Deep Reinforcement Learning Based Decision-Making Strategy of Autonomous Vehicle in Highway Uncertain Driving Environments

Huifan Deng, Youqun Zhao, Qiuwei Wang, Anh-Tu Nguyen

Deep Reinforcement Learning Based Decision-Making Strategy of Autonomous Vehicle in Highway Uncertain Driving Environments

Article *in* Automotive Innovation · August 2023

DOI: 10.1007/s42154-023-00231-6

4 authors:

Huifan Deng
Nanjing University of Aeronautics & Astronautics
21 PUBLICATIONS **134** CITATIONS
SEE PROFILE

You Qun Zhao
Nanjing University of Aeronautics & Astronautics
113 PUBLICATIONS **1,047** CITATIONS
SEE PROFILE

Qiuwei Wang
Nanjing University of Aeronautics & Astronautics
21 PUBLICATIONS **131** CITATIONS
SEE PROFILE

Anh-Tu Nguyen
Université Polytechnique Hauts-de-France
140 PUBLICATIONS **2,084** CITATIONS
SEE PROFILE

Deep Reinforcement Learning Based Decision Making to Improve Multi-Lane Highway Traffic Under Uncertain Driving Environments

Abstract

Dealing with environmental uncertainties and improving vehicle decision-making ability are among the most important issues for autonomous highway driving. To this end, we propose a vehicle decision-making framework based on heuristic reinforcement learning while considering environmental uncertainties. In particular, a future integrated risk assessment model is used to solve the environmental uncertainty. First, this paper predicts the uncertain environment based on a long short-term memory model, including predicting driving intention and vehicle motion trajectory. A future integrated risk assessment model in uncertain environments is proposed in the reinforcement learning framework. Moreover, to solve the exploration and exploitation dilemma in reinforcement learning, a heuristic decaying state entropy (HDSE) deep reinforcement learning algorithm is proposed, which effectively shortens the training time period of the agent. A path tracking model and a rule-based vehicle decision model are built to deal with the path tracking problem and the interaction decision problem of surrounding vehicles. Finally, the vehicle decision framework is validated in both low-density and high-density traffic scenarios. The obtained results show that the proposed vehicle decision-making framework based on HDSE deep reinforcement learning considering environmental uncertainty improves the traffic efficiency while ensuring the vehicle safety.

Keywords: Automated driving, decision making, uncertain driving environment, reinforcement learning, multi-lane traffic, integrated risk assessment.

Abbreviations

AUC	Area under curve
CL	Changing the lane
DARPA	Defense advanced research projects agency
DQN	Deep Q-network
EB	Emergency braking
FCNN	Fully connected neural network
FD	Drive faster
HDSE	Heuristic decaying state entropy
IDM	Intelligent driver model
LC	Change to the left lane
LK	Lane keeping
LSTM	Long short-term memory
MAE	Mean absolute error
MOBIL	Overall braking induced by lane changes
MSE	Mean square error
NGSIM	Next generation simulation
NONE	Maintain speed and lane
POMDP	Partially observable Markov decision process
RC	Change to the right lane
ROC	Receiver operating characteristic
SD	Drive slower
TTC	Time-to-collision

1. Introduction

Over the past two decades, autonomous vehicles have rapidly developed and attracted much attention in academic and industry research [1]. About 94% of traffic accidents are caused by drowsiness, distraction, and poor decision-making, which can be solved by autonomous driving [2]. An autonomous driving system can be divided into four key parts: perception, decision-making, planning and control [3]. A decision-making strategy is regarded as the human brain, which is crucial for autonomous driving [4]. Under uncertain environments, an autonomous vehicle needs to understand the driving intentions of surrounding vehicles to cooperate with them via reasonable driving behaviors. The corresponding behavior decision-making ability largely determines the driving performance of autonomous vehicles.

Vehicle decision-making methods can generally be divided into rule-based and data-driven methods. The most famous rule-based decision-making method can be found in the defense advanced research projects agency (DARPA) Urban Challenge, where the Junior team won the 2005 DARPA championship with a finite state machine [5]; in 2007, the Knigh Rider team used a hierarchical state machine to solve parking task [6]. However, these methods ignore the dynamics and uncertainty of the environment. Moreover, the traditional rule-based systems cannot solve the decision-making problems with many characteristics of driving scenes.

Recently, with the rapid developments of artificial intelligence, using machine learning to solve the vehicle decision-making problem has become a hotspot research topic. The most common one is reinforcement learning, which is different from supervised learning and unsupervised learning. In reinforcement learning, the agent continuously interacts with the environment to improve its performance on

specified tasks, so this learning scheme is very similar to the learning of human beings [7]. Since reinforcement learning has the advantages of active learning and the ability to react to changes in the environment, reinforcement learning-based vehicle decision making has applications in highways [8], roundabouts [9], highway exits [10], on-ramp merging [11], and other scenarios.

