Optimizing Vehicle Trajectories and Ensuring Pedestrian Safety in Complex Traffic

Zhenwei Xu[1,*], Jiaqi Zeng[2], Yaoyong Zhou[2], Qing Yu[1,†] and Wushouer Silamu[1]

[1]School of Computer Science and Technology, Xinjiang University (XJU), No. 777, Huarui Street, Shuimogou District, Urumqi, 830017, Xinjiang Uygur Autonomous Region, People's Republic of China

[2]School of Software, Xinjiang University (XJU), 499 Xibei Road, Urumqi, 830091, Xinjiang Uygur Autonomous Region, People's Republic of China

Abstract

In tackling the complexities of vehicle trajectory planning at intricate traffic intersections, where prioritizing safety and efficiency is crucial, this research introduces an innovative model dubbed Adaptive Hierarchical Traffic Intersection Reinforcement Learning (AHTRL). This model is underpinned by a hierarchical deep deterministic policy gradient algorithm, which is structured into two principal layers: the upper layer is tasked with policy decision-making, while the lower layer focuses on the generation of precise waypoints. Steering and throttle adjustments are meticulously managed by a foundational PID controller, ensuring pinpoint accuracy in vehicular control. The model is further enhanced by integrating recurrent neural networks for the analysis of historical vehicle trajectory data, alongside a spatio-temporal variational autoencoder (ST-VAE) for the prediction of pedestrian future movements, thereby markedly improving interactive safety measures. Rigorous testing conducted on the Carla simulation platform has demonstrated the model's exceptional performance, outstripping existing methodologies across several critical metrics. Relative to the optimal baseline model, AHTRL has achieved a commendable 11.9% boost in total average rewards, a notable 13% increase in average transit velocity, and a significant 36% decrease in collision occurrences, affirming its dominance in ensuring safety and enhancing efficiency within the realm of complex intersection trajectory planning. These outcomes not only underscore the model's superior safety and efficiency but also its remarkable adaptability and practicality in navigating complex traffic scenarios. By melding advanced hierarchical reinforcement learning frameworks with cutting-edge deep learning technologies, this study substantially elevates the caliber of vehicle trajectory planning at convoluted traffic intersections. It paves the way for novel, efficacious solutions for fostering safe and efficient interactions within intelligent transportation systems. This contribution is not only academically innovative but also sets a robust foundation for real-world applications.

Keywords

Autonomous Driving Technology, Pedestrian Trajectory Prediction, Vehicle Trajectory Planning, Hierarchical Reinforcement Learning

1. Introduction

In modern society, with the acceleration of urbanization and the continuous increase in the number of vehicles, traffic intersections have become one of the most complex and challenging parts of the urban transportation system. Vehicle trajectory planning at traffic intersections is of great practical significance for improving traffic efficiency, reducing traffic congestion, and lowering the rate of accidents. With the development of autonomous driving technology, how to achieve safe and efficient vehicle trajectory planning in complex traffic intersection environments has become a key issue in the field of autonomous driving research.

In traditional vehicle trajectory planning research, the vast majority of methods depend on accurate environmental models and predefined rules [1]. These approaches can exhibit good performance in dealing with simple or predefined scenarios. However, the complexity of traffic intersections mainly

ICCBR AI Track'24: Special Track on AI for Socio-Ecological Welfare at ICCBR2024, July 1, 2024, Mérida, Mexico

† Corresponding Author.

zhenweixu@stu.xju.edu.cn (Z. Xu); 2433260364@qq.com (J. Zeng); 107552104337@stu.xju.edu.cn (Y. Zhou); zhenweixuelvis@gmail.com (Q. Yu); wushour@126.com (W. Silamu)

https://github.com/the-low-key-emperor (Z. Xu); https://ieeexplore.ieee.org/author/37397678900 (W. Silamu)

0009-0003-4156-1226 (Z. Xu); 0000-0001-7116-9338 (J. Zeng); 0009-0007-2571-9731 (Y. Zhou); 0009-0006-7944-1889 (W. Silamu)

arises from the dynamic changes of participants, the unpredictability of pedestrian behavior, and the variability of traffic signals. These elements render traditional trajectory planning methods insufficiently adaptable to the complexity and uncertainty inherent in traffic intersection scenarios.

In recent years, learning-based methods, especially reinforcement learning, have become effective means for solving the problem of vehicle motion trajectory planning at traffic intersections. Reinforcement learning, by interacting with the environment, learns the mapping from observed states to actions taken, demonstrating strong potential [2]. However, despite this potential, early end-to-end reinforcement learning approaches still face several challenges when dealing with complex traffic intersection scenarios. These challenges include the opacity of the decision-making process, difficulty in generating stable behaviors, and low sample efficiency, among others [3, 4]. These challenges highlight areas that require further exploration and optimization when applying reinforcement learning in complex environments.

Addressing the challenges mentioned above, this chapter introduces a hierarchical reinforcement learning framework, AHTRL, specifically designed for the traffic intersection scenario. The core idea of the AHTRL method is to decompose the decision-making problem in complex traffic intersections into multiple sub-problems and solve them hierarchically, thereby enhancing the overall efficiency and stability of decision-making. Moreover, this study pays special attention to the safe interaction between vehicles and pedestrians. To accurately predict pedestrians' future trajectories and plan the vehicle's driving path based on this, the chapter utilizes a Spatio-Temporal Variational Encoder (ST-VAE) to process the historical trajectory data of pedestrians.

In terms of technical implementation, this study employs a hierarchical approach known as Hierarchical Deep Deterministic Policy Gradient (HDDPG). The model is structured into three layers (as shown in Figure 1), with the top layer, the selection layer, responsible for making vehicular action decisions, such as stopping, turning left, turning right, and decelerating. These decisions are based on an analysis and evaluation of the current state of the vehicle, environmental conditions, and objectives. For instance, when approaching an intersection, the selection layer decides whether to continue straight, decelerate, or stop, considering traffic rules, the condition of the intersection, and vehicle objectives. This layer's decisions provide guidance for the vehicle, ensuring that its actions comply with safety and efficiency requirements. Guided by high-level decisions, the trajectory planning layer then generates specific waypoint trajectories that detail how the vehicle should move from its current position to the destination, considering the vehicle's dynamic limitations, road conditions, and obstacle avoidance needs to ensure the trajectory is both safe and practical. In this way, the trajectory planning layer bridges high-level policy decisions and low-level execution controls, providing a clear and feasible path. Finally, the Proportional-Integral-Derivative (PID) controller operates at the trajectory tracking layer, ensuring the vehicle accurately follows the planned waypoint trajectory by dynamically adjusting the vehicle's state (such as speed and direction) to minimize deviations. The PID controller adjusts its three parameters (proportional, integral, derivative) in response to any deviations from the trajectory, ensuring the vehicle closely follows the predetermined path. Moreover, by analyzing pedestrians' historical trajectories and predicting their future movements, this study can plan vehicle trajectories more accurately, ensuring safe and effective interaction between vehicles and pedestrians at traffic intersections.

