



南京大學

NANJING UNIVERSITY

RINC



基于深度强化学习的移动机器人导航研究

- 答辩人: 高可攀 MG1833020
- 导 师: 申富饶 教授



目录

CONTENTS

- 1 研究背景
- 2 研究内容
 - 机器人分层强化学习导航
 - 局部强化学习规划迁移
- 3 实际应用
 - 办公室场景下的机器人导航系统
- 4 研究总结



第一部分

Research Background 研究背景



- 机器人导航：给出起点到终点的**可行**路径，控制机器人从起点**尽快**移动至终点，并**智能地避开障碍物**。
- 应用场景：



工业场景



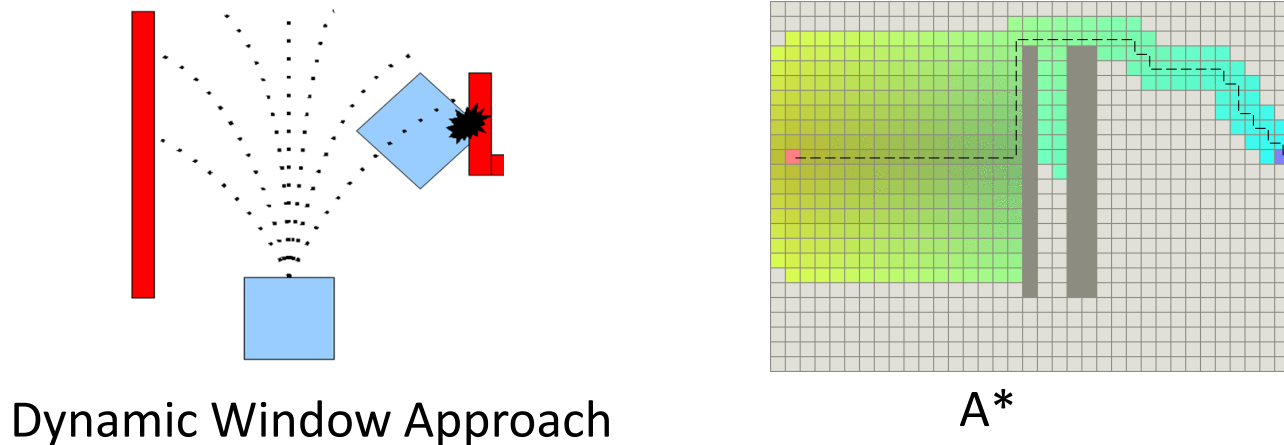
服务场景



家用场景

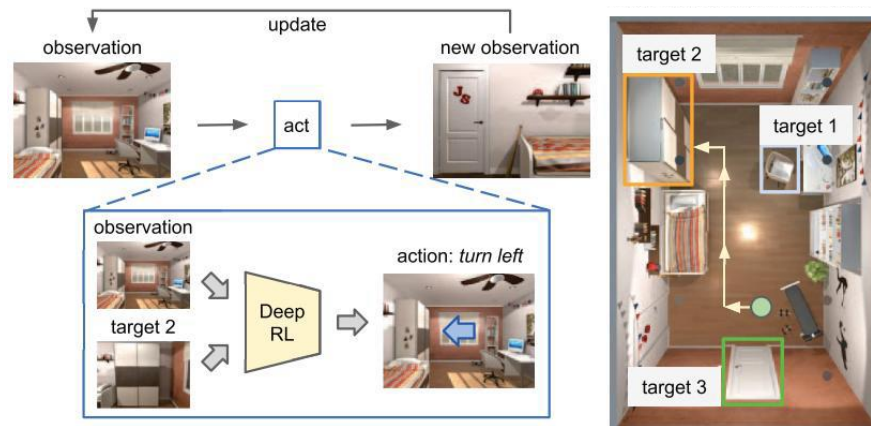
■ 传统方法:

- 依赖人工设计估价函数
- 依赖障碍物地图
- 依赖运动模型



■ 深度学习方法:

