

中图法分类号: TP29; U46 文献标识码: A 文章编号: 1006-8961(2026)03-0719-26

论文引用格式: Li C X, Hu H X, Guo D B, Cheng R M, Wu H K and Yuan C. 2026. Trajectory planning for autonomous driving: a survey. Journal of Image and Graphics, 31(3):0719-0744(李程翔, 胡海霄, 郭大波, 陈荣敏, 吴宏坤, 袁畅. 2026. 自动驾驶轨迹规划综述. 中国图象图形学报, 31(3):0719-0744)[DOI:10.11834/jig.240774]

自动驾驶轨迹规划综述

李程翔, 胡海霄, 郭大波*, 陈荣敏, 吴宏坤, 袁畅

三亚学院信息与智能工程学院, 三亚 572000

摘要: 自动驾驶技术是交通领域重要的前沿技术, 其发展对交通安全和效率具有变革性意义。本文试图从传统算法、强化学习和模仿学习 3 个方面介绍自动驾驶轨迹规划领域的前沿成果, 对主要相关文献进行了梳理。首先, 介绍了目前较为主流的自动驾驶框架: 模块化和端到端, 并论述了轨迹规划在两种框架中的使用; 其次, 对传统算法在轨迹规划中的最新运用进行梳理; 在此基础上, 总结了混合传统算法的发展方向; 对当下较为新颖的强化学习和模仿学习路径规划算法进行总结, 列举了针对强化学习算法中安全性不足和样本效率低等核心问题提出的众多典型解决方案, 如虚拟仿真驾驶平台、终身学习等; 同时总结了模仿学习的改进措施, 如引入自学习、元学习等; 还对如今热门的视觉语言模型和视觉语言动作模型路径规划进行了分析, 阐述强化学习在其中的作用; 概述了以扩散模型结合强化学习的未来发展趋势, 以及世界模型在自动驾驶中的运用; 最后分析了高质量自动驾驶存在的挑战。本文旨在通过系统的梳理, 凝练出自动驾驶轨迹规划领域中存在的挑战以及发展趋势, 期望为后续研究者提供研究方向及参考。

关键词: 自动驾驶; 轨迹规划; 传统规划算法; 强化学习(RL); 模仿学习(IL); 视觉语言动作模型; 扩散模型(DM); 世界模型

Trajectory planning for autonomous driving: a survey

Li Chengxiang, Hu Haixiao, Guo Dabo*, Cheng Rongmin, Wu Hongkun, Yuan Chang

School of Information and Intelligent Engineering, University of Sanya, Sanya 572000, China

Abstract: Autonomous driving technology represents a cornerstone of innovation within the transportation sector, with the potential to dramatically enhance the safety and efficiency of transportation systems on a global scale. This technological advancement has received substantial policy backing from numerous governments worldwide, while the automotive industry has significantly ramped up its commitment to research and development efforts, concurrently initiating processes aimed at establishing standardized protocols for autonomous vehicles. Autonomous driving is highly complex, encompassing a vast array of interdisciplinary fields and technological nuances. This work summarizes the most recent and cutting-edge findings in autonomous driving trajectory planning, including traditional algorithmic approaches, reinforcement learning methodologies, and imitation learning methodologies, to systematically collate, dissect, and synthesize the relevant academic literature, providing a comprehensive overview of the current state and future directions of this field. First, this paper provides

收稿日期: 2025-01-03; 修回日期: 2025-08-03; 预印本日期: 2025-08-10

* 通信作者: 郭大波 dabo_guo@sxu.edu.cn

基金项目: 海南省高等学校教育教学改革研究项目(Hnjg2025ZC-89); 三亚学院容淳铭院士工作站; 三亚学院 2023“产教融合”研究项目(USY23CJR01, USY23CJR02)

Supported by: Hainan Province Higher Education Teaching Reform Research Project(Hnjg2025ZC-89); Sanya University of Sanya Academician Rong Chunming Workstation

an in-depth examination of the two most widely recognized frameworks for autonomous driving: the modular framework and the end-to-end framework. Additionally, it acknowledges the existence and relevance of other, less commonly utilized but no less classical frameworks, offering a nuanced analysis of their respective strengths and weaknesses. This analysis focuses on the practical implementation of trajectory planning within these frameworks, with a particular emphasis on identifying potential areas for future enhancement and optimization within each framework's architecture. Afterward, the paper presents a detailed survey of traditional planning algorithms that are currently employed in autonomous driving trajectory planning. The survey is meticulously organized, categorizing and summarizing the state-of-the-art traditional algorithms at the forefront of trajectory planning research. The algorithms under consideration include graph-based search methodologies, rule-based local planning approaches, intelligent optimization techniques, static and dynamic sampling methods, artificial potential field-based algorithms, and various model control strategies, among others. The paper provides a thorough synthesis of the application domains, advantages, and limitations of these algorithms while also highlighting the innovative optimization strategies and emerging trends that have been proposed by the academic community. Building upon this foundation, the paper introduces the concept of hybrid traditional algorithms, which combine local planning algorithms with global planning algorithms or integrate multiple local and global algorithms to create new and potentially more effective trajectory planning methods. The paper engages in speculative analysis regarding the future trends and directions of these hybrid traditional algorithms and discusses how such algorithms can address the deficiencies of traditional approaches. Moreover, the paper scrutinizes the impact and effectiveness of these hybrid algorithms on the path planning capabilities of autonomous driving systems. The paper concludes this section by delineating the problems and limitations that are associated with hybrid traditional trajectory algorithms. Evidently, individual traditional trajectory planning algorithms and their hybrid counterparts have numerous issues and shortcomings that need to be considered carefully. In response to these challenges, the paper focuses on the integration of novel reinforcement learning algorithms as a means to overcome the limitations of traditional trajectory planning methods. These reinforcement learning algorithms, which include value-based, policy-based, and actor-critic-based approaches, are classified, explored, and summarized. This paper discusses the operational principles, application domains, advantages, and disadvantages of each reinforcement learning algorithm in detail, supported by the research findings and scholarly contributions in the field. The paper's focus then shifts to the critical aspects of reinforcement learning in the context of autonomous driving, such as efficiency, safety, and function optimization. It addresses the difficulties encountered by existing reinforcement learning techniques and presents innovative solutions proposed by scholars to tackle core issues such as the security and sample efficiency of reinforcement learning algorithms. These solutions include the establishment of virtual simulation driving platforms, the application of imitation learning algorithms, lifelong learning techniques, hierarchical learning methods, and multi-agent reinforcement learning algorithms. Also, this paper introduces a reinforcement learning approach based on the world model and summarizes the latest progress, along with the latest research progress of VLM and VLA and the role of reinforcement learning. This paper also provides a comprehensive evaluation of popular virtual simulation platforms, assessing their suitability for research in autonomous driving on the basis of 12 metrics: scenario building, vehicle dynamics, sensor models, traffic participants, weather and time simulation, data logging, API interfaces, realism and performance, scalability, community support, gym interface compatibility, and support for multi-agent reinforcement learning. The paper synthesizes the research applications and distinctive features of these platforms, offering a valuable resource for researchers selecting appropriate simulation environments for their work. Furthermore, the paper studies the application of imitation learning techniques in autonomous driving, including inverse reinforcement learning, behavioral cloning, and generative imitation learning, along with optimization strategies proposed by the academic community. It explores how these techniques can refine reinforcement learning policy networks, enhancing the sample efficiency and optimality of the learning strategies. This paper also introduces the VLA method based on the diffusion model, which is currently popular. On the basis of traditional imitation learning, more novel directions such as online imitation learning, self-imitation learning, state-only imitation learning, multi-agent imitation learning, generative imitation learning, and meta-imitation learning are also expounded. In the final sections of the paper, the authors critically analyze the current limitations and future development trends of reinforcement learning algorithms and imitation learning algorithms in autonomous driving. Finally, the future research directions and

challenges are elaborated, including the VLA framework based on pre-trained VLM, the diffusion model combined with reinforcement learning, and the reinforcement learning based on the world model. In this regard, a detailed analysis of the causes was conducted, and the pain points of the above three research directions were identified: the pain points of the world model (virtual to reality), the pain points of VLA architecture (data requirements), and the pain points of the diffusion model (computing power bottleneck). Other papers were reviewed to provide ideas and directions for solving these pain points. In conclusion, this paper offers a comprehensive overview of the trajectory planning algorithms in autonomous driving, synthesizing the development trends and challenges of traditional algorithms, reinforcement learning methods, and imitation learning algorithms. It aims to provide future researchers with a solid foundation for understanding the current state of the art, as well as a roadmap for future research directions in the field of autonomous driving technology. This paper also emphasizes the need for interdisciplinary collaboration and the integration of diverse methodologies to advance the state of autonomous driving systems and ensure their safe, efficient, and widespread adoption in the transportation sector.

Key words: autonomous driving; trajectory planning; traditional programming algorithms; reinforcement learning (RL); imitation learning (IL); visual language action model; diffusion model (DM); world model

0 引言

自动驾驶是人类由来已久的梦想,在提高交通效率和促进社会经济发展上有着重要作用,这使得自动驾驶成为社会较为关注的领域,而轨迹规划作为其技术核心(Claussmann等,2020;Dinneweth等,2022;Zhao等,2020),发挥着至关重要的作用。本文将从传统算法和基于深度强化学习两个方面论述自动驾驶轨迹规划。传统算法多采用机器人的行为算法,如启发式搜索算法、图搜索算法、元启发式算法、虚拟力场算法以及动力学约束的局部路径规划算法等。这些算法能够使目标车辆在简单路面上找到抵达目的地的最短距离并自动行驶到目的地,但是传统算法对于较为复杂的道路情况,很难有较好的成效(Reda等,2024)。近年来,强化学习(reinforcement learning, RL)逐渐被学者们关注和认可。RL算法适合解决序贯决策问题,算法可让车辆能够通过学习来迭代更新模型,从而逐步改善车辆在复杂路面上的行驶能力,使自动驾驶车辆在行驶过程中更为稳定(Aradi,2022)。而正在兴起的深度强化学习(deep reinforcement learning, DRL)具有强大的建模能力,使得复杂道路下的自动驾驶轨迹规划成为可能(Kiran等,2022)。另一方面,模仿学习(imitation learning, IL)因为其能够直接利用专家示范数据(如真实场景中的人类驾驶数据、虚拟仿真中的模拟驾驶数据等)的特性,避免了强化学习低效探索的问题而备受关注,并且模仿学习可以通过专家行为进行隐式编码,以此绕过显式奖励函数设计(褚端峰

等,2024),避免因为奖励函数设计不佳导致的学习效率低。事实上,模仿学习和强化学习技术可共存,模仿学习可以使自动驾驶车辆快速学习并达到接近人类驾驶者的水平,而由强化学习来辅助探索并解决人类驾驶记录中不常出现的“长尾问题”,因此越来越多的学者尝试使用深度强化学习结合模仿学习的方法实现自动驾驶轨迹规划。随着大模型技术的快速发展,许多学者尝试将机器人领域视觉语言动作运用到自动驾驶领域,并尝试联合扩散模型和基于世界模型的强化学习实现多策略协同,以提高自动驾驶效率。

1 自动驾驶架构

自动驾驶架构(autonomous driving system, ADS)是由汽车工程师协会(Society of Automotive Engineers, SAE)提出并规定的,协会定义能够可持续执行动态驾驶任务(dynamic driving task, DDT)的软硬件系统称为ADS。如今,许多著名的汽车公司和工程研究机构相继推出了自家的ADS(Yurtsever等,2020)。这些ADS往往建立于几种架构模式,即模块化架构、端到端架构、集中式架构、分布式架构和混合架构等,其中最为热门和最受推崇的是传统的模块化架构及较为新颖的端到端架构。本节将重点介绍这两种自动驾驶架构。

1.1 模块化架构

模块化架构,相较于端到端架构拥有较好的可解释性,其架构逻辑为多个独立的子模块,每个模块负责特定的任务(Li等,2023b)。这种结构使得问题

更容易定位和调试。每个模块的功能是明确且可知的,有助于分析系统的决策过程。模块化结构灵活性与可扩展性强,在不影响其他模块的前提下可单独修改或替换某个模块。模块化架构下各模块之间的接口清晰,利于系统安全验证和测试,以及后期维护,使系统具有较强的整体鲁棒性和可靠性(Zhao等,2024)。但是,过多的模块也导致了系统的复杂性高,模块与模块之间的交互频繁,造成了庞大的通信量,从而降低了系统的性能。最后,模块化架构需要考虑如何进行后期的优化,由于各模块独立工作,仅靠数据通信,使得全局优化变得困难,优化一个模块时可能会降低其他模块的性能,因此难以平衡各模块以达到整体上最佳性能,图1所示为自动驾驶中传统的模块化架构图。

1.2 端到端架构

端到端架构比传统的模块化架构相对简单,系统直接将各类传感器得到的数据输入到模型中,再通过多模态算法生成控制指令,从而大幅减少了中间各类模块的传输通信(Xiao等,2022)。这种简洁的设计降低了系统复杂性和出错率。模块化设计的全局优化难度高,反观端到端设计可以直接从数据中学习最优的映射关系,从而实现全局最佳,无需

人为考虑和协调各个模块,同时也减少了人为干预造成的误差。相比于模块化模型,在面对新环境时,端到端模型通过训练可快速适应新情况,无需重新设计整个系统。端到端的优势在于其实时性,模块化架构由于有较多的中间处理环节,带来了不必要的消耗,拖累了响应速度,而端到端架构抑制甚至规避了这一问题,具有更快的响应速度和更低的延迟(Tampuu等,2022)。对于驾驶汽车这类需要快速反应的应用场景,端到端架构可能更具发展前景,近年来,国外的特斯拉,国内的蔚来、小鹏、商汤等公司相继在端到端自动驾驶中取得了重要成果(陈妍妍等,2024)。但是,基于端到端的ADS并非完美,其缺乏解释性,通常为黑盒式,甚至开发者也难以理解其内部的具体决策过程,因此系统的透明度低;其次无论是基于强化学习还是基于模仿学习的端到端模型,都需要依靠大量的标注数据,实际应用中这一问题往往难以解决,且需要大量的成本。这些数据的质量和准确性、多样性直接或间接影响模型的性能,因此对于收集到的数据往往还需进行复杂的预处理;模型的泛化能力过于依赖训练数据的覆盖范围,架构安全性略逊于模块化。图2为ADS端到端架构图。

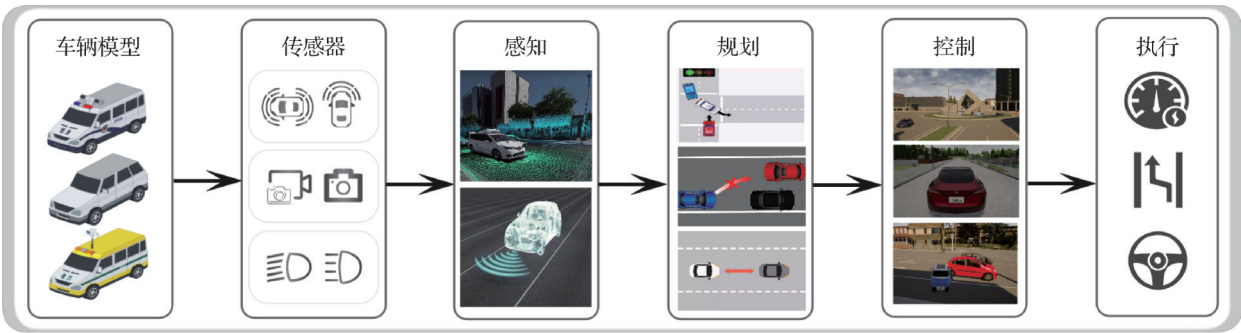


图1 ADS模块化架构图(Reda等,2024)

Fig. 1 ADS modular architecture diagram(Reda et al. ,2024)

