基于拓扑博弈的自动驾驶：受 AlphaZero 范式启发的路网自博弈决策框架

Lanhaijian　大模型辅助

摘要

当前主流自动驾驶决策训练方案高度依赖海量真实驾驶数据与高性能计算集群，典型方案包括端到端视觉模仿学习、人工定制场景强化学习两类，普遍存在模型可解释性差、长尾边缘场景覆盖不足、对陌生路网泛化能力弱等缺陷。在数据与算力资源受限的落地条件下，这类高数据、高算力消耗的开发路线实施难度极大。针对该痛点，本文借鉴 AlphaZero 自博弈范式，提出一套基于拓扑博弈的自动驾驶决策框架。

该框架将复杂城市路网拆解为标准化道路基元，支持模块化、可重置道路博弈环境程序化生成；将车辆驾驶决策从传统连续物理控制任务，重构为结构化路网拓扑上的时序多目标组合决策问题。框架摒弃暴力像素级数据拟合，采用图神经网络提取路网结构化拓扑嵌入；借助蒙特卡洛树搜索（MCTS）实现多步前瞻决策推理，并设计分层多目标奖励体系平衡行车安全、交规合规、通行效率与乘坐舒适性。通过仿真自博弈生成高质量驾驶轨迹，结合策略蒸馏实现车端轻量化部署。

此外，本文引入专用泊车基元、细分离散动作集与位姿优化奖励函数，将拓扑博弈建模从道路行驶场景拓展至泊车场景；基于统一博弈决策机制，在单网络架构内完成行驶-泊车跨尺度任务处理，探索全场景自动驾驶统一建模潜力。

本文属于基于学习的运动规划框架类研究，大幅降低对人工枚举交互场景、大规模人工标注驾驶数据的依赖，使车辆交互行为依托拓扑结构先验与自博弈迭代自主涌现。理论分析表明，该框架借助结构化道路先验压缩高维状态空间，具备决策逻辑可解释、场景生成模块化可扩展、陌生路网拓扑泛化能力强、交互驾驶行为自主涌现、行驶泊车任务统一建模等优势。本文提出一套假设驱动的研究框架，支撑自动驾驶决策模型的方案设计、分阶段验证与迭代优化，为数据算力受限条件下智能驾驶研发提供全新技术路径。

关键词：自动驾驶；拓扑博弈；路网基元；行驶泊车一体化；运动规划；AlphaZero；蒙特卡洛树搜索；自博弈强化学习；图神经网络

1　引言

1.1　现有自动驾驶范式核心瓶颈

现阶段自动驾驶智能决策系统主流分为两条技术路线：端到端视觉模仿学习、基于人工场景的强化学习。两类方案虽在工业落地中取得阶段性成果，但范式底层固有缺陷限制自动驾驶通用能力进一步提升。

1.1.1　端到端视觉模仿学习范式缺陷

以大规模车队学习方案为代表的端到端视觉模仿，依靠海量真实行车轨迹与高性能算力集群，从像素级传感数据拟合人类驾驶行为，存在三大核心短板：1. 依赖大规模数据积累与高端算力资

源，无同等车队数据支撑的科研、工业团队难以复现；2. 模型属于黑盒系统，无显式拓扑推理逻辑，故障溯源困难，无法满足高阶自动驾驶安全可解释要求；3. 模型仅学习视觉表观特征，忽略路网内在连通拓扑结构，面对陌生城市道路布局泛化性能极差。

1.1.2 人工场景强化学习范式缺陷

依托 CARLA、LGSVL 等仿真平台搭建路口、高速、城市道路人工场景训练决策模型，核心局限为场景覆盖不足：人工设计场景无法系统性覆盖真实路网海量拓扑组合；长尾极端场景仅能通过实车路测随机采样，无法保证决策模型结构鲁棒性，易对固定预设场景过拟合。

传统基于规则、基于采样的运动规划方法（A*、RRT、晶格规划器等）依靠人工代价函数与路径模板生成无碰撞确定轨迹，但无法建模多车协商、礼让、并道等社会性驾驶交互，难以适配真实交通动态博弈交互场景。