Image-based solutions have been used to obtain the information about the vehicle's surrounding environment [12, 13]. However, these observation models (image, radar, or grid-based) are all unstructured datasets and need to be processed by convolutional neural networks. This approach requires a large number of samples and time convergence [14], resulting in a long learning process [15]. Therefore, it is necessary to preprocess the data and use structured datasets to speed up the convergence. For example, Fu et al. proposed an emergency braking strategy based on the deep deterministic policy gradient, which considered efficiency, accuracy, and passengers' comfort in the reward function [16]. Hoel et al. [17] extended the AlphaGo Zero algorithm to the continuous state-space domain and applied it to the field of autonomous driving decision-making. In [18], the dueling deep Q-network algorithm was used to solve the decision-making problem under the highway driving conditions.

There are multiple uncertainties in the driving process of autonomous vehicles. Most papers use the partially observable Markov decision process (POMDP) to solve the vehicle decision-making problem under an uncertain environment. Hubmann et al. [19] adopted POMDP to solve the vehicle decision-making problem at the intersection by considering the uncertainty of the intention of surrounding vehicles and the uncertainty of sensors. However, the designed action space is only longitudinal action, which is not extended to the lateral motion of the vehicle, so it is not suitable for highway decision-making. Zhang et al. [20] combined the POMDP model with heuristics and proposed a probabilistic modeling framework to explain environmental uncertainty. Galceran et al. [21] proposed a multi-strategy decision framework based on POMPD, which models the vehicle behavior of the self-vehicle and surrounding vehicles as a set of discrete closed-loop policies. However, due to insufficient samples of intentions and inaccurate initial behavior predictions, the risk-aware objective may not reflect the policy evaluation [22]. POMPD provides a general mathematical framework to solve uncertainty problems. However, due to the "curse of dimensionality" issue, POMDP is challenging in terms of computation [23]. Although various simplification and discretization methods have been adopted, the existing methods are not efficient enough to deal with highly dynamic driving scenarios [24].

Autonomous vehicles are faced with many unavoidable uncertainties, such as the uncertainty of surrounding vehicles' behavior, the uncertainty of surrounding vehicles' motion, and the uncertainty of interaction between self-vehicle and surrounding vehicles [25]. These complex and uncertain factors are unavoidable and widespread in highway environments, bringing severe challenges to autonomous vehicles' behavior decision-making system. An unreasonable decision-making behavior lead to significant traffic accidents.

The vehicle observation space describes the information obtained by the autonomous vehicle. It generally includes the state of the vehicle (position, speed and heading angle), the topology information, and other traffic participants, e.g., surrounding vehicles, obstacles [26]. The selection of the vehicle state space is critical. Table 1 shows the state space and the corresponding action space in related works. In the observation space, v_{ego} is the speed of the self-vehicle, $lane_{ego}$ is the lane index of the self-vehicle, v_{des} is the ideal speed of the self-vehicle, $behavior_{ego}$ is the behavior of the self-vehicle, $heading_{ego}$ is the heading angle of the self-vehicle, $lane_s$ is the lane index of the surrounding vehicles, dx_s and dy_s are the relative longitudinal and lateral distances between the surrounding vehicles and the self-vehicle, respectively. dv_s and dvy_s are the relative longitudinal and lateral speeds of the surrounding vehicles and the self-vehicle, x_s, y_s, v_s and $heading_s$ are the longitudinal displacement, lateral displacement, speed, and heading angle of the self-vehicle, respectively. Moreover, *LK* represents the lane keeping, *LC* is the change to the left lane, *RC* is the change to the right lane, *FD* is to drive faster, *SD* is to drive slower, *EB* represents the emergency braking, *CL* represents the action of changing the lane, and *NONE* corresponds to maintain the speed and the lane.

From Table 1, it can be seen that the state of the self-vehicle and the surrounding vehicles are chosen as the observation space in most of related works. Motivated by the above issues, we propose a new decision-making framework for autonomous driving considering the uncertainty of driving environments. Moreover, a new future integrated risk assessment state model is proposed as the observation space. How to obtain an efficient exploration is one of the main challenges in reinforcement learning [27]. Moreover, the essence of the exploration and exploitation dilemma is how to make the algorithm achieve better convergence in a limited time. Therefore, we propose a novel exploration method to balance the exploration and the exploitation. Specifically, the main contributions of this paper are summarized as follows.

