Proactive and Safe Trajectory Planning via a Reinforcement Learning–Optimization Hybrid

Yujia Lu and Chong Wei

***Abstract*— Ensuring safe and reliable trajectory planning for autonomous driving remains a critical challenge. This paper proposes a conceptual learning-based hybrid framework for proactive trajectory planning in dynamic and uncertain traffic environments. The framework integrates deep reinforcement learning for adaptive decision-making and velocity planning, complemented by optimization-based components to ensure feasibility and safety. Building on our team's prior research experience, future work will focus on refining reward design and conducting comprehensive evaluations.**

I. Problem Statement and Motivation

In autonomous vehicle trajectory planning, one major source of safety risk is the uncertainty in the future motions of surrounding traffic participants, especially in situations where surrounding agents may exhibit sudden and hazardous behaviors, which, although not frequent, pose severe safety risks. To address this, this paper explores a research direction that leverages proactive trajectory planning to enhance safety in emergency obstacle avoidance scenarios.

A. Proactive Obstacle Avoidance for Trajectory Planning

Reactive obstacle avoidance strategies typically involve emergency braking or abrupt steering maneuvers to change lanes near an obstacle. Such passive responses often compromise ride comfort and, in extreme cases, may even result in severe collisions. In contrast, proactive trajectory planning enables autonomous vehicles to take preventive measures when encountering unexpected events—for example, slowing down or switching to a clear lane in advance. Reinforcement learning (RL), with its predictive and adaptive capabilities, has emerged as a promising approach to support such proactive driving strategies.

Autonomous driving scenarios are highly dynamic and uncertain, requiring vehicles to continuously interact with complex traffic environments to learn optimal policies. RL, centered on interaction and feedback, is particularly suited to this setting. For instance, [1] employs an LSTM-TD3 network for lane-change trajectory planning, improving overall performance by balancing driving efficiency and fuel consumption. Experiments conducted on the CARLA platform show that this approach effectively reduces lane-change time while lowering fuel usage. Similarly, [2] presents an integrated method using deep RL to plan target lanes for autonomous vehicles, with numerical results demonstrating the ability to accurately perform advanced driving behaviors such as overtaking and maintaining desired speed.

Compared with imitation learning, RL offers greater adaptability in proactive trajectory planning. Imitation learning relies on expert demonstrations, and its performance is constrained by the coverage of the dataset, making it challenging to handle unforeseen scenarios. By contrast, RL continuously explores and optimizes policies through interaction with the environment, allowing it to learn effective strategies even in previously unseen situations. This capability makes RL particularly suitable for rare but high-risk emergency avoidance scenarios in autonomous driving.

B. Learning-Based Hybrid Planning Frameworks

Despite its advantages, RL is not without limitations. Due to the black-box nature of deep neural networks, learning-based methods may introduce uncertainty into trajectory planning, which can pose non-negligible safety risks. A promising solution is to design hybrid frameworks that combine learning-based adaptability with the verifiability and interpretability of optimization- or rule-based approaches. For example, in [3], the RL agent does not directly output motion parameters such as velocity or angular velocity. Instead, it learns the input parameters of a classical motion planner, thereby incorporating the safety and interpretability of traditional methods into the system. In [4], for a left-turn scenario at an unsignalized intersection, the system falls back on a Monte Carlo tree search combined with MPC if the agent-generated candidate trajectories fail a safety check. Likewise, a study from Tsinghua University [5] employs large language models to automatically generate driving strategies, which are then implemented as rule-based decision trees, ensuring executability, interpretability, and modifiability.

However, hybrid frameworks also introduce new challenges. Optimization modeling in complex traffic scenarios often results in highly nonlinear formulations [6,7], reducing computational efficiency. Careful modeling and efficient approximations are therefore essential. Our previous work [8] proposes a MIQP-based approach that introduces only a small number of binary variables, maintaining obstacle avoidance reliability while significantly reducing computational overhead. Building on [9], this method further applies linear approximations to model vehicle geometry, decreasing the number of binary variables. Experiments with real-world data demonstrate that this approach efficiently generates safe lane-change trajectories even in extreme emergency situa-

Y. Lu is the corresponding author and is with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China (e-mail: 23111256@bjtu.edu.cn).

