Hierarchical Reinforcement Learning with Spatial-Temporal Attention for Ramp Cut-In/Cut-Out

Pao-Kai Wang, Yu-Chen Lin *, Bo-Yu Wei, Wen-De Xiao
Department of Automatic Control Engineering, Feng Chia University, Taichung 40724, Taiwan, R.O.C
* E-mail(s): yuchlin@fcu.edu.tw (Y. C. Lin).

***Abstract*—Turbulence in highways or expressways often arises from vehicles merging onto the main road from entrance ramps (cut-in) or exiting onto exit ramps (cut-out). These maneuvers, dictated by road geometry, significantly disrupt traffic flow, leading to speed variations, turbulence, and reduced capacity. To address these challenges, this paper proposes a spatial-temporal attention-based framework for Cut-in/Cut-out lane-changing strategies. The framework integrates graph attention networks (GAT), safety potential fields, and long short-term memory (LSTM) networks to extract dynamic traffic interactions and risk levels as spatiotemporal features. Built on these features, a hierarchical reinforcement learning (HRL) structure is adopted, dividing the lane-changing task into high-level decision-making, which determines the timing and feasibility of lane changes, and low-level trajectory planning, which generates smooth and collision-free paths. Prioritized experience replay is employed to improve learning efficiency, while a multi-level reward function balances lane-changing success, safety, and efficiency. Experiments conducted on the CARLA simulation platform validate the framework's effectiveness in optimizing lane-changing strategies, demonstrating significant improvements in traffic safety and operational efficiency.**

***Keywords*—cut-in, cut-out, spatial-temporal, safety potential field, hierarchical reinforcement learning, CARLA**

I. Introduction

Autonomous driving technology is rapidly advancing and is poised to bring significant changes to transportation. Through advanced sensors, cameras, and algorithms, autonomous driving systems enable vehicles to operate without human intervention, thereby enhancing road safety, reducing traffic congestion, and improving fuel efficiency [1]. The integration of deep learning and computer vision has further driven autonomous navigation and accident mitigation, positioning autonomous driving technology as a cornerstone of future intelligent transportation systems [2]. However, the implementation of this technology still faces challenges.

In autonomous driving systems, the management of cut-in and cut-out maneuvers, particularly during ramp merging and exiting, is crucial for enhancing road safety and vehicle control strategies. Unlike regular lane changes, ramp maneuvers involve unique challenges such as high-speed differentials, limited merging spaces, and the need to respond to match the speed of traffic while navigating a confined space, making it difficult to merge smoothly into the traffic flow. Exit ramps (cut-out) similarly present challenges, requiring vehicles to decelerate while avoiding interference with ongoing traffic, all within a limited time frame. Necessitating targeted research to improve the safety and responsiveness of autonomous driving systems [3], hierarchical reinforcement learning (HRL) is a key technological framework that effectively enhances learning efficiency by structuring the decision-making process into hierarchical levels. The advantage of HRL lies in its ability to decompose complex tasks into more manageable sub-tasks and sub-goals, allowing the autonomous driving system to flexibly respond to complex road scenarios [4]. Additionally, the Prioritized Experience Replay (PER) technique in HRL is significantly improving learning efficiency and effectiveness. For example, a deep Q-Network (DQN) utilizing PER has outperformed standard DQN in most Atari games, demonstrating the effectiveness of this approach in complex environments [5]. The introduction of new frameworks like Prioritized Sequence Experience Replay (PSER) has further strengthened the potential of HRL in autonomous driving applications [6].

Spatiotemporal analysis techniques are crucial for road condition prediction and navigation in autonomous driving. By analyzing data that changes across space and time, we can better understand and apply this information to improve the decision-making accuracy of autonomous driving systems. The graph attention network (GAT), as an advanced graph neural network (GNN), enhances the precision of node representations by focusing on the most relevant neighbors during the information aggregation process through an attention mechanism, addressing the over-smoothing issue of traditional GNNs [7], [8]. Meanwhile, the long short-term memory (LSTM) network effectively captures long-term dependencies in sequential data, demonstrating excellent learning and performance in time series forecasting and regression tasks. The integration of these technologies provides a more intelligent and safer navigation system for autonomous vehicles.

