机器人工程与智能制造南充市重点实验室
石油天然气装备教育部重点实验室
2026 年度联合开放基金申请书

资助类别：☑经费资助项目　☐自筹资金项目

项目名称：面向不确定环境的可靠停车位搜索策略算法研究

申 请 人：陈浩铭

联系电话：**18282217554**

承担单位：

起止时间：

学科门类：

通讯地址：

填报日期：

机器人工程与智能制造南充市重点实验室

石油天然气装备教育部重点实验室

2026 年制

填报说明

1、申报书适用于机器人工程与智能制造南充市重点实验室与石油天然气装备教育部重点实验室2026年度联合开放基金；

2、申报书打印要求用A4纸双面打印。一式四份，立项后统一报送至机器人工程与智能制造南充市重点实验室；

3、有意申请者请于2026年3月1日前将申报材料电子档发送到指定邮箱，具体联系方式如下：

联系人：林老师

联系电话：0817-2641202

邮箱：wlin@swpu.edu.cn

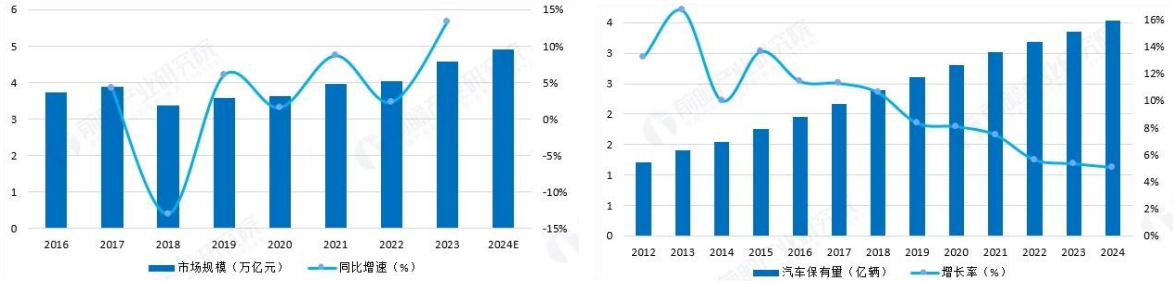
地址：四川省南充市顺庆区油院路二段1号西南石油大学南充校区完井楼208

4、未尽事宜，可另附材料说明。

（一）立项依据：

1.1 研究意义

智能停车位搜索系统能够突破城市复杂交通环境下对停车效率与驾驶体验的双重制约，在大型商业综合体、交通枢纽、高密度住宅区等动态停车场景中展现出巨大潜力。当前，我国面临城市汽车保有量持续攀升、停车资源供需矛盾日益尖锐、因寻找车位引发的交通拥堵与碳排放增加等严峻挑战（如**图 1** 所示），国家已出台《交通强国建设纲要》、《关于推动城市停车设施发展的意见》等政策法规，聚焦智慧交通基础设施建设与出行服务品质升级。因此，动态与不确定环境下的可靠停车位搜索策略研究不仅对缓解城市交通压力、提升居民出行效率具有重要价值，同时也是契合国家智慧城市战略、推动自动驾驶技术落地与应用的重要支撑。



(a) 中国汽车零部件行业市场规模　　(b) 中国汽车保有量及增长

图 1 中国汽车行业以及保有量现状

本项目所研究的可靠停车位搜索任务具有如下特点及挑战：

(1) 动态性：停车环境的时变特性。在实际停车过程中，车位的占用状态、道路的通行状况以及其他车辆的行驶轨迹均处于实时变化之中。这种高度的动态性削弱了传统基于静态地图或离线规划算法的有效性，导致预先规划的路径极易失效，迫切需要具备在线学习与实时适应能力的搜索策略。

(2) 不确定性：搜索过程中的随机扰动。受传感器噪声、通信延迟以及行人或其他车辆的不可预测行为影响，停车位搜索过程充满了随机扰动。经典算法往往假设环境信息完全已知或确定，难以应对这种内生的不确定性，容易导致车辆在搜索过程中陷入徘徊或决策震荡。

(3) 可靠性：对搜索时间波动性的严格约束。

传统的路径规划算法通常以“平均搜索时间最小化”为单一目标，忽略了搜索时间的波动范围。然而，对于赶时间的驾驶员或自动驾驶系统而言，搜索时间的稳定性往往比单纯的平均速度更为关键。一个忽快忽慢、极不稳定的搜索策略会严重降低用户信任

度，需优先保障在可控时间波动范围内的任务成功率。

因此，本项目拟针对动态停车场景下可靠搜索任务中的动态性、不确定性与可靠性等挑战，深入剖析强化学习在非线性目标下的优化机理，重点突破尚未解决的关键基础性问题。具体包括：建立停车位搜索任务的马尔可夫决策过程（MDP）模型，揭示搜索时间均值与方差的内在耦合关系；提出面向非线性目标函数的均值-标准差近端策略优化算法；设计基于在线学习的方差约束机制；最终形成一套“稳中求快”的可靠停车位搜索策略，并通过仿真环境与真车实验验证其有效性。本项目将为不确定环境下的自动驾驶决策规划提出有针对性的研究思路，提升算法在复杂城市交通场景中的适应性与鲁棒性，助力智慧停车管理能力提升，并为自动驾驶技术的创新与应用提供重要支撑。

1.2 国内外研究现状及发展动态分析

智能停车位搜索作为智能交通系统（ITS）和自动驾驶技术的一个重要研究方向，在缓解城市交通拥堵、降低碳排放、提升出行效率以及优化自动驾驶“最后一公里”体验等领域具有广泛应用价值。近年来，国内外学者在该领域取得了一系列重要进展，研究热点集中在停车位检测与状态预测、路径规划与导航策略、以及基于强化学习的决策优化等方面。本项目研究动态不确定场景下的可靠停车位搜索策略，主要内容包括：构建停车位搜索的马尔可夫决策过程模型；提出面向非线性目标的均值-标准差策略优化算法；设计在线学习与方差约束机制。因此，本节将围绕 (1) 停车位检测与状态预测技术；(2) 传统路径规划与启发式搜索策略；(3) 基于强化学习的智能决策与可靠性优化，梳理国内外研究进展，并分析相关技术发展动态与演变趋势。

（1）停车位检测与状态预测技术研究现状

停车位状态的精准感知与预测是实现高效搜索的基础，决定着搜索算法的输入质量。现有研究主要从基于传感器检测、计算机视觉检测和数据驱动预测三个角度切入。

在传感器检测方面，早期的研究主要依赖地磁、超声波等固定传感器网络，通过物联网技术实时上传车位状态。例如，欧美等国在早期智慧城市建设中广泛部署了基于 ZigBee 或 LoRa 的车位检测网络[1,2]，实现了区域级的车位引导。然而，这种依赖基础设施的方法部署成本高、维护困难，且难以覆盖所有路侧停车位。

计算机视觉技术的发展推动了基于摄像头的车位检测研究。基于深度学习的目标检测算法[3]被广泛应用于监控视频或车载摄像头数据中，以识别空闲车位。例如，有研究利用车载鱼眼相机结合 SLAM 技术[4]，实现了停车场内的实时语义地图构建与

车位识别。这类方法虽然降低了对基础设施的依赖，但在光照变化剧烈、遮挡严重等复杂环境下，检测精度往往大幅下降。

数据驱动的状态预测技术旨在解决实时信息的滞后性问题。研究者利用 LSTM、Transformer[5,6] 等时间序列模型，基于历史停车数据预测未来时段的车位占用率。虽然预测模型在宏观调度上表现良好，但难以精确捕捉微观层面（如某个具体车位在几秒后的状态）的随机变化。此外，现有预测方法大多假设环境具有平稳性，难以应对突发事件或极端天气导致的非平稳动态变化。

（2）传统路径规划与启发式搜索策略研究现状

在获取车位信息后，如何规划最优搜索路径是核心问题。现有方法主要包括：基于图论的最短路径算法、概率图模型以及启发式搜索策略。

基于图论的算法[7]以距离或预计行驶时间为权重，寻找全局最优路径。这类算法在静态环境中表现优异，但在车位状态实时变化的动态场景中，预规划路径极易失效，导致车辆频繁重规划，计算开销大且效率低下。

概率图模型引入了不确定性描述。例如，有些研究将停车搜索问题建模为时间依赖的旅行商问题[8]或概率路图[9]，结合贝叶斯推理更新车位空闲概率。这类方法在一定程度上考虑了环境的随机性，但计算复杂度随路网规模呈指数级增长，难以满足车载计算单元的实时性要求。

启发式搜索策略通过设计经验规则或元启发式算法[10]来指导车辆搜索。这类方法计算量小，反应速度快，但缺乏理论保证，容易陷入局部最优，且策略通常是固定的，无法根据环境反馈进行在线调整。更重要的是，上述传统方法大多以“期望时间最小化”为单一目标，忽略了搜索时间的波动性，无法满足用户对“可靠性”和“稳定性”的需求。

（3）基于强化学习的智能决策与可靠性优化研究现状

强化学习（RL）因其强大的在线学习和决策能力，近年来成为解决动态停车搜索问题的新兴热点。研究者通常将停车搜索建模为马尔可夫决策过程（MDP），利用 DRL 算法进行优化。

在策略生成方面，基于值函数的方法[11]和基于策略梯度的方法[12]已被应用于自动泊车和路径规划中。例如，有学者利用 DDPG 算法实现了复杂停车场环境下的端到端自动驾驶泊车，通过与仿真环境交互学习避障和搜索策略[13]。然而，现有的 DRL 方法大多采用“风险中性”的优化目标，即最大化累积奖励的期望值（Expectation），