同时，行业现有行驶泊车一体化方案本质是独立行驶、泊车模块功能堆叠，两套决策系统分属不同逻辑框架，算法层面未实现统一空间拓扑建模，无法原生支撑两类场景通用决策。

根本上，现有主流技术路线未充分利用路网结构属性：城市路网并非无结构纯视觉场景，而是由有限基础拓扑单元组合而成的组合系统。传统方法将驾驶简化为纯物理运动控制，忽视道路拓扑先验，引发高维状态空间爆炸、学习效率低下、泛化性差，无法统一宏观道路通行与微观车位占用决策。

1.2 AlphaZero 博弈范式迁移可行性与局限

AlphaZero 依托规则约束自博弈、自主涌现最优策略，无需大量人类专家棋谱数据，核心组件包含结构化封闭博弈环境、离散时序决策动作、策略-价值双网络评估、蒙特卡洛树搜索多步前瞻、迭代自博弈数据生成。该范式降低对人工专家数据依赖，在固定规则约束下自主探索收敛至有效策略。

驾驶场景与棋类游戏存在逻辑相似性：路网具备相对固定拓扑结构与交通法规约束，驾驶行为可抽象为带约束时序决策过程，为 AlphaZero 自博弈思路迁移至自动驾驶决策提供基础可行性。但真实驾驶与棋类博弈存在本质差异：棋类规则固定、最优结果清晰唯一；自动驾驶是兼顾安全、效率、舒适性的多目标均衡任务，不存在全局唯一最优策略。该核心差异决定无法直接照搬棋类自博弈流程，需针对性设计奖励函数与多智能体交互约束。

基于上述分析，本文借鉴 AlphaZero 自博弈核心思想而非完整复刻棋类训练管线：基于道路拓扑基元搭建模块化博弈环境，将高层驾驶意图抽象为时序决策走子，依托交通法规定义博弈约束，通过仿真自博弈迭代生成驾驶策略；同时行驶、泊车均可抽象为空间尺度、约束密度不同的拓扑空间占用博弈，为跨场景统一决策建模提供可行基础。

1.3 本文核心贡献

本文聚焦系统级框架设计与场景建模创新，不提出全新基础算法，主要贡献总结如下：1. **拓扑博弈建模框架**：构建拓扑驱动自博弈决策框架，将连续驾驶控制任务重构为模块化路网拓扑上的时序组合决策问题，为高数据消耗模仿学习、人工枚举场景强化学习提供替代技术路线；2. **模块化道路拓扑基元设计**：定义 7 类典型道路基元用于生成可扩展城市行驶场景，配套拓展泊车拓扑基元，支持宏观路网、微观泊车场地模块化生成；3. **跨尺度拓扑编码设计**：引入拓扑感知混合专家网络 TopoMoE 骨干，自适应编码稀疏宏观路网拓扑与稠密微观泊车拓扑，实现行驶、泊车场景统一特征提取；4. **算法原生行驶泊车一体化架构**：区别于行业独立模型功能堆叠方案，本框架采用统一图状态表征、统一 MCTS 博弈搜索、统一自博弈迭代逻辑，探索原生跨场景决策能力；

5. **假设驱动分阶段验证体系**：针对高层离散动作有效性、图编码器拓扑泛化、多智能体交互行为稳定涌现、跨尺度拓扑适配四大核心科学假设设计分阶段实验验证方案，指导框架迭代与性能评估。

2 相关工作

2.1 面向道路拓扑建模的几何深度学习

几何深度学习（GDL）支持图结构空间数据的结构感知特征学习，已广泛应用于自动驾驶路网图表征、交通流量预测、车辆轨迹预测任务。图神经网络 GNN 可有效编码车道连通性、路口拓扑，面对陌生道路布局时泛化性能优于基于图像的卷积网络。

但现有自动驾驶 GNN 应用大多聚焦静态拓扑感知与轨迹预测，仅将路网拓扑视作固定环境特征，而非用于时序决策、多智能体交互规划的可编程博弈棋盘；极少工作将宏观路网拓扑、微观泊车拓扑纳入同一套图嵌入框架，导致行驶、泊车模型分离设计。

区别于现有感知导向 GNN 管线，本文将拓扑图表征从被动环境特征提取拓展至主动拓扑博弈决策：将道路图构建为可生成博弈环境，行驶、泊车行为建模为时序拓扑走子，依托拓扑结构先验支撑自博弈策略优化。

2.2 面向多尺度图学习的拓扑感知混合专家网络

拓扑感知混合专家网络 TopoMoE 依据图结构属性实现稀疏专家路由，对异质拓扑子图自适应特征提取。对比固定结构 GNN，TopoMoE 可动态分配不同专家模块处理密度、连通性差异显著的图区域，提升异质空间结构表征多样性。

