

附件 2

# 江苏省研究生实践创新计划

## 项目申报书

申 请 人: 吴麟

专业学位类别代码: 0852

专业学位类别名称: 工程硕士

实践研究方向: 控制工程

申请项目名称: 视觉 SLAM 中动态目标检测与静态场景恢复

指 导 教 师 : 孙长银 教授

所 在 学 校 : 东南大学 (盖章)

江 苏 省 学 位 委 员 会 制表  
江 苏 省 教 育 厅

## 填 表 说 明

一、填写本表前，应先仔细阅读有关通知文件及本说明，务必实事求是填写。

二、填写本表栏目时，如需要可加附页。

三、专业学位类别参见：

<http://www.cdgdc.edu.cn/xwyyjsjyxx/gjjl/cjwt/276470.shtml>

四、本表所有信息必须全部填写，不存在的内容一律填“无”。

项目概况	项目名称	视觉 SLAM 中动态目标检测与静态场景恢复						
	项目类别	A.人文社科项目 <input type="checkbox"/> B.自然科学项目 <input checked="" type="checkbox"/>		起止年限	2020 年 6 月 至 2021 年 6 月			
申请人	姓名	吴麟	性别	男	出生年月	1997.04	博(硕)士入学年月	2019.09
	所在院系	自动化学院		联系电话	17721726716	电子信箱	wulin@seu.edu.cn	

一、立项依据（包括项目来源，实践研究意义，国内外实践研究现状、水平和发展趋势等）

## 1 项目来源及研究意义

随着智能信息技术和传感器技术的飞速发展，移动机器人也逐渐在工业制造、城市安防、家用服务等多种场景下得到广泛应用，典型的如仓储运货的 AGV 小车、商城巡航的安保机器人、家用的扫地机器人等[1]。与此同时，AR/VR、无人机、自动驾驶等领域的研究热度持续升温，创新成果层出不穷，展现出这些新兴智能产业技术的巨大潜力。而这些相关领域蓬勃发展的背后都无可避免地涉及到同时定位与地图构建技术（Simultaneous Localization and Mapping, SLAM），在诸多不同类型的 SLAM 系统中，基于视觉的 SLAM 由于获取环境信息丰富且成本较低等特点而格外受到关注。

视觉 SLAM 是装载着摄像头等视觉传感器的智能机器人在未知环境进行自感知的重要手段，目前大部分视觉 SLAM 系统工作时都基于一个重要假设，即观测环境是静态的或者观测信息的变化只依赖于相机自身运动[2]，然而真实环境往往具有动态性。一方面，环境中经常出现如行走的人、运动的车辆等动态物体；另一方面，环境本身也会因季节、天气等因素而发生变化，如雨天环境基于特征点匹配的视觉 SLAM 系统非常容易出现定位误差。尽管目前已有很多算法对小的运动物体引入的离群点进行处理，如随机采样一致性（RANSAC），但当动态物体在图像占据较大面积时，视觉 SLAM 系统存在较大误差[3]。

智能机器人相比于传统机器人具有更多独特的特征，这也对智能机器人提出了更高的要求，如智能感知、自主决策、运动控制等。而这些任务的完成 SLAM 技术必不可少，但由于系统工作环境从静态稳定不断向动态复杂延拓，感知任务从粗略单一到精细多样扩展，如此种种均对 SLAM 的高效性、鲁棒性以及智能化程度提出了更高要求与挑战。

因此本项目研究视觉 SLAM 系统中针对场景动态性的特殊处理方法，着眼于动态场景下的动态目标检测、去除以及静态背景重建，对提高 SLAM 系统的鲁棒性与准确性具有重要价值。此外，由于机器学习算法研究的深入，机器人还可以从图像中获取场景的高级特征，如语义类别，来进行更高等级的场景理解[4]，或通过多模态信息融合提高数据关联的准确性[5]，也同样为动态场景下的视觉 SLAM 系统提供了新的研究思路，故而基于几何和学习的方法逐步成为研究热点，对机器人未知环境下的“感、知、控”具有推动意义。