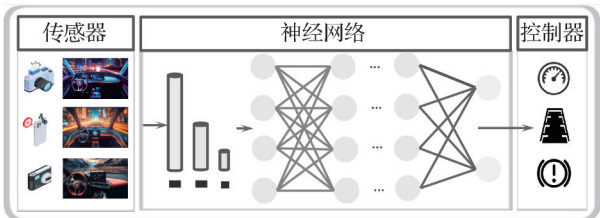


图2 ADS端到端架构图

Fig. 2 ADS end-to-end architecture diagram

1.3 轨迹规划

无论是行业内较为新颖的端到端架构还是较为传统的模块化架构,都离不开规划(Wu等,2024)。模块化架构中规划模块需要制定对应的规划算法实现这些功能,轨迹规划在模块化架构中需要接收来自传感器模块和地图模块的信息,根据收到的信息,利用写好的规划算法计算出当前情况下

的最优路径。接着利用模块中计算出的最优路径转化为一系列的轨迹点,发送给控制模块。而在端到端框架中,输入端需要根据传感器送来的信息,提取出所需的特征,供模型计算出未来的行驶轨迹并进行轨迹优化,以此得出一条相对安全的行驶轨迹或轨迹点,将其传送给输出端。不难发现,规划在指示目标车辆行驶中起到了至关重要的作用,如何优化规划算法已成为学术探索的热点。大数据的兴起,伴随着数据驱动算法的狂潮,RL

也因其算法强大的表征能力和优化有效性进入研究者的视野。但不可否认的是,传统的自动驾驶轨迹规划算法依然在许多方面要强于如今的RL算法,而且其解释性更佳,因此需要在研究RL的同时注重传统算法的创新,尝试取二者的长处,舍弃二者的不足。

图3是本文综述提纲,总结了自动驾驶轨迹规划算法的分类和特例,包括传统算法、强化学习和模仿学习三大类,以下逐一论述。

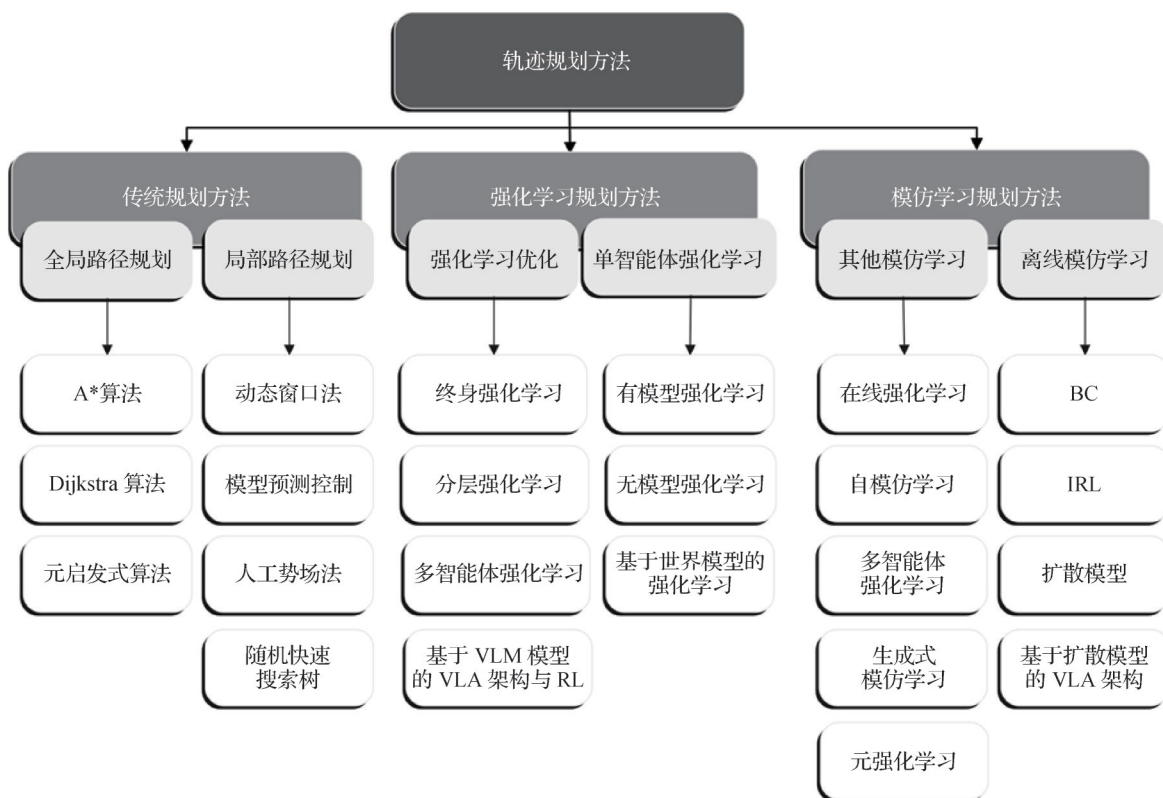


图3 规划算法划分图

Fig. 3 Planning algorithm partition diagram

2 基于传统算法的轨迹规划

轨迹规划作为自动驾驶中重要的一环,其核心目标是依靠现有数据计算出一条安全且舒适的最短轨迹,引导驾驶车辆完成行驶任务。其中安全指的是车辆在行驶过程中不会与其他障碍物发生碰撞;舒适指的是车辆的行驶稳定,不会出现突然的急刹车或者车辆突然加速,给予乘客舒适的乘坐体验。轨迹规划中最为重要的一点还是车辆能够找到耗时最短的路径,而寻找最短路径的方法有许多。下面将介绍传统的

经典方法,以及对于这些经典算法的融合和优化。

2.1 传统轨迹规划算法

2.1.1 全局路径规划

经典的全局路径规划算法有 A*算法、Dijkstra 算法、可视图方法、细胞自动机、蚁群算法、遗传算法、模拟退火算法等。其中 A*算法使用启发式函数评估从当前所在节点到目标所在节点的代价;Dijkstra 算法一种基础的单源最短路径算法;可视图方法通过构建可视图,判断两个节点之间是否存在障碍物,若不存在,则可构成一条边,以此构建从起点到终点的最短路径;细胞自动机模型将时空和状态离

散化,模型由演化规则、元胞空间等元素组成,各元胞根据其自身状态和邻居状态,按照局部规则进行更新,依此模拟复杂系统时空演化,预测行驶路径;蚁群算法模拟蚂蚁觅食行为,通过模拟信息素标识路径质量;遗传算法是一种受自然选择过程启发所实现的搜索算法,可以用于优化路径规划问题;模拟退火算法通过模拟冶金过程中的退火阶段,控制参数平衡进行探索,避免陷入局部最优解。由表1所

示,将全局路径规划算法进行分类,并对算法进行优缺点分析。

上述全局路径规划思路是在已知环境地图的情况下,计算出从起点到终点的大致路径。这也使得这些算法对动态障碍物和其他实时变化的因素敏感性差,难以做出及时的调整。这意味着单一使用全局规划算法在自动驾驶轨迹规划这类实时变化较高的场景中难以达到很好的效果。

表1 经典全局规划算法分析表
Table 1 Analysis table of classic global planning algorithms

算法	优点	不足	分类
A*	搜索效率高,快速找到最优路径	计算复杂度高、高维空间中的计算量大	基于图搜索的方法
D*	动态环境中高效,快速响应环境变化,适应性强	计算复杂度高、存储空间大、对初始路径依赖	基于图搜索的方法
Dijkstra	适用于动态环境的路径规划	计算复杂度高、不适用于高维空间、计算量大	基于图搜索的方法
可视图方法	直观,便于理解和调试	耗时久、大规模数据处理能力差	基于图搜索的方法
细胞自动机	适合于处理具有局部规则的复杂系统	复杂环境的适应性不足、计算量大	基于局部规则的方法
蚁群算法	快速发现较好的解决方案	容易陷入局部最优解、且参数调整可能较为困难	基于元启发式方法
遗传算法	搜索空间范围广,灵活性和适应性强	计算复杂度较高、参数敏感,结果随机性强	基于元启发式方法
模拟退火	适用于复杂优化问题	可能导致陷入局部最小值、执行时间长	基于元启发式方法

2. 1. 2 局部路径规划

经典的局部路径规划算法有:动态窗口法、模型预测控制、人工势场法、随机快速搜索树(rapidly-exploring random tree, RRT)等。动态窗口法根据行驶车辆当前的运动状态和运动学模型,模拟一定时间内的运动轨迹,并通过评价函数对轨迹评分,选择最优速度来控制车辆;模型预测控制方法利用现有的动力学模型对未来可能出现的状态进行预测,并进行优化和微调,生成最佳轨迹;人工势场法的核心是将目标点视为吸引场,将障碍物视为排斥场,通过模拟力的作用生成路径;RRT和RRT*是基于随机采样的路径规划算法,这类算法的优点是适合高维度的空间或复杂的环境,与自动驾驶领域的适配性好。表2对局部路径规划算法进行了分析。

对比可知,局部路径规划相比于全局路径规划,更加关注如何在当前环境中实现某种方法来避开障碍物,能够实时处理传感器传递的数据,并快速做出反应。自动驾驶中不但需要有初始路径,通常还需

要进一步优化这条路径,使之更平滑,平滑的轨迹才能提供安全的驾驶环境。

2. 2 传统算法的新发展趋势

传统算法在特定场景下展现出各自的优势,但它们在实际应用中普遍存在一些局限性。这些算法在处理大规模或复杂路网时,往往面临计算复杂度高、实时性难以保证等问题,尤其是在动态变化的交通环境中,难以准确预测未来状态并生成适当的路径。针对这些局限性,未来的研究方向包括怎样提高计算效率、增强算法对动态环境的适应性、开发数据驱动的方法以减少对大量训练数据的依赖,以及混合算法的开发。这些研究方向有望通过技术创新,显著提高自动驾驶车辆路径规划的性能和可靠性,下面先从单一算法的优化开始介绍。

2. 2. 1 基于图搜索的算法

A*和Dijkstra算法是基于图搜索算法内最为著名的两种方法,许多研究者对这两种算法进行了优化。早期的改进算法,包括改进A*(赵晓等,2018),

表 2 经典局部规划算法分析表
Table 2 Analysis table of classic local planning algorithms

算法	优点	不足	分类
随机快速搜索树	快速探索高维空间,适应力强	生成的路径可能不是最优的,且不会收敛到渐进最优解	基于静态采样方法
动态窗口法	计算量小,反应快,适用于未知环境避障	可能会在复杂环境中陷入局部停滞,动态避障性能有待提高	基于动态采样方法
人工势场法	简单且反应快,适用于快速避障场景	容易陷入局部最优,目标附近有障碍物时,车辆目标不可达	基于人工势场方法
模型预测控制	轨迹平滑且符合物理约束	计算复杂度较高,对实时性要求高,计算开销大	基于模型控制方法

在路径搜索时减少搜索点的数量;aleph*(Keselman等,2018)实现了可学习的启发式A*算法;Hybrid A*(Tu等,2019)解决了非完整约束车辆的平滑路径问题;Neuarl A*(Yonetani等,2021)结合了神经网络和A*;蜂巢栅格A*(陶哲等,2020)实现了在蜂巢网络中使用A*,提升了A*的搜索性能;改进A*(杨明亮和李宁,2022)降低了A*算法的转折次数和转角幅度。近年来也有不少科研人员提出了较为新颖的改进算法,如iADA*(improved anytime dynamic A*)(Maw等,2020)、EBS-A*(expansion, bidirectional, smoothing, improved A* algorithm)(Wang等,2022)、RLDA(reverse labeling Dijkstra algorithm)(Zhu和Sun,2021)、FC-A*(fuel consumption-A*)(Liu和Zhang,2022)。前面几种方法主要是对算法的性能进行了优化,EBS-A*不仅实现了并行化,还在并行化的基础上实现批处理机制,以批处理的方式模拟节点扩展,减少重复操作和提高缓存利用率;iADA*结合了Anytime的思想,算法提供一个可行解,随着时间逐步改进解的质量,这使得行驶过程中即使系统并没有充分全面搜索也能做出反应。后面几种算法则是考虑了其他驾驶因素,RLDA在Dijkstra的基础上引入了交叉口属性,对行驶路面尤其是交叉口进行了深度研究;FC-A*则是对评价指标做出了修改,将汽车油耗纳入考虑。

2.2.2 基于人工势场的算法

人工势场法(artificial potential field, APF)作为经典的机器人轨迹算法,许多研究者在其基础上提出新的优化方式,ISR-APF(icy and snowy roads-APF)是一个较为典型的例子(Zhai和Pei,2023)。该算法考虑到冰雪路面这一特征,将其作为参数优化人工势场函数,使算法在冰雪环境中也有较好的

性能;针对传统APF出现引力与斥力合力为零而陷入局部最优的问题,VF-APF(virtual force-APF)(张佳尚和陈志华,2023)引入虚拟力辅助人工势场法以摆脱局部最优;也有团队将研究重点放在了研究人工势场法的特性上,其思路是通过人工势场法生成路径长度,譬如用于自动驾驶汽车路径规划的人工势场法的变体(Paliwal,2023),该变体在各种复杂环境中生成的路径比使用原始算法生成的路径短。表3总结了基于人工势场法的算法。

表 3 基于人工势场的算法
Table 3 Algorithms based on APF

算法名称	优化领域	核心思想
APF	通用路径规划	设计思路简单,计算量相对小,实时性强,适应性佳
ISR-APF	冰雪路面	优化人工势场函数,提高恶劣环境适应性
VF-APF	避免局部最优	引入虚拟力辅助人工势场法
IAPF	生成路径长度	在各种复杂环境中生成的路径比使用原始算法生成的路径要短

2.2.3 基于静态采样的算法

随机快速搜索树(RRT)中比较经典的改进算法有RRT*、Goal-bias RRT(Jin等,2020)、Bi-RRT(bidirectional RRT)、RRT-Connect、Extend RRT、Local-tree-RTT、Dynamic-RRT等。其中RRT*是最为经典的改进算法,其思想是在RRT的基础上,引入重写模块和随机重连模块,凭借两个额外模块优化路径,并保证渐进最优性。如今,RRT*出现了不少优秀的改进算法,如D-RRT*(Ganesan等,2022)、Bi-RRT*(Fan等,2024)等。RRT*-Smart受Informed

RRT* 算法的启发,采用椭圆形的采样点范围约束以提高采样效率,同时加入信标节点使算法能够更有效地在障碍物附近进行采样;Ganesan 等人(2022)提出了 D-RRT* 算法,其关键的思想是通过减小采样空间优化算法的性能,利用起始配置到形成的简单椭圆的启发式算法关注目标的方向;Bi-RRT* 算法将 Goal-Bias 和 RRT-Connect 中的目标偏置和双向扩展的思想融入到了 RRT* 中,并加入了采样约束与自适应偏置的功能。表 4 分别对基于 RRT 和 RRT* 的算法性能进行分析,并总结了几种算法在目标偏差、双向搜索、渐进最优性、计算时间以及路径质量等几个方面的性能优势和不足。

2.2.4 基于动态采样的算法

动态窗口法(dynamic window approaches, DWA)根据车辆当前的位置、速度、目标以及周围环境,在速度空间内采样多个线速度和角速度组合,模拟出短期运动轨迹,通过评价函数评估,选择出能够避开障碍物的最优局部路径。对于 DWA 算法,早期提出了 DWA* 算法(Lin 等,2009),通过广度优先搜索(breadth first search, BFS)改善 DWA 的搜索能力,但是并没有取得突破性的进步。近年,一些优秀算法相继提出,如: Improved-DWA(Yan 等,2022), Bounded-DWA(Gong 等,2023), FDWA(Pan 等,2024)。Improved-DWA 联想到人为因素在车辆行驶过程中的影响,将其作为重要参数调整模型;

