

北京理工大学

本科生毕业设计（论文）

基于深度强化学习的自动驾驶避障技术研究

Research on Automatic Driving Obstacle Avoidance Based on  
Deep Reinforcement Learning

学 院:	自动化学院
专 业:	自动化
学生姓名:	黄宸睿
学 号:	1120181506
指导教师:	宋春雷

2022 年 5 月 7 日

## 原创性声明

本人郑重声明：所呈交的毕业设计（论文），是本人在指导老师的指导下独立进行研究所取得的成果。除文中已经注明引用的内容外，本文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

特此申明。

本人签名：

日期：

年

月

日

## 关于使用授权的声明

本人完全了解北京理工大学有关保管、使用毕业设计（论文）的规定，其中包括：①学校有权保管、并向有关部门送交本毕业设计（论文）的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存本毕业设计（论文）；③学校可允许本毕业设计（论文）被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换本毕业设计（论文）；⑤学校可以公布本毕业设计（论文）的全部或部分内容。

本人签名：

日期：

年

月

日

指导老师签名：

日期：

年

月

日

# 基于深度强化学习的自动驾驶避障技术研究

## 摘 要

本文……。

关键词：北京理工大学；本科生；毕业设计（论文）

## **Research on Automatic Driving Obstacle Avoidance Based on Deep Reinforcement Learning**

### **Abstract**

In order to study……

**Key Words: BIT; Undergraduate; Graduation Project (Thesis)**

## 目 录

摘 要 .....	I
Abstract .....	II
第 1 章 绪论 .....	1
1.1 研究背景与意义 .....	1
1.2 国内外研究现状 .....	1
1.3 本文主要研究内容及章节安排 .....	1
第 2 章 强化学习相关算法理论分析 .....	2
2.1 强化学习算法 .....	2
2.2 Q-learning 算法 .....	2
2.3 DQN 算法 .....	2
2.4 改进的 DQN 算法 .....	2
2.5 本章小结 .....	2
第 3 章 基于 DQN 算法的自动驾驶避障算法设计 .....	3
3.1 仿真环境 .....	3
3.1.1 Highway-Env .....	3
3.1.2 Metadrive .....	3
3.2 基于 DQN 算法的控制策略设计 .....	3
3.2.1 网络结构设计 .....	3
3.3 基于 DQN 算法的控制器设计 .....	3
第 4 章 基于 DQN 算法的自动驾驶避障策略测试 .....	4
4.1 Highway-Env .....	4
4.2 Metadrive .....	4
4.3 改进的网络结构 .....	4
4.4 结果对比 .....	4
结 论 .....	5
参考文献 .....	6
附 录 .....	8
附录 A L <sup>A</sup> T <sub>E</sub> X 环境的安装 .....	8
附录 B B <sup>I</sup> Thesis 使用说明 .....	8
致 谢 .....	9

## 第 1 章 绪论

### 1.1 研究背景与意义

自动驾驶系统的决策模块需要先进的决策算法保证安全性、智能性、有效性。目前传统算法的解决思路是以价格昂贵的激光雷达作为主要传感器，依靠人工设计的算法从复杂环境中提取关键信息，根据这些信息进行决策和判断。该算法缺乏一定的泛化能力，不具备应有的智能性和通用性。深度强化学习的出现有效地改善了传统算法泛化性不足的问题，这给智能驾驶领域带来新的思路。

目前，强化学习在策略选择的理论和算法方面已经取得了很大的进步，然而直接从高维感知输入（如图像、语音等）中提取特征，学习最优策略，对强化学习来说依然是一个挑战。

深度强化学习 (Deep Reinforcement Learning, DRL) 结合了深度神经网络和强化学习的优势，可以用于解决智能体在复杂高维状态空间中的感知决策问题<sup>[1][2]</sup>。2016 年，基于深度强化学习和蒙特卡洛树搜索的 AlphaGo 击败了人类顶尖职业棋手，引起了全世界的关注<sup>[3]</sup>。2017 年，DeepMind 在《Nature》上公布了最新版 AlphaGo 论文，介绍了更强的围棋人工智能：AlphaGo Zero。它不需要人类专家知识，只使用纯粹的深度强化学习技术和蒙特卡罗树搜索，经过 3 天自我对弈就以 100 比 0 击败了上一版本的 AlphaGo。AlphaGo Zero 证明了深度强化学习的强大能力，也必将推动以深度强化学习为代表的人工智能领域的进一步发展。基于深度强化学习在棋局与游戏上的成功，最近的研究大多注重于深度强化学习在各个领域中的扩展与应用。