In summary, this paper effectively addresses the vehicle trajectory planning problem in the complex scenario of traffic intersections by proposing a hierarchical reinforcement learning framework specifically designed for this context. Simulation results demonstrate that, compared to existing methods, the framework proposed in this chapter performs better in adapting to the complexity and uncertainty of traffic intersections, successfully reducing merging time and collision rates under the premise of ensuring safety, and generating smoother trajectories. These achievements not only showcase the application value of hierarchical reinforcement learning in the field of autonomous driving but also offer new perspectives and methodologies for future trajectory planning research in complex traffic environments. The main contributions are as follows:

- For the complex scenario of traffic intersections, this paper introduces a hierarchical decision

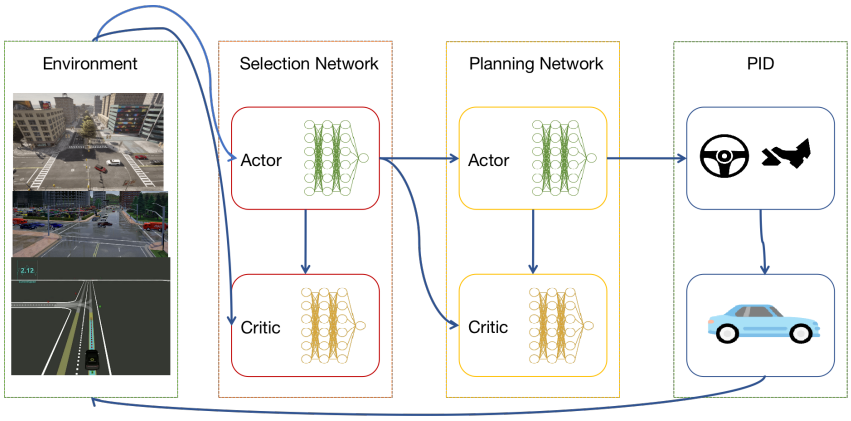

Figure 1: The three-layer hierarchical architecture (the top layer being the selection layer, the bottom layer the trajectory planning layer, and finally, the planned trajectories are precisely followed through a Proportional-Integral-Derivative (PID) controller).

model, AHTRL, which effectively solves the vehicle motion trajectory planning problem. This approach decomposes the complex decision-making problem into multiple sub-problems and solves them layer by layer, significantly improving the efficiency and stability of the decision-making process.

- This study incorporates a Spatio-Temporal Variational Autoencoder (ST-VAE) for processing pedestrian historical trajectory data, with the objective of precisely forecasting future pedestrian movements. Leveraging these predictions, the AHTRL model meticulously crafts vehicle driving paths to ensure a safe interaction between vehicles and pedestrians. This approach not only enhances the precision of trajectory predictions but also furnishes vehicles with robust decision-making support for autonomous navigation in intricate environments.
- This research employs a hierarchical approach using Deep Deterministic Policy Gradient (DDPG) methodology, coupled with a Proportional-Integral-Derivative (PID) controller, to realize a comprehensive process from high-level policy decision-making through to detailed trajectory planning and precise trajectory tracking. This technical implementation not only elevates the accuracy of trajectory planning but also ensures the smoothness and safety of the path.
- Testing on the Carla simulation platform has demonstrated that the AHTRL model surpasses existing methodologies in both efficiency and safety. Compared to the optimal baseline model, it achieved an 11.9% increase in total average rewards, a 13% improvement in average transit speed, and a 36% reduction in collision rates. These results validate the superior performance and practical value of the AHTRL model in complex intersection trajectory planning.

2. Related Work

This section provides a systematic overview of the existing research in the vehicle trajectory planning domain, with a particular focus on traditional methods, the application of reinforcement learning, and the progress in hierarchical reinforcement learning frameworks. These discussions lay the theoretical and technical groundwork for this paper, aiming to offer robust background support for further exploration and resolution of vehicle trajectory planning challenges.

2.1. Non-reinforcement Learning Methods

In the field of autonomous driving, trajectory planning is one of the key technologies ensuring safe and efficient vehicle operation. In recent years, non-reinforcement learning (NRL) methods have made significant progress in addressing trajectory planning problems. These methods can be broadly categorized into four groups: classical planning methods, heuristic-based methods, supervised learning

approaches, and statistical methods. Classical planning methods, such as Rapidly-exploring Random Trees (RRTs) [5] and its variant RRT* [6], are widely used sampling-based planning techniques in trajectory planning. They construct a tree-like structure through random sampling points to effectively explore the environmental space, searching for paths from start to end. RRTs are suitable for complex environments with obstacles but may not guarantee the optimal solution. RRT* improves the path's optimization level and efficiency by refining the path selection process. Heuristic-based methods, including Time-To-Collision (TTC [7]) and slot-based approaches [8], rely on predefined rules and heuristic strategies for trajectory planning. TTC [7] assesses safety by estimating vehicle arrival and collision times, while slot-based methods [8] check the safety of the target lane or intersection. These methods are effective in specific contexts but have limited generalization ability in unknown or variable environments. Supervised learning approaches learn trajectory planning strategies by analyzing expert drivers' driving data, including deep imitation learning [9, 10, 11], which learns driving strategies in urban scenarios through offline learning and enhances driving safety with safety control modules. While these methods can mimic human driver behavior, they require large amounts of high-quality driving data and may face challenges in unseen scenarios. Statistical methods [12] predict vehicle behaviors and decision-making strategies by analyzing historical data, such as using change point-based approaches for predicting and decision-making in autonomous driving vehicles. Bayesian change point detection estimates the target vehicle's strategy and simulates interactions between vehicles based on these predictions. These methods can provide probabilistic forecasts of future behaviors but may require complex calculations and extensive historical data.

In summary, non-reinforcement learning methods for trajectory planning each have their advantages and limitations. Classical planning methods and heuristic-based approaches are effective in specific scenarios but may lack flexibility and generalization capability. Supervised learning and statistical methods can handle more complex scenarios but rely heavily on large amounts of data and computational resources. In contrast, reinforcement learning methods offer a more dynamic and adaptive trajectory planning solution through learning from interactions with the environment. Nonetheless, in-depth research into non-reinforcement learning methods remains crucial for understanding and improving the trajectory planning of autonomous vehicles.

