

学校代码: 10385

分类号: \_\_\_\_\_

研究生学号: 20013080034

密 级: \_\_\_\_\_



华侨大学  
HUAQIAO UNIVERSITY

# 硕士学位论文

低速履带式挖掘机自动驾驶端到端决策算法研究  
**Research on End-to-End Decision Algorithm for Automatic Driving  
of Low-Speed Crawler Excavator**

作者姓名: 马荣华

指导教师: 陈其怀 副教授

学 科: 机械工程

研究方向: 智能电液控制与节能技术

所在学院: 机电及自动化学院

论文提交日期: 二〇二三年五月二十七日

## 摘 要

随着国家战略的转移和行业需求的上升,智能化、无人化正逐渐成为工程机械行业的主要发展趋势之一。工程机械电动化、线控底盘技术的成熟,为实现智能化控制奠定了技术基础;人工智能、5G技术、物联网、大数据、云计算等高新技术为工程机械开展智能化提供了技术支持。目前智能化、无人化技术研究主要集中于汽车领域,对于工程机械的智能化研究相对较少,工程机械作业环境复杂且工况众多,常作业于矿山、港口、山地等崎岖的非结构化道路,施工困难,人力成本巨大,人身安全有着极大的危险,工程机械领域自身迫切需要开展智能化。

本文对工程机械智能化的研究背景及意义进行了系统论述,对智能车和工程机械自动驾驶技术整体国内外研究现状进行了详细阐述,对自动驾驶决策技术研究现状进行了分析。对比了目前自动驾驶主流传感器,针对工程机械自身特点,以履带式挖掘机为例,提出了基于双目图像的多模态决策网络、基于激光雷达点云的决策网络,在数据集上进行对比分析,并开展实车测试,实现了履带式挖掘机特定场景的自动驾驶行走功能。

论文基于双目摄像头,以轻量级语义分割模型 **BiseNet** 为基础框架,构建语义分割信息提取子网络、**RGB** 图像信息提取子网络,融合图像信息的同时尽量保证网络运行速度,同时嵌入坐标注意力机制优化决策网络性能,保证精度。在开源数据集和真实场景数据集进行测试,搭建的模型在最初的基本模型之上各项指标都有较大的提升,基本满足对于履带式挖掘机左右履带控制信号预测的需求。

论文基于激光雷达,搭建自定义激光雷达点云决策网络,并嵌入经典残差连接操作提升网络精度。基于开源数据集和真实场景数据集进行测试,并与基于双目图像的决策网络进行对比。试验结果表明以点云作为输入的端到端决策网络表现尚可,但各项指标均比不上双目图像决策网络。

为验证所提网络在工程机械真实场景下的性能,实现履带式挖掘机自动驾驶行走,论文选择所提更优的双目图像决策网络,利用现有试验条件开展了实车测试。选择工程机械典型工况设计试验方案,搭建相关试验平台,并进行必要的数据采集。实车试验证明,本文所提的双目图像决策网络能够在非结构化道路较

好的完成履带式挖掘机的自动行走，并进行自主避障，对搭建工程机械智能化系统具有较大的参考价值。

**关键词：**深度学习；自动驾驶；挖掘机；端到端；决策

## Abstract

With the transfer of national strategy and the rise of industry demand, intelligence and unmanned are gradually becoming one of the main development trends of the construction machinery industry. The maturity of construction machinery electrification and wire-controlled chassis technology has laid a technical foundation for the realization of intelligent control; artificial intelligence, 5G technology, Internet of Things, big data, cloud computing, and other high-tech technologies also provide technical support for the development of intelligent construction machinery. At present, the research on intelligent and unmanned technology is mainly concentrated in the field of automobiles. There is relatively little research on the intelligentization of construction machinery. The operating environment of construction machinery is complex and there are many working conditions. Roads and construction are difficult, labor costs are huge, and personal safety is in great danger. The field of construction machinery itself urgently needs to develop intelligence.

This paper systematically discusses the research background and significance of the intelligentization of construction machinery, elaborates on the overall research status of intelligent vehicles and construction machinery automatic driving technology at home and abroad, and analyzes the research status of automatic driving decision-making technology. Comparing the current mainstream sensors for autonomous driving, according to the characteristics of construction machinery, taking crawler excavators as an example, a multi-modal decision network based on binocular images and a decision network based on LiDAR point clouds are proposed, and compared on the data set Analyzed and carried out real vehicle tests, and realized the automatic driving and walking function of crawler excavators in specific scenarios.

Based on the binocular camera, the paper uses the lightweight semantic segmentation model BiSeNet as the basic framework to construct the semantic segmentation information extraction sub-network and the RGB image information extraction sub-network. While fusing the image information, the network operation

speed is guaranteed as much as possible, and the coordinate attention mechanism is embedded at the same time. Optimize the performance of the decision-making network to ensure accuracy. Tested on open-source datasets and real-scene datasets, the built model has a greater improvement in various indicators than the original basic model, basically meeting the demand for the prediction of the left and right crawler control signals of crawler excavators.

Based on LiDAR, the paper builds a custom LiDAR point cloud decision network and embeds classic residual connection operations to improve network accuracy. Tested on open source datasets and real scene datasets, and compared with the decision network based on binocular images. The experimental results show that the performance of the end-to-end decision network with point cloud as input is acceptable, but the indicators are not as good as the binocular image decision network.

To verify the performance of the proposed network in the real scene of construction machinery and realize the automatic driving of crawler excavators, the paper chooses the proposed better binocular image decision network and uses the existing experimental conditions to carry out real vehicle tests. Select the typical working conditions of construction machinery to design the experimental scheme, build the relevant experimental platform, and collect the necessary data. The actual vehicle test proves that the binocular image decision-making network proposed in this paper can better complete the automatic walking of the crawler excavator on the unstructured road, and perform autonomous obstacle avoidance, which has a great influence on building an intelligent system of construction machinery reference value.

**Key words:** Deep learning; Autonomous driving; Excavator; End-to-end; Decision

# 目 录

第 1 章 绪论.....	1
1.1 课题背景及意义.....	1
1.2 自动驾驶技术整体研究现状 .....	2
1.2.1 智能车自动驾驶技术研究现状.....	2
1.2.2 工程机械自动驾驶技术研究现状.....	5
1.3 自动驾驶决策技术研究现状 .....	9
1.3.1 传统的自动驾驶算法研究现状.....	10
1.3.2 基于学习的自动驾驶决策算法研究现状 .....	11
1.4 课题的提出及研究内容.....	13
1.4.1 课题的提出.....	13
1.4.2 课题研究内容.....	14
1.5 本章小结.....	14
第 2 章 端到端工程机械自动驾驶决策系统方案设计 .....	15
2.1 自动驾驶常用传感器.....	15
2.1.1 摄像头.....	15
2.1.2 雷达.....	16
2.2 端到端自动驾驶系统.....	20
2.2.1 神经网络与自动驾驶.....	20
2.2.2 自动驾驶开源数据集.....	21
2.2.3 自动驾驶开源仿真平台 .....	23
2.3 工程机械决策系统方案设计 .....	25
2.3.1 环境感知传感器选择.....	25
2.3.2 自动驾驶决策方案.....	27
2.4 本章小结.....	27
第 3 章 基于双目图像的工程机械自动驾驶决策 .....	29
3.1 深度神经网络.....	29
3.1.1 卷积神经网络.....	29
3.1.2 注意力机制.....	31
3.2 多模态决策网络架构.....	34
3.2.1 嵌入坐标注意力机制的 ResNet18 网络.....	34
3.2.2 多模态特征融合网络.....	35
3.3 图像决策网络的试验设计与结果分析 .....	37
3.3.1 数据集介绍.....	37
3.3.2 网络训练配置.....	38

3.3.3 基于开源数据集的试验结果分析.....	39
3.3.4 基于非结构化真实场景数据集的试验结果分析 .....	41
3.4 本章小结.....	44
第 4 章 基于激光雷达点云的工程机械自动驾驶决策 .....	45
4.1 点云相关算法介绍.....	45
4.1.1 点云的下采样.....	45
4.1.2 点云神经网络.....	46
4.2 基于激光雷达点云的网络决策架构 .....	48
4.2.1 基于 X-Conv 搭建的自定义点云网络架构.....	48
4.2.2 嵌入残差连接的点云决策网络.....	49
4.3 点云决策网络的试验设计与结果分析 .....	50
4.3.1 基于开源数据集的试验结果分析.....	51
4.3.2 基于非结构化真实场景数据集的试验结果分析 .....	52
4.3.3 点云采样点数对网络性能的影响.....	54
4.4 本章小结.....	54
第 5 章 基于非结构化道路的决策实车试验 .....	57
5.1 试验方案设计.....	57
5.1.1 试验原理.....	57
5.1.2 试验流程.....	58
5.2 实车平台搭建.....	60
5.3 数据采集.....	61
5.4 工程机械决策避障试验.....	63
5.4.1 静态障碍物避障测试.....	63
5.4.2 动态障碍物避障测试.....	65
5.5 本章小结.....	67
第 6 章 总结与展望 .....	69
6.1 课题总结.....	69
6.2 课题的创新性.....	70
6.3 工作展望.....	70
参考文献.....	73

## 第1章 绪论

### 1.1 课题背景及意义

工程机械是综合机械化施工工程所必需的一切机械装备，如土石方施工工程，路面施工与维护，移动式起重装卸作业，以及各类建筑工程所需的机械装备，均称为工程机械。传统工程机械以内燃机为动力，在城市建设、农田水利、露天采矿等工程任务中应用广泛，普遍存在污染物排放高、噪声大、效率低等问题<sup>[1]</sup>；电动式工程机械在攻克上述难题的基础上，推动了工程机械智能化的发展，具有零排放、低噪音、高传动效率等优点。为应对各方面不断上升的需求，智能化、无人化逐渐引领工程机械的主要潮流。

2015年我国提出《中国制造2025》的发展战略，给整个中国各行各业带来了众多理念，工程机械智能制造也得到指引，2021年我国颁布“十四五”规划，提出“推动智能工程机械产业研发的发展和产业化”，“培育先进制造业集群，推动工程机械产业创新发展”。在5G技术、物联网、大数据、云计算等新一代信息技术的推动下，工程机械智能化进程将进一步加速。

工程机械作业环境复杂且工况众多，大部分工程机械运用于矿山、高原、南北极等环境比较恶劣的地形以及地震、泥石流等救灾抢险的恶劣工况中。复杂的地形和工况对驾驶员的身体各方面的素质要求非常高，在该环境下进行工作非常危险，随时可能给驾驶员生命造成危险。甚至有些极端环境，如高浓瓦斯、高辐射或超高温的环境，驾驶无法进入，很难进行正常的工作。部分专家提出远程遥控的概念，但受操控距离以及远距离操作信号不好等因素的制约，在部分环境无法正常进行，极大的影响工作效率；此外远程遥控所需的硬件设备以及能达到的效果也不能很好的满足要求。

综上所述，随着国家战略、行业需求和工程机械本身作业任务的迫切要求，工程机械的智能化、无人化逐渐成为主要的发展潮流。为应对工程机械不断提高的作业精度及效率要求，适应危险、恶劣的环境，工程机械自动驾驶的研究迫在眉睫，具有重大的意义。



## 1.2 自动驾驶技术整体研究现状

自动驾驶技术在汽车领域已相对比较成熟，众多企业和高校开展研究，已经取得不错的成果，朝着更高级别的自动驾驶逐渐前进。而工程机械自动驾驶起步相对较晚，目前只有一些大的工程机械公司和团队在做研究，并且只是针对像矿山、港口等特定的场景以及压路机、推土机等部分简单的工程机械<sup>[2]</sup>。

### 1.2.1 智能车自动驾驶技术研究现状

国外研究智能汽车较早，可追溯到上世纪 50 年代。1953 年，世界上第一辆自主引导车由美国贝瑞特电子公司研制；到 1966 年，美国斯坦福大学研制出开创自主导航先河的简易轮式机动机器人，继而各类高校和科研院所都纷纷投入对自动驾驶的研究。最具代表性的是美国卡内基梅隆大学于 1984 年开始研发的 NavLab 系列智能车，其中 1995 年研发出的 NavLab-5 自主车辆首次完成了横跨美国 4586 km 的自主行驶测试<sup>[3]</sup>。

为了更深远地发展自动驾驶，激发研究人员的积极性，美国国防高级研究计划局（The Defense Advanced Research Projects Agency, DAPRA）在 2004、2005 和 2007 年连续三年举办了三届智能汽车自动驾驶挑战赛。2004 年比赛在美国 Mojave 沙漠举行，参赛队伍来自五湖四海，没有一支队伍完成比赛。2005 年 5 支车队完赛，斯坦福大学的 Stanley 自动驾驶赛车夺冠<sup>[4]</sup>。2007 年，在废弃的军事基地内举办了第三届挑战赛，要求参赛的自动驾驶车辆在遵守交通规则的同时与其他车辆进行互动和避让，由来自加拿大卡内吉梅隆大学的 “Boss”<sup>[5]</sup> 获得了第一名，斯坦福大学的 “Junior”<sup>[6]</sup> 和弗吉尼亚州大学的 “Odin”<sup>[7]</sup> 分别获得亚军和季军。图 1.1 是三届智能汽车自动驾驶挑战赛的部分获奖车辆。



(a) Stanley



(b) Boss



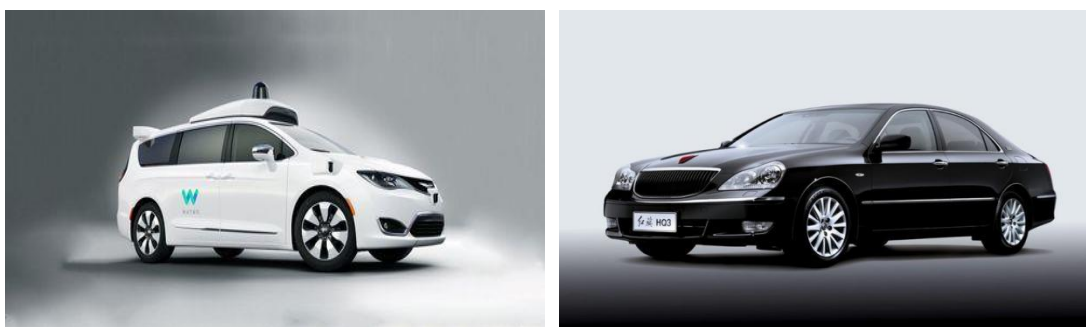
(c) Junior

(d) Odin

图 1.1 智能汽车自动驾驶挑战赛部分获奖车辆

2007 年以后汽车自动驾驶开始进入高速发展阶段，各大车企和互联网公司都开始投入智能车的研发，如谷歌、通用、福特等公司。其中最具代表性的是谷歌公司，2010 年谷歌试验室以丰田汽车为基础开发出谷歌首款自动驾驶汽车，截至 2020 年 1 月，该公司旗下 Waymo 自动驾驶汽车的续航里程达到 3220 万公里，在汽车自动驾驶领域已经独占鳌头，如图 1.2 (a) 所示。

与国外相比，国内智能车研究相对较晚。大约在上世纪 80 年代，国内部分高校和科研机构开始加入到智能车的研究，清华大学从 1986 年开始从事智能车的研究<sup>[8]</sup>，1987 年起国防科技大学从事智能汽车研究。我国第一辆真正意义上的自动驾驶汽车是国防科技大学 2003 年研制成功的，随后至 2011 年，其开发的红旗 HQ3 智能轿车在长沙至武汉完成了 286 公里高速全程自动驾驶试验，标志着中国汽车自动驾驶技术已达到国际相对先进的水平，如图 1.2 (b) 所示。



(a) Waymo

(b) HQ3

图 1.2 自动驾驶汽车

从 2009 年开始，为推动我国汽车自动驾驶技术的研究，国家自然科学基金

委员会连续举办了十二届“中国智能车未来挑战赛”。作为国内最具权威的智能车自动驾驶比赛，众多高校、科研机构以及企业都陆续参加此项赛事，并取得了不错的成效。从最开始的交通标志识别能力测试到立交、高架桥路段考验，由封闭道路走向真实道路环境，以及考察车路协同、避让行人等任务，比赛的难度逐渐增加。到 2019 年，设置城乡道路环境赛程和高架道路环境赛程，考察多个丰富场景。截至 2022 年，比赛开始考虑进行自动驾驶的出行服务，任务难度越来越大，图 1.3 是“中国智能车未来挑战赛”中一些场景模拟。

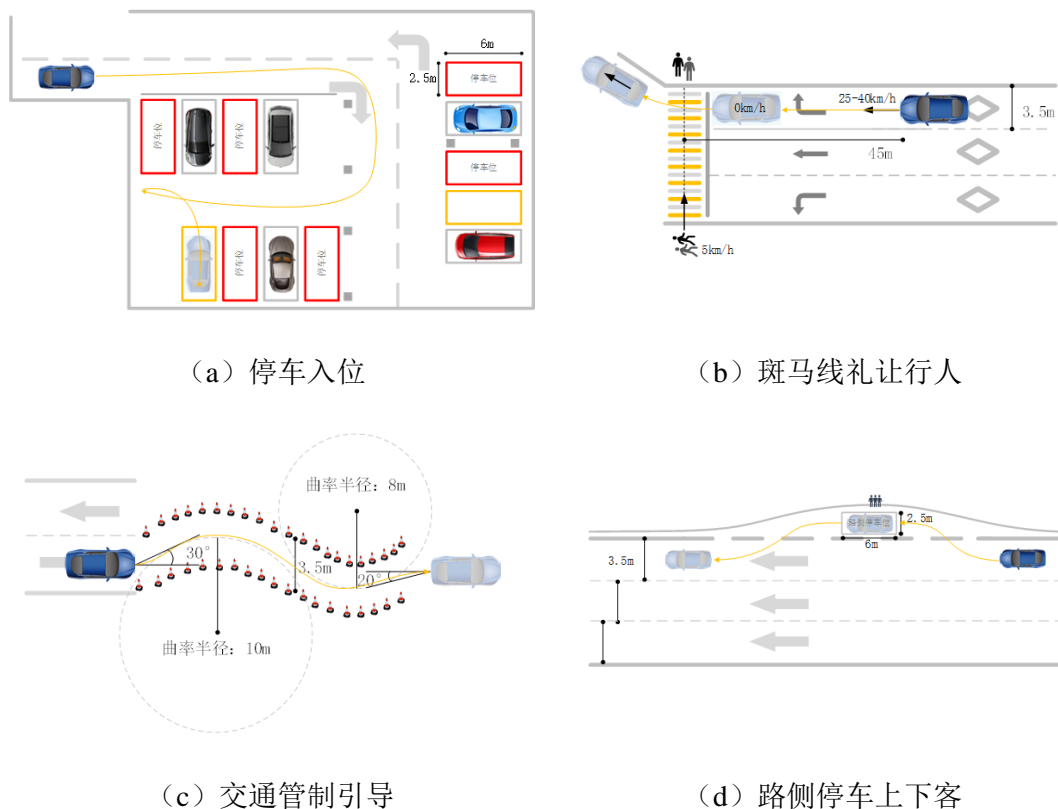


图 1.3 自动驾驶场景模拟

总体上，自动驾驶技术在汽车行业已经取得一定的成果。但与汽车实现自动驾驶有较大不同，工程机械的作业工况非常复杂，作业场景也很复杂，一般都是非结构化道路甚至是崎岖的山地。因此，虽然实现汽车自动驾驶的思想和理念可以借鉴，但汽车绝大部分自动驾驶技术无法直接移植到工程机械行业。

### 1.2.2 工程机械自动驾驶技术研究现状

工程机械自动驾驶研究起步较晚，主要有美国卡特彼勒、日本小松、三一重工、徐工、山推等企业在开展研究。

20 世纪 80 年代末，卡特彼勒公司开始研制矿用自动驾驶自卸车，1999 年，卡特彼勒公司针对矿区独特的作业环境，开发出充分利用矿区作业特点的自动推土机和地下自动铲车。之后这些早期工程机械智能化产品开始投入使用，固化作业场景，降低工程机械产品落地难度<sup>[9]</sup>。2000 年，韩国建筑技术学院的 Kim 联合美国威斯康星大学的 Jeffrey，提出了智能工程施工系统框架，将智能施工分为任务识别规划、资源定位和作业执行三部分，其中作业执行部分负责整合资源分布数据、定位数据和环境数据，通过运动模型完成自动行驶，轨迹规划完成自动作业任务，整个框架涉及多学科和多系统，对系统软硬件性能要求较高<sup>[10]</sup>。随后，澳大利亚机器人中心针对自动作业任务进行了轨迹规划及控制算法的研究，并于 2003 年通过改造小松微型挖掘机进行试验，其轨迹作业精度可以控制在 20 cm 以内<sup>[11]</sup>。



图 1.4 卡特彼勒“MineStar”系统及无人矿山作业现场

2008 年，卡特彼勒与知名大学机器人研究所合作，开发出大型自动驾驶卡车，以提高整合矿山运营的核心技术。2012 年，卡特彼勒研发出完整的“MineStar”系统，“MineStar”系统由五个功能模块组成：车辆管理系统（Fleet）、生产现场管理系统（Terrain）、安全检测系统（Detect）、设备诊断系统（Health）和调度协调指挥系统（Command）。车辆管理系统可参照结合作业计划予以实时的设备跟踪、车辆调配和选用统计管理，生产现场管理系统可做到良好的管理，安全检测系统依靠对环境敏感的检测器和报警系统来从旁协助操作以避免碰撞，设备诊断系统能够对设备进行诊断分析，加载实时数据，显示设备的运行状态和运行数据。“MineStar”系统能够共享各功能模块的相关数据信息，促进达成环境信息感知，如装车、运车、卸车、规划装卸地点和车辆行驶路线等<sup>[12]</sup>，如图 1.4