### 1.2 国内外研究现状

### 1.3 本文主要研究内容及章节安排

## 第 2 章 强化学习相关算法理论分析

### 2.1 强化学习算法

强化学习 (Reinforcement Learning, RL) 通过与环境交互, 学习状态到行为的映射关系。如图2-1所示, 在一个离散时间序列  $t = 0, 1, 2, \dots$  中, 智能体需要完成某项任务。在每一个时间  $t$ , 智能体都能从环境中接受一个状态  $S_t$ , 并通过动作  $a_t$  与环境继续交互, 环境会产生新的状态  $S_{t+1}$ , 同时给出一个立即回报  $r_{t+1}$ 。如此循环下去, 智能体与环境不断地交互, 从而产生更多数据 (状态和回报), 并利用新的数据进一步改善自身的行为。

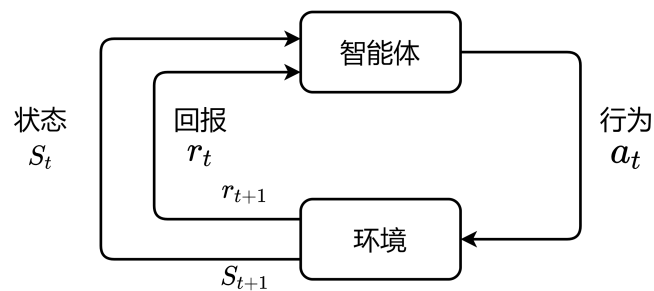


图 2-1 强化学习原理

### 2.2 Q-learning 算法

### 2.3 DQN 算法

### 2.4 改进的 DQN 算法

### 2.5 本章小结

## 第3章 基于 DQN 算法的自动驾驶避障算法设计

### 3.1 仿真环境

#### 3.1.1 Highway-Env

Highway-Env 的状态值设计

111

Highway-Env 的动作值设计

Highway-Env 的奖惩值函数设计

#### 3.1.2 Metadrive

Metadrive 的状态值设计

Metadrive 的动作值设计

Metadrive 的奖惩值函数设计

### 3.2 基于 DQN 算法的控制策略设计

#### 3.2.1 网络结构设计

### 3.3 基于 DQN 算法的控制器设计



## 第 4 章 基于 DQN 算法的自动驾驶避障策略测试

### 4.1 Highway-Env

### 4.2 Metadrive

### 4.3 改进的网络结构

### 4.4 结果对比

## 结 论

本文结论……。<sup>[4]</sup>

结论作为毕业设计（论文）正文的最后部分单独排写，但不加章号。结论是对整个论文主要结果的总结。在结论中应明确指出本研究的创新点，对其应用前景和社会、经济价值等加以预测和评价，并指出今后进一步在本研究方向进行研究工作的展望与设想。结论部分的撰写应简明扼要，突出创新性。阅后删除此段。

结论正文样式与文章正文相同：宋体、小四；行距：22 磅；间距段前段后均为 0 行。阅后删除此段。

## 参考文献

### 参考文献书写规范

参考国家标准《信息与文献参考文献著录规则》【GB/T 7714—2015】，参考文献书写规范如下：

#### 1. 文献类型和标识代码

普通图书：M      会议录：C      汇编：G      报纸：N

期刊：J      学位论文：D      报告：R      标准：S

专利：P      数据库：DB      计算机程序：CP      电子公告：EB

档案：A      舆图：CM      数据集：DS      其他：Z

#### 2. 不同类别文献书写规范要求

##### 期刊

[序号] 主要责任者. 文献题名 [J]. 刊名, 出版年份, 卷号 (期号): 起止页码.

[1] 唐振韬, 邵坤, 赵冬斌, 等. 深度强化学习进展: 从 AlphaGo 到 AlphaGo Zero[J]. 控制理论与应用, 2017, 34(12): 18.

[2] Li Y. Deep Reinforcement Learning: An Overview[J]., 2017.

[3] Babbar S. Review - Mastering the game of Go with deep neural networks and tree search[J]., 2017.