## 2 国内外实践研究现状

近三十年来，随着机器视觉以及相关软硬件技术的发展，视觉 SLAM 也持续高速地发展，业已成为机器人领域的关键技术，视觉 SLAM 的框架类型与分类方式也呈现出多元化特点。按照视觉相机的类型，可以分为单目、双目、RGB-D；按照建模方式，视觉 SLAM 可以分为基于滤波器的方法和基于非线性优化的方法；以使用数据的方式分类，视觉 SLAM 可以分为直接法、半直接法和特征点法；以地图构建的稀疏程度分类，视觉 SLAM 算法可以分为稀疏法、半稠密法和稠密法。因为涉及庞杂，故以下简要从静态场景以及动态环境下的视觉 SLAM 技术两个方面进行介绍。

### (1) 基于静态场景假设的视觉 SLAM 研究现状

任何一个 SLAM 系统都包括两个主要任务：定位与建图，同时，这也是决定机器人运动过程以及构建准确的环境地图的重要问题，因此自从 SLAM 的概念提出以来便饱受关注。2002 年，S. Thrun 等人在《Probabilistic robotics》中将 SLAM 过程数学化为两个方程：运动方程和观测方程[6]，SLAM 早期的工作也正是基于该框架，将位置和路标视为服从某种分布的变量。从而，将 SLAM 问题描述为：当机器人获得控制数据和观测数据时，如何得到位置和路标的分布。

此后，概率统计方法一度成为机器人定位与建图领域的主导方法。1986 年 R. Smith 利用扩展卡尔曼滤波（Extended Kalman Filter, EKF）求解状态方程，取得了不错的效果[7]。随之，研究人员发展了各种各样的滤波器，如无迹卡尔曼滤波（Unscented Kalman Filter, UKF）[8]和能够避免线性化的粒子滤波（Particle Filter, PF）[9]。早期的 SLAM 系统主要使用超声波、激光雷达等传感器，随着机器视觉算法的演进，视觉 SLAM 技术应运而生。

2007 年，帝国理工学院 Davision 等人提出利用单目相机进行同时定位与建图的 Mono-SLAM 算法[10]，开启了实时的视觉 SLAM 研究新的篇章。其后，Klein 等人提出了并行跟踪与建图算法（Parallel Tracking and Mapping, PTAM）[11]，该算法创造性地将整个 SLAM 系统分为两线程：跟踪与地图构建，一度成为后来研究者的基准。2014 年，Engel 等人提出 LSD-SLAM 算法[12]，阐明了像素梯度与直接法的关系。与此同时，Forster 等人发展一种新的半直接法单目视觉里程计（Semi-direct monocular Visual Odometry, SVO），它将特征点与直接跟踪光流方法很好的结合在一起[13]。之后，视觉 SLAM 框架如 DSO、VINS-Mono 等都沿用直接法，这种方法在跟踪和匹配过程有效节约计算机资源，且对环境光照等因素较为敏感，这对视觉 SLAM 的实际应用构成极大挑战。

因此，为了构建一个完善稳定的视觉 SLAM 框架，特征点的使用是不可或缺的。这方面典范的工作如 Leutenegger 等人提出的 OKVIS 视觉惯性里程计框架，Mur-Artal 等人推出的 ORB-SLAM 以及 ORB-SLAM2 等[14-15]。其中 ORB-SLAM2 框架是经典 SLAM 线程模型的继承者。为了完成 SLAM 系统，它使用 ORB 特征点和三个主要并行线程：跟踪线程实时跟踪特征点，局部建图线程构建局部束调整（Bundle Adjustments, BA）建图，闭环检测线程纠正累积漂移并执行位图优化。它可以使系统在大场景下长时间运行，保证轨迹和地图的全局一致性。ORB-SLAM2 是同时定位和地图构建的最佳框架之一。然而，在处理动态环境问题时还存在许多不足之处，有必要进行进一步的探索。

## (2) 动态场景下的视觉 SLAM 研究现状

由于 SLAM 应用场景的泛化，动态场景在所难免。基于特征点法视觉 SLAM 认为，如果在位姿估计的解算过程中只保留静态特征，则该位姿估计是准确的。基于这个思路，许多研究者在运动分割方面做了大量探索。当前的运动分割方法主要包括：光流法、几何约束法、深度学习技术等，其与 SLAM 技术结合产生了众多鲁棒的视觉 SLAM 方案。