- 机动性差
- 适用环境小
- 泛化性能差
- 采样量巨大



Target-driven visual navigation



第二部分

Proposed Methods

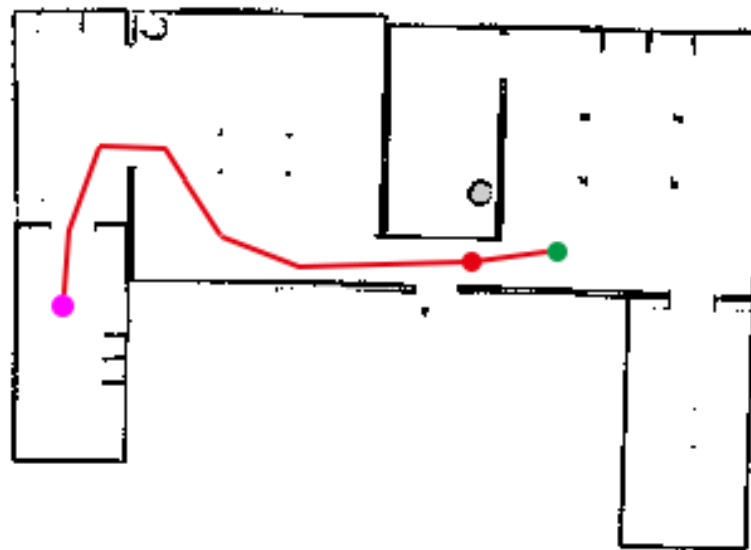
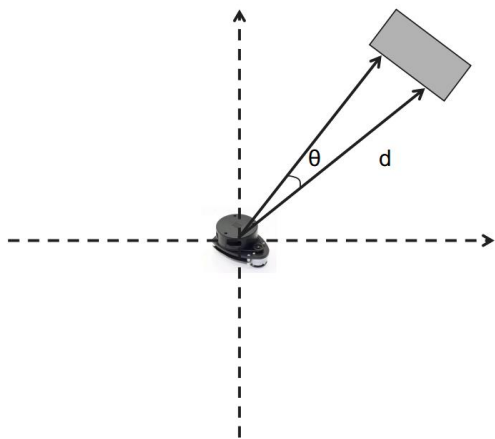
提出方法

- 机器人分层强化学习导航
- 局部强化学习规划迁移



机器人分层强化学习导航

- 局部规划实现传感器视野内的局部避障
- 全局规划找出全局路径，拆分为局部规划任务

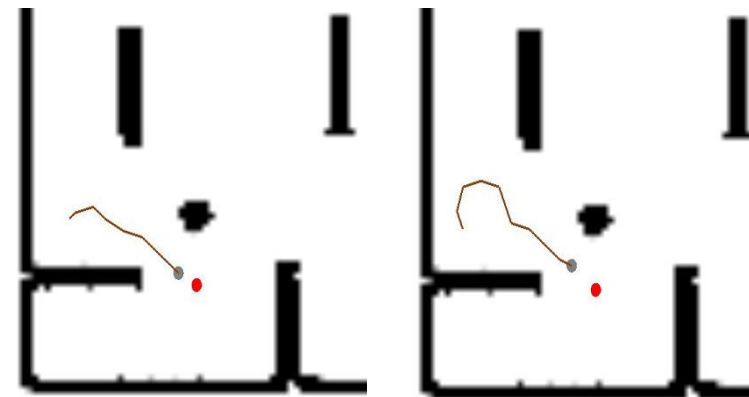




局部强化学习导航

- 传感器采样、相对位姿、运动状态构成状态空间，实现无地图特性
- 连续动作空间提升机动性
- 避免稀疏奖赏，辅助训练

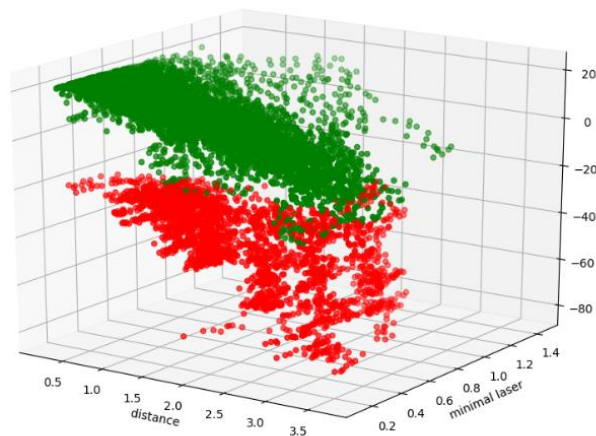
$$reward(P_{now}, P_{goal}) = \begin{cases} R_s & success \\ R_m * ||P_{now} - P_{goal}||_2 & move \\ R_f & fail \end{cases}$$