##### 普通图书

[序号] 主要责任者. 文献题名 [M]. 出版地: 出版者, 出版年. 起止页码. [5]

[4] 李成智, 李小宁, 田大山. 飞行之梦: 航空航天发展史概论[M]. 北京: 北京航空航天大学, 2004.

[5] Raymer, Daniel P. Aircraft design: A Conceptual Approach[M]. Reston, Virginia: American Institute of Aeronautics, 1992.

##### 会议论文集

[序号] 析出责任者. 析出题名 [A]. 见 (英文用 In): 主编. 论文集名 [C]. (供选择项: 会议名, 会址, 开会年) 出版地: 出版者, 出版年. 起止页码. [6]

[6] 孙品一. 高校学报编辑工作现代化特征[C]//张为民. 中国高等学校自然科学学报研究会. 科技编辑学论文集 (2). 北京: 北京师范大学出版社, 1998: 10-22.

##### 专著中析出的文献

[序号] 析出责任者. 析出题名 [A]. 见 (英文用 In): 专著责任者. 书名 [M]. 出版地: 出版者, 出版年. 起止页码. [7]

[7] 罗云. 安全科学理论体系的发展及趋势探讨[M]//白春华, 何学秋, 吴宗之. 21 世纪安全科学与技术的发展趋势. 北京: 科学出版社, 2000: 1-5.

##### 学位论文

[序号] 主要责任者. 文献题名 [D]. 保存地: 保存单位, 年份. [8][9]

[8] 张和生. 嵌入式单片机系统设计[D]. 北京: 北京理工大学, 1998.

[9] Sobieski I P. Multidisciplinary Design Using Collaborative Optimization[D]. United States – California: Stanford University, 1998.

### 报告

[序号] 主要责任者. 文献题名 [R]. 报告地: 报告会主办单位, 年份. [10][11]

[10] 冯西桥. 核反应堆压力容器的 LBB 分析[R]. 北京: 清华大学核能技术设计研究院, 1997.

[11] Sobieszczanski-Sobieski J. Optimization by Decomposition: A Step from Hierarchic to Non-Hierarchic Systems[R]. NASA CP-3031, 1989.

### 专利文献

[序号] 专利所有者. 专利题名 [P]. 专利国别: 专利号, 发布日期. [12]

[12] 姜锡洲. 一种温热外敷药制备方案: 88105607[P]. 中国. 1989-07-26.

### 国际、国家标准

[序号] 标准代号. 标准名称 [S]. 出版地: 出版者, 出版年. [13]

[13] GB/T 16159—1996. 汉语拼音正词法基本规则[S]. 北京: 中国标准出版社, 1996.

### 报纸文章

[序号] 主要责任者. 文献题名 [N]. 报纸名, 出版年, 月 (日): 版次. [14]

[14] 谢希德. 创造学习的思路[N]. 人民日报, 1998-12-25(10).

### 电子文献

[序号] 主要责任者. 电子文献题名 [文献类型/载体类型]. 电子文献的出版或可获得地址 (电子文献地址用文字表述), 发表或更新日期/引用日期 (任选). [15]

[15] 姚伯元. 毕业设计 (论文) 规范化管理与培养学生综合素质[EB/OL]. 中国高等教育网教学研究. [2013-03-26]. <http://www.cnnic.net.cn/hlwfzyj/hlwxyzbg/201201/P020120709345264469680>.

关于参考文献的未尽事项可参考国家标准《信息与文献参考文献著录规则》(GB/T 7714—2015)

## 附 录

附录相关内容...

### 附录 A L<sup>A</sup>T<sub>E</sub>X 环境的安装

L<sup>A</sup>T<sub>E</sub>X 环境的安装。

### 附录 B B<sup>I</sup>Thesis 使用说明

B<sup>I</sup>Thesis 使用说明。

附录是毕业设计（论文）主体的补充项目，为了体现整篇文章的完整性，写入正文又可能有损于论文的条理性、逻辑性和精炼性，这些材料可以写入附录段，但对于每一篇文章并不是必须的。附录依次用大写正体英文字母 A、B、C……编序号，如附录 A、附录 B。阅后删除此段。

附录正文样式与文章正文相同：宋体、小四；行距：22 磅；间距段前段后均为 0 行。阅后删除此段。

## 致 谢

值此论文完成之际，首先向我的导师……

致谢正文样式与文章正文相同：宋体、小四；行距：22 磅；间距段前段后均为 0 行。阅后删除此段。