Moreover, the management of on-ramps and off-ramps is crucial in traffic flow estimation and stability studies. Data-driven transfer learning frameworks have been applied to address traffic estimation problems in areas lacking physical sensors, revealing that on-ramps negatively impact traffic stability, while off-ramps help alleviate traffic congestion [9], [10]. These findings have a direct impact on the path planning and vehicle control strategies of autonomous driving systems. Building on this technical background, this study proposes an innovative autonomous driving system architecture aimed at addressing key challenges in current technologies. The main contributions are as follows:

1. **Spatiotemporal Data Analysis and Safety Field Application:** Combining graph attention network (GAT), long short-term memory (LSTM) network, and Safety Field technology for spatiotemporal data analysis, enhancing the autonomous driving system's ability to perceive, understand, and predict road environments, thereby improving path planning, risk assessment, and dynamic adjustment capabilities.
2. **Enhancement and Application of Hierarchical Reinforcement Learning and Prioritized Experience Replay:** This study applies hierarchical reinforcement learning (HRL) to autonomous driving decision-making and control systems, decomposing complex driving tasks into manageable sub-tasks to improve decision-making efficiency and stability in handling challenging road scenarios. Additionally, the integration of prioritized experience replay (PER) prioritizes and samples experiences based on their significance, further enhancing learning efficiency and improving the system's adaptability in complex and dynamic environments.
3. **Traffic Impact Studies of On-Ramps and Off-Ramps:** Analyzing the impact of on-ramps and off-ramps on traffic flow and stability, proposing more accurate traffic estimation methods, and improving the path planning and vehicle control strategies of autonomous driving systems.

II. Related Work

A. Spatial-Temporal Feature Extraction

In the development of autonomous driving systems, various potential hazards that vehicles face during driving are significant challenges that cannot be ignored. These hazards include sudden lane changes by other vehicles, pedestrians crossing the road, and overly narrow road boundaries. Autonomous driving systems must be able to perceive and assess the safety of the surrounding environment in a timely manner and respond appropriately to ensure driving safety. However, most existing methods still fall short in handling temporal correlations, particularly in situations involving pedestrian interactions, which may pose challenges in practical applications. In this context, the concept of safety fields [11], [12] has garnered increasing attention. Safety fields aim to quantify the safety level of each position on the road based on dynamic changes in the surrounding environment, thereby providing more precise environmental perception for autonomous driving systems. However, relying solely on spatial information to construct safety fields is often insufficient to handle complex driving scenarios. With the development of autonomous driving technology, spatiotemporal feature extraction methods that combine temporal correlations have become a research hotspot. Attention mechanisms [13], [14] have proven to be significantly advantageous in handling spatial data. Through attention mechanisms, autonomous driving systems can effectively identify and focus on potential hazardous areas, thereby providing higher accuracy in environmental perception. At the same time, to fully utilize temporal information, time series models such as LSTM and GRU are widely used to process sequential data, excelling in capturing long-term dependencies and temporal correlations.

B. Path Planning

Although the aforementioned methods have achieved some success in capturing behavioral features and spatial interactions, they still have limitations in addressing dynamic and complex scenarios, especially in changing traffic environments, where more flexible and adaptive decision-making processes are needed. Qiao [15], Zhang [16] and Peng [17] *et al.* introduced hierarchical reinforcement learning into autonomous driving behavior planning, while Naveed et al. [18]combined it with Proportional-Integral-Derivative (PID) controllers for trajectory planning. The application of HRL helps to decompose the tasks of autonomous vehicles into sub-goals, supporting advanced network learning options and low-level trajectory planner strategies. The introduction of sub-goals not only reduces convergence time but also makes the learned strategies reusable in other scenarios, thereby enhancing the system's flexibility and adaptability. Additionally, to further improve the efficiency of reinforcement learning models, experience prioritized replay techniques have been introduced [19], [20]. This technique prioritizes the selection of more valuable experience samples for training based on TD error, accelerating the convergence process and improving learning performance in complex scenarios.