全局PRM-强化学习规划

- 利用DDPG算法拟合的Q值对PRM路标点间连通性进行快速判定
- 避免路网图稀疏、减少建图耗时

$$Q_{\pi}(s, a) = E_{\pi} \left[\sum_{t=0}^{n-1} \gamma^t r_{t+1} \mid s_0 = s, a_0 = a \right]$$



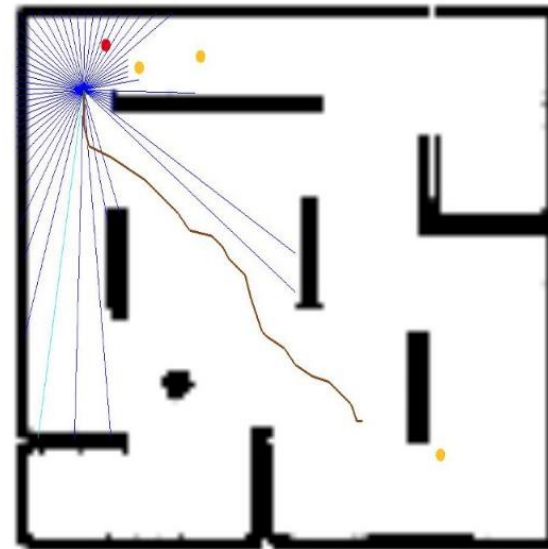


定制化仿真环境构建

- 静态/动态障碍物、传感器、运动模拟
- 可视化界面、OpenAI Gym支持

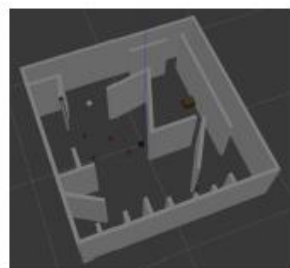
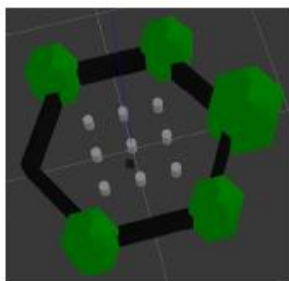
运行场景	随机策略	模型推理	模型训练
无动态障碍物	625x	417x	137x
1个动态障碍物	139x	109x	62x
3个动态障碍物	126x	104x	57x

训练时长：12天->2小时





实验验证



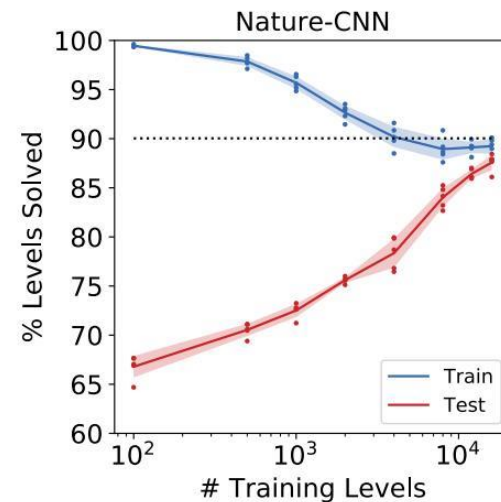
动态障碍物数量		0 个	1 个	3 个
场景 1	RL	99.5%	98.1%	94.1%
	DWA	96.7%	96.2%	93.5%
场景 2	RL	84.6%	80.8%	72.5%
	DWA	65.5%	61.0%	51.8%
场景 3	RL	61.8%	59.6%	56.6%
	DWA	41.3%	39.0%	35.4%

方法	建图耗时	导航成功率	导航平均耗时
直连接结	30 秒	联通 92.5%/不连通 53.0%	74.1s
访问联结	604 分	85.0%	68.0s
值函数联结	641 秒	91.6%	72.8s



