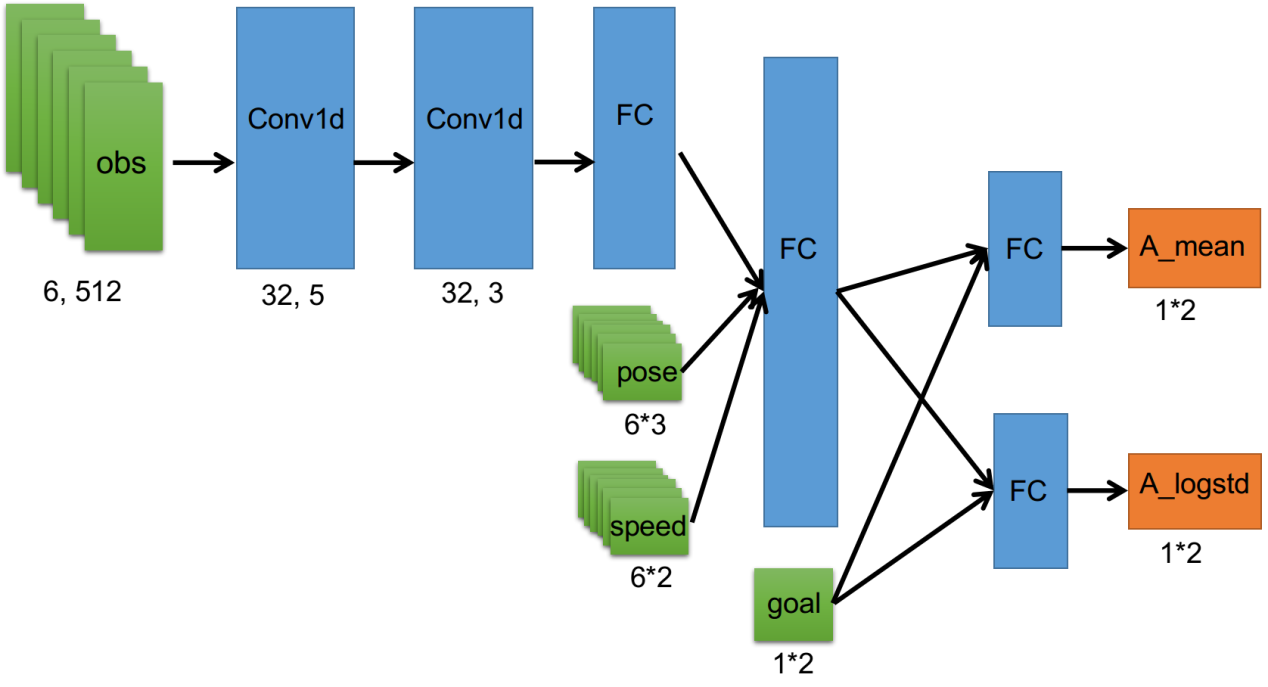
室内复杂场景下的机器人避障导航策略研究

（机器人的高度；真实环境的复杂性要求模型的泛化－－多场景，数据增广；训练速度和样本利用效率－－分布式和off-policy方法）

生活中大对数可以用于护理，导购，教育等用途的机器人都需要进行必要的人机交互，使得机器人需要具有一定的高度，相比于扫地机器人的应用场景，这无疑又对机器人的避障导航功能又提出了新的要求。针对室内的复杂场景下的机器人避障导航任务，普通的单线激光雷达只能获取一个平面的障碍物距离信息，无法探测到完整的障碍物，而多线雷达的成本太高，限制了其应用的推广。本文提出了基于深度相机的障碍物探测方法，相比于激光雷达360度的观测视野，其可观测的视野约８０度左右；**为了解决视野受限问题，本文采用基于深度强化学习的方法，利用长短时记忆网络（ｌｓｔｍ）构建策略网络，利用其时间上的记忆功能，以弥补视野的不足**（为了同时解决静态和动态的场景的避障问题，本文利用连续的多针观测和速度位姿，连续的观测是用于记忆适应静态场景，连续的速度和位姿是为了达到对动态场景的适应性）；为了保证控制的精细和连续，本文利用的深度强化学习算法直接输入连续的状态空间，输出连续的动作空间，这就对处理网络的泛化能力提出了要求；本文利用分布式技术，在在多场景下用多机器人进行蒙特卡洛采样，加快训练速度，提升模型的泛化能力。

