Topological Game-Based Autonomous Driving: A Road Network Self-Play Decision Framework Inspired by the AlphaZero Paradigm

Lanhaijian
Independent Researcher
Preprint Published on Hongqiao University Tech Academic Site
Global Personal Academic Platform

This paper is a preprint framework proposal, focusing on theoretical modeling and system design. This paper is originally released, and all subsequent updated versions are published on Hongqiao University Tech Academic Site.

1 Abstract

Current mainstream training paradigms for autonomous driving decision-making heavily rely on massive real-world driving data and high-performance computing clusters. Typical approaches include end-to-end visual imitation learning and manually customized scenario-based reinforcement learning, which suffer from poor model interpretability, insufficient coverage of long-tail edge cases, and weak generalization to unseen road layouts. Such data- and computation-intensive development paths are challenging to implement under limited data and computing resources. Against this background, this paper proposes a **topological game-based autonomous driving decision framework inspired by the AlphaZero self-play paradigm**.

The proposed framework decomposes complex urban road networks into a set of defined road primitives, supporting the programmatic generation of modular, resettable road game environments. It reformulates vehicle driving decision-making from a conventional continuous physical control task into a **temporal multi-objective combinatorial decision problem defined on structured road topologies**. The framework abandons brute-force pixel-level data fitting, and adopts graph neural networks to extract structured topological embeddings of road networks. Monte-Carlo Tree Search (MCTS) is employed to implement multi-step lookahead decision reasoning, while a hierarchical multi-objective reward system balances driving safety, traffic compliance, travel efficiency and ride comfort. High-quality driving trajectories are generated through simulation self-play, and policy distillation is adopted to support lightweight deployment on vehicle-end devices.

Furthermore, this framework extends topological game modeling from road driving scenarios to parking scenarios by introducing dedicated parking primitives, fine-grained discrete action sets and pose optimization reward functions. A unified game decision mechanism is constructed to handle cross-scale driving-parking tasks within one network architecture, exploring the potential of unified modeling for full-domain autonomous driving.

This work belongs to the category of **learning-based motion planning frameworks**. It reduces reliance on manually enumerated interactive scenarios and large-scale human-labeled driving data, enabling interactive driving behaviors to emerge from topological structural priors and self-play iterations. Theoretical analysis indicates that the proposed framework compresses high-dimensional state space via structured road priors, and exhibits several desirable characteristics: interpretable decision logic, modularly extensible scenario generation, improved generalization across unseen road topologies, emergent interactive driving behaviors, and unified modeling for driving and parking tasks. This paper provides a hypothesis-driven research framework for the design, staged verification and iterative optimization of autonomous driving decision models, offering an alternative technical path for intelligent driving development under constrained data and computing conditions.

Keywords: Autonomous Driving; Topological Game; Road Network Primitives; Integrated Driving and Parking; Motion Planning; AlphaZero; Monte-Carlo Tree Search; Self-Play Reinforcement Learning; Graph Neural Networks

2 Introduction

2.1 Core Bottlenecks of Existing Autonomous Driving Paradigms

At present, intelligent decision systems for autonomous driving mainly follow two mainstream technical routes: **end-to-end visual imitation learning** and **scenario-based reinforcement learning**. Although both approaches have achieved staged progress in industrial applications, their inherent paradigm limitations restrict further improvement of generalizable autonomous driving capabilities.

The **end-to-end visual imitation learning paradigm**, represented by large-scale fleet-based learning schemes, depends heavily on massive real-world driving trajectories and high-performance computing clusters to fit human driving behaviors from pixel-level sensory data. This paradigm has three prominent limitations. First, it requires large-scale data accumulation and high-end computing resources, which are difficult to replicate for research and industrial teams without equivalent fleet data support. Second, the model operates as a black-box system without explicit topological reasoning logic, making it difficult to trace decision failures and meet the interpretability requirements of high-level autonomous driving safety. Third, such models mainly learn visual appearance features rather than the inherent connectivity of road topologies, leading to poor generalization when encountering unfamiliar urban road layouts.

The **manual scenario-based reinforcement learning paradigm** implemented on simulation platforms such as CARLA and LGSVL trains decision models based on manually constructed scenarios including intersections, highways and urban roads. Its core limitation lies in **limited scenario coverage**: manually designed scenarios cannot systematically cover the diverse topological combinations of real-world road networks.

Long-tail edge cases can only be sampled randomly during road testing, which cannot guarantee the structural robustness of decision models and may lead to overfitting to fixed predefined scenarios.

