基于强化学习的医院运输机器人自主避障研究

张乐宇 黄一展
广州科技贸易职业学院 广东广州 511400

摘 要: 针对医院室内运输场景中人员密集、环境动态变化、障碍物类型复杂等问题,传统运输机器人避障技术存在适应性差、决策滞后等局限。本文提出一种基于改进深度 Q 网络（DQN）的医院运输机器人自主避障方法，通过构建融合医院场景特征的强化学习框架，优化状态空间、动作空间与奖励函数，实现机器人在复杂环境下的实时避障决策。首先，基于医院实地调研数据与 ROS 搭建高保真场景仿真平台；其次，引入动态权重因子改进 DQN 算法，提升模型对突发障碍物的响应速度；最后，通过仿真测试与实地实验验证算法性能。结果表明，该方法在医院常见场景下的避障成功率达 96.7%，平均避障决策时间缩短至 0.19s，优于传统 A ★ 算法与基础 DQN 算法，可为医院运输机器人的智能化应用提供技术支撑。

关键词： 强化学习；医院运输机器人；自主避障；深度 Q 网络

前言

随着智慧医疗的快速发展，医院运输机器人作为医疗物资流转的核心设备，已广泛应用于药品配送、标本转运、器械运输等场景。多数运输机器人不具备智能自主避障功能，躲避障碍物需依赖人工操作，因此需要提高机器人在医院物流的运营效率 [1]。医院环境的特殊性是避障技术的核心挑战：一方面，人员流动具有随机性（如患者突然横穿通道、医护人员推车快速通行）；另一方面，环境存在动态障碍物（如临时摆放的医疗设备、清洁车作业），且部分通道狭窄（如走廊宽度仅 1.2–1.5m），对机器人的实时决策能力及自动化程度提出极高要求 [2]。传统避障技术难以适配医院场景需求，基于红外、超声波传感器的方法易受光线、电磁干扰，在人群密集区域误判率超 20%[2]；基于预设地图的路径规划算法（如 A*、Dijkstra）面对突发障碍物时需重新规划路径，决策滞后时间达 0.5–1s [3]；现有强化学习应用多聚焦于工业或家居场景，未考虑医院的医疗优先级（如急救推车需优先避让）与环境动态特征，适配性不足。

在移动机器人避障领域，宋海荦 [4] 提出多模态 DRL 避障方法，采用了一种双线性融合模块能够充分捕获不同模态数据间的互补信息从而提升避障性能，在一定程度上提升了移动机器人的避障性能。鲁志等 [5] 提出了一种融合改进 A* 算法与 DWA 算法，该方法在 A* 算法中引入全局障碍物占比，在 DWA 算法中加入目标点代价子函数，从而实现移动机器人的动态避障。曾俊杰等 [6] 提出一种基于内在动机强化学习方法，结合内在动机取向函数的奖励权重和外部动机奖励函数的奖励权重计算其综合奖赏值，将此奖赏值作为强化学习算法奖励机制，通过不断学习和训练实现运输机器人自主避障。DeepMind 团队的 Mnih 等 [7] 提出了深度 Q 学习算法（Deep Q-learning Network, DQN），使用卷积神经网络处理高维信息输入，使用全连接层拟合状态价值函数，并提出了经验回放池技术，提高了数据的利用效率，使用目标网络技术稳定训练过程，使得基于值函数的深度强化学习逐渐成为研究热点，并且许多优质改进算法被提出，从改善值函数的过估计问题、改进采样方式和利用值函数的分布等多方面改进了 DQN[8–10]。综上，可针基于强化学习算法对环境动态性、医疗优先级等特征进行专项优化，形成适配医院场景的避障方案。