所示,是卡特彼勒的“MineStar”系统及无人矿山作业现场。

2013年,福特斯克金属集团联合卡特彼勒公司,签订了首批8辆自动化793F矿用卡车合同,用于所罗门(Solomon)铁矿区<sup>[13]</sup>。澳大利亚所罗门铁矿目前有60辆自动驾驶汽车在运行,累计产量超过2.5亿吨,其中12辆属于有人驾驶。无人驾驶卡车每天工作23.5小时,检查加油0.5小时,每天多工作2-3小时,一年工作近7000小时,相比有人驾驶卡车可提高生产效率20~30%<sup>[14]</sup>。目前,卡特运营的自动驾驶卡车总数已超过200辆。

2014年,瑞典Atlas Copco Rock Drills AB公司提出了对工程机械的位置状态参数实时观测的多机协同系统,以提高操作人员工作的安全性,但各工程机械之间在实际操作过程中无法很好地协同工作,工作效率低下。所以,大规模的推广未能进行<sup>[15]</sup>。

2016年,小松公司开发出第一辆矿用卡车,用于自动驾驶。小松公司生产的卡车自动驾驶主要通过自动化运输系统(Automated Haulage Systems, AHS)来实现<sup>[16]</sup>,该系统将原来坐在驾驶室外的司机换成了控制装置、GPS卫星、无线通讯技术和软件。同年9月,小松又开发出一款自卸车,以应对矿山的这一危险环境。该矿自卸车车体全长15米,额定载重230吨,最大功率2700马力,最高时速可达64 km/h。整车装上了相机、雷达等传感器,取消了驾驶室的设计,节省了空间,对于在矿山这种环境下对于物料的运输、装卸、避障等的难题起到了较好的作用。

2019年11月,徐工开发的徐工自动驾驶压路机集群在京雄高速进行智能化施工,图1.5为徐工无人驾驶压路机京雄高速施工现场。徐工将互联网和5G技术结合,真正实现了智能化监控并解决了一些人工技术难题,使得施工质量更高。该压路机在作业过程中,可实现远距离操控,覆盖广播电台和微波两公里,更远的距离可按需配置<sup>[17]</sup>。压路机碾压直线精确度在5 cm以内,重叠精确度在5 cm以内,每隔一段时间就会记录轨迹,随时可以查询。

日立(Hitachi)集团基于其EHAC-3系列矿卡提供自动驾驶改装方案,目前与澳大利亚电力公司Stanwell Corporation Limited合作,在Meandu煤矿建立了一个试验场进行原型试验。与此同时,BELAZ与矿山数字化公司VIST集团合作的2辆130吨载重自动驾驶矿卡,在俄罗斯的露天矿进行了小型试验。截至2020年3月4日,已经完成超过5000公里的运输测试,但目前没有声称可以进行大规模部署。



图 1.5 徐工无人驾驶压路机京雄高速施工现场



图 1.6 绅宝露天煤矿无人矿山现场

2020 年 5 月，绅宝集团、汇拓、航天重工就绅宝露天煤矿智能化改造展开深入交流与合作，图 1.6 是绅宝露天煤矿的无人矿山现场。10 月底，绅宝露天煤矿各项智能性能指标均达到安全运行要求，并通过了中国自动化学会组织的安全评审、执行授权的测试、检查和其他工作。至此，绅宝露天煤矿智能化改造不仅有效克服了极寒天气的严峻挑战，且自动驾驶的速度已经超过了人工驾驶的速度，促进达成了安全与效率的有机驱动。同时，该项目工业试运行状态稳定，未发生安全事故。

2020 年 11 月，以“科技引领绿色可靠，智慧赢得未来”为主题的中国工程机械工业协会铲土机械分会年会在山东济宁隆重召开，会上，山推公司发布了国内首款自动驾驶推土机。搭载高性能车载计算平台、北斗高精度定位系统，DH17C2U 自动驾驶推土机采取使用模块化、系统化设计和应用，支持 5G 智能通信，构建做到通用的“云”数字智能空间、时间、状态等服务，为集群运营奠定了技术和产品的基础<sup>[18]</sup>。推土机可广泛用于避免污染、辐射和危险，改善工人的工作环境，降低劳动强度，从而进一步提升冶金、化工、矿山、港口散货装卸、消防作业的生产效率等，在稳定的环境下，社会效益显著。

截止到 2020 年底，小松的 FrontRunner 自主运输系统已经在包括铜、铁、油砂、煤矿在内的三大洲 11 座矿山上运送了超过 30 亿吨的物料。车辆以及系统在使用过程中安全事故为零，安全性、生产力、环境耐受性以及系统灵活性等方面都经受了考验。操作员只需在 AHS 中设置好矿场的倾卸位置，就可以操作多达 30 辆卡车，极大地降低了人力成本。

此外，小松还总结了在矿山应用自动驾驶应当具备以下条件：

(1) 功能区域分离：必须在卡车的自动驾驶工作区域设置适当的标志和路障，以确保只有经过授权的人员和设备才能进入工作区域；

(2) 授权操作：只有经过 FrontRunner 操作和安全培训的人员才能授权操作；

(3) 简单的固定线路：固定线路使 FrontRunner 可以更好地引导自动驾驶卡车并优化生产效率；

(4) 灵活的路口规划：例如，小松发现使用“Y”形路口有助于自动驾驶卡车保持速度并减少自动驾驶卡车之间的冲突，从而提高效率；

(5) 宽阔的道路和排土场；

(6) 保养良好的运输道路；

(7) 最小化载人车辆与自动驾驶车辆的交互。

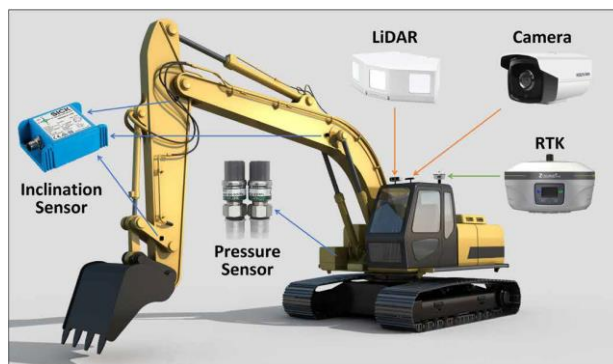


图 1.7 百度公司挖掘机自动作业测试车

百度机器人自动驾驶试验室与马里兰大学帕克分校合作，在 2020 年创设了新的自动挖掘机系统<sup>[19]</sup>，图 1.7 是百度公司装备自动挖掘机系统的挖掘机自动作业测试车。自动挖掘机系统包含感知、规划、控制等，其中，感知模块集成了多个传感器，其中包含激光雷达（LiDAR）和摄像头（Camera）、倾角传感器（Inclination Sensor）、压力传感器（Pressure Sensor）等，同时还具备实时动态载波相位差分技术（RTK）。这些传感器能够感知 3D 环境并识别目标材料，还能借助先进的算法（例如除尘神经网络）生成清晰的图像。同时，这套系统还能够让挖掘机做到长时间独立的装料作业，无需任何人工干预，提供媲美经验丰富的体力劳动者的工作效率。

2021 年 12 月，东风公司与中国移动、中国远洋联合多家行业合作伙伴共同打造“智慧港口 2.0”，成为全国首个全自动化集装箱码头，并正式在厦门远海码

头开始商业运营。“智慧港口 2.0”推动 5G 高、中、低频立体网络、无人集运车开放场景混运、北斗高精度定位、多传感器融合改造等关键技术系统性提升，同时基于 5G 进行港机远程控制<sup>[20]</sup>。在“5G 智慧港口全场景应用”的基础上，不断创新发展港口自主集装箱车辆标准，并开展了智慧港口的商业运营。“智慧港口 2.0”借助实际现场试验和方案对比，具备投资小、建设周期短、绿色低碳、适用性广等特点，对传统集装箱码头智能化升级具备示范作用。

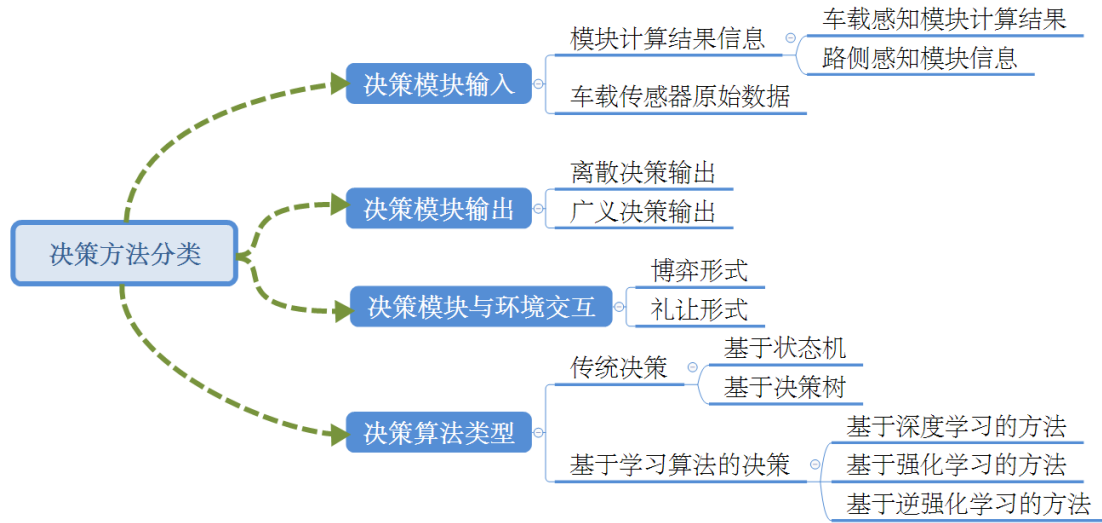
综上所述，目前工程机械自动化的实现集中在简单设备，针对的场景多为矿山、港口等单一区域，针对挖掘机、装载机等复杂机型的智能化还存在较多技术难点，实现工程机械自动驾驶的道路还很漫长，当前在工程机械自动驾驶领域的研究尚在初级阶段，开展针对工程机械的自动驾驶技术研究具有较高的学术意义和工程意义。

### 1.3 自动驾驶决策技术研究现状

自动驾驶技术主要可分为四大部分：感知、规划、决策和控制。决策是自动驾驶的关键技术之一，对于自动驾驶决策的研究最早可以追溯到 1985 年 MichonJA 提出的行为决策分解模型，即战略层、战术层和操作层<sup>[21]</sup>。

图 1.8 是自动驾驶的决策方法分类。考虑从环境感知的信息输入到决策算法输出的信号，可以将决策模块从输入、输出、与环境的交互、决策所采用的算法进行分类<sup>[22]</sup>。按决策输入信息是否计算来分，分为两大类：一类是输入车载模块的计算结果，另一类是输入车载传感器的原始数据；根据决策模块的输出信息来分，分为离散决策输出和广义决策输出两类；根据决策模块与外界环境的交互方式可以分为基于博弈形式的决策和基于礼让形式的决策；最后从决策所采用的算法来分，可以分为传统的决策和基于学习算法的决策两大类。下面将对传统决策和基于学习算法的决策方法进行全面概述。





### 1.3.1 传统的自动驾驶算法研究现状

传统决策系统按各子模块间的连接方式分为串联式、并联式、混联式三种决策方式，其中一种最具代表性的是有限状态机(FSM)，简单地说就是离散输入输出的抽象数学模型。有限状态机是由开始状态，输入和转换成下一状态的函数组成。具体过程是：当前状态接收外部输入的事件，满足状态转移函数后就会执行相应的动作，并迁移到新的状态。有限状态机法因其使用方便、稳定性好而在国内外广泛使用，其中卡耐基梅隆大学的 Boss、斯坦福大学的 Junior、弗吉尼亚理工大学的 Odin、麻省理工大学的 Talos 自动驾驶车辆都采用有限状态机法组成的自动驾驶决策系统。在国内，国防科技大学研制的红旗 CA7460 和中国科技大学研制的智能先锋 II 号也均采用有限状态机法。

针对当前的自动驾驶汽车采用交通法规和交通参与物混合处理的方式来进行决策，但决策系统各层并不能很好的分隔，陈永尚<sup>[23]</sup>提出了一种基于多目标优选的无人驾驶行为决策方案，让汽车能更好的遵守交通规则并且躲避障碍物。试验证明这种方法可以很好的应对交叉路口、超车、车道变换等场景，通过微观交通仿真系统建立复杂的交通场景，该方法具有一定的适应性。但该方法只能针对简单场景取得较好的效果，在复杂的真实场景下的效果还未可知。

传统的决策方法逻辑性较强，大多具有严密的规则，可以根据具体的场景制定合理的规则以便做出决策，场景和决策因果关系明确，系统的可解释性和稳定

性较强，系统一旦发生问题，可以比较容易的定位故障所在并进行相关调试；且可拓展性较强，可以根据不同的场景制定不同的规则和方案。但是现实生活中场景是不断变化的，传统的决策方法难以覆盖所有的工况，任务也难以穷尽。

### 1.3.2 基于学习的自动驾驶决策算法研究现状

传统的决策方法架构非常冗余，算法结构比较复杂，难以适用于工程机械较为复杂多变的工况。随着科学技术的发展，尤其是人工智能的兴起，使得简单高效、收敛快的基于学习的算法逐渐走进研究学者的视野。基于学习算法的自动驾驶决策系统直接依据传感器原始数据，在行驶过程中借助神经网络对车辆控制信号进行预测，避免了按规则控制的传统决策系统出现的检测等模块的精度问题，简化了系统结构，缓解了整个自动驾驶系统的计算代价<sup>[24]</sup>。

2018 年，剑桥的 Wayve 公司介绍在虚拟环境中预训练 DRL 模型完成车道跟随，然后将虚拟训练后的网络移植到现实环境中可以在一天内完成训练并进行车道跟随控制<sup>[25]</sup>。具体原理如图 1.9 所示。

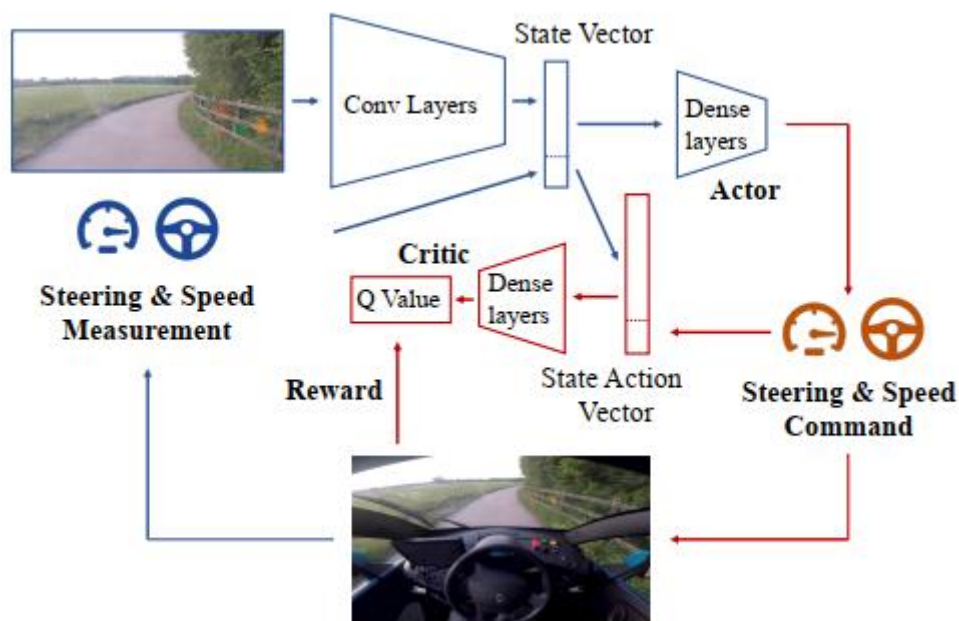


图 1.9 基于摄像头的深度强化学习端到端框架

Valiente R 等<sup>[26]</sup>考虑到图像前后帧之间的关系，在传统的 CNN 网络基础上附加一个时间序列网络，以此来构建端到端网络，最后利用 V2V 通信共享技术获取现在和未来的图像，更为精确的预测车辆的转向角度。试验证明该相互通信

并共享数据的模型展现出了更好的性能,预测转向角结果更为准确,但该模型未在真实数据中应用,仅在模拟数据中进行了验证。

针对现有端到端自动驾驶模型输入数据类型单一导致预测精确度低的问题,赵祥模等<sup>[27]</sup>选取 RGB 图像、深度图像和车辆历史连续运动状态序列作为多模态输入,提出一种基于 MM-STConv 的端到端自动驾驶行为决策模型。研究结果表明,该模型与其它现有主流模型相比,该模型有更强的稳定性和泛化能力。

李凌云<sup>[28]</sup>运用图像翻译的方法缩小虚拟与现实之间的差距,使用 CycleGAN 网络将虚拟图像转换成与真实图像相似的视觉图像,将生成的图像作为深度强化学习算法的输入,将虚拟环境中学习到的策略直接迁移到现实世界,从而实现一个全新的端到端决策方法。试验证明运用迁移学习的方法,在制作的真实数据集当中,可以取得比原始端到端方法更高的一个准确率,但真实世界是复杂多变的,该方法并没有在真实车辆和真实场景当中得到验证。

Hook J 等<sup>[29]</sup>提出了一种基于数据驱动的多智能体的强化学习决策方法,在虚拟的多智能体场景中测试强化学习算法。但是此方法只运用到了虚拟世界,而现实的多智能体的环境非常复杂,只有用到真实世界才具有更为广泛的意义。

深度学习已被用于从原始传感器输入到自动驾驶汽车控制的端到端神经网络学习。一方面,激光雷达传感器可以提供可靠准确的信息,但处理 3D 数据需要大量内存占用和计算成本,因此,现有的端到端驾驶解决方案主要基于摄像头;另一方面,提高系统的鲁棒性也较为关键,然而,由于基于采样的方法的成本,估计模型的不确定性具有很大挑战。基于上述问题,LIU Z 等<sup>[30]</sup>在 2021 年提出了一个高效且鲁棒的基于激光雷达的端到端导航框架,如图 1.10 所示。基于稀疏卷积核优化和硬件感知模型设计的 Fast-LiDARNet,从单一的向前传递,提出了混合证据融合,直接估计预测的不确定性,然后智能地融合控制预测。论文在一辆真实车辆上评估了系统,展示了车道的稳定导航能力。此外,在出现紧急事件(例如传感器故障)时,很少需要人工接管现实世界中的车辆,系统保持着较高的鲁棒性。

综上所述,传统的决策方法尽管稳定性、可解释性、拓展性较强,但其架构冗余,算法结构复杂,难以应对工程机械复杂的场景和工况。而基于学习算法的自动驾驶决策方法由于其简单高效和强大的表征能力,正逐渐成为自动驾驶领域的主流方法。然而基于学习算法的决策方法还不够成熟,决策精度较低,且大都依托一个单目相机来进行端到端的决策,传感器获取的信息不足,网络设计简

单，难以做出比较安全、可靠的决策。

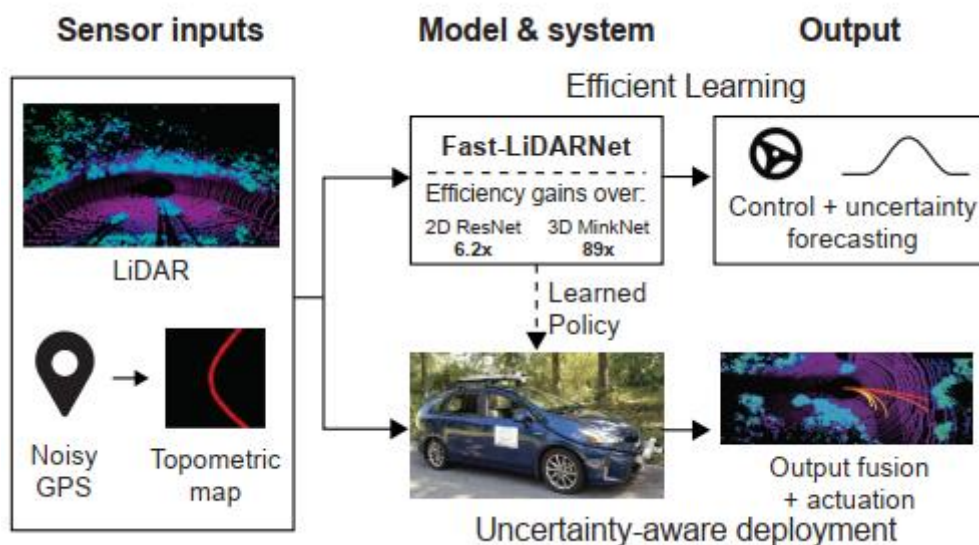


图 1.10 基于激光雷达的端到端导航框架

## 1.4 课题的提出及研究内容

### 1.4.1 课题的提出

随着国家的需求、行业变革，智能化、无人化逐渐成为工程机械行业的主要发展趋势之一。工程机械电动化、线控底盘技术的成熟，为实现智能化控制奠定了技术基础；人工智能、5G 技术、物联网、大数据、云计算等高新技术也为工程机械开展智能化提供了支持。同时工程机械领域自身也迫切需要开展智能化，作业环境复杂、工况众多，常作业于矿山、港口、山地等崎岖的非结构化道路，施工困难，人力成本巨大，人身安全有着极大的危险。