这在本质上仍然是追求平均性能，而忽视了性能的分布特征。

针对可靠性与风险规避问题，风险敏感强化学习（Risk-Sensitive RL）开始受到关注。部分研究尝试引入风险度量[14]作为约束条件，或者在目标函数中加入方差惩罚项。然而，在停车搜索这一特定领域，关于“均值-标准差”复合目标优化的研究尚处于起步阶段。主要挑战在于：标准差具有非线性与不可加性，导致传统的贝尔曼方程（Bellman Equation）失效，难以直接应用现有的时序差分（TD）学习方法进行价值估计。虽然已有少量理论研究探讨了方差的递归估计方法，但将其成功应用于大规模、高动态的停车位搜索场景，并实现从仿真到真车的部署，目前仍鲜有报道。

1.3 研究现状分析与总结

综上所述，国内外学者在智能停车领域已取得显著成果，但在应对动态不确定环境下的可靠搜索任务时，仍存在以下不足：

(1) 针对搜索过程波动性的优化不足。现有算法多以最小化平均搜索时间为目标，忽视了搜索时间的方差。对于风险厌恶型用户，搜索过程的稳定性往往比单纯的“快”更重要。目前缺乏能同时优化均值与标准差的成熟搜索策略。

(2) 应对非线性目标函数的理论与算法框架尚不完善。引入标准差作为优化目标后，破坏了传统强化学习中累积回报的线性可加性，导致经典的策略梯度算法难以直接适用。亟需建立面向“均值-标准差”非线性目标的广义强化学习框架。

(3) 仿真到真车的迁移验证存在鸿沟。现有基于强化学习的停车研究大多停留在简化的网格仿真阶段，缺乏在真实物理环境（考虑传感器噪声、控制延迟、动态障碍物）中的验证与部署，导致算法的工程实用性存疑。

因此，本项目拟在 PPO 算法的基础上，提出均值-标准差近端策略优化算法。通过推导方差贝尔曼方程，建立均值与方差的联合优化机制，并结合真车实验平台进行验证，旨在解决动态环境下停车位搜索的可靠性难题，填补该领域的理论与应用空白。

主要参考文献

[1]Ding, Han et al. “ Design On Parking Space Management System Based On ZigBee. ” *2023 2nd International Conference on Artificial Intelligence and Computer Information Technology (AICIT)* (2023): 1-4.

[2]Yazıcı, Alper Berkin et al. “LoRa Technology Overview and Smart Parking System Design.” *2024 32nd Signal Processing and Communications Applications Conference (SIU)* (2024): 1-4.

[3]Ahad, Abdul, and Farhan Ahmad Kidwai. "YOLO based approach for real-time parking detection and dynamic allocation: Integrating behavioral data for urban congested cities." *Innovative Infrastructure Solutions* 10.6 (2025): 252.

[4]Li, Ye, et al. "AVM-SLAM: Semantic visual SLAM with multi-sensor fusion in a bird’ s eye view for automated valet parking." *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024.

[5]Jin, Bowen, Yu Zhao, and Jing Ni. "Sustainable transport in a smart city: Prediction of short-term parking space through improvement of LSTM algorithm." *Applied Sciences 12.21 (2022)*: 11046.

[6]Zheng, Han, et al. "MultiPark: Multimodal Parking Transformer with Next-Segment Prediction." *arXiv preprint arXiv*:2508.11537 (2025).

[7]Wang, Wei, You Xu, and Zhi Meng. "Design of parking guidance path in parking lot based on Dijkstra algorithm." *International Symposium on Robotics, Artificial Intelligence, and Information Engineering (RAIIE 2022)*. Vol. 12454. SPIE, 2022.

[8]Hong, Lingrui, Lin Zhou, and Roberto Baldacci. "The traveling salesman problem with drone based on vehicle mobile parking for customer self-pickup." *Transportation Research Part E: Logistics and Transportation Review* 203 (2025): 104347.

[9]Li, Yicheng, et al. "Intelligent vehicle localization and navigation based on intersection fingerprint roadmap (IRM) in underground parking lots." *Measurement Science and Technology* 35.3 (2024): 036301.

[10]Dou, Nan et al. “Parking Space Matching and Path Planning Based on Wolf Feeding Decision Algorithm in Large Underground Garage.” 6GN (2023).

[11]Xie, Jun, Zhaocheng He, and Yiting Zhu. "A DRL based cooperative approach for parking space allocation in an automated valet parking system." *Applied Intelligence* 53.5 (2023): 5368-5387.

[12]Albilani, Mohamad, and Amel Bouzeghoub. "Dynamic adjustment of reward function for proximal policy optimization with imitation learning: Application to automated parking systems." 2022 *IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022.

[13]Li, Jia, et al. "An integrated model for coordinating adaptive platoons and parking decision-making based on deep reinforcement learning." *Computers & Industrial Engineering* 203 (2025): 110962.

[14]Qureshi, Kalim. "A risk-sensitive and semantic-aware task execution framework for mobility-coupled vehicular networks with CVaR and DRL." *Computing* 108.1 (2026): 2.

（二）研究内容：

2.1 研究内容

本项目旨在解决动态不确定环境下停车位搜索效率低、搜索时间波动大以及决策可靠性不足的问题。项目将围绕“环境建模—算法设计—系统验证”这一主线展开，重点研究基于均值-标准差的强化学习决策机制。各研究内容之间的内在关系如**图 2** 所示。

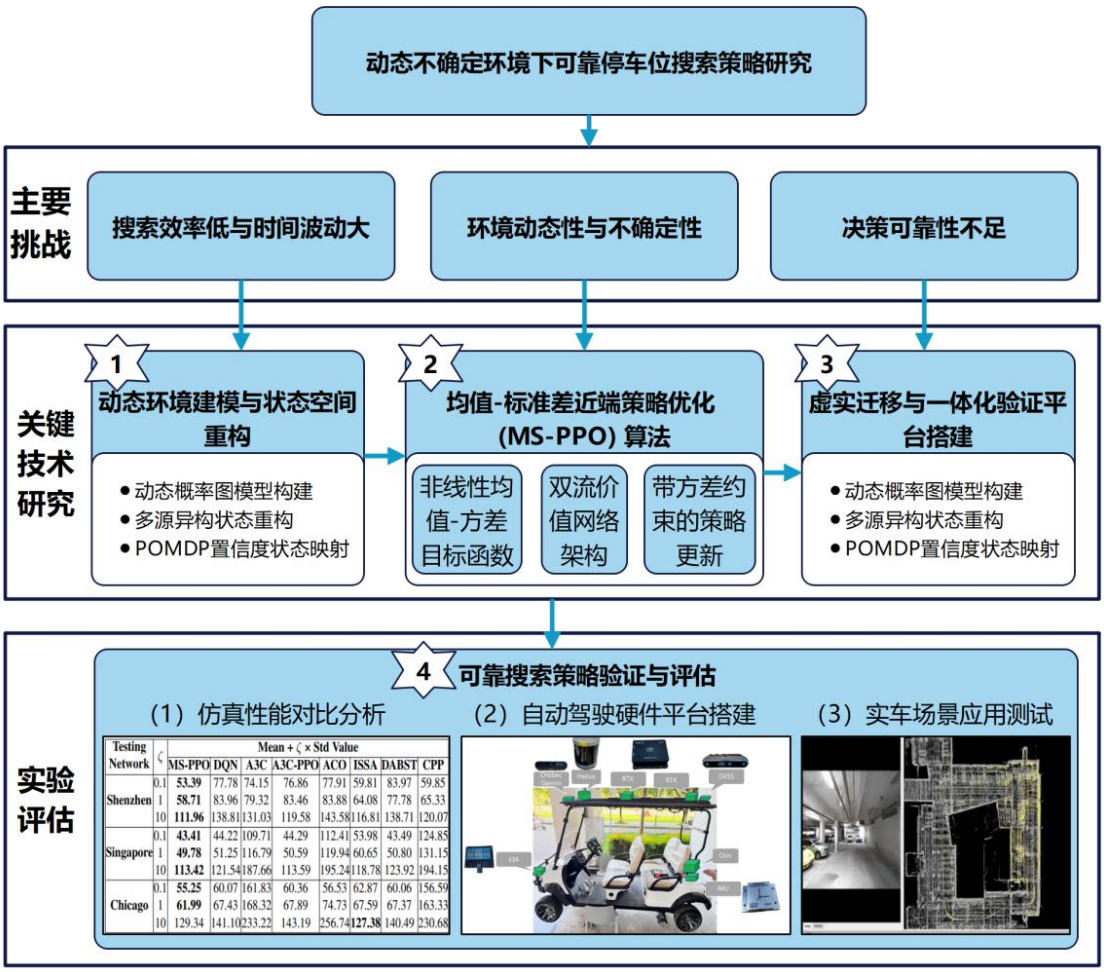


图 2 本项目研究内容框架图

(1) 研究内容一：动态不确定环境下的停车位搜索过程建模与状态空间重构

针对现有模型对环境动态性与部分可观测性描述不足的问题，本部分拟构建能够精确表征停车位搜索特性的马尔可夫决策过程（MDP）模型。

①动态概率图模型的构建：研究停车位占用状态的时空演化规律，建立基于泊松过程或历史数据的车位状态转移概率模型。将停车场拓扑结构抽象为概率图，节点属性不仅包含位置信息，还包含该车位“当前被占用”的动态概率估计。