局部强化学习规划迁移

- 强化学习泛化性能差



- 复用模型

提升泛化能力—— 扩充MDP数量、增强随机化

策略迁移 —— 引入迁移学习，进行状态空间映射迁移



基于非配对数据集的状态空间映射与改进

■ 循环一致误差

$$\mathcal{L}_{cc_x}(G, F, X, Y) = E_{x \sim X} [\|F(G(x)) - x\|_1]$$

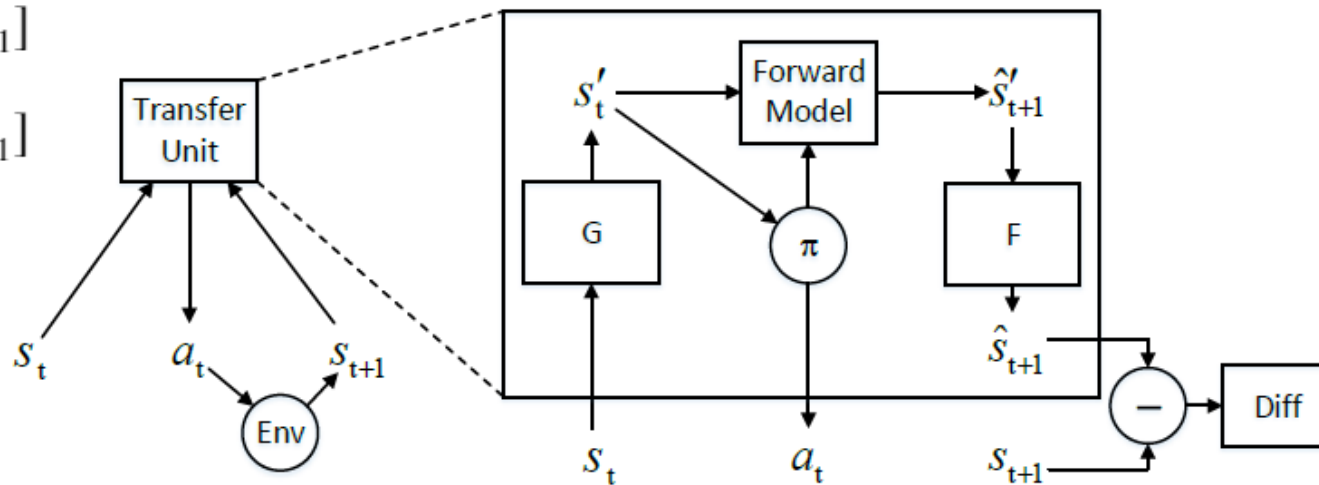
$$\mathcal{L}_{cc_y}(G, F, X, Y) = E_{y \sim Y} [\|G(F(y)) - y\|_1]$$

■ 时序状态预测

$$\hat{s}'_{t+1} = FM(s'_t, \pi)$$

■ 单步状态重构误差

$$\mathcal{L}_{osr}(G, F) = \|F(FM(G(s)|\pi)) - s'_{t+1}\|_1$$

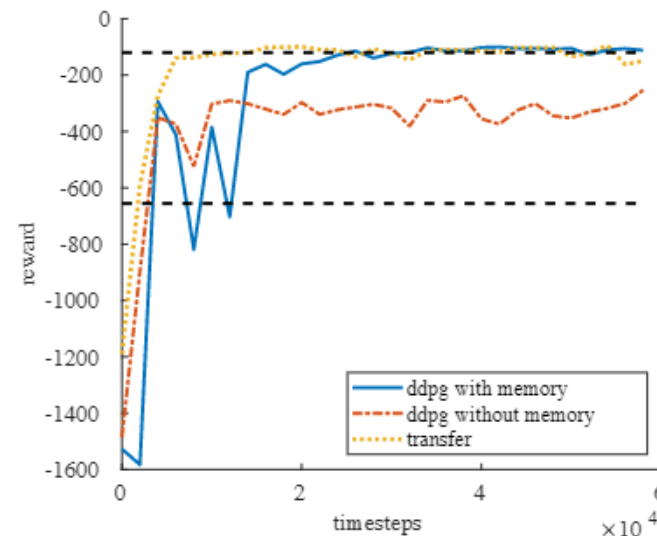




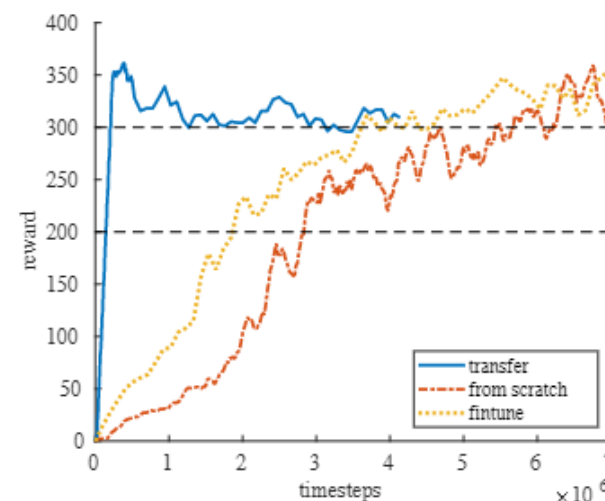
实验验证

- 以少量采样得到相当的效果
- 迁移算法在图像任务上验证

实验场景	策略复用	策略迁移	
	成功率	成功率	采样量
运动参数变更	70.45%	84.7%	32000(1.25%)
激光雷达变更	-	85.1%	14000(0.55%)
仿真环境变更	64.50%	82.2%	8000(0.31%)
虚拟到现实	55.0%	72.5%	6107