综上所述，本文针对当前工程机械自动驾驶中的决策技术难点开展研究，提出了基于双目图像的多模态决策算法、基于激光雷达点云的决策算法，在开源数据集和非结构化真实场景数据集上开展试验，进行对比，并基于提出的算法选择开展了实车测试，基于端到端实现了履带式挖掘机特定场景的自动驾驶行走功能。

### 1.4.2 课题研究内容

本文根据工程机械的复杂工况需要,结合试验室现有情况,设计开发了履带式工程机械专用的自动驾驶决策系统方案,实现了履带式挖掘机非结构化道路自动驾驶行走,具体包括以下方面的研究:

#### (1) 自动驾驶决策系统方案设计

针对工程机械本身特点,基于端到端方案,本文分析和对比了几种主流传感器。依据端到端决策算法的需要,选取双目摄像头和激光雷达作为主要传感器开展研究,依托某履带式挖掘机试验平台,基于该样机设计了自动驾驶决策系统方案。

#### (2) 基于双目图像的工程机械自动驾驶决策算法

为确保能够提取工程机械行走时复杂环境的特征,实现端到端自动驾驶,本文首先采用双目相机,并将多种图像信息进行融合,引入注意力机制,构建了一种基于多模态的工程机械自动驾驶决策网络,同时使用自动驾驶开源数据集和非结构化真实场景数据集进行试验和分析。

#### (3) 基于激光雷达点云的工程机械自动驾驶决策算法

为了寻找适合工程机械自动驾驶行走的方案,引入激光雷达点云与双目图像决策网络进行对比。基于 X-Conv 构建点云端到端网络,并嵌入残差连接,最后使用自动驾驶开源数据集和真实场景数据集进行试验和分析。

#### (4) 非结构化道路实车试验与分析

为进一步验证所搭建的基于双目图像和激光雷达点云决策网络的有效性,以履带式挖掘机为例,进行了实车试验。在校园某非结构化试验场地内搭建试验平台,进行数据采集工作,最后针对典型的工程机械自动驾驶工况进行了实车测试和分析。

## 1.5 本章小结

本章对所研究课题国内外背景进行了详细的介绍。首先介绍了智能车、工程机械的自动驾驶技术的整体研究现状,接着详细介绍自动驾驶领域决策技术发展现状,指出了智能车自动驾驶决策技术和工程机械自动驾驶决策技术需求的差异,对工程机械决策难点进行了分析,引出了本文的研究重点。

## 第2章 端到端工程机械自动驾驶决策系统方案设计

工程机械的自动驾驶与智能车有一定差异，可以分为自动驾驶行走和自动作业两部分，自动行走和智能车相类似，大致可以分为感知、规划、决策、控制四个部分，传感器传入数据到感知系统经过计算和分析，之后传入规划和决策系统，最后经由计算平台输出控制信号到整车控制器。而端到端方案可以实现由周围环境信息到车载横、纵向控制信号的直接映射，步骤简单，本章基于端到端方案，针对工程机械本身的特点，从传感器、数据集、和决策算法三方面出发对工程机械专用自动驾驶决策系统方案进行了分析和讨论。

### 2.1 自动驾驶常用传感器

传感器是实现自动驾驶硬件不可或缺的一部分，在端到端方案中更尤为重要，可以为实现工程机械自动驾驶行走提供丰富的环境感知信息及定位信息。数据决定自动驾驶的上限，不同的传感器可以获得不同的信息，针对工程机械不同的环境和工况，需要选择合适的传感器，下面对常用的传感器进行介绍。

#### 2.1.1 摄像头

摄像头在自动驾驶中是最常用的传感器，成本相对较低，可以划分为单目、双目、以及多目摄像头，如图 2.1 所示，是比较典型的用于自动驾驶的双目和多目摄像头。在自动驾驶任务中，相机主要用来拍摄道路的图像，根据目标检测、深度估计等机器视觉技术获得环境感知信息以及定位信息，然后进行规划决策<sup>[31]</sup>；也可直接根据相机获得的图像信息进行端到端推理，通过车载计算平台输出控制信号到整车端。

单目相机可以分析周围环境的变化，并通过相机的视野对终端或云端的处理器进行反馈结果的判断。单目摄像机结构简单，造价较低，便于标定和使用，但不能同时兼顾测距范围和距离，两者存在着不可调和的矛盾，即摄像机所能获得的精确距离越短、视角越窄。通过相机的运动形成轨迹，单目相机可以测量物体的相对深度，但无法确定图像中物体的真实尺度。

双目视觉是以视差原理为基础，通过多幅图像获取物体三维几何信息的一



种重要的机器视觉形式。双目立体视觉技术的实现大致可以分为以下几个步骤：获取图像、相机标定、特征提取、立体匹配以及三维重建<sup>[32]</sup>，在自动驾驶任务中，可以用双目相机来代替激光雷达，获取三维信息，特斯拉正是采用了这样的一种彻底的方式。相较于单目摄像头而言，双目摄像头可以运用到室内和室外，基线距离越大，能够测量的距离就越远，但配置和标定流程较为复杂，深度量程和精度会受到双目基线距离与分辨率的限制，同时在双目立体视觉技术中，双目视差计算相当耗费资源，需要一些加速设备进行加速。

多目摄像头在自动驾驶中应用相对较少，可以通过使用不同的相机，将不同范围的场景覆盖在周围环境中。例如，用来看近处环境的广角镜头，30 米左右的环境可以由 80 度镜头负责，60 度镜头负责中远距离，40 度镜头远距离观察。但多目摄像机与其他类型摄像机相比，成本、安装位置、精度等方面的要求本身就不是一个完美的解决方案。



(a) 双目摄像头



(b) 多目摄像头

图 2.1 摄像头

### 2.1.2 雷达

在自动驾驶任务中，雷达是非常重要的一类传感器，可以获得对周围环境的感知信息，主要有三大类：激光雷达、毫米波雷达和超声波雷达<sup>[33]</sup>。激光雷达是一种传感器，可用于精确测量周围环境的三维位置信息。一般由光学发射模组、光电接收模组、机芯模组以及讯号处理模组等构成。其工作原理是光学发射模组按照一定频率发射侦测讯号，接收模组在侦测讯号遇到目标物反射时，将反射回来的讯号收集起来，透过相应的计算转换，即可取得周围环境中的空间资讯，以及目标对象的距离、方位、尺寸等参数，以达到侦测、辨识、追踪特定区域内的环境与目标。

激光雷达的分类方式众多，常用的分类方式按扫描方式划分，可以分为机械

旋转式激光雷达、固态激光雷达和混合式激光雷达。机械式激光雷达是最早应用于自动驾驶的激光雷达产品，原理简单，易于驱动，容易实现周围 360 度扫描，因此，直到今天仍被广泛应用于自动驾驶领域。代表厂家有 Velodyne、Ouster、禾赛科技等公司，如图 2.2 所示为激光雷达厂家的代表性产品，64 线激光雷达 HDL-64E 以及 128 线激光雷达 VLS-128。但机械式激光雷达也有自身的缺陷，其内部含有大量的活动部件，容易受到车辆震动的影响，长时间使用可靠性较差，容易在驾驶环境中受到磨损，而且其造价较高，很难落地。



(a) HDL-64E



(b) VLS-128

图 2.2 激光雷达

为降低激光雷达的生产成本，增加其结构的稳定性，同时保证其较好的探测性能，混合固态式激光雷达和固态式激光雷达应运而生。混合式固态激光雷达是基于微机电系统（Micro-Electro-Mechanical System, MEMS）研制的一类激光雷达。MEMS 激光雷达主要由二维 MEMS 微镜、激光发射模块、光束探测系统和 APD 传感器组成，采用微振镜结构进行激光束偏转，微型化、电子化设计机械结构，避免机械激光雷达整体大幅度转动，有效降低车辆行驶过程中的功耗和整套系统发生意外的几率。MEMS 激光雷达也有缺陷，振镜机械扫描角度较小，要实现大视场扫描需要增加角度扩展系统，对震动敏感，很难控制。图 2.3 是 MEMS 激光雷达的工作原理图。





18

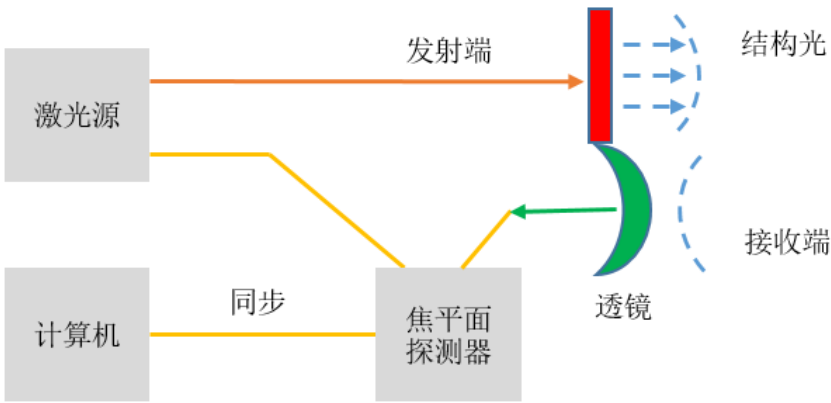


图 2.5 Flash 激光雷达工作原理图

毫米波雷达在自动驾驶领域应用也非常广泛，与激光雷达相比，毫米波雷达技术更成熟，已经广泛应用于汽车的安全系统中<sup>[34]</sup>。图 2.6 为毫米波雷达实物拆解图，毫米波雷达是车载的核心传感器，按照频段可将毫米波雷达分为 3 类：24 GHz 频段、77 GHz 频段、76~81 GHz 频段，不同频段的毫米波雷达应用在车载的不同位置 and 不同任务。24 GHz 频段的毫米波雷达探测范围较广，但频率低、带宽窄，探测距离较近，主要应用于车载的盲区检测、编辅辅助等，作为侧向雷达使用；77 GHz 频段的毫米波雷达频率、带宽比较高，常用做前向雷达，测量本车与前方车辆的相对距离和速度，主要用于自适应巡航、避障等；76~81 GHz 频段的毫米波雷达则具备极高的距离分辨率、带宽非常高，主要应用于对于精细物体的检测。

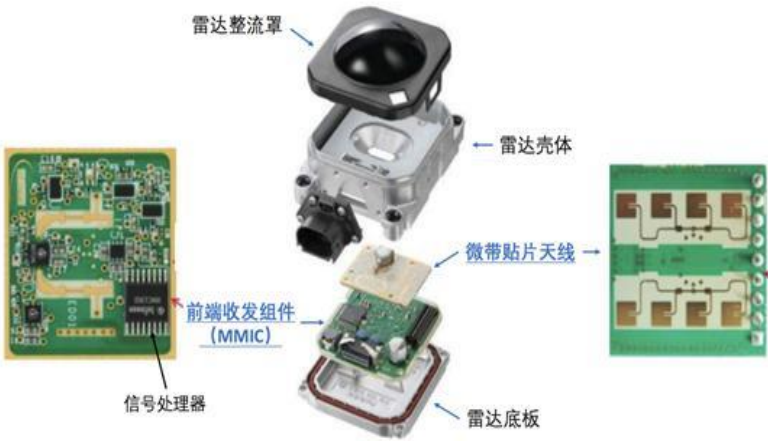


图 2.6 毫米波雷达实物拆解图

超声波雷达常见的主要有两种，一种是用来测量车辆前后障碍物的，叫做倒车雷达(UPA)；另一种用来测量左右障碍物的距离，叫做 APA。超声波传播距离远，测距简单，成本也比激光雷达和毫米波雷达要低，但测距准确度在车辆行驶速度较高时较低。

## 2.2 端到端自动驾驶系统

端到端在神经网络中直接输入传感器原始数据，行驶过程中借助神经网络实时对车辆控制信号进行预测。根据工程机械的复杂工况需求，本节对端到端组成、相关的数据集和仿真平台进行分析，选择合适的数据集便于后续工程机械自动驾驶决策方案的验证。

### 2.2.1 神经网络与自动驾驶

神经网络的概念很早就被提出，神经网络主要指的是深度神经网络，也被称为深度学习，深度学习从属于机器学习，是人工智能领域的重要分支，2006 年由多伦多大学教授 Hinton<sup>[35]</sup>正式提出。随着计算机、大数据等技术的蓬勃发展，深度学习发展迅猛，逐渐影响图像识别、语音识别、自动驾驶、智慧医疗等领域。

深度学习的发展推动着自动驾驶技术的发展，主要影响着自动驾驶的环境感知技术。随着深度神经网络模型的发展，基于学习算法的自动驾驶决策由于其简单高效和强大的表征能力，也在自动驾驶汽车领域受到了广泛关注及应用。最早的是由 NVIDIA 提出的一种基于端到端的 PilotNet 网络架构<sup>[36]</sup>，根据前方道路图像直接输出转向角度，实现了自动驾驶汽车在高速公路和平坦道路上的车道保持，开创了端到端自动驾驶的先河，如图 2.7 所示，是 PilotNet 网络的具体结构。

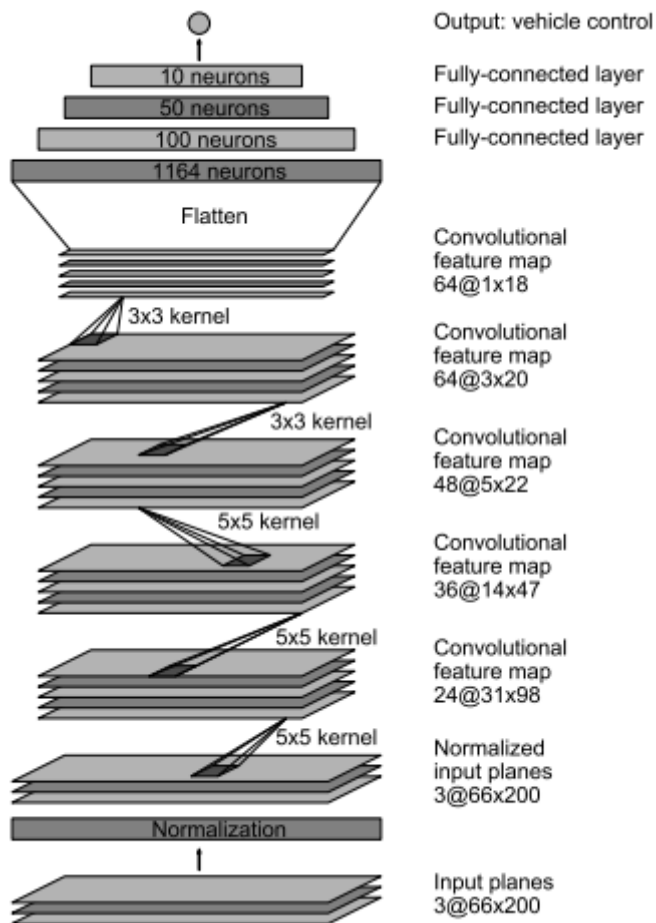


图 2.7 PilotNet 网络具体结构

端到端自动驾驶的大致步骤是先人为采集图像、激光雷达点云、车辆的控制信号，同时保证时间同步，控制信号在汽车领域主要包括方向盘转角、油门、刹车、速度等信号，在工程机械领域，如履带式挖掘机，则需要采集左右履带的信息以及速度等。将图像、激光雷达点云等传感器原始数据直接输入到端到端自动驾驶模型，类似于计算机视觉领域，车辆的控制信号作为网络训练的标签，经过数据集训练、验证之后，依据保存的模型参数加载模型实现对于车载的横、纵向控制，从而完成自动驾驶行走、自动作业等任务。

### 2.2.2 自动驾驶开源数据集

数据集在自动驾驶技术中尤为重要，可以对算法进行验证，加快研究进度。在进行实车试验之前，许多的自动驾驶公司 and 研究团队均会利用数据集训练并

验证算法在各个评估指标的表现,节约成本,在研究前期快速完成各项算法的迭代,表 2.1 是主流自动驾驶数据集在场景和数据信息方面的对比。

(1) KITTI<sup>[37]</sup>数据集。目前全球最大的自动驾驶各种场景的计算机视觉算法评价数据集,由德国卡尔斯鲁厄理工学院和丰田美国理工学院联合创设建立。该数据集可用于评估 3D 成像 (Stereo)、光流 (Optical Flow)、视觉测距 (Visual Odometry)、3D 物体检测 (Object Detection) 和 3D 跟踪 (Tracking) 等任务。KITTI 数据集包含从城市、城镇和高速公路等乘用车场景以及具备不同程度遮挡和截断的场景收集的真实图像数据,每幅图像最多包含 15 辆汽车和 30 名行人。整个数据集涵盖 389 对立体图像和光流图,39.2 km 的视觉距离测量序列,同时以 10 Hz 的频率在数据层面同步。

(2) NuScenes<sup>[38]</sup>自动驾驶多模式数据集。是多个自动驾驶场景下第一个拥有全自动驾驶传感器的数据集,包含六部相机、五部雷达与一部激光雷达,都是 360 度的全景视图。自动驾驶数据集含有 1000 个场景,每个场景约 8 秒长的视频,标注 23 类对象与 8 个属性类别的 3D 轮廓,比 KITTI 数据集多七倍的标注信息与 100 倍的图像数据,拥有最大规模的 3D 标注框数据。

(3) Waymo 自动驾驶数据库。自动驾驶数据集中规模最大、内容最全、种类最丰富的一种。Waymo 是谷歌母公司 Alphabet 旗下的自动驾驶公司,该公司宣布将团队的自动驾驶数据库开源,用于帮助学术界在机器感知和自动驾驶技术上实现技术进步。公开的数据集包含了 Waymo 开发的涵盖了多个地点、1000 个行驶路段所收集的数据以及自驾汽车在各种条件下收集到的高精度标注数据。

(4) DBNet<sup>[39]</sup>数据集。厦门大学 SCSC 试验室李军教授团队和上海交通大学 MVIG 试验室卢策武教授团队专门发布的研究驾驶学习策略的大规模驾驶行为数据集。DBNet 数据集记录了合格驾驶员 (10 年以上驾驶经验) 的视频、激光雷达点云和相应的真实驾驶行为。

(5) NGSIM 高速公路数据集。美国联邦公路局发起的一项数据采集项目,被学术界学者广泛用于车辆跟随、换道等驾驶行为研究、交通流分析、微观交通模型构建、车辆运动轨迹预测、驾驶员意图识别、自动驾驶决策规划等。所有数据由摄像头采集,均为美国高速公路国道 101 上采集的行驶轨迹数据。数据文件包含 3 个 TXT 文档,三个文档对应三个不同的时段,每个时段 15 分钟。

(6) Lyft L5 自动驾驶数据集。是 Lyft 公司提供的包含丰富的高质量语义地图的 L5 级自动驾驶数据集,能够检测各种物体。该数据集包含了大约 3000 条

驾驶场景、超过 4000 条道路、197 条人行横道、60 个停止标志物和 54 个停车区域等信息。数据集为 NuScenes 格式，通过两类不同版本的汽车进行采集，两类试验车均搭载 7 个不同的摄像头和 3 个不同的雷达。

表 2.1 主流自动驾驶数据集

数据集	场景	地点	数据信息
KITTI	城市、乡村、高速	美国	摄像头、激光雷达、GPS、IMU
NuScenes	城市、乡村、高速	新加坡、波士顿	摄像头、激光雷达、雷达、GPS、IMU
Waymo	城市、高速	美国	摄像头、激光雷达
DBNet	城市、高速	中国	摄像头、激光雷达、转角、车速
NGSIM	城市、高速	美国	摄像头、车辆轨迹参数
Lyft L5	城市、高速	美国	摄像头、激光雷达

### 2.2.3 自动驾驶开源仿真平台

自动驾驶仿真平台对于自动驾驶的研究也必不可少。实车测试需要花费极大的人力物力，对于许多高校和企业来说，成本、门槛较高。在实车测试之前利用仿真平台对自动驾驶各项算法进行测试，不仅节约了运作成本，同时极大地减少了试验的安全问题，许多自动驾驶公司不惜耗费巨大的资金来搭建一个相对比较完善和逼真的仿真平台，图 2.8 为常用自动驾驶仿真平台的一些图片实示例。

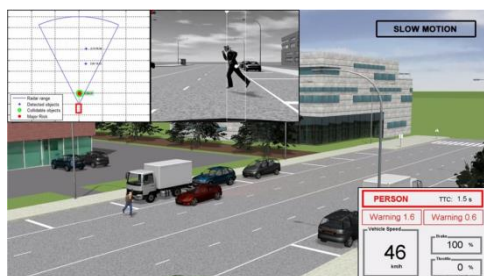
PreScan 是西门子公司根据物理模型开发和评估汽车的一款仿真软件。具备模型在环（MIL）、软件实时在环（SIL）、硬件在环（HIL）等各式各样选用模式，其中包含 134 个场景和 19 个传感器模型，主要涵盖摄像头、雷达、激光雷达、GPS 等，可开发多种应用功能，如 ADAS 开发、模拟智能车载系统、V2V/V2I 车辆通信等，支持与 Simulink、ROS、Autoware、FMI 等联合仿真。迄今为止，PreScan 已广泛应用于智能汽车系统的各项任务当中。

CarMaker 是一款专门针对乘用车和商用车的车辆动力学模拟软件，由德国 IPG 公司开发。首先，CarMaker 是一款出色的动力学模拟软件，不仅提供了精

密的车辆车身模型(引擎、底盘、转向等),还创造了闭环模拟系统,包括车辆、司机、道路和交通环境。其次,CarMaker 作为平台软件,可以与 ADAMS、AVLCRUISE、RFPRO 等众多第三方软件进行整合,也可以和其它诸如 DSPACE、CarSim 等车载模拟软件一样,与 MATLAB、Simulink 等进行联合模拟。