Table 1. Selection of the state space and the corresponding action space in related works.

Method	Problem	Observation Space	Action Space
Alizadeh et al. [28]	Highway	ego: $\{v_{ego}, lane_{ego}\}$ surrounding: $\{lane_s, dx_s, dy_s, dv_s\}$	*Discrete*: $\{LK, LC, RC\}$
Ye et al. [29]	Vehicle-following	ego: $\{v_{ego}\}$ surrounding: $\{dx_s, dv_s\}$	*Continuous*: $\{a_x\}$
Hoel et al. [30]	Highway	ego: $\{v_{ego}, lane_{ego}\}$ surrounding: $\{lane_s, dx_s, dv_s\}$	*Discrete*: $\{LC, RC, FD, LK, SD, EB\}$
Xu et al. [8]	Highway	ego: $\{v_{ego}, lane_{ego}, v_{des}\}$ surrounding: $\{lane_s, dx_s, v_s\}$	*Discrete*: $\{LK, LC, RC\}$
Wolf et al. [31]	Highway	ego: $\{behavior_{ego}, v_{ego}, lane_{ego}\}$ surrounding: $\{lane_s, x_s, y_s, v_s, heading_s\}$	*Discrete*: $\{LC, RC, FD, SD\}$
Nageshrao et al. [32]	Highway	surrounding: $\{dx_s, dy_s, dv_s, dvy_s\}$	*Discrete*: $\{LK, LC, RC, FD, NONE, SD, EB\}$
Aradi et al. [33]	Highway	ego: $\{v_{ego}, y_{ego}, heading_{ego}\}$ surrounding: $\{dx_s, dy_s, dv_s\}$	*Discrete*: $\{LC, RC, FD, SD\}$
Yu et al. [34]	Highway	ego: $\{v_{ego}, lane_{ego}\}$ surrounding: $\{dx_s, dy_s, dv_s\}$	*Discrete*: $\{LK, CL\}$

1) Differently from [18, 35], the design of the new decision-making framework for autonomous driving takes into account the environmental uncertainty in terms of driving intention of surrounding vehicles, future driving risk in each lane, and interaction between surrounding vehicles and self-vehicle.

2) A heuristic decaying state entropy (HDSE) deep reinforcement learning algorithm is proposed to solve the well-known exploration and exploitation dilemma.

3) A future integrated risk assessment model is developed to effectively deal with environmental uncertainties.

This paper is organized as follows. In Section 2, the algorithm framework is presented. Moreover, the uncertain driving environment is discussed, including the prediction of lane change intention, the vehicle motion trajectory, and the future integrated risk assessment model. In Section 3, the problem of vehicle decision-making in an uncertain environment is proposed, and the HDSE deep reinforcement learning algorithm is designed to solve the Markov problem. Section 4 shows the relevant results of a series of case studies. Section 5 concludes the paper.

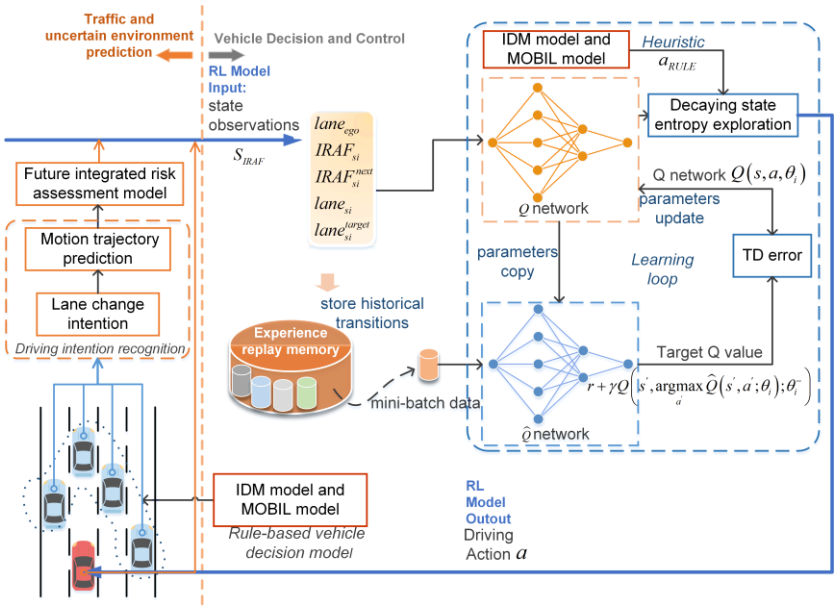

Fig. 1 System framework for vehicle decision system.