表 4 基于 RRT 和 RRT* 的算法

Table 4 Algorithms based on RRT and RRT*

算法名称	双向搜索	渐进最优性	计算时间	路径质量
RRT	否	否	较短	一般
Goal-bias RRT	否	否	较短	较高
Bi-RRT	是	否	较短	较高
RRT-Connect	是	否	较短	一般
Extend RRT	否	否	较长	较高

算法名称	双向搜索	渐进最优性	计算时间	路径质量
Local-tree-RRT	否	否	较长	较高
Dynamic RRT	否	是	较长	较高
RRT*	否	是	较长	较高
B-RRT*	是	是	较长	较高
D-RRT*	否	是	较短	较高

Bounded-DWA 方法遵循 Ackermann 运动学约束,通过将约束与 DWA 采样窗口集成,缩小和限制采样范围,提高了路径规划效率和车辆移动效率;FDWA 具有解决多智能体协同运动的实时路线规划问题的能力,能够在车辆较多的场景中做出较好的规划,是一种具有模糊自适应参数调整的改进算法,能实时进行权重参数评估。对基于动态窗口法的总结分析如表 5 所示。

表 5 基于动态窗口法的算法

Table 5 Algorithms based on dynamic window

算法名称	改进内容	解决问题
DWA*	根据车辆状态和环境信息选择最优路径	提高路径规划的适应性和效率
FDWA	实时路线规划	多智能体协同运动的实时路线规划问题
Bounded-DWA	提高车辆移动时 DWA 的效率	提高路径规划效率和遵循车辆运动学约束
Improved-DWA	允许在车辆行驶过程中人为干涉修改路径	车辆行驶过程中的人为干涉和路径修改问题

2.2.5 基于元启发式算法

元启发式算法是解决复杂优化问题的一类通用策略。其主要思想是通过高效搜索策略在解空间中寻找近似最优解。通过全局探索和局部利用的平衡,可以避免陷入局部最优,不被具体问题所影响。在自动驾驶中常用算法类型包括:1)基于进化机制的算法:遗传算法、差分进化算法等;2)基于群体智能的算法:粒子群算法、蚁群算法、灰狼算法和鲸鱼

算法等;3)基于物理现象的算法:模拟退火、引力搜索算法等。元启发式算法作用通常体现在优化其他传统算法以提升生成路径效果,但有时也可以通过单一算法直接形成路径。如今,此类算法已经具备通用性强、全局搜索能力强、无需梯度信息和易于与其他方法结合等优势,但依然存在计算成本较高、收敛性不确定、理论分析困难和高维问题性能下降等不足,有待进一步提升。一些团队对于元启

发式算法在自动驾驶中的运用进行了尝试,如通过引力搜索算法进行高维避障的自动驾驶路径规划(张爱国等,2020);利用平衡鲸鱼优化算法的无人车路径规划算法(蔡雨岑和杜鹏桢,2021);为了保证车辆到目标点的途中生成的路径平滑且无碰撞以及 Ait-Saadi 等人(2022)提出的基于混沌 Aquila 与 SA(simulated annealing)的优化方法 CAO-SA(chaos-simulated annealing)。针对轨迹规划中

的时间优化问题,Xu 等人(2023)提出一种改进 SA 的鲸鱼优化算法 IWOA-SA(improves the whale optimization algorithm with simulated annealing)和自适应权重迭代策略,可防止算法陷入局部最优;也有团队尝试利用模拟退火算法进行双目标优化的路径规划算法(田志伟等,2025)。表 6 所示为元启发式算法的分析表,总结了上面提到的几种算法的特点。

表 6 基于元启发式算法分析
Table 6 Analysis based on meta-heuristic algorithms

算法名称	针对问题	改进内容
基于引力搜索	针对传统算法车辆实时避障能力差的问题	采用引力搜索算法增强避障能力
基于平衡鲸鱼	针对全局探索和局部利用协调问题	引入动态平衡和种群重构机制
IWOA-SA	针对局部最优状态的问题	结合 SA 和 IWOA 并改良自适应权重
CAO-SA	针对算法利用率低的问题	结合混沌 Aquila 优化与 SA 的优化方案
基于模拟退火	针对距离时间双目标优化问题	基于模拟退火算法和线性加权法的优化模型

2.3 混合算法的发展趋势

根据上面的分析不难看出,单一算法不再能满足轨迹规划的发展,因此出现了许多较为新颖的混合方式。这些算法将全局规划算法和局部规划算法相结合,能够吸取两种算法的优点。

2.3.1 涉及 A* 的混合规划的算法

双向 A* 用做全局路径规划、DWA 作为其局部路径规划策略的新式算法(Xu 等,2024);在 A* 和 DWA 基础上加入跳点技术(jump point search, JPS)的 JBS-A*B(jump point search-A* with b-spline smoothing)算法(Dang 和 Tan,2024);结合改进 A* 算法和 APF(artificial potential field)的路径规划算法(Yang 等,2024);A* 和多子目标人工势场法相融合的轨迹算法(Sang 等,2021);以及将 Hybrid A* 扩展到 3 个维度,并用 3D 矢量约束 Hybrid A* 的 3D-Hybrid A* 算法(Zhang 等,2024)。

2.3.2 涉及 Dijkstra 的混合规划的算法

由 Dijkstra 用做全局路径规划、DWA 作为其局部路径规划的混合轨迹规划算法 DWA-Dijkstra(Liu 等,2021);Ait 等人(2023)提出一种将 A* 的启发式函数思想与 Dijkstra 算法相结合的混合算法 Dijkstra-A* 算法;采取两步路径规划策略的 Dijkstra-RRT 算法先利用 RRT 算法缩小规划区域,将较小规划区中的路径规划问题转变为图论问题,接着利用 Dijkstra

算法解决图论问题(Chen 等,2022);基于 Dijkstra 算法和 Lyapunov 的控制系统的 Dijkstra-LbCS(Dijkstra-Laypunov based control system)算法(Parkash 等,2024)在保证实现了高效路径寻路方案的同时确保了导航的稳定性;结合了新式拓扑和 Dijkstra 的最优路径规划算法 Dijkstra-CDT(Dijkstra-convex dissection topolpgy)算法(Liu 等,2023)通过计算二维连续连续空间中从一个点到所有其他点的全局最优路径,进而实现极高速度返回任意目标点的全局最优路径。

2.3.3 涉及 APF 的混合规划的算法

由 APF 提供全局规划并使用 MPC(model predictive control)作为运动轨迹辅助的混合算法(Zuo 等,2019);为满足自动驾驶汽车动态变道轨迹规划(dynamic lane trajectory planning, DLTP)的多目标需求而提出的基于改进 APF 和 RRT 的新型 DLTP 方法(Zhao 等,2025);基于部分可观测马尔可夫决策过程(partially observable Markov decision process, POMDP)的 APF 无人机路径规划策略;一种新颖的基于动态增强烟花算法和 APF 的自动驾驶汽车实时路径规划方法(Li 等,2022);为解决人工势场法在无人驾驶变道规划中问题而提出的基于势能重构和人工势场的变道轨迹规划算法(胡丹丹和张琪,2022)。

2.3.4 涉及RRT和RRT*的混合规划算法

RRT*与APF简单融合的算法(Li等,2024);结合变概率目标偏置策略和APF的RRT*增强算法(Tao等,2024);RRT*-TEB(RRT*-time elastic band)算法,其思想是RRT*用于生成全局路径,接着对移动目标进行运动学建模并优化时间弹性带(time elastic band,TEB)算法(Yin等,2024);通过APF方法被集成到Quick-RRT*算法中,利用目标点的吸引力和障碍物的排斥优化树扩展过程的新式离散化人工势场IDAPF-QRRT*(improved APF with Quick RRT*)(Liu等,2024)算法。

2.3.5 涉及ACO、SA和GA的混合规划算法

用于解决多冲突目标问题的混合轨迹算法GA-A*(genetic algorithm-A*)(Beed等,2020);提升路径规划收敛速度的基于蚁群优化算法结合模糊控制的全局路径规划算法ACO-MPC(ant colony optimization-model predictive control)(Tao等,2021);轨迹优化算法混合PSO-SA(particle swarm optimization-simulated annealing)(Lin等,2023);在扩展随机树的过程中引入蚁群优化(ACO)使规划路径渐近最优的方案ACO-RRT*(Wang等,2024d)。

2.4 传统算法总结

各类混合算法的提出,解决了传统单一算法在某些情况下效率不高的问题,同时也使得这些算法能够

适用于动态环境中的快速路径规划,但传统轨迹规划算法在适应性和优化问题上具有天然的劣势,所以许多研究者考虑在自动驾驶轨迹规划中引入学习算法。RL就是其中之一,传统轨迹规划算法适应性较差,通常需要预定义规则或者建立各类假设,对于复杂多变的真实驾驶环境效果差,而RL能够通过自适应学习,与环境的互动,不断学习和调整策略,以达到更好地适应各种驾驶情况的目的。传统的轨迹规划模型优化较为复杂,想要在复杂环境找到最优解,往往需要同时考虑多个约束条件,这使得参数调整的过程更为冗长和烦琐,而RL算法能够自主探索状态空间,并依此寻找在给定奖励函数下的最优策略。只需要设置合适的奖励机制,模型便可自主在多参数之间找到平衡点,而不需要进行烦琐的调参步骤。

3 基于强化学习的轨迹规划

3.1 基础强化学习算法

3.1.1 经典强化学习

自动驾驶路径规划中,常将强化学习分为两类,一类是无模型强化学习方法(model-free RL, MFRL),一类是有模型强化学习方法(model-based RL, MBRL),其区别在于是否需要对场景进行显式建模。图4展示了自动驾驶中的无模型强化学习。

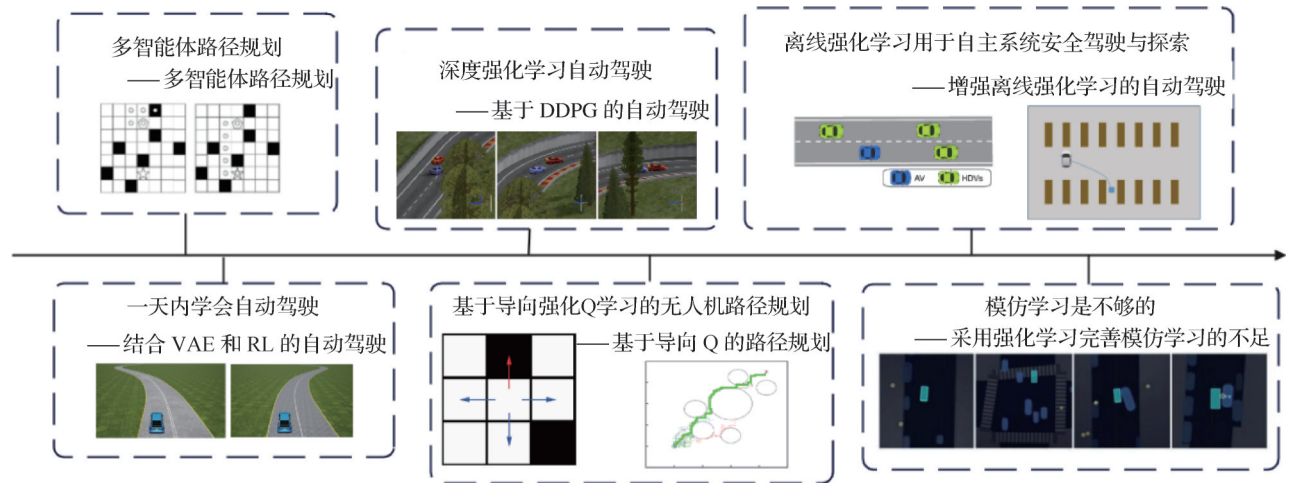


图4 自动驾驶中的无模型强化学习

Fig. 4 Model-free reinforcement learning in autonomous driving

MFRL无需对环境进行建模,直接通过交互学习来迭代策略和价值函数,这样的学习方式数据利用效率较低,但是计算复杂度低,仅需要优化策略网

络和价值网络。

无模型强化学习一般分为3类:1)以Q-learning和SARAS(state action reward state action)为例,这一

系列的算法通过建立价值函数,并使用价值函数 $Q(s, a)$ 评估在 s 状态下采取 a 动作的预期回报,因这种计算,此类算法称为基于价值的算法。这类算法与深度学习算法适配性佳,经过不断发展、演变。逐渐形成了深度强化学习算法,如 DQN(deep Q network)(Mnih 等,2013)。2)以 Reinforce 为例,这一系列算法的核心是建立可微的策略函数,通过与环境交互评估策略的性能,接着使用策略梯度算法来优化算法,因此这一系列算法称为基于策略的算法,信任域策略优化(trust region policy optimization, TRPO)(Schulman 等,2015)和近端策略优化(proximal policy optimization, PPO)(Schulman 等,2017)是此类算法的典型代表。3)以 SAC(soft actor-critic)为例,这一系列算法结合了上述两种系列算法的特点,分别建立两个模块:actor,负责生成动作;critic,负责评估生成的动作。算法结合了策略优化的灵活性和价值估计的稳定性,使其在复杂环境中也能高效地学习。这一系列较为经典的算法包括:A3C(asynchronous advantage actor-critic)(Mnih 等,2016)、DDPG(deep deterministic policy gradient)(Lillicrap 等,2016)、TD3(twin delayed deep deterministic policy gradient)(Fujimoto 等,2018)、DSAC-T(distributional soft actor-critic with three refinements)(Duan 等,2023)等。图5所示包含了上面3种系列无模型强化学习算法的框架结构和发展进程。

MBRL 需要显式地建立动态环境模型,这样的方式可以提高数据利用率,但是需要额外优化环境模型,增加了计算复杂度。MBRL 适合可建模的确定性或者低随机性任务,如 Dyna-Q(Peng 和 Williams,1993)、MBPO(model-based policy optimization)(Janner 等,2019)、MuZero(Schrittwieser 等,2020)等算法。

3.1.2 经典强化学习自动驾驶案例

上述3类算法都有着各自的优势和不足,有的因为其算法的特性与自动驾驶磨合不佳,而也有许多的经典强化学习算法在自动驾驶中的运用已经较为成熟,其中有一些经典的框架包括结合变分自编码器(variational auto-encoders, VAE)和RL的强化学习自动驾驶(Kendall 等,2019)、基于DDPG的自动驾驶(Wang 等,2018)、基于离线强化学习的自动驾驶(Shi 等,2021)、基于结合模仿学习和强化学习的自动驾驶(Lu 等,2023),还有一些较为新颖的研究方

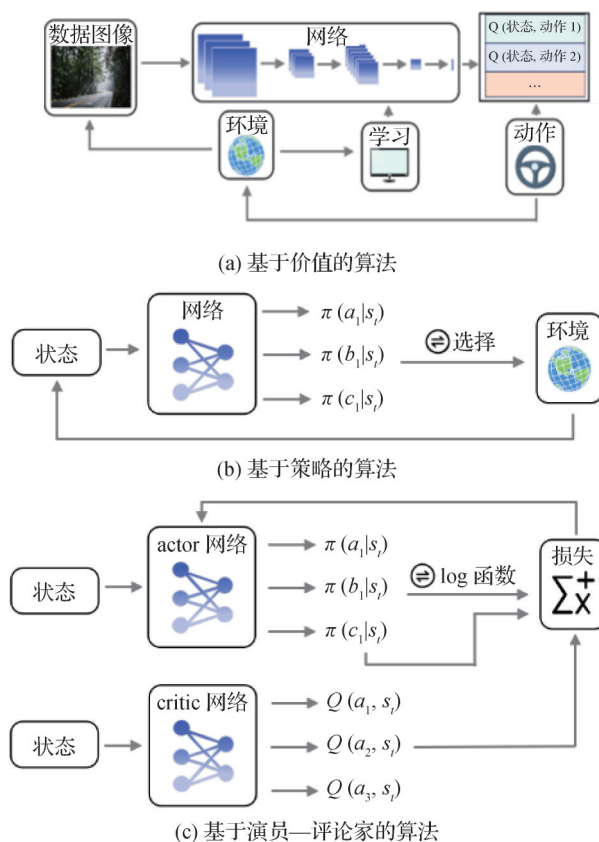


图5 3类无模型强化学习算法

Fig. 5 Three types of model-free reinforcement learning algorithms ((a) value-based algorithm; (b) policy-based algorithm; (c) based on the actor-critic algorithm)