CarSim 软件是一款由专业研究汽车动力学模拟软件 MSC 公司开发的模拟软件。主要针对于四轮汽车和轻型卡车,从整车角度进行仿真,不需要进行复杂的调参,用户可以快速使用。同时,CarSim 的运行速度非常快,能够实现对驾驶者操控的实时模拟车辆的反应以及 3D 路面和空气动力学输入;可以模拟附着系数、直弯道等路况车辆的加速行驶、制动和方向盘转向输入与调整,模拟环境高度接近真实车辆;可用于对车辆的操纵稳定性、制动性、平顺性、动力性、经济性等方面进行预测和模拟;对于 CarSim 自带标准的 MATLAB 和 Simulink 接口,可方便地与 MATLAB 和 Simulink 进行联合仿真,用于控制算法的开发和后续利用 MATLAB、Excel 等工具进行分析或可视化。RT 版本的 CarSim 还可支持 DSPACE 和 NI 的系统等主流的 HIL 测试系统,方便 HIL 联合模拟。

CARLA<sup>[40]</sup>是在西班牙巴塞罗那自治大学计算机视觉中心的指导下开发的用于自动驾驶研究的开源模拟器。与 AirSim 一样,Carla 通过采用高度可扩展的服务器和多客户端架构,选用虚幻引擎开发。在场景方面,CARLA 创设建立了自动驾驶场景的开源数字资源(其中包含城市布局、建筑和车辆),以及根据这些资源构建了多个自动驾驶训练、测试场景。同时,CARLA 选用 VectorZero 的修路软件 RoadRunner,创设建立与车辆周围环境相匹配的高精度地图,并予以简单的地图编辑器。CARLA 可任意选配传感器,使用者可配置包括相机、雷达、深度传感器、GPS 等多种传感器套件,还能够调节车辆周围环境的光线和天气,用于方便与自动驾驶系统进行联合仿真,完成决策系统和端到端的强化学习训练,控制场景中的车辆和信号灯。

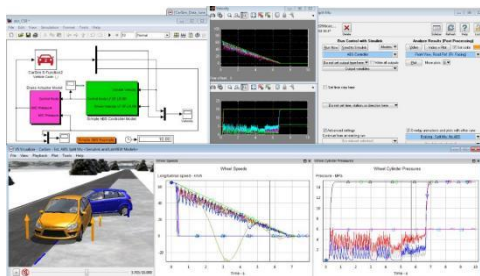


(a) PreScan

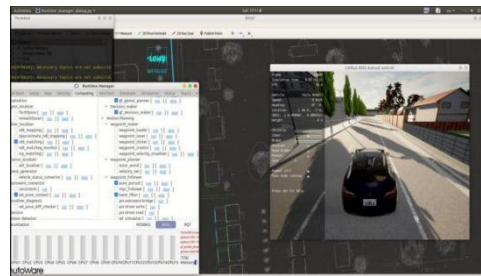


(b) CarMaker





(c) CarSim



(d) CARLA

图 2.8 自动驾驶仿真平台图片实例

## 2.3 工程机械决策系统方案设计

本节依据前文所述，并参考工程机械自身的特点，对本文的自动驾驶决策系统在环境感知传感器和决策方案两方面进行设计。

### 2.3.1 环境感知传感器选择

工程机械作业环境复杂且工况众多，大部分工程机械运用于矿山、高原、南北极等环境比较恶劣的地形以及地震、泥石流等救灾抢险的恶劣工况中。这类环境通常含有大量粉尘、金属等颗粒，对雷达的影响比较严重，而摄像头则不适用于雨雪天气和夜间的情况。在工程机械作业场景中，相机主要用来拍摄道路的图片，根据目标检测、距离估计的机器视觉技术进行定位，进行下一步的路径规划。基于相机的感知可以获得大量的外部信息，但一部相机并不能满足要求，而多部摄像头，会产生大量的计算，无法保证工程机械行驶的实时性。此外，相机受外部光照、天气等环境的影响非常大，且鲁棒性不够好。基于雷达也可以实现定位和导航，用的最多的是激光雷达，激光雷达主要通过发射装置发射激光，然后再接收激光，就可以根据时间差来计算物体的距离。激光测距解析度高，抗干扰能力强，也具有较好的隐蔽性和定向性。但是激光雷达在大雨、大雪等恶劣环境下，也会受到天气的影响，传播距离也会出现明显衰减。而且在大气环流的影响下，激光雷达也会受到不同程度的影响。因此，需要两类传感器进行互补，对于两类传感器都有研究的必要。

摄像头有单目、双目以及多目的区分，单目和双目摄像头都是目前自动驾驶领域的主流。单目摄像头结构简单，成本较低，方便标定和使用，但在测距的范



围和距离方面，有着不可调和的矛盾；多目摄像头可以用不同的镜头来覆盖不同的场景范围，但本身成本较高、安装位置难以确定、算法处理更为复杂，目前也没有完美的解决方案。相比之下，双目摄像头尽管在传统的技术上有着获取图像、相机标定、特征提取、立体匹配和三维重建等复杂的步骤，但采用端到端的自动驾驶方案可以避免繁琐的步骤，直接根据采集的双目图像信息实现由周围环境信息到履带式挖掘机左右履带控制信号的直接映射。

表 2.2 各类传感器

名称	摄像头	激光雷达	毫米波雷达	超声波雷达
功能	利用计算机视觉判别周围环境与物体	障碍检测、跟踪、路面检测、定位和导航、环境建模等	感知大范围内车辆的运行情况，多用于自适应巡航系统	探测低速环境、常用于自动泊车系统
优势	成本低、硬件技术相对成熟、可识别物体属性	精度极高、分辨率高、抗干扰性强、探测范围大	全天候工作，探测距离远，性能稳定，分辨率较高	成本低，近距离探测精度高，不受光线条件的影响
劣势	易受恶劣天气影响，难以精确测距	成本高，容易受雨雪、大雾等天气影响	探测距离受频段损耗制约，感知行人能力弱，探测角度小	受信号干扰，探测距离短

雷达在自动驾驶领域主要使用的有激光雷达、毫米波雷达和超声波雷达，激光雷达的使用更为常见。毫米波雷达也是自动驾驶的核心传感器，相对于激光雷达来说，价格更为便宜，但毫米波雷达更多的用于局部的自动驾驶，如车载的盲区检测、自适应巡航、避障等功能；超声波雷达与毫米波雷达类似，常用来测量距离，也更多的实现的是自动驾驶的局部功能，如表 2.2 所示是各类传感器的功能和优劣对比。

综上所述，本文选择双目摄像头和激光雷达作为端到端自动驾驶试验的环境感知传感器，分别探索双目摄像头和激光雷达对决策算法的影响，探索端到端的方案如何实现工程机械的自动驾驶行走。为后续开展多传感器融合、结合学习算法和基于规则的决策奠定基础。

### 2.3.2 自动驾驶决策方案

与乘用车、港口机械、环卫车不同，工程机械工作环境恶劣，对驾驶员的身体各方面的素质要求非常高，随时可能给驾驶员的生命造成危险。通过自动驾驶技术对车辆进行控制，在行走和作业时自主获取周围环境信息并做出决策，极大降低了操作人员的作业风险，减少了劳动力的浪费，提高了作业效率。

依据 1.3 节所述，目前无人驾驶决策技术主要应用在汽车领域，主要可分为基于规则的决策和基于学习算法的端到端决策两大类，基于学习算法的端到端决策主要是基于深度学习和深度强化学习的方法。基于规则的决策非常简单，只需根据驾驶场景制定一系列的规则，逻辑明确，但是规则的制定无法穷尽，且无法适用于较为复杂的场景；基于深度强化学习的自动驾驶决策方法通过和环境不断交互获得数据进行训练，无需制作数据集和标注。但是深度强化学习算法存在收敛慢，难以设计合适的奖励函数等问题，目前主要运用在游戏领域，多在仿真环境中进行测试。而端到端自动驾驶决策方法可利用网络结构简化算法，具有对场景深度遍历的优势，具有自学习功能。现有的自动驾驶仿真系统主要服务于汽车领域，对于工程机械并不适用，工程机械目前仿真系统发展也并不完善，故本文基于端到端自动驾驶的思想，只选择开源的 DBNet 数据集以及采集的自定义数据集，对双目摄像头和激光雷达进行对比，构建决策网络实现特定作业场景下的工程机械自动驾驶行走。

## 2.4 本章小结

本章以自动驾驶端到端方案为基准，首先介绍了目前自动驾驶领域主流的传感器，并对传感器的适用场景和优缺点进行了分析；然后详细介绍了端到端自动驾驶系统，包括常见的开源数据集和开源自动驾驶平台；最后基于工程机械自身和主流传感器的特点，选择使用双目摄像头和激光雷达构建端到端决策网络，确定了课题研究路线，为后续的研究工作做好准备。



## 第3章 基于双目图像的工程机械自动驾驶决策

端到端自动驾驶简化自动驾驶技术的整个流程,实现由周围环境信息到车载控制信号的直接映射。随着计算机、大数据等技术的提升,人工智能领域的自动驾驶、远程遥控等许多新兴技术迅猛发展,基于学习算法的自动驾驶决策由于其简单高效和强大的表征能力,在自动驾驶汽车领域受到了广泛关注及应用。与汽车领域不一样,目前工程机械端到端研究相对较少,工程机械作业环境、工况复杂多变,常行驶于非结构化道路,甚至是崎岖的山地,相较于汽车而言,更难由端到端网络进行学习,且两者的整车控制系统存在较大差异,无法直接移植。针对上述问题,本章采用双目相机,将多种图像信息进行融合,同时引入注意力机制,构建了一种基于多模态的工程机械自动驾驶决策网络,完成履带式挖掘机的端到端行走决策<sup>[41]</sup>。

### 3.1 神经网络

神经网络也叫深度学习(Deep Learning),在人工智能领域是非常重要的技术。迄今为止,神经网络的技术已广泛应用于计算机视觉、自然语言处理等领域,并不断向多模态多任务方向发展。本文的工作基于神经网络进行,为了更好的论述本文的工作,本节将在卷积神经网络、注意力机制等方面展开阐述。

#### 3.1.1 卷积神经网络

卷积神经网络是一种前馈性神经网络,核心是卷积计算。20世纪60年代,Hubel和Wiesel发现猫的中枢神经系统的神经元有着相当独特的网络连接,基本结构能够有效降低反馈神经网络的复杂性,由此提出了卷积神经网络(Convolutional Neural Networks, CNN)<sup>[42]</sup>。到80年代,Fukushima等学者提出了第一个用卷积实现的神经网络,后面多位学者对其进行了改进研究。到1998年,开始应用卷积神经网络,LeCun提出用于手写数字识别的经典卷积神经网络LeNet<sup>[43]</sup>。由于当时计算机、大数据等技术的限制,卷积神经网络逐渐没落。直到2012年,LeCun所在团队又提出经典卷积神经网络AlexNet<sup>[44]</sup>,力压众多传

统算法，获得当时举办的 ImageNet 视觉识别比赛冠军，随后以卷积为主的深度神经网络迅猛发展，涌现了诸如 VGGNet<sup>[45]</sup>、GoogleNet<sup>[46]</sup>、ResNet<sup>[47]</sup>、DenseNet<sup>[48]</sup> 等非常经典的图像分类网络。

卷积神经网络通常由卷积层、池化层、全连接层、激活函数等组成，在卷积神经网络中，卷积层和全连接层主要用来进行特征提取，池化层进行上、下采样，激活函数给整个网络提供强大的非线性能力，使得网络可以拟合任意函数。

### (1) 卷积层

卷积在神经网络领域主要是二维离散卷积的运算，本质是进行内积的操作，即衡量两个向量之间的相似度，对于图像而言，是卷积模板与图像某一区域计算相似度，卷积结果越大图像中的某一位置与该卷积模板越相似。卷积最大的特点就是局部连接和权值共享，局部连接指网络各层之间局部连接，权值共享指同一卷积的卷积核参数共享，卷积核与输入图片不同区域做卷积，检测相同区域，而不同的卷积核对应不同的权值参数，检测不同的特征。二维卷积计算公式如下：

$$y(i, j) = \sum_{k_1} \sum_{k_2} x(k_1, k_2) K(i - k_1, i - k_2) \quad (3.1)$$

其中， $x$  为输入图像， $k_1, k_2$  为卷积核  $K$  的尺寸， $y$  为卷积计算结果。

二维卷积通常指的是标准卷积，经过不断更新迭代，到现在为止，已经出现各种各样的卷积操作，最常见的二维卷积有深度可分离卷积、分组卷积、空洞卷积等。

### (2) 池化层

池化层常用于卷积神经网络，一般和卷积层配合使用，可以调节网络数据的维度、抑制噪声、减少信息冗余、降低模型计算量、防止过拟合等，同时池化层没有可训练的参数，类似于激活函数，可以给模型提供非线性。池化操作在卷积神经网络中经常使用的包括最大池化、平均池化、反池化以及全局平均池化。

### (3) 全连接层

在卷积神经网络中，全连接层常用在网络比较靠后的位置，提取全局特征。本质是把卷积提取到的局部特征进行加权求和重新组成完整的图；在图像领域，全连接层将输入图像在经过卷积和池化操作后提取的特征进行压缩，并根据压缩的特征完成各项任务。相对于卷积层的局部连接，全连接层各层神经元均一一相连，也可以进行特征提取，但计算量过大，且容易过拟合，所以常使用全局平均池化层替代全连接层。

#### (4) 激活函数

卷积神经网络中激活函数常和卷积层、归一化层一起使用，作为卷积单元。神经网络中的每个神经元节点接受上一层神经元的输出值作为本神经元的输入值，并将输入值传递给下一层，而输入层的神经元节点将直接传递输入属性值到下一层（隐藏层或输出层）。在多层神经网络中，上层节点的输出与下层节点的输入之间存在一种函数关系，函数称为激励函数（也称为激活函数）。激活函数最大的作用就是给模型提供非线性，增强模型的表征能力，使得模型可以拟合任意复杂曲线。

#### 3.1.2 注意力机制

注意力机制（Attention Mechanism）<sup>[49]</sup>的提出没有严格的数学定义，比如传统的局部图像特征提取和滑动窗口方法都可以认为是一种注意力。在神经网络中，注意力机制通常是一个额外的神经网络，只要能从大量的信息中过滤出有用的信息，就能称为注意力机制。注意力机制能够严格选择输入的某些部分或为输入的不同部分分配不同的权重<sup>[50]</sup>，由此，针对于注意力机制的引起方式，可以分为两类，一种是非自主提示，另一种是自主提示。其中非自主提示指的是由于物体本身的特征十分突出引起的注意力倾向，自主提示指的是经过先验知识的介入下，对具有先验权重的物体引起的注意力倾向。从域对注意力机制来进行划分，可以分为通道域、空间域和混合域注意力。

2017年，《Attention Is All You Need》论文首次提出注意力机制的思想，提出一种新的简单的网络架构 Transformer，完全避免循环和卷积。提出的自注意力机制为键值对模式，根据以上参数进行一系列矩阵运算得到自注意力特征 Attention，计算公式如下：

$$Q = W_q \times I \quad (3.2)$$

$$K = W_k \times I \quad (3.3)$$

$$V = W_v \times I \quad (3.4)$$

$$\text{Attention}(Q, K, V) = \text{soft max} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (3.5)$$

式中  $I$  代表输入， $Q$  代表 query，后续会和每一个  $K$  进行匹配， $K$  代表 key，

后续会被每个  $Q$  匹配,  $V$  代表提取得到的信息,  $w_q$ ,  $w_k$ ,  $w_v$  为变换矩阵参数, 通过网络训练得到, 并且参数共享<sup>[51]</sup>。

多头注意力机制在自注意力机制的基础上改进, 能够联合来自不同 head 部分学习到的信息。将  $Q$ ,  $K$ ,  $V$  均分为  $n$  份,  $n$  为 head 的数目, 每一个 head 都进行一次上述公式的计算, 即每一个 head 有一个自注意力模块, 将每个 head 得到的结果进行拼接, 将拼接后的结果通过  $w^o$  进行融合, 再经过 Dropout 层得到多头自注意力模块, 计算公式如下:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^o \quad (3.6)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (3.7)$$

其中,  $w^o$  是一个可学习参数,  $w_i^Q$ ,  $w_i^K$ ,  $w_i^V$  参数通过映射得到, 也可直接进行均分得到, head 表示进行自注意力模块的数目。

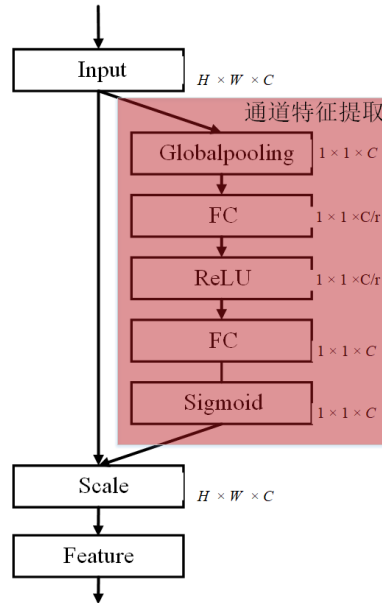


图 3.1 SE 注意力机制网络结构

SE 注意力模块出自于 2017 年 SENet 网络, 斩获 ImageNet 比赛的冠军<sup>[52]</sup>。如图 3.1 所示, 其基本工作原理是借助一个权重值向量空间, 从一个通道域的角度, 对显示图像的不同位置赋予不同的权重值, 获取更重要的与外部特征所关联的信息。首先对输入特征图进行全局池化, 将特征图变成  $1 \times 1 \times$  通道数, 然后连接全连接层和激活函数, 对  $1 \times 1 \times$  通道数的特征图进行调整, 变成每一个特征图

的权重，最后与输入的特征进行相乘。SE 模块即插即用，属于较早的将注意力引入到 CNN 中，并进行模块化设计，缺点是没有引入空间信息。

CBAM 注意力机制由通道注意力机制和空间注意力机制组成<sup>[53]</sup>，是一个简单而有效的前馈卷积神经网络注意力模块。如图 3.2 所示，给定一个中间特征图，模块沿着通道和空间维度顺序推理注意力图，然后将注意力图乘以输入特征图以进行自适应特征细化。

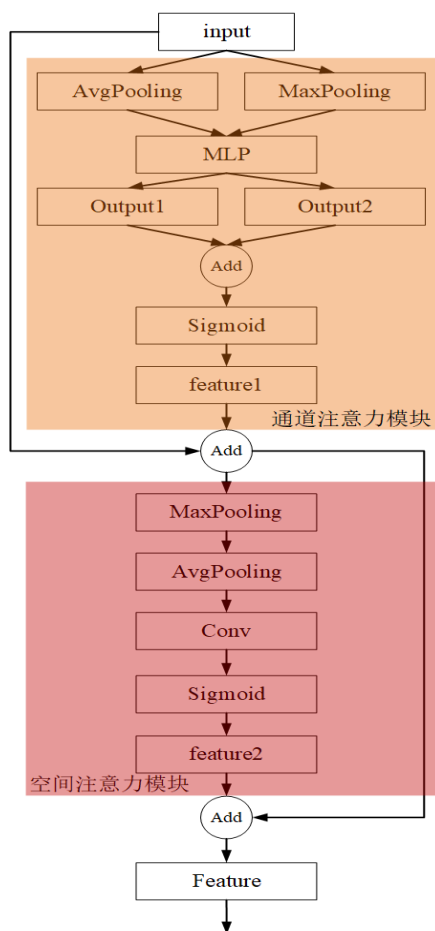


图 3.2 CBAM 注意力机制网络结构

CA (Coordinate Attention) 坐标注意力机制<sup>[54]</sup>基于 SE 和 CBAM 改进而来，作者认为 SE 模块没有考虑空间信息，CBAM 模块通过对每个位置的通道上进行池化，经多层卷积和下采样后，特征图的每个位置映射回原图，只包含原图的一个局部区域，这种作法只将局部区域信息考虑在内。为此，作者提出了 CA，CA 通过将位置信息嵌入到通道注意，避免了引入较大的开销，从而使神经网络能够



获得更大区域的信息。由于引入 2D 全局池化会使得位置信息损失，提出分解通道注意力为两个并行的 1D 特征编码以有效地将空间坐标信息集成到生成的注意力图中。

## 3.2 多模态决策网络架构

工程机械作业环境、工况复杂多变，大部分工程机械运用于矿山、高原、南北极等环境比较恶劣的地形以及地震、泥石流等救灾抢险的恶劣工况中，较难获取道路特征，控制系统与汽车有较大差异，汽车的自动驾驶技术无法直接移植。针对上述问题，本节采用双目相机，并将多种图像信息进行融合，同时引入注意力机制，构建了一种基于多模态的工程机械自动驾驶决策网络。本文提出的多模态决策网络整体架构可分为语义分割信息提取子网络、RGB 图像信息提取子网络。语义分割信息提取子网络主要使用轻量级语义分割模型 BiSeNet<sup>[55]</sup>作为整体框架，backbone 使用 ResNet18，同时在 backbone 网络中嵌入坐标注意力机制提升决策方法的性能；RGB 图像信息提取子网络尝试过 EfficientNet<sup>[56]</sup>、ConvNext<sup>[57]</sup>等一些比较流行的分类网络，效果不如经典的 ResNet 网络，故最终也选择 ResNet18 作为基准网络。