②多源异构状态空间的重构：摒弃仅依赖车辆位置的简单状态定义，构建包含车辆自身状态（位置）、环境感知状态（局部障碍物分布）以及全局信息状态（目标车位剩余数量、预测占用率）的高维混合状态空间。

③部分可观测性（POMDP）处理： 考虑到传感器视距限制和通信延迟，研究将部分可观测马尔可夫决策过程（POMDP）映射为置信度状态（Belief State）的方法，利用贝叶斯滤波器实时更新对全局车位状态的认知，为后续决策提供准确输入。

(2) **研究内容二：面向可靠性目标的均值-标准差近端策略优化（MS-PPO）算法研究**

这是本项目的核心理论研究内容。针对传统强化学习仅优化期望累积回报（风险中性）的局限，研究如何将搜索时间的方差（风险）纳入优化目标，实现“快且稳”的搜索策略。

①非线性目标函数的构建与求解：定义包含累积回报均值与标准差的复合目标函数 ,针对标准差导致的贝尔曼方程失效问题,推导适用于方差估计的方差贝尔曼方程。

双流价值网络架构设计： 设计一种双流（Two-Stream）神经网络架构，一路用于估计传统的状态价值,另一路专门用于估计回报的二阶矩,通过共享底层特征提取网络，实现均值与方差特征的联合学习。

②MS-PPO 算法的策略更新机制： 在近端策略优化（PPO）的框架下，重新推导引入方差约束后的优势函数（Advantage Function）计算方法。研究如何在保证策略更新单调不减的前提下，利用信赖域约束（Trust Region）平衡均值最大化与方差最小化之间的冲突，确保算法收敛。

(3) **研究内容三：算法验证平台搭建与实车部署测试**

为了验证所提理论与算法的有效性及工程实用性，本项目将采用仿真模拟与物理实验相结合的方式进行验证。

①高保真仿真环境构建：基于 SUMO 或 CARLA 仿真平台，二次开发构建包含动态行人、随机车辆干扰及多样化停车场拓扑结构的测试环境。在此环境中进行大规模蒙特卡洛模拟，对比 MS-PPO 与 Dijkstra、A*、标准 PPO 等基线算法在平均搜索时间、搜索方差及成功率上的性能差异。

②Sim-to-Real 迁移技术研究： 针对仿真与现实世界的“域差异（Domain Gap）”，研究域随机化（Domain Randomization）技术，在仿真中引入传感器噪声、控制延迟和物理摩擦系数扰动，提高模型的泛化能力。

③微缩车实物平台部署与测试： 搭建基于 Jetson Orin/Nano 嵌入式计算平台的智能微缩车系统。在实验室环境下构建物理沙盘，进行实车搜索实验。重点评估算法在计算资源受限、感知数据有噪条件下的实时决策能力与鲁棒性，验证“智寻”系统的实际应用潜力。

2.2 研究目标

本项目拟围绕智能停车场景下搜索策略的实时适应性与可靠性不足的问题，研究并突破非线性标准差可加性数学建模、均值-标准差联合优化策略梯度推导、在线学习与方差约束机制等关键技术；将研究成果转化为可在自动驾驶车辆上部署的智能停车算法，并在高保真仿真环境与真实地下停车场物理场景中验证算法的有效性、稳定性与可移植性，为解决城市停车难问题提供理论依据与技术支撑。具体包括：

（1）建立面向“均值-标准差”双目标的强化学习数学模型：突破传统强化学习仅关注累积回报均值的局限，推导方差贝尔曼方程，解决标准差非线性可加的数学难题。建立能够同时精准表征搜索效率与搜索波动性的马尔可夫决策过程模型，为可靠停车搜索任务提供坚实的数学表征基础。

（2）构建均值-标准差近端策略优化（MS-PPO）算法框架：设计包含均值-方差在线估计（MS-TD）、策略梯度计算（MS-PG）及优势函数融合（MS-AC）的模块化算法架构。引入方差约束与裁剪机制，实现对停车搜索策略的在线学习与动态调整，解决传统算法在动态环境中由于高方差导致的训练震荡问题，提升算法的收敛稳定性。

（3）搭建“软硬一体”的自动驾驶停车搜索实验平台：基于激光雷达、深度相机及高性能计算单元，构建具备环境感知、SLAM 建图及路径规划能力的自动驾驶硬件平台。开发从仿真环境到真车的全流程验证系统，实现无模型算法在真实物理环境中的零样本迁移与部署，验证算法在传感器噪声与动态干扰下的鲁棒性。

（4）形成兼顾搜索效率与稳定性的可靠停车位搜索策略：实现“稳中求快”的搜索效果，在保证平均搜索时间优于主流基准算法（如 DQN、A3C）的同时，显著降低搜索过程的时间波动性。预期在动态复杂场景下，搜索稳定性指标降低 15%以上，有效缓解用户停车焦虑。

2.3 关键科学问题

本项目遵循“理论突破与算法创新并重”、“仿真验证与实车部署互补”的研究思路，围绕智能停车搜索中环境动态性、目标非线性与部署迁移性的挑战，聚焦可靠停车位搜索策略研究，拟解决以下三条关键科学问题：

(1) **如何建立面向非线性“均值-标准差”复合目标的强化学习数学模型**

传统强化学习基于贝尔曼方程处理累积回报的期望（均值），具有良好的线性可加性。然而，本项目提出的“可靠停车搜索”将优化目标定义为搜索时间的均值与标准差之和。标准差本质上是非线性的，无法简单地将各步标准差相加，导致传统的价值函数递归定义失效。因此，如何突破标准差非线性可加的数学障碍，推导方差贝尔曼方程，建立能够同时表征搜索效率（均值）与波动性（标准差）的马尔可夫决策过程模型，是本项目首要解决的关键科学问题。本项目拟引入广义优势估计（GAE）的双维度扩展，构建均值与方差的联合递归表达，为后续策略优化提供理论基石。

(2) **如何探究动态停车环境下均值与方差联合优化的策略梯度收敛机理**

在动态变化的停车场景中，环境概率转移矩阵未知且时变。若直接将均值与方差作为联合优化目标，由于方差估计误差会通过平方项放大，极易导致策略更新过程中的高方差与震荡，使得算法难以收敛。因此，如何设计有效的策略梯度算法，在无模型（Model-Free）条件下同时实现均值最大化与方差最小化，并从数学上证明其收敛性，是本项目的第二个关键科学问题。本项目拟提出 MS-PPO 算法框架，通过引入裁剪机制约束策略更新幅度，并利用重要性采样复用历史数据，探究在多目标联合优化下的梯度平滑与收敛机制。

(3) **如何揭示从仿真到真实物理环境的“感知-决策”零样本迁移规律**

现有的智能停车研究多依赖于高精度的仿真环境或离线数据集，而真实地下停车场存在传感器噪声、光照变化、行人动态干扰及车辆控制延迟等复杂因素（Sim-to-Real Gap）。仅在仿真中表现优异的策略往往难以直接适应真实物理环境。因此，如何克服仿真与现实之间的鸿沟，揭示策略模型在不同物理约束下的泛化规律，实现算法的零样本迁移与在线适应，是本项目的第三个关键科学问题。本项目拟搭建“软硬一体”的自动驾驶实验平台，通过真实感知数据与决策回路的闭环验证，探索提升算法在非结构化真实场景中鲁棒性与可移植性的内在规律。

（三）研究基础：

3.1 总体研究思路与技术路线

本项目拟采用“理论推导—算法设计—仿真验证—实车部署”相结合的闭环研究范式，聚焦解决智能停车搜索中忽视时间波动性与缺乏实时适应性的关键问题。总体

方案遵循“问题转化—理论构建—模块设计—功能增强”的四层架构，旨在突破传统强化学习仅优化期望值的局限，建立面向风险规避的无模型强化学习框架(MS-PPO)。

具体而言，本项目首先将停车搜索问题建模为马尔可夫决策过程（MDP），推导方差贝尔曼方程，解决标准差非线性可加的理论难题；其次，设计包含均值-方差时序差分（MS-TD）、策略梯度（MS-PG）及执行者-评判者（MS-AC）的模块化算法架构，并引入广义优势估计（GAE）与裁剪机制以保障训练稳定性；最后，构建“软硬一体”的自动驾驶实验平台，开展从多场景仿真到真车实地测试的全流程验证，形成如**图 3** 所示的研究方案整体框架。