Traditional **rule-based and sampling-based motion planning methods** (e.g., A*, RRT, Lattice planners) generate deterministic collision-free trajectories via manually designed cost functions and path templates. However, these methods lack the ability to model social driving interactions such as multi-vehicle negotiation, yielding and lane merging, making it difficult to adapt to dynamic game-like interactions in real traffic.

In addition, most industry-level integrated driving-parking solutions are essentially **functional stacking of independent driving and parking modules**. Driving and parking decision systems belong to separate logical frameworks without unified spatial topological modeling at the algorithm level, failing to achieve native generalizable decision-making across these two scenarios.

Fundamentally, existing mainstream paradigms **underutilize the structural properties of road networks**. Urban road networks are not purely unstructured visual scenes but combinatorial systems assembled from a limited set of basic topological units. Conventional methods frame driving as a pure physical motion control problem and ignore road topological priors, resulting in high-dimensional state space explosion, low learning efficiency, poor generalization, and difficulty in unifying macro-scale road traversal and micro-scale parking-space occupancy decision-making.

2.2 Feasibility and Limitations of Migrating the AlphaZero Game Paradigm

AlphaZero achieved remarkable performance in board games through **rule-constrained self-play and emergent optimal strategies**, without relying on large amounts of human expert game records. Its core components include a structured closed game environment, discrete sequential decision actions, policy-value dual-network evaluation, multi-step lookahead via Monte-Carlo Tree Search, and iterative self-play data generation. This paradigm reduces dependence on human expert data and converges to effective strategies through constrained autonomous exploration within defined game rules.

Driving scenarios share partial logical similarities with board games: road networks have relatively fixed topological structures and traffic rules, and driving behaviors can be abstracted as sequential constrained decision-making processes. These characteristics provide the basic feasibility for migrating the AlphaZero self-play idea to autonomous driving decision-making. However, there exist essential differences between board games and real-world driving: board games have fixed rules and relatively clear optimal outcomes, while autonomous driving is a multi-objective balancing task involving safety, efficiency and comfort, and there is no globally unique optimal strategy. This fundamental difference imposes key constraints on the direct transplantation of board game self-play mechanisms, which requires targeted design of reward functions and multi-agent interaction constraints.

Based on this analysis, this paper draws on the core idea of AlphaZero-style self-play rather than directly replicating the board game training pipeline: **constructing**

modular game environments based on road topology primitives, abstracting high-level driving intentions as sequential decision moves, defining game constraints via traffic rules, and generating driving strategies through simulation self-play.

Meanwhile, both driving and parking can be abstracted as **topological space occupancy games with different spatial scales and constraint densities**, which provides a feasible basis for exploring unified cross-scene decision modeling.

2.3 Core Contributions of This Work

This work focuses on **system-level framework design and scenario modeling innovation**, rather than proposing novel fundamental algorithms. The main contributions are summarized as follows:

1. **Topological Game Modeling Framework**: A topology-driven self-play decision framework is constructed, reformulating continuous driving control tasks into sequential combinatorial decision problems defined on modular road topologies, providing an alternative to mainstream data-intensive imitation learning and manually enumerated scenario reinforcement learning.

2. **Modular Road Topology Primitive Design**: Seven typical road primitives are defined to generate extensible urban driving scenarios, together with an extended parking topology primitive, enabling modular generation of macro road networks and micro parking field environments.

3. **Cross-Scale Topological Encoding Design**: A Topology-aware Mixture-of-Experts (TopoMoE) backbone is introduced to adaptively encode sparse macro road topologies and dense micro parking topologies, supporting unified feature extraction across driving and parking scenes.

4. **Algorithm-Native Integrated Driving-Parking Architecture**: Different from industry-style functional stacking of separate driving and parking models, this framework adopts unified graph state representation, unified MCTS-based game search and unified self-play iteration logic to explore native cross-scene decision-making.

5. **Hypothesis-Driven Staged Verification System**: Four core scientific assumptions are proposed regarding the effectiveness of discrete high-level actions, topological generalization of graph encoders, stable emergence of multi-agent interactive behaviors, and cross-scale topology adaptation. Corresponding staged experimental verification schemes are designed to guide subsequent framework iteration and performance evaluation.

3 Related Work

3.1 Geometric Deep Learning for Road Topology Modeling

Geometric deep learning (GDL) supports structure-aware feature learning on graph-structured spatial data, and has been widely applied to road graph representation, traffic flow prediction and vehicle trajectory forecasting in autonomous driving. Graph neural networks (GNNs) can effectively encode lane connectivity and intersection topology, showing better generalization than image-based convolutional neural networks when facing unseen road layouts.