现有自动驾驶、机器人领域 TopoMoE 研究多用于场景分类、图分割、自适应环境感知，极少用于解决稀疏道路交通场景与稠密泊车场景间跨尺度拓扑异质性问题，亦未部署至长时序运动规划任务。

本文定制面向决策的 TopoMoE 骨干，统一双尺度拓扑建模：宏观稀疏路网拓扑、微观稠密泊车拓扑自动路由至对应专家模块，无需双分支模型堆叠即可完成行驶泊车统一编码，作为拓扑博弈规划跨场景泛化核心中间层。

2.3 拓扑强化学习与 GRPO 优化

PPO、A2C 等传统强化学习算法在交互交通场景训练方差大、多智能体信用分配不稳定。组相对策略优化 GRPO 依托组内轨迹排序优化策略优势，无需独立价值网络，大幅降低计算开销、提升训练稳定性；拓扑感知 GRPO 进一步将图结构约束纳入策略更新，使强化学习具备空间邻接、拓扑差异感知能力。

但现有拓扑 GRPO 研究大多局限于多智能体通信优化、简单网格世界控制任务，未与长时序蒙特卡洛树搜索自博弈系统性结合用于自动驾驶运动规划。

本文框架中，MCTS 生成高质量多步拓扑博弈轨迹作为教师监督信号；拓扑 GRPO 集成至自博弈迭代管线，稳定策略更新、缓解多智能体自博弈收敛至极端纳什均衡问题，作为策略优化模块而非替代 MCTS 搜索机制。

2.4 与现有工作的差异化总结

现有几何深度学习、TopoMoE、拓扑强化学习研究大多单独解决拓扑感知、多尺度编码、策略优化子问题，极少将上述组件系统性整合为面向全场景行驶泊车任务的统一拓扑博弈决策框架。

本文搭建三层技术体系：1. 几何深度学习提供拓扑结构先验与基于图的状态表征；2. TopoMoE 实现行驶、泊车拓扑特征自适应多尺度统一；3. 拓扑 GRPO 稳定基于 MCTS 的自博弈迭代、完成拓扑博弈策略优化。

对比传统基于学习的运动规划方法，本框架将自动驾驶从高数据消耗模仿、规则依赖确定性规划，转变为拓扑结构化、自博弈演化的博弈规划，旨在提升决策可解释性、结构泛化能力与跨场景通用性。

3 自动驾驶拓扑博弈建模系统

3.1 路网基元：可生成博弈环境基础单元

拆解城市道路、高速、乡村路网典型拓扑结构，抽象 7 类核心行驶道路基元作为可生成博弈环境基础模块；真实公共路网可通过基元组合、延伸、嵌套搭建：1. 直道单元：线性基础结构，用于匀速跟车、直行通行场景；2. 正交十字路口单元：标准城市四岔路口拓扑，覆盖直行交汇、车流冲突、转向交互场景；3. T 型岔路单元：三向分支拓扑，适配主干道汇入、侧向路口通行；4. 环岛单元：多径向入口闭合环形拓扑，代表无序多车交互典型场景；5. 鱼骨带状单元：主干分支带状拓扑，适配沿河、交通干线狭长城市道路；6. 放射枢纽单元：中心节点放射拓扑，对应老城核心放射状路网；7. 分支树状单元：无闭环分层分支拓扑，适配城郊、山区、乡村非结构化路网。

为拓展框架跨场景泛化能力，本文定义第八类拓展基元：停车场拓扑单元，用于建模封闭、高密度、高约束微观空间泊车博弈场景。

3.1.1 停车场拓扑单元（通用拓展基元）

停车场与宏观公共道路拓扑差异显著，属于高密度、窄通道结构化拓扑，带终端泊车约束，可纳入本框架统一图拓扑体系：

- 拓扑定义：由主通行边、次级通道边、车位终端节点构成网格拓扑；
- 车位建模：垂直、平行、斜列车位统一抽象为带边界约束的终端拓扑分支；
- 场地特征：通道狭窄、位姿精度要求高、可通行空间紧凑；
- 博弈特征：多车抢位、通道会车、多次调整挪车、安全距离维持，典型高密度长时序决策任务；
- 尺度特征：定义为微观拓扑博弈场景，与宏观行驶拓扑配合实现从大型路网到小型停车场的全场景可扩展覆盖。