**光流法** 光流定义了从两幅连续图像计算出的亮度模式的视运动[16]。一般来说，它与图像中的运动场相对应，因此可以用来分割运动的物体。Klappstein[17]基于光流计算的运动度量定义了运动对象的可能性。运动度量度量的是在场景中有运动物体时光流被破坏的程度。采用基于运动度量的图像分割算法对运动目标进行分割。

Alcantarilla 等人基于场景流(光流的 3D 版本)中的三维运动矢量模量，通过残差运动概率对运动对象进行[18]分割。在基于稠密光流和立体重建的场景流计算中，马氏距离被用来考虑测量的不确定性。如果残差很低，那么特征点很可能属于静态对象。通过对残差运动概率的阈值化，将运动目标上的特征点从 SLAM 过程中剔除，使视觉测程估计更加稳健。Derome 等人[19]通过计算预测图像与立体相机观测图像之间的残差来计算光流。通过时间倒序处理，将当前的立体帧转换为前一帧，利用估计的相机自我运动，计算出预测的图像。然后通过检测残差场中的斑点来观察运动物体。

**几何约束** 依赖于几何约束的技术利用对极几何特性[20]来分割静态和动态特性。它们基于这样一个事实，即动态特性将违反静态多视图几何结构中定义的标准约束。约束条件可以从极线方程、三角剖分、基本矩阵估计或重投影误差等推导出来。

Lin 等人[21]通过将运动对象错误地分类为静态对象并将其合并到姿态估计中来检测运动对象，这将显著降低 SLAM 的性能。他们计算了两种不同情况下的位姿估计的差异，一种是在不添加检测到的新特征的情况下，另一种是在假设新特征是平稳的情况下加入新特征的情况下。通过计算两个结果之间的距离，设置阈值，并通过二进制贝叶斯滤波器对其进行积分，可以对静止和运动特征进行高精度的分割。

另一种几何方法是利用重投影误差。Tan 将前一帧的特征投影到当前帧中，并测量被跟踪特征的距离。静态和动态特征的分类是由它们的重投影距离的大小决定的。Tan 等[22]也使用了类似的投影原理来检测动态特征。然而，他们也考虑遮挡处理来提供健壮的视觉 SLAM。将特征投影到当前帧后，使用外观差异进行检查图像的一部分是否发生了变化。如果外观发生明显变化，则很有可能是由于视角改变而被动态对象或静态对象遮挡。被这些条件遮挡的三维点将被保留，并被用来稳健地估计相机的姿态。

**深度学习** 深度神经网络(Deep Neural Networks, DNNs)在赢得 ImageNet 对象识别比赛后，通过将分类错误减少一半，与最先进的技术相比[23]，深度神经网络(Deep Neural Networks, DNNs)在计算机视觉领域得到了广泛的应用。

Lin 和 Wang[24]构建了一个网络来显式分割图像空间中的移动对象。他们使用重构独立分量分析(RICA)自编码程序[2]来学习时空特征。然而，由于时空特征无法学习运动的三维几何，因此仍然使用几何特征来帮助分割运动。将几何和时空特征输入递归神经网络(RNNs)进行最终的运动分割。Fragkiadaki 等人使用另一种方法，通过对给定 RGB 图像和光流的客观度评分进行回归，对移动对象进行分割。构造两个类似于 AlexNet 的并行 CNNs，

在将 RGB 图像和光流送入回归网络并生成运动建议之前对其进行处理。

清华大学学者提出 DS-SLAM[4]，在 ORB-SLAM2 框架下进行改进，增加了语义分割和稠密地图，在语义分割线程，利用 SegNet 网络对输入图像进行语义分割、进而剔除动态物体，相比于 ORB-SLAM2、DS-SLAM 在动态环境下的精度大大提升。Bescos[26]等人提出 DynaSLAM，利用 Mask R-CNN 网络检测动态目标、同时采用背景修复技术，能够在动态环境下剔除动态物体的同时，修补动态目标遮挡区域，构建出高精度的静态地图，但该算法耗时严重，还不能满足实时性需求。