However, most existing GNN applications in autonomous driving focus on **static topological perception and trajectory prediction**. Road topology is treated as a fixed environmental feature rather than a programmable game board for sequential decision-making and multi-agent interaction planning. Few existing works unify macro-scale road topology and micro-scale parking topology within a single graph embedding framework, leading to separated model designs for driving and parking tasks.

Different from these existing perception-oriented GNN pipelines, this work extends topological graph representation from passive environmental feature extraction to **active topological game decision-making**. Road graphs are constructed as generable game environments, and driving and parking behaviors are modeled as sequential topological moves, enabling topological structural priors to support self-play policy optimization.

3.2 Topology-Aware Mixture-of-Experts for Multi-Scale Graph Learning

Topology-aware Mixture-of-Experts (TopoMoE) architectures implement sparse expert routing according to graph structural properties, enabling adaptive feature extraction for heterogeneous topological subgraphs. Compared with fixed-structure GNNs, TopoMoE can dynamically assign different network experts to process graph regions with distinct density and connectivity, improving representation diversity for heterogeneous spatial structures.

In existing autonomous driving and robotics research, TopoMoE is mainly applied to scene classification, graph segmentation and adaptive environmental perception. Few works apply TopoMoE to solve the **cross-scale topological heterogeneity** between sparse road-level traffic scenarios and dense parking-field scenarios, nor are they deployed for long-horizon sequential motion planning tasks.

In this framework, a decision-oriented TopoMoE backbone is customized to unify dual-scale topological modeling. Macro sparse road topologies and micro dense parking topologies are automatically routed to corresponding expert modules, realizing unified encoding for driving and parking without two-branch model stacking, which serves as a key intermediate layer for cross-scene generalization in topological game planning.

3.3 Topological Reinforcement Learning and GRPO Optimization

Traditional reinforcement learning algorithms such as PPO and A2C often suffer from high training variance and unstable multi-agent credit assignment in interactive traffic scenarios. Group Relative Policy Optimization (GRPO) optimizes policy advantages via intra-group trajectory ranking without relying on an independent critic network, effectively reducing computational overhead and improving training stability. Topology-aware GRPO further incorporates graph structural constraints into policy updates, making reinforcement learning aware of spatial adjacency and topological differences.

Nevertheless, existing topological GRPO research is mostly limited to multi-agent communication optimization and simple grid-world control tasks, and has not been systematically integrated with long-horizon Monte-Carlo tree search self-play for autonomous driving motion planning.

In the proposed framework, MCTS generates high-quality multi-step topological game trajectories to serve as teacher supervision signals. Topology-GRPO is integrated into the self-play iteration pipeline to stabilize policy updates and mitigate the problem of multi-agent self-play converging to extreme Nash equilibria. It acts as a policy optimization module rather than replacing the MCTS search mechanism.

3.4 Differences from Existing Works

In summary, existing geometric deep learning, TopoMoE and topological reinforcement learning methods mostly address separate sub-problems including topological perception, multi-scale encoding and policy optimization. Few studies systematically integrate these components into a unified topological game decision framework targeting full-domain driving and parking tasks.

This work constructs a three-layer technical system:

1. **Geometric Deep Learning provides topological structural priors and graph-based state representation**;
2. **TopoMoE realizes adaptive multi-scale unification of driving and parking topological features**;
3. **Topology-GRPO stabilizes MCTS-based self-play iteration and topological game policy optimization**.

Compared with conventional learning-based motion planning methods, this framework transforms autonomous driving from data-intensive imitation or rule-dependent deterministic planning into **topology-structured, self-play-evolving game planning**, aiming to improve decision interpretability, structural generalization and cross-scene universality.

4 Topological Game Modeling System for Autonomous Driving

4.1 Road Network Primitives: Basic Building Blocks of Generable Game Environments

By decomposing typical topological structures of urban roads, highways and rural road networks, seven core driving road primitives are abstracted as the basic building blocks of the proposed generable game environments. Real public road networks can be constructed through the combination, extension and nesting of these primitives:

1. Straight road unit: Basic linear structure for constant-speed car-following and straight traversal;

2. Orthogonal cross intersection unit: Standard four-way urban intersection topology covering straight crossing, conflicting traffic and turning interaction scenarios;
3. T-junction unit: Three-way branching topology adapted to branch merging into main roads and lateral intersection traversal;
4. Roundabout unit: Closed circular topology with multiple radial entrances, representing typical unordered multi-vehicle interaction scenarios;
5. Fishbone strip unit: Trunk-branch strip topology adapted to long and narrow urban roads along rivers and traffic arteries;
6. Radial hub unit: Central-node radial topology corresponding to radial road structures in old urban cores;
7. Branching tree unit: Hierarchical branching topology without closed loops, adapted to suburban, mountainous and rural unstructured road networks.