本质上，行驶路网、停车场是空间尺度、密度层级不同的结构化拓扑空间。本框架自适应调整图卷积感受野，实现宏观路网、微观泊车统一特征提取，构成算法原生行驶泊车一体化决策核心基础。

依托以上 8 类基元，框架可程序化生成宏观行驶博弈棋盘、微观泊车博弈棋盘、连接行驶与泊车的过渡场景棋盘，模块化可扩展覆盖自动驾驶典型场景。本拓扑博弈设计思路不限于道路行驶，具备拓展至其他带约束结构化交通空间决策问题的潜力。

3.2 博弈系统核心映射关系

借鉴 AlphaZero 棋类抽象博弈逻辑，建立行驶泊车统一博弈映射体系，如表1所示：

表 1: 棋类博弈与自动驾驶拓扑博弈概念映射

AlphaZero 棋类概念	自动驾驶拓扑博弈统一定义
固定博弈棋盘	道路/泊车基元程序化生成多尺度拓扑博弈场地
离散走子	宏观高层驾驶决策走子 + 细粒度泊车连续调整走子
博弈规则	交通法规、车辆动力学、空间边界、泊车约束
博弈奖励	安全高效通行 / 无碰撞精准泊车完成
专家棋谱	仿真自博弈迭代生成时序轨迹
棋类定式	道路基元、泊车基元拓扑结构特征

3.3 博弈状态结构化表征

摒弃高维原始图像输入，采用图结构状态表征定义全局博弈状态，压缩原始高维状态空间，提升拓扑泛化能力。全局状态定义：

$$S = (G_{road}, \text{Ego}_{state}, \text{Traffic}_{state}, \text{Light}_{state})$$

式中：G_{road}：多尺度拓扑属性图（宏观路网、微观停车场统一图结构）；Ego_{state}：自车运动与位姿状态；Traffic_{state}：动态交通流与障碍物状态；Light_{state}：信号灯规则与场地边界约束。

3.4 离散时序驾驶动作空间

基于统一高层走子抽象，构建基础驾驶动作集 + 细粒度泊车动作集双层动作体系，维持范式一致性同时适配不同场景特征。

3.4.1 基础驾驶动作集（核心）

包含 6 类纵向速度意图、7 类横向车道意图，组合得到 42 种合法高层驾驶决策；通过非法动作掩码施加硬性规则约束，高层决策意图最终在车辆执行层平滑转换为连续控制输出。

3.4.2 细粒度泊车动作集（拓展）

针对泊车场景小幅位移、多次迭代调整、大角度转向、分步入库特征，定义细粒度泊车走子适配微观拓扑博弈需求：

- P0：厘米级向前微调；
- P1：厘米级向后微调；
- P2：大角度转向入库动作；
- P3：原地/小范围泊车姿态修正动作。

泊车场景下可配置更深 MCTS 搜索深度、更多仿真采样，适配入库、姿态调整长时序多步决策特征。

3.4.3 统一约束机制与动作层级说明

行驶场景通过掩码禁止闯红灯、实线变道等非法动作；泊车场景屏蔽超出车位边界、压线、占用障碍物区域的动作。

关键说明：本文定义离散时序走子为高层行为决策层抽象，非底层连续运动控制指令。该设计匹配 MCTS 树搜索时序推理逻辑，避免车辆阶跃式不真实离散运动；所有高层走子决策通过连续轨迹插值、模型预测控制 MPC 平滑处理，保证行驶连续性与乘坐舒适性。

3.5 分层多目标奖励函数体系

在分层驾驶奖励基础上增设专用泊车奖励分支，实现行驶、泊车场景差异化优化目标。

3.5.1 行驶奖励（核心）

设计四层加权目标：安全硬约束、交规合规约束、通行效率、乘坐舒适性。

3.5.2 专用泊车奖励（拓展）

1. 位姿精度奖励：车身与车位平行偏差、中心偏移最小化正向奖励；2. 边界安全奖励：与墙体、邻车、立柱维持安全距离正向奖励；3. 泊车完成奖励：无过多重复调整、精准入库的效率奖励；4. 平顺约束奖励：抑制频繁大幅打方向、无效反复挪车行为。