2. System Framework and Prediction of Uncertain Driving Environment

2.1. System Framework

The proposed vehicle decision-making framework is illustrated in Fig. 1, which can be divided into uncertain environment prediction and vehicle decision system. The driving intention identification module identifies the lane-changing intention and the vehicle motion trajectory of the surrounding vehicles based on long short-term memory (LSTM). The future integrated risk assessment function and the future target lane of surrounding vehicles are calculated using a prediction of the driving environment and the integrated risk assessment function, see Section 2.4. The result is used as the input of the vehicle decision system together with the current state of the vehicle and the surrounding vehicles. In addition, the data obtained from the interaction with the environment is stored in the experience replay memory and are extracted to input the reinforcement learning model during the training process. The vehicle decision system adopts the HDSE deep reinforcement learning algorithm to learn the optimal driving behavior by interacting with the driving environment. In particular, the optimal action is heuristically generated by a_{RULE}, which is determined by the intelligent driver model (IDM) in [36] and minimizes the overall braking induced by lane changes (MOBIL) [37]. By this way, the vehicle decision system obtains optimal decision actions considering environment uncertainties, which include different driving actions, e.g., change to the left lane, change to the right lane, drive faster, drive slower, maintain speed and lane.

2.2. Lane Change Intention Prediction

Driving intent prediction is one of the core technologies of intelligent vehicles, which can infer the future intent of the driver to predict the likelihood of a possible collision and also take measures to avoid accidents in advance [38]. This paper uses data from the Next Generation Simulation (NGSIM) program conducted by the Federal Highway Administration for lane change intention prediction and driving trajectory prediction [39]. The NGSIM project was undertaken to study microscopic vehicle driving behavior and included the US-101 and I-80 highway datasets. For this paper, 543 lane change scenarios and 870 lane-keeping scenarios were extracted to construct the dataset. We used 11 data items from the NGSIM data, including dataset id, Vehicle id, Frame index, Local X, Local Y, Lane id, v_length, v_Width, v_Class, Velocity, Acceleration. Moreover, the dataset is divided into a training set and a test set with a ratio of 8:2. Each dataset has a trajectory duration of 8 s, and here we use 4 s of track history and a 4 s prediction horizon. The inputs of both the lane change intention prediction and the vehicle motion trajectory prediction can be expressed as

$$\boldsymbol{I}^{(t)} = \left[\boldsymbol{S}_{\mathrm{e}}^{(t)}, \boldsymbol{E}_{si}^{(t)} \right] \quad t = \left(T - T_{\mathrm{p}}, \cdots, T-1, T \right) \tag{1}$$

where $\boldsymbol{S}_{\mathrm{e}}^{(t)} = \left\{ x_{ego}^{(t)}, y_{ego}^{(t)}, v_{ego}^{(t)}, lane_{ego}^{(t)} \right\}$ is the history information for the predicted vehicle, $\boldsymbol{E}_{si}^{(t)} = \left\{ dx_{si}^{(t)}, dv_{si}^{(t)}, lane_{si}^{(t)} \right\}$ is the history information for the surrounding vehicle, T_p is the history time horizon. Moreover, the output of the intention prediction module is the probability of changing lanes P_{lane_change}.

Since the driving intention is influenced by various

potential factors, we choose the LSTM model with a high nonlinear fitting ability and data feature extraction capability for its prediction. After the training periods, the lane change decision model summarizes the lane change decision behavior's pattern in the NGSIM data, predicting the driver's lane change or lane-keeping choice under current traffic environments. We selected the same training and test sets from the NGSIM dataset to test the intention recognition model under the fully connected neural network (FCNN), and LSTM network, respectively, and the results are shown in Table 2. The structure of both the FCNN network and the LSTM network have the same number of network nodes. The structure of the LSTM network is shown in Fig. 2, where Relu and Sigmoid are the activation functions, and the FC means the fully connected layers. The Area Under Curve (AUC) is the area under the Receiver Operating Characteristic (ROC) curve, which is the evaluation index of the two-class classification. F1-score is the harmonic mean of precision and recall. As can be seen from Table 2, the accuracy of the LSTM method is higher than that of the neural-network-based method. Moreover, the accuracy and loss evolutions during the training process are shown in Fig. 3.

Table 2. Performance comparison for lane change intention.