To expand the cross-scene generalization of the framework, this paper defines an eighth extended primitive: **Parking Lot Topology Unit**, which is designed to model closed, high-density, highly constrained micro-space parking game scenarios:

4.1.1 Parking Lot Topology Unit (Extended General Unit)

Different from macro public road topologies, parking lots belong to **high-density, narrow-channel structured topologies with terminal parking constraints**, which can be incorporated into the unified graph topology system of this framework:

- Topology Definition: Grid topology composed of main access edges, secondary channel edges and parking-space terminal nodes;
- Parking Space Modeling: Vertical, parallel and angled parking spaces are uniformly abstracted as **terminal topological branches with boundary constraints**;
- Field Characteristics: Narrow passages, high pose accuracy requirements and compact traversable space;
- Game Characteristics: Multi-vehicle parking competition, channel passing, multiple adjustment maneuvers and safe distance maintenance, belonging to typical high-density, long-sequence decision tasks;
- Scale Characteristics: Defined as **micro topological game scenes**, forming extensible full-scene coverage together with macro driving topologies, ranging from large-scale road networks to small-scale parking fields.

Essentially, driving road networks and parking lots are **structured topological spaces with different spatial scales and density levels**. The proposed framework adaptively adjusts graph convolution receptive fields to realize unified feature extraction for both macro road networks and micro parking fields, which forms the core basis for exploring algorithm-native integrated driving and parking decision-making.

Based on the above eight primitives, the framework can programmatically generate **macro driving game boards, micro parking game boards, and transition scenario boards connecting driving and parking**, achieving modularly extensible coverage of typical autonomous driving scenarios.

The core idea of this topological game design is not limited to road driving scenarios, but is potentially applicable to other structured, constraint-based traffic space decision problems.

4.2 Core Mapping Relationship of the Game System

Referring to the abstract game logic of AlphaZero-style board games, a unified game mapping system for driving and parking decision-making is established:

AlphaZero Board Game Concept	Unified Definition in Topological Game Autonomous Driving
Fixed Game Board	Multi-scale topological game fields programmatically generated by road/parking p
Discrete Moves	Macro high-level driving decision moves + fine-grained sequential parking adjustm
Game Rules	Traffic regulations, vehicle dynamics, spatial boundaries and parking constraints
Game Reward	Safe and efficient driving / accurate collision-free parking completion
Expert Game Records	Sequential trajectories generated by simulation self-play iterations
Board Game Patterns	Topological structural features of road primitives and parking primitives

4.3 Structured Representation of Game States

Abandoning high-dimensional raw image input, **graph-structured state representation** is adopted to define the global game state, compressing the original high-dimensional state space and facilitating topological generalization. The global state is defined as:

$$S = (G_{road}, Ego_{state}, Traffic_{state}, Light_{state})$$

where:

- G_{road} denotes the multi-scale topological attribute graph (unified graph structure for macro road networks and micro parking lots);
- Ego_{state} denotes the ego-vehicle motion and pose state;
- $Traffic_{state}$ denotes the dynamic traffic and obstacle state;
- $Light_{state}$ denotes traffic signal rules and field boundary constraints.

4.4 Discrete Sequential Driving Action Space

Based on the unified high-level move abstraction, a two-layer action system of **basic driving action set + fine-grained parking action set** is constructed to maintain paradigm consistency while adapting to the characteristics of different scenarios.

4.4.1 Basic Driving Action Set (Core)

It includes 6 types of longitudinal speed intentions and 7 types of lateral lane intentions, which are combined into 42 legal high-level driving decisions. An illegal action mask is applied to enforce hard rule constraints, and these high-level decision intentions are finally converted into smooth continuous control outputs at the vehicle execution layer.

4.4.2 Fine-Grained Parking Action Set (Extension)

Aiming at the characteristics of parking scenarios including **small displacement, multiple iterative adjustments, large steering angles and sequential entry maneuvers**, a set of fine-grained parking moves are defined to adapt to micro topological game requirements:

- P0: Centimeter-level forward fine-tuning
- P1: Centimeter-level backward fine-tuning
- P2: Large-angle steering entry action
- P3: In-situ / small-range parking pose adjustment action

In parking scenarios, MCTS can be configured with deeper search depth and more simulation samples to adapt to the long-sequence multi-step decision characteristics of parking entry and pose adjustment.