### (3) 小结

视觉 SLAM 技术是智能机器人感知与决策过程的必然要求，动态场景是智能机器人工工作时面临的必然挑战。而推动 SLAM 处理动态场景技术发展的措施主要分两种，其一，为了克服硬件上的限制，视觉 SLAM 技术所使用的硬件设备从单目相机到多目以及多传感器集成发展，如 RGB-D 相机；其二，在传统 SLAM 线程中加入新的模块，如动态目标检测与外点滤除等。随着深度学习技术的发展，其已在目标检测与识别、语义分割等多方面体现出独特的优势，利用这种优势，提升视觉 SLAM 系统在静态与动态场景下的表现，已经成为视觉 SLAM 发展的新方向。本课题开展动态场景下的鲁棒的视觉 SLAM 研究，针对目前深度学习在动态目标检测中的不完备性，发展一种新的动态检测与动态目标去除的方法；同时，针对传统动态去除算法难以处理较大面积动态区域的弱点，发展一种新的动态场景到静态场景的恢复方法，并将其嵌入视觉 SLAM 框架内，服务于静态地图构建任务。

### (4) 参考文献

- [1]. 李云天, 穆荣军与单永志, 无人系统视觉SLAM技术发展现状简析. 控制与决策: 第1-10页.
- [2]. Huang, B., J. Zhao and J. Liu, A Survey of Simultaneous Localization and Mapping with an Envision in 6G Wireless Networks. 2019.
- [3]. 张金凤, 石朝侠与王燕清, 动态场景下基于视觉特征的SLAM方法. 计算机工程: 第1-8页.
- [4]. Yu, C., et al., DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments. 2018.
- [5]. 闫光, 基于多传感器信息融合的动态目标检测与识别, 2015, 北京理工大学. 第 87页.
- [6]. Smith, R.C. and P. Cheeseman, On the Representation and Estimation of Spatial Uncertainty. The International Journal of Robotics Research, 1986. 5(4): p. 56-68.
- [7]. Smith, R., M. Self and P. Cheeseman. Estimating uncertain spatial relationships in robotics. 1987: IEEE.
- [8]. Wang, X. and H. Zhang. A UPF-UKF Framework For SLAM. 2007: IEEE.
- [9]. Li, X., et al. Distributed parallel processing of mobile robot PF-SLAM. 2012. Stevenage, UK: IET.
- [10]. Davison, A.J., et al., MonoSLAM: Real-Time Single Camera SLAM. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007. 29(6): p. 1052-1067.
- [11]. Klein, G. and D. Murray. Parallel Tracking and Mapping for Small AR Workspaces. 2007: IEEE.
- [12]. Engel, J., J. Stuckler and D. Cremers. Large-scale direct SLAM with stereo cameras. 2015: IEEE.
- [13]. Forster, C., M. Pizzoli and D. Scaramuzza, SVO: Fast semi-direct monocular visual odometry, 2014 IEEE International Conference on Robotics and Automation (ICRA). 2014. p. 15-22.
- [14]. Mur-Artal, R., J.M.M. Montiel and J.D. Tardos, ORB-SLAM: A Versatile and Accurate Monocular

- SLAM System. IEEE Transactions on Robotics, 2015. 31(5): p. 1147-1163.
- [15].Mur-Artal, R. and J.D. Tardos, ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. IEEE Transactions on Robotics, 2017. 33(5): p. 1255-1262.
- [16].Horn, B.K.P. and B.G. Schunck, Determining optical flow. Artificial Intelligence, 1981. 17(1): p. 185-203.
- [17].Gu, J.D., Moving object segmentation and tracking for intelligent video surveillance system. 2009, ProQuest Dissertations Publishing.
- [18].Alcantarilla, P.F., et al. On combining visual SLAM and dense scene flow to increase the robustness of localization and mapping in dynamic environments. 2012: IEEE.
- [19].Derome, M., et al. Real-time mobile object detection using stereo. 2014: IEEE.
- [20].Hartley, R. and A. Zisserman, Multiple view geometry in computer vision. 2003, Cambridge [England];New York: Cambridge University Press.
- [21].Lin, K. and C. Wang. Stereo-based simultaneous localization, mapping and moving object tracking. 2010: IEEE.
- [22].Zou, D. and P. Tan, CoSLAM: Collaborative Visual SLAM in Dynamic Environments. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013. 35(2): p. 354-366.
- [23].Krizhevsky, A., I. Sutskever and G. Hinton, ImageNet classification with deep convolutional neural networks. 2017, ACM. p. 84-90.
- [24].Lin, T. and C. Wang, Deep learning of spatio-temporal features with geometric-based moving point detection for motion segmentation, 2014 IEEE International Conference on Robotics and Automation (ICRA). 2014. p. 3058-3065.
- [25].Le, Q.V. Building high-level features using large scale unsupervised learning. 2013: IEEE.
- [26].Bescos, B., et al., DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes. IEEE Robotics and Automation Letters, 2018. 3(4): p. 4076-4083
- .