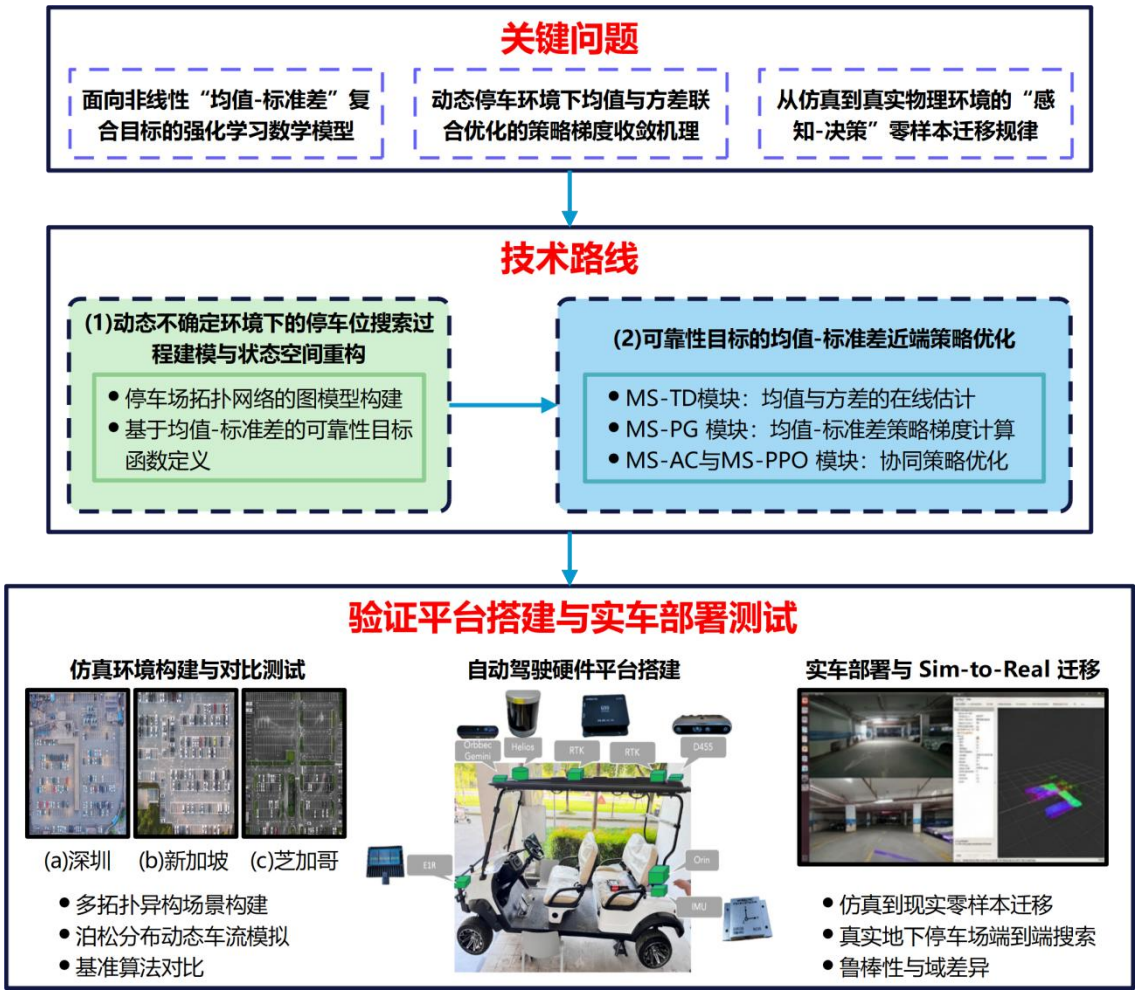


图 3 本项目总体研究方案图

3.2 详细研究方案

(1) **研究内容一：动态不确定环境下的停车位搜索过程建模与状态空间重构**

现有的智能停车方法普遍依赖固定模型，且目标函数仅关注最小化平均搜索时间，忽略了搜索过程中的不确定性（即方差），导致系统可靠性不足。为此，本项目拟构建基于均值-标准差联合优化的数学模型，从理论层面重构价值函数的递归定义。

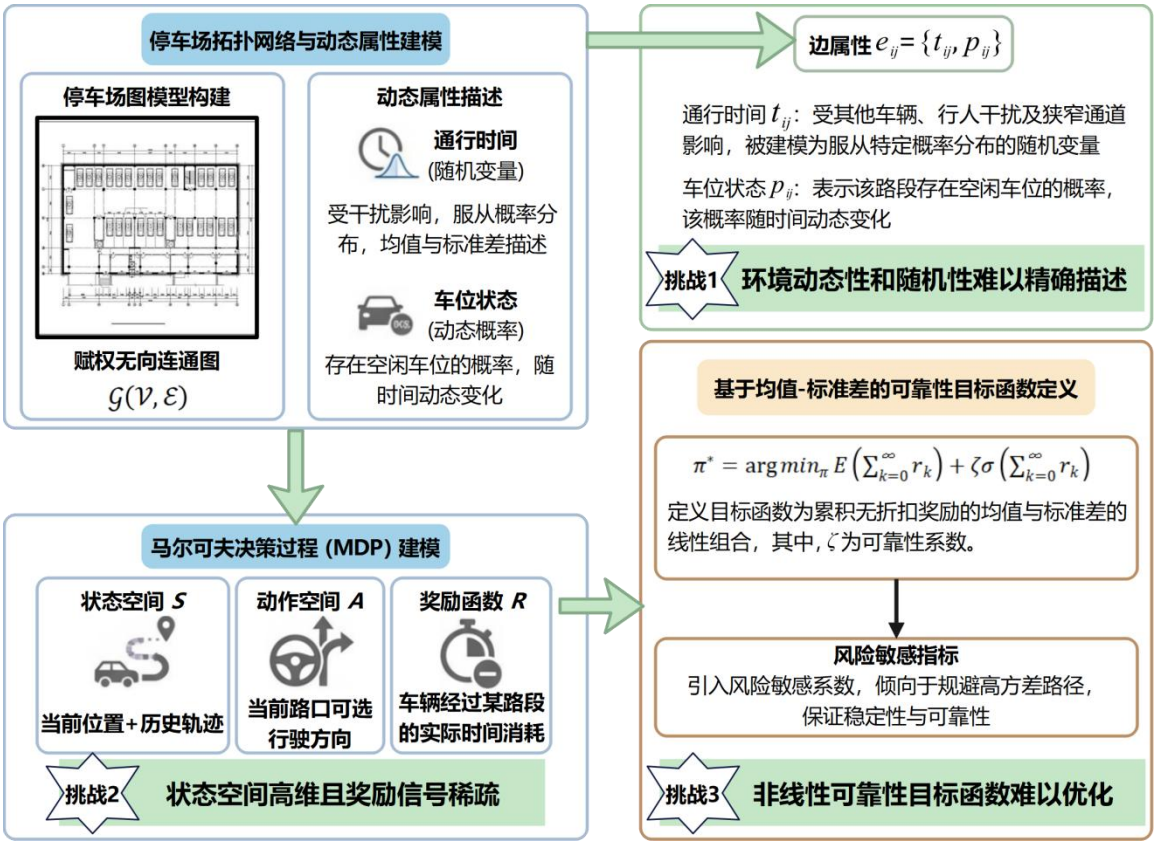


图 4 研究内容一：动态不确定环境下的停车位搜索过程建模的研究方案图

①停车场拓扑网络的图模型构建

针对结构化停车场环境，首先将其抽象为赋权无向连通图 $G(V,\mathcal{E})$，节点集V代表道路交叉口、停车区域入口或关键决策点。边集 $\mathcal{E}$代表连接各节点的行车通道。动态属性描述：不同于静态路径规划，本研究重点考虑环境的随机性。对于任意一条边 $e_{ij} \in \mathcal{E}$，其属性包含两个关键随机变量：通行时间t_{ij}：受其他车辆、行人干扰及狭窄通道影响，通行时间被建模为服从特定概率分布（如高斯分布或对数正态分布）的随机变量，其统计特性由均值μ_{ij}和标准差σ_{ij}描述；车位状态p_{ij}：表示该路段或区域存在空闲车位的概率，该概率随时间动态变化，需通过历史数据统计或实时感知进行更新。

②基于均值-标准差的可靠性目标函数定义

传统的路径规划算法通常仅以最小化期望时间$\mathbb{E}_{tot}$为目标，这往往导致车辆被规划到“平均时间短但波动极大”的高风险路径（例如容易发生拥堵的捷径）。为了满足用户对到达时间可预期的需求，本研究引入风险敏感指标。将停车场景抽象为马尔可夫决策过程，定义状态空间S为车辆当前位置及历史轨迹(避免环路)，动作空间$A(s)$为当前路口的可选行驶方向，奖励函数r为车辆经过某路段的实际时间消耗。为体现风险敏感特性，定义目标函数为累积无折扣奖励的均值与标准差的线性组合：

$$\pi^* = \arg min_\pi E\left(\sum_{k=0}^{\infty} r_k\right) + \zeta\sigma\left(\sum_{k=0}^{\infty} r_k\right) \quad (1)$$

其中，$\zeta > 0$ 为可靠性系数（Reliability Coefficient）。当ζ增大时，策略将更加倾向于规避高方差的路径，从而保证停车过程的稳定性与可靠性。

(2) **研究内容二：面向可靠性目标的均值-标准差近端策略优化算法研究**

这是本课题的核心创新点。传统的强化学习算法仅优化累积回报的期望值，无法直接优化方差。本项目拟设计 MS-PPO 算法，通过四个核心模块的协同工作，实现对停车策略的在线学习与鲁棒优化，技术路线如**图 5** 所示。

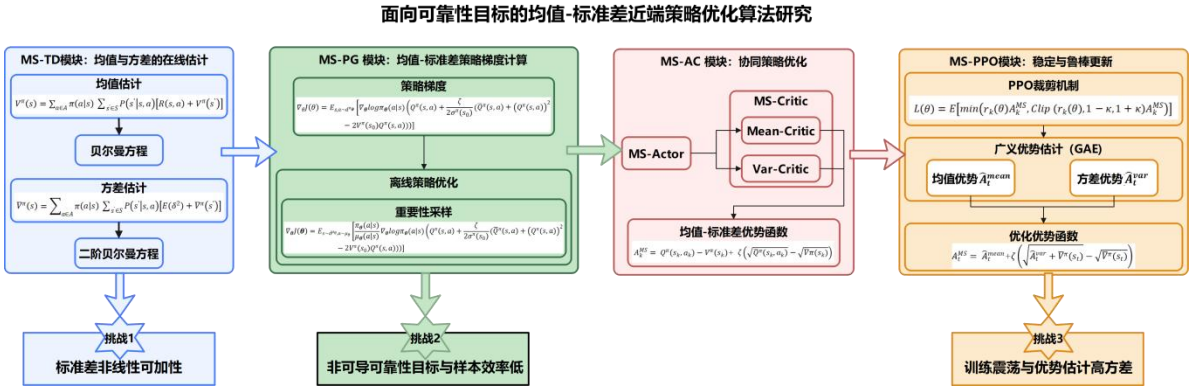


图 5 研究内容二：面向可靠性目标的均值-标准差近端策略优化算法研究方案图

①MS-TD 模块：均值与方差的在线估计

在强化学习中，累积回报的期望（均值）具有线性可加性，满足标准的贝尔曼方程：

$$V^{\pi}(s) = \sum_{a\in A}\pi(a|s)\sum_{s'\in S}P(s'|s,a)\left[R(s,a) + V^{\pi}(s')\right] \tag{2}$$