Therefore, by combining attention mechanisms for spatial feature extraction, using models like LSTM and GRU for time series processing, and employing experience prioritized replay techniques to accelerate the reinforcement learning process, the challenges faced by autonomous driving systems in complex and dynamic traffic scenarios can be addressed more effectively. The integrated application of these technologies, particularly the introduction of hierarchical reinforcement learning, not only enhances the decision-making efficiency of the system in variable scenarios but also contributes to the flexibility and robustness of trajectory planning. The research framework of this paper will focus on spatiotemporal feature extraction, initially exploring spatial feature extraction methods based on graph attention networks and safety fields, followed by introducing LSTM for temporal feature extraction, and finally integrating experience prioritized replay techniques and hierarchical reinforcement learning to construct an comprehensive autonomous driving decision and control system.

III. Methodolgy

In this paper, we propose a novel lane-changing architecture for autonomous vehicles, as depicted in Fig. 1. This architecture integrates spatiotemporal feature extraction, hierarchical reinforcement learning, and prioritized experience replay mechanisms, aiming to enhance the efficiency and safety of lane-change decisions in highway ramp merging scenarios. The following sections provide a detailed explanation of each module within the architecture.

A. Spatial-Temporal Feature Extraction

• *Feature Calculation - GAT and Safety Field*

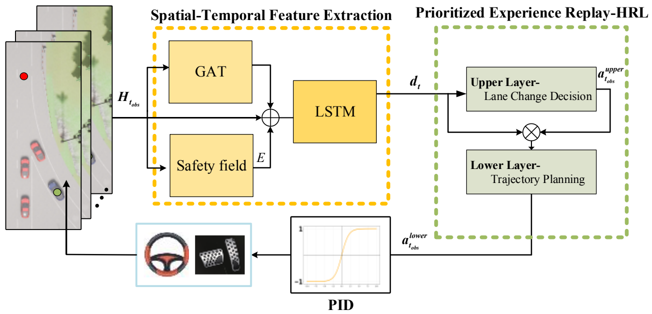

Fig. 1. The architecture of hierarchical reinforcement learning.

In Fig. 1, a Graph Attention Network (GAT) is used to compute the spatial features between vehicles, effectively capturing their dynamic interactions. To evaluate the safety risks in current traffic scenarios, we define an overall safety field that accounts for road boundaries, static obstacles, and surrounding vehicles. It comprises three components, such as lane, static, and kinematic fields, computed as follows:

$$E = E_l + E_s + E_k \tag{1}$$

Next, a brief introduction of field E is shown as below

1) Lane Field E_l

The lane field evaluates the relative distance between the vehicle and the road boundary (e.g., lane markings), reflecting the risk of deviating from the lane center. It is defined as:

$$E_l = M_l \times R_l \times (D/2 - |d_1|)^{k_1} \tag{2}$$

where M_l is the lane type factor (higher for solid lines), R_l is the road condition factor, D is the lane width, d_l is the distance from the vehicle to the lane marking, and k_l controls sensitivity to deviation.

2) Static Field E_s

The static field measures the risk posed by static obstacles such as guardrails, traffic signs, or buildings. Larger objects at closer distances contribute to higher risk. It is calculated as:

$$E_s = M_s \times R_s \times \left(1/|d_s| \right)^{k_s} \tag{3}$$

where M_s is the obstacle volume, R_s is the road condition factor, d_s is the distance to the static obstacle, and k_s adjusts the impact of distance.

3) Kinematic Field E_k

The kinematic field reflects the interaction risk with other moving vehicles, taking into account not only the distance but also the relative speed and direction of those vehicles. It is defined as:

$$E_k = M_k \times R_k \times \left(1/|d_k| \right)^{k_{k1}} \times \exp(k_{k2} \times |v_k| \times \cos\theta_k) \tag{4}$$

where M_k is the volume of the other vehicle, R_k is the road condition factor, d_k is the distance, v_k is its speed, θ_k is the relative angle, and k_1, k_2 are sensitivity constants.