Method	Actual classification	Predicted lane changing	Predicted lane keeping	Accuracy	Total accuracy	F1	AUC
LSTM	lane changing	267	9	96.7%	96.8%	0.878	0.97
	lane keeping	65	1978	96.8%			
FCNN	lane changing	262	14	94.9%	91.5%	0.670	0.91
	lane keeping	244	1799	88.1%			

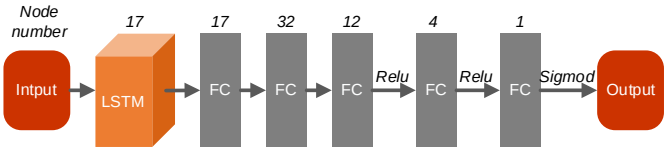

Fig. 2. Network structure of the LSTM method in lane change intention prediction.

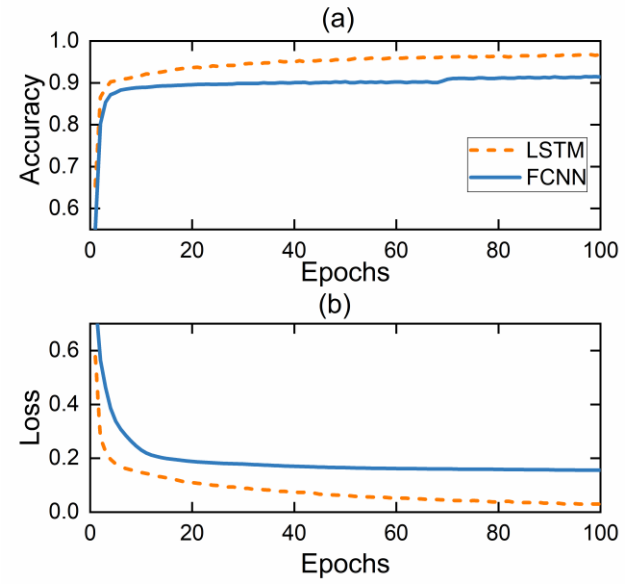

Fig. 3. Accuracy and loss evolutions during the training process.

2.3. Vehicle Motion Trajectory Prediction

The vehicle motion trajectory is generally considered a time series. Hence, we built a vehicle motion trajectory prediction based on the LSTM model which is suitable for time series prediction. The network structure of the model is shown in Fig. 4, where R-dropout 0.2 means that the dropout probability. With the particular structure of LSTM, the cell states are transferred to the subsequent training with minimal time loss, which can avoid the gradual decay of the leading output. The output of the trajectory prediction module is defined as

$$\boldsymbol{O}^{(t)} = \left[x_{ego}^{(t)}, y_{ego}^{(t)} \right], \quad t = \left(T+1, T+2, \cdots, T+T_f \right) \tag{2}$$

where T_f is the predicted time horizon.

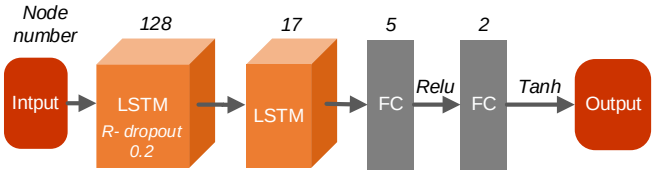

Fig. 4. Network structure of the LSTM method in lane change trajectory prediction.

After the vehicle motion trajectory prediction training is completed, we randomly selected 10 different initial positions and vehicle speeds for testing from the test set, and the results are shown in Table 3. As shown in Table 3, the proposed model can effectively predict the vehicle motion trajectory. The mean absolute error (MAE) of the lateral position is significantly better than that of the longitudinal position, and the prediction accuracy decreases as the prediction time horizon increases. The prediction results of both root mean square error (RMSE) and MAE can meet the requirements of motion trajectory prediction.

Table 3. Vehicle nomenclature.

Prediction horizon	RMSE (m)	MAE of *x* (m)	MAE of *y* (m)
1	0.70	0.45	0.15
2	1.79	1.20	0.32
3	3.22	2.37	0.40
4	5.57	3.56	0.51

2.4. Future Integrated Risk Assessment Model

Time-to-collision (TTC) is commonly used for risk assessment, but it is inaccurate to only use TTC for risk assessment [40]. Hence, this paper defines a newly integrated risk assessment function based on the reference [41] to more comprehensively assess the future risk. Differently from TTC, the integrated risk assessment function considers both longitudinal and lateral risks of the vehicle. The vehicle nomenclature is given in Table 4. According to the vehicle kinematics model, the lateral risk and the longitudinal risk are expressed as

Table 4. Vehicle nomenclature.