2.2. Reinforcement Learning Methods

Although non-reinforcement learning methods can directly learn from human driving behaviors, their reliance on large amounts of manually annotated data and the uncertainty brought about by differences in decision-making among drivers limit their application in complex tasks. To reduce the dependency on labeled data, researchers have turned to reinforcement learning (RL) for autonomous decision-making and planning. The advent of Deep Reinforcement Learning (DRL), which combines deep learning techniques, has shown immense potential in handling complex decision-making and planning problems, especially achieving breakthrough progress in areas like intelligent gaming, natural language processing, and autonomous driving. DRL is capable of exploring a wide range of possibilities, including hazardous scenarios, and has the potential to achieve performance beyond human capabilities. The study by Mnih et al. [13] marked a significant breakthrough in combining deep learning with reinforcement learning by using an end-to-end Q-Learning framework to learn control signals directly from screen captures, employing a deep learning approach based on Q-learning for the first time. Subsequently, Wolf et al. [14] introduced the Q-learning method to the field of intelligent vehicles, defining various driving actions in the Gazebo simulator and making action decisions based on image information to enhance the processing of high-dimensional sensory inputs. Kendall et al. [15] successfully applied the Deep Deterministic Policy Gradient (DDPG) algorithm in actual intelligent vehicles, using monocular images as the sole input to teach the agent lane-keeping strategies, demonstrating performance comparable to human drivers in a 250-meter road test. To improve the efficiency of exploration in continuous spaces, Liang et al. [16] combined imitation learning with DDPG, introducing an adjustable gating mechanism to selectively activate different control signals for central signal control of the model. Addressing the limitations of learning efficiency in RL methods, Tian et al. [17] designed a new strategy

that integrates human prior knowledge into reinforcement learning. Huang et al. [18] proposed a human-guided reinforcement learning method, enhancing the learning efficiency and performance in complex scenarios through an innovative priority experience replay mechanism.

Reinforcement learning offers an effective approach for vehicle trajectory planning, but to fully leverage this technology's potential, challenges such as low sample efficiency and handling complex tasks must be overcome. Future research directions include exploring more efficient learning algorithms and integrating advanced perception and decision-making mechanisms to achieve effective trajectory planning in even more complex environments.

2.3. Hierarchical Reinforcement Learning

Hierarchical Reinforcement Learning (HRL) simplifies the learning process by dividing the overall problem into a hierarchical structure of multiple subtasks. Each subtask is assigned specific objectives and strategies, and these subtasks are organized hierarchically, with higher-level subtasks providing guidance and context to lower-level ones. This layered approach allows agents to focus on narrower problem scopes, reducing the complexity of the learning task and making the problem more solvable.

Chen et al. [19] proposed a dual-layer architecture for lane-changing tasks, where the upper layer network decides whether to perform a lane change, and the lower layer network learns the specific strategy for executing the chosen action. Building on this, Shi et al. [20] and Li et al. [21] further developed a two-stage hierarchical reinforcement learning method. Shi et al. [20] employed a pure tracking strategy to follow trajectory points, while Li et al. [21] enhanced the performance of the lower-level controller by integrating vehicle position, speed, and heading information. These methods offer robust solutions for building efficient and safe autonomous driving systems. Lu et al. [22] introduced a hierarchical reinforcement learning method for autonomous decision-making and motion planning in complex dynamic traffic scenarios. Duan et al. [23] decomposed the navigation task into three modules, where the main policy network, trained to select appropriate driving tasks, significantly improved the model's versatility and efficiency. Building on the work of Duan et al. [23], the introduction of Cola-HRL [24] aimed to further enhance decision quality in complex scenarios. These studies highlight the immense potential of Hierarchical Reinforcement Learning (HRL) in simplifying complex decision-making and motion planning tasks, paving new pathways and perspectives for the development of autonomous driving technology.

Hierarchical Reinforcement Learning (HRL) has garnered attention for its significant advantages in simplifying complex tasks and improving learning efficiency. However, it faces considerable challenges in ensuring pedestrian safety, a critical area. Particularly in unpredictable traffic intersection scenarios, HRL shows certain limitations in dealing with dynamic environments involving pedestrian interactions. The model's insufficient predictive ability often struggles to accurately capture pedestrian movement trajectories and intentions, limiting autonomous vehicles' capacity for safe and efficient passage in complex traffic situations. Moreover, the generalization ability of HRL strategies is also challenged. Faced with new traffic scenes or previously unseen pedestrian behaviors, even well-trained models may falter, struggling to respond appropriately, which compounds the difficulty of ensuring pedestrian safety. This limitation in generalization, stemming from HRL's inherent hierarchical structure, restricts its adaptability in unknown environments. Thus, despite HRL's notable advantages in handling complex tasks, ensuring pedestrian safety in critical scenarios like traffic intersections remains fraught with multiple challenges. Future developments should focus on enhancing the model's predictive accuracy regarding environmental changes and strengthening strategy generalization capabilities, aiming to ensure pedestrian safety without compromising efficiency and fluidity in complex traffic environments. This requires not only deepening technical research but also continuously exploring new ideas and methods to overcome existing limitations, pushing autonomous driving technology towards higher levels of safety and intelligence.

3. Method

3.1. Architecture Overview

Figure 2 depicts a Hierarchical Reinforcement Learning framework tailored for vehicle motion planning at traffic intersections, named AHTRL. The AHTRL model is divided into three levels, each performing different trajectory decision-making and planning tasks. Both the high-level selector and the low-level planner utilize the Actor-Critic structure of the Deep Deterministic Policy Gradient (DDPG). The high-level selector undertakes key decision-making tasks, such as stopping, turning, and decelerating, based on a comprehensive analysis of the current situation of vehicles and pedestrians as well as the surrounding environment. The lower-level trajectory planning layer generates precise waypoint trajectories according to decisions made by the top level, taking into account the vehicle's dynamic characteristics, current traffic conditions, and obstacle avoidance needs, ensuring the safety and practicality of the trajectory. The trajectory tracking layer ensures that the vehicle can accurately follow the planned waypoints through a PID controller. Additionally, the framework integrates a Spatio-Temporal Variational Autoencoder (ST-VAE) model, specifically for predicting pedestrian future trajectories, which is crucial for ensuring safe interactions between vehicles and pedestrians at traffic intersections.

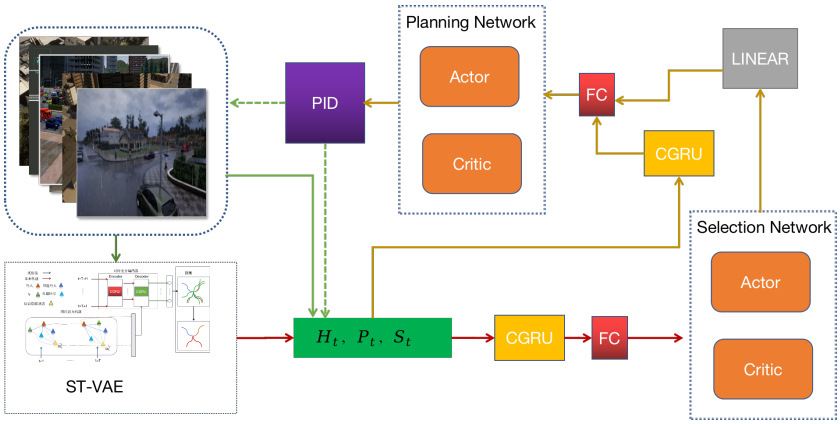

Figure 2: AHTRL model overall architecture (The framework consists of three layers: a decision selector, a trajectory planner, and a trajectory tracker. Both the selector and the planner are based on the Actor-Critic architecture, making and executing decisions based on the vehicle's historical trajectory (H_t), the predicted future trajectory of pedestrians (P_t), and environmental information (S_t).)