• *Feature Fusion and LSTM Processing*

The system first applies GAT to extract spatial relationships between vehicles, using an attention mechanism to emphasize key neighbors based on position, distance, and velocity. Simultaneously, a safety field is computed to evaluate risks from nearby obstacles and road boundaries. These spatial and safety features are then concatenated with the vehicle's state to form a comprehensive feature set, which is fed into the LSTM to capture temporal dependencies in vehicle behavior. By integrating both raw and processed features, the system improves the prediction of future movements, supporting accurate decision-making in dynamic traffic scenarios.

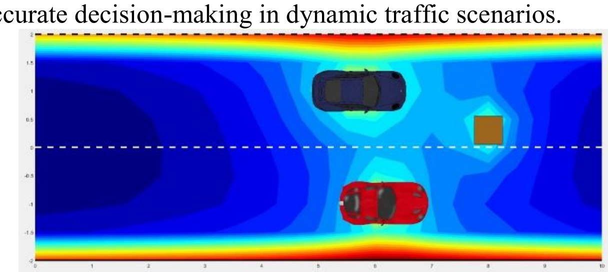

Fig. 2. Field strength of corresponding driving safety field.

B. Prioritized Experience Replay Hierarchical Reinforcement Learnin

1) Upper-Level Decision Module

The hidden vectors output by the LSTM are further input into the upper-level decision module of the hierarchical reinforcement learning framework. This module is responsible for determining whether to perform a lane-change operation based on the current temporal information and road conditions. The upper-level decision module is trained using reinforcement learning methods, continuously learning and adapting to the ever-changing traffic environment to optimize lane-change strategies.

2) Lower-Level Target Selection Module

The lane-change decision from the upper level, along with the LSTM hidden vectors, is input into the lower-level target selection module. This module determines specific driving targets, such as the position of the target lane and the desired speed of the vehicle, based on the upper-level strategy. The lower-level module also employs reinforcement learning techniques to achieve fine-grained control of the vehicle's movement, ensuring that it can safely and accurately reach the designated target as dictated by the upper-level decision.

3) Prioritized Experience Replay

This framework employs a prioritized experience replay mechanism, where the replay priority of experience samples is determined by the magnitude of the TD-error. In the hierarchical reinforcement learning structure, the upper-level and lower-level decisions each generate their respective TD-errors, δ_i^{upper} and δ_i^{lower}. To account for the impact of both levels, we calculate a combined TD-error using the following formula:

$$\delta_i^{comboned} = \lambda \cdot \delta_i^{upper} + (1-\lambda) \cdot \delta_i^{lower} \tag{5}$$

where parameter λ is computed based on the weighted average of the upper and lower TD-errors relative to their previous errors. The specific formula is:

$$\lambda_i = \frac{\Delta\delta_i^{upper}}{\Delta\delta_i^{upper} + \Delta\delta_i^{lower}} \tag{6}$$

It ensures that λ_i reflects the relative changes in TD-errors of the upper and lower layers, and it combines the influences of both layers. The combined TD-error is then used to compute the priority of experience samples, thereby selecting the most valuable data for strategy learning and enhancing the efficiency and performance of the reinforcement learning model.

IV. Experiments

A. Environment Settings

The ramp simulation scenario used in this paper is based on the "Town04" map from the open-source autonomous driving simulation software CARLA Simulator. In addition to simulating vehicle dynamics, we integrated V2V communication into the simulation environment, allowing vehicles to share real-time data, such as speed and lane intensions. This feature was used to access the impact of V2V on improving decision-making accuracy and reducing collision rates in complex ramp merging scenarios.

The sensor configurations for autonomous driving can be flexibly adjusted to test the performance of different algorithms. The agent is required to perform lane-changing maneuvers, such as merging onto and exiting from the ramp. If a collision occurs, the episode will end prematurely. The training simulation diagram is shown in Fig. 3 and Fig. 4.