向,如多智能体强化学习(multi-agent reinforcement learning, MARL)路径规划(王毅然 等,2019)和基于导向强化学习的路径规划(周彬 等,2021)。另外,深度强化学习方面的综述(刘全 等,2018),可供研究者对深度强化学习辅助自动驾驶的理论进行深入了解。

3.2 基于世界模型的强化学习

随着自动驾驶和强化学习的发展,世界模型的研究得到了显著推进,并在复杂动态环境建模任务中展现出优异效果。基于世界模型的强化学习(world-based RL, WBRL)属于 MBRL,但其更像是 MBRL 的进阶范式, Dyna-Q 利用表格记录状态—动作转移矩阵,而 MBPO 则通过神经网络预测下一状态分布,但这样的形式无法处理高维观测数据。Google 团队提出的深度规划网络 PlaNet(deep planning network)利用循环状态空间模型(recurrent state space model, RSSM)将环境观测数据压缩编码为低维潜空间,通过循环神经网络(recurrent neural net-

work, RNN)预测状态转移和奖励,最后利用交叉熵在潜空间搜索最优动作序列,生成多步预测,选择最优策略。这种方式解决了MBPO无法处理高维观测数据的问题。DreamerV1(Hafner等,2020)则是PlaNet的全面进化,引入了潜空间想象和梯度传播,即通过价值网络(critic)和策略网络(actor)在潜空间中优化行为,剔除在线规划模块,利用反向传播将长期奖励信号传递至策略网络,提升决策对未来的判断准确性。DreamerV2(Hafner等,2021)在DreamerV1的基础上加入了分类变量表示状态,也就是将潜状态进行离散化,而DreamerV3(Hafner等,2023)则引入了符号编码、KL(Kullback-Leibler)平衡技术,以解决过拟合问题。Dreamer系列算法的提出使得WBRL成为重要的自动驾驶研究方向。

WBRL在路径规划中可分为两类。一是隐式世界模型驱动规划,包括:1)Dreamer系列通过反向传播直接优化未来轨迹;2)基于Transformer的世界模型将图像拆分为Token,捕捉时空依赖以预测路径,如Transformer-XL(Robine等,2023)。二是多模态生成式规划,包括:1)通过扩散模型生成多模态轨迹,如GAIA-1(Bu等,2021),Vista(Gao等,2024);2)利用Transformer解码器生成路径点,如:DrivingGPT(driving generative pre-trained Transformer)(Chen等,2024)。如何建立世界模型,已逐渐成为强化学习在自动驾驶中的重要研究方向。

3.3 自动驾驶虚拟仿真系统

3.3.1 虚拟仿真系统概述

近几年,为了解决实车训练样本不足、实地训练安全性差等问题,业内提出了虚拟仿真驾驶环境的概念,随之出现了一系列开源或闭源的虚拟仿真驾驶平台,这些虚拟仿真驾驶平台大量运用在科研领域或者商业领域。根据自动驾驶轨迹规划所需要解决的问题,以及虚拟驾驶仿真平台应有的功能,本文提出了12个关键指标评估各仿真平台的优势和不足,以确定这些虚拟平台适用的科研方向。1)场景构建,在平台中创建和修改环境的支持度,驾驶场景可修改性越好,生成的数据集越全面;2)车辆动力学模型,车辆行为模拟的准确性,需要能精确地模拟车辆的加速、制动和转向等行为;3)传感器数据是自动驾驶系统感知环境的关键组件,高级的传感器模型将协助研究人员更精确地测试感知算法;4)交通参与者,适用于测试车辆算法的避障能力和应对突发

情况的能力;5)天气与时间变化数据同交通参与者,保证数据集的全面性和多样性;6)数据记录,是自动驾驶中的重中之重,其需要不断记录仿真过程的数据,并根据回放得到的数据进行分析以调整算法;7)API(application programming interface)接口,许多情况下,实验者常需要平台搭配其他软件或硬件共同开发;8)真实度与性能,仿真的逼真程度决定了测试环境的准确性,平台的运行效率影响开发时间成本和效率;9)开放性与可扩展性,平台是否开源,以及用户自定义内容的程度,个人学者和资金并不富裕的企业,更倾向于选择一些开源的平台;10)社区支持,算法的研发往往会遇到各式各样的问题,因此开发者社区的支持情况及活跃度也是平台评价的指标之一;11)是否有Gym接口是RL引入的一个接入点,现在许多开源的RL算法都在Gym环境下运行;12)是否支持MARL,多智能体RL也是如此,使用者往往考虑所用平台是否有能力测试多智能体RL算法的能力。上述指标中,指标1)、4)、5)充分考虑了平台的数据集多样性,以使生成的模型泛化性较强;指标2)、3)、8)则是考虑数据的准确性和有效性;指标6考虑到了数据集的生成和记录;指标7)、11)、12)考虑了平台接口的全面性;指标9)、10)考虑了平台的社区支持度。表7是根据CarSim、CarMaker、PreScan、PTV、TESS-NG、NVIDIA DRIVE、Carla和Apollo等主流虚拟仿真平台在上述评价指标下的评测结果。

3.3.2 虚拟仿真系统适用研究领域

结合上述通用指标和各个平台的优缺点进行分析,可得出如下结论:1)CarSim能够提供高精度的车辆动力学系统,并且生成多样化场景,编程语言适配MATLAB,适合熟悉MATLAB的用户使用,但是界面操作复杂,专业化程度高。2)CarMaker同样具有高精度车辆动力学系统,支持硬件在环测试,适合自动驾驶控制策略验证,但是操作界面较为简陋,对平台学习成本高。3)PreScan与CarSim兼容,支持各种传感器,适合自动驾驶算法测试,适配MATLAB和Simulink。4)PTV Vissim支持微观交通流,支持多样化交通道路,适合进行城市交通规划和交通要素交互,但是学习成本不低,复杂功能需要专业化学习。5)TESS-NG(TESS next generation),提供Python和C++ API,支持二次开发,设置所需的环境,使用的人数较少,社区支持差,适合相关专业和Python/C++编程语言使用者。6)NVIDIA DRIVE Sim各方面支持

表 7 主流自动驾驶虚拟仿真平台评测表

Table 7 Evaluation table of mainstream autonomous driving virtual simulation platforms

功能/平台	CarSim	CarMaker	PreScan	PTV	TESS-NG	NVIDIA	Carla	Apollo	LGSVL	Sumo	Gazebo	JunSim
场景构建	有限	中等	高级	中等	中等	高级	高级	高级	高级	高级	中等	高级
车辆动力学	高级	高级	中等	有限	有限	高级	中等	中等	中等	有限	中等	中等
传感器模型	有限	中等	高级	有限	有限	高级	高级	高级	高级	有限	有限	高级
交通参与者	有限	中等	高级	高级	中等	高级	高级	高级	高级	高级	有限	中等
天气与时间	有限	有限	中等	有限	有限	高级	高级	高级	高级	有限	有限	中等
数据记录	中等	高级	高级	中等	有限	高级	高级	高级	高级	高级	中等	高级
API接口	中等	高级	高级	中等	有限	高级	高级	高级	高级	高级	高级	高级
真实度与性能	高级	高级	高级	中等	中等	高级	中等	中等	中等	有限	中等	高级
可扩展性	封闭	封闭	封闭	封闭	封闭	封闭	开源	开源	开源	开源	开源	封闭
社区支持	低	中	中	中	低	中	高	高	中	高	中	低
Gym接口	无	无	无	无	无	有	有	有	无	无	有	无
支持MARL	无	无	无	无	无	有	有	有	无	无	有	无

较好,适合端到端全栈开发,但是生态封闭并且对硬件需求高。7)Carla 支持 ROS 和 Python 二次开发,各方面支持性好,版本迭代快,与其他虚拟仿真驾驶平台适配性好,视觉渲染效果极佳,适合端到端全栈开发以及学术研究,但是大规模仿真可能出现性能瓶颈,对硬件要求高。8)Apollo 与 NVIDIA DRIVE Sim 相似,各方面支持好,适合端到端全栈开发,但是生态封闭且需要专业的计算平台。9)LGSVL Simulator 同 Carla 兼容主流框架 Autoware、ROS 等,适合自动驾驶算法验证,但是社区支持一般,功能不够完善。10)Sumo 非常适合交通流控制和生成,并且支持 Python 脚本控制和跨平台使用,但是缺乏精细的视觉效果,建议与其他虚拟仿真平台联合使用。11)Gazebo 的 ROS 集成效果佳,适合机器人开发,能够模拟复杂的机械交互,但是对自动驾驶的支持有限。12)JunSim 可以进行高效的大规模并行仿真,各方面支持性佳,适合端到端的自动驾驶开发和测试,但是社区支持度较低。图 6 展示了上述 12 种经典的虚拟驾驶平台。

综上所述,CarSim、CarMaker 二者适合车辆动力学与控制领域研究;LGSVL、PreScan、Carla、NVIDIA DRIVE Sim、JunSim、Apollo 由于其逼真的视觉效果适合传感器与感知算法研究或者端到端算法训练与测试。而 Sumo、PTV Vission、TESS-NG 常用于研究大规模路网和交通流仿真;Gazebo 与 ROS 系统适配性好,常用于机器人任务。其中 NVIDIA、

Carla、Apollo、Gazebo 非常适合多智能体自动驾驶。

3.3.3 虚拟仿真系统混合开发

随着各大仿真平台的迭代更新,一些平台为研究者提供了与其他仿真平台进行混合开发的接口,如 Carla 仿真器在新版本中附带了与 Sumo、PTV Vissim、CarSim 等虚拟平台的联合仿真示例,利用 Sumo 等仿真器加强仿真中的交通流效果,因此研究者可以根据实际需求选择其中几种虚拟仿真器进行联合仿真以达到所需的效果。

3.4 强化学习算法研究方向

3.4.1 安全强化学习

生成安全路径是轨迹规划的重中之重,即使在放弃效率的情况下,自动驾驶也要保证行车安全。因此,RL 算法必须遵守行车安全约束,根据这一点,有科研团队提出在 RL 基础上加入无监督学习算法 K-means 聚类作为安全机制(Hsu 等,2022);也有团队结合无模型 RL 和有模型控制器,提出联合控制的新思路,以此提高生成轨迹的安全性(Gangopadhyay 等,2023);Wang 等人(2023)提出了基于 actor-critic 的框架 SRL-TR2(safe reinforcement learning based trajectory tracker framework),采用自控制和专家控制相结合,防止自控制出现安全问题,提高行车稳定性。

3.4.2 高效强化学习

强化学习需要花时间成本训练,才能收敛到较好的性能,因此如何提升样本效率是强化学习长期

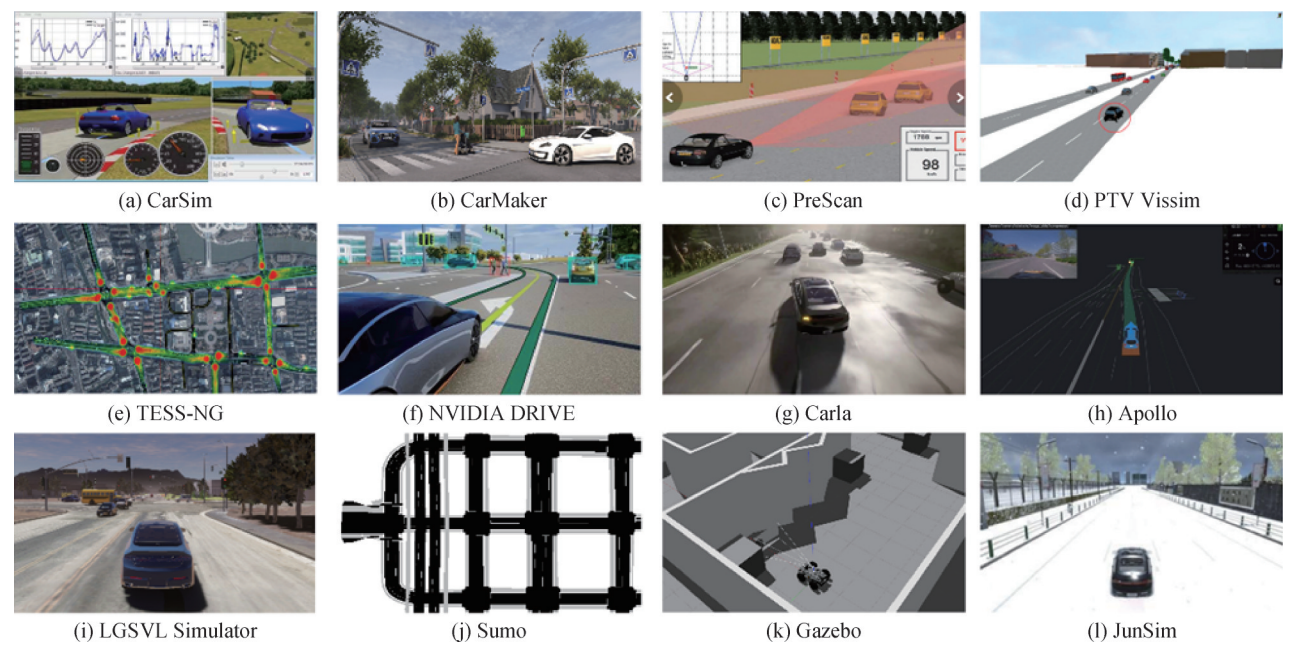


图6 12种经典的虚拟驾驶平台

Fig. 6 12 classic virtual driving platforms ((a) CarSim; (b) CarMaker; (c) PreScan; (d) PTV Vissim; (e) TESS-NG; (f) NVIDIA DRIVE; (g) Carla; (h) Apollo; (i) LGSVL Simulator; (j) Sumo; (k) Gazebo; (l) JunSim)

研究的方向之一。针对这一点,提出了一种用于自动PID(proportional integral differential)调整的熵最大化双延迟深度确定性策略梯度的方法,该方法能快速收敛到全局最优,提高样本效率(Chowdhury等,2023);同样是针对样本效率问题,也有团队通过建立具有轨迹辨别的世界模型与其对应的轨迹判别器、学习动态信息,并根据状态值估计权重,快速优化策略(Qu等,2024);Xu等人(2024)从轨迹数据集覆盖率低的问题入手,提出基于控制理论和规则约束的数据增强方法。

3.4.3 函数优化强化学习

在车辆学习过程中,奖励函数设计是关键,好的奖励函数方可指导车辆做出正确的决策动作。但针对不同环境,奖励函数的设计也有所不同,其设计难度极大。针对这一问题,Goto等人(2023)提出了一种强泛化的奖励函数,考虑复杂的行车环境,在奖励函数中加入速度保持、超车动机、碰撞避免等参数。随着大语言模型的兴起,Han等人(2024)提出利用大语言模型设置奖励函数的方法。

3.5 强化学习算法的新发展趋势

3.5.1 终身强化学习

终身强化学习(lifelong reinforcement learning, LRL),指模型能够像人类一样,不断学习新知识,同时不会遗忘之前学到的知识。这种算法的难点在于

如何解决“灾难性遗忘”,即在学习新任务的过程中如何保持对旧任务的记忆,对此,Diaz-Ruiz等人(2022)提供了可用于终身学习的数据集Ithaca365,该数据集创新点在于通过历史经验复用机制为自动驾驶车辆构建记忆;Meng等人(2025)提出了基于贝叶斯非参数模型的机器人终身强化学习框架LEGION(language embedding-based generative incremental off-policy reinforcement learning framework with non-parametric Bayes)。