然而，标准差不具备线性可加性（即总标准差不等于各阶段标准差之和），导致无法直接建立递归关系。为解决此问题，本项目拟从方差的数学性质入手，通过分解即时方差贡献与未来方差贡献，推导出方差贝尔曼方程：

$$\overline{V}^{\pi}(s) = \sum_{a\in A}\pi(a|s)\sum_{s'\in S}P(s'|s,a)\left[E(\delta^2) + \overline{V}^{\pi}(s')\right] \tag{3}$$

其中，$\delta = r + V^{\pi}(s') - V^{\pi}(s)$ 为时序差分误差，$\overline{V}^{\pi}(s)$表示策略π下车辆在状态s时的回报方差（方差状态值函数），$E(\delta^2)$为方差期望项，实现了方差的递归建模。

考虑到日常停车状况的动态性和不确定性，环境条件概率难以准确获取，同时为了解决传统系统对固定模型和离线数据的依赖，使系统能够灵活应对动态变化的停车环境并且在线学习，本研究采用无模型的强化学习范式。在无模型设置下，时序差分学习能够通过与环境的实时交互逐步估计均值与方差回报，无需依赖环境模型，故而被选为核心学习算法。TD 算法利用 Bellman 方程的递归特性，通过比较当前估计值与单步更新后的估计值差异（即时序差分误差）来迭代修正价值函数估计。采用单步 TD 误差更新均值状态值/动作值：

$$V^{\pi}(s_k) = V^{\pi}(s_k) + \alpha_k(r_{k+1} + V^{\pi}(s_{k+1}) - V^{\pi}(s_k)) \tag{4}$$

同时，利用“TD 误差平方”近似即时方差贡献，更新方差状态值/动作值：

$$\overline{V}^{\pi}(s_k) = \overline{V}^{\pi}(s_k) + \overline{\alpha}_k(\delta_k^2 + \overline{V}^{\pi}(s_{k+1}) - \overline{V}^{\pi}(s_k)) \tag{5}$$

其中，$\delta_k = r_{k+1} + V^{\pi}(s_{k+1}) - V^{\pi}(s_k)$，$0 < \alpha_k$，$\overline{\alpha_k} < 1$ 为学习率。通过上述迭代过程，价值函数估计将收敛至当前策略与环境下的真实值。

②MS-PG 模块：均值-标准差策略梯度计算

有了 MS-TD 算法模块估计出的均值和方差的状态值/动作值，为了确定优化方向，采用 MS-PG 算法，通过梯度上升最大化累积回报期望，让智能体学习“该选什么动作”的最优决策规则。基于目标函数 $J(\theta)= V^{\pi}(s_0) + \zeta\sigma^{\pi}(s_0)$（其中 $\sigma^{\pi}(s_0) = \sqrt{\overline{V}^{\pi}(s_0)}$），推导出策略梯度表达式：

$$\nabla_{\theta}J(\theta) = E_{s,a\sim d^{\pi\boldsymbol{\theta}}}\left[\nabla_{\boldsymbol{\theta}} log\pi_{\boldsymbol{\theta}}(a|s)\left(Q^{\pi}(s,a) + \frac{\zeta}{2\sigma^{\pi}(s_0)}(\overline{Q}^{\pi}(s,a) + \left(Q^{\pi}(s,a)\right)^2 - 2V^{\pi}(s_0)Q^{\pi}(s,a)))\right] \tag{6}$$

其中，$d^{\pi\boldsymbol{\theta}}$ 为策略诱导的状态-动作分布。该梯度同时融合了“均值贡献”与“方差贡献”，确保策略更新过程兼顾搜索效率与稳定性。

为了进一步提升效率并于 MS-TD 模块保持一致，本研究进一步推导了离线策略 MS-PG 算法。离线策略的 MS-PG 算法使得执行者（Actor）能够利用行为策略 μ_{θ} 生成的数据进行学习，与评论家（Critic，即 MS-TD）的运行机制保持一致，进而使整个演员-评论家架构能够从经验回放缓冲区（Replay Buffer）的高样本效率中获益。通过重要性采样原理，可将在线策略期望转换为离线策略等效形式，推导结果如下：

$$\nabla_{\boldsymbol{\theta}}J(\boldsymbol{\theta}) = E_{s\sim d^{\mu\theta},a\sim\mu_{\theta}}\left[\frac{\pi_{\boldsymbol{\theta}}(a|s)}{\mu_{\boldsymbol{\theta}}(a|s)}\nabla_{\boldsymbol{\theta}} log\pi_{\boldsymbol{\theta}}(a|s)\left(Q^{\pi}(s,a) + \frac{\zeta}{2\sigma^{\pi}(s_0)}(\overline{Q}^{\pi}(s,a) + \left(Q^{\pi}(s,a)\right)^2 - 2V^{\pi}(s_0)Q^{\pi}(s,a)))\right] \tag{7}$$

推导过程中的关键近似为 $d^{\pi_\theta}(s) \approx d^{\mu_\theta}(s)$，即利用行为策略的状态访问分布替代目标策略的状态访问分布[1]。这是离线策略演员-评论家方法中的常见假设，只要行为策略 μ_{θ} 与目标策略 $\pi_{\boldsymbol{\theta}}$ 不存在显著差异，该假设通常成立。

③MS-AC 与 MS-PPO 模块：协同策略优化

MS-AC 算法采用执行者-评判者架构：评判者模块首先估计状态价值函数 $V^{\pi}(s_t)$

或动作价值函数$Q^{\pi}(s_t, a_t)$，并使用时序差分算法更新价值函数参数；执行者模块则基于 MS-TD 模块得到的均值与方差状态价值，通过“均值-标准差优势函数”更新策略。该优势函数定义为：

$$A_k^{MS} = Q^{\pi}(s_k, a_k) - V^{\pi}(s_k) + \zeta\left(\sqrt{\overline{Q^{\pi}}(s_k, a_k)} - \sqrt{\overline{V}\pi(s_k)}\right) \tag{8}$$

用于量化特定动作相对于当前状态平均水平的综合优势。

为避免策略更新幅度过大导致的训练震荡，本研究在 MS-AC 基础上引入 MS-PPO 模块。MS-PPO 采用裁剪后的代理目标函数确保单调改进：

$$L(\theta) = E\left[min\left(r_k(\theta)A_k^{MS}, Clip\left(r_k(\theta), 1-\kappa, 1+\kappa\right)A_k^{MS}\right)\right] \tag{9}$$

其中，$r_k(\theta) = \frac{\pi_\theta(a_k|s_k)}{\pi_{\boldsymbol{\theta}_{old}}(a_k|s_k)}$为评估策略与行为策略间的重要性采样比率，$0 < \kappa < 1$为裁剪参数。该裁剪机制具有双重作用：当$A_k^{MS} > 0$ 时，防止策略过大幅度更新；当$A_k^{MS} < 0$ 时，避免过度抑制动作概率。

MS-PPO 虽通过裁剪机制缓解策略更新震荡，但初始优势函数A_k^{MS}依赖单步 TD 估计，存在高方差缺陷——单步信息未融入多步累积数据，导致优势估计噪声大，且噪声会通过方差值函数更新(依赖δ_t^2)进一步放大，干扰策略梯度方向。为此，本研究进一步引入广义优势估计(Generalized Advantage Estimation, GAE)，通过融合多步 TD 误差的指数加权信息，平衡“偏差-方差”，提升优势估计精度与优化稳定性。

GAE 本质是“TD 误差的指数加权平均”，通过参数$\lambda \in [0, 1]$调节信息侧重：趋近 1 时侧重多步信息（接近蒙特卡洛，偏差小），趋近 0 时侧重单步信息（接近单步 TD，方差小）。针对 MS-PPO“均值-标准差”双目标，GAE 扩展至两个维度：

- **均值优势 $GAE(\widehat{A}_t^{mean})$**：基于 MS-TD 的均值值函数，定义单步 TD 误差$\delta_t^V = r_{t+1} + V^{\pi}(s_{k+1}) - V^{\pi}(s_k)$（$r_{t+1}$为停车搜索时间消耗，$V^{\pi}(s_k)$为当前状态均值值函数）。均值优势 GAE 理论上是未来 TD 误差的指数加权和，推导递归形式：

$$\widehat{A}_t^{mean} = \delta_t^V + \gamma\lambda\widehat{A}_{t+1}^{mean} \tag{10}$$