4.4.3 Unified Constraint Mechanism and Explanation of Action Hierarchy

For driving scenarios, illegal actions such as red-light intersection entry and solid-line lane changes are prohibited. For parking scenarios, actions exceeding parking boundaries, crossing parking lines and occupying obstacle areas are blocked via an Action Mask.

Key Explanation: The discrete sequential moves defined in this study are **abstractions at the high-level behavioral decision layer**, rather than direct low-level continuous motion control commands. This design matches the sequential reasoning logic of MCTS tree search, avoiding unrealistic step-like discrete vehicle motion. All high-level move decisions are smoothed via continuous trajectory interpolation and model predictive control (MPC) to ensure driving continuity and ride comfort.

4.5 Hierarchical Multi-Objective Reward Function System

On the basis of hierarchical driving reward design, a dedicated parking reward branch is added to realize differentiated optimization objectives for driving and parking scenarios.

4.5.1 Driving Rewards (Core)

Four-level weighted objectives are designed, including safety hard constraints, traffic compliance constraints, traffic efficiency and ride comfort.

4.5.2 Dedicated Parking Rewards (Extension)

1. **Pose Accuracy Reward**: Positive reward for minimizing parallelism deviation and center offset between the vehicle body and the parking space;
2. **Boundary Safety Reward**: Positive reward for maintaining safe distances from walls, adjacent vehicles and structural pillars;
3. **Parking Success Reward**: Efficiency reward for accurate parking completion without excessive repeated adjustments;
4. **Smoothness Constraint Reward**: Suppress frequent large steering wheel oscillations and ineffective parking adjustment behaviors.

A hierarchical weighted reward scheme is adopted for initial framework verification with manually configured weights. Future work will conduct **reward weight sensitivity analysis experiments** and explore **Inverse Reinforcement Learning (IRL)** to calibrate reward functions and reduce the subjectivity of manual parameter tuning.

5 Overall Architecture and Core Algorithm Design

5.1 Overview of the Five-Layer Overall Architecture

1. Road Network Topology Generation Layer: Generate multi-scale game road networks via modular primitive assembly;
2. Topology-Aware Encoding Layer: Extract topological features of road network graphs via GNN and TopoMoE modules;
3. Topological Game Decision Layer: **GNN feature input → MCTS-UCT multi-step search → Policy/Value Network output** (core decision module);
4. Policy Distillation Lightweight Layer: Distill high-quality MCTS teacher policies into lightweight vehicle-end planning networks;
5. Sim2Real Simulation Transfer Layer: Output decision-layer trajectories to connect with the perception and low-level control modules of real vehicles.

The overall framework forms a closed-loop training pipeline supported by a three-layer progressive topological learning system: **Geometric Deep Learning** constructs non-Euclidean topological representation priors to support combinatorial generalization of modular road networks; **TopoMoE Topological Mixture of Experts** adaptively matches sparse macro driving topologies and dense micro parking topologies to realize unified driving-parking encoding; **Topology GRPO Topological Group Relative Policy Optimization** stabilizes MCTS self-play iteration, suppresses training variance and extreme Nash equilibrium convergence in multi-agent game interactions, and finally supports interpretable, generalizable and full-domain unified learning-based motion planning on structured road networks.

5.2 Topology-Aware Encoding Algorithm (GNN Topological Embedding)

5.2.1 Graph Structure Definition

The road network topology graph is defined as $G = (V, E, \mathcal{X}_V, \mathcal{X}_E)$

- V: Node set (intersections, road segment endpoints, parking-space terminal nodes);
- E: Edge set (lane segments, access channels, parking-space branches);
- $\mathcal{X}_V$: Node features (node type, intersection priority, traffic light status, congestion occupancy);
- $\mathcal{X}_E$: Edge features (road grade, lane count, speed limit, solid/dashed lines, merging segment attributes).

5.2.2 Multi-Scale Graph Convolutional Encoding

Graph convolutional networks are adopted to aggregate local topological subgraph features:

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

where $\mathcal{N}(v)$ denotes the neighbor nodes of node v;

- Driving mode: Set a large neighborhood aggregation radius to capture long-distance road network correlations;
- Parking mode: Set a small neighborhood aggregation radius to capture refined parking boundary constraints.

The final output is the global topological embedding feature z_{topo}, which serves as the input of the policy network and value network.

To avoid computational explosion caused by full-graph encoding of large-scale global road networks, the model adopts a **local topological subgraph perception strategy centered on the ego vehicle**, only encoding the topological regions relevant to local decision-making to support real-time inference.