### 3.2.1 嵌入坐标注意力机制的 ResNet18 网络

ResNet 是 2016 年何凯明提出的用于计算机视觉任务所设计的一个模型。通过在卷积神经网络中引入残差连接可以有效解决深层网络的退化问题，同时还能有效缓解神经网络模型训练过程中出现的梯度消失问题。随着端到端自动驾驶决策方案的兴起，大量的计算机视觉领域的模型被迁移至自动驾驶汽车领域，但在工程机械智能化领域还应用较少，工程机械复杂的环境、工况，对于端到端决策模型而言，需要提取的周围环境的信息更多，对模型的鲁棒性要求更强。同时相较于汽车，工程机械行驶速度较低，对于算法运行速度要求不高，有着更大的空间来改善网络结构，更好的实现精度和速度的权衡。

工程机械驾驶场景复杂多变，大多为非结构化道路，需要模型提取的信息众多，并非所有的信息都同等重要。受注意力机制的启发，本文引入坐标注意力模块 CA 到残差网络 ResNet18 中，称为 ResNet-CA。ResNet-CA 由 4 组残差块组成，每组残差块包括 4 个 3×3 卷积，称为 Res18-Residual，CA 模块加入到后 3 组

残差块当中。带 CA 模块的 ResNet18 网络部分结构如图 3.3 所示。

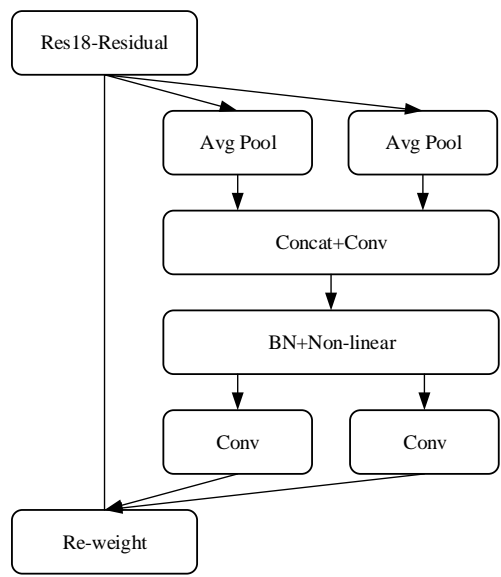


图 3.3 带 CA 模块的 ResNet18 网络部分结构

坐标注意力模块 CA 模块先分别在水平方向和垂直方向进行平均池化得到两个 1 维向量，同时在空间维度上拼接并且通过  $1 \times 1$  卷积来压缩通道维度，再通过 BN 层<sup>[58]</sup>和 Non-linear 层来编码垂直方向和水平方向的空间信息，接下来各自通过  $1 \times 1$  卷积得到和输入一样的通道数，最后通过 Sigmoid 激活函数进行归一化加权。

### 3.2.2 多模态特征融合网络

RGB 图像可以提取丰富的颜色、纹理特征，但工程机械自动驾驶的周围环境复杂多变，仅依赖 RGB 图像特征不足以理解自动驾驶周围的驾驶场景。场景的语义信息可以更好的表达周围环境对象之间的关系，研究显示场景的语义信息对于理解自动驾驶的周围环境有着不小的帮助。网络的融合主要可以分为三种：前融合、后融合和特征级融合，前融合是数据层面的融合，将点云和图像等传感器原始信息通过投影等方式直接融合，存在着较大的信息损失，难以做到较为精准的时间和空间对齐；后融合是对各自算法输出的结果进行融合，在复杂场景中需要较为复杂的后处理；特征级融合是在网络结构层面进行融合，无需经过任何处理。因此本文摒弃复杂的前融合和后融合，进行了特征级融合，结合 RGB 图像特征和语义信息提出多模态特征融合网络。

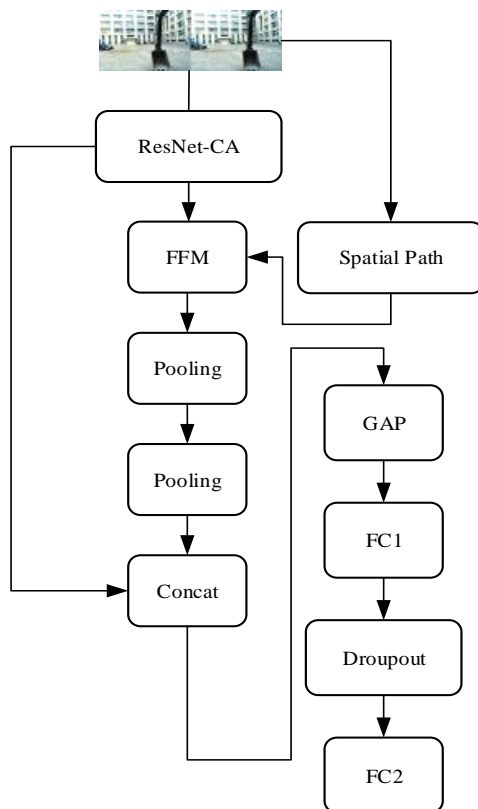


图 3.4 多模态特征融合网络

本文搭建多模态特征融合网络主要由语义分割信息提取子网络、RGB 图像信息提取子网络组成。语义分割信息提取子网络使用语义分割轻量级模型 **BiseNet** 作为基本框架，backbone 使用前文所述的 **ResNet-CA**，在保证决策模型实时性的同时，可以为模型提供丰富的语义信息。**RGB** 图像信息提取子网络由 **ResNet-CA** 直接引出提供图像 **RGB** 特征，并与语义分割子网络获得的特征图进行拼接。

图 3.4 是所提的多模态特征融合网络，使用带 **CA** 模块的 **ResNet18** 作为 **BiseNet** 的 backbone，通过特征融合 **FFM** 模块拼接空间路径分支 **Spatial Path** 提供语义分割特征图。**FFM** 模块为一个较为浅层的全卷积网络，首先是一个拼接操作融合输入，后接一个批归一化层平衡特征的尺度，接着类似于前文所述的 **SE** 注意力模块，把相连接的特征池化为一个特征向量，并计算权重向量，重新加权特征，完成特征选择和结合，具体网络结构如图 3.5 所示。

**Spatial Path** 也由多组卷积组成，包含 3 个卷积层以及对应的 **BN** 层、**ReLU**

激活函数，捕获空间信息的同时编码了绝大多数的丰富细节信息。最终获得的语义分割特征图再通过池化层进行图像尺寸的调整，与带 CA 模块的 ResNet18 输出的 RGB 特征进行多模态特征融合，通过全局平局池化、Droupout 层、全连接层进行处理，保证特征的充分融合，并且缓解了模型的过拟合。

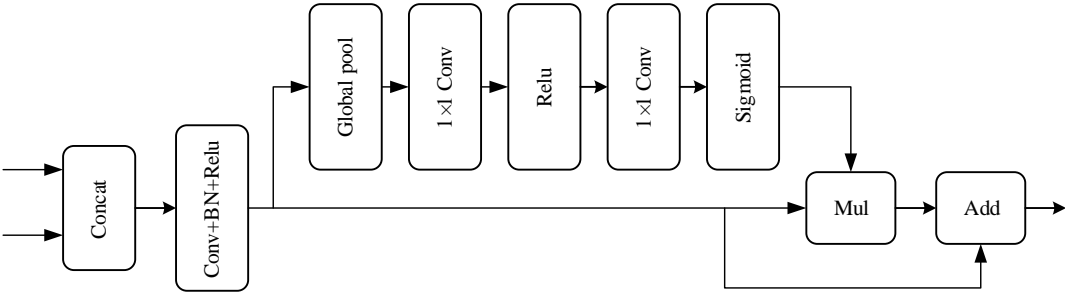


图 3.5 FFM 模块

### 3.3 图像决策网络的试验设计与结果分析

本小节将介绍所使用的数据集和网络训练所使用的环境及参数，在自动驾驶的开源数据集和真实场景的数据集上分别对提出的模型进行验证，并分析试验结果。

#### 3.3.1 数据集介绍

本文所使用的开源数据集 DBNet 数据集由上海交通大学和厦门大学联合创建，提供了大规模、高质量的点云数据、视频以及驾驶者的行为，分为训练集、验证集和测试集，包含超过 200 km 的真实世界驾驶数据。本文只采用该数据集一部分，该数据集为指定的挑战数据集：数据集共 8893 张，并且已经划分了训练、测试数据集，训练集 6549 张，测试集 2344 张。

本文所采用的特定场景的真实数据集由华侨大学的某试验大楼空地采集，图 3.6 为履带式挖掘机数据采集的试验场地，道路类型为非结构化道路，与工程机械的真实作业场景类似。为保证图像和激光雷达数据正常，采集工作在白天进行，光照正常，激光雷达和双目相机固定安装在电动履带式挖掘机上，采集时保证时间和空间上的同步。



图 3.6 履带式挖掘机数据采集的试验场地

该数据集共先后采集大约 4000 多帧数据，包含激光雷达、双目图像和履带式挖掘机的左右履带、动臂、斗杆等的 CAN 总线信号，为方便叙述，统称为控制信号。对于履带式挖掘机的控制信号，由于试验暂时不涉及自动作业和泊车的任务，只使用了履带式挖掘机的左右履带前进的控制信号，并以此作为激光雷达和双目图像每一帧数据对应真值。划分的数据集如表 3.1 所示：

表 3.1 数据集划分

	训练集	验证集
总帧数	3248	1135
占比	74%	26%

### 3.3.2 网络训练配置

本文训练试验使用台式机电脑，电脑硬件详细信息如表 3.2 所示：

表 3.2 计算机硬件信息

名称	信息
CPU	Inter Corei7-11700K-3.60 GHz
GPU	NVIDIA GTX3090 24 GB
CUDA	CUDA11.1

对于开源的数据集和非结构化真实场景数据集，为保证对比，训练和验证使用超参数尽量保持一致。训练中优化器选用 AdamW<sup>[59]</sup>，学习率采用余弦周期退

火<sup>[60]</sup>，初始学习率设置为 0.0001，批次设定为 8。

神经网络的训练依赖于标签的制作，本文选用控制信号预测值和真实值的误差作为标准。作为一个回归任务，本试验中，直接使用最为常见的 MSE 作为损失函数，同时，为了更好的评估网络，引入容差阈值对网络的精度进行分析，容差阈值是认为模型预测正确的情况下，预测值与真实值在设定的阈值范围内所能容纳的最大偏差<sup>[61]</sup>。网络精度计算公式如下：

$$I = \begin{cases} 1 & (|p_{pred} - p_{true}| < M) \\ 0 & \text{其他} \end{cases} \quad (3.8)$$

式中， $P_{pred}$  代表网络的预测值， $P_{true}$  代表数据集的真实值，在开源数据集中， $M$  代表转向角和速度阈值；在自定义的非结构化真实场景数据集中， $M$  代表履带式挖掘机左、右履带控制信号阈值。

### 3.3.3 基于开源数据集的试验结果分析

该数据集需要预测转向角和速度，故进行一定的预处理；由于自动驾驶任务的特殊性，数据增强并不能产生很好的效果，甚至一些数据增强可能会破坏图像原有的空间位置信息，因此本文只对训练和验证数据集进行标准化和归一化操作，加快决策网络的收敛。

本文以均方误差作为客观标准来评价神经网络的优劣，并以神经网络的预测值与真实值之间的误差作为主观评价标准来衡量预测值与真实值之间的偏差，同时引入容差阈值在一个批量图像数据内对转向角和速度预测进行定量计算。

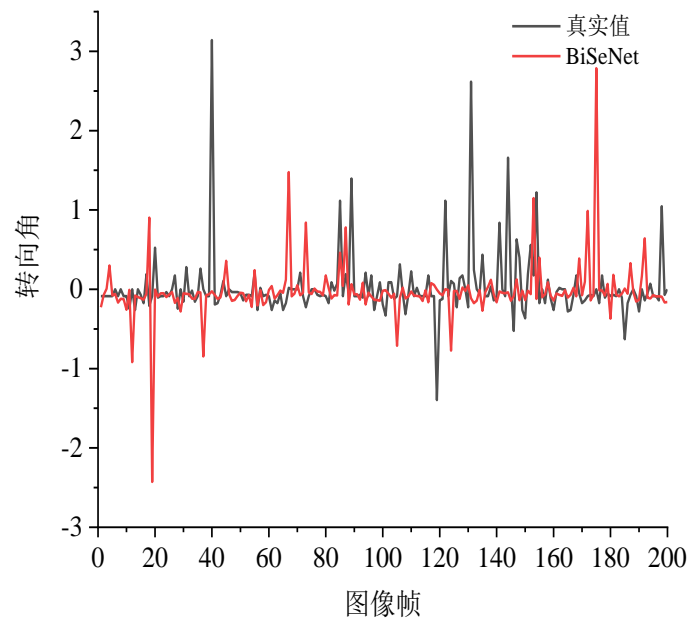
如表 3.3 所示，是多种网络模型在开源数据集上的多个评估指标的对比，表中的平均损失是整个数据集验证集的 MSE 损失，是方向盘转角和速度两个损失函数误差之和。从表中可以看出经过改进的模型在验证集上的损失值远低于 ResNet18 和 BiSeNet 两个 Baseline 模型，表明所搭建的多模态特征融合网络和嵌入的坐标注意力机制有利于模型的收敛，有着更强的泛化能力。

对模型的方向盘转角和速度进行预测比较，计算出模型在 DBNet 挑战数据集验证集转向角在 20 度容差阈值内，速度在 15 km/h 容差阈值内的平均精度。所提模型有着更高的一个精度，其中 ResNet18+BiSeNet 模型对于速度预测更为准确，达到了 93.72%，而 ResNet18+BiSeNet+CA 模型则在整体上更优，对转向角预测更为准确，最大误差更低，表明该模型预测性能更为稳定。

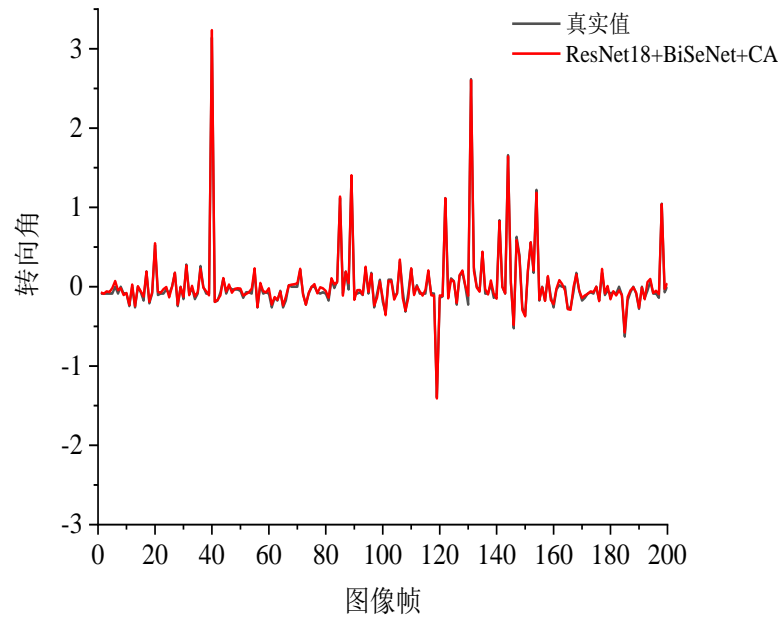
表 3.3 图像决策网络在开源数据集的试验结果对比

模型	转向角和速度平均精度(%)		转向角和速度平均误差		转向角和速度最大误差		平均损失
ResNet18	80.96	84.98	3.61	2.10	18.00	12.60	0.011819
BiSeNet	76.04	84.56	4.68	2.27	53.12	128.87	0.034045
ResNet18+BiSeNet	88.05	93.72	2.09	0.79	12.62	4.26	0.002321
ResNet18+BiSeNet+CA	91.04	91.21	1.59	1.17	10.60	4.83	0.003074

为了更为直观的对各个模型的预测性能进行对比, 选用验证集连续 200 多帧图像, 做出如图 3.7 所示转向角路径跟随曲线图。可以看出相对于最初的模型, 经过改进之后的模型对于验证集连续的图像, 归一化后的转向角和速度的整体预测量都更为准确, 对于突变的值也能预测的很好, 跟随曲线基本吻合。



(a) BiSeNet 的转向角跟随曲线



(b) ResNet18+BiSeNet+CA 的转向角跟随曲线

图 3.7 开源数据集转向角跟随曲线

### 3.3.4 基于非结构化真实场景数据集的试验结果分析

非结构化真实场景的控制信号数据集由同一时间轴的视频信息和控制信号构成，将分段采集的数据分为两部分，第一部分为训练部分，第二部分为验证部分，同自动驾驶汽车 DBNet 数据集一样，只对训练和验证数据集进行标准化和归一化，同时引入容差阈值进行评估。为验证所提决策网络的有效性，保证网络训练超参数和在 DBNet 数据集使用的一致，训练中优化器选用 AdamW，学习率采用余弦周期退火，初始学习率设置为 0.0001，批次设定为 8，选择均方误差为损失函数。

如表 3.4 所示，是多种网络模型在非结构化真实场景数据集上的多个评估指标的对比，表中的平均损失是整个数据集验证集的 MSE 损失，表征挖掘机左右履带控制信号的误差之和。从表中可以看出经过改进的模型在特定场景的验证集上的损失值、平均误差、最大误差均远低于 ResNet18 和 BiSeNet 两个 Baseline 模型，表明所搭建的多模态特征融合网络有着强大的泛化能力，能够适应不同的数据集。同时左右履带控制信号平均精度相差不大，说明本文所搭建的多模态特

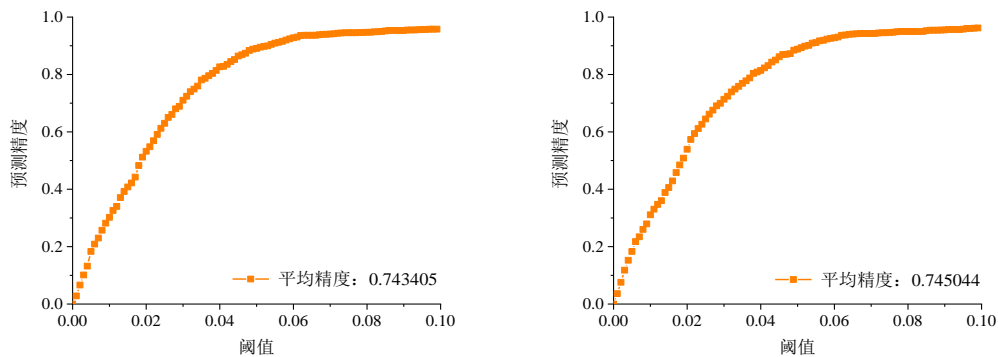


征融合网络比较稳定，数据的采集较为合理。

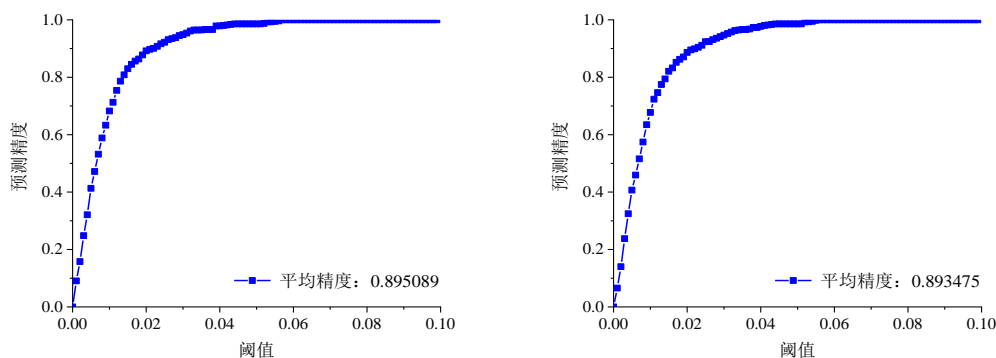
表 3.4 图像决策网络在真实场景数据集的试验结果对比

模型	左右履带控制信号平均精度(%)		左右履带控制信号平均误差		左右履带控制信号最大误差		平均损失
ResNet18	84.46	84.11	0.02	0.01	0.12	0.12	0.002204
BiSeNet	74.34	74.60	0.03	0.03	0.14	0.15	0.005348
ResNet18+BiSeNet	87.35	86.05	0.01	0.01	0.08	0.09	0.001282
ResNet18+BiSeNet+CA	89.51	89.35	0.01	0.01	0.06	0.06	0.000749

图 3.8 分别是验证集归一化之后的左、右履带控制信号预测精度曲线图，左、右履带控制信号容差阈值为 0.1。在阈值为 0.02 时，最初的 BiSeNet 模型左右履带验证集精度均只有 40% 左右，而经过改进的模型精度达到了 80% 至 90%，有着巨大的提升；整体上，左、右履带控制信号在容差阈值范围内的平均精度也有接近 15% 的提升，改进的模型对于左、右履带控制信号的预测已经较为准确。



(a) BiSeNet 的左、右履带控制信号预测曲线



(b) ResNet18+BiSeNet+CA 的左、右履带控制信号预测曲线

图 3.8 左右履带控制信号预测精度曲线

为了更为直观的对各个模型的挖掘机左右履带预测性能进行对比, 选用测试集连续 1000 多帧图像做出如图 3.9 所示的转向角路径跟随曲线图。所使用的履带式挖掘机与汽车不同, 左右履带的控制信号代表着速度, 而控制信号的差值表示转向角的大小, 故只例举转向角路径跟随曲线图。可以看出, 最初的 BiSeNet 模型只在连续帧的起始阶段有较好的跟随效果, 在中期和后期阶段, 不能很好的跟随, 出现较大误差; 而经过改进的模型, 无论是在整体上, 还是在突变的情况下, 都要优于最初的 BiSeNet 模型。