其他相关的工作概览

1. 强化学习在机器人避障导航中的应用
2. 强化学习基本原理介绍
3. 深度强化学习的对比和选择（从平衡探索，样本效率等方面着手）
4. 网络结构设计（针对要解决的问题设计相应的结构等，ｌｓｔｍ，ａｔｔｅｎｔｉｏｎ等）



1. Ｒｅｗａｒｄ设计
2. 训练方法（探索利用

为了平衡探索与利用，本文采用SAC(soft actor critic)算法进行训练机器人的避障任务。SAC算法在一定程度上平衡了探索与利用，并且作为off-policy的方法，其样本利用率较高，训练速度较快，在多项强化学习的任务中都取得了state-of-the-art的效果。

为了加速训练的速度，本文采用分布式方式进行训练。设计可以同时运行上百个机器人的场景，静态环境是静态的障碍物，机器人之间是相互的动态障碍物。同时开启与机器人数目相同的进程数，每个进程则负责控制每个机器人与环境的交互采样，收集用于更新策略的数据。

训练的伪代码如下：

If 进程号==0

初始化Ｑ1网络(theta1)，Q2网络(theta2)，policy网络(phi)

初始化目标网络Q1-target=Q1 ,Q2-target=Q2

初始化一个空的样本池R

For 回合数＝1,2,……Ｍ　do

While(没有碰到障碍物) do

进程0收集其他进程机器人的状态

If 进程==０

at = pi(at|st) 按照策略网络进行动作采样

将得到的输出速度分发给其他进程的机器人

s\_t+1 = p(s\_t+1 | st,at)　从环境中得到下一个状态

If 进程==0

R<--R　将新的样本加进样本池

If 进程==0

//更新参数

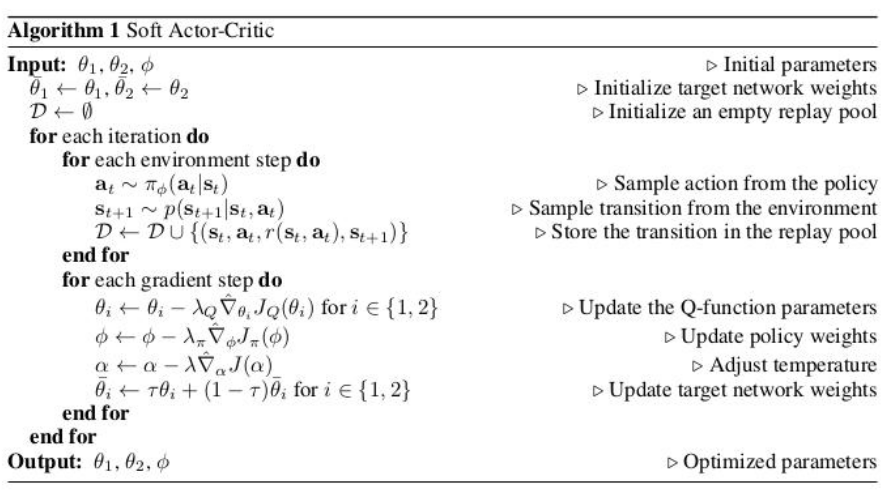
Theta1 , theta2 =

Phi =

Alpha =

Theta\_target =

End for



1. 实验设计和结果（主要展示设计的训练和测试环境，以及避障的效果，ｒｅｗａｒｄ曲线）
2. 实验平台说明

本文为了验证上述方法的可行性，先在仿真环境下设计训练环境进行训练，学到应对各种场景从策略，为了测试模型的泛化能力，并证明模型从仿真到真实的迁移的有效性，则将训练好的模型移植至真实机器人上进行测试。本文的仿真平台，采用的是player/stage环境，真机是在某公司提供的机器人上进行改造实验的。

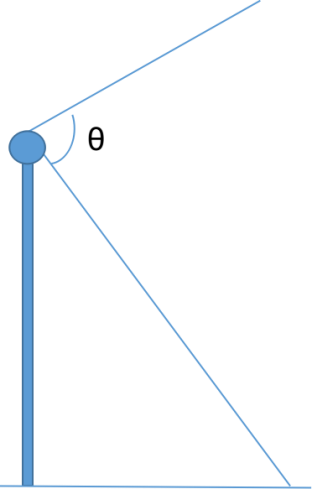
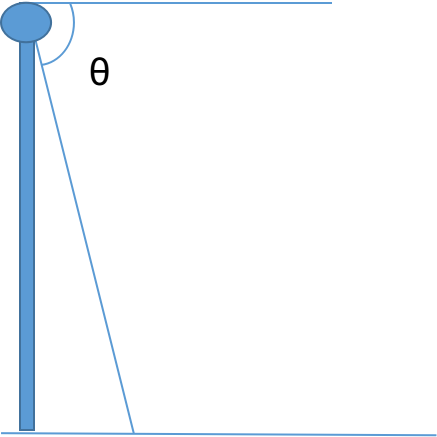
1. 仿真环境（训练场景，测试场景等）
2. 真机实验（包括深度相机数据的获取，真机的实际效果）