1 医院运输机器人避障系统总体设计

本文设计的自主避障系统采用 “感知 – 决策 – 执行” 三层架构，如图 1 所示。

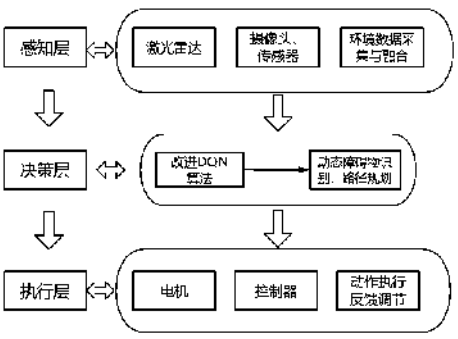

图 1 避障系统总体构架

1.1 感知层

激光雷达：采集 10m 范围内障碍物距离与轮廓数据，扫描频率 25Hz，精度 ±5mm；

高清摄像头：识别障碍物类型（如人、推车、设备），帧率 30fps；

红外传感器：辅助检测近距离障碍物（0.1–0.8m），避免激光雷达盲区遗漏。

传感器数据经 ROS 节点融合后，输出 “障碍物位置 – 类型 – 移动速度” 的结构化信息，为决策层提供输入。

1.2 决策层

核心为改进 DQN 强化学习模型，接收感知层数据后，完成障碍物优先级排序、避障路径规划与动作决策，输出 “前进 / 后退 / 转向角度 / 速度” 指令。

1.3 执行层

由轮式驱动模块（两个直流电机 + 差速控制）与运动控制器（STM32F407）组成，接收决策层指令后驱动机器人执行动作，并通过编码器实时反馈运动状态，形成闭环控制。

2 医院场景特征建模

为提升算法适配性，需先明确医院运输场景的核心特征，模拟选取 3 家医院，采集早高峰、午间、晚间三个时段数据，提炼关键特征如表 1：

表 1 医院运输场景特征

特征类型	具体表现	对避障的影响
人员流动	医护人员推速度 0.8–1.2m/s，患者步行速度 0.4–0.6m/s，突发横穿概率 15%	需区分人员类型，优先避让快速移动目标
障碍物类型	静态障碍（病床、器械柜）占比 60%，动态障碍（清洁车、急救推车）占比 40%	动态障碍需实时跟踪，急救推车优先级最高
空间约束	走廊宽度 1.2–1.5m，电梯口、护士站区域拥堵概率 30%	路径规划需预留 0.3m 安全距离，避免拥堵
环境干扰	电磁设备（MRI、CT 等）导致局部传感器信号衰减，光亮度变化范围 50–800lux	需增强数据抗干扰能力，避免误判

基于上述特征，构建医院场景的状态空间模型，将 “机器人位置、障碍物信息、环境干扰系数” 作为核心状态变量，确保算法能精准捕捉环境动态变化。

3 基于改进 DQN 的避障算法设计

3.1 传统 DQN 算法原理

DQN（Deep Q–Network）通过深度神经网络近似 Q 函数，利用经验回放（Experience Replay）与目标网络（Target Network）解决传统 Q–learning 算法在高维状态空间下的收敛性问题 [8]。其核心公式为：

$$Q(s,a) \leftarrow Q(s,a)+\alpha\,[r+\gamma \max_{a'} Q'(s',a')-Q(s,a)] \quad (1)$$

式中：s 为当前状态，a 为执行动作，r 为奖励值，γ 为折扣因子，α 为学习率，Q′ 为目标网络输出。

但传统 DQN 在医院场景中存在局限：一是奖励函数未考虑障碍物优先级，对急救推车与普通人员无差别处理；二是状态空间未纳入环境动态特征，导致模型对突发障碍响应滞后。

3.2 DQN 改进算法

3.2.1 状态空间优化

扩展传统 DQN 的状态向量，构建包含医院场景特征的高维状态空间：

$$S = [x_{robot}, y_{robot}, \theta_{robot}, v_{robot}, \{x_i, y_i, v_i, t_i, p_i\}_{i=1..n}, \delta] \quad (2)$$