3.5.2 分层强化学习

分层学习(hierarchical learning, HL),通过实现层次结构,将复杂问题分解为多个子任务简化学习过程,每个层次可以有不同的时间尺度和训练目标,提高学习效率和性能。其中高层策略负责决定智能体选择要执行的子任务或动作序列,低层策略负责在每个子任务内部进行具体的行动选择,为了实现HL和RL相结合,Mao等人(2024)提出了基于技能的分层强化学习方式。相比于传统RL提高了相似任务的学习效率。

3.5.3 多智能体学习

多智能体强化学习(MARL),指多个智能体共享同一环境,智能体之间互相协作和竞争来解决问题,多个智能体相互作用构建多智能体系统,每个智能体都可以根据环境和任务做出相应的决策。目

前主流的自动驾驶强化学习方法主要聚焦于单一智能体的奖励函数优化,其局限性在于未能充分建模多车辆交互的动态博弈过程。MARL通过构建分布式协同决策框架,为解决该领域的群体智能协调问题提供了新的理论视角和技术路径。潘振华等人(2024)详细介绍了无人机集群博弈模型,涉及多智能体合作、对抗等模型,在多智能体自动驾驶研究中有着重要的参考价值;Wang等人(2024a)为了解决高速公路车辆合并时的优化问题,改良了多智能体强化学习算法使用方式,实现平稳和安全的交通流。终身学习、分层学习和多智能体学习架构如图7所示。

3.5.4 基于预训练VLM的VLA框架

VLM (vision-language-model) 与 VLA (vision-language-action) 作为跨模态推理的典型架构,最初应用于机器人交互决策,近年被迁移至自动驾驶端到端模型并实现功能解耦:VLM 聚焦于环境建模与语义理解,通过视觉编码器提取图像特征,并利用语言模型对齐文本指令,实现交通场景元素识别与导航指令解析,但其输出严格限定于语义表征层;而VLA在继承VLM多模态感知能力的基础上,创新引入分层动作框架,通过宏观轨迹规划器生成粗粒度路径点,再经微观控制器(neural MPC)输出转向角与加速度的连续控制信号,由此构建“感知—决策—执行”的闭环链路。图8是本文总结的2019年—2024年期间较为经典的VLM、VLA以及对应数据集时间表。包括:ViLBERT(vision-and-language bidirectional encoder representations from Transformers) (Lu等, 2019)、LXMERT(learning cross-modality encoder representations from Transformers) (Tan 和 Bansal, 2019)、CLIP、Flamingo(Arai等, 2025)、BLIP-2(Li等, 2023a)、Kosmos-2(Peng等, 2023)、RT-2(robotics Transformer-2) (Brohan等, 2023)、Qwen2-VL(Qwen2-vision language) (Wang等, 2024c)、Gemini-1.5(Gemini Team, 2024)、OpenVLA(Kim等, 2025)、MindVLA、CoVLA(Arai等, 2025)、HybridVLA(Liu等, 2025a)、iRe-VLA、AlphaDrive(Jiang等, 2025)等。其中,ViLBERT与LXMERT最先提出视觉—语言联合表征,二者分别通过共同注意力机制和层次化跨模态交互对齐视觉图像特征与文本特征,但二者都局限于固定模态输入;CLIP(contrastive language-image pre-training)突破性地引入对比学习,在大规

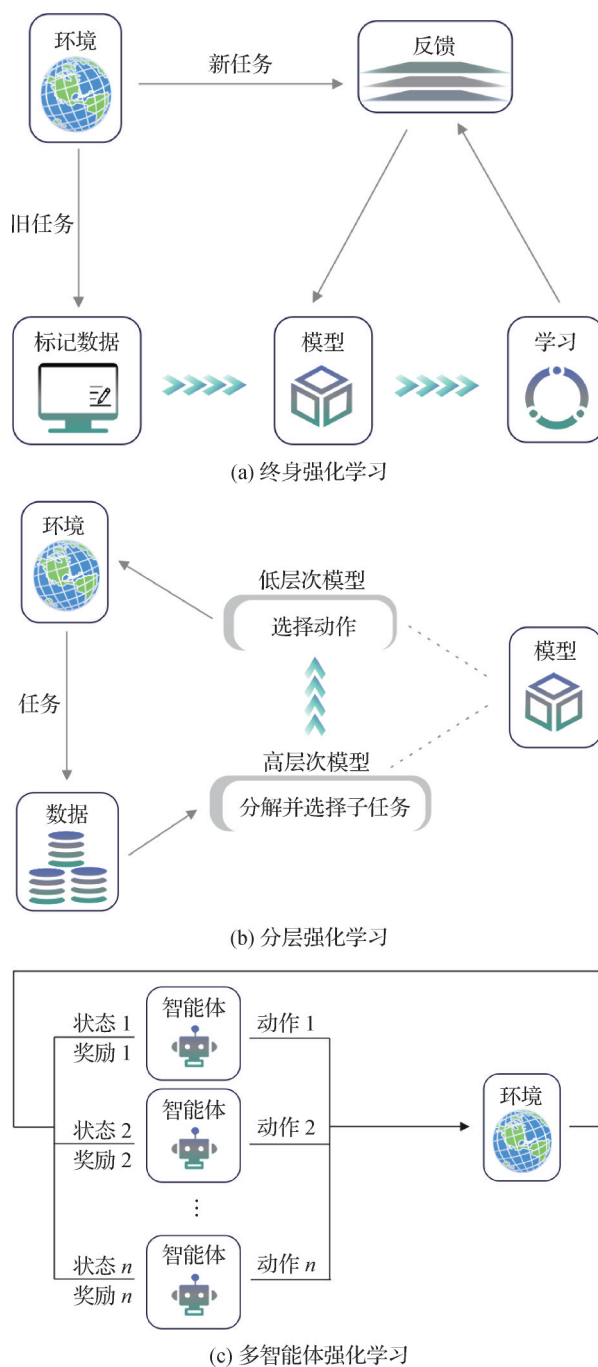


图7 终身强化学习、分层强化学习、多智能体强化学习架构

Fig. 7 The architectures of lifelong reinforcement learning, hierarchical reinforcement learning and multi-agent reinforcement learning ((a) lifelong reinforcement learning; (b) hierarchical reinforcement learning; (c) multi-agent reinforcement learning)

模图一文对上进行训练,实现图文的语义对齐,为后续模型的通用视觉编码器打下基础;Flamingo首次支持交错图文输入,算法通过压缩视觉特征,结合门控交叉注意力将多模态信息注入到语言模型中,自此获得上下文学习能力;BLIP-2提出Q-Former桥接

冻结的视觉编码器与语言模型,实现零样本任务迁移,降低了VLM框架的训练成本;同年,RT-2将视觉语言模型扩展至机器人控制,将连续动作离散化为语义 token,通过生成指令驱动机器人动作,推动了VLA范式的发展;Qwen2-VL引入动态分辨率机制与多模态旋转位置编码,支持所有分辨率图像和视频时空建模;Gemini-1.5则是专注于更新效率,基于混合专家架构实现百万级 token 长上下文处理,在视频理解与多模态推理中接近人类真实表现。随着具身智能技术的发展,越来越多的VLA算法相继提出,OpenVLA整合多样化演示数据,在RT-2动作 token

化的基础上引入扩散策略提升泛化性;MindVLA提出融合3D高斯与扩散模型,赋予自动驾驶空间推理与博弈能力;HybridVLA在单一语言模型中整合自回归与扩散策略,通过联合噪声动作编码与置信度融合机制,超越传统VLA方法;CoVLA则是较为前沿的自动驾驶综合数据集,联合VLA信号,推动真实场景下的闭环控制研究;最新的AlphaDrive模型则是在之前VLA框架的基础上引入强化学习机制,通过在线策略优化与场景树搜索提升自动驾驶规划安全性,在复杂博弈场景中实现SOTA(state of the art)。

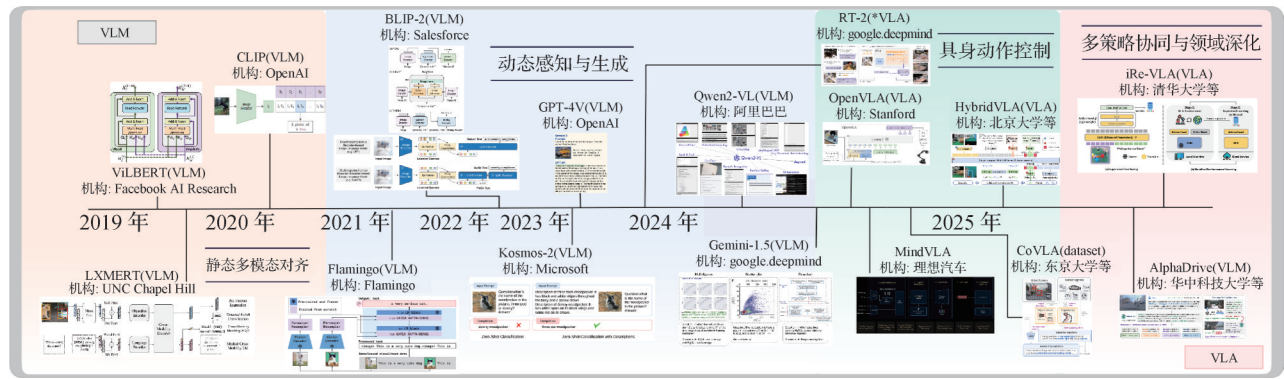


图8 VLM和VLA历史经典算法时间表

Fig. 8 Timeline of VLM and VLA historical classical algorithms

上述VLA框架大多基于多模态输入,通常是视觉图像信号和语言指令,依此生成可执行的动作序列。其架构通常包含:1)视觉编码器,用于提取图像或视频的视觉特征;2)语言模型,用于解析自然语言指令(常用的大语言模型);3)动作生成模块,用于将获取的多模态信息映射为物理动作。强化学习在VLA中有4种新颖的用法:1)模型微调:通过RL动态调整模型策略,优化VLM的推理能力,相比于监督微调(supervised fine-tuning, SFT),提升了对于复杂任务的泛化性,并减少了数据依赖,如Visual-RFT(reinforcement fine-tuning)(Liu等,2025b),这一用法出现于大语言模型(large language model, LLM)的强化学习微调中,如DeepSeek-R1(DeepSeek-AI, 2025)。2)提升泛化:通过RL和监督学习的交替使用提升模型对环境的识别能力和重塑能力,其中在线RL阶段需要冻结VL主干以训练动作头,而监督学习阶段负责融合专家数据与RL的探索数据,如:iRe-VLA(Guo等,2025)。3)联合扩散:结合扩散模型和RL模型的特性,根据扩散模型生成候选轨迹,

并通过RL的奖励函数筛选最优路径。如: Mind-VLA。4)驾驶风格:利用驾驶数据与基于人类反馈的强化学习(reinforcement learning with human feedback, RLHF)技术对齐驾驶人员偏好,如特斯拉通过影子模式采集人类驾驶数据,构建偏好数据集,结合RL生成符合驾驶习惯的驾驶策略。基于上述RL的应用,为VLA在复杂视觉推理、动态环境交互及多模态决策等场景提供了多维度突破,标志着RL技术正在成为推动具身智能发展的重要范式。

4 基于模仿学习的轨迹规划

基于RL的轨迹规划能够探索全部的搜索空间,但是这样探索的效率是极低的,而模仿学习可以直接根据专家示范,如人类驾驶轨迹数据,虚拟仿真驾驶平台中生成的驾驶轨迹数据,快速掌握任务策略,而无需通过试错探索以积累经验。在自动驾驶中,模仿学习的这一特性可以使算法跳过强化学习的稀疏奖励探索阶段,而直接进入接近人类驾驶员的驾

驶策略,实现自动避障和路径选择,显著缩短自动驾驶模型的训练周期。

4.1 经典模仿学习

模仿学习中较为主流且经典的算法分为行为克隆、逆强化学习和生成式对抗模仿学习3种。1)行为克隆(behavior cloning, BC):其思想是直接模仿专家策略,通过监督学习从专家数据中找到一种从状态到动作的最佳映射,优点是无需设计奖励函数,直接利用高质量的专家数据,但是其复合误差问题严重,在自动驾驶中可能导致策略在未知状态下出现崩溃;2)逆强化学习(inverse reinforcement learning, IRL):其思想是从专家数据中反推隐含的奖励函数,再通过强化学习优化策略,这种方式的解释性优势在于奖励函数能揭示任务本质(如自动驾驶中辨识关键决策要素),但存在奖励函数多解性及潜在歧义性问题;3)生成式对抗模仿学习(generative adversarial imitation learning, GAIL):结合生成式对抗网络(generative adversarial network, GAN)的模仿学习方式,通过对抗训练使策略生成的轨迹分布更为逼近专家轨迹分布,其中对抗训练是指生成器生成虚拟的状态动作试图骗过判别器,而判别器区分专家数据和生成数据,这样的优点是策略更为逼近专家轨迹分布,但是训练不稳定,而且对参数敏感。模仿学习BC和IRL的架构如图9所示。

4.2 扩散模型

扩散模型(diffusion model, DM)是一种生成模

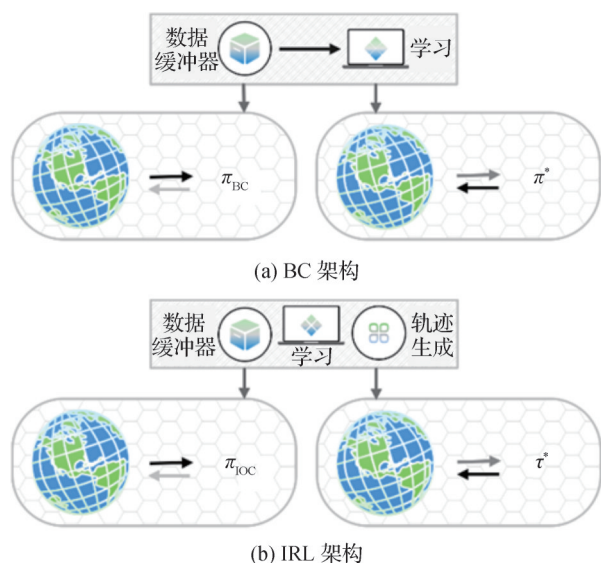


图9 模仿学习BC和IRL架构图

Fig. 9 Imitation learning BC and IRL architecture diagrams
(a) BC architecture; (b) IRL architecture)

型,常用于图像生成。在自动驾驶中,扩散模型常用于从专家示范数据中生成行为策略,这一点符合模仿学习的特性,因此可以将其在自动驾驶中的运用归类为模仿学习,扩散模型在自动驾驶规划中主要以目标路线、感知层实时数据等作为输入在潜在空间生成候选路径,通过正向制造噪声和反向去除噪声的形式迭代并优化路径,如HE-Drive(Wang等,2024b)。基于扩散模型路径选择还可以结合强化学习方法进行路径评分,提高选取最优路径概率。图10展示了几种较为经典的模仿学习方法以及在自动驾驶中的运用,包括:Chauffeurnet(Bansal等,2019)、GAN-based(Kuefler等,2017)、3D Diffusion Policy(Ze等,2024)、IRL-based(Phan-Minh等,2022)、Diffusion-based(Zheng等,2025)。

4.3 模仿学习算法研究方向

4.3.1 在线模仿学习

在线模仿学习中较为经典的算法如:DAgger(dataset aggregation)算法,旨在解决离线模仿学习,尤其是BC算法中因训练数据与测试数据分布不匹配导致的问题(Ross等,2011),在自动驾驶中,这可能是因为两个道路或状态空间过于相似,导致策略选择了错误的动作。DAgger利用在线同时标记专家动作和预测动作减少分布偏移和复合误差,以解决BC中存在的问题,然而这样的做法会较大地提升计算复杂度。