Notation	Description
Δy	Relative lateral position of the self-vehicle and the surrounding vehicle
c_0	Distance between the self-vehicle's distance sensor and the side of the vehicle
v_e	Vehicle speed of the self-vehicle
φ_e	Heading angle of the self-vehicle
v_s	Vehicle speed of the surrounding vehicle
φ_s	Heading angle of the surrounding vehicle
Δx	Relative longitudinal position of the self-vehicle and the surrounding vehicle
a_0	Distance between the self-vehicle's distance sensor and the leading vehicle
τ_1	Reaction time of the driver
d_s	Deceleration of the self-vehicle and surrounding vehicle
v_e^{next}	Self-vehicle speed at the next time
v_{sx}^{next}	Longitudinal vehicle speeds at the next time
v_{sy}^{next}	Lateral vehicle speeds at the next time

$$\zeta_y = \frac{\Delta y - c_0}{v_s \sin\varphi_s - v_e \sin\varphi_e} \tag{3}$$

$$\zeta_x = \frac{v_e \cos\varphi_e (\tau_1 + \tau_2)}{D_{es}(t)} + \frac{(v_e \cos\varphi_e)^2}{2 d_e D_{es}(t)} - \frac{(v_s \cos\varphi_e)^2}{2 d_s D_{es}(t)} d_e \tag{4}$$

where $D_{es}(t) = (\Delta x - a_0) + (v_s \cos\varphi_s - v_e \cos\varphi_e)\zeta_y(t)$. The specific derivations for vehicle longitudinal and lateral risks can be found in [41], which is omitted here for brevity. Combining lateral risk (3) and the longitudinal risk (4), the integrated risk assessment function can be defined as follows, where the exponential and weight function is added to make the threat risk obvious.

$$IRAF = S_r \exp(\zeta_x(t)) \cdot \left(1 + w_l \left(\frac{t_b}{\zeta_y(t)}\right)\right) \tag{5}$$

where $t_b = \frac{v_{ego}}{2g}$, w_l is the weight of the lateral risk. The deceleration of automobiles is 7~26ft/s2, according to the Transportation Research Board [42]. Therefore, we select the values of d_e and d_s as 0.75*g*, where *g* is the gravitational acceleration. The system reaction time τ_1 is around 0.3~1.2s, here we take 0.7s. The brake reaction time τ_2 is about 0.15s, $S_r = 2.03$ is selected according to the road conditions. Moreover, we can obtain the future integrated risk assessment with the predicted results for the uncertain environment. With the predicted trajectories Δx_s^{next} and Δy_s^{next}, the predicted future lateral risk is expressed as

$$\xi_y^{next} = \frac{\Delta y^{next} - c_0}{v_{sy}^{next} - v_e^{next} \sin\varphi_e} \tag{6}$$

where $\Delta y^{next} = \Delta y + \Delta y_s^{next} - v_e^{next} \sin\varphi_e \Delta t$, $v_e^{next} = v_e - d_e \Delta t$, $v_{sy}^{next} = \Delta y_s^{next} / \Delta t$, Δt is the time interval from the current time to the next moment. Similarly, the predicted future longitudinal risk is expressed by

$$\begin{aligned} \xi_x^{next} &= \frac{D_{sf}^{next}}{D_{es}^{next}} \\ &= \frac{v_e^{next} \cos\varphi_e \tau + \left(v_e^{next} \cos\varphi_e\right)^2 / 2 d_e}{D_{es}^{next}} - \frac{\left(v_{sx}^{next}\right)^2 / 2 d_s}{D_{es}^{next}} \end{aligned} \tag{7}$$

with $D_{es}^{next}(t) = \left(v_s^{next} \cos\varphi_s - v_e^{next} \cos\varphi_e\right)\xi_y^{next}(t) + \Delta x^{next} - a_0$, $\Delta x^{next} = \Delta x + \Delta x_s^{next} - v_e^{next} \cos\varphi_e \Delta t$ and $v_{sx}^{next} = \Delta x_s^{next} / \Delta t$. Combining the predicted future lateral and longitudinal risk expressions (6) and (7), the future integrated risk assessment function can be defined as

$$IRAF^{next} = S_r \exp\left(\xi_x^{next}\right)\left(1 + w_l \left(\frac{t_b^{next}}{\xi_y^{next}}\right)\right) \tag{8}$$

with $t_b^{next} = v_e^{next} / 2g$.

3. Heuristic Decaying State Entropy Deep Reinforcement Learning Algorithm

This section proposes a vehicle decision-making algorithm based on HDSE deep reinforcement learning in uncertain environments. First, we establish commonly used vehicle decision models and uncertainty vehicle decision models that consider the future vehicle uncertainty based on the future integrated risk assessment function (8). Then, based on these

models, the decision-making problem are solved by the HDSE deep reinforcement learning algorithm.