Pendulum 迁移



Breakout 灰度迁移



第三部分

Applications 实际应用

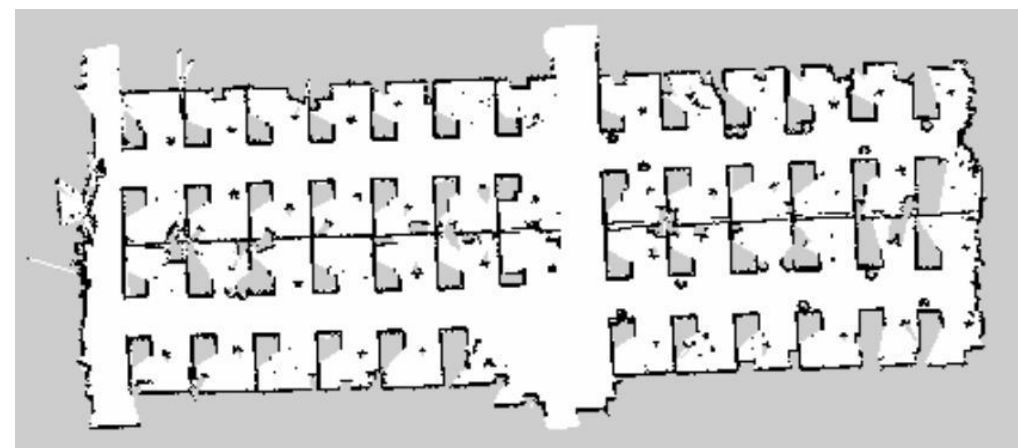
- 办公室场景下的机器人导航系统



机器人导航系统

■ 导航要求

- 静态/动态障碍物避障
- 狭窄过道通行
- 长距离规划
- 状态恢复与重规划