4.3.2 自模仿学习

自模仿学习(self-imitation learning, SIL)并非传统意义上的模仿学习,而是结合了RL的混合学习方式,SIL舍弃了直接从已知数据中获取专家数据,而是从RL与环境交互中直接筛选出高奖励轨迹作为专家数据,通常策略优化需要同时考虑RL和IL的损失,重构混合损失函数(Oh等,2018)。这种方式的优点在于高效利用相关数据,且无需人工标注和专家示范,减少了RL的盲目探索,但是抑制RL的探索智能体可能导致过早地收敛到次优策略,陷入局部最优。

4.3.3 仅状态模仿学习

仅状态模仿学习(imitation from observations, IFO)顾名思义是指仅通过专家状态进行模仿学习而无需专家动作数据(Torabi等,2018)。这种方式已经在机器人领域取得了一定的成功,如:DeepMind在连续视频中的学习框架(Carreira等,2024)。

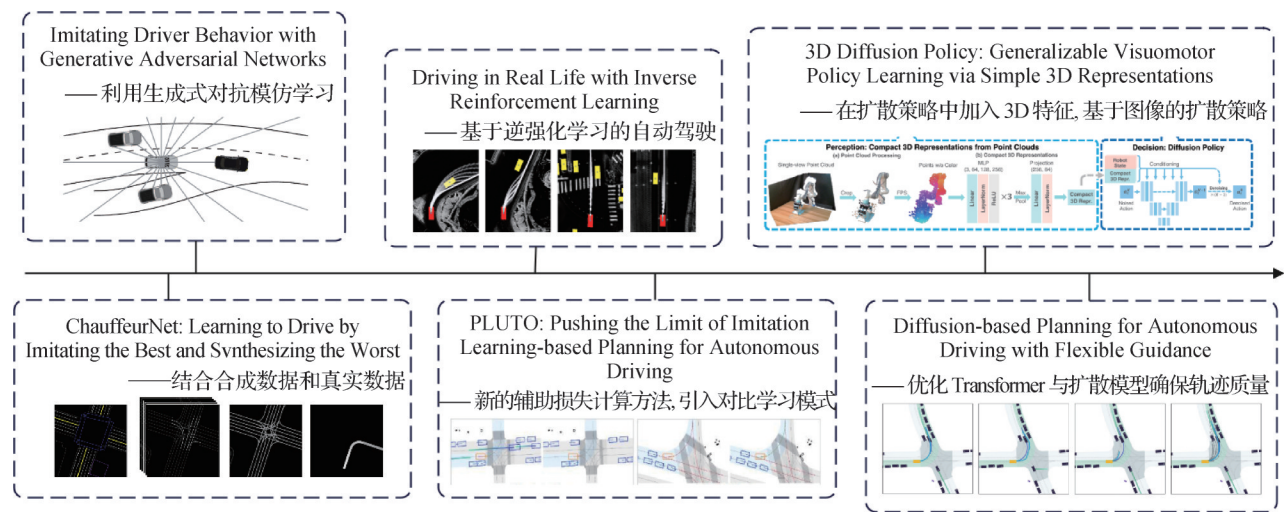


图 10 自动驾驶中的经典模仿学习
Fig. 10 Classical imitation learning in autonomous driving

如此的学习方式优势明显,可以直接将驾驶视频、图像作为状态输入,而无需记录驾驶行为如油门、转向等,避免了动作抖动,如人类驾驶过程中方向盘无意义的轻微移动。然而这种学习方式可能会加重模仿学习的因果混淆问题。

4.3.4 多智能体模仿学习

多智能体模仿学习 (multi-agent imitation learning, MAIL) 与多智能体强化学习同属于多智能体问题,MAIL 通过观察专家协同数据,学习联合协作策略。学习方式采用集中训练和分散执行,常采用分层策略架构,高层策略分配驾驶路径,低层策略生成转向控制。MAIL 可以提升智能体的协作能力,但是大幅度提升了通信开支,需要全局进行输出传输,数据量大,常被运用于自动驾驶路口或者变道协作中,如王焱宁等人 (2025) 提出的基于混合模仿学习的多智能体追捕决策方法。

4.3.5 生成式模仿学习

生成式模仿学习 (generative imitation learning, GIL) 是一种与扩散模型相仿的模仿学习方式,但是侧重点略有不同,基于扩散模型的路径规划可以直接生成候选路径,而生成式模仿学习则利用生成模型生成更为逼真的专家数据以增强模仿学习的效果,如:3D Diffusion Policy。这里的扩散模型也可以替换为生成式对抗网络 (GAN) 或变分自编码器 (VAE) 等生成模型,前者即 GAIL 算法。但是 GIL 方法极大地依赖数据质量,若是生成模型训练不充分,反而将阻碍后期的策略学习。

4.3.6 元模仿学习

元模仿学习 (meta imitation learning, MIL) 旨在通过少量专家演示就能适应新任务,与迁移学习的特性一致,元模仿学习需要在训练阶段让智能体“学会模仿” (Finn 等, 2017)。元模仿学习可同时进行多个路径规划任务,每个任务提供少量专家数据,首先针对每个独立任务调整策略,接着统筹调整全局策略,以实现调整后的策略在所有任务上达到最佳,如: PLUTO (pushing the limit of imitation learning-based planning for autonomous driving) (Cheng 等, 2024)。这样的学习方式跨任务泛化能力强,但是依赖任务分布和数据量,若是智能体并未“学会模仿”,会导致策略效果降低。

4.4 基于扩散模型的 VLA 框架

在之前的论述中,提到了基于预训练 VLM 的 VLA 框架,本节将介绍一种基于扩散模型的 VLA 架构,这一类型的架构分为 3 个部分: 1) 多模态编码器: 类似于 VLM 中的 VL 部分,需要通过视觉编码器和文本编码器对图像特征和文本指令进行编码,并通过跨模态注意力机制进行融合。2) 扩散策略头: 将融合后的特征向量作为条件输入,通过扩散模型生成连续动作。例如, Diffusion Policy 通过时间序列扩散 Transformer, 生成高维动作序列, 并支持后退视界控制 (Chi 等, 2025)。3) 动作解码器: 解码器负责将去噪后的潜在变量解码为车辆执行器所需的信号, 如转向或油门等。以扩散模型为基础的 VLA 模型能够捕捉动作内部多模态分布特性, 避免了自回

归方法中动作离散化而导致的可能存在的量化误差,但是缺失LLM的推理能力,对于复杂文本指令的理解能力有限,依赖于预定义的模板文本指令,并且并不像基于VLM的VLA模型那样将指令进行分割,而是用自然语言形式表示。

5 发展趋势与挑战

对于自动驾驶轨迹规划甚至端到端模型可能呈现多样化的趋势,并没有一种框架能够在这场自动驾驶博弈中取得碾压式的实力,未来的自动驾驶竞争将更为激烈和细化,但是不可避免的是,企业和研究者会将注意力从基于规则的路径规划向基于学习模型的路径规划转变,以下是当下3个重要研究方向:

1)基于预训练VLM的VLA框架。VLA能够利用大语言模型的推理能力,为原本为黑箱操作的自动驾驶路径决策提供了自然语言解释,譬如:前方出现行人,需紧急制动。这一点对于现有的黑箱式端到端模型框架是重要的突破点。

2)扩散模型结合强化学习。结合扩散模型和强化学习的自动驾驶范式,扩散模型作为新式的模仿学习算法与强化学习进行联合,利用扩散模型的多模态动作生成能力和强大的环境建模能力,结合强化学习的奖励机制和微调能力,可实现更为优异的自动驾驶效果。

3)基于世界模型的强化学习。建立世界模型后可以显著提高样本效率,减少强化学习交互需求,加速算法策略优化进程。基于大型世界模型的强化学习算法的多任务泛化能力要高于无模型强化学习。基于世界模型的强化学习可以适应突发变化,迅速更换候选路径,而无模型强化学习只能等待环境交互反馈。

上述3个研究方向并不是分离的,未来可能实现技术互补,将扩散模型的动作生成能力与VLA的多模态理解能力相结合,并且扩散模型强大的建模能力也可以构建世界模型,实现三者技术上的统一,而VLA也可以根据世界模型生成的合成数据弥补数据稀缺的问题。然而为了实现高质量的自动驾驶路径规划还需要解决下列几个问题:

1)世界模型中存在的问题(虚拟到现实)。如今许多成果基于虚拟仿真系统,而如何将研究成果快

速转换为现实成果是自动驾驶现有重要研究方向之一,也是亟待需要攻克的问题。对于此方面的内容,Zhao等人(2020)对于域随机化、域适应、模仿学习、元学习和知识蒸馏等虚拟到现实技术进行了详细的总结,而胡学敏等人(2024)则是将虚拟到现实的方法归结为虚实迁移和平行智能两类,并分别概述上述二者如何解决在现实环境中的适配问题,从而对于解决虚拟到现实问题提供了良好的思路。

2)VLA架构中存在的问题(数据需求)。VLA依赖多模态数据,而真实世界的复杂状态空间如:极端天气、突发情况,远超合成数据所能覆盖的范围,长尾问题依然存在。虽然可以通过世界模型合成数据进行缓解,但真实且多样的真实数据难以完全覆盖。对此,PLUTO框架采用对比学习方式,在训练时对原始数据进行扰动,引导模型学习各场景中的不变特征。例如,PLUTO框架通过此方法将长尾场景避让成功率提升了27%,为未来发展提供了一条路径。

3)扩散模型中存在的问题(算力瓶颈)。基于扩散模型的算法的迭代时间成本高,难以满足车辆的实时需求;生成高精度轨迹可能需多次采样,因为车载芯片的算力限制必然迫使减小模型复杂度,从而导致生成结果损失部分精度。如何突破此瓶颈将是工业界和学术界长期面临的挑战。

为了解决上述问题并实现更为高效的自动驾驶,一方面需要对现有的车载系统进行更新,以满足更为庞大的计算需求;另一方面还需要持续探索软件及框架上的待优化点,以求用较小的资源消耗获得较高的效益。

6 结 语

本文对自动驾驶轨迹规划领域进行了探讨,首先介绍了自动驾驶领域的两大框架:模块化以及端到端,并指出轨迹规划是自动驾驶框架中最为重要的功能。而后分别分析了历年文献中基于传统算法的轨迹规划、基于强化学习的轨迹规划以及基于模仿学习的轨迹规划的现有成果和未来的发展趋势。为了方便后续工作者能够更为便捷地在虚拟仿真平台中对代码进行测试,本文提出了12个通用指标评估当前虚拟仿真系统的性能、优缺点和适合的研究方向,期望为未来的研究者提供参考。

研究发现,传统算法在特定环境下效果较好,但

在复杂场景中受限,并且依赖高精地图;强化学习算法(无模型强化学习),在处理连续动作空间和复杂决策问题上展现出巨大潜力,并取得了一定成就。但是RL算法盲目探索时间过长,策略收敛过慢,并且在安全性和样本效率方面的挑战尚未完全解决,其在实际交通环境中的泛化能力仍需进一步验证;IL解决了RL盲目探索的问题,可以根据现有专家数据以较快的速度达到较好的策略,跳过盲目搜索阶段,但是可能存在因果混淆的问题且无法解决自动驾驶中的长尾问题。

未来的研究方向应包括:1)优化RL算法,尝试利用终身强化学习、分层强化学习和多智能体强化学习等新型的学习方式,针对性地克服传统强化学习在动态环境适应性、任务复杂性和多智能体协作方面的局限性;2)改进IL算法,尝试使用在线模仿学习、自模仿学习、仅状态模仿学习、多智能体模仿学习、生成式模仿学习和元模仿学习等新式模仿学习算法改善IL存在的问题,以增强IL的能力;3)引入扩散模型,并进行世界建模,从而实现基于世界模型的强化学习,引入VLA框架,以扩散模型和强化学习作为动作头,在前端插入视觉语言模型模块,以达成3项技术统一的VLA架构,并且需在更为复杂的交通环境中测试和验证所提出的上述VLA模型,以提高其在实际应用中的泛化能力,同时对现有的硬件设施及车载设施进行性能升级,尝试使用虚拟到现实技术解决虚拟仿真系统和现实之间的差距,以此将算法从虚拟仿真平台上转移到实际生活中。通过这些努力,期望自动驾驶轨迹规划能够更加高效、安全,且适应各种复杂的交通环境。

参考文献(References)

- Ait Ben Mouh L, Ouhda M, El Mourabit Y and Baslam M. 2023. A hybrid approach of Dijkstra's algorithm and A* search, with an optional adaptive threshold heuristic//Proceedings of the 8th International Conference on Business Intelligence. Istanbul, Turkey: Springer: 117-133 [DOI: 10.1007/978-3-031-37872-0_9]
- Ait-Saadi A, Meraihi Y, Soukane A, Ramdane-Cherif A and Benmessaoud Gabis A. 2022. A novel hybrid chaotic aquila optimization algorithm with simulated annealing for unmanned aerial vehicles path planning. Computers and Electrical Engineering, 104: #108461 [DOI: 10.1016/j.compeleceng.2022.108461]
- Aradi S. 2022. Survey of deep reinforcement learning for motion planning of autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 23 (2): 740-759 [DOI: 10.1109/TITS.2020.3024655]
- Arai H, Miwa K, Sasaki K, Watanabe K, Yamaguchi Y, Aoki S, et al. 2025. CoVLA: comprehensive vision-language-action dataset for autonomous driving//Proceedings of 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Tucson, USA: IEEE: 1933-1943 [DOI: 10.1109/WACV61041.2025.00195]
- Bansal M, Krizhevsky A and Ogale A. 2019. ChauffeurNet: learning to drive by imitating the best and synthesizing the worst//15th Robotics: Science and Systems 2019. Breisgau, Germany: [s.n.]
- Beed R S, Sarkar S, Roy A, Biswas S D and Biswas S. 2020. A hybrid multi-objective carpool route optimization technique using genetic algorithm and A* algorithm [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2007.05781.pdf>
- Brohan A, Brown N, Carbajal J, Chebotar Y, Chen X, Choromanski K, et al. 2023. RT-2: vision-language-action models transfer web knowledge to robotic control//Proceedings of the 7th Conference on Robot Learning. Atlanta, USA: PMLR: 2165-2183
- Bu X Y, Peng J R, Yan J J, Tan T N and Zhang Z X. 2021. GAIA: a transfer learning system of object detection that fits your needs//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 274-283 [DOI: 10.1109/CVPR46437.2021.00034]
- Cai Y C and Du P Z. 2021. Path planning of unmanned ground vehicle based on balanced whale optimization algorithm. Control and Decision, 36(11): 2647-2655 (蔡雨岑, 杜鹏桢. 2021. 基于平衡鲸鱼优化算法的无人车路径规划. 控制与决策, 36(11): 2647-2655) [DOI: 10.13195/j.kzyjc.2020.0416]
- Carreira J, King M, Pătrăucean V, Gokay D, Ionescu C, Yang Y, et al. 2024. Learning from one continuous video stream//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 28751-28761 [DOI: 10.1109/CVPR52733.2024.02716]
- Chen R N, Hu J and Xu W C. 2022. An RRT-Dijkstra-based path planning strategy for autonomous vehicles. Applied Sciences, 12(23): #11982 [DOI: 10.3390/app122311982]
- Chen Y T, Wang Y Q and Zhang Z X. 2024. DrivingGPT: unifying driving world modeling and planning with multi-modal autoregressive transformers [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2412.18607.pdf>
- Chen Y Y, Tian D X, Lin C M and Yin H B. 2024. Survey of end-to-end autonomous driving systems. Journal of Image and Graphics, 29(11): 3216-3237 (陈妍妍, 田大新, 林椿昞, 殷鸿博. 2024. 端到端自动驾驶系统研究综述. 中国图象图形学报, 29(11): 3216-3237) [DOI: 10.11834/jig.230787]
- Cheng J, Chen Y and Chen Q. 2024. PLUTO: pushing the limit of imitation learning-based planning for autonomous driving [EB/OL].