3.2. Spatio-Temporal Variational Autoencoder

The ST-VAE (Spatio-Temporal Variational Autoencoder) model is dedicated to predicting the future movement trajectories of entities within a scene. In a complex setting containing N pedestrians, the ST-VAE model can predict their future location distribution over a future time span H by analyzing each pedestrian's spatial position over a time span T. The ST-VAE model particularly focuses on extracting features from pedestrians' social behaviors and independently predicting each pedestrian's future trajectory. This model is adaptable to scenarios of varying scales, especially in urban scenes where dynamic changes occur rapidly and the pedestrian's surrounding environment frequently changes. Even when a pedestrian's local neighborhood is difficult to continuously monitor, the ST-VAE model can accurately predict their future positions. This predictive capability not only aids in enhancing the efficiency of autonomous driving and intelligent surveillance systems but also provides new perspectives and tools for understanding and analyzing crowd dynamic behaviors.

Figure 3 showcases the overall architecture of the model proposed in this study, centrally featuring a time-based variational autoencoder integrated with the Complex Gated Recurrent Unit (CGRU) structure designed in previous research [25], for sequence prediction relying on the CGRU state variables in an autoregressive model. Diverging from traditional methods that directly predict time series data, this

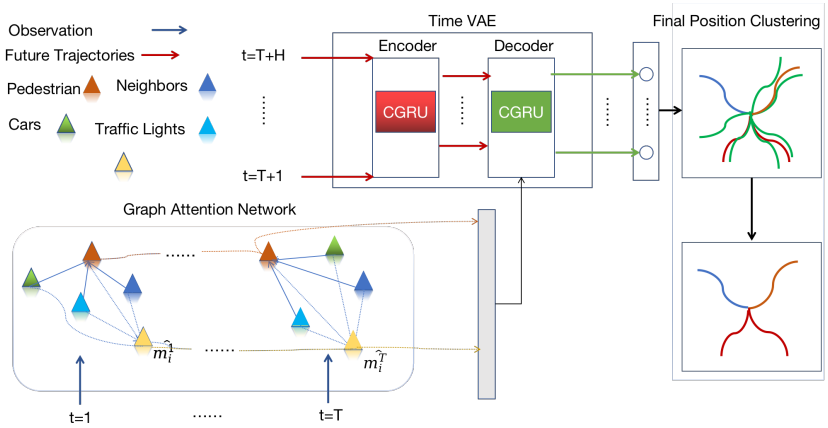

Figure 3: ST-VAE model structure (ST-VAE is used to capture pedestrian behavior patterns in both time and space. At its core, ST-VAE employs a Variational Autoencoder (VAE), with a Graph Attention Network (GAT) responsible for assessing the spatial influence between individuals. Meanwhile, a Convolutional Gated Recurrent Unit (CGRU) focuses on improving the processing of temporal features. Finally, ST-VAE utilizes Final Position Clustering (FPC) to enhance the diversity of trajectory predictions.)

model innovatively utilizes historical observational data as conditional variables. Moreover, the model does not predict the future absolute positions $x_i^{T+1:T+H}$ but rather the sequence of positional changes $d_i^{T+1:T+H}$, where $d_i^{t+1} \triangleq x_i^{t+1} - x_i^t$. Through this approach, the target probability distribution of the displacement sequence is precisely defined (as in Equation 1), providing a more nuanced and dynamic perspective for pedestrian trajectory prediction.

$$p(d_i^{T+1:T+H}|O_i^{1:T}) = \prod_{\tau=1}^{H} p(d_i^{T+\tau}|d_i^{T+1:T+\tau-1}, O_i^{1:T}). \tag{1}$$

To effectively capture the complex interactions between pedestrians and other traffic participants in crowded scenes, the ST-VAE model employs a Graph Attention Network (GAT). Traffic participants within the scene (such as pedestrians, traffic lights, vehicles, and crosswalks) are represented as nodes in a graph, with edges between nodes indicating interactions between pedestrians and their environment. GAT dynamically assigns importance weights to different nodes, thereby effectively aggregating information from neighboring nodes. This enables the model to understand the nuanced patterns of pedestrian interactions within social environments more intricately. This approach not only considers the influence of other pedestrians but also takes into account the impact of the surrounding environment on pedestrian behavior, addressing the limitations of traditional LSTM models in capturing such complex social interactions. Recent studies have shown that in scenarios where the behaviors of each traffic participant are interconnected with surrounding pedestrians and environmental factors, GAT models provide a more comprehensive perspective for pedestrian trajectory prediction by mapping these interactions onto a graph structure.

By employing a Graph Attention Network (GAT), we can acquire information on all pedestrians from time step 1 to T. To synthesize the information of each pedestrian i at every time point $O_i^{1:T}$, the ST-VAE model utilizes a Long Short-Term Memory network (LSTM) to aggregate information across time steps (as indicated in Equation 2).

$$O_i^{1:T} = \text{LSTM}(\hat{m}_i^1, \hat{m}_i^2, \ldots, \hat{m}_i^T) \tag{2}$$

Where $\hat{m}_i^t$ represents the comprehensive hidden state of the i^{th} pedestrian at time t, which is obtained after processing m_i^t (the state information of the i^{th} pedestrian at time step t) through two layers of graph attention.

Another key feature of the ST-VAE model is its ability to generate stochastic predictions, enhancing the model's flexibility and the randomness of its forecasts. This is achieved by introducing latent

variables during the sequence generation process, each of which updates its state via the CGRU network to reflect the complexity of pedestrian movement states and environmental interactions (as indicated in Equation 3).

$$\hat{m_i^t} = \vec{g}\left(\psi_{zd}\left(Z_i^t, d_i^t\right), \hat{m_i^t}\right) \tag{3}$$

The training process of the ST-VAE model follows the standard VAE training objective of maximizing the Evidence Lower Bound (ELBO), while also considering the potential accumulation of errors in the final trajectory generated through displacement sequences. By employing the reparameterization trick with Gaussian distributions and adjustments to the training loss, the model can compensate for predictive errors from previous time steps, thereby enhancing the overall accuracy of the predictions (as indicated in Equation 4).

$$\mathbb{E}_i\left[\frac{1}{H}\sum_{t=T+1}^{T+H}\mathbb{E}_{d_i^t\sim p_\xi(\cdot|z_i^t,\hat{m_i^t})}\mathbb{E}_{z_i^t\sim q_\phi(\cdot|b_i^t,\hat{m_i^t})}\left[\left\|x_i^t - x_i^T - \sum_{\tau=T+1}^{t} d_i^\tau\right\|^2\right]\right] + q_\phi(z_i^t|b_i^t,\hat{m_i^t}) - p_\theta(z_i^t|\hat{m_i^t}). \tag{4}$$

Where p_ξ, q_ϕ, and p_θ are parameterized through network parameters, forming Gaussian distributions.