B. State Space

During the lane-changing process, the selection is made to better match the real-world data collection and environmental perception needs of autonomous driving systems. In the given scenario, the state space is used to comprehensively describe the environmental conditions during the land change process, and can be represented by tuple $\vec{s}$ shown below:

$$\vec{s}^i = \left[p_x^i, p_y^i, v_x^i, v_y^i, a_x^i, a_y^i, l_{ID}^i \right], i = 1, 2, ..., N$$

where N indicates the amount of vehicles, $p_x^i, p_y^i, v_x^i, v_y^i, a_x^i, a_y^i$ are the longitudinal and lateral positions, velocities and accelerations, l_{ID}^i is the lane ID of vehicles.

C. HRL structure for Decision Making

The HRL framework in this study divides the decision-making process into two levels: upper and lower, each responsible for different levels of decision tasks. The upper level focuses on strategic decisions, specifically whether ego vehicle should change lanes or maintain the current lane, and the upper-level action space is defined as: a^{upper} ={lane change; lane keep}. In the lower-level planning process, whether the vehicle is maintaining the lane or preparing for a lane change, the system plans the path based on the driving situation. When the ego vehicle chooses to maintain or wait in the lane, the lower-level trajectory planner generates variable-length sub-trajectories according to traffic conditions and transmits the final waypoint and target speed to the PID controller to achieve smooth tracking. In the context of a lane change, the system uses an ε-greedy strategy to select waypoints in the target lane, ensuring operational stability and smoothness. The lower-level selected target points are adjusted based on different driving needs, with the action space defined as: a^{lower} ={danger; moderate; safe}. Danger-level waypoints are used for scenarios requiring lane changes at higher speeds or when advanced paths

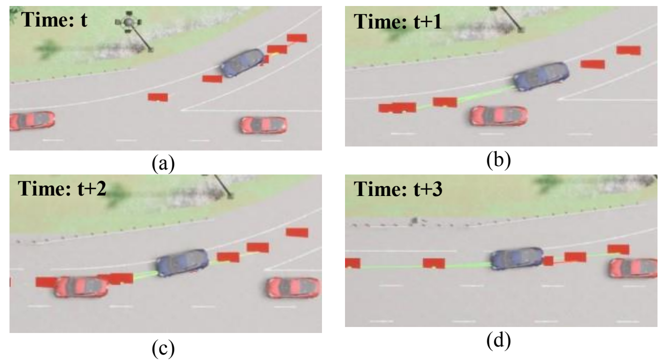

Fig. 3. The architecture of cut-in simulation.

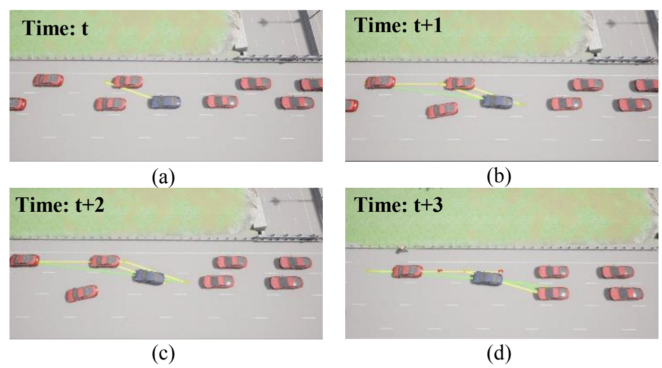

Fig. 4. The architecture of cut-out simulation.

are needed. These waypoints enable gradual adjustments in vehicle direction over greater distances, minimizing sharp turns or abrupt directional changes and ensuring vehicle stability and safety. Moderate-level waypoints are suitable for general traffic flow conditions, striking a balance between speed and stability. Safe-level waypoints are employed in low-speed or sharp turn situations, allowing the vehicle to maneuver more flexibly and precisely in complex environments.

Additionally, during model training, we considered the impact of learning rates on model stability and convergence speed. A high learning rate may lead to unstable learning while a low rate could result in slow convergence. Therefore, we selected a moderate learning rate to balance learning speed with stability, ensuring that the model can make accurate and efficient decisions across various driving scenarios.