- [2025-01-03]. <https://arxiv.org/pdf/2404.14327.pdf>
- Chi C, Xu Z J, Feng S Y, Cousineau E, Du Y L, Burchfiel B, et al. 2025. Diffusion policy: visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10/11): 1684-1704 [DOI: 10.1177/02783649241273668]
- Chowdhury M A, Al-Wahaibi S S S and Lu Q G. 2023. Entropy-maximizing TD3-based reinforcement learning for adaptive PID control of dynamical systems. *Computers and Chemical Engineering*, 178: #108393 [DOI: 10.1016/j.compchemeng.2023.108393]
- Chu D F, Wang R K, Wang J Y, Hua Q Z, Lu L P and Wu C Z. 2024. End-to-end autonomous driving: advancements and challenges. *China Journal of Highway and Transport*, 37(10): 209-232 (褚端峰, 王如康, 王竞一, 花俏枝, 陆丽萍, 吴超仲. 2024. 端到端自动驾驶的研究进展及挑战. *中国公路学报*, 37(10): 209-232) [DOI: 10.19721/j.cnki.1001-7372.2024.10.019]
- Claussmann L, Revilloud M, Gruyer D and Glaser S. 2020. A review of motion planning for highway autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 21(5): 1826-1848 [DOI: 10.1109/TITS.2019.2913998]
- Dang T V and Tan P X. 2024. Hybrid mobile robot path planning using safe JBS-A*B algorithm and improved DWA based on monocular camera. *Journal of Intelligent and Robotic Systems*, 110(4): #151 [DOI: 10.1007/s10846-024-02179-z]
- DeepSeek-AI. 2025. DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2501.12948.pdf>
- Diaz-Ruiz C A, Xia Y Y, You Y R, Nino J, Chen J N, Monica J, et al. 2022. Ithaca365: dataset and driving perception under repeated and challenging weather conditions//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 21351-21360 [DOI: 10.1109/CVPR52688.2022.02069]
- Dinnethew J, Boubezoul A, Mandiau R and Espié S. 2022. Multi-agent reinforcement learning for autonomous vehicles: a survey. *Autonomous Intelligent Systems*, 2(1): #27 [DOI: 10.1007/s43684-022-00045-z]
- Duan J L, Wang W X, Xiao L M, Gao J X and Li S E. 2023. DSAC-T: distributional soft actor-critic with three refinements [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2310.05858.pdf>
- Fan H H, Huang J H, Huang X Z, Zhu H J and Su H C. 2024. BI-RRT*: an improved path planning algorithm for secure and trustworthy mobile robots systems. *Heliyon*, 10(5): #e26403 [DOI: 10.1016/j.heliyon.2024.e26403]
- Finn C, Yu T H, Zhang T H, Abbeel P and Levine S. 2017. One-shot visual imitation learning via meta-learning//*Proceedings of the 1st Annual Conference on Robot Learning*. Mountain View, USA: PMLR: 357-368
- Fujimoto S, Hoof H and Meger D. 2018. Addressing function approximation error in actor-critic methods//*Proceedings of the 35th International Conference on Machine Learning*. Stockholm, Sweden: PMLR: 1587-1596
- Ganesan S, Natarajan S K and Srinivasan J. 2022. A global path planning algorithm for mobile robot in cluttered environments with an improved initial cost solution and convergence rate. *Arabian Journal for Science and Engineering*, 47(3): 3633-3647 [DOI: 10.1007/s13369-021-06452-3]
- Gangopadhyay B, Dasgupta P and Dey S. 2023. Safe and stable RL (S²RL) driving policies using control barrier and Control Lyapunov Functions. *IEEE Transactions on Intelligent Vehicles*, 8(2): 1889-1899 [DOI: 10.1109/TIV.2022.3160202]
- Gao S Y, Yang J Z, Chen L, Chitta K, Qiu Y H, Geiger A, et al. 2024. Vista: a generalizable driving world model with high fidelity and versatile controllability//*Proceedings of the 38th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: #2906 [DOI: 10.5555/3737916.3740822]
- Gemini Team. 2024. Gemini 1.5: unlocking multimodal understanding across millions of tokens of context [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2403.05530.pdf>
- Gong K, Xu Z Y and Zhang X Y. 2023. Bounded-DWA: an efficient local planner for Ackermann-driven vehicles on sandy terrain//*Proceedings of 2023 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. Datong, China: IEEE: 632-637 [DOI: 10.1109/RCAR58764.2023.10249383]
- Goto T, Kizumi Y and Iwasaki S. 2023. Design of reward function on reinforcement learning for automated driving. *IFAC-PapersOnLine*, 56(2): 7948-7953 [DOI: 10.1016/j.ifacol.2023.10.902]
- Guo Y J, Zhang J K, Chen X Y, Ji X, Wang Y J, Hu Y C, et al. 2025. Improving vision-language-action model with online reinforcement learning [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2501.16664.pdf>
- Hafner D, Lillicrap T, Ba J and Norouzi M. 2020. Dream to control: learning behaviors by latent imagination//*Proceedings of the 8th International Conference on Learning Representations*. Addis Ababa, Ethiopia: OpenReview.net
- Hafner D, Lillicrap T P, Norouzi M and Ba J. 2021. Mastering Atari with discrete world models//*Proceedings of the 9th International Conference on Learning Representations*. [s.l.]: OpenReview.net
- Hafner D, Pasukonis J, Ba J and Lillicrap T. 2023. Mastering diverse domains through world models [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2301.04104.pdf>
- Han X, Yang Q N, Chen X D, Cai Z H, Chu X W and Zhu M X. 2024. AutoReward: closed-loop reward design with large language models for autonomous driving. *IEEE Transactions on Intelligent Vehicles*: 1-13 [DOI: 10.1109/TIV.2024.3485964]
- Hsu H L, Huang Q H and Ha S. 2022. Improving safety in deep rein-

- forcement learning using unsupervised action planning//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). Philadelphia, USA: IEEE: 5567-5573 [DOI: 10.1109/ICRA46639.2022.9812181]
- Hu D D and Zhang Q. 2022. Lane changing trajectory planning of driverless vehicle based on PER-APF algorithm. *Computer Measurement and Control*, 30(6): 229-234, 241 (胡丹丹, 张琪. 2022. 基于PER-APF算法的无人驾驶汽车换道轨迹规划. *计算机测量与控制*, 30(6): 229-234, 241) [DOI: 10.16526/j.cnki.11-4762/tp.2022.06.035]
- Hu X M, Huang T Y, Yu Y L, Ren J J, Xie W and Chen L. 2024. Decision technologies of simulation to reality for autonomous driving: a survey. *Journal of Image and Graphics*, 29(11): 3173-3194 (胡学敏, 黄婷玉, 余雅澜, 任佳佳, 谢微, 陈龙. 2024. 仿真到现实环境的自动驾驶决策技术综述. *中国图象图形学报*, 29(11): 3173-3194) [DOI: 10.11834/jig.230780]
- Janner M, Fu J, Zhang M and Levine S. 2019. When to trust your model: model-based policy optimization//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #1122 [DOI: 10.5555/3454287.3455409]
- Jiang B, Chen S Y, Zhang Q, Liu W Y and Wang X G. 2025. AlphaDrive: unleashing the power of VLMs in autonomous driving via reinforcement learning and reasoning [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2503.07608.pdf>
- Jin X J, Yan Z Y, Yang H, Wang Q K and Yin G D. 2020. A goal-biased RRT path planning approach for autonomous ground vehicle//Proceedings of the 4th CAA International Conference on Vehicular Control and Intelligence (CVCI). Hangzhou, China: IEEE: 743-746 [DOI: 10.1109/CVCI51460.2020.9338597]
- Kendall A, Hawke J, Janz D, Mazur P, Reda D, Allen J M, et al. 2019. Learning to drive in a day//Proceedings of 2019 International Conference on Robotics and Automation (ICRA). Montreal, Canada: IEEE: 8248-8254 [DOI: 10.1109/ICRA.2019.8793742]
- Keselman A, Ten S, Ghazali A and Jubeh M. 2018. Reinforcement learning with A* and a deep heuristic [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/1811.07745.pdf>
- Kim M J, Pertsch K, Karamcheti S, Xiao T, Balakrishna A, Nair S, et al. 2025. OpenVLA: an open-source vision-language-action model//Proceedings of the 8th Conference on Robot Learning. Munich, Germany: PMLR: 2679-2713
- Kiran B R, Sobh I, Talpaert V, Mannion P, Sallab A A A, Yogamani S, et al. 2022. Deep reinforcement learning for autonomous driving: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6): 4909-4926 [DOI: 10.1109/TITS.2021.3054625]
- Kuefler A, Morton J, Wheeler T and Kochenderfer M. 2017. Imitating driver behavior with generative adversarial networks//2017 IEEE Intelligent Vehicles Symposium (IV). Los Angeles, USA: IEEE: 204-211 [DOI: 10.1109/IVS.2017.7995721]
- Li H C, Liu W J, Yang C, Wang W D, Qie T and Xiang C L. 2022. An optimization-based path planning approach for autonomous vehicles using the DynEFA-artificial potential field. *IEEE Transactions on Intelligent Vehicles*, 7(2): 263-272 [DOI: 10.1109/TIV.2021.3123341]
- Li J N, Li D X, Savarese S and Hoi S. 2023a. BLIP-2: bootstrapping language-image pre-training with frozen image encoders and large language models//Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: PMLR: 19730-19742
- Li X, Bai Y Q, Cai P L, Wen L C, Fu D C, Zhang B, et al. 2023b. Towards knowledge-driven autonomous driving [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2312.04316.pdf>
- Li X, Li G and Bian Z J. 2024. Research on autonomous vehicle path planning algorithm based on improved RRT* algorithm and artificial potential field method. *Sensors*, 24(12): #3899 [DOI: 10.3390/s24123899]
- Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Tassa Y, et al. 2016. Continuous control with deep reinforcement learning//Proceedings of the 4th International Conference on Learning Representations. San Juan, Puerto Rico: [s.n.]
- Lin S W, Liu A, Wang J G and Kong X Y. 2023. An intelligence-based hybrid PSO-SA for mobile robot path planning in warehouse. *Journal of Computational Science*, 67: #101938 [DOI: 10.1016/j.jocs.2022.101938]
- Lin Y C, Chou C C and Lian F L. 2009. Indoor robot navigation based on DWA*: velocity space approach with region analysis//2009 ICCAS-SICE. Fukuoka, Japan: IEEE: 700-705
- Liu J M, Chen H, An P J, Liu Z Y, Zhang R R, Gu C Y, et al. 2025a. HybridVLA: collaborative diffusion and autoregression in a unified vision-language-action model [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2503.10631.pdf>
- Liu J Y, Fu M L, Zhang W, Chen B, Prapakovich R and Sychou U. 2023. CDT-Dijkstra: fast planning of globally optimal paths for all points in 2D continuous space//Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Detroit, USA: IEEE: 2224-2231 [DOI: 10.1109/IROS55552.2023.10341744]
- Liu L S, Lin J F, Yao J X, He D W, Zheng J S, Huang J, et al. 2021. Path planning for smart car based on Dijkstra algorithm and dynamic window approach. *Wireless Communications and Mobile Computing*, 2021(1): #8881684 [DOI: 10.1155/2021/8881684]
- Liu Q, Zhai J W, Zhang Z Z, Zhong S, Zhou Q, Zhang P, et al. 2018. A survey on deep reinforcement learning. *Chinese Journal of Computers*, 41(1): 1-27 (刘全, 翟建伟, 章宗长, 钟珊, 周倩, 章鹏, 等. 2018. 深度强化学习综述. *计算机学报*, 41(1): 1-27) [DOI: 10.11897/SP.J.1016.2018.00001]
- Liu T B and Zhang J D. 2022. An improved path planning algorithm

- based on fuel consumption. *The Journal of Supercomputing*, 78(11): 12973-13003 [DOI: 10.1007/s11227-022-04395-6]
- Liu W H, Wu H Y, Xiong W T, Li X P, Cai B F, Yu S D, et al. 2024. Path planning for mobile robots based on the improved DAPF-QRRT⁺ strategy. *Electronics*, 13(21): #4233 [DOI: 10.3390/electronics13214233]
- Liu Z Y, Sun Z Y, Zang Y H, Dong X Y, Cao Y H, Duan H D, et al. 2025b. Visual-RFT: visual reinforcement fine-tuning [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2503.01785.pdf>
- Lu J S, Batra D, Parikh D and Lee S. 2019. ViLBERT: pretraining task-agnostic visiolinguistic representations for vision-and-language tasks//*Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: #2 [DOI: 10.5555/3454287.3454289]
- Lu Y R, Fu J, Tucker G, Pan X L, Bronstein E, Roelofs R, et al. 2023. Imitation is not enough: robustifying imitation with reinforcement learning for challenging driving scenarios//*Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Detroit, USA: IEEE: 7553-7560 [DOI: 10.1109/IROS55552.2023.10342038]
- Mao Z Q, Liu Y and Qu X B. 2024. Integrating big data analytics in autonomous driving: an unsupervised hierarchical reinforcement learning approach. *Transportation Research Part C: Emerging Technologies*, 162: #104606 [DOI: 10.1016/j.trc.2024.104606]
- Maw A A, Tyan M and Lee J W. 2020. iADA*: improved anytime path planning and replanning algorithm for autonomous vehicle. *Journal of Intelligent and Robotic Systems*, 100(3): 1005-1013 [DOI: 10.1007/s10846-020-01240-x]
- Meng Y, Bing Z S, Yao X T, Chen K J, Huang K, Gao Y, et al. 2025. Preserving and combining knowledge in robotic lifelong reinforcement learning. *Nature Machine Intelligence*, 7(2): 256-269 [DOI: 10.1038/s42256-025-00983-2]
- Mnih V, Badia A P, Mirza M, Graves A, Lillicrap T, Harley T, et al. 2016. Asynchronous methods for deep reinforcement learning//*Proceedings of the 33rd International Conference on Machine Learning*. New York, USA: PMLR: 1928-1937 [DOI: 10.48550/arXiv.1602.01783]
- Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, et al. 2013. Playing Atari with deep reinforcement learning [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/1312.5602.pdf>
- Oh J, Guo Y J, Singh S and Lee H. 2018. Self-imitation learning//*Proceedings of the 35th International Conference on Machine Learning*. Stockholm, Sweden: PMLR: 3878-3887
- Paliwal P. 2023. Improved artificial potential field method for motion-planning of autonomous vehicles//*Proceedings of 2023 International Conference for Advancement in Technology (ICONAT)*. Goa, India: IEEE: 1-6 [DOI: 10.1109/ICONAT57137.2023.10080077]
- Pan Z H, Xia Y Q, Bao H, Wang R Z and Yu T T. 2024. Research progress in decision-making for unmanned intelligent swarm system and control. *Journal of image and Graphics*, 29(11): 3195-3215 (潘振华, 夏元清, 鲍泓, 王睿哲, 于婷婷. 2024. 无人智能集群系统决策与控制研究进展. *中国图象图形学报*, 29(11): 3195-3215) [DOI: 10.11834/jig.230766]
- Pan Z P, Cheng W Z, Zou G, Liu G L, Liu B and Ma Y H. 2024. Multi-agent real-time route planning based on fuzzy adaptive DWA//*Proceedings of the 43rd Chinese Control Conference (CCC)*. Kunming, China: IEEE: 1873-1878 [DOI: 10.23919/CCC63176.2024.10662575]
- Peng J and Williams R J. 1993. Efficient learning and planning within the Dyna framework//*Proceedings of 1993 IEEE International Conference on Neural Networks*. San Francisco, USA: IEEE: 168-174 [DOI: 10.1109/ICNN.1993.298551]
- Peng Z L, Wang W H, Dong L, Hao Y R, Huang S H, Ma S M, et al. 2023. Kosmos-2: grounding multimodal large language models to the world [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2306.14824.pdf>
- Phan-Minh T, Howington F, Chu T S, Lee S U, Tomov M S, Li N X, et al. 2022. Driving in real life with inverse reinforcement learning [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2206.03004.pdf>
- Prakash S, Sami A P and Sharma B. 2024. Enhancing path planning with obstacles via a combined Dijkstra-LbCS methodology//*Bansal J C, Borah S, Hussain S and Salhi S, eds. Computing and Machine Learning*. Singapore: Springer: 47-59 [DOI: 10.1007/978-981-97-6588-1_4]
- Qu T, Duan F Q, Zhang J G, Zhao B and Huang W Z. 2024. Data-efficient model-based reinforcement learning with trajectory discrimination. *Complex and Intelligent Systems*, 10(2): 1927-1936 [DOI: 10.1007/s40747-023-01247-5]
- Reda M, Onsy A, Haikal A Y and Ghanbari A. 2024. Path planning algorithms in the autonomous driving system: a comprehensive review. *Robotics and Autonomous Systems*, 174: #104630 [DOI: 10.1016/j.robot.2024.104630]
- Robine J, Höftmann M, Uelwer T and Harmeling S. 2023. Transformer-based world models are happy with 100k interactions//*Proceedings of the 11th International Conference on Learning Representations*. Kigali, Rwanda: OpenReview.net
- Ross S, Gordon G and Bagnell D. 2011. A reduction of imitation learning and structured prediction to no-regret online learning//*Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*. Fort Lauderdale, USA: JMLR: 627-635
- Sang H Q, You Y S, Sun X J, Zhou Y and Liu F. 2021. The hybrid path planning algorithm based on improved A* and artificial potential field for unmanned surface vehicle formations. *Ocean Engineering*, 223: #108709 [DOI: 10.1016/j.oceaneng.2021.108709]
- Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, Schmitt S, et al. 2020. Mastering Atari, Go, chess and shogi by planning