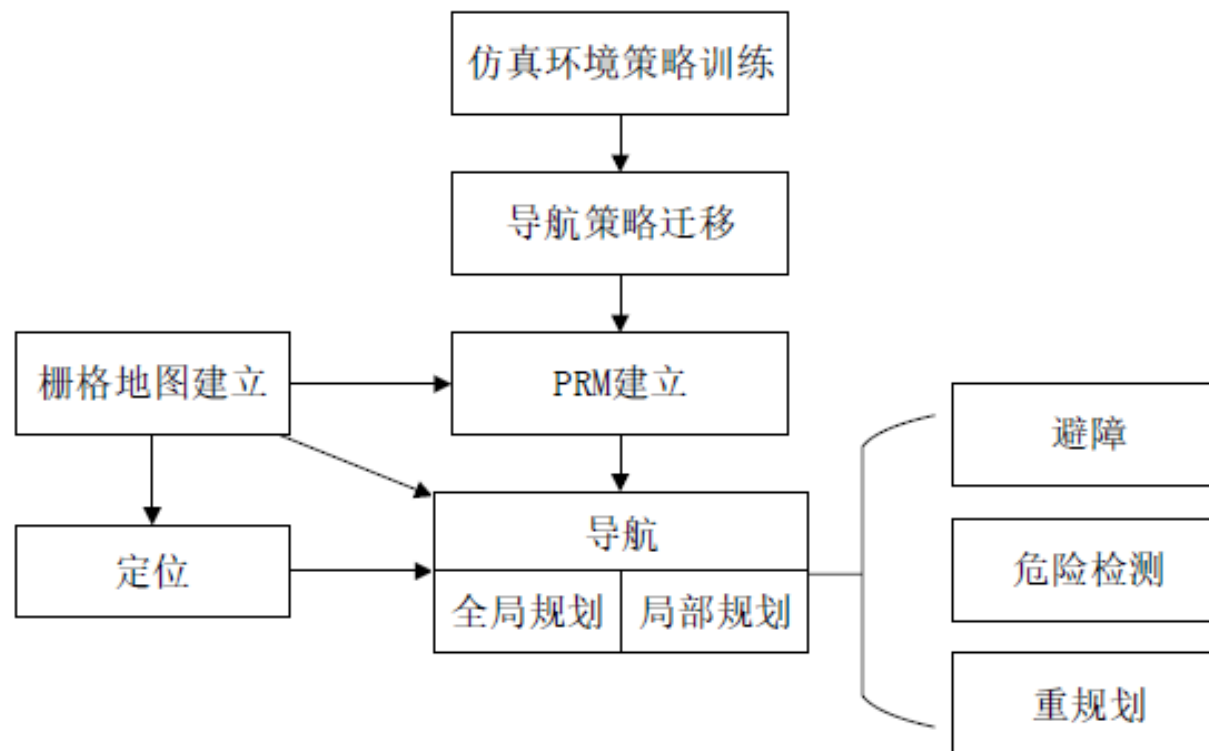


20m*9m



机器人导航系统

- 结合定位、建图组件构建完整的导航系统

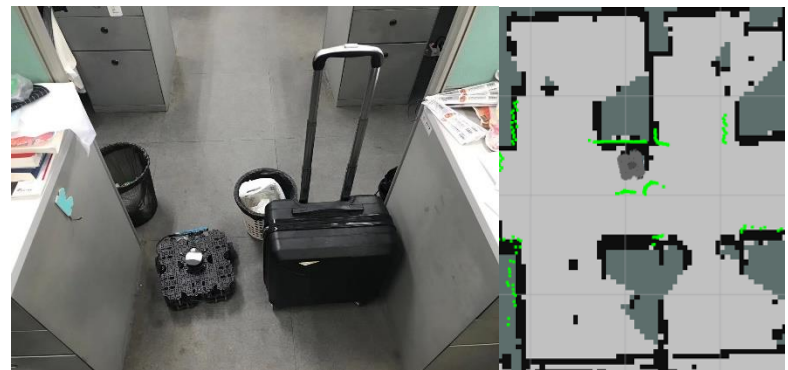
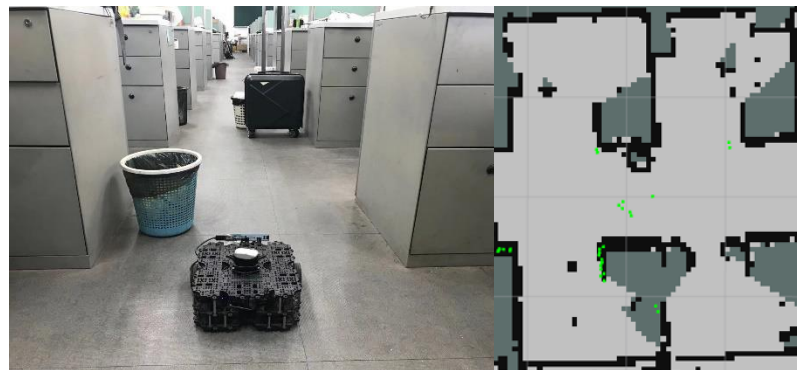
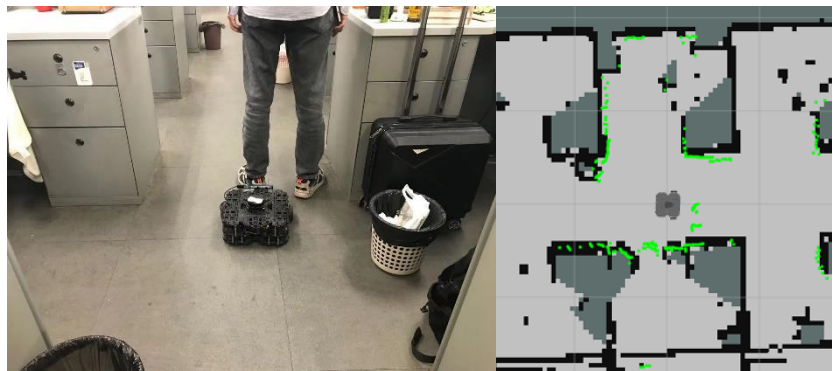