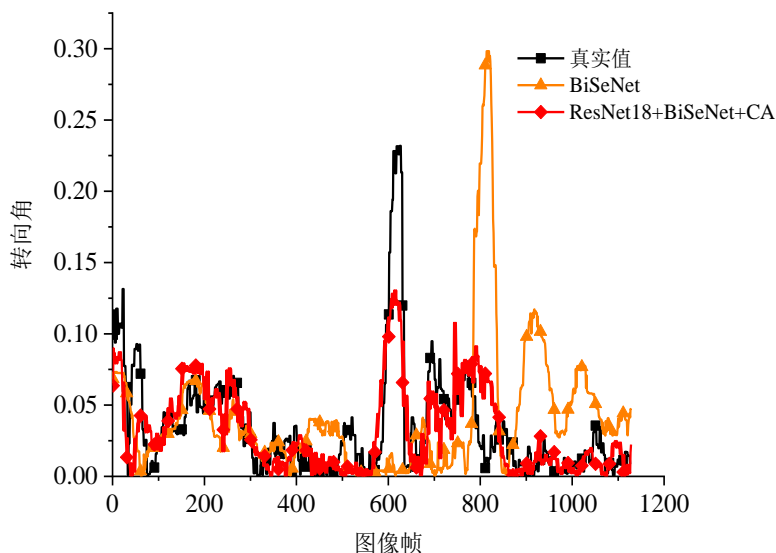


图 3.9 左右履带路径跟随曲线

### 3.4 本章小结

本章先介绍了搭建决策网络所需的卷积神经网络和注意力模块，对所使用的履带式挖掘机行走决策方案进行了分析。接着针对履带式挖掘机独特的作业环境和控制系统，搭建了基于多模态的双目图像决策网络，并嵌入注意力机制提升网络精度，最后基于开源自动驾驶数据集和自己采集的真实场景数据集完成了训练和验证。试验结果表明，无论是在开源数据集还是非结构化真实场景数据集，本文搭建的模型在最初的基本模型之上都有较大提升，基本满足对于履带式工程机械左右履带控制信号预测的需求。

## 第4章 基于激光雷达点云的工程机械自动驾驶决策

上一章利用双目图像基于开源数据集和非结构化真实场景数据集进行了试验。工程机械作业场景复杂多变,仅依靠图像难以应对所有情况,本章以激光雷达作为主要传感器,先介绍点云常用的相关算法,参考搭建决策网络,并嵌入残差连接,基于开源数据集和非结构化真实场景数据集对所提网络进行验证,并与双目图像决策网络进行对比。

### 4.1 点云相关算法介绍

激光雷达点云本身具有稀疏、无序的特性,每一帧原始点云数据量巨大。点云的处理比较复杂,存在的方式多样,常用的包括点云滤波、点云上、下采样等,同时利用点云进行应用的方式也多种多样,如:基于多视角投影的卷积神经网络、基于体素的卷积神经网络、基于图谱的卷积神经网络以及基于端到端的神经网络。本节内容对所要用到的相关技术进行介绍和分析。

#### 4.1.1 点云的下采样

从某种角度看,对于大规模点云,直接从原始点云中提取特征能够比较完整地保留点云的三维信息,但是鉴于点云本身的稀疏性和无序性,需要进行各种邻域搜索操作,从而需要高计算成本直接处理点云。一种常见的解决方案是对点云进行下采样,将所有原始点云的操作转化为下采样得到的关键点,以减少计算量。或是进行点云曲面重建时,所获得的点云数量稀缺,则要对点云进行上采样操作增加点云数量,以便更好的计算曲面特征。常用的点云采样方法包括体素下采样、随机下采样、均匀采样、曲率采样等。

体素下采样是把三维空间体素化,然后在每个体素里采样一个点,通常可用中心点或最靠近中心的点作为采样点。一般最常用的体素下采样方法是体素化网格的方法,即基本保留了点云三维空间信息的同时,减少点的数量,保持点云的形态特征基本不变。体素下采样效率较高,通过控制网格的大小,可以把点和点之间的距离控制住,但采样点个数无法精确控制,应用较多的算法如点云配准、曲面重构、形态识别等。

均匀采样指每隔固定的点数采样一次，样本按点的顺序执行，始终选择从第 1 个点开始，而不是随机选择。均匀采样的原理类似于体素化网格采样方法，同样是将点云空间进行划分，常用做下采样，其中最常见的是最远点采样(FPS)，大致步骤是：首先随机初始化起始采样点，并添加到采样点集，接着计算其余点和起始点的距离，选择距离最远的点添加到采样点集，重复前面过程，直到选择的点数量满足要求。相比于体素的采样方法，点云均匀采样后的点数是固定可控的，但算法复杂度较高、效率低。

随机下采样即在原始点云中随机采样定点数的点。随机采样非常简单，首先指定下采样的点数，然后进行随机点去除执行采样操作，可借助 Open3d、PCL 等工具直接调用，可以自定义随机采样的方式，更加灵活。随机下采样能控制输出点云的数量，同时能自定义实现，方便部署到履带式挖掘机上，故本文选择随机采样作为处理、搭建点云网络的方式之一。

#### 4.1.2 点云神经网络

近年来，越来越多的人关注点云学习，点云学习在计算机视觉、自动驾驶、机器人等领域得到普遍应用。深度学习被认为是领先的人工智能技术，已成功应用于解决各种二维视觉问题。但是点云深度学习仍处于起步阶段，并且鉴于深度神经网络在处理点云时面临的独特挑战，直至最近才出现了几种解决问题的方法。主要可以概括为 3 种方法：Projection-based 的方法、Voxel-based 的方法以及 Point-based 的方法，表 4.1 是 3 种方法的具体对比。

Projection-based 即基于投影的方法，类似于 2D 图像的处理思路，将点云投影为 2D 图像，然后进行处理。基于投影的点云处理主要可以分为两种，第一种是构建俯视图(BV)，构建方式是将点云栅格化，形成三维栅格，每一层栅格作为通道，再加上反射率(Intensity)和密度(Density)的信息，即每一层栅格都是这个栅格内雷达点的最高高度；二是构筑前视图(FV)，将雷达点云投射到柱坐标系，再栅格化，在柱坐标系中形成二维栅格，构造出高度、反射率、密度等信息。MVCNN<sup>[62]</sup>是基于投影方法一个开创性的工作，简单地将多视图特性汇集到一点，在视图池化层进行聚合后，将结果送入卷积模块，最终完成对点云的处理。

Voxel-based 即基于体素化的方法，通常将点云体素化为三维网格，然后将卷积神经网络应用于体素表示进行形状分类。比较经典的网络有 VoxelNet<sup>[63]</sup>、Second<sup>[64]</sup>、Pointpillars<sup>[65]</sup>等，基于体素的方法表现出相对较好的性能，但由于体

素的稀疏性,该方法需要消耗大量内存,会导致网络在大量非覆盖区域内进行卷积,从而造成大量不必要的计算。

**Point-based** 即基于原始点云的方法,直接对原始无序点云进行操作,无需任何处理。鉴于三维点云固有的数据无序性,典型的二维图像深度学习方法不能直接应用于三维点云。最具代表性的如 **PointNet** 系列<sup>[66]</sup>, **PointNet** 直接以点云作为输入,利用数个 **MLP** 层独立学习点云空间特征,利用最大池化层对全局特征进行提取。本质上, **PointNet** 在一点或所有点上运行,并没有从点云中提取局部特征,很难学习精细特征,在分割和检测方面有局限性,适用于小范围内的点云分类。由此 **PointNet++**<sup>[67]</sup>应运而生, **PointNet++**核心理念在局部区域重复迭代使用 **PointNet**,在小区域选用 **PointNet** 形成新的点,新的点定义新的小区域,进行多层次的特征学习。

简而言之,基于点的方法直接对原始点云进行处理,不进行任何体素化或投影,不会丢失显式信息,适用场景广泛、步骤简单、适合部署,因此本文选择了基于原始点云的方法进行决策网络设计。

表 4.1 点云学习方法对比

点云神经网络	基于多视角投影	基于体素	基于原始点云
适用场景	三维建模识别	三维建模识别	三维建模识别、三维模型语义分割
优势	数据规模小,可直接利用传统、深度学习网络	结构规则,仅需对传统、深度学习网络进行维度扩展	步骤简单,可通过三维扫描仪直接获得
劣势	图像的表达方式难以完全表达三维的几何细节	计算代价大,难以表示高精度的三维模型	缺少点之间的信息,对三维表面细节表示能力不足

## 4.2 基于激光雷达点云的网络决策架构

### 4.2.1 基于 X-Conv 搭建的自定义点云网络架构

本节搭建的点云决策网络结构参考 PointCNN<sup>[68]</sup>网络，PointCNN 网络类似于 PointNet 系列，直接使用点云原始数据，借鉴在计算机视觉的二维卷积网络处理规则数据，实现了对无序点云的处理，整体结构如图 4.1 所示：

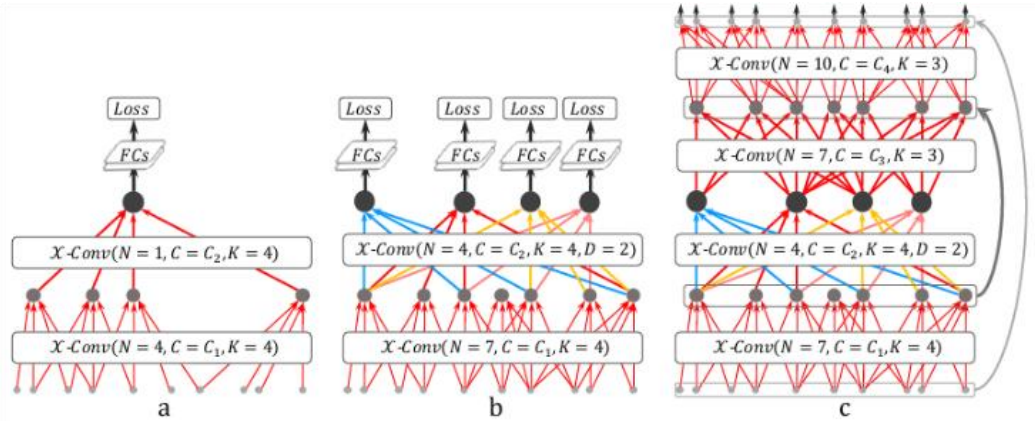


图 4.1 PointCNN 网络整体结构

PointCNN 网络的核心之一是 X-Conv，X-Conv 和普通卷积的不同点有两个方面：其一，局部特征的提取方式，在图像中是一个像素区域，而在点云中是  $K$  近邻点；其二，是局部信息的学习方式，X-Conv 不仅对点云自身特征提取，也对通道维度、空间信息进行处理，实现了数据的规则化。此外，X-Conv 的计算量与  $K$  有关，而不是普通二维卷积的二次增长和三维卷积的三次增长。

X-Conv 的输入是前一点的点集  $P$  和对应的特征集  $F$ ，输出的更加具有表达性的点  $p$ ，以及  $X$  变换矩阵  $K$ ，输出是点  $p$  的特征集  $F_p$ 。具体的步骤为：

- (1) 将输入点集转化到相对于输出点的局部坐标系，这是因为 X-Conv 是设计来对局部区域的点云进行卷积的，输出不应该受点的绝对位置影响；
- (2) 输出的特征集不仅仅包含输入特征集，还应包含局部的坐标关系，因此使用多层感知机 MLP 将局部坐标升维到  $C_\delta$ ，得到局部特征集  $F_\delta$ ；
- (3) 串接输入特征集  $F$  和局部特征集  $F_\delta$ ；
- (4) 再次通过多层感知机 MLP 学习得到  $X$  变换矩阵；
- (5) 使用  $X$  变换对串接的特征进行加权和变换，得到规范表达的特征集合；
- (6) 在规范表达的特征集合上执行普通卷积。

本节基于 X-Conv 搭建激光雷达点云决策网络，对于使用的 X-Conv 组数，进行了探索，最终考虑到网络模型大小和精度，使用了 7 组卷积操作，探索的卷积网络结构如下表 4.2 所示。整个点云决策网络结构最深的维度为 512，升至最高维度后与图像决策网络保持统一，直接通过两个全连接层迅速降维，保证较小的计算量，最后取均值调整维度完成输出。

表 4.2 点云网络组数

阶段	类型	输入尺寸	输出尺寸
0	X-Conv	8x128x3	8x128x32
1	X-Conv	8x128x32	8x128x64
2	X-Conv	8x128x64	8x128x96
3	X-Conv	8x128x96	8x120x128
4	X-Conv	8x120x128	8x120x192
5	X-Conv	8x120x192	8x120x256
6	X-Conv	8x120x256	8x120x512
7	FC	8x120x512	8x120x256
8	FC	8x120x256	8x120x2
9	Mean	8x120x2	8x2

#### 4.2.2 嵌入残差连接的点云决策网络

自定义的网络结构，网络层数较深容易发生梯度消失和梯度爆炸，为了缓解这种现象，便于梯度的传播，在自定义网络架构中嵌入残差连接的操作，将浅层特征连接到深层特征之中。

使用了常用的残差连接操作 Concat 和 Add，将这种嵌入残差连接操作的网络结构称为 PointCNN+ShortCut，具体网络结构如图 4.2 所示，首先使用两组 X-Conv 操作提取浅层点云特征，接着保存提取的特征与深层特征使用多次 Concat 和 Add 操作，进行多次残差连接，得到最终的代表周围环境信息的特征向量，最后通过两层全连接输出最终结果。PointCNN+ShortCut 网络没有多余的分支，没有一些碎片化的操作，没有使用特殊的算子，整个架构串行运行，符合常见框架的优化，易于部署。



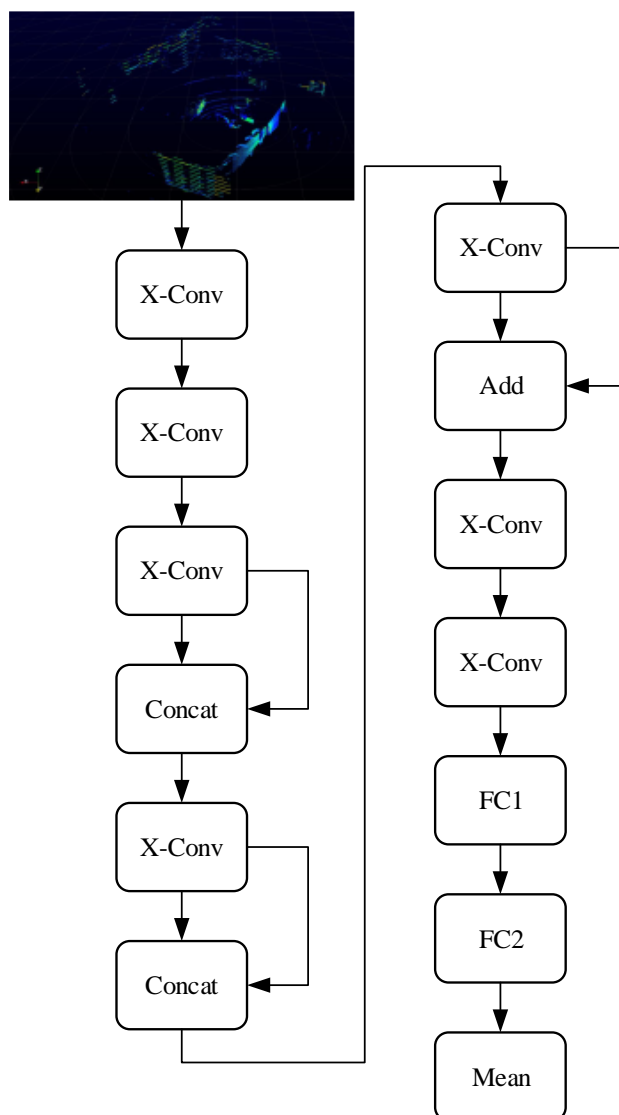


图 4.2 PointCNN+ShortCut 网络结构

### 4.3 点云决策网络的试验设计与结果分析

本小节我们将仍在 3.4.1 节介绍的数据集上对搭建的点云决策网络进行验证。为了公平比较，网络训练配置与 3.4.2 节所述保持一致，基于开源数据集和自己采集的真实场景数据集进行试验结果的对比和分析，并分析点云采样点数对网络性能的影响。

#### 4.3.1 基于开源数据集的试验结果分析

本次试验与 3.4.3 节一致, 基于 DBNet 数据集, 依然采用均方根误差评价点云决策网络, 衡量预测值和真实值之间的偏差, 并引入容差阈值。将网络在验证集上的预测结果与前文使用图像作为唯一输入的 BiSeNet 和 BiSeNet+ResNet+CA 网络进行对比。

表 4.3 点云决策网络在开源数据集的试验结果对比

模型	转向角和速度平均精度(%)		转向角和速度平均误差		转向角和速度最大误差		平均损失
BiSeNet	76.04	84.56	4.68	2.27	53.12	128.87	0.034045
ResNet18+BiSeNet+CA	91.04	91.21	1.59	1.17	10.60	4.83	0.003074
PointCNN	60.10	56.49	10.98	6.73	68.74	29.85	0.168490
PointCNN+ShortCut	68.77	68.86	6.80	4.68	50.08	28.24	0.079474

表 4.3 是点云决策网络在开源数据集的试验结果, 在 DBNet 挑战数据集上, 点云网络相对于使用图像作为输入的两种模型, 不仅转向角和速度平均精度相对较低, 转向角和速度的平均误差、最大误差也远低于图像作为输入的网络。可见对于相同大小的数据集, 在预测转向角和速度这种多任务预测量时, 以点云作为输入的网络比不上以图像作为输入的网络, 点云作为一种复杂的传感器数据, 在完成自动驾驶端到端任务的潜能还有待挖掘。

更直观地, 依据 DBNet 挑战数据集的 500 多帧连续验证集图像, 绘制在一定容差阈值范围内网络的转向角跟随曲线。如图 4.3 所示, 可以看出基于激光雷达点云作为输入的模式, 对于转向角的预测, 大多都在 0 值附近, 与真值相差较大, 基本体现不出预测效果, 不如前文所述的双目图像决策网络的跟随曲线。

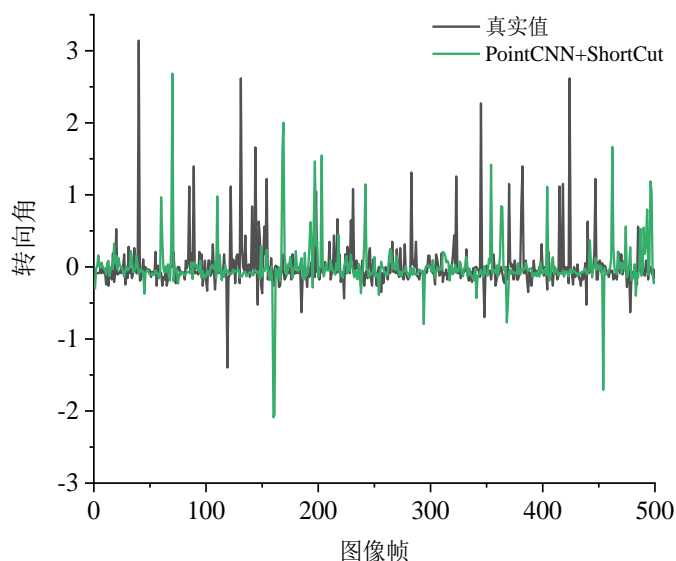


图 4.3 左右履带开源数据集路径跟随曲线

#### 4.3.2 基于非结构化真实场景数据集的试验结果分析

本小节仍然使用 3.4.1 节介绍的真实场景数据集，网络训练配置与基于双目图像的决策网络保持一致。采集的原始数据每一帧有 10 万多点，数据量巨大，为加快网络训练速度，对数据集的原始点云进行离线采样，采用随机采样的方式进行采样，每一帧保留 2048 个点。为保证训练速度，点云网络训练时默认基准使用 128 个采样点进行试验，加快网络迭代速度。

与以图像作为输入的部分网络对比结果如表 4.4 所示，相对于点云网络在开源自动驾驶数据集的表现，在真实场景数据集上，点云决策网络左右履带控制信号平均精度依然低于以图像为输入的网络，平均误差、最大误差、平均损失方面相差不大。说明对于真实场景数据集，点云决策网络相对收敛更快、模型更为鲁棒，当然，和数据集的大小、制作等有一定关系。

表 4.4 点云决策网络在真实场景数据集的试验结果对比

模型	左右履带控制信号 平均精度(%)		左右履带 控制信号 平均误差		左右履带 控制信号 最大误差		平均 损失
BiSeNet	74.34	74.60	0.03	0.03	0.14	0.15	0.005348
ResNet18+BiSeNet+CA	89.51	89.35	0.01	0.01	0.06	0.06	0.000749
PointCNN	76.26	76.10	0.02	0.02	0.11	0.11	0.003441
PointCNN+ShortCut	81.14	81.66	0.02	0.02	0.14	0.13	0.002526

为了更好的分析点云决策网络在工程机械非结构化道路的表现,挑选验证集的 1000 多帧连续图像,绘制多个网络归一化后的履带式挖掘机左右履带控制信号的路径跟随曲线。如图 4.4 所示,以激光雷达点云为输入的决策网络转向角曲线与真值曲线差值均在一定范围之内,平均精度只略低于图像决策网络,但并不能实现很好的跟随,可能和点云本身的稀疏性和无序性有关,与图像决策网络相比有着较大差距,履带式挖掘机左右履带的控制信号预测效果不佳。