式中：$x_{robot}, y_{robot}, \theta_{robot}, v_{robot}$：机器人的位置、航向角与速度；$x_i, y_i, v_i, t_i, p_i$：第 i 个障碍物的位置、速度、类型（$t_i$=1 为人员，$t_i$=2 为推车，$t_i$=3 为设备）与优先级（$p_i$=3 为急救推车，$p_i$=2 为医护推车，$p_i$=1 为普通障碍）；δ：环境干扰系数（0–1，值越大表示干扰越强）。通过状态空间扩展，模型可精准识别障碍物类型与优先级，提升决策针对性。

3.2.2 奖励函数设计

为引导机器人优先避让高优先级障碍、避免碰撞与拥堵，设计多维度奖励函数：

$$r = r_{collision}+r_{priority}+r_{efficiency}+r_{safety} \quad (3)$$

碰撞惩罚（$r_{collision}$）：若发生碰撞，$r_{collision}=-100$；未碰撞则为 0；

优先级奖励（$r_{priority}$）：避让高优先级障碍时，$r_{priority}= p_i*5$；未避让则为 $-p_i*3$；

效率奖励（$r_{efficiency}$）：若机器人沿目标方向移动，$r_{efficiency}= v_{robot}* 0.1$；偏离方向则为 −0.5；

安全奖励（r_{safety}）：若与障碍物距离 > 0.3m 安全阈值，$r_{safety}=2$；否则为 −1。

通过多维度奖励机制，模型可在 “安全”与 “效率”之间实现平衡，符合医院运输需求。

3.2.3 动态权重因子引入

针对医院环境动态变化的特点，在 Q 函数更新中引入动态权重因子 ω，根据障碍物移动速度与环境干扰系数调整学习率：

$$\omega = 0.5 + 0.3 \times \max(vi) + 0.2 \times \delta \quad (4)$$

$$\alpha\ \text{dynamic} = \alpha * \omega \quad (5)$$

当环境中存在快速移动障碍（如急救推车）或干扰较强时，ω 增大，α dynamic 提升，加快模型学习速度，缩短决策滞后时间。

3.3 算法训练流程

改进 DQN 算法的训练流程如图 2 所示，具体步骤如下。

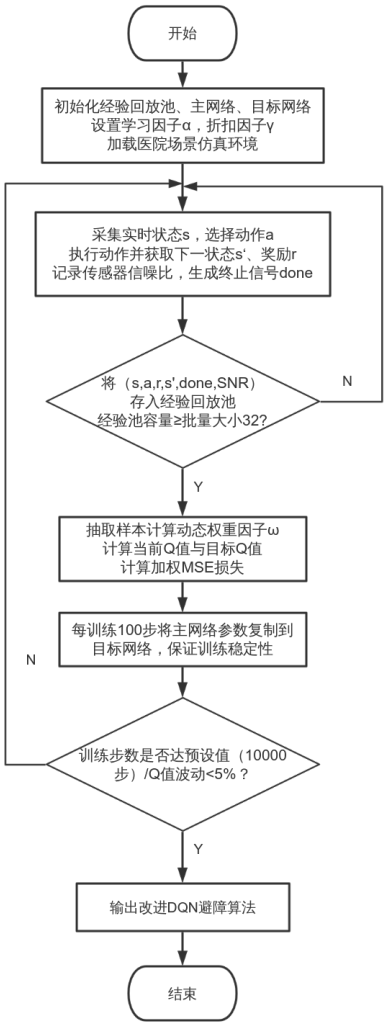

图 2 改进 DQN 算法训练流程

1. 初始化经验回放池（容量 10000）、主网络（Q–Network）与目标网络（Target Q–Network），设置学习率 α =0.001，折扣因子 γ =0.9；2. 基于 ROS 搭建医院场景仿真环境（包含走廊、电梯口、仓库、护士站等典型区域），随机生成障碍物；3. 机器人在仿真环境中执行动作，采集状态 s、动作 a、奖励 r、下一状态 s’，存入经验回放池； 4. 当经验回放池数据量达 1000 时，随机抽取 32 条样本进行训练，更新主网络参数；5. 每训练 100 步，将主网络参数复制到目标网络，保证目标网络稳定性；6. 重复步骤 2–5，直至模型收敛（Q 值波动范围 <5%）。