- **方差优势 $GAE(\widehat{A}_t^{var})$**：针对方差值函数更新特性，定义单步 TD 误差$\overline{\delta}_t = \delta_t^2 + \overline{V}\pi(s_{k+1}) - \overline{V}\pi(s_k)$（$\overline{V}\pi$为当前状态方差值函数）。类比均值维度，方差优势 GAE 的递归形式为：

$$\widehat{A}_t^{var} = \overline{\delta}_t + \gamma\lambda\widehat{A}_{t+1}^{var} \tag{11}$$

与均值维度共享 λ，确保双维度优化一致性。将均值$GAE(\widehat{A}_t^{mean})$与方差$GAE(\widehat{A}_t^{var})$代入原优势函数，得到优化后的均值-标准差优势函数：

$$A_t^{MS} = \widehat{A}_t^{mean} + \zeta\left(\sqrt{\widehat{A}_t^{var} + \overline{V}^{\pi}(s_t)} - \sqrt{\overline{V}^{\pi}(s_t)}\right) \tag{12}$$

该函数通过多步信息融合降低方差，且因方差组件的平方根特性形成非线性递归关系，更贴合标准差非可加性需求。这种基于 GAE 的优势估计器（记为A_t^{MS}）取代了 MS-PPO 裁剪代理目标中原始的单步优势估计器。这一整合产生了更稳健的算法，通过利用多步信息降低方差，为策略优化提供更可靠的优化信号，从而在停车场网络场景中实现更稳定、更有效的策略优化。

(3) **研究内容三：算法验证平台搭建与实车部署测试**

为了全方位验证 MS-PPO 算法的有效性，本研究设计了“仿真训练—虚实迁移—实车验证”的闭环测试方案，全方位论证 MS-PPO 在可靠停车空间搜索任务中的可行性，实验设计聚焦“数据代表性、配置严谨性、指标全面性”三大核心，严谨全面的验证了 MS-PPO 算法的可行性和优越性，具体技术路线如**图 6** 所示。

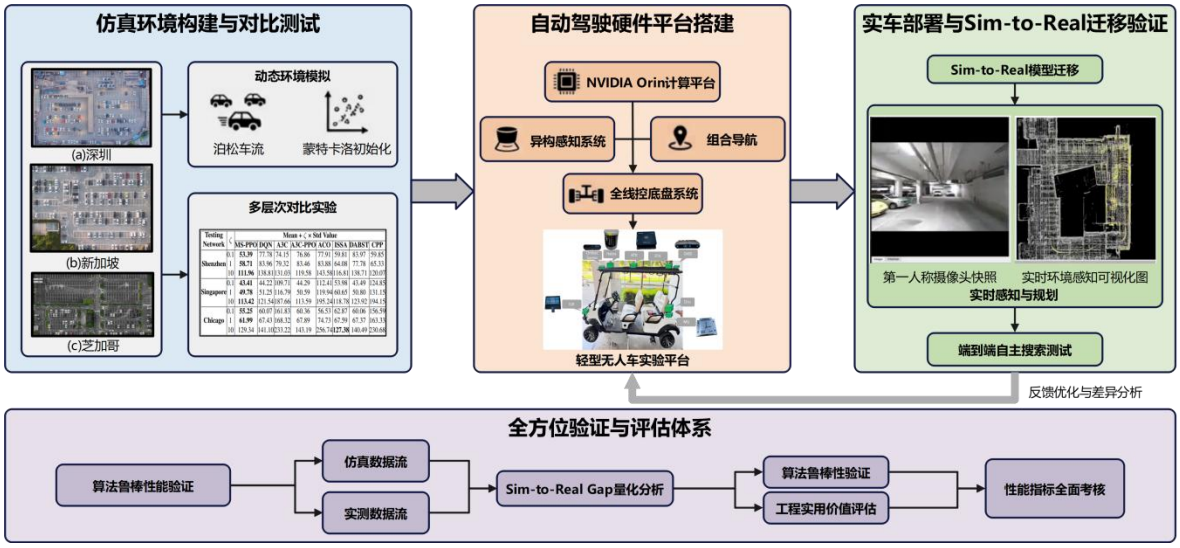


图 6 研究内容三：算法验证平台搭建与实车部署测试方案图

①仿真环境构建与对比测试

为实现对不同拓扑复杂度下停车搜索策略的泛化能力验证，本项目拟基于真实世界的物理网络构建高保真拓扑仿真环境。首先，选取深圳（20 节点、28 边）、新加坡（28 节点、37 边）及芝加哥（30 节点、43 边）三组具有代表性的真实停车场数据

集（如**图 7** 所示），构建三个异构仿真场景：将深圳数据集映射为空间受限、路口密集的“紧凑型 Urban 停车场”；将新加坡数据集重构为结构复杂、存在层级转换的“多层中型停车场”；将芝加哥数据集还原为路网稀疏、覆盖面积广的“大型地面停车场”。在此模型中，精确定义节点为停车场入口或内部交叉口，边为行车通道，确保仿真拓扑与真实物理环境的连通性高度一致。其次，设计符合泊松分布的随机车流生成机制，以模拟真实场景中因潮汐车流、临时卸货或行人穿行导致的道路通行时间波动，为算法提供具有非平稳特性的动态测试环境。最后，引入蒙特卡洛初始化方法，在每个仿真场景中随机生成 50 个停车空间搜索起始点，模拟车辆进入停车场时的不同初始状态，以全面覆盖从“入口直达”到“深层搜索”的各类长尾场景，贴合真实停车过程中面临的高度不确定性。

本项目将在上述仿真环境中开展多层次的正交实验，涵盖基准性能对比、风险敏感度分析及极端场景测试，以系统评估 MS-PPO 算法的综合效能。首先，在标准交通流密度下，针对上述三种不同拓扑复杂度的场景，分别运行 MS-PPO 算法与 DQN、A3C、PPO 等基准算法，记录平均搜索时间、时间标准差及路径规划成功率，验证本算法在不同规模网络中的基础搜索效率与收敛稳定性。其次，开展风险偏好敏感性实验，通过调整风险系数ζ，评估算法在“激进搜索”（追求最短时间但容忍高波动）与“保守搜索”（追求稳定时间规避拥堵风险）模式下的策略差异，分析其对不同用户心理预期的适应性。最后，模拟高动态突变场景（如主干道突发拥堵或目标车位被抢占），测试算法在环境状态剧烈变化下的实时重规划能力，验证 MS-PPO 在复杂动态约束下的鲁棒性与决策优势。

(a)深圳停车场　　(b)新加坡停车场　　(c)芝加哥停车场

图 7 真实停车场的实景鸟瞰图

②自动驾驶硬件平台搭建

为了支撑 MS-PPO 算法从仿真环境向物理世界的 Sim-to-Real 迁移验证，项目组将

参照 L4 级自动驾驶系统标准，搭建一套具备高机动性与高扩展性的线控自动驾驶实验平台（如**图 8** 所示）。该平台采用模块化设计理念，通过异构传感器融合与高性能边缘计算的深度协同，确保在复杂的停车场光照条件与受限空间内实现精准决策。具体硬件架构详述如下：

- **NVIDIA Orin 高性能计算平台：**项目选用 NVIDIA Jetson AGX Orin 作为核心计算单元。该平台提供高达 275 TOPS 的 AI 算力，支持高并发的深度神经网络推理。在本项目中，Orin 将承担双重核心任务：一方面，利用其强大的 GPU 并行处理能力，实时运行 MS-PPO 算法，在毫秒级时延内完成从状态输入到动作输出的端到端推理；另一方面，负责处理多源传感器数据的时空同步与融合，运行基于 ROS 架构的通信中间件，确保感知、规划与控制模块间的高效数据流转，为应对动态障碍物提供充足的算力冗余。
- **Helios 激光雷达与 Gemini 深度相机：**针对停车场环境光线昏暗、立柱密集及死角众多的特点，构建“远近结合”的异构感知系统：搭载 RoboSense Helios 新一代多线机械激光雷达，利用其高密度的点云扫描能力，实现对车辆周围 360° 范围内建筑物、停放车辆及行人的高精度三维建模，构建实时占据栅格地图（Occupancy Grid Map），有效解决远距离障碍物检测问题。在车头及盲区配置 Orbbec Gemini 双目结构光深度相机。该相机具备优秀的抗环境光干扰能力，负责近距离的语义特征提取（如车位线识别、地锁检测）及 RGB-D 深度信息补全，弥补激光雷达在近距离纹理缺失的短板，确保车辆在狭窄车位泊入过程中的安全性。
- **RTK-GNSS + 高精度 IMU 组合导航：**为解决室内外场景切换及地下停车场无 GPS 信号的定位难题，采用紧耦合（Tightly-coupled）的组合导航方案：在室外露天停车场场景，利用 RTK（实时动态差分） 系统接收卫星信号，提供厘米级的绝对定位坐标。在进入室内或地下停车场信号遮挡区域时，系统无缝切换至基于高精度工业级 IMU（惯性测量单元） 的航位推算模式。结合轮速计数据与激光雷达里程计（LiDAR Odometry），通过扩展卡尔曼滤波（EKF）算法融合多源位姿数据，有效抑制长时间运行产生的累积误差，确保车辆在无卫星信号环境下仍能保持高精度的自我定位能力。
- **全线控底盘系统：**选用具备完整 CAN 总线通讯协议的开发型线控底盘，作为算法落地的物理载体。该底盘支持 Ackermann（阿克曼）转向几何，能够真实模拟

乘用车的运动学约束。通过解析上层规划指令，对驱动电机与电子液压制动系统进行精细化扭矩控制，实现平滑的加减速与定点停车。采用高响应精度的电动助力转向系统，支持大角度转向操作，满足在狭窄停车场通道内的掉头与直角转弯需求。此外，底层控制器内置硬件级安全逻辑，当上层算法输出异常或通信中断时，可立即接管车辆进入紧急制动状态，保障实验过程的物理安全。