- with a learned model. *Nature*, 588 (7839): 604-609 [DOI: 10.1038/s41586-020-03051-4]
- Schulman J, Levine S, Moritz P, Jordan M and Abbeel P. 2015. Trust region policy optimization//Proceedings of the 32nd International Conference on International Conference on Machine Learning — Volume 37. Lille, France: JMLR. org: 1889-1897 [DOI: 10.5555/3045118.3045319]
- Schulman J, Wolski F, Dhariwal P, Radford A and Klimov O. 2017. Proximal policy optimization algorithms [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/1707.06347.pdf>
- Shi T Y, Chen D, Chen K A and Li Z J. 2021. Offline reinforcement learning for autonomous driving with safety and exploration enhancement [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2110.07067.pdf>
- Tampuu A, Matiisen T, Semikin M, Fishman D and Muhammad N. 2022. A survey of end-to-end driving: architectures and training methods. *IEEE Transactions on Neural Networks and Learning Systems*, 33(4): 1364-1384 [DOI: 10.1109/TNNLS.2020.3043505]
- Tan H and Bansal M. 2019. LXMERT: learning cross-modality encoder representations from transformers//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics: 5100-5111 [DOI: 10.18653/v1/D19-1514]
- Tao F Z, Ding Z W, Fu Z G, Li M Y and Ji B F. 2024. Efficient path planning for autonomous vehicles based on RRT* with variable probability strategy and artificial potential field approach. *Scientific Reports*, 14(1): #24698 [DOI: 10.1038/s41598-024-76299-9]
- Tao Y, Gao H, Ren F, Chen C Y, Wang T M, Xiong H G, et al. 2021. A mobile service robot global path planning method based on ant colony optimization and fuzzy control. *Applied Sciences*, 11(8): #3605 [DOI: 10.3390/app11083605]
- Tao Z, Gao Y F, Zheng T J, Li J J and Li B Z. 2020. Research on path planning in honeycomb grid map based on A* algorithm. *Journal of North University of China (Natural Science Edition)*, 41(4): 310-317 (陶哲, 高跃飞, 郑天江, 李俊杰, 李保在. 2020. 基于A*算法在蜂巢栅格地图中的路径规划研究. 中北大学学报(自然科学版), 41(4): 310-317) [DOI: 10.3969/j.issn.1673-3193.2020.04.005]
- Tian Z W, Qian R L and Ye J K. 2025. Research on the application of dual-objective optimization model based on simulated annealing algorithm in waste collection path planning. *Computer Knowledge and Technology*, 21(2): 23-25, 33 (田志伟, 钱儒亮, 叶俊康. 2025. 基于模拟退火算法的双目标优化模型在垃圾收集路径规划中的应用研究. 电脑知识与技术, 21(2): 23-25, 33) [DOI: 10.14004/j.cnki.ckt.2025.0001]
- Torabi F, Warnell G and Stone P. 2018. Generative adversarial imitation from observation (in Chinese with English abstract) [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/1807.06158.pdf>
- Tu K B, Yang S S, Zhang H and Wang Z P. 2019. Hybrid A* based motion planning for autonomous vehicles in unstructured environment//2019 IEEE International Symposium on Circuits and Systems (ISCAS). Sapporo, Japan: IEEE: 1-4 [DOI: 10.1109/ISCAS.2019.8702779]
- Wang C Y, Wang L H, Lu Z M, Chu X H, Shi Z R, Deng J Y, et al. 2023. SRL-TR²: a safe reinforcement learning based TRajjectory TRacker framework. *IEEE Transactions on Intelligent Transportation Systems*, 24 (6): 5765-5780 [DOI: 10.1109/TITS.2023.3250720]
- Wang H W, Lou S J, Jing J, Wang Y S, Liu W and Liu T M. 2022. The EBS-A* algorithm: an improved A* algorithm for path planning. *PLoS One*, 17(2): #0263841 [DOI: 10.1371/journal.pone.0263841]
- Wang J M, Zhang X Y, Xing Z B, Gu S E, Guo X Y, Hu Y, et al. 2024b. HE-Drive: human-like end-to-end driving with vision language models [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2410.05051.pdf>
- Wang P, Bai S, Tan S N, Wang S J, Fan Z H, Bai J Z, et al. 2024c. Qwen2-VL: enhancing vision-language model's perception of the world at any resolution [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/2409.12191.pdf>
- Wang S, Jia D Y and Weng X S. 2018. Deep reinforcement learning for autonomous driving [EB/OL]. [2025-01-03]. <https://arxiv.org/pdf/1811.11329.pdf>
- Wang W J, Li J Y, Bai Z N, Wei Z H and Peng J X. 2024d. Toward optimization of AGV path planning: an RRT* -ACO algorithm. *IEEE Access*, 12: 18387-18399 [DOI: 10.1109/ACCESS.2024.3359748]
- Wang X, Dai B, Miao Q G, Nie Y M and Shang E K. 2024a. An efficient deep reinforcement learning-based car-following method via rule-constrained data augmentation. *IEEE Transactions on Intelligent Vehicles*, 2024, 9 (10): 6381-6393 [DOI: 10.1109/TIV.2024.3367720]
- Wang Y N, Zhang F D, Xiao D M, and Sun Z Q. 2025. Multi-agent pursuit decision-making method based on hybrid imitation learning. *Computer Science*, 52(1): 323-330 (王焱宁, 张锋镝, 肖登敏, 孙中奇. 2025. 基于混合模仿学习的多智能体追捕决策方法. 计算机科学, 52(1): 323-330) [DOI: 10.11896/jsjx.240800072]
- Wang Y R, Jing X C, Tian T, Sun Y Q and Cong S J. 2019. Multi-agent path planning based on reinforcement learning. *Computer Applications and Software*, 36(8): 165-171 (王毅然, 经小川, 田涛, 孙运乾, 从帅军. 2019. 基于强化学习的多Agent路径规划方法研究. 计算机应用与软件, 36(8): 165-171) [DOI: 10.3969/j.issn.1000-386x.2019.08.029]
- Wu J D, Huang C, Huang H L, Lyu C, Wang Y T and Wang F Y. 2024. Recent advances in reinforcement learning-based autonomous driving behavior planning: a survey. *Transportation Research*

- Part C: Emerging Technologies, 164: #104654 [DOI: 10.1016/j.tre.2024.104654]
- Xiao Y, Codevilla F, Gurram A, Urfalioglu O and López A M. 2022. Multimodal end-to-end autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 23 (1): 537-547 [DOI: 10.1109/TITS.2020.3013234]
- Xu D H, Yang J, Zhou X Q and Xu H T. 2024. Hybrid path planning method for USV using bidirectional A* and improved DWA considering the manoeuvrability and COLREGs. *Ocean Engineering*, 298: #117210 [DOI: 10.1016/j.oceaneng.2024.117210]
- Xu R Y, Zhao C H, Li J X, Hu J W and Hou X L. 2023. A hybrid improved-whale-optimization — simulated-annealing algorithm for trajectory planning of quadruped robots. *Electronics*, 12 (7): #1564 [DOI: 10.3390/electronics12071564]
- Yan X Z, Ding R C, Luo Q H, Ju C Y and Wu D. 2022. A dynamic path planning algorithm based on the improved DWA algorithm//2022 Global Reliability and Prognostics and Health Management (PHM-Yantai). Yantai, China: IEEE: 1-7 [DOI: 10.1109/PHM-Yantai55411.2022.9942106]
- Yang M L and Li N. 2022. Study on mobile robot path planning based on improved A* algorithm. *Mechanical Science and Technology for Aerospace Engineering*, 41 (5): 795-800 (杨明亮, 李宁. 2022. 改进A*算法的移动机器人路径规划. *机械科学与技术*, 41 (5): 795-800) [DOI: 10.13433/j.cnki.1003-8728.20220005]
- Yang Y L, Luo X Y, Li W, Liu C D, Ye Q and Liang P. 2024. AAPF*: a safer autonomous vehicle path planning algorithm based on the improved A* algorithm and APF algorithm. *Cluster Computing*, 27 (8): 11393-11406 [DOI: 10.1007/s10586-024-04287-9]
- Yin X, Dong W T, Wang X M, Yu Y X and Yao D J. 2024. Route planning of mobile robot based on improved RRT star and TEB algorithm. *Scientific Reports*, 14 (1): #8942 [DOI: 10.1038/s41598-024-59413-9]
- Yonetani R, Tanai T, Barekatin M, Nishimura M and Kanezaki A. 2021. Path planning using neural A* search//Proceedings of the 38th International Conference on Machine Learning. [s. l.]: PMLR: 12029-12039
- Yurtsever E, Lambert J, Carballo A and Takeda K. 2020. A survey of autonomous driving: common practices and emerging technologies. *IEEE Access*, 8: 58443-58469 [DOI: 10.1109/ACCESS.2020.2983149]
- Ze Y J, Zhang G, Zhang K N, Hu C Y, Wang M H and Xu H Z. 2024. 3D diffusion policy: generalizable visuomotor policy learning via simple 3D representations//2nd Workshop on Dexterous Manipulation: Design, Perception and Control (RSS). Delft, Netherlands: [s.n.]
- Zhai S Z and Pei Y L. 2023. The dynamic path planning of autonomous vehicles on icy and snowy roads based on an improved artificial potential field. *Sustainability*, 15 (21): #15377 [DOI: 10.3390/su152115377]
- Zhang A G, Zhang Y Y and Fan W Q. 2020. Research on obstacle avoidance path planning of autonomous vehicles based on gravitational search algorithm. *Automation and Instrumentation*, (10): 35-38 (张爱国, 张迎娅, 范文强. 2020. 基于引力搜索算法的自动驾驶车辆避障路径规划研究. *自动化与仪器仪表*, (10): 35-38) [DOI: 10.14016/j.cnki.1001-9227.2020.10.035]
- Zhang J S and Chen Z H. 2023. The improved artificial potential field algorithm based on pre-added virtual force. *Journal of Ordnance Equipment Engineering*, 44 (3): 219-225 (张佳尚, 陈志华. 2023. 基于预添加虚拟力的改进人工势场算法. *兵器装备工程学报*, 44 (3): 219-225) [DOI: 10.11809/bqzbgcxb2023.03.031]
- Zhang T L, Xu S T, Gao Y, Wang P, Schonfeld P, Zou Y, et al. 2024. 3D constrained hybrid A*: improved vehicle path planning algorithm for cost-effective road alignment design. *Automation in Construction*, 166: #105645 [DOI: 10.1016/j.autcon.2024.105645]
- Zhao J Y, Zhao W Y, Deng B, Wang Z H, Zhang F, Zheng W X, et al. 2024. Autonomous driving system: a comprehensive survey. *Expert Systems with Applications*, 242: #122836 [DOI: 10.1016/j.eswa.2023.122836]
- Zhao S E, Leng Y, Zhao M J, Wang K, Zeng J and Liu W L. 2025. A novel dynamic lane-changing trajectory planning for autonomous vehicles based on improved APF and RRT algorithm. *International Journal of Automotive Technology*, 26 (2): 451-461 [DOI: 10.1007/s12239-024-00153-0]
- Zhao W S, Queralta J P and Westerlund T. 2020. Sim-to-real transfer in deep reinforcement learning for robotics: a survey//2020 IEEE Symposium Series on Computational Intelligence (SSCI). Canberra, Australia: IEEE: 737-744 [DOI: 10.1109/SSCI47803.2020.9308468]
- Zhao X, Wang Z, Huang C K and Zhao Y W. 2018. Mobile robot path planning based on an improved A* algorithm. *ROBOT*, 40 (6): 903-910 (赵晓, 王铮, 黄程侃, 赵燕伟. 2018. 基于改进A*算法的移动机器人路径规划. *机器人*, 40 (6): 903-910) [DOI: 10.13973/j.cnki.robot.170591]
- Zheng Y N, Liang R M, Zheng K X, Zheng J L, Mao L Y, Li J X, et al. 2025. Diffusion-based planning for autonomous driving with flexible guidance//Proceedings of the 13th International Conference on Learning Representations. Singapore, Singapore: [s.n.]
- Zhou B, Guo Y, Li N and Zhong X J. 2021. Path planning of UAV using guided enhancement Q-learning algorithm. *Acta Aeronautica et Astronautica Sinica*, 42 (9): #325109 (周彬, 郭艳, 李宁, 钟锡健. 2021. 基于导向强化Q学习的无人机路径规划. *航空学报*, 42 (9): #325109) [DOI: 10.7527/S1000-6893.2021.25109]
- Zhu D D and Sun J Q. 2021. A new algorithm based on Dijkstra for vehicle path planning considering intersection attribute. *IEEE Access*, 9: 19761-19775 [DOI: 10.1109/ACCESS.2021.3053169]
- Zuo Z Q, Yang X, Zhang Z C and Wang Y J. 2019. Lane-associated

MPC path planning for autonomous vehicles//Proceedings of 2019 Chinese Control Conference (CCC). Guangzhou, China: IEEE: 6627-6632 [DOI: 10.23919/ChiCC.2019.8866609]

作者简介

李程翔,男,硕士研究生,主要研究方向为自动驾驶。
E-mail: chengxiangli@sanyau.edu.cn
郭大波,通信作者,男,教授,主要研究方向为强化学习。
E-mail: dabo_guo@sxu.edu.cn

胡海霄,男,硕士研究生,主要研究方向为自动驾驶。
E-mail: haixiaohu@sanyau.edu.cn
陈荣敏,男,硕士研究生,主要研究方向为多智能体通信。
E-mail: rongminchen@sanyau.edu.cn
吴宏坤,男,硕士研究生,主要研究方向为自动驾驶。
E-mail: hongkunwu@sanyau.edu.cn
袁畅,女,学士,中级记者,主要研究方向为广播电视编导。
E-mail: yy4343@126.com