采用分层加权奖励方案用于框架初期验证，权重人工配置；后续工作开展奖励权重敏感性实验，探索逆强化学习 IRL 校准奖励函数，降低人工调参主观性。

4 整体架构与核心算法设计

4.1 五层整体架构概述

1. 路网拓扑生成层：模块化基元组装生成多尺度博弈路网；
2. 拓扑感知编码层：GNN+TopoMoE 模块提取路网图拓扑特征；
3. 拓扑博弈决策层：GNN 特征输入 →MCTS-UCT 多步搜索 → 策略/价值网络输出（核心决策模块）；
4. 策略蒸馏轻量化层：将 MCTS 高质量教师策略蒸馏为轻量化车端规划网络；
5. 仿真实车迁移层：决策层轨迹输出对接实车感知、底层控制模块。

整套框架形成三层递进拓扑学习体系支撑闭环训练管线：几何深度学习构建非欧拓扑表征先验，支撑模块化路网组合泛化；TopoMoE 拓扑混合专家自适应匹配稀疏宏观行驶拓扑、稠密微观泊车拓扑，实现行驶泊车统一编码；拓扑 GRPO 拓扑组相对策略优化稳定 MCTS 自博弈迭代，抑制训练方差、多智能体博弈收敛至极端纳什均衡，最终在结构化路网实现可解释、强泛化、全场景统一的基于学习运动规划。

4.2 拓扑感知编码算法（GNN 拓扑嵌入）

4.2.1 图结构定义

路网拓扑图定义为 $G=(V,E,X_V,X_E)$：V：节点集合（路口、路段端点、车位终端节点）；E：边集合（车道段、通行通道、车位分支）；X_V：节点特征（节点类型、路口优先权、信号灯状态、拥堵占用）；X_E：边特征（道路等级、车道数、限速、虚实线、汇流段属性）。

4.2.2 多尺度图卷积编码

采用图卷积网络聚合局部拓扑子图特征：

$$h_v^{(l)}=\sigma\left(W^{(l)}\cdot \mathrm{AGG}\left(h_v^{(l-1)},\left\{h_u^{(l-1)}|u\in\mathcal{N}(v)\right\}\right)\right)$$

$\mathcal{N}(v)$ 代表节点 v 邻域节点集合：

- 行驶模式：设置大邻域聚合半径，捕捉远距离路网关联；
- 泊车模式：设置小邻域聚合半径，捕捉精细车位边界约束。

最终输出全局拓扑嵌入特征 z_{topo}，作为策略网络、价值网络输入。

为避免大规模全局路网全图编码计算爆炸，模型采用自车中心局部拓扑子图感知策略，仅编码与局部决策相关拓扑区域，保障实时推理。

4.3 AlphaZero 式自博弈核心决策算法

4.3.1 网络结构设计

1. **策略网络 PolicyNet** 输入：拓扑嵌入特征 z_{topo} + 自车动态状态特征输出：全部候选高层驾驶动作先验概率分布 $\pi(a|S)$，结合动作掩码屏蔽非法动作。

2. **价值网络 ValueNet** 输入与策略网络保持一致输出：当前路网博弈状态长期累积价值估计 $v(S)\in[-1,1]$。

4.3.2 改进 MCTS-UCT 选择公式（核心算法公式）

采用树置信上界 UCT 作为 MCTS 节点选择策略，平衡探索与利用：

$$a^*=arg\max_a\left(\frac{Q(s,a)}{N(s,a)}+c_{puct}\cdot\pi(a|S)\cdot\frac{\sqrt{\sum_b N(s,b)}}{1+N(s,a)}\right)$$

式中：$Q(s,a)$：节点 (s,a) 累积奖励；$N(s,a)$：动作访问次数；c_{puct}：探索系数，控制智能体对陌生动作探索程度；$\pi(a|S)$：策略网络输出动作先验概率。

单次 MCTS 搜索分为四步：选择 → 扩展 → 评估 → 反向传播。

4.3.3 单轮自博弈流程算法（动力学约束版本）

算法 1：拓扑博弈自动驾驶单轮自博弈流程输入：路网生成器 RoadGenerator、策略网络 PolicyNet、价值网络 ValueNet、最大轮次 T_{max} 输出：单轮驾驶轨迹数据集 trajectory

1. 初始化：road graph = RoadGenerator.generate topo() // 生成模块化组装路网博弈棋盘
2. current state = RoadBoardState(road graph)，trajectory = []