D. Rewards

When the vehicle agent operates in a dense traffic environment, its goal is to maintain a high driving speed while minimizing unnecessary lane changes and ensuring the stability of overtaking maneuvers [21]. The Hybrid Reward Mechanism [15] is used to our reward structure, rewards are given when the agent successfully changes lanes and reaches its destination, while penalties are imposed for collisions, failure to reach the destination within the time limit, and negative impacts on the average traffic speed. The reward function is defined as follows:

$$R_{total} = R_{LC\_success} + R_{LC\_fail} + R_{goal\_success} + R_{goal\_fail} + R_{collision} + R_{efficiency} + R_{speed\_impact} \quad (7)$$

where $R_{LC\_success}$ and R_{LC_fail} indicate that the vehicle success or fail lane change, $R_{goal\_success}$ and R_{goal_fail} indicate that the vehicle has or has not arrive the destination, $R_{collision}$ indicates the collision happened. $R_{efficiency}$ indicates that the reward function for the efficiency of the ego vehicle reaching the set target, which is the ratio of the time required for the ego vehicle to reach the target point to the time limit. R_{speed_impact} means that the impact of the average speed of other vehicles in the current lanes. In the training process, a higher penalty weight was assigned to rewards related to reaching the destination and maintaining average traffic speed, ensuring both driving safety and traffic flow stability.

E. Results

Fig. 3 and Fig.4 illustrate the vehicle's behavior during the cut-in and cut-out maneuver. Fig. 5(a) shows the speed variation of the ego vehicle and the rear target vehicle during the cut-in process. It can be observed that while the ego vehicle's speed fluctuates as it approaches the ramp entrance, it remains within a certain range. Meanwhile, the rear target vehicle, after an initial acceleration, gradually stabilizes its speed, slightly decelerating to coordinate with the ego vehicle. To further evaluate the model's robustness, we simulated a more complex scenario by placing a stationary vehicle, such as a construction vehicle, as an obstacle in the target lane. In this situation, the ego vehicle must perform more precise path planning to safely merge while avoiding the stationary vehicle. This setup mimics a real-world hazardous scenario, providing a robust test of the model's ability to handle dynamic traffic conditions while navigating around stationary hazards. Fig. 5(b) depicts the trajectories of the ego vehicle and the rear target vehicle, clearly showing the ego vehicle's curved trajectory as it gradually merges from a lower position into the main lane, while the rear target vehicle maintains a relatively straight path. These illustrations reveal how the ego vehicle adjusts its speed and direction to successfully merge into the main lane during the cut-in process, while the rear target vehicle adjusts its speed and position to maintain a safe distance and avoid collision. These charts provide an in-depth understanding of the dynamic behavior of vehicles during the cut-in maneuver, laying the groundwork for subsequent analysis.

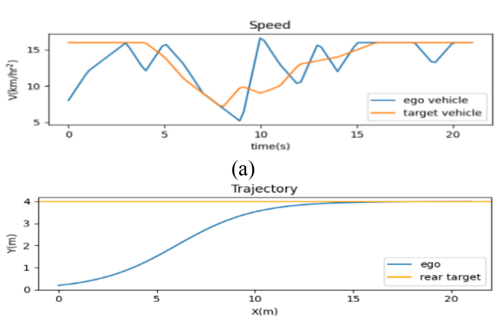

Fig. 5. Simulation results of speed and trajectory in vehicle cut-in scene.

To evaluate the effectiveness of our hierarchical reinforcement learning approach in highway ramp merging and exiting, we compared it with HIDS-DRL [16], D3QN-DDPG [17], and Robust-HRL [18], and conducted ablation studies. All models were trained for 1500 rounds to ensure convergence. The model combines a spatiotemporal graph attention module, consisting of GAT, a safety field, and LSTM, with prioritized experience replay. Experiments assess the effect of removing each component and compare performance with Robust-HRL using metrics such as driving speed, lane change frequency, and ramp merging success rate. To comprehensively assess the performance of different models in autonomous driving scenarios involving highway on-ramp and off-ramp maneuvers, we selected two key metrics for comparison:

1. **Success Rate (%)**: The success rate denotes the proportion of successful highway on-ramp or off-ramp maneuvers completed by the model. A higher success rate means the performs more reliably and stably in traffic situations.