本文真机的实验，一是为了验证仿真环境下训练的避障的策略足够可靠，具备很强的泛化能力，可以无缝的迁移至真实机器人上；二是为了提现本文提出的用深度相机作为传感器的解决方案，用于具备一定高度的机器人的可行性和可靠性。

选择激光雷达进行实验的优势在于，首先，单线激光雷达的观测范围仅限于一个平面，无法探测到如桌子凳子这样障碍物的全部信息，不可避免的会导致避障失败，而深度相机可以探测到视野中的所有障碍物信息，对障碍物信息的提取更为准确；其次，多线激光雷达的成本过高，目前还不利于推广，而深度相机与单线激光雷达的成本则相当。

**相机安装与标定**

本文在进行仿真时，利用的传感器观测数据是一个平面内的激光雷达的扫描数据，为了在具有一定高度的真机上做模型的迁移，则需要对深度相机的点云数据进行地面平面上的投影，以到达平面激光雷达的数据同样的效果，保证可移植性，同时也保留了全部的障碍物信息。

考虑到相机的安装方式会影响最后得到的观测向量，如图ａ为正常平视的观测的视角，会存在较大的视觉盲区，导致检测脚底的障碍物不充分，而图ｂ，则将相机向下倾斜一定角度，会有助于减小这样的视觉盲区。

为了在安装时有一定的倾角的相机下得到准确的反应障碍物信息的向量，则需要考虑到将图ｂ时相机坐标系下得到的点云数据变换到机器人的坐标系下。这就要求对相机安装的倾角做较为准确的标定。标定的方法如下：

障碍物检测算法的基本步骤:

1. 获取深度相机的点云数据（ｗ×ｈ）
2. 标定相机安装的倾角

2，根据几何关系，求出相机坐标系下的平面方程

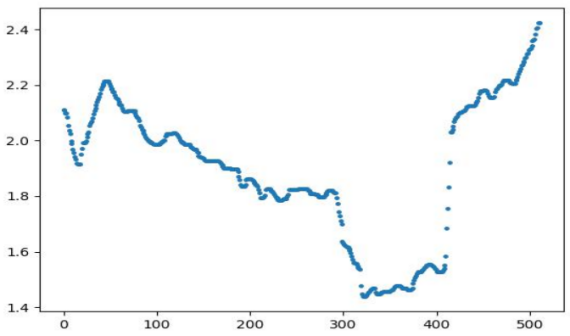
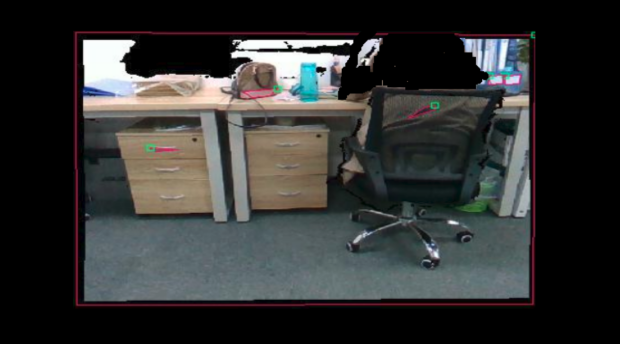
3，遍历点云数据（ｗ×ｈ），计算每个点到地面平面的距离h

4，设置障碍物的高度阈值h1，若h>h1，则判定该点存在障碍物

5，将所有符合障碍物条件的点，变换到机器人的坐标系下

6，将每列的ｈ个点中，距离机器人最近的点作为障碍物，则可得到ｗ维向量，再在ｗ中均匀抽样得到与模型输入格式一致的512维的向量，即可表示所观测到的障碍物的信息。

如图所示，是用上述方法得到的观测向量示意。由于深度相机自带的测量误差，数据会带有一定　波动性，但最终的实际效果表明，真机的效果并不受影响。



１，针对有一定高度的机器人的感知问题（简化了二维信息，方便训练；实测效果可靠）

２，分析了强化学习算法应用于移动机器人动态避障上的优势；

３，训练速度，采样效率，探索利用