## 二、研究目标、内容和拟解决的主要问题

### 1 研究目标

以视觉 SLAM 系统为对象，拟解决其在动态场景下基于特征点提取与匹配位姿估计方法的不稳定等问题，包括以下几个方面：

#### (1) 为实现动态场景下鲁棒的定位与建图、提出图像动态检测与静态场景恢复算法

基于特征点提取与匹配的视觉 SLAM 系统在面对动态场景时往往存在较大误差，而误差的来源则主要是系统引入了大量的动态特征点。因此，图像动态检测尤为关键，而现有的处理方法大多采用几何约束或深度学习进行动态内容识别或语义分割，当算法检测为动态点后往往简单地将其去除，致使部分图像像素缺失，而本项目则希望利用背景图像将整个静态场景恢复，从而实现静态地图地准确构建，如图 1 所示。



图 1 动态检测与静态场景恢复（左：动态场景；中：静态场景；右：动态检测掩膜）

#### (2) 提出一种利用动态到静态场景转换提升视觉 SLAM 精度的算法

在视觉 SLAM 系统中增加一个动态处理模块以提高 SLAM 效率与精度已被研究人员广泛采用，本项目基于动态场景到静态场景地转换模块，如图 2，一方面提升了视觉 SLAM 的动态处理能力，另一方面对几何难以恢复的场景，本项目提出的静态场景恢复算法也能重建静态区域。此外，根据动态到场景转换，提出一种提升机器人重定位精度的算法，同样可以嵌入视觉 SLAM 系统，增强智能机器人的场景适应性。

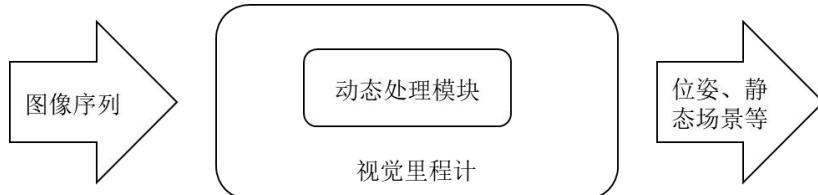


图 2 动态到静态场景转换模块与视觉里程计融合

#### (3) 算法集成及实验验证

将上述方法集成在机器人系统之中，并完成实验验证，实现智能机器人在复杂场景下的自主稳定的定位与建图。

## 2 研究内容

本项目围绕动态场景下的视觉 SLAM 系统展开研究，结合深度学习技术，对传统视觉 SLAM 框架下的视觉里程计、闭环检测等模块进行补充和延拓，具体包括以下几点：

### (1) 基于深度学习的图像动态检测与静态场景恢复方法研究

针对传统的视觉 SLAM 方法的场景适应性不足进行研究，即其只能适用于静态场景，并且场景需要一定条件（如适宜光照、明显的纹理结构和帧间图像有重叠等），目前已有许多利用深度学习进行（潜在）动态目标检测的成果，但其检测任务并非直观作用于动态、静态的分割，而是作用于语义分割和目标识别。因而，本项目着力于对图像中潜在动态与静态区域进行直接分割，同时在分割的基础上，对动态区域进行图像恢复，利用生成对抗网络等深度网络强大的图像生成能力，实现动态场景到静态场景的转化。

### (2) 基于动态场景到静态场景转换的视觉 SLAM 方案研究

针对动态场景到静态场景转换的效果应用进行研究。在视觉里程计中，静态场景恢复对视觉 SLAM 精度的提升与实时性损耗；在闭环检测中，静态场景恢复对重定位精度的影响；在视觉语义 SLAM 系统中，静态场景恢复为对语义一致性的影响。