C. Wei is with the School of Mechanical and Electrical Engineering, Hainan University, Hainan 570228, China (e-mail: weichong@hainanu.edu.cn).

tions while ensuring smooth longitudinal velocity changes and highly smooth lateral maneuvers.

II. Conceptual Framework and Modeling

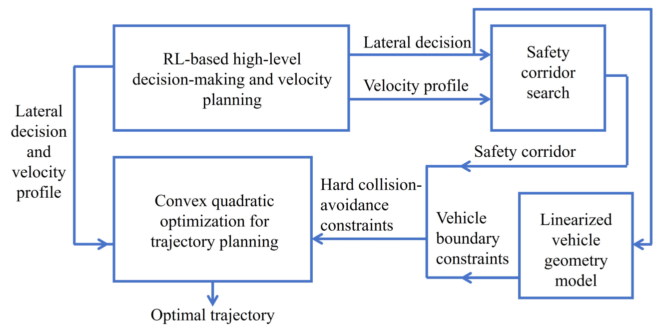

Fig. 1. A Learning-based Hybrid Framework

Motivated by the need for proactive and reliable trajectory planning in dynamic traffic conditions, this paper proposes a learning-based hybrid framework for autonomous vehicle trajectory planning, as illustrated in Fig. 1. First, the perception module provides the ego-vehicle's states and surrounding environmental information. Based on these inputs, an RL-trained agent performs high-level decision-making and velocity planning. Subsequently, a fast heuristic method searches for a safe corridor based on the decision and velocity profile, defining feasible drivable boundaries for obstacle avoidance. Meanwhile, a linearized vehicle geometry model computes the ego-vehicle's geometric boundary. These two types of boundaries collectively serve as hard constraints in the optimization model. Finally, a convex optimization problem generates the ego-vehicle's optimal trajectory while balancing multiple objectives, including safety, comfort, kinematic feasibility, and compliance with traffic regulations.

The deep RL task for decision and velocity planning is formulated as follows. The state space comprises three components: (i) the ego-vehicle's current and historical motion states, including position coordinates, lateral/longitudinal velocity, and acceleration; (ii) surrounding vehicles' current and historical motion states, capturing interactive behaviors and traffic flow dynamics; and (iii) environmental features, such as lane markings and static obstacles. The action space combines high-level driving decisions (e.g., left turn, straight, right turn) with longitudinal control variables (acceleration and velocity), allowing the agent to select appropriate driving intentions while generating a consistent velocity profile.

Designing the reward function remains a critical hurdle in applying deep RL to autonomous vehicle decision-making and trajectory planning. One emerging approach is Inverse Reinforcement Learning (IRL), which aims to infer the underlying reward function from expert demonstrations. For instance, [10] employs multi-agent adversarial IRL to recover reward functions for both the ego vehicle and following vehicles before performing trajectory planning and prediction. This provides planners and predictors with a principled evaluation metric and controllable generation capability. Similarly, [11] proposes a learning-from-demonstration approach that balances behavior cloning and RL exploration, yielding a trajectory planning method tailored for highly constrained environments.

Another common approach involves carefully crafted, experience-driven reward functions. This paper adopts this strategy due to its flexibility in explicitly prioritizing safety, efficiency, and comfort while allowing fine-tuned control over multiple objectives in complex traffic scenarios. The reward function integrates multiple objectives: severe penalties for collisions ensure safety, maintaining desired velocity promotes efficiency, alignment with surrounding vehicle speeds fosters traffic harmony, and smooth lateral and longitudinal accelerations and jerks enhance driving comfort. In studies with predefined trajectory goals, reward functions may also incorporate execution completeness [1]. Additionally, adherence to traffic rules, a vital objective, poses difficulties due to significant variations across driving scenarios, making it hard to encapsulate within a single reward function. Therefore, achieving robust and broadly applicable trajectory planning may necessitate integrating rule-based methods alongside the reward function, adding complexity to the design process.