3.3. AHTRL:Hierarchical Driving Model for Planning

3.3.1. Overview of AHTRL

In designing a hierarchical reinforcement learning model for trajectory planning, this paper utilizes two independent network layers: a high-level selection network and a low-level planning network. This layered approach allows for specific optimizations at different decision-making levels and facilitates modular learning of complex tasks.

A. The high-level selection network

The high-level selection network is responsible for formulating long-term strategies and objectives, functioning to generate sub-goals g_t that guide the lower-level planning network. This network, by observing the current environmental state s_t, generates a high-level action or sub-goal g_t to indicate the macro objective the vehicle aims to achieve. In the context of autonomous driving, these sub-goals could be specific intersection behaviors such as "stop," "turn left," "turn right," or "decelerate." The selection network's responsibility for generating sub-goals g_t can be represented by the following equation:

$$g_t = \pi_O(s_t|\theta_O) \tag{5}$$

Where θ_O represents the parameters of the selection network's policy, and s_t is the current state.

The objective of the policy is to maximize the expected return, and the value function Q_O of the selection network is represented as:

$$Q_O(s_t, g) = \mathbb{E}_{\pi_O}\left[\sum_{k=t}^{T}\gamma^{k-t}r_k|s_t, g_t = g\right] \tag{6}$$

The policy parameters θ_O are updated through gradient ascent:

$$\theta_O \leftarrow \theta_O + \alpha\nabla_{\theta_O}Q_O(s_t, g_t) \tag{7}$$

And updated using the actor-critic method:

$$\pi_P \leftarrow \pi_P + \alpha_P\nabla_{\pi_P}Q_P(s_t, o_t, \pi_P(s_t, o_t)) \tag{8}$$

$$Q_P(s_t, o, p) \leftarrow Q_P(s_t, o, p) + \alpha_P\left(r_t + \gamma Q_P(s_{t+1}, o_t, \pi_P(s_{t+1}, o_t)) - Q_P(s_t, o, p)\right) \tag{9}$$

B. A low-level planning network

The low-level planning network receives the sub-goal g_t passed down from the selection network and generates specific actions a_t based on this. This layer's network focuses on calculating short-term, specific waypoint trajectories that adhere to the vehicle's dynamic constraints, take into account road conditions and obstacle avoidance requirements, to ensure the safety and practicality of the trajectory. The planning network is responsible for outputting actions a_t under the guidance of the given sub-goal g_t, which can be represented by the following equation:

$$a_t = \pi_P(s_t, g_t|\theta_P) \tag{10}$$

Where θ_P represents the parameters of the planning network's policy.

The objective of the policy is to maximize the expected return, and the value function Q_P of the planning network is represented as follows:

$$Q_P(s_t, g, a) = \mathbb{E}_{\pi_P}\left[r_t + \gamma Q_P(s_{t+1}, g, a_{t+1})|s_t, g_t, a_t = a\right] \tag{11}$$

The policy parameters θ_P are updated through gradient ascent:

$$\theta_P \leftarrow \theta_P + \alpha \nabla_{\theta_P} Q_P(s_t, g_t, a_t) \tag{12}$$

Here, γ is the discount factor, α is the learning rate, and r_t represents the immediate reward.

And updated using the actor-critic method:

$$\pi_P \leftarrow \pi_P + \alpha_P \nabla_{\pi_P} Q_P(s_t, o_t, \pi_P(s_t, o_t)) \tag{13}$$

$$Q_P(s_t, o, p) \leftarrow Q_P(s_t, o, p) + \alpha_P\left(r_t + \gamma Q_P(s_{t+1}, o_t, \pi_P(s_{t+1}, o_t)) - Q_P(s_t, o, p)\right) \tag{14}$$

3.3.2. The Neural Network

The network structures of the high-level selector and the low-level planner, as shown in Figure 4, take S_t (i.e., 2D LIDAR images and BEV semantic images processed through a 32×256 preprocessing layer) as environmental inputs. The network architecture includes a self-attention mechanism, an input layer (MLP), a Long Short-Term Memory network layer (LSTM), an output layer (MLP), and a dense layer used for generating action behaviors. The self-attention layer aims to identify the most critical parts within the fused features for the current task, thereby enhancing the precision of the high-level selector and the low-level planner in the goal generation process. The input layer (MLP) further processes the feature vectors, typically integrating multiple fully connected layers and nonlinear activation functions to refine key features and construct a basis for decision-making. The LSTM layer, essential for handling time-series data, is crucial for capturing time-dependent states and actions, such as the vehicle's historical movement trajectories. After processing through the LSTM layer, data flows to the output layer (MLP), which maps features to the action space, laying the groundwork for generating deterministic actions. Finally, the dense layer transforms output layer data into continuous action values, directly guiding the vehicle's movements in the environment.

3.4. Reward function

The design of the reward mechanism is based on several factors, including longitudinal speed, penalties for collisions (in this chapter, collisions between vehicles and pedestrians are specifically identified with $P_{\text{collision}}$ and given the maximum penalty), lane deviation, large steering angles, speeding, and significant lateral acceleration. It is designed based on a similar environmental setup used by Chen et al. [26] in 2019. The overall reward function is as follows:

$$r = \alpha_1 r_{\text{P\_collision}} + \alpha_2 r_{\text{collision}} + \alpha_3 r_{\text{longspeed}} + \alpha_4 r_{\text{exceed}} + \alpha_5 r_{\text{out}} + \alpha_6 r_{\text{steer}} + \alpha_7 r_{\text{latspeed}} + \alpha_8 \tag{15}$$

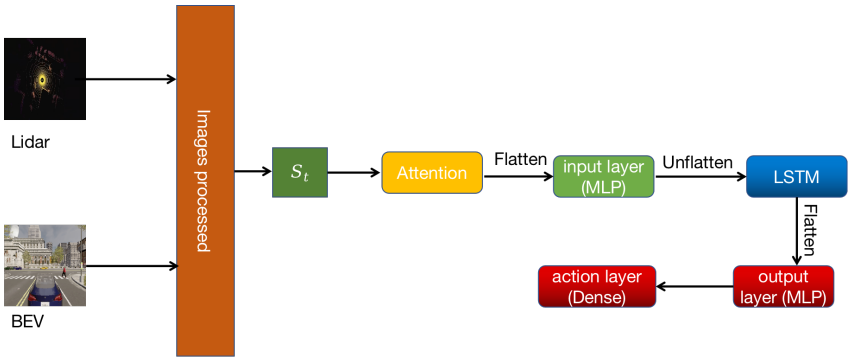

Figure 4: Selection Network And Planning Network Architecture

In the work of this chapter, the corresponding weights for each factor are designed as follows: $\alpha_1 = 1000$, $\alpha_2 = 200$, $\alpha_3 = 1$, $\alpha_4 = 10$, $\alpha_5 = 1$, $\alpha_6 = 5$, $\alpha_7 = 0.2$, $\alpha_8 = -0.1$.