图 8 轻型无人车实验平台

③实车部署与 Sim-to-Real 迁移

为验证 MS-PPO 算法在物理世界中的泛化能力与实战性能，本研究将实施严格的 Sim-to-Real（仿真到现实） 迁移验证实验。实验流程遵循“虚实映射-模型迁移-实地验证-差异分析”的闭环逻辑，具体方案如下：

首先，构建与真实地下停车场拓扑结构高度一致的高保真仿真环境。在仿真中引入传感器噪声模型与动力学摩擦系数，对 MS-PPO 智能体进行大规模强化学习训练。待策略网络收敛并表现出稳定的寻位能力后，冻结模型参数，将训练好的网络模型进行轻量化封装，并无缝部署至搭载 NVIDIA Orin 平台、Helios 激光雷达及 Gemini 深度相机的自动驾驶实验车上，实现从虚拟训练场到物理计算平台的零样本迁移。

其次，在实车运行过程中，感知系统实时处理多源数据。如**图 9** 所示，系统展示了车辆在搜索过程中的双重视角：左侧展示了车辆导航过程中的第一人称摄像头快照，反映了地下停车场光照条件复杂、纹理特征重复及存在动态遮挡的真实环境挑战；右侧展示了基于 RViz 工具生成的实时环境感知可视化图。算法将激光雷达采集的高频

点云数据实时转化为占据栅格地图，清晰勾勒出墙体边界、停放车辆及立柱障碍物，直观呈现了算法对物理环境的几何建模能力与障碍物检测精度。

然后，实验场地选定为结构复杂的真实地下停车场，在非管制状态下进行，以保留真实社会车辆与行人的干扰因素。实验设计涵盖直线巡航、直角转弯、会车避让等多种工况。在每次测试中，车辆需在无任何人工干预及模型参数微调的前提下，完全依赖车载传感器与 MS-PPO 算法，自主完成“入口初始化-全局路径规划-局部动态避障-车位识别-精准停靠”的端到端全流程。

最后，将真实场景的实验数据与仿真环境下的基准数据进行对比，量化分析 Sim-to-Real Gap，探究光照变化、地面摩擦力差异及传感器误差对算法性能的具体影响，从而验证 MS-PPO 算法在非结构化真实环境中的鲁棒性与工程实用价值。

图 9 搜索过程快照及局部感知图

3.3 可行性分析

本项目聚焦动态不确定场景下的智能停车位可靠搜索任务需求，涵盖面向非线性“均值-标准差”复合目标的强化学习数学建模、均值-标准差近端策略优化（MS-PPO）算法设计、以及在线学习与方差约束机制等研究内容，并在高保真仿真环境与真实地下停车场物理场景进行算法验证，研究任务明确。尽管当前专门针对兼顾效率与稳定性的停车搜索策略研究成果有限，但风险敏感强化学习与自动驾驶决策规划领域的最新进展可以为本项目提供研究思路。本项目的研究内容与研究方案是在现有理论与方法的基础上，结合申请人前期研究成果提出的，具有较强的理论支撑。

(1) 面向非线性“均值-标准差”复合目标的强化学习建模可行性分析

技术路线（1）中涉及的马尔可夫决策过程（MDP）建模方法具备数学定义严谨、状态转移描述清晰等特点，近年来已广泛应用于自动驾驶路径规划、交通信号控制等领域，在提升决策效率与优化资源分配方面展现出良好效果。文献[1, 2]针对自动驾驶泊车与路径规划问题，通过深度强化学习（DRL）实现了端到端的策略学习。该方法证明了将车辆状态与环境感知信息映射为动作空间的可行性。由此为启发，本项目拟在经典 MDP 的基础上引入方差惩罚项，以增强其对搜索过程波动性的刻画能力。此外，本项目将推导方差贝尔曼方程，以指导 MS-PPO 解决其在非线性目标函数建模上的局限性，从而提升对“快且稳”搜索任务的表达能力。针对动态停车场景下非线性可靠搜索目标函数与环境概率转移矩阵未知的挑战，项目将构建基于广义优势估计（GAE）的双维度扩展框架，以弥补传统 TD 学习在方差估计精度方面的不足。

(2) 均值-标准差近端策略优化（MS-PPO）算法设计可行性分析

技术路线（2）中涉及的策略优化方法是一种基于在线交互数据的智能体自适应优化和稳定更新方法，由于其具备高效样本利用率与稳定收敛性，近年来在机器人控制、游戏 AI、自动驾驶等众多领域被广泛应用，能够提升任务执行效率、增强系统稳定性。文献[3]针对传统策略优化方法难以平衡探索与利用的问题，提出将模仿学习与 PPO 相结合的方法。该方法通过引入专家示范数据引导策略更新，实现了在复杂泊车环境下的快速收敛，证明了 PPO 算法框架在停车任务中的适用性。

此外，申请人还提出了一种基于裁剪机制的方差约束方法，能够在同一套训练框架下平衡均值最大化与方差最小化，相关成果已申请国家发明专利。本项目将在申请人已有的算法设计与优化技术基础上，结合在线学习机制，对动态变化的停车环境进行实时适应。同时，利用 MS-AC 架构对网络进行在线优化，深入开展面向动态停车场景的可靠搜索策略研究，解决经典算法在非线性目标下容易陷入震荡的难题，提升车辆在复杂环境中的决策稳定性。

(3) 仿真到真车（Sim-to-Real）的迁移验证可行性分析

技术路线（3）中的 Sim-to-Real 迁移是自动驾驶与机器人领域实现算法落地应用的核心技术，具有降低实车测试风险、加速算法迭代和增强系统鲁棒性等优点。该方法已成功应用于需要高精度感知与决策的复杂任务，如无人机穿越、机械臂抓取等，为本项目提供了重要参考。文献[4]提出了一种基于语义 SLAM 的泊车位识别与定位方法，将视觉感知与激光雷达数据融合，实现了复杂光照条件下的精准定位。该方法

证明了多传感器融合技术在弥补单一传感器缺陷、提升环境感知可靠性方面的有效性。本项目拟依据文献[5]提出的感知融合思路，构建高保真仿真环境与真实物理环境的映射关系，确保仿真中的传感器噪声模型与真实世界保持一致。通过该方法，车辆根据仿真训练出的策略在真实环境中做出的决策将与仿真环境一致,从而实现“感知-决策”回路的有效闭环。在面向物理世界非结构化特征（如地面摩擦系数变化、动态行人干扰）的情况下，本项目将引入域随机化（Domain Randomization）技术，在仿真训练阶段引入多样化的扰动，确保算法在部署到真车时具备足够的泛化能力。

(4) 研究平台与实验基础可行性分析

本项目涉及深度强化学习、计算机视觉、SLAM 建图定位、路径规划、车辆控制等多个领域的研究。申请人在过去 2 年里，专注于智能交通系统、自动驾驶决策规划、强化学习应用等方面的研究，发表人工智能顶级会议 1 篇，并申请 1 项发明专利，为本项目顺利开展奠定坚实的学术基础。项目将依托四川大学机器学习与工业智能应用教育部工程研究中心，该平台在智能机器人控制、复杂环境感知、自动化装备等关键技术方面拥有扎实的研究基础。团队拥有一批经验丰富的科研人员，并在嵌入式系统开发、多传感器融合、车辆动力学控制等方面积累了丰富的研究经验，这为本项目提供了强有力的研究平台与实验基础支撑，确保项目的顺利推进。

团队与新加坡国立大学相关课题组建立了紧密的合作关系，并已购置了 RoboSense 激光雷达、Orbbec 深度相机、NVIDIA Jetson AGX Orin 计算平台以及线控底盘等关键硬件设备。这些硬件平台覆盖了环境感知、边缘计算及运动控制等核心环节，能够适应地下停车场光线昏暗、空间狭窄的复杂环境，为项目的实验验证提供了坚实的硬件基础。基于上述软硬件平台，申请人近期提出了一种基于 ROS 架构的自动驾驶模块化开发方案，为本项目中算法的集成与调试提供了核心架构支持。该方案通过标准化的消息接口，有效解决了感知、规划与控制模块之间的数据流转问题。相关研究成果已在校内开放道路测试，验证了系统的稳定性与实时性。本项目将进一步结合项目组现有的实验场地，基于申请人提出的 MS-PPO 算法，开展动态停车场景下可靠搜索策略的物理实验验证，为算法的实用化与工程化奠定基础。

综上所述，本项目的研究方案是在广泛阅读相关文献、深入理解现有理论与方法并结合申请人前期研究成果提出的，因此具有较强的合理性。申请人在前期工作基础上，综合考虑了依托单位的科研条件和国内外相关技术的发展现状及发展趋势，根据

动态停车场景的特点将总问题拆解为多个彼此独立且紧密关联的子问题开展研究，研究目标明确，研究方案具体。如“详细研究方案”章节所述，本项目为每个子问题提供了详尽可靠的研究思路与技术路线，详细地分析了每个子问题的研究方案中可能涉及到的理论知识及相关技术，确保项目的研究方案和技术路线具有可行性。

参考文献

[1] Zhang, Zheyu et al. “Automated Parking Trajectory Generation Using Deep Reinforcement Learning.” 2025 6th *International Conference on Computer Engineering and Application (ICCEA)* (2025): 516-520.