### (3) 算法集成设计及验证

将上述理论方法集成并应用在实际的视觉 SLAM 系统中的算法及软硬件实现问题，以及相应的实际实验设计和验证工作。

## 3 拟解决的主要问题

本项目拟解决的关键问题为：智能机器人同时定位与地图构建过程中，考虑到动态因素的影响，致使机器人位姿估计产生较大误差，以及难以构建准确的静态地图等问题。

(1) 随着视觉 SLAM 应用场景的升级，由于动态场景下基于视觉特征的 SLAM 系统具有误差大、稳定性差等缺点，因而需要减少动态特征点对系统的影响。另外，由于场景中的动态物体的随机运动，对于基于视觉的智能机器人重定位提出了极大的挑战，因为提取特征点的迥异非常容易诱导机器人做出实际同样地点但算法输出为非同样地点的误判。

(2) 动态图像场景到静态场景的转换一旦完成，由于静态图像中具有一定的伪成分，如何使用带有伪成分的静态场景图像辅助视觉 SLAM 系统进行位姿估计或闭环检测，便成为影响视觉 SLAM 系统定位精度与静态地图构建能力的重要制约。因此，利用动态静态转化算法，并将其融入现有的视觉 SLAM 框架成为一个关键性问题。

(3) 现有的动态视觉 SLAM 框架中几何信息与语义信息耦合较为松散，主要是由于这些系统删除动态特征时，要么只依赖于几何信息，要么只依赖于语义信息，要么以简单方式将动态特征删除结果组合起来，没有对两者进行细致的讨论，因此不论是利用几何还是深度学习结果，都需要结合动态场景下视觉 SLAM 的特殊任务，从而进行准确的动态分割和静态地图构建。

### 三、项目的研究思路与方法、技术路线、试验方案（含创新性）及其可行性分析

#### 1 研究思路与方法

##### (1) 基于深度学习的动态区域检测与静态背景恢复技术

深度神经网络在计算机视觉领域展现出强大的图像处理能力，随着多年的发展，多层卷积网络、生成对抗网络、残差神经网络、长短期记忆网络等层出不穷，其中生成对抗网络在图像生成任务中表现优异，尤其是条件生成对抗网络，可以根据给定的标签根据网络经验去生成期望的图像，因此考虑采用其框架，对图像的动态区域采用特殊处理方式，进而生成动态、静态分割掩膜，融入到静态背景恢复的统一框架中。

##### (2) 利用动态、静态场景转换结果辅助定位与建图方法

典型的视觉 SLAM 框架由传感器数据读取、视觉里程计、后端优化、地图构建和回环检测。理论上，动态场景到静态场景的转换方法可以应用到位姿估计和静态地图构建，即涉及视觉里程计、地图构建和回环检测三个环节。在视觉里程计环节，动态检测可以滤除动态特征点的影响；在建图环节，静态图可以辅助构建静态地图；回环检测环节，考虑利用静态恢复后的路标进行定位，减少重定位误差。

##### (3) 算法融入现有视觉 SLAM 框架及仿真、试验

在目前先进的视觉 SLAM 算法基础上进行改进融合，如 ORB-SLAM，LSD-SLAM 等。通过各种评价指标，来评估算法的效率和价值。在视觉 SLAM 系统工作过程中，考虑基于几何和深度学习两种途径，对动态目标进行检测，因为从单张图像中只能识别潜在动态的模式，而在实际的 SLAM 系统中，考虑利用几何的模式与深度学习的模式进行联合。

#### 2 技术路线

本项目提出动态场景下鲁棒的视觉 SLAM 系统的设计路线，主要为四个部分构成。基于无人驾驶模拟器进行数据集生成、基于深度学习的动态区域检测与静态背景恢复技术、基于深度学习的动态区域检测与静态背景恢复技术以及将算法融入现有视觉 SLAM 框架及仿真试验

第一部分是根据 CARLA 进行动态场景与静态场景数据集的获取。CARLA 是用于自动驾驶研究的开源模拟器，旨在支持自动驾驶系统的开发，训练和验证。除了开源代码和协议，CARLA 还提供了为此目的而创建且可以免费使用的开放数字资产（城市布局，建筑物，车辆）。该仿真平台支持传感器套件的灵活规范，环境条件，对所有静态和动态参与者的完全控制，地图生成等等。

第二部分的动态区域检测与静态背景修复是本项目的基础也是核心，其研究思路是在当前深度神经网络研究的基础上进行升级以适配本项目的任务，利用条件生成对抗网络，以及 CARLA 无人驾驶模拟器生产的数据集，构建完整的端到端的框架，将潜在动态内容的检测设计到整个网络中，形成动态到静态的转换过程。