4. Experiments

4.1. Experimental Environment Design

This paper utilized the CARLA simulator [27] in conjunction with the OpenAI Gym interface, based on the setup by Chen et al. [26], to conduct an in-depth study on vehicle motion planning at traffic intersections. Three maps with distinct traffic intersection characteristics (Town02, Town03, and Town04) were selected to comprehensively evaluate the driving capabilities of vehicles in diverse urban environments. These maps encompass a variety of traffic scenarios, ranging from simple T-intersections to complex junctions involving five lanes. In each simulation map, to construct a dynamic and realistic traffic environment, this study not only introduced 100 background traffic vehicles that can interact with the reinforcement learning (RL) controlled agent vehicle but also added 100 pedestrian models active at the intersections. These setups are aimed at enhancing the realism and complexity of the simulation environment, providing a comprehensive and challenging test platform for evaluating the proposed AHTRL model.

In the initial phase of the experiments, this paper conducted a random policy for 10,000 steps on each map to initialize the experience replay buffer. Subsequently, 30,000 steps of training were carried out, enabling the agent vehicle to effectively plan driving paths, avoid collisions, maintain lanes, and appropriately interact with other vehicles and pedestrians in various intersection environments. Moreover, to enhance the vehicle's performance in pedestrian-dense traffic intersections, this study added 100 pedestrian models at each intersection. On top of the original 30,000 steps of training, an additional 10,000 steps of specialized training focused on vehicle-pedestrian interactions were conducted.

This paper particularly emphasizes the interactive planning between vehicles and pedestrians by introducing pedestrian models with diverse behavior patterns. It designs algorithms that enable RL-controlled vehicles to recognize pedestrians, predict their actions, and take appropriate measures to avoid them when necessary, ensuring pedestrian safety. Through this series of experimental setups and training, the approach not only improves the vehicle's motion planning capabilities at intersections but also significantly enhances its safety performance in pedestrian-dense environments.

Ultimately, this paper compared the proposed method with several baseline algorithms, validating its effectiveness in complex traffic environments, particularly in ensuring the safe coexistence of vehicles and pedestrians. This research not only showcases the application potential of hierarchical reinforcement learning technology in the field of autonomous driving but also provides valuable insights and practical guidance for the safe operation of autonomous driving systems in complex urban environments.

In the experiments conducted in this paper, the proposed AHTRL model (hereafter referred to as Our_HRL) was compared with the original Deep Q-Network (hereafter referred to as DQN), H_DQN

[3], the original Deep Deterministic Policy Gradient (hereafter referred to as DDPG), and atHRL [28]. This comparison aimed to demonstrate the performance of the method proposed in this paper among different types of reinforcement learning approaches. Below is a brief overview of these comparative methods and their significance:

1. DQN (Deep Q-Network): A cornerstone algorithm in deep reinforcement learning that combines Q-learning with deep neural networks to handle high-dimensional state spaces. Comparing with DQN allows for evaluating the effectiveness of hierarchical approaches against a foundational deep learning-based method.

2. H_DQN [3]: An extension of the DQN that incorporates hierarchical structures to manage complex decision-making processes by breaking down the problem into manageable sub-tasks. This comparison highlights the advantages of different hierarchical approaches and their efficacy in complex environments.

3. DDPG (Deep Deterministic Policy Gradient): A model-free, off-policy actor-critic algorithm that can operate over continuous action spaces, making it highly relevant for real-world applications like autonomous driving. Comparing with DDPG showcases the benefits of hierarchical modeling in environments where precise control over actions is crucial.

4. atHRL [28]: A state-of-the-art hierarchical reinforcement learning approach that focuses on learning abstract representations and temporal abstractions. This method serves as a benchmark for advanced hierarchical models, allowing for an assessment of the proposed method's novelty and performance improvements.

By comparing Our_HRL with these methods, the paper aims to underscore the improvements in learning efficiency, decision-making quality, and adaptability to complex scenarios brought about by the proposed hierarchical reinforcement learning model, particularly in the context of autonomous vehicle navigation and pedestrian safety.

4.2. Results and Discussion

Figure 5 displays the rewards obtained during the training process by different algorithms, where the reward value comprehensively considers penalties for various factors including collisions with pedestrians, other types of collisions, lane deviations, speeding, as well as large steering angles and high lateral accelerations. Thus, the comparison of reward values effectively reflects the performance of each method. The results indicate that after 30,000 steps of training, the strategy proposed in this study achieved the highest reward among all four comparative methods. Figure 6 further demonstrates that the method proposed by this study performs exceptionally well in pedestrian-dense traffic intersection scenarios, obtaining the highest reward after 10,000 steps of training. Table 1 shows that the method proposed by this study surpasses all other comparison methods in terms of average reward and average speed. While the average reward reflects the overall performance of the agent, comparing average speeds helps to verify the rationality of the vehicle driving strategy. Compared to the optimal baseline model atHRL [28], the method proposed in this study increased the average reward by 11.9% and improved the average speed by 13.0%, indicating that the strategy proposed in this study not only ensures safety but also enhances driving efficiency.

Table 1
algorithm performance

Algorithm	Average Reward	Average Speed (m/s)
DQN	56	2.1
DDPG	141	2.7
H_DQN	183	3.7
atHRL	217	4.6
Our_HRL	**243**	**5.2**

The hierarchical reinforcement learning method proposed in this paper was compared with the original DQN and DDPG algorithms. The results show that the use of a hierarchical planner significantly