4 结果分析

4.1 仿真实验环境

基于 ROS Noetic 与 Gazebo 11 搭建医院场景仿真平台，场景包含：

核心区域：长 50m、宽 1.5m 的走廊，2 个护士站（面积 10m^2），3 部电梯口；

障碍物设置：随机生成 10–20 个动态障碍物（人员、推车）与 5–8 个静态障碍物（器械柜）；

性能指标：避障成功率、平均决策时间、路径偏差率（实际路径与目标路径的夹角均值）。

4.2 不同算法避障仿真实验结果

选取 2 类典型算法作为对比：

传统路径规划算法：A * 算法（医院场景常用算法）；强化学习算法：标准 DQN 算法（未优化医院场景特征）。

表 2 不同算法下避障仿真实验数据

算法	避障成功率（%）	平均决策时间（s）	路径偏差率（° ）
A * 算法	82.5	0.68	12.3
标准 DQN	90.2	0.35	8.7
改进 DQN	96.7	0.19	3.5

由表 2 可知，本文改进 DQN 算法的避障成功率较 A * 算法提升 15.8 个百分点，平均决策时间缩短 82.4%，路径偏差率降低 8.8%，表明算法在动态环境下的适应性与决策效率显著优于传统方法。

4.3 消融实验结果

为验证各改进模块的有效性，设计消融实验（基于仿真环境），结果如表 3 所示。

表 3 消融实验数据

实验组	改进模块	避障成功率（%）	平均决策时间（s）
1	无改进（标准 DQN）	90.2	0.35
2	仅优化状态空间	93.5	0.32
3	仅设计奖励函数	94.1	0.31
4	仅引入动态权重	92.8	0.2
5	全改进（本文算法）	96.7	0.19

5 结论

（1）本文提出的基于改进 DQN 的医院运输机器人自主避障算法，通过优化状态空间、设计多维度奖励函数、引入动态权重因子，有效优化了医院场景中障碍物动态变化、优先级差异等问题，在测试中表现出良好性能。

（2）实验结果表明，改进算法的避障成功率达 96.7%，平均决策时间 <0.2s，优于传统 A* 算法与标准 DQN 算法，可满足医院运输机器人的实际运行需求。

参考文献：

[1] 刘朝阳 , 程维国 . 多措并举提高机器人在医院物流的运营效率 [J]. 中国物流与采购 ,2025,(10):67-68.

[2] 姜朋 . 基于强化学习的室内移动机器人避障策略研究 [D]. 浙江大学 ,2023.

[3] 李明 , 王强 . 医院运输机器人避障技术现状与展望 [J]. 机器人技术与应用 ,2022,35 (4):45-52.

[4] 宋海荦 . 基于多模态深度强化学习的移动机器人避障方法研究 [D]. 中国科学技术大学 ,2021.

[5] 鲁志 , 刘莹煌 , 张绪坤 , 等 . 融合 A* 与 DWA 算法的移动机器人动态避障研究 [J]. 电子测量技术 ,2025,48(8):34-45.

[6] 曾俊杰 , 秦龙 , 徐浩添 , 等 . 基于内在动机的深度强化学习探索方法综述 [J]. 计算机研究与发展 ,2023,60(10):2359-2382.

[7]Mnih V,,Kavukcuoglu K,Silver D,et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[8]Van Hasselt H,Guez A,Silver D. Deep reinforcement learning with double q learning[C]. Proceedings of the AAAI conference on Artificial Intelligence. 2016, 30: 2094-2100.

[9]Wang Z, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning[C]. International Conference on Machine Learning. PMLR, 2016: 1995-2003.

[10] Schaul T,Quan J,Antonoglou I, et al. Prioritized experience replay[J]. arXiv preprint arXiv: 1511.05952, 2015.

作者简介：张乐宇（1992—），男，汉族，湖北，广州科技贸易职业学院，硕士，助教，主要研究方向：自动化装备设计与制造、嵌入式开发。