2. **Speed Impact (ΔV)**: This metric measures the change in speed on the target lane after a lane change. Ideally, the model should minimize the impact on the speed of other vehicles in the target lane after a lane change, with a smaller speed change indicating a smoother driving experience.

From Tables I and II, it can be observed that our proposed model outperforms other models, including D3QN-DDPG and HIDS-DRL, in both cut-in and cut-out scenarios, demonstrating higher success rates and reduced speed impacts. These results highlight the model's ability to optimize decision-making, minimize collision risks, and maintain lane speed

TABLE I. CUT-IN SCENE COMPARISON OF RESULTS OBTAINED FROM DIFFERENTALGORITHMS

Method	***Robust-HRL (2021)***	***D3QN-DDPG(2022)***	***HIDS-DRL (2023)***	***PER-HRL(Our)***
Success Rate %	94.6%	95.1%	95.7%	96.7%
Speed Impact (ΔV) (km/h)	-4.6	-4.2	-3.6	-3.0

TABLE II. CUT-OUT SCENE COMPARISON OF RESULTS OBTAINED FROM DIFFERENTALGORITHMS

Method	***Robust-HRL (2021)***	***D3QN-DDPG(2022)***	***HIDS-DRL (2023)***	***PER-HRL(Our)***
Success Rate %	95.4%	95.6%	96.0%	97.0%
Speed Impact (ΔV) (km/h)	-4.4	-4.0	-3.4	-2.8

TABLE III. CUT-IN SCENE ABLATION COMPARISON RESULTS

Method	*Without GAT*	*Without safety field*	*Without LSTM*	*PER-HRL(Ours)*
Success Rate %	95.1%	95.6%	96.0%	96.7%
Speed Impact (ΔV) (km/h)	-4.3	-3.8	-3.6	-3.0

TABLE IV. CUT-OUT SCENE ABLATION COMPARISON RESULTS

Method	*Without GAT*	*Without safety field*	*Without LSTM*	*PER-HRL(Ours)*
Success Rate %	95.2%	95.7%	96.1%	96.8%
Speed Impact (ΔV) (km/h)	-4.3	-3.9	-3.6	-3.1

stability in highway ramp scenarios. In contrast, the D3QN-DDPG model faces challenges in adapting to dynamic traffic interactions, while the HIDS-DRL model lacks the integration of critical components such as GAT, Safety Field, or LSTM, leading to difficulties in maintaining speed consistency and preventing collisions. This emphasizes the importance of these modules in enhancing safety and performance. Tables III and IV present the ablation studies for the two scenarios, confirming the critical role of each module in improving system stability and decision-making. Removing individual components significantly impacts success rates and safety metrics, while the inclusion of all modules enables our proposed model to effectively navigate complex environments.

V. CONCLUSION

This paper proposes a hierarchical reinforcement learning-based decision-making framework specifically designed for cut-in and cut-out maneuvers in highway ramp scenarios. By integrating advanced components such as graph attention networks (GAT), safety field, and long short-term memory (LSTM) networks, the system efficiently captures spatiotemporal features and dynamically evaluates traffic interactions and risk levels. These capabilities significantly enhance decision-making efficiency and safety, particularly in complex traffic environments with high vehicle density and dynamic flow conditions. Experimental results validate the proposed framework's effectiveness in achieving high success rates during both cut-in and cut-out scenarios, demonstrating its ability to maintain smooth traffic flow and reduce disruptions to surrounding vehicles. Furthermore, the hierarchical structure, which decomposes tasks into high-level decision-making and low-level trajectory planning, ensures the model's adaptability to varying traffic conditions, supporting more stable and precise lane-changing maneuvers. Future work will incorporate constraints such as speed and following distance during cut-in maneuvers.

ACKNOWLEDGMENT

This Research was supported by National Science and Technology Council, R.O.C., under the Grant NSTC 113-2218-E-035-001 and NSTC 112-2628-E-035 -001 -MY3

REFERENCES

[1] K. Yuan, "Analysis of the Current Development and Future Prospect of Autonomous Driving," *Apply Computing and Engineering*, vol. 45, pp. 147–151, 2024.