5.3 AlphaZero-Style Self-Play Core Decision Algorithm

5.3.1 Network Structure Design

1. **PolicyNet**
   Input: Topological embedding feature z_{topo} + ego-vehicle dynamic state features
   Output: Prior probability distribution $\pi(a|S)$ of all candidate high-level driving actions, combined with Action Mask to block illegal actions.

2. **ValueNet**
   Input: Consistent with the policy network
   Output: Long-term cumulative value estimation $v(S) \in [-1, 1]$ of the current road network game state.

5.3.2 Improved MCTS-UCT Selection Formula (Core Algorithm Formula)

The Upper Confidence Bound for Trees (UCT) is adopted as the MCTS node selection strategy to balance exploration and exploitation:

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

where:

- $Q(s,a)$: Cumulative reward of node (s,a);
- $N(s,a)$: Action visit count;
- c_{puct}: Exploration coefficient controlling the agent's exploration degree of unfamiliar actions;
- $\pi(a|S)$: Prior action probability output by the policy network.

A single MCTS search consists of four steps: **Select** $\rightarrow$ **Expand** $\rightarrow$ **Evaluate** $\rightarrow$ **Backpropagate**.

5.3.3 Single Episode Self-Play Flow Algorithm (Dynamics-Constrained Version)

Algorithm 1: Single Episode Self-Play Flow for Topological Game Autonomous Driving

```
Input: RoadGenerator, PolicyNet, ValueNet, Maximum episode steps T_max
Output: Single episode driving trajectory dataset trajectory
Initialize: road_graph = RoadGenerator.generate_topo() // Generate
    modular assembled road game board
current_state = RoadBoardState(road_graph) trajectory = [] while
 current_state is not terminated and steps < T_max do
    // Step 1:  Encode current topological state via
        GNN
    state_emb = GNN_Encoder(current_state) // Step 2:
        Execute MCTS-UCT multi-step lookahead game
        search
    root_node = MCTS_Search(root_state =
        current_state, state_emb, PolicyNet, ValueNet) // Step 3:
        Obtain optimal high-level decision move
        according to visit frequency
    action_dist = get_action_distribution(root_node)
        best_action = sample_action(action_dist) // Step 4:
        Record current state and action distribution
        as training samples
    trajectory.append((current_state, action_dist)) // Step 5:
        Execute move decisions constrained by vehicle
        bicycle dynamics
    // Generate physically feasible smooth
        trajectories and update vehicle poses and road
        game states
    current_state, step_reward = current_state.step(best_action)
Obtain the final episode reward final_value
Label each sample in trajectory with the target value final_value
return trajectory
```

Supplementary Explanation of Dynamic Constraints: The discrete moves output by the agent are only high-level sequential decision intentions. The action execution process strictly embeds **vehicle bicycle dynamic constraints** to ensure that each trajectory satisfies physical boundaries such as vehicle speed, steering angle and jerk, avoiding non-physical teleportation and step-like motion, which fully meets the physical feasibility requirements of autonomous driving motion planning.

5.3.4 Network Iterative Update Loss Functions

$$\mathcal{L}_{policy} = -\sum_{a} \rho_{mcts}(a) \log \pi_{net}(a) \tag{1}$$

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

$$\mathcal{L}_{total} = \mathcal{L}_{policy} + \lambda \cdot \mathcal{L}_{value} \tag{3}$$

where λ denotes the weight coefficient of value loss.

5.3.5 Policy Distillation Algorithm (Teacher-Student Network)

Algorithm 2: MCTS Teacher Policy Distillation for Lightweight Student Network

Input: MCTS self-play generated teacher trajectory dataset $D_{teacher}$
Output: Lightweight student planning network $StudentNet$
Initialize $StudentNet$ (lightweight MLP / shallow Transformer)
foreach *teacher sample* $(state\_emb, \rho_{mcts})$ **do**
    Student network predicts action distribution $\pi_{student}$
    Calculate distillation loss: $L_{distill} = -\sum \rho_{mcts} \log \pi_{student}$
Optimize $StudentNet$ via gradient descent to fit the optimal decision distribution from MCTS
After distillation, $StudentNet$ can be directly used for real-time vehicle-end planning output

5.4 Dynamic Search Strategy for Different Scenarios

1. **Driving Scenarios**: Set a relatively shallow MCTS search depth, focusing on short-term lane-level interactions and intersection turning decisions;

2. **Parking Scenarios**: Increase MCTS search depth and simulation counts to adapt to the long-sequence decision requirements of multi-step parking adjustments and sequential parking entry;

3. Adaptive switching of both scenarios can be realized by adjusting the MCTS search depth parameter d_{search}.

6 Differentiated Characteristics of the Proposed Framework

Compared with mainstream autonomous driving training paradigms, the proposed topological game framework has several differentiated characteristics:

1. **Reduced Dependence on Large-Scale Real-World Driving Data**
High-quality driving samples are generated through rule-constrained simulation self-play rather than relying on massive real-world labeled trajectories, which

is suitable for research scenarios with limited data accumulation and high-end computing resources.