Policy learning is conducted using the actor–critic-based Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm. The agent continuously updates its policy through interaction with the environment, which consists of two components. First, the downstream optimization module transforms the agent's outputs into executable trajectories, ensuring physical feasibility and constraint satisfaction. Second, a SUMO simulation platform provides a dynamic traffic environment, allowing the agent to explore and learn in multi-vehicle interaction scenarios. Once trained, the policy network can directly output optimal high-level decisions and velocity plans for a given state, achieving end-to-end adaptive driving capabilities.

The remaining components of the framework—including safety corridor search, linearized vehicle geometry modeling, and the optimization module—are further refined based on [8], retaining the same modeling principles. As these components are not the primary focus of this study, they are only briefly referenced.

III. Conclusion

This paper focuses on the safety challenges posed by unpredictable and potentially hazardous behaviors in dynamic traffic scenarios, and proposes a learning-based hybrid framework for proactive and reliable trajectory planning. The framework leverages deep reinforcement learning to achieve proactive decision-making and velocity planning, while incorporating an optimization-based trajectory refinement module to ensure feasibility and safety. One of the key challenges is designing an effective reward function that balances multiple objectives, enabling the agent to avoid collisions while maintaining efficiency and comfort. Future work will focus primarily on this challenge.

REFERENCES

[1] S. Jing, Y. Feng, F. Hui, et al., "Efficient and Eco Lane-Changing Trajectory Planning for Connected and Automated Vehicles: Deep Reinforcement Learning-Based Method," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 7, pp. 9882–9892, 2025.

[2] S. Li, C. Wei, and Y. Wang, "Combining Decision Making and Trajectory Planning for Lane Changing Using Deep Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16110–16136, 2022.

[3] Z. Xu, G. Dhamankar, A. Nair, et al., "APPLR: Adaptive Planner Parameter Learning from Reinforcement," pp. 6086–6092.

[4] L. Zhang, S. Cheng, Z. Wang, et al., "A Safety-Enhanced Reinforcement Learning-Based Decision-Making and Motion Planning Method for Left-Turning at Unsignalized Intersections for Automated Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 11, pp. 16375–16388, 2024.

[5] F. Zeng, S. Wang, C. Zhu, et al., "ADRD: LLM-Driven Autonomous Driving Based on Rule-based Decision Systems," arXiv:2506.14299, June 2025.

[6] J. Fan, N. Murgovski, J. Liang, et al., "Exact Obstacle Avoidance for Autonomous Vehicles in Polygonal Domains," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 10, pp. 5964–5976, 2024.

[7] L. Schafer, S. Manzinger, and M. Althoff, "Computation of Solution Spaces for Optimization-Based Trajectory Planning," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 216–231, 2023.

[8] Y. Lu, C. Wei, and L. Ma, "Reliable and Real-Time Highway Trajectory Planning via Hybrid Learning-Optimization Frameworks," arXiv:2508.04436, August 2025.

[9] Y. Wang, C. Wei, S. Li, et al., "A Convex Trajectory Planning Method for Autonomous Vehicles Considering Kinematic Feasibility and Bi-State Obstacles Avoidance Effectiveness," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 7, pp. 9575–9590, July 2024.

[10] K. Chen, Y. Luo, M. Zhu, et al., "Human-Like Interactive Lane-Change Modeling Based on Reward-Guided Diffusive Predictor and Planner," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 3, pp. 3903–3916, 2025.

[11] W. Lu, L. Chen, Y. Wang, et al., "Demonstration Data-Driven Parameter Adjustment for Trajectory Planning in Highly Constrained Environments," *IEEE Robotics and Automation Letters*, vol. 9, no. 12, pp. 11641–11648, 2024.