第三部分是根据第一阶段的结果，分别在视觉 SLAM 的多个环节进行改进。视觉里程计环节，利用对极几何、重投影误差以及深度网路检测的动态内容，对获取的图像特征进行静态与动态的分割，若动态区域较大，则考虑使用静态恢复的结果辅助位姿估计，否则引入太多的外点将显著降低位姿估计的精度；建图环节，采用 RGB-D 相机建立稠密地图时，采用滤除动态点以及静态恢复后的图像生成，而不是掺杂动点的输入图像；在闭环检测阶段，若动态区域较大，则采用静态恢复后的场景图像进行重定位。

第四部分主要对前面两个部分的算法进行仿真试验以及实际系统开发，基于 ORB-SLAM2 等先进的基准算法，并在此基础上进行开发，使用公开数据集如 TUM、KITTI 进行算法评测，使用真实轨迹与计算轨迹的误差来衡量定位精度。并将相应处理模块，在硬件平台上进行测试，以评测算法的性能。

### 3 试验方案及创新性分析

**试验方案：**在基于深度学习的动态区域检测与静态背景恢复技术研究中，使用常见的深度学习框架 Pytorch 或 Tensorflow，通过其搭建相应的深度神经网路，包含动态区域的检测模块，静态场景的恢复模块，通过 CARLA 无人驾驶模拟器使用 Python 脚本生成一一对应的静态场景与动态场景组合，构建一个适合动态检测与静态场景恢复的专用数据集，并以此训练所搭建的深度神经网络，不断优化其恢复效果。而后，在 ORB-SLAM 的基础上，在其视觉里程计阶段加入动态滤除，在其闭环检测阶段根据动态区域占据的面积来选择性地使用静态图像进行重定位，在地图构建环节，若连续帧运动无法获得某些像素点地真实的三位坐标，则使用静态场景生成图像进行辅助构建静态地图。最后，在机器人平台开展相应的试验，将机器人置身于动态场景下，观察其位姿估计过程和轨迹跟踪效果，验证本项目的研究成果的可靠性和实用性。

**创新性：**本项目的研究及试验的创新主要包括三个方面

(1) 同一地点动态场景与静态场景目前暂无公开的数据集可供使用，而 CARLA 无人驾驶模拟器支持视觉、激光雷达等传感器，对静态和动态的参与者可以进行完全控制，且可以定制不同数量不同类型的动态目标甚至天气等情况，基于此可以设计生成本项目所需的数据集。

(2) 从深度学习的角度在动态场景下进行动态目标的检测滤除，摒弃了直接使用目标检测、语义分割结果，如提取出行人、车辆所在区域，将其滤除，而是以动态内容的检测为指导思想，将动态静态检测整合为前景与背景的二分类问题，优化了图像中潜在动态区域的检测过程。

(3) 从静态场景到动态场景转换的结果出发，衍生出一系列的应用渠道。将动态静态分割结果用于位姿估计，比传统的动态检测具有理论上更准确的效果，而针对动态内容区域较大，本算法的作用体现更为明显，因为该转换算法可以生成较为逼真的静态场景图像，对其提取的特征点，比基于动态点提取的特征点具有更好的稳定性。同样，该算法可以用来辅助生成静态地图，这在目前视觉 SLAM 建图任务中研究较少。

## 4 可行性分析

### (1) 理论上可行

随着多年的发展，视觉 SLAM 已经形成传感器数据获取、视觉里程计、后端优化、地图构建、回环检测这样完善的框架。其中前端部分，特征点法、直接法多元多样，后端非线性优化方法，粒子滤波、图优化理论层出不穷。在过去的几十年里，在解决动态环境中的视觉同步定位和重建问题方面取得了重大进展。基于语义分割的动态视觉 SLAM、基于静态点权重的视觉 SLAM、稠密 RGB-D 背景重建动态 SLAM、基于八叉树的对象级多实例动态 SLAM 等方法的提出为本项目提供了重要的理论基础。