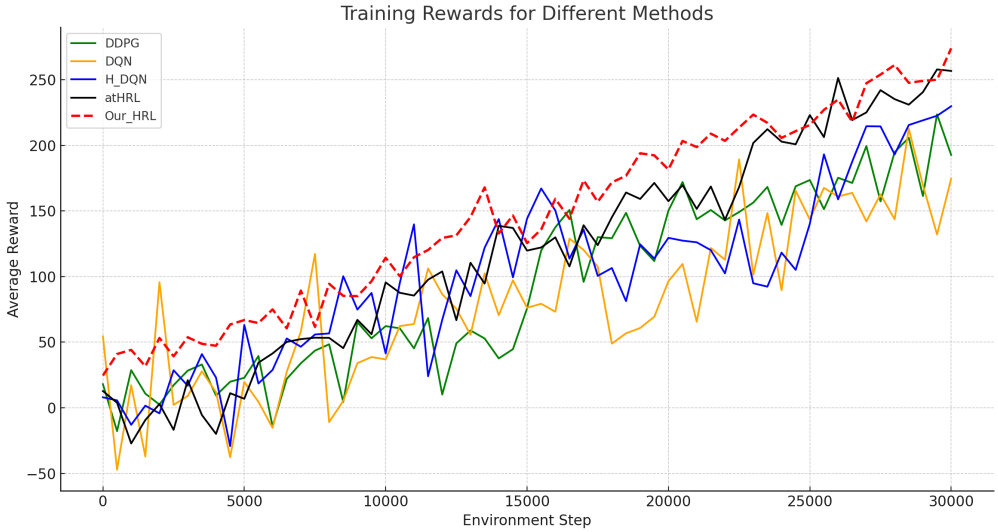

Figure 5: 30,000 routine training average reward

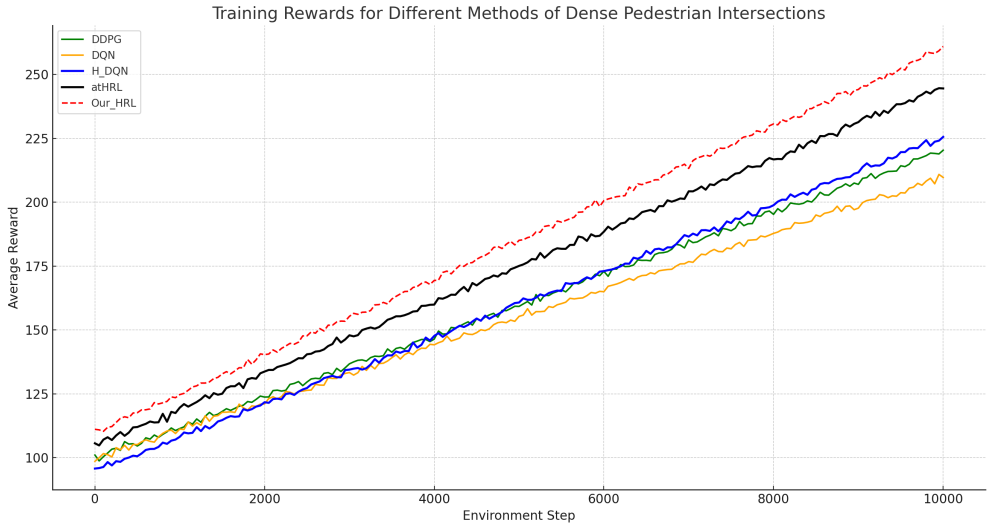

Figure 6: Intensive pedestrian scene 10,000-step training average reward

improved the performance of the reinforcement learning agent in specific urban driving scenarios. Compared to the original DDPG, the average reward increased by 73%, and the average speed of vehicle passage increased by 70.3%. This performance enhancement is not only reflected in a higher overall average reward, indicating that the agent can effectively avoid collisions and ensure safe driving in the simulated environment, but also in a faster average driving speed, demonstrating the agent's enhanced robustness in handling complex scenarios. The introduction of a high-level decision planner enhanced the stability of motion control and reduced the risk of collisions. In contrast, learning low-level control commands directly from observations could lead to unstable control and difficulties in learning strategies for different tasks. Additionally, the method proposed in this paper performed better than H_DQN, which has a similar trajectory planning structure. Although H_DQN is effective for simple decisions, it is insufficient for more complex urban driving task combinations. Compared to the traditional hierarchical DDPG model, atHRL, the average reward increased by 11.9%, and the average speed of

vehicle passage increased by 13%. This indicates that by incorporating pedestrian trajectory prediction, the method in this chapter more effectively managed interactions between pedestrians and vehicles at traffic intersections. Furthermore, the introduction of a self-attention network into the actor-critic model enabled the vehicle to focus on key environmental elements, thereby improving the overall effectiveness of policy learning, leading to superior results in complex urban driving scenarios.

4.3. Safety Analysis

Safety analysis of vehicle trajectory planning is crucial in the development of autonomous driving systems. By comprehensively assessing the system's safety performance under various conditions, this paper gains a deep understanding of the system's behavior and promptly identifies potential safety hazards. This provides essential guidance for system design, such as determining the appropriate planning algorithms, route selection strategies, and parameter settings. Moreover, safety analysis supports the optimization of planning strategies, aiding in the identification and resolution of existing issues and shortcomings within the system. Furthermore, the results of the safety analysis offer targeted recommendations and support for decision-making, thereby ensuring the safe operation of autonomous driving systems in all circumstances. In summary, safety analysis of vehicle trajectory planning is an indispensable part of the autonomous driving system development process, vital for ensuring the system's safety performance, guiding system design and optimization, and supporting decision-making. To evaluate vehicle safety, this paper employs two key metrics: collision rate and success rate.

1. Collision Rate: The percentage of test events in which collisions occur.

2. Success Rate: The percentage of test events where the test vehicle successfully completes its trajectory from start to end without any collisions.

The evaluation of the trained policy consists of 500 test episodes that cover the vehicle safely navigating through various types of traffic intersections, including roundabouts, crossroads, T-junctions, and intersections without traffic lights.

Table 2
Safety Performance Analysis

Algorithm	Collision Rate %	Success Rate %
DQN	3.4	96.6
DDPG	3.2	96.8
H_DQN	2.6	97.4
atHRL	2.2	97.8
Our_HRL	**1.4**	**98.6**

As shown in Table 2, across 500 test scenarios, the algorithm proposed in this paper demonstrated a lower collision rate and a higher success rate. Compared to the best baseline model atHRL, the collision rate decreased by 36.3%. The Our_HRL(AHTRL) algorithm exhibits significant advantages in vehicle trajectory planning over traditional reinforcement learning algorithms. First, Our_HRL(AHTRL) can handle continuous action spaces, producing smoother and more natural trajectories compared to DQN, which enhances driving comfort and safety. Second, by incorporating a hierarchical structure and recurrent neural networks, Our_HRL (AHTRL) achieves more flexible and precise decision-making and more accurately predicts pedestrian behavior compared to DDPG, improving safety when navigating through traffic intersections. Compared to hierarchical DQN (H_DQN) and hierarchical DDPG (atHRL), Our_HRL (AHTRL) not only calculates continuous target waypoints but also adopts a mixed reward mechanism and reward-driven exploration strategy, thereby improving learning efficiency and convergence speed. In summary, the low collision rate and high success rate demonstrated by Our_HRL (AHTRL) in navigating traffic intersections are attributed to its adaptability in continuous action spaces, the stability of deterministic policy gradients, the efficiency of its hierarchical structure, the predictive capability of recurrent neural networks, and the learning efficiency of its mixed reward mechanism.

5. Conclusions and future works

This paper addresses the vehicle trajectory planning problem at complex traffic intersections by proposing a hierarchical reinforcement learning-based model, AHTRL. By decomposing the decision-making process, predicting pedestrian behavior, and integrating the pedestrian trajectory prediction model, ST-VAE, the model effectively enhances vehicle trajectory planning performance in variable traffic environments, significantly reducing collision rates and improving safety and efficiency. Experimental results demonstrate the method's exceptional performance across various traffic intersection environments, particularly in pedestrian-dense traffic intersection scenarios, where it significantly enhances safety and reduces collision risks. These research achievements not only highlight the application value of hierarchical reinforcement learning in autonomous driving technology but also provide new perspectives and methods for future research on trajectory planning in complex traffic environments.