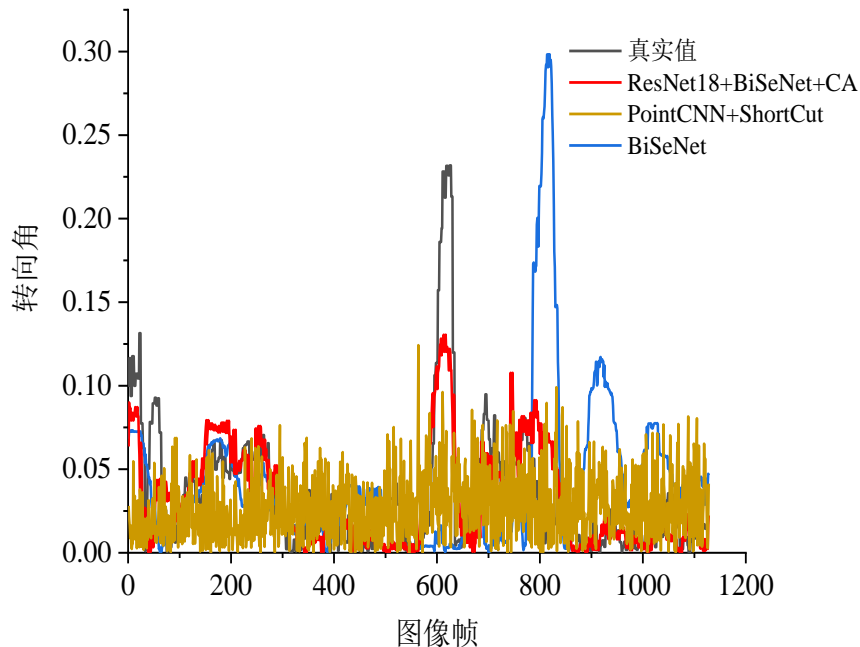


图 4.4 左右履带真实场景数据集路径跟随曲线

### 4.3.3 点云采样点数对网络性能的影响

在前文中，对于 PointCNN-ShortCut 网络，默认使用随机采样的 128 个点。本次试验，仍然采用随机采样的方式，将点云决策网络的输入尺寸限制为 16384 个点，对于出现的空间点不足的情况，采用常见的补零操作进行填充。网络的输入尺寸极大的影响网络的训练速度和推理速度，因此，本小节基于点云的采样点数对网络的影响展开试验，本次试验只使用非结构化真实场景数据集，试验结果如表 4.5 所示：

表 4.5 点云采样点数对网络的影响

模型	左右履带控制信号 平均精度(%)		左右履带控 制信号平均 误差		左右履带控 制信号最大 误差		平均 损失
PointCNN+ShortCut- 128	81.14	81.66	0.02	0.02	0.14	0.13	0.002526
PointCNN+ShortCut- 160	82.44	82.06	0.02	0.02	0.14	0.14	0.002551
PointCNN+ShortCut- 192	83.01	83.10	0.02	0.02	0.12	0.13	0.002057
PointCNN+ShortCut- 256	83.85	83.49	0.02	0.02	0.14	0.14	0.002050
PointCNN+ShortCut- 512	84.97	84.49	0.01	0.01	0.13	0.12	0.001861

从结果可以看出，随着采样点数的增加，左右履带控制信号的平均误差和最大误差基本没有变化，但点云决策网络在验证集平均精度表现越来越好。点云采样点数为 192 时损失下降最为明显，随着采样点数到达 256 和 512，网络的平均精度增长缓慢，网络训练和推理的时间会大幅增加，训练时间会成倍增长。为了平衡训练和推理效果，选择 192 点云采样点数作为最终的点云决策网络输入尺寸。

## 4.4 本章小结

本章使用激光雷达点云作为网络输入，首先介绍了处理点云数据常使用的

采样方式以及一些经典的点云端到端网络；参考并搭建了自定义的点云端到端决策网络，基于开源数据集和真实场景数据集进行测试，并与基于双目图像的决策网络进行对比。试验结果表明，以点云作为输入的端到端决策网络，无论是在开源数据集，还是真实场景数据集上，都比不上双目图像决策网络，尤其体现在对于非结构化路径的跟随任务上面。



## 第 5 章 基于非结构化道路的决策实车试验

为了进一步验证所提网络在真实场景下的性能,实现履带式挖掘机的自动驾驶行走功能,本章利用现有试验条件进行了实车测试。首先设计了某履带式挖掘机进行自动行走的试验方案,然后搭建相关试验平台,并进行必要的数据采集,最后依据试验方案完成了真实场景下的自动驾驶决策实车试验。

### 5.1 试验方案设计

为了更好的验证本文提出的决策网络的性能,本文针对典型的工程机械工况进行试验部分的设计。试验共分为两个工况:静态避障和动态避障,任务设置主要考虑功能和安全为主,在实际行驶中会设置静态障碍物和动态障碍物模拟真实避障的情况。

主要针对某特定非结构化场景,测试本文提出的网络对挖掘机控制信号的实时预测能力,不同的工况任务设计如下:

(1) 静态避障。目的地设定在起点正前方,场景中放置静态障碍物,障碍物尺寸约为 $1\times 1\times 1.5\text{ m}^3$ 大小,由于试验条件的限制,用钢板代替工程机械真实场景中常见对象,平均行驶距离 20 m 左右,测试履带式挖掘机是否能够躲避静态障碍物,并进行自主回正。

(2) 动态避障。试验场景选择场地中央,设置动态参与者,平均行驶距离 20 m 左右,测试履带式挖掘机是否能够躲避动态障碍物,并进行自主回正。

#### 5.1.1 试验原理

试验所使用的硬件主要包括双目摄像头、图为科技 T609、逆变器、远程图传模块、履带式挖掘机平台等;软件部分包含了双目图像决策网络及其相关程序。与汽车不同,本次履带式挖掘机自动驾驶端到端试验,首先规划驾驶任务,将感知模块的图像等传感器原始数据直接输入到端到端自动驾驶网络,类似于计算机视觉领域,车辆的控制信号作为网络训练的标签,经过数据集训练、验证之后,依据保存的网络参数加载网络、实时输出控制信号,实现对于车载的横、纵向控制,从而完成自动驾驶行走、自动作业等任务。图 5.1 是履带式挖掘机自动驾驶



端到端原理图。

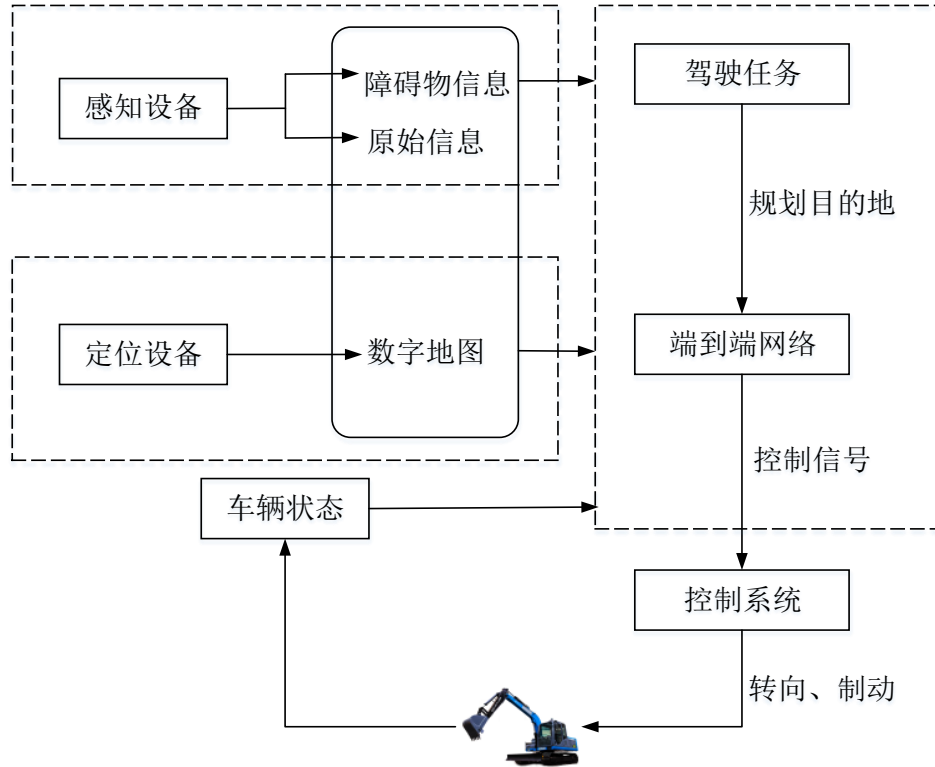


图 5.1 履带式挖掘机自动驾驶端到端原理图

### 5.1.2 试验流程

依托履带式挖掘机平台和 PCAN-View 软件采集实车驾驶数据集并完成实车试验，如图 5.2 所示，是本文试验的基本流程。首先利用所构建的双目图像决策网络对开源汽车自动驾驶数据集进行迭代训练，对网络进行初步验证，接着使用采集的真实场景数据集对所提网络进一步验证，最后利用 RMSE 评估结果是否合格，不合格则进行反复迭代训练，反复调整网络超参数，对网络进行微调，直至取得最优权重，最后在计算平台上搭建好相关算法环境和配置，将所提算法和权重移植到计算平台上。

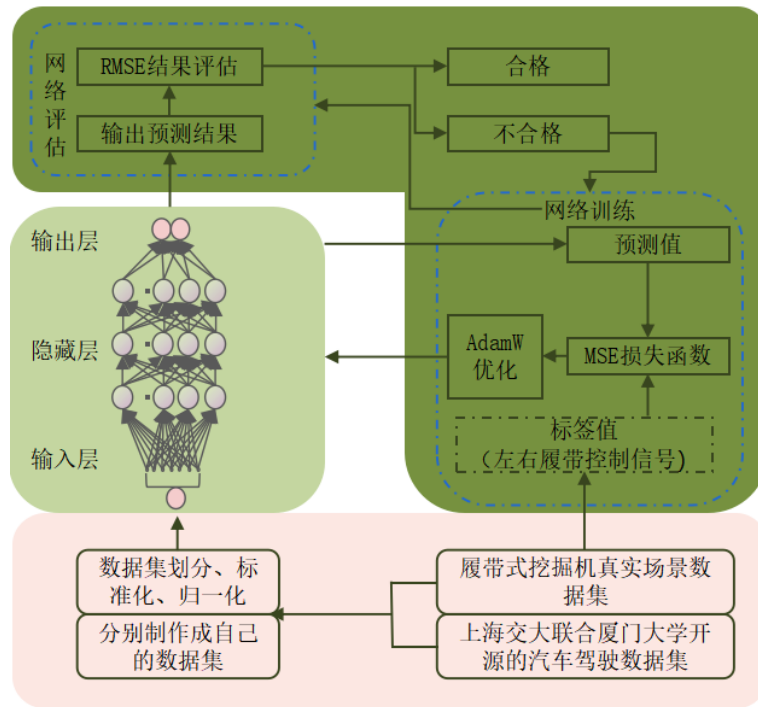


图 5.2 试验流程

履带式挖掘机在进行自动驾驶试验时首先根据权重加载模型，在行驶之前对挖掘机当前状态进行检测，先进行硬件自检，查看各传感器设备是否正常工作，若无异常则进行软件自检，检测决策网络及其相关模块等当前状态是否正常，若无异常则开始进行驾驶任务，若有异常，通过远程图传模块控制挖掘机停止，基本流程如图 5.3 所示。接着 T609 计算平台通过逆变器接电运行，USB 接口双目摄像头图像大小为  $640 \times 240$  的视频流实时传入作为算法的输入，核心算法程序运行。最后通过 CAN 总线在挖掘机行驶过程中进行通信，将决策网络输出的信号值进行换算，得到左、右履带式 CAN 信号，输入整车控制单元后，产生 PWM 脉冲信号，接收到电信号的比例减压阀输出先导压力给多路阀，从而实现最终输出。

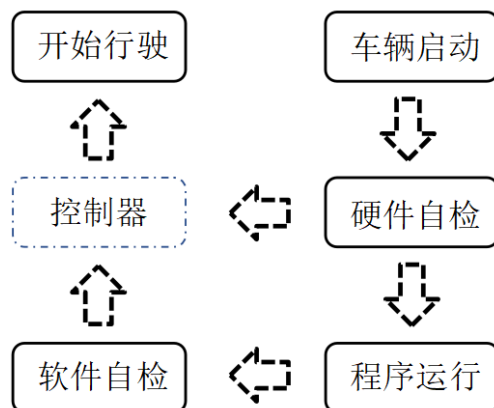


图 5.3 履带式挖掘机实时行驶流程图

## 5.2 实车平台搭建

为了满足算法的验证需求，本章基于实车试验对算法进行验证，实车平台已具有线控底盘执行系统，为试验室前期研究成果，不再作过多赘述，实车试验在某履带式挖掘机上进行。本文自动驾驶决策方法采用端到端的方式，视觉信息和激光雷达点云的信息作为网络输入，因此核心传感器为双目摄像头和激光雷达。本次摄像头采用 USB 接口高清双目摄像头，像素 200 万，最大分辨率  $1280 \times 480$ ，高清、高帧工业摄像头，专门用于深度学习、机器人视觉；雷达采用的是威力登的 128 线束激光雷达，水平视角 360 度，垂直视角 40 度，帧率 5~20 Hz，测量范围高达 245 m，且集成 Web 服务器，便于监控和配置。其次，设计了定制支架将双目摄像头和激光雷达一起固定在驾驶室前端，保证空间位置的对齐。

在整车的计算平台方面，考虑到计算的实时性，且试验需在履带式挖掘机上完成，因此要求计算平台可以配置在车辆上，具有良好的密闭性，且体积，功耗不可太大，同时需要满足神经网络计算能力和内存开销方面的需求<sup>[69]</sup>。本文对比了目前市面已出现的几款产品，如：NVIDIA Jetson TX2、NVIDIA Jetson Xavier、NVIDIA Jetson Orin 等产品，综合考虑计算平台性能及试验室当前情况，最终选择了图为科技产品 T609。图 5.4 是图为科技的 T609，是一款面向无人驾驶车载系统的计算平台，内置 NVIDIA Jetson Xavier 芯片，NVIDIA Jetson Xavier 部分参数如表 5.1 所示。同时 T609 预装 Ubuntu 18.04 操作系统，具备 32 TOPS 浮点运算的 AI 处理能力，采用超强固轻型铝合金材料设计，传导被动散热，具备优秀的散热能力，支持 USB、CAN、RS232/485、GPIO、同步信号等丰富 IO 接口

类型，内置 4G 通信模块和 WIFI 模块，支持同步信号输入输出，8 路 GMSL2 摄像头输入。预留便于坚固安装支架，具有计算能力强、可靠性高、集成度高、功耗低特点，可应用于无人清洁车、无人配送车、智能巡检、AGV 等无人驾驶领域。



图 5.4 图为科技 T609

表 5.1 NVIDIA Jetson Xavier 参数表

计算平台	Jetson Xavier
GPU	NVIDIA Volta 架构，384 CUDA 核心和 48 个 Tensor Core
CPU	6 核 NVIDIA Carmel ARMv8.2 64 位 CPU 6 MB L2+4 MB L3
视频	4K x 2K 60 Hz 编码（HEVC） 4K x 2K 60 Hz 解码（12 位支持）
内存	8 GB 128 位 LPDDR 4x 59.7 GB/s
显示器	2 个多模 DP 1.4/eDP 1.4/HDMI 2.0
PCIE	1 个 x1（PCIE3.0）+1 个 x4（PCIE4.0）
存储	32 GB eMMC 5.1
网络	10/100/1000 BASE-T 以太网

5.3 数据采集

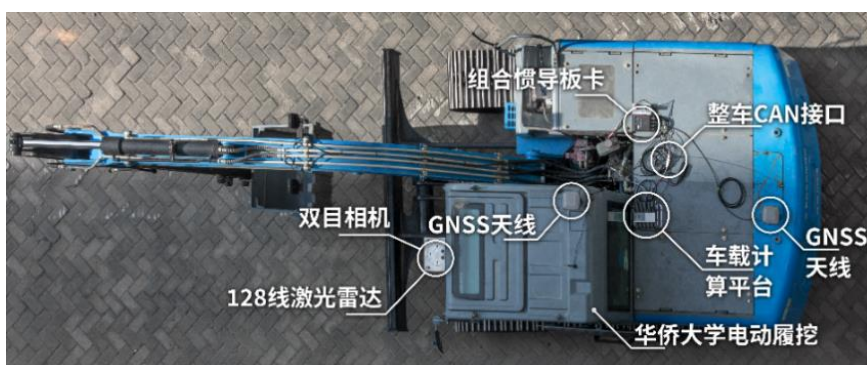
关于数据采集，在 3.4.1 节已经做了简单介绍，下面介绍详细的数据采集方案。试验场所地段路面为砖铺地面，四周为试验大楼，工程机械作业环境多变，不同作业任务场景差别较大，但都属于非结构化道路，所以该试验场景具有较大的参考意义。

根据前文所述，基于激光雷达点云决策网络的精度和控制信号跟随效果比

不上以双目图像作为输入的决策网络，故本次实车测试只使用双目相机开展试验。进行数据采集的履带式挖掘机实车平台如图 5.5 所示，采集场景图像信息、及履带式挖掘机行走控制信号构建数据集。履带式挖掘机行走时的最高行驶速度一般在每小时 10 公里以下，在作业过程中，履带式挖掘机通常保持在每小时 2 至 3 公里的低速度，本文将此数据采集过程控制在每小时 3 公里以下，按照一定履带式挖掘机的正常作业速度进行<sup>[70]</sup>。基于当前的场景，选择试验大楼一角为起点，每一趟采集的数据基本保持同一起点和同一终点，且行驶的路径、速度基本相同，以维持数据分布，采集多次静态避障和动态避障数据，并分别存储。



(a) 实车平台-轴测图



(b) 实车平台-俯视图

图 5.5 实车平台



## 5.4 工程机械决策避障试验

本节将第三章所述模型部署到载有 NVIDIA Jetson Xavier 的图灵科技计算平台。试验主要分为两部分试验内容：第一为静态障碍物避障部分，旨在测试试验挖掘机能否较好的躲避静态障碍物，并进行回正；第二位动态障碍物避障部分，旨在测试试验挖掘机是否能较好的躲避动态障碍物，并进行回正。

### 5.4.1 静态障碍物避障测试

图 5.6 是挖掘机自动行驶过程中躲避静态障碍物的部分连续图像。使用双目图像决策网络控制挖掘机，在行驶过程中机身产生微震荡，但能较好的完成直行、转弯，实现对静态障碍物的躲避并实现回正，证明本文提出的决策网络能够有效的完成车辆的横纵向控制，实现对静态障碍物的躲避。

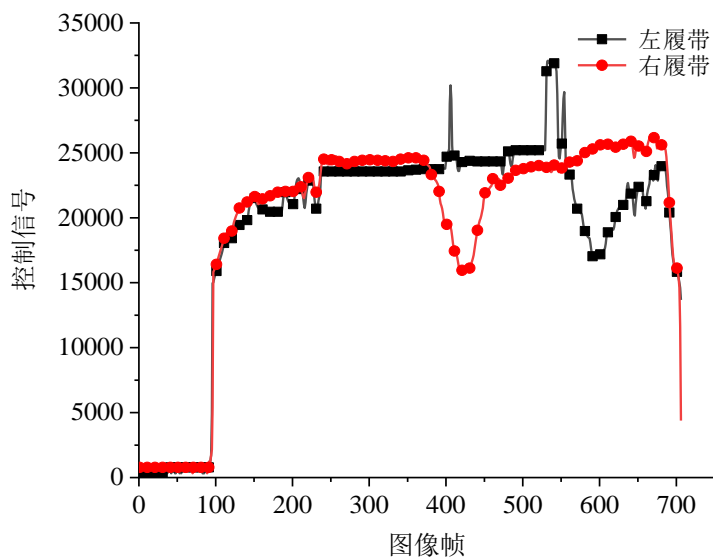
在试验过程中发现一些现象，第一，决策网络模型在实车测试时对于起始点的要求有较高的要求，如果起始点离数据采集的起始点相距较远，挖掘机行驶效果不佳；第二，如果要同时多条路径行驶躲避静态障碍物，挖掘机必须朝向所行驶路径，否则，挖掘机无法辨别该行驶哪条路径；第三，相较于转弯，所提决策网络在挖掘机直行时表现更好，可能与制作的数据集有关，数据集中直行的部分偏多。