3. while 当前状态未终止且步数 $< T_{max}$ do

   (a) 步骤 1：GNN 编码器对当前拓扑状态编码 state emb = GNN Encoder(current state)

   (b) 步骤 2：执行 MCTS-UCT 多步前瞻博弈搜索 root node = MCTS Search(root state = current state, state emb, PolicyNet, ValueNet)

   (c) 步骤 3：依据访问频次获取最优高层决策走子 action dist = get action distribution(root node), best action = sample action(action dist)

   (d) 步骤 4：记录当前状态、动作分布作为训练样本 trajectory.append((current state, action dist))

   (e) 步骤 5：在车辆自行车动力学约束下执行走子决策，生成物理可行平滑轨迹，更新车辆位姿与路网博弈状态 current state, step reward = current state.step(best action)

4. 获取本轮最终奖励 final value，使用 final value 标记轨迹内所有样本

5. return trajectory

动力学约束补充说明：智能体输出离散走子仅为高层时序决策意图；动作执行过程严格嵌入车辆自行车动力学约束，保证每条轨迹满足车速、转向角、加加速度等物理边界，杜绝非物理瞬移、阶跃运动，完全满足自动驾驶运动规划物理可行性要求。

4.3.4 网络迭代更新损失函数

$$\mathcal{L}_{policy} = -\sum_{a} \rho_{mcts}(a) \log \pi_{net}(a)$$

$$\mathcal{L}_{value} = \left(v_{net}(S) - v_{final}\right)^2$$

$$\mathcal{L}_{total} = \mathcal{L}_{policy} + \lambda \cdot \mathcal{L}_{value}$$

λ 为价值损失权重系数。

4.3.5 策略蒸馏算法（师生网络）

算法 2：面向轻量化学生网络的 MCTS 教师策略蒸馏输入：MCTS 自博弈生成教师轨迹数据集 D_{teacher} 输出：轻量化车端规划学生网络 StudentNet

1. 初始化 StudentNet（轻量化多层感知机/浅层 Transformer）

2. foreach 教师样本 (state emb, ρ_{mcts}) do

   (a) 梯度下降优化 StudentNet，拟合 MCTS 输出最优决策分布

   (b) 学生网络预测动作分布 π_{student}

   (c) 计算蒸馏损失：$\mathcal{L}_{\text{distill}} = -\rho_{mcts} \log \pi_{\text{student}}$

蒸馏完成后 StudentNet 可直接用于车端实时规划输出。

4.4 分场景动态搜索策略

1. 行驶场景：设置较浅 MCTS 搜索深度，侧重短周期车道交互、路口转向决策；2. 泊车场景：增大 MCTS 搜索深度与仿真采样数，适配多步挪车、入库长时序决策需求；3. 通过调整 MCTS 深度参数 d_{search} 实现两类场景自适应切换。

5 本框架差异化特征

对比主流自动驾驶训练范式，拓扑博弈框架具备多项差异化优势：1. **降低对大规模真实行车数据依赖**：依托规则约束仿真自博弈生成高质量驾驶样本，无需海量真实标注轨迹，适配数据积累、高端算力有限的科研场景；2. **模块化可扩展场景生成**：依托有限拓扑基元组合生成典型路网场景，系统性覆盖各类基础路口、道路拓扑，减少极端场景采样对实车路测的依赖；3. **拓扑驱动决策逻辑可解释**：决策来源于显式路网拓扑、交通法规、MCTS 多步搜索树，而非纯黑盒像素拟合，决策过程可溯源，满足高阶自动驾驶安全可解释需求；4. **跨路况拓扑泛化能力提升**：模型学习路网通用结构特征而非场景专属像素特征，在不同城市布局陌生组装路网自适应性能更优；5. **行驶泊车任务统一建模**：区别于行业独立模型堆叠方案，依托统一图拓扑先验、统一 MCTS 博弈搜索、统一自博弈迭代机制探索原生跨场景决策，旨在算法层面搭建通用智能驾驶决策基础。

与传统基于规则规划核心差异：传统规则运动规划依赖人工设计代价函数；本框架通过 MCTS 实现多步博弈前瞻，依靠自博弈隐式学习多车交互偏好，可自主涌现人工规则难以穷举的复杂社会性驾驶行为（无保护左转、环岛多车汇流、车位争抢交互等），动态交互交通场景适配性更强。