In future research, plans are in place to optimize and extend the AHTRL model proposed in this paper from multiple dimensions. Firstly, efforts will be dedicated to the optimization and improvement of the algorithm by exploring more efficient training strategies and novel reward mechanisms to enhance the algorithm's convergence speed and stability while fine-tuning the balance between safety and efficiency. Secondly, considering the complexity of real-world traffic environments, research on multi-agent collaboration will become a focus. This includes studying interaction strategies between vehicles as well as between vehicles and pedestrians, thereby enhancing the coordination and efficiency of the overall traffic system. Furthermore, transferring the model from simulation environments to real-world applications and conducting tests with actual vehicles will be a critical step in verifying the model's practicality and robustness. At the same time, exploring the model's cross-scenario generalization ability aims to develop more versatile trajectory planning models adaptable to diverse traffic environments and geographical locations. Finally, considering the social impact of autonomous driving technology, future work will also address ethical and responsibility issues in artificial intelligence, ensuring that technological development adheres to moral standards while safeguarding the rights of users and the public. These research directions will not only drive technological progress in the field of autonomous driving but also contribute to realizing a safer and more efficient autonomous driving future.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62166043), the Joint Funds of the National Natural Science Foundation of China (U1603262), and the "Intelligent Information R&D Project" (Project Number: 202104140010). We thank all anonymous commenters for their constructive comments.

References

[1] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, E. Frazzoli, A survey of motion planning and control techniques for self-driving urban vehicles, IEEE Transactions on intelligent vehicles 1 (2016) 33–55.

[2] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, P. Pérez, Deep reinforcement learning for autonomous driving: A survey, IEEE Transactions on Intelligent Transportation Systems 23 (2021) 4909–4926.

[3] K. B. Naveed, Z. Qiao, J. M. Dolan, Trajectory planning for autonomous vehicles using hierarchical reinforcement learning, in: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), IEEE, 2021, pp. 601–606.

[4] Z. Gu, L. Gao, H. Ma, S. E. Li, S. Zheng, W. Jing, J. Chen, Safe-state enhancement method for autonomous driving via direct hierarchical reinforcement learning, IEEE Transactions on Intelligent Transportation Systems (2023).

[5] S. M. LaValle, J. J. Kuffner, Rapidly-exploring random trees: Progress and prospects: Steven m. lavalle, iowa state university, a james j. kuffner, jr., university of tokyo, tokyo, japan, Algorithmic and computational robotics (2001) 303–307.
[6] S. Karaman, E. Frazzoli, Sampling-based algorithms for optimal motion planning, The international journal of robotics research 30 (2011) 846–894.
[7] D. N. Lee, A theory of visual control of braking based on information about time-to-collision, Perception 5 (1976) 437–459.
[8] C. R. Baker, J. M. Dolan, Traffic interaction in the urban challenge: Putting boss on its best behavior, in: 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2008, pp. 1752–1758.
[9] A. Sadat, S. Casas, M. Ren, X. Wu, P. Dhawan, R. Urtasun, Perceive, predict, and plan: Safe motion planning through interpretable semantic representations, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16, Springer, 2020, pp. 414–430.
[10] C. Yan, J. Qin, Q. Liu, Q. Ma, Y. Kang, Mapless navigation with safety-enhanced imitation learning, IEEE Transactions on Industrial Electronics 70 (2022) 7073–7081.
[11] H. Du, Y. Sun, Y. Pan, Z. Li, P. Siarry, A lane-changing trajectory re-planning method considering conflicting traffic scenarios, Engineering Applications of Artificial Intelligence 127 (2024) 107264.
[12] H. Song, D. Luan, W. Ding, M. Y. Wang, Q. Chen, Learning to predict vehicle trajectories with model-based planning, in: Conference on Robot Learning, PMLR, 2022, pp. 1035–1045.
[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing atari with deep reinforcement learning, arXiv preprint arXiv:1312.5602 (2013).
[14] P. Wolf, C. Hubschneider, M. Weber, A. Bauer, J. Härtl, F. Dürr, J. M. Zöllner, Learning how to drive in a real world simulation with deep q-networks, in: 2017 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2017, pp. 244–250.
[15] A. Kendall, J. Hawke, D. Janz, P. Mazur, D. Reda, J.-M. Allen, V.-D. Lam, A. Bewley, A. Shah, Learning to drive in a day, in: 2019 international conference on robotics and automation (ICRA), IEEE, 2019, pp. 8248–8254.
[16] X. Liang, T. Wang, L. Yang, E. Xing, Cirl: Controllable imitative reinforcement learning for vision-based self-driving, in: Proceedings of the European conference on computer vision (ECCV), 2018, pp. 584–599.
[17] Y. Tian, X. Cao, K. Huang, C. Fei, Z. Zheng, X. Ji, Learning to drive like human beings: A method based on deep reinforcement learning, IEEE Transactions on Intelligent Transportation Systems 23 (2021) 6357–6367.
[18] Z. Huang, J. Wu, C. Lv, Efficient deep reinforcement learning with imitative expert priors for autonomous driving, IEEE Transactions on Neural Networks and Learning Systems (2022).
[19] Y. Chen, C. Dong, P. Palanisamy, P. Mudalige, K. Muelling, J. M. Dolan, Attention-based hierarchical deep reinforcement learning for lane change behaviors in autonomous driving, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0.
[20] T. Shi, P. Wang, X. Cheng, C.-Y. Chan, D. Huang, Driving decision and control for automated lane change behavior based on deep reinforcement learning, in: 2019 IEEE intelligent transportation systems conference (ITSC), IEEE, 2019, pp. 2895–2900.
[21] J. Li, L. Sun, J. Chen, M. Tomizuka, W. Zhan, A safe hierarchical planning framework for complex driving scenarios based on reinforcement learning, in: 2021 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021, pp. 2660–2666.
[22] Y. Lu, X. Xu, X. Zhang, L. Qian, X. Zhou, Hierarchical reinforcement learning for autonomous decision making and motion planning of intelligent vehicles, IEEE Access 8 (2020) 209776–209789.
[23] J. Duan, S. Eben Li, Y. Guan, Q. Sun, B. Cheng, Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data, IET Intelligent Transport Systems 14 (2020) 297–305.
[24] L. Gao, Z. Gu, C. Qiu, L. Lei, S. E. Li, S. Zheng, W. Jing, J. Chen, Cola-hrl: Continuous-lattice hierarchical reinforcement learning for autonomous driving, in: 2022 IEEE/RSJ International

Conference on Intelligent Robots and Systems (IROS), IEEE, 2022, pp. 13143–13150.
[25] Z. Xu, Q. Yu, W. Slamu, Y. Zhou, Z. Liu, S-cgru: An efficient model for pedestrian trajectory prediction, in: International Conference on Neural Information Processing, Springer, 2023, pp. 244–259.
[26] J. Chen, B. Yuan, M. Tomizuka, Model-free deep reinforcement learning for urban autonomous driving, in: 2019 IEEE intelligent transportation systems conference (ITSC), IEEE, 2019, pp. 2765–2771.
[27] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, V. Koltun, Carla: An open urban driving simulator, in: Conference on robot learning, PMLR, 2017, pp. 1–16.
[28] X. Lu, F. X. Fan, T. Wang, Action and trajectory planning for urban autonomous driving with hierarchical reinforcement learning, arXiv preprint arXiv:2306.15968 (2023).