨永刚，武楚健，杨正全. 基于融合改进 RANSAC 光流法的无人机视觉 SLAM 研究[J]. 半导体光电，2023，44(2)：277-283.

Yang Yonggang, Wu Chujian, Yang Zhengquan. Research on UAV visual SLAM based on fusing improved RANSAC optical flow method[J]. Semiconductor Optoelectronics, 2023, 44(2): 277-283.

[7] Wen S, Li P, Zhao Y, et al. Semantic visual SLAM in dynamic environment[J]. Autonomous Robots, 2021, 45(4): 493-504.

[8] Su P, Luo S, Huang X. Real-time dynamic SLAM algorithm based on deep learning[J]. IEEE Access, 2022, 10: 87754-87766.

[9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. USA: IEEE, 2016: 779-788.

[10] Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[11] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]// Proc. of the IEEE Inter. Conf. on Computer Vision. USA: IEEE, 2017: 2961-2969.

[12] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network [C]// IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). USA: IEEE, 2017: 6230-6239.

[13] Li G, Liu Z, Ling H. ICNet: information conversion network for RGB-D based salient object detection[J]. IEEE Trans. on Image Processing, 2020, 29: 4873-4884.

[14] Yu C, Liu Z, Liu X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C]// 2018 IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems (IROS). IEEE, 2018: 1168-1174.

[15] Bescos B, Fácil J M, Civera J, et al. DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes [J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076-4083.

[16] Long X, Zhang W, Zhao B. PSPNet-SLAM: A semantic SLAM detect dynamic object by pyramid scene parsing network[J]. IEEE Access, 2020, 8: 214685-214695.

[17] Cheng S, Sun C, Zhang S, et al. SG-SLAM: A real-time RGB-D visual SLAM toward dynamic scenes with semantic and geometric information [J]. IEEE Trans. on Instrumentation and Measurement, 2022, 72: 1-12.

[18] Wei W, Huang K, Liu X, et al. GSL-VO: A geometric-semantic information enhanced lightweight visual odometry in dynamic environments[J]. IEEE Trans. on Instrumentation and Measurement, 2023, 72: 2522513.

[19] Mur-Artal R, Tardós J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras [J]. IEEE Trans. on Robotics, 2017, 33(5): 1255-1262.

[20] Campos C, Elvira R, Rodríguez J J G, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. IEEE Trans. on Robotics, 2021, 37(6): 1874-1890.

[21] Chen W, Shang G, Ji A, et al. An overview on visual SLAM: From tradition to semantic[J]. Remote Sensing, 2022, 14(13): 3010.

[22] Pazhani A A J, Vasanthanayaki C. Object detection in satellite images by faster R-CNN incorporated with enhanced ROI pooling (FrRNet-ERoI) framework[J]. Earth Science Informatics, 2022, 15(1): 553-561.

[23] Hu K, Lu F, Lu M, et al. A marine object detection algorithm based on SSD and feature enhancement [J]. Complexity, 2020, 2020: 1-14.

[24] Shao F, Chen L, Shao J, et al. Deep learning for weakly-supervised object detection and localization: A survey[J]. Neurocomputing, 2022, 496: 192-207.

[25] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2014: 580-587.

[26] Girshick R. Fast R-CNN[C]// Proc. of the IEEE Inter. Conf. on Computer Vision, 2015: 1440-1448.

[27] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[28] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]// Computer Vision-ECCV 2016: 14th European Conf., Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part Ⅰ 14. Springer International Publishing, 2016: 21-37.

[29] Zhong F, Wang S, Zhang Z, et al. Detect-SLAM: Making object detection and SLAM mutually beneficial[C]// 2018 IEEE Winter Conf. on Applications of Computer Vision (WACV). IEEE, 2018: 1001-1010.

[30] Gong H, Gong L, Ma T, et al. AHY-SLAM: Toward faster and more accurate visual SLAM in dynamic scenes using homogenized feature extraction and object detection method [J]. Sensors, 2023, 23(9): 4241.

[31] Wu W, Guo L, Gao H, et al. YOLO-SLAM: A semantic SLAM system towards dynamic environment with geometric constraint[J]. Neural Computing and Applications, 2022: 1-16.

[32] Wang Z, Zhang Q, Li J, et al. A computationally efficient semantic SLAM solution for dynamic scenes[J]. Remote Sensing, 2019, 11(11): 1363.

[33] Liu Y, Miura J. RDS-SLAM: Real-time dynamic SLAM using semantic segmentation methods[J]. IEEE Access, 2021, 9: 23772-23785.

[34] Wang Y, Bu H, Zhang X, et al. YPD-SLAM: A real-time VSLAM system for handling dynamic indoor environments [J]. Sensors, 2022, 22(21): 8561.

[35] Hoang T M, Zhou J, Fan Y. Image compression with encoder-decoder matched semantic segmentation[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops, 2020: 160-161.

[36] Esparza D, Flores G. The STDyn-SLAM: A stereo vision and semantic segmentation approach for VSLAM in dynamic outdoor environments[J]. IEEE Access, 2022, 10: 18201-18209.

[37] Zhao Y, Xiong Z, Zhou S, et al. KSF-SLAM: A key segmentation frame based semantic SLAM in dynamic environments[J]. J. of Intelligent & Robotic Systems, 2022, 105(1): 3.

[38] Li X, Belaroussi R. Semi-dense 3D semantic mapping from monocular SLAM[J]. arXiv preprint arXiv: 1611.04144, 2016[2023-11-22]. http:// arSiv. labs. arxiv. org/html/1611.04144.

[39] Han S, Xi Z. Dynamic scene semantics SLAM based on semantic segmentation[J]. IEEE Access, 2020, 8: 43563-43570.

[40] Schorghuber M, Steininger D, Cabon Y, et al. SLAMANTIC-leveraging semantics to improve VSLAM in dynamic environments[C]// Proc. of the IEEE/CVF Inter. Conf. on Computer Vision Workshops, 2019: 3759-3768.

[41] Lai D, Li C, He B. YO-SLAM: A robust visual SLAM towards dynamic environments[C]// 2021 Inter. Conf. on Communications, Information System and Computer Engineering (CISCE). IEEE, 2021: 720-725.

[42] Yu N, Gan M, Yu H, et al. DRSO-SLAM: A dynamic RGB-D SLAM algorithm for indoor dynamic scenes[C]// 2021 33rd Chinese Control and Decision Conference (CCDC). IEEE, 2021: 1052-1058.

[43] Liu Y, Miura J. RDMO-SLAM: Real-time visual SLAM for dynamic environments using semantic label prediction with optical flow[J]. IEEE Access, 2021, 9: 106981-106997.

[44] Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. IEEE Trans. on Robotics, 2015, 31(5): 1147-1163.

作者简介：

罗　元(1972—)，女，湖北省宜昌市人，博士，教授，硕导，主要研究方向为移动机器人 SLAM、机器视觉；

沈吉祥(2000—)，男，安徽省芜湖市人，硕士研究生，主要研究方向为视觉 SLAM。