[2] T. Zhang, "The Application of Deep Learning in Autonomous Driving," *Apply Computing and Engineering*, vol. 50, pp. 144–148, 2024.

[3] C. Chen, J. Guo, C. Guo, C. Chen, Y. Zhang, and J. Wang, "Adaptive Cruise Control for Cut-In Scenarios Based on Model Predictive Control Algorithm," *Applied Science*, vol. 11, no. 11: 5293, pp. 1–29 2021.

[4] C. Diuk and M. L. Littman, "Hierarchical Reinforcement Learning," *IGI Global*, pp. 825–830, 2020.

[5] B. Saglam, F. B. Mutlu, D. C. Cicek, and S. S. Kozat, "Actor Prioritized Experience Replay," *Journal of Artificial Intelligence Research (JAIR)*, vol. 78, pp. 639–672, 2023.

[6] M. Brittain, J. Bertram, X. Yang, and P. Wei, "Prioritized Sequence Experience Replay," *arXiv*: 1905.12726, pp.1–18, 2020.

[7] M. Eliasof and E. Treister, "Global-Local Graph Neural Networks for Node-Classification," *arXiv*: 2406.10863, pp.1–8, 2024.

[8] Chen, Zhao, "Graph Adaptive Attention Network with Cross-Entropy," *Entropy*, Vol. 26, no. 7, pp. 576, 2024.

[9] J. Zhang, C. Song, Z. Mo and S. Cao, "A Transfer Learning-Based Approach to Estimating Missing Pairs of On/Off Ramp Flows," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 2, pp. 1247–1262, 2024.

[10] R. Cheng, X. An, and Y. Cheng, "Nonlinear Analysis of an Extended Heterogeneous Lattice Hydrodynamic Model Considering on/off-Ramps," *International Journal Applied Mechanics,* vol. 16, issue 3, pp. 1–22, 2024.

[11] B. Zhu, J. Han, J. Zhao and H. Wang, "Combined Hierarchical Learning Framework for personalized Automatic Lane-Changing", *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 10, pp. 6275–6285, 2020.

[12] J. Wang, J. Wu and Y. Li, "The Driving Safety Field Based on Driver–Vehicle–Road Interactions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2203–2214, 2015.

[13] A. Vemula, K. Muelling, and J. Oh, "Social Attention: Modeling Attention in Human Crowds," *IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, pp 1–7, 2018.

[14] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "STGAT: Modeling Spatial-Temporal Interactions for Human Trajectory Prediction," *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), pp. 6271–6280, 2019.

[15] Z. Qiao, Z. Tyree, P. Mudalige, J. Schneider and J. M. Dolan, "Hierarchical Reinforcement Learning Method for Autonomous Vehicle Behavior Planning," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, pp. 6084–6089, 2020.

[16] X. Zhang, J. Sun, Y. Wang and J. Sun, "A Hierarchical Framework for Multi-Lane Autonomous Driving Based on Reinforcement Learning," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 4, pp. 626-638, 2023.

[17] J. Peng, S. Zhang, Y. Zhou and Z. Li, "An Integrated Model for Autonomous Speed and Lane Change Decision-Making Based on Deep Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 21848–21860, 2022.

[18] K. B. Naveed, Z. Qiao and J. M. Dolan, "Trajectory Planning for Autonomous Vehicles Using Hierarchical Reinforcement Learning," *IEEE International Intelligent Transportation Systems Conference (ITSC)*, Indianapolis, IN, USA, pp. 601–606, 2021.

[19] X. Tao and A. S. Hafid, "DeepSensing: A Novel Mobile Crowdsensing Framework with Double Deep Q-Network and Prioritized Experience Replay," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11547–11558, 2020.

[20] D. Fährmann, N. Jorek, N. Damer, F. Kirchbuchner and A. Kuijper, "Double Deep Q-Learning with Prioritized Experience Replay for Anomaly Detection in Smart Environments," *IEEE Access*, vol. 10, pp. 60836–60848, 2022.

[21] K. Min, H. Kim and K. Huh, "Deep Distributional Reinforcement Learning Based High-Level Driving Policy Determination," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 3, pp. 416–424, 2019.