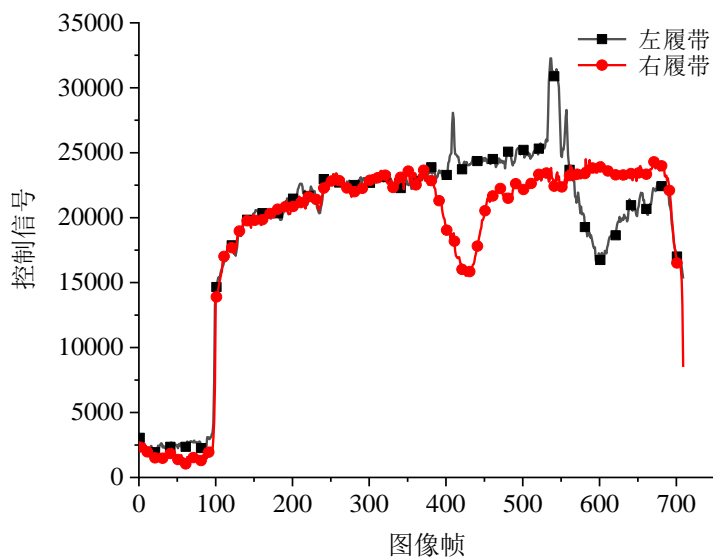
图 5.6 挖掘机静态避障过程的部分连续图像

为了更为准确的描述挖掘机自动行驶的情况，保存了数据采集时其中一趟挖掘机左右履带控制信号和挖掘机自动行驶时左右履带的控制信号，绘制成跟随曲线分别如图 5.7 所示。从跟随曲线图可以看出，刚开始挖掘机左右履带控制信号相差不大，保持直行，随后挖掘机右履带控制信号逐渐减小，挖掘机开始执行转弯。转弯完毕后，左履带控制信号逐渐减小，挖掘机开始回正，最后保持直行，直至挖掘机停止。挖掘机的行驶轨迹与数据采集时实际行驶轨迹基本一致，

且左右履带控制信号和实际行驶时相差不大，只在挖掘机开始行走有略微差异。本文提出的决策网络对于静态障碍物具有较好的实际效果。



(a) 静态避障数据采集



(b) 静态避障网络预测

图 5.7 挖掘机完成静态避障过程的自动行驶跟随曲线

### 5.4.2 动态障碍物避障测试

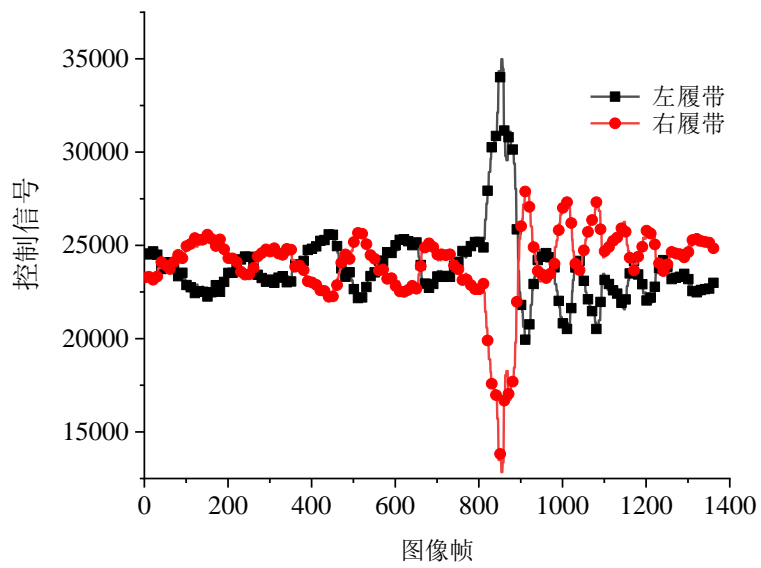
图 5.8 是挖掘机自动行驶过程中躲避动态障碍物的部分连续图像。实际测试过程中,决策网络执行,挖掘机开始行驶,保持直行,当有行人出现在双目摄像头视野范围之内,挖掘机开始转弯,行人离开摄像头视野,挖掘机开始回正,最终完成避障。在试验过程中发现,利用所提决策网络进行端到端避障并非每次都能成功。第一,与行人行走的速度和位置相关,行人与数据采集时所行走轨迹相近时避障容易成功;第二,与数据集大小有关,动态避障数据采集的趟数越多,越容易成功。



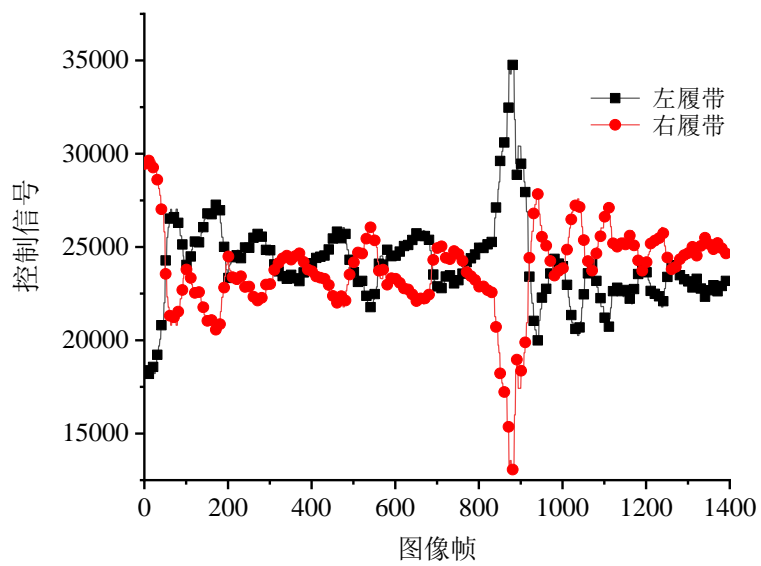
图 5.8 挖掘机动态避障过程的部分连续图像

与静态避障试验过程类似,使用采集数据中的一趟动态避障数据与挖掘机进行动态避障时左右履带的控制信号进行对比,绘制成跟随曲线分别如图 5.9 所示。从跟随曲线图可以看出,履带式挖掘机一开始保持直行,当行人出现在双目摄像头视野,左履带控制信号逐渐增大,右履带控制信号逐渐减小,直至转弯完成。随后挖掘机开始回正,左履带控制信号逐渐减小,右履带逐渐增大,直至完全回正,最后保持直行。从整体上看,挖掘机的行驶轨迹与数据采集时实际行驶轨迹大致一致,本文提出的多模态决策网络对于动态障碍物具有较好的效果,只在最开始时控制信号有较大差异,可能与履带式挖掘机自身的非线性系统有关。





(a) 动态避障数据采集



(b) 动态避障网络预测

图 5.9 挖掘机完成动态避障过程的自动行驶跟随曲线

## 5.5 本章小结

本章在某履带挖掘机上搭建自动驾驶平台,进行数据采集,并将前文所述的决策网络部署到了搭建的平台上,完成了履带挖掘机的自主避障试验。试验表明,本文搭建的决策网络能够在非结构化道路,较好的完成履带挖掘机的自动行走,并进行自主避障。



## 第6章 总结与展望

### 6.1 课题总结

随着国家战略的实施和行业变革,智能化、无人化逐渐成为工程机械行业的主要发展趋势之一。对于自动驾驶决策技术,目前主要集中在汽车领域,企业和高校大多采用传统技术;在工程机械领域,极少数大企业刚开始研究工程机械自动驾驶,工程机械智能化研究相对较少。本文针对工程机械非结构化的场景,对工程机械自动驾驶决策技术进行了研究,主要完成了以下几方面内容:

#### (1) 端到端自动驾驶决策方案研究

针对工程机械端到端方案,对目前自动驾驶主流使用的传感器进行对比和分析,根据工程机械场景、工况特点确定使用双目相机和激光雷达作为主要传感器开展研究。分析了目前自动驾驶存在的开源数据集和开源仿真平台,根据工程机械需要只选择了 DBNet 数据集作为验证。最后分析了为什么选择基于端到端来对工程机械进行自动驾驶决策,确定了最终的决策方案。

#### (2) 基于双目图像的工程机械自动驾驶决策研究

介绍了网络搭建所用到的卷积神经网络和注意力模块,对工程机械自动驾驶行走方案进行了分析。以履带式挖掘机为例,针对履带式挖掘机独特的作业环境和控制方式,搭建了基于多模态的双目图像决策网络,基于开源自动驾驶数据集和某特定场景数据集开展了试验。试验结果表明,无论是在开源数据集还是特定场景数据集,本文搭建的模型在最初的基本模型之上,转向角的预测精度、误差等指标都有较大的提升,基本满足履带式挖掘机左右履带控制信号预测的需求。

#### (3) 基于激光雷达点云的工程机械自动驾驶决策研究

分析了搭建激光雷达点云决策网络的常用采样方式和常用的端到端网络,选择了比较简单的随机采样,参考 PointCNN 网络搭建了激光雷达点云决策网络,并嵌入经典的残差连接操作提升网络精度。基于开源自动驾驶数据集和真实场景数据集进行测试,并与基于双目图像的决策网络进行对比,试验表明,激光雷达点云决策网络对于转角的预测、平均误差、最大误差等指标的表现均比不上所

提的双目图像决策网络。

#### (4) 基于非结构化道路的决策实车试验

根据实车测试需要,在某非结构化道路场景,根据工程机械典型工况设计了试验方案,基于履带式挖掘机搭建了实车测试平台,并完成了数据采集。将在 PC 机上训练好的决策网络权重保存,部署网络到带有 NVIDIA Jetson Xavier 芯片的图灵科技计算平台上,完成了某履带挖掘机的自主避障试验。试验结果表明,本文搭建的决策网络在非结构化道路上,能够较好的完成履带挖掘机的自动行走,并进行自主避障。

## 6.2 课题的创新性

(1) 针对目前主要采用单目相机作为传感器进行决策,获取信息不足,提出了一种基于多模态的工程机械自动驾驶决策网络。主要由语义分割信息提取子网络、RGB 图像信息提取子网络两部分组成,融合图像信息且尽量保证网络运行速度,同时嵌入坐标注意力机制进一步提升决策网络性能。

(2) 针对目前大多以相机来进行端到端决策,难以适应工程机械工况和环境,提出了自定义搭建的激光雷达点云决策网络。嵌入经典残差连接操作提升网络精度,并对原始点云的采样点数进行探索,实现较好的速度与精度权衡,使得以原始点云作为输入实现履带式挖掘机的行走成为可能。

## 6.3 工作展望

工程机械的作业环境、工况复杂多变,经常行驶在非结构化道路和特征不明显的山地,获取道路特征的难度比汽车的行驶环境更大,两者的整车控制系统差异较大,针对工程机械的自动驾驶研究相对较少。本文针对工程机械非结构化场景,设计了基于双目图像和激光雷达点云的端到端自动驾驶决策网络,并完成了实车避障测试。通过本论文的研究和探索,觉得今后的工作开展可从以下几个方面进行:

#### (1) 进一步完善履带式挖掘机实车试验

本文已在履带式挖掘机真实场景数据集上进行了验证,且完成了自主避障。下一步可以结合试验室自主作业、环境感知、路径规划方向的成果,完成更为复杂的自动驾驶任务。

(2) 多传感器融合

不同种类的传感器具有各自的优势,工程机械环境、工况复杂度多变,需要不同的传感器来支撑决策模型的鲁棒性。今后可将激光雷达、毫米波雷达、GPS等传感器加入,利用多传感器融合技术处理图像信息、点云信息、距离信息等,从而实现工程机械躲避障碍、自动行走等功能。

(3) 基于学习算法和规则相结合

尽管本文以双目摄像头图像、激光雷达点云作为输入数据,利用提取的语义分割特征图和RGB特征进行多模态融合提高模型预测能力,但仍属于端到端决策的范畴。因此后续可以考虑根据实车场景制定一系列的规则,与基于学习算法的决策结合,从而实现对于工程机械大规模场景的决策。



## 参考文献

- [1] GE L, QUAN L, ZHANG X, et al. Power matching and energy efficiency improvement of hydraulic excavator driven with speed and displacement variable power source[J]. Chinese journal of mechanical engineering, 2019, 32(1):1-12.
- [2] LI A, ZHANG J, ZHOU N, et al. A model for evaluating the production system of an intelligent mine based on unascertained measurement theory[J]. Journal of Intelligent & Fuzzy Systems, 2020, 38(2): 1865-1875.
- [3] BERTOZZI M, BROGGI A, FASCIOLI A. Vision-based intelligent vehicles: State of the art and perspectives[J]. Robotics and Autonomous Systems, 2000, 32(1): 1-16.
- [4] THRUN S, MONTEMERLO M, DAHLKAMP H, et al. Stanley: The robot that won the DARPA Grand Challenge[J]. Journal of Field Robotics, 2006, 23(9): 661-692.
- [5] URMSOON C, ANHALT J, BAGNELL D, et al. Autonomous driving in urban environments: Boss and the urban challenge[J]. Journal of Field Robotics, 2008, 25(8): 425-466.
- [6] MONTEMERLO M, BECKER J, BHAT S, et al. Junior: The Stanford entry in the Urban Challenge[J]. Journal of Field Robotics, 2008, 25(9): 569-597.
- [7] BACHAA A, BAUMAN C, FARUQUE R, et al. Odin: Team victortango's entry in the DARPA urban challenge[J]. Journal of Field Robotics, 2008, 25(8): 467-492.
- [8] ZHANG D, LI K, WANG J. Radar-based target identification and tracking on a curved road[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of automobile engineering, 2012, 226(1): 39-47.
- [9] CANNON H N. Extended earthmoving with an autonomous excavator[D]. Pittsburgh: Carnegie Mellon University, 1999.
- [10] NGUYEN Q H, HA Q P, RYE D C, et al. Force/position tracking for electrohydraulic systems of a robotic excavator[C]// Proceedings of the 39th IEEE Conference on Decision and Control. Piscataway: IEEE Computer Society, 2000: 5224-5229.
- [11] KIM S, RUSSELL J S. Framework for an intelligent earthwork system: Part I. System architecture[J]. Automation in Construction, 2003, 12(1): 1-13.
- [12] ANONYMOUS. Caterpillar launches MineStar Edge management platform[J]. Canadian Mining Journal, 2020, 141(05).
- [13] 张瑞新, 毛善君, 赵红泽, 等. 智慧露天矿山建设基本框架及体系设计[J]. 煤炭科学技术, 2019, 47(10): 1-23.
- [14] 武讲, 郑群飞. 哈尔乌素露天矿无人驾驶方案研究[J]. 金属矿山, 2021(02): 167-172.
- [15] LUNDHEDE O. Multi machine control[C]// Proceedings of the 6th International Platinum Conference. Johannesburg: SAIMM, 2014: 237-246.



- [16] GABER T, EL JAZOULI Y, ELDESOUKY E, et al. Autonomous haulage systems in the mining industry: Cybersecurity, communication and safety issues and challenges[J]. Electronics, 2021, 10(11): 1357.
- [17] 徐工. 徐工无人驾驶压路机与平地机3D找平技术成功应用[J]. 工程机械文摘, 2019(04): 16.
- [18] 杨雪吟. 开启智能新时代 山推纯电推土机、无人驾驶推土机盛装下线[J]. 今日工程机械, 2020(06): 70-71.
- [19] ZHANG L, ZHAO J, LONG P, et al. An autonomous excavator system for material loading tasks[J]. Science Robotics, 2021, 6(55): 3164.
- [20] 李明. 1.4万TEU超大型集装箱船厦门海沧港区远海码头靠泊操纵实践[J]. 上海: 中国港口, 2021(09): 41-44.
- [21] MICHON J A. A critical view of driver behavior models: what do we know, what should we do?[M]// Human behavior and traffic safety. Boston: Springer, 1985: 485-524.
- [22] 胡益恺, 王春香, 杨明. 智能车辆决策方法研究综述[J]. 上海交通大学学报, 2021.
- [23] 陈永尚. 智能汽车城区复杂交通情景的驾驶行为决策方法研究[D]. 哈尔滨: 吉林大学, 2019.
- [24] TALEBPOUR A, MAHMASSANI H S, HAMDAR S H. Modeling lane-changing behavior in a connected environment: A game theory approach[J]. Transportation Research Procedia, 2015, 7: 420-440.
- [25] KENDALL A, HAWKE J, JANZ D, et al. Learning to drive in a day[C]// 2019 International Conference on Robotics and Automation(ICRA). Montreal: IEEE, 2019: 8248-8254.
- [26] VALIENTE R, ZAMAN M, OZER S, et al. Controlling steering angle for cooperative self-driving vehicles utilizing cnn and lstm-based deep networks[C]// 2019 IEEE Intelligent Vehicles Symposium (IV). Paris: IEEE, 2019: 2423-2428.
- [27] 赵祥模, 连心雨, 刘占文, 等. 基于MM-STConv的端到端自动驾驶行为决策模型[J]. 中国公路学报, 2020, 33(03): 170-183.
- [28] 李凌云. 基于深度强化学习的端到端自动驾驶技术研究[D]. 北京: 中国科学院大学(中国科学院人工智能学院), 2020.
- [29] HOOK J, EL-SEDKY S, SILVA V D, et al. Learning data-driven decision-making policies in multi-agent environments for autonomous systems[J]. Cognitive Systems Research, 2021, 65: 40-49.
- [30] LIU Z, AMINI A, ZHU S, et al. Efficient and robust lidar-based end-to-end navigation[C]// 2021 IEEE International Conference on Robotics and Automation (ICRA). Xian: IEEE, 2021: 13247-13254.
- [31] WANG P. Research on Comparison of LiDAR and Camera in Autonomous Driving[C]// Journal of Physics: Conference Series(JPCS). Zhuhai: IOP Publishing, 2021: 012-032.

- [32] LIN X, WANG J, LIN C. Research on 3D reconstruction in binocular stereo vision based on feature point matching method[C]// 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE). Dalian: IEEE, 2020: 551-556.
- [33] ABDU F J, ZHANG Y, FU M, et al. Application of deep learning on millimeter-wave radar signals: A review[J]. Sensors, 2021, 21(6): 1951.
- [34] YONEDA K, HASHIMOTO N, YANASE R, et al. Vehicle localization using 76GHz omnidirectional millimeter-wave radar for winter automated driving[C]// 2018 IEEE Intelligent Vehicles Symposium (IV). Changshu: IEEE, 2018: 971-977.
- [35] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [36] BOJARSKI M, DEL TESTA D, DWORAKOWSKI D, et al. End to end learning for self-driving cars[EB/OL]. [2016-04-25]. <https://arxiv.org/pdf:1604.07316>.
- [37] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Providence: IEEE, 2012: 3354-3361.
- [38] CAESAR H, BANKITI V, LANG A H, et al. nuscenes: A multimodal dataset for autonomous driving[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle: IEEE, 2020: 11618-11628.
- [39] CHEN Y, WANG J, LI J, et al. Lidar-video driving dataset: Learning driving policies effectively[C]//: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City: IEEE, 2018: 5870-5878.
- [40] DOSOVITSKIY A, ROS G, CODEVILLA F, et al. CARLA: An open urban driving simulator[C]// Conference On Robot Learning(CoRL). Mountain View: PMLR, 2017: 1-16.
- [41] FANG Z, LIN T, LI Z, et al. Automatic Walking Method of Construction Machinery Based on Binocular Camera Environment Perception[J]. Micromachines, 2022, 13(5): 671.
- [42] FUKUSHIMA K, MIYAKE S. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological cybernetics, 1980, 36(4): 193-202.
- [43] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [44] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [45] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]// The 3rd International Conference on Learning Representations (ICLR), San Diego: OpenReview.net, 2015.
- [46] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the

- IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Boston: IEEE, 2015: 1-9.
- [47] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- [48] HUANG G, LIU Z, LAURENS V, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu: IEEE, 2017: 4700-4708.
- [49] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in neural information processing systems(NIPS). Long Beach: MIT Press, 2017: 5998-6008.
- [50] 王星圆, 苏楚奇, 黄元毅. 基于改进UNet++网络的路面类型识别技术研究[J]. 武汉理工大学学报, 2022, 44(08): 76-83.
- [51] 陈其怀, 马荣华, 林添良, 等. 一种基于特征融合和时空约束的无人驾驶工程机械决策方法: 中国, CN202111043186.5[P]. 2021-12-28.
- [52] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City: IEEE, 2018: 7132-7141.
- [53] WOO S, PARK J, LEE J, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 3-19.
- [54] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR): IEEE, 2021: 13713-13722.
- [55] YU C, WANG J, PENG C, et al. Bisenet: Bilateral segmentation network for real-time semantic segmentation[C]// Proceedings of the European conference on computer vision (ECCV). Munich: Springer, 2018: 325-341
- [56] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]// International conference on machine learning(ICML). Long Beach: PMLR, 2019: 6105-6114.
- [57] LIU Z, MAO H, WU C, et al. A convnet for the 2020s[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). New Orleans: IEEE, 2022: 11976-11986.
- [58] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International conference on machine learning(ICML). Lile: PMLR, 2015: 448-456.
- [59] LOSHCHILOV I, HUTTER F. Fixing weight decay regularization in adam[EB/OL]. [2017-11-14]. <https://arxiv.org/pdf/1711.05101>.
- [60] LOSHCHILOV I, HUTTER F. Sgdr: Stochastic gradient descent with warm restarts[EB/OL].

- [2016-08-13]. <https://arxiv.org/pdf/1608.03983>.
- [61] 连心雨. 融合时空特征与注意力机制的端到端自动驾驶行为决策方法研究[D]. 西安: 长安大学, 2020.
- [62] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3d shape recognition[C]// Proceedings of the IEEE international conference on computer vision (ICCV). Santiago: IEEE, 2015: 945-953.
- [63] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3d object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018: 4490-4499.
- [64] YAN Y, MAO Y, LI B. Second: Sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.
- [65] LANG A H, VORA S, Caesar H, et al. Pointpillars: Fast encoders for object detection from point clouds[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 12697-12705.
- [66] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 652-660.
- [67] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.
- [68] LI Y, BU R, SUN M, et al. Pointcnn: Convolution on x-transformed points[J]. Advances in neural information processing systems, 2018, 31.
- [69] 许文杰. 基于深度学习的工程机械无人驾驶方法研究[D]. 厦门: 华侨大学, 2020.
- [70] 林添良, 姚瑜, 许文杰, 等. 基于环境识别的电动工程机械无人驾驶行走方法[J]. 机械工程学报, 2021: 1-8.