6 四大核心科学假设与分阶段验证方案

框架有效性依托四项核心科学假设，需通过分阶段仿真实验验证优化，明确适用边界与后续改进方向。

6.1 假设 1：高层离散走子可有效表征安全驾驶策略

核心风险：连续车辆控制离散化可能引发轨迹抖动、决策滞后、驾驶性能劣化，无法匹配 MPC 等连续控制方案平顺性。验证方案：固定单路口无动态车流场景，定量对比离散 MCTS 高层决策与连续 MPC 控制的通行效率、轨迹平顺度、启停抖动指标；若存在明显缺陷，可升级为参数化轨迹段走子，平衡离散搜索特性与连续行驶平顺性。

6.2 假设 2：GNN 可在组装拓扑场景间泛化

核心风险：GNN 局部邻域编码无法捕捉多基元组装产生的全局交互行为，单基元训练策略迁移至混合拓扑路网时性能下滑。验证方案：划分独立训练、测试集：仅在独立单道路基元训练，在未见过的混合组装拓扑上测试；观测智能体决策稳定性，验证模型是否学习底层拓扑本质特征而非记忆训练场景。

6.3 假设 3：多智能体自博弈可涌现合理社会性驾驶行为

核心风险：无约束多智能体自博弈易收敛至极端纳什均衡，出现过度保守礼让导致通行效率低下，或激进驾驶频繁碰撞、策略收敛不稳定。验证方案：初期引入规则背景智能体作为行为锚点，约束自博弈策略收敛区间；长期可引入策略种群机制维持驾驶风格多样性，促进涌现贴合真实交通特征的混合博弈策略。

6.4 假设 4：行驶泊车跨尺度拓扑映射一致性

核心风险：百米级宏观行驶拓扑、米级微观泊车拓扑空间密度、节点间距、特征尺度差异巨大，单一统一图编码器难以稳定学习双尺度博弈特征，跨场景迁移性能衰减。验证方案：1. 设计

多尺度消融实验：仅行驶训练、仅泊车训练、行驶泊车混合训练三组对照；2. 引入动态聚合半径多尺度图卷积，自适应调整不同场景感受野；3. 量化跨场景迁移精度、决策稳定性、任务成功率，验证框架尺度泛化能力。

6.5 辅助优化风险说明

针对人工设定奖励权重主观性问题，开展奖励权重敏感性扫描实验，量化不同权重配置影响边界；后续工作探索逆强化学习 IRL，依托少量高质量人类驾驶轨迹推导最优奖励函数，降低人工调参主观性。

7 分阶段实验研发路线图

为实现框架稳定迭代、分阶段验证，制定五阶段最小可行研究路径，从基础能力验证逐步推进至全场景自博弈与仿真实车迁移：1. **基础验证阶段**：搭建固定单路口拓扑博弈棋盘，无动态车流，验证模型可自主学习信号灯通行、车道保持、抵达目的地等基础规则；2. **单智能体交互阶段**：引入 IDM 规则车流、非机动车智能体，依托分层奖励体系验证跟车、减速礼让等基础交互行为；3. **拓扑泛化验证阶段**：启用路网自动生成器搭建多基元组装路网，测试模型在陌生拓扑布局下泛化性能；4. **多智能体自博弈阶段**：部署多个同构决策智能体，验证社会性驾驶策略涌现效果，优化博弈收敛均衡；5. **蒸馏与域迁移阶段**：完成 MCTS 策略蒸馏，对接高保真仿真与实车智能驾驶管线，探索仿真训练策略向真实场景迁移。

8 学术价值与工程应用前景

8.1 学术价值

本文回应当前自动驾驶领域过度依赖大规模数据、黑盒模型拟合的研究趋势，构建一套假设驱动、面向智能驾驶运动规划的拓扑组合博弈框架；为自动驾驶决策模型提供结构化设计、分阶段验证范式，聚焦路网拓扑结构泛化、自博弈驱动交互行为自主涌现。本文补充基于学习的智能驾驶领域拓扑结构建模、自博弈迭代优化研究视角。

依托几何深度学习、TopoMoE 多尺度拓扑编码、拓扑 GRPO 博弈优化三层技术体系，系统性解决拓扑结构化表征、行驶泊车跨尺度统一、多智能体自博弈训练不稳定三大核心挑战，为几何人工智能与自博弈运动规划交叉领域提供全新融合框架。