3.1. Vehicle Decision Model

We describe the vehicle decision problem as a Markov decision process, which is generally defined as a tuple (*S*, *A*, *P*, *R*, γ). This tuple is composed of a state space *S*, an action space *A*, a state transition probability *P*, a reward function *R*, and a discount factor γ which is used to calculate the cumulative reward of the whole process.

1) Vehicle Actual Motion State Model: To reflect its representativeness, the first observation space of this paper is selected as

$$S_{actual} = \{x_{ego}, y_{ego}, lane_{ego}, d_{si}, dv_{si}, lane_{si}\},\ i = 1,2,3,4. \quad (9)$$

Since the surrounding vehicles are not always present around the observable range when the vehicle is running. To ensure that the same number of features is input to the deep neural network, assume that the *i*th vehicle is at the observable range boundary when it is absent, i.e., d_{si}=200 and dv_{si}=30.

2) Future Integrated Risk Assessment State Model: The state space should be as efficient and straightforward as possible to effectively reduce the training difficulty and to improve the algorithm's performance. Compared with the actual vehicle motion state model, we propose an observation space model that considers the intention of surrounding vehicles and the future risk assessment of vehicles. Based on the future integrated risk assessment model, we define an observation space that takes into account the driver intention and the future risk assessment as

$$S_{IRAF} = \{lane_{ego}, IRAF_{si}, IRAF_{si}^{next}, lane_{si}, lane_{si}^{target}\}, \quad (10)$$

for $i = 1,2,3,4$. $lane_{si}^{target}$ is the driver's intention, $IRAF_{si}$ is the integrated risk assessment function defined in (5), $IRAF_{si}^{next}$ is the future integrated risk assessment function calculated from (8), and *i* represents the leading vehicle on the *i*th lane. Similarly, if there is no leading vehicle in on the *i*th lane, it is assumed at the boundary of the observation range, i.e., $IRAF = 0$, $IRAF_{si}^{next} = 0$ and $lane_{si}^{target} = lane_{si}$.

3) Action Space: To improve the convergence speed, the action space of the reinforcement learning output is the vehicle's behavioral decisions. Then, the vehicle's action is defined as the action space that solves all highway driving tasks, that is

$$A = \{LC, RC, FD, SD, NONE\} \quad (11)$$

To ensure the predictability of the vehicle behavior decision, the interval between two behavior decisions of the vehicle will not be too short, the execution interval is 1s.

4) Reward Function: Complex reward functions, e.g., to guarantee the distance between the ego-vehicle and the vehicle ahead, may reduce the agent's exploration ability [4]. Hence, the goal here is to use straightforward reward functions to solve the vehicle decision-making problem. To this end, we construct the following rewards to guarantee a high-speed driving with collision avoidance:

$$R(s,a)_{all} = R(s,a)_{speed} + R(s,a)_{collision} \quad (12)$$

$$R(s,a)_{speed} = C_{speed}\frac{v - v_{\min}}{v_{\max} - v_{\min}} \quad (13)$$

$$R(s,a)_{collision} = \begin{cases} C_{collision} & \text{vehicle collision} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where *s* and *a* represent the current state and action, C_{speed} is the reward constant for vehicle speed in the desired range, $C_{collision}$ is the penalty factor for a vehicle collision.

3.2. Heuristic Decaying State Entropy Dueling Double DQN

The agent interacts and explores with the environment in real-time and directly learns from the obtained experiential data to eventually maximize the cumulative returns or to achieve a specific goal. However, the high dimensionality of the input features and the computational complexity bring challenges to traditional reinforcement learning methods [43]. Mnih et al. of the Google DeepMind team proposed a deep Q-Network (DQN) algorithm based on convolutional neural networks [44], which solves the challenges of traditional reinforcement learning by exploiting deep learning features to fit functions and characterize learning properties nonlinearly. Then, double-DQN, which allows avoiding overestimation, and Dueling DQN, based on Advantage Learning, have been proposed to bring new developments to deep reinforcement learning. Meanwhile, Prioritized Experience Replay improves the learning efficiency of the agents. Moreover, existing vehicle decision models can provide heuristic guidance for agents and avoid meaningless exploration. In this paper, a deep reinforcement learning algorithm combines the above advantages and a HDSE deep reinforcement learning is proposed for vehicle decision making.