[2] A. R., Vimal Kumar and Raghu Ram Theerthala Theerthala. “Reinforcement Learning Based Parking Space Egress for Autonomous Driving.” *SAE Technical Paper Series* (2024): n. pag.

[3] Thunyapoo, Baramee et al. “Self-Parking Car Simulation using Reinforcement Learning Approach for Moderate Complexity Parking Scenario.” 2020 17th *International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (ECTI-CON) (2020): 576-579.

[4] Shao, Xuan et al. “ Towards a Robust Visual-Inertial-Surround-View SLAM System for Autonomous Indoor Parking.” *ACM Transactions on Multimedia Computing, Communications and Applications* 21 (2025): 1 - 23.

[5] Liu, Kai et al. “Efficient Topology-Aware Motion Planning for AVP in Large-Scale Occupancy Map.” *IEEE Transactions on Intelligent Transportation Systems* 27 (2026): 924-937.

（四）最终成果形式及成果预期水平

4.1 年度研究计划

本项目计划在两年内完成，其研究和实施过程分为以下阶段：相关文献查阅与调研；具体研究内容的技术攻关；搭建算法验证与测试的高保真仿真环境，评估并反馈 MS-PPO 算法的可靠性、实时性及收敛性；物理场景下的自动驾驶停车搜索算法实机整合、部署、运行与评估；地下停车场模拟场景下的实车搜索演习（外场实验演习）；总结各阶段研究成果并发表论文、申请国家发明专利。项目申请人近年来系统研究了强化学习、自动驾驶决策规划的国内外的主要研究成果，项目组成员多数是正在或曾经从事过深度强化学习、SLAM 建图与路径规划等领域的研究工作，研读了大量的相关文献资料，积累了丰富的研究经验。因此，本项目的文献查阅和调研阶段的工作不需要单独规划时间，整体工作具体进度安排如**图 10** 所示。

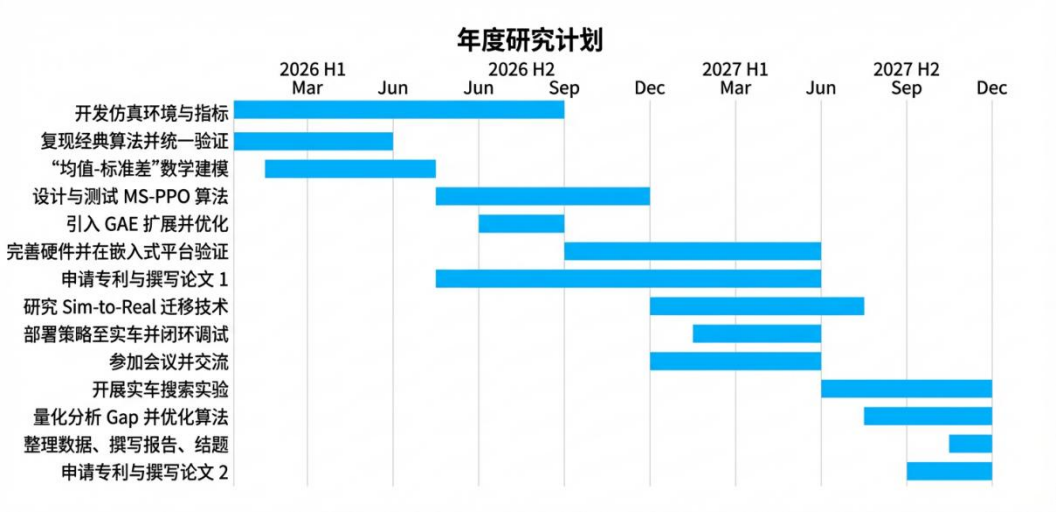


图 10 研究计划甘特图

工作具体安排如下：

2026 年 3 月 1 日至 2026 年 6 月 30 日

(1) 基于项目组已采集的地下停车场激光雷达 3D 点云数据、车位状态历史数据和已构建的栅格地图，开发对应的面向动态停车场景的生成式仿真环境（基于 SUMO 或 CARLA），提出可以衡量停车搜索算法可靠性、搜索效率（均值与方差）的数字化指标；

(2) 复现当前智能停车领域的经典算法，在生成式仿真环境中进行统一验证并给出对应基准算法的可靠性、搜索时间波动性等数字评价指标；

(3) 完成非线性“均值-标准差”复合目标的数学建模，推导方差贝尔曼方程，解决标准差非线性可加的理论难题。

2026 年 7 月 1 日至 2026 年 12 月 31 日

(1) 设计均值-标准差近端策略优化（MS-PPO）算法框架，实现 MS-TD（均值与方差在线估计）与 MS-PG（策略梯度计算）模块，在仿真环境中测试算法在车位状态动态变化情况下的收敛性；

(2) 引入广义优势估计（GAE）的双维度扩展，平衡均值与方差的估计偏差，设计在线学习与方差约束机制，优化奖励函数设计；

(3) 完善已有硬件平台，整合激光雷达、深度相机与线控底盘驱动，在实验室沙盘环境下验证 MS-PPO 算法在嵌入式平台（NVIDIA Orin）上的运行实时性与可行性；

(4) 围绕“均值-标准差”强化学习建模方法申请 1 项发明专利，并撰写 1 篇智能交通或机器人领域高水平会议（如 ITSC, IV 等）或期刊论文。

2027 年 1 月 1 日至 2027 年 6 月 30 日

(1) 研究 Sim-to-Real 迁移技术，引入域随机化（Domain Randomization）机制，在仿真中模拟传感器噪声、控制延迟及动态行人干扰，提升算法的泛化能力；

(2) 完善已搭建的自动驾驶实验车，将训练好的 MS-PPO 策略模型部署至实车，进行“感知-决策-控制”闭环调试；

(3) 参加智能交通或机器人领域学术会议，交流讨论已有学术成果，邀请专家研讨 Sim-to-Real 迁移中的难点，为项目后续推进提供指导意见。

2027 年 7 月 1 日至 2027 年 12 月 31 日

(1) 在真实地下停车场开展实车搜索实验，测试算法在光照变化、动态障碍物干扰下的鲁棒性；

(2) 采集实车实验数据，量化分析 Sim-to-Real Gap，对比仿真与实车环境下的搜索效率与稳定性指标，对算法进行微调与优化；

(3) 整理项目研究数据与代码，撰写项目结题报告，准备结题、评估和验收；

(4) 围绕可靠停车位搜索策略申请 1 项发明专利，并撰写 1 篇 SCI 检索期刊论文。

在项目研究过程中，项目组会尽可能按照原计划对项目各研究内容进行高效展开与推进，在发生一些不可预知的困难制约项目进展时，也会及时调整和改进。因此，上述研究计划在实际过程中会根据需要及时调整，以获得理想的研究成果。

4.2 预期研究成果

本项目针对动态不确定场景下的智能停车位可靠搜索问题展开研究，深入分析停车环境动态性与感知不确定性对搜索效率的影响，以及可靠性搜索任务的非线性目标函数，建立面向“均值-标准差”双目标的马尔可夫决策过程模型。针对搜索时间波动性大的问题，设计并优化均值-标准差近端策略优化（MS-PPO）算法；针对仿真到真车迁移难的特点，构建 Sim-to-Real 迁移技术与高保真测试平台。最终，形成一套“稳中求快”的可靠停车位搜索策略，在生成式仿真环境与真实地下停车场物理场景中进行系统性验证。具体成果如下：

(1) **理论成果**

- 建立面向动态不确定环境的停车位搜索任务“均值-标准差”马尔可夫决策过程（MDP）模型；
- 设计并优化均值-标准差近端策略优化（MS-PPO）算法，推导方差贝尔曼方程与广义优势估计（GAE）公式；

- 构建基于在线学习的方差约束机制，解决非线性目标函数下的策略震荡问题；
- 发表中科院 SCI/EI 检索论文 2-3 篇，人工智能或智能交通领域旗舰会议论文如 ITSC, IV, IROS 等 1-2 篇；

(2) **系统成果**

- 研发一套面向动态停车场景的自动驾驶可靠搜索系统，包括环境感知、SLAM 建图、MS-PPO 决策规划模块；
- 构建一套可用于评价、标定智能停车算法性能的高保真仿真测试平台（基于 SUMO/CARLA）；
- 搭建一台具备完整自动驾驶能力的线控实验车平台（搭载激光雷达、深度相机、NVIDIA Orin）；
- 基于可靠停车位搜索算法与系统应用,申请国家发明专利 2 项,软件著作权 1 项；

（五）项目组成员

序号	姓名	职称	学位	出生年月	单位	现从事专业	分工	签字
1							项目负责人	

六、资金预算表

预算科目名称	金额 （元）	预算根据或理由
设备费	0	本项目依托现有实验室设备，无需购置大型设备
材料费	1,500	购买实验用电子元器件
测试化验加工费	0	无需外协测试
燃料动力费	0	依托实验室现有条件
差旅费	1,000	用于项目组成员参加学术会议的注册费
会议费	0	不主办会议
国际合作与交流费	0	无此类计划
出版/文献/信息传播/知识产权事务费	2,000	论文版面费
劳务费	500	支付给项目成员的津贴
专家咨询费	0	预算有限，暂不列支
管理费	0	由学校统筹
其他费用	0	无
总经费合计	5000	伍仟元整

七、单位科研部门审查意见

负责人：　　（单位盖章）

年　月　日

八、机器人工程与智能制造南充市重点实验室学术委员会审查意见

学术委员会主任：　　（签字）

年　　月　　日

九、机器人工程与智能制造南充市重点实验室审批意见

实验室主任：　　（签字）

年　　月　　日