2. **Modularly Extensible Scenario Generation**
   Typical road network scenes are generated via combinations of finite topological primitives, supporting systematic coverage of various basic intersection and road topologies and reducing reliance on random real-world testing for edge case sampling.

3. **Interpretable Topology-Driven Decision Logic**
   Decisions are derived from explicit road topologies, traffic rules and MCTS multi-step search trees rather than pure black-box pixel fitting. The decision process is traceable and supports safety reasoning, which is beneficial for meeting the interpretability requirements of high-level autonomous driving.

4. **Improved Topological Generalization Across Road Conditions**
   The model learns common structural features of road networks rather than scene-specific pixel features, which helps improve adaptive performance on unseen assembled road networks in different urban layouts.

5. **Unified Modeling for Driving and Parking Tasks**
   Different from the mainstream industry scheme of stacking separate driving and parking models, this framework explores native cross-scene decision-making via **unified graph topological priors, unified MCTS game search and unified self-play iteration mechanisms**, aiming to build a general intelligent driving decision foundation at the algorithm level.

Core Difference from Traditional Rule-Based Planning: Different from traditional rule-based motion planning relying on manually designed cost functions, this framework realizes multi-step game foresight via MCTS and implicitly learns multi-agent interaction preferences through self-play. It can potentially emerge complex social driving behaviors that are difficult to exhaustively define via handcrafted rules, such as unprotected left turns, multi-vehicle roundabout merging and parking-space competition interactions, showing better adaptability to dynamic interactive traffic scenarios.

7 Core Scientific Assumptions and Staged Verification Scheme

The effectiveness of the proposed framework relies on four core scientific assumptions, which need to be verified and optimized in staged simulation experiments to clarify the applicable boundary and subsequent improvement directions:

7.1 Assumption 1: Discrete High-Level Moves Can Effectively Represent Safe Driving Strategies

Core Risk: Discretization of continuous vehicle control may lead to trajectory jitter, decision lag and suboptimal driving performance, failing to match the ride smoothness of continuous control methods such as Model Predictive Control (MPC).
Verification Scheme: Fix a single intersection scenario without dynamic traffic flow, and quantitatively compare driving efficiency, trajectory smoothness and start-stop jitter indicators between discrete MCTS high-level decisions and continuous MPC control. If obvious defects exist, the framework can be upgraded to **parameterized trajectory segment moves** to balance discrete search characteristics and continuous driving smoothness.

7.2 Assumption 2: GNN Can Generalize Across Assembled Topological Scenes

Core Risk: Local neighborhood encoding of GNNs may fail to capture global interactive behaviors emerging from multi-primitive assembly, resulting in poor transfer performance when applying strategies learned from single primitives to mixed topology road networks.
Verification Scheme: Adopt separated training and test sets: train only on independent single road primitives, and test on unseen assembled mixed topologies. Observe the stability of agent decisions to verify whether the model learns essential topological features rather than memorizing specific training scenes.

7.3 Assumption 3: Multi-Agent Self-Play Can Emerge Reasonable Social Driving Behaviors

Core Risk: Unconstrained multi-agent self-play may converge to extreme Nash equilibria, leading to over-conservative yielding with low traffic efficiency or aggressive driving with frequent collisions and unstable strategy convergence.
Verification Scheme: Adopt rule-based background agents as behavioral anchors in the early stage to constrain the convergence range of self-play strategies. In the long term, a strategy population mechanism can be introduced to maintain diverse driving styles and promote the emergence of mixed game strategies consistent with real-world driving characteristics.

7.4 Assumption 4: Cross-Scale Topology Mapping Consistency Between Driving and Parking

Core Risk: Large-scale driving topologies (hundreds of meters) and small-scale parking topologies (meters) have large differences in spatial density, node spacing and feature scales. A single unified graph encoder may struggle to stably learn dual-scale game features, leading to performance degradation during cross-scene transfer.
Verification Scheme:

1. Design **multi-scale ablation experiments**: three groups including driving-only training, parking-only training and mixed driving-parking training;
2. Introduce **dynamic aggregation radius multi-scale graph convolution** to adaptively adjust the receptive field for different scenarios;
3. Quantify cross-scene transfer accuracy, decision stability and task success rate to verify the scale generalization capability of the framework.