机器人导航局部避障效果

动态避障

狭窄过道通行





机器人导航长距离规划效果



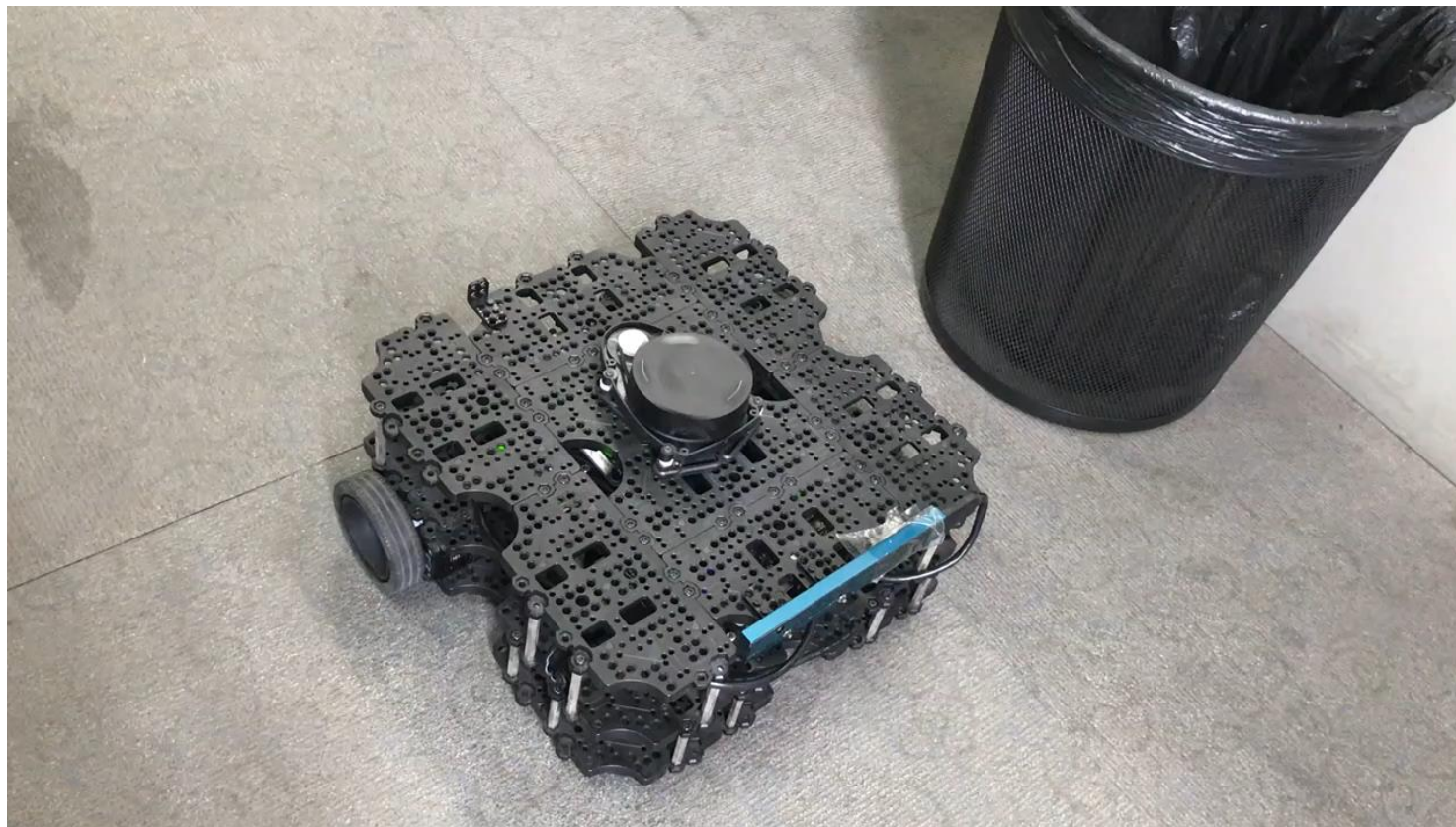
仅局部规划



局部规划+全局规划



机器人导航系统避障展示





第四部分

Summary 研究总结



研究内容总结

- 提出了端到端、无地图的强化学习局部规划算法
- 利用值函数改进PRM算法进行全局规划
- 提出高效的导航策略迁移算法
- 将导航方法成功应用于实际导航系统中



项目

- AGV柔性导航，负责融合定位系统

专利

- 申富饶，**高可攀**，刘小亮，李俊，赵健. “一种融合UWB和LiDAR的室内定位方法”
(202011520518.X)

荣誉奖项

- 南京大学二等学业奖学金



南京大學
NANJING UNIVERSITY



谢谢!