### (2) 技术上可行

技术上采用无人驾驶模拟器生成所需训练测试数据集，在 CARLA 中可以设置场景中出现动态目标的类别和数量，如指定若干个行人、车辆等，CARLA 提供了丰富的接口以供用户调用，在服务器模式，其可以输出场景的 RGB 图像、深度图、语义分割图以及此时无人车的位置和姿态等信息。随后使用 Pytorch、Tensorflow 等深度学习开源框架来构建所需深度神经网络，以实现动态内容检测与静态场景识别任务。当该模块建构完毕，将其融合至现有的视觉 SLAM 框架中是较为便利可行的。

## 四、研究工作的总体安排及进度

本项目计划一年完成（2020.6—2021.6），年度进度安排如下：

**2020 年 6—8 月**，研究视觉 SLAM 系统中动态场景下的动态内容检测与静态场景重建算法，利用无人驾驶模拟器生成的合成数据集进行训练及测试，并做阶段性总结。

**2020 年 9—12 月**，研究动态场景到静态场景的转化模块的应用方式，基于此设计视觉里程计中的动态检测模块以及优化闭环检测中的重定位方法。

**2021 年 1—3 月**，研究视觉 SLAM 系统的地图构建，主要是静态地图的构建方法，为智能机器人路径规划等提供便利，并做阶段性总结。

**2021 年 4—6 月**，将上述算法模块在实际系统中试验，实现智能机器人鲁棒的视觉 SLAM 任务，并对整个项目做全面的总结工作。

上述计划视实际研究完成情况，在内容安排上可能进行穿插或作部分调整。

## 五、研究工作的预期成果及成果提交形式

### (1) 理论方面

针对动态场景下鲁棒的视觉同时定位与地图构建任务，提出动态内容检测、静态场景恢复等方法，给出其模型训练效果与仿真分析。基于动态静态场景转换方法，给出其在视觉里程计环节的融合以及闭环检测的融合方法，此外，根据静态场景恢复结果给出完整静态地图构建的方案，并对算法达到的性能指标分析，发表至少一篇期刊论文。

### (2) 应用方面

将所提的动态静态转换与视觉 SLAM 融合算法利用实际的视觉智能机器人硬件平台进行试验验证，对比当前一般场景下视觉 SLAM 系统，研究各方法异同和优缺点，基于此申请实用新型专利或国家发明专利至少一项。

## 六、研究基础和工作条件（申请人与本项目有关的成果，含承担或参与的项目情况；现有的主要仪器设备、研究技术及协作条件等）

申请人为硕士研究生，主研方向为视觉 SLAM、深度学习。申请人具备较好的软硬件理论与实践基础，本科至目前在中国自动化大会等会议上发表学术论文两篇，公开国家发明专利一项，曾主持“科博达”公司自主创新研究基金项目一项，曾在小米科技（全球 500 强企业）进行过为期四个月的短期实习，曾获“华为杯”研究生数学建模竞赛国家二等奖、“高教社”全国大学生数学建模竞赛国家二等奖、美国大学生数学建模竞赛 Meritorious 奖等多项国家级奖项。

申请人所在课题组东南大学学习机器研究中心，长期针对控制理论、网络通信、机器学习与模式识别展开研究，在智能机器人、无人机、多智能体系统方面具有深厚的理论与实践基础。课题组成员先后承担多项相关纵向科研课题研究。自主无人系统协同控制理论及应用（国家创新群体项目），人在回路的混合增强智能理论与方法（国家“人工智能专项”重大项目），机器人集群的智能协同控制理论与方法（国家自然科学基金委-深圳机器人基础研究中心项目），多智能体事件驱动分布式优化控制（国家自然科学基金重点项目），基于端到端学习的道路配准与匹配算法及其在非 GPS 无人机定位中的应用（国家自然科学基金青年基金项目），事件驱动的非线性网络控制系统自适应优化控制策略研究（国家自然科学基金青年基金项目）等，表面课题组立足学科前沿、面向国家重大需求、重视产业化应用的深技术研究与实验开发能力。

本项目指导老师孙长银教授，中国自动化学会会士（CAA Fellow），国家二级教授，东南大学特聘教授、博士生导师，西安交通大学兼职教授，现任东南大学学习机器研究中心主任、自动化学院学术委员会主任。2008 年入选教育部新世纪优秀人才计划，2011 年获得国家杰出青年科学基金，中国科协九大代表，国家“万人计划”科技创新领军人才，