8.2 工程应用场景

1. **自动驾驶预仿真验证平台**：支撑各类决策算法路网适配性低成本批量验证，降低大规模实车测试成本；2. **复杂交互路口博弈决策模块**：为环岛、无信号路口、匝道汇流等高难度交互场景提供针对性决策支撑；3. **行驶泊车一体化通用决策训练底座**：本框架消除算法层面行驶、泊车领域边界；依托统一拓扑博弈机制、统一 GNN 编码、统一自博弈训练逻辑，支撑高速、城市道路、环岛匝道、室内外停车场全场景决策训练，探索算法原生一体化自动驾驶潜力，对比传统软硬件堆叠方案具备一定参考价值。

8.3 未来研究方向

后续迭代聚焦三大方向：优化连续自适应 MCTS 策略，提升高层离散决策执行平顺性；构建动态拓扑感知机制，实现未知路网实时理解与自适应决策；开源道路基元仿真库，推动轻量化智能驾驶科研生态搭建。

9 结论

本文借鉴 AlphaZero 自博弈范式，提出一套基于拓扑博弈的自动驾驶决策框架，为高数据高算力消耗端到端模仿学习、行驶泊车模块分离设计提供全新替代开发路径。定义 7 类核心道路基元支撑典型行驶场景模块化生成，新增拓展泊车拓扑基元，将拓扑博弈建模能力拓展至微观停车场。依托统一图拓扑表征、统一 MCTS 博弈搜索机制、统一自博弈迭代训练体系，搭建覆盖宏观路网、微观停车场的全场景通用自动驾驶博弈框架。

在几何深度学习、TopoMoE 多尺度专家编码、拓扑 GRPO 拓扑博弈优化支撑下，整套框架形成理论设计、算法管线闭环体系。作为假设驱动预印本框架，本文为通用自动驾驶决策模型分阶段验证、安全迭代、跨场景泛化提供结构化研究基础，为智能驾驶运动规划提供全新拓扑博弈研究视角。

10 参考文献

1. M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," IEEE Signal Process. Mag., vol. 34, no. 4, pp. 18–42, Jul. 2017.

2. W. Chen, X. Han, and B. Wang, "Graph representation learning: A survey," APSIPA Trans. Signal Inf. Process., vol. 9, p. e15, 2020.

3. I. Chami, Z. Ying, C. Ré, and J. Leskovec, "Hyperbolic graph convolutional neural networks," in Adv. Neural Inf. Process. Syst., vol. 32, 2019.

4. T. Shao et al., "Group relative policy optimization," arXiv preprint arXiv:2406.09279, 2024.

5. Y. Cang et al., "Graph-GRPO: Stabilizing multi-agent topology learning via group relative policy optimization," arXiv preprint arXiv:2603.02701, 2026.

6. D. Silver et al., "Mastering the game of Go without human knowledge," Nature, vol. 550, no. 7676, pp. 354–359, Oct. 2017.

7. D. Silver et al., "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," Science, vol. 362, no. 6419, pp. 1140–1144, Dec. 2018.

8. C.-J. Hoel, K. Driggs-Campbell, K. Wolff, L. Laine, and M. J. Kochenderfer, "Combining planning and deep reinforcement learning in tactical decision making for autonomous driving," arXiv preprint arXiv:1905.02680, 2019.

9. M. Cusumano-Towner et al., "Robust autonomy emerges from self-play," arXiv preprint arXiv:2502.03349, 2025.

10. M. Elbanhawi and M. Simic, "Sampling-based robot motion planning: A review," IEEE Access, vol. 2, pp. 56–77, 2014.

11. D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," IEEE Robot. Autom. Mag., vol. 4, no. 1, pp. 23–33, Mar. 1997.

12. A. Faust et al., "PRM-RL: Long-range robotic navigation tasks by combining reinforcement learning and sampling-based planning," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2018, pp. 5113–5120.

13. S. Podgorski et al., "TANGO: Traversability-aware navigation with local metric control for topological goals," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2025, pp. 6789–6795.

14. W. C. Bandara, J. M. J. Valanarasu, and V. M. Patel, "Spin road mapper: Extracting roads from aerial images via spatial and interaction space graph reasoning for autonomous driving," in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2022, pp. 2456–2462.

15. J. M. Álvarez and A. M. López, "Road detection based on illuminant invariance," IEEE Trans. Intell. Transp. Syst., vol. 11, no. 1, pp. 95–105, Mar. 2010.