7.5 Auxiliary Optimization Risks

Aiming at the subjectivity of manually set reward weights, conduct **reward weight sensitivity scanning experiments** to quantify the influence boundary of different weight configurations. Future work will explore inverse reinforcement learning (IRL) to infer optimized reward functions from a small amount of high-quality human driving trajectories and reduce the subjectivity of manual parameter tuning.

8 Phased Experimental R&D Roadmap

To realize steady iteration and staged verification of the framework, a five-stage minimum feasible research path is formulated, progressing from basic capability verification to full-domain self-play and simulation-to-real transfer:

1. **Basic Verification Stage**: Build a fixed single-intersection topological game board without dynamic traffic flow, verify that the model can autonomously learn basic rules such as signalized traversal, lane keeping and destination arrival;
2. **Single-Agent Interaction Stage**: Introduce IDM rule-based traffic flows and non-motor vehicle agents, verify basic interactive behaviors such as car-following, deceleration and yielding under the hierarchical reward system;
3. **Topological Generalization Verification Stage**: Enable the automatic road network generator to construct road networks assembled from multiple primitives, test the generalization ability of the model on unseen topological layouts;
4. **Multi-Agent Self-Play Stage**: Deploy multiple homogeneous decision agents, verify the emergence effect of social driving strategies and optimize the balance of game convergence;
5. **Distillation and Domain Transfer Stage**: Complete MCTS policy distillation, connect high-fidelity simulation and real-vehicle intelligent driving pipelines, and explore the transfer of simulation-trained strategies to real-world scenarios.

9 Academic Value and Application Prospects

9.1 Academic Value

This work responds to the current research tendency of over-reliance on large-scale data and black-box model fitting in autonomous driving, and constructs a **hypothesis-driven topological combinatorial game framework for intelligent driving motion planning**. It provides a structured design and staged verification paradigm for autonomous driving decision models, focusing on the structural generalization of road topologies and self-play-based interactive behavior emergence. This work supplements the research perspective of topological structure modeling and self-play iterative optimization in the field of learning-based intelligent driving.

Relying on the three-layer technical system of **Geometric Deep Learning, Topo-MoE Multi-Scale Topological Encoding, Topology GRPO Game Optimization**, this work systematically addresses three core challenges: topological structured representation, cross-scale unification of driving and parking scenes, and unstable multi-agent self-play training. It provides a novel fusion framework for the intersection of geometric artificial intelligence and self-play-based motion planning.

9.2 Engineering Application Scenarios

1. **Pre-simulation Verification Platform for Autonomous Driving**: Support low-cost batch verification of road network adaptability for various decision algorithms, helping reduce the cost of large-scale real-vehicle testing;

2. **Game Decision Module for Complex Interactive Intersections**: Provide targeted decision support for high-difficulty interactive scenarios such as roundabouts, unsignalized intersections and ramp merging;

3. **General Decision Training Base for Integrated Driving and Parking**
   This framework explores the elimination of algorithm-level boundaries between driving domains and parking domains. Relying on unified topological game mechanisms, unified GNN encoding and unified self-play training logic, it can support full-scenario decision training for highways, urban roads, roundabout ramps, indoor and outdoor parking lots, exploring the potential of algorithm-native integrated driving and parking autonomous driving, which has certain reference value compared with traditional hardware-software stacking schemes.

9.3 Future Research Directions

Future work will focus on three iterative directions: optimizing continuous-adaptive MCTS strategies to improve the smoothness of high-level discrete decision execution; constructing dynamic topology perception mechanisms to realize real-time understanding and adaptive decision-making for unknown road networks; building an open-source road primitive simulation library to promote the construction of lightweight intelligent driving research ecosystems.

10 Conclusion

This paper proposes a **topological game-based autonomous driving decision framework inspired by the AlphaZero self-play paradigm**, providing an alternative development path beyond data-computing intensive end-to-end imitation learning and separated driving-parking module design. Seven core road primitives are defined to support the modular generation of typical driving scenarios, and an extended parking topology primitive is added to expand the topological game modeling capability to micro parking fields. Based on unified graph topological representation, unified MCTS game search mechanism and unified self-play iterative training system, this paper constructs a full-domain general autonomous driving game framework covering **macro road networks and micro parking fields**.

Supported by geometric deep learning, TopoMoE multi-scale expert encoding and Topology GRPO topological game optimization, this framework forms a closed-loop system of theoretical design and algorithm pipeline. As a hypothesis-driven preprint framework, it provides a structured research basis for the staged verification, safety iteration and cross-scene generalization of general autonomous driving decision models, and offers a new topological game-based research perspective for intelligent driving motion planning.

References

[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.