In reinforcement learning, the agent's goal is to maximize the expectation of the cumulative reward. The value function uses the expectation of return to evaluate the agent's performance under the current state or a specific state and action. To solve the reinforcement learning problem, we define the following optimal state action-value function $Q^{\pi}(s_t, a_t)$ under optimal strategy π as the maximum expected return obtained for a specific action in a specific state:

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}\left[r + \gamma \max_{a'} Q^{\pi}(s', a')\right] \quad (15)$$

where *s*, *a* and *r* are the state, action and reward of each period respectively, s' is the state of the next period, a' is the

action of the next period, γ is the discount factor, $\mathbb{E}$ is the expected value. The optimal strategy is obtained by maximizing the optimal action-value function as

$$\pi(s) = \arg\max_a Q(s,a) \tag{16}$$

The observation space is generally a continuous high-dimensional state space for the autonomous vehicle decision problem. Hence, we use a parameterized continuous function $Q(s,a,\theta)$ to estimate the state-action value, where θ is the parameter of the evaluate *Q*-network. There are two neural networks in the DQN algorithm, i.e., the evaluate *Q*-network and the target *Q*-network $\hat{Q}$. The parameters of the evaluate *Q*-network are obtained by minimizing the loss function $L(\theta)$

$$L(\theta) = \mathbb{E}_{(s,a,r,s')}\left[\left(y^Q - Q(s,a,\theta)\right)^2\right] \tag{17}$$

where $y^Q = r + \gamma \max_a Q(s',a',\theta^-)$, the parameter of the target *Q*-network θ^- is copied from the *Q*-network at a certain number of iterations during the training process. However, the target *Q*-network y^Q is prone to overestimation of values, then the target y^Q is replaced with

$$y^{DDQN} = r + \gamma Q\left(s', \arg\max_a Q(s',a,\theta), \theta^-\right). \tag{18}$$

Here, the actions are selected based on the evaluate *Q*-network, and then the *Q*-value is determined based on the target *Q*-network. The double DQN uses two parameters θ and θ^-, where the parameter θ is used to select the action, and the parameter θ^- is used to evaluate the state-action value function. The new loss function can be defined from (18) as

$$L(\theta) = \mathbb{E}_{(s,a,r,s')}\left[\left(y^{DDQN} - Q(s,a,\theta)\right)^2\right] \tag{19}$$

Minimizing the loss function (19), we obtain the parameters of the evaluate Q-network. To estimate more accurately the Q values, we use dueling networks to divide the Q-network into two parts, i.e., the value function part and the advantage function part. The Q-value function is defined as

$$Q(s,a) = V(s) + A(s,a) \tag{20}$$

where $V(s)$ is the value function, $A(s,a)$ is the advantage function. The detailed structure of the Q-network established in this paper is shown in Fig. 5. The network input is the future integrated risk assessment state S_{IRAF} described in (10), and the output is the state-action value function $Q(s,a)$.

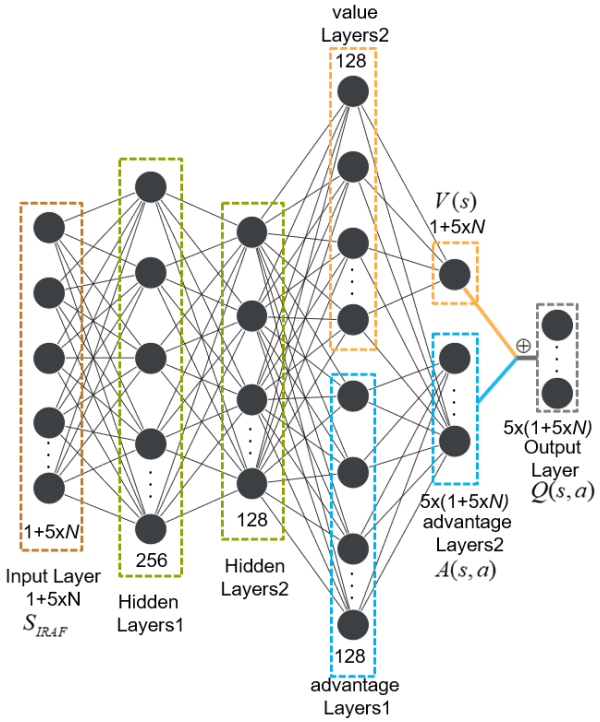

Fig. 5. Structure of the Q-network.

The traditional ε-greedy exploration has an equal probability for each exploration in the same episode, without considering that the weights of exploration should be varied for different values of $Q(s,a)$. Here, we define the action probability distribution $p_s(a)$ based on $Q(s,a)$ as

$$p_s(a) = \frac{\exp\left(Q(s,a) - \max_{\tilde{a}\in A} Q(s,\tilde{a})\right)}{\sum_{b\in A}\exp\left(Q(s,b) - \max_{\tilde{a}\in A} Q(s,\tilde{a})\right)} \tag{21}$$

where *A* is the action space. The entropy $H(s)$ of a discrete random variable *s* with probability distribution $p_s(a)$ is defined as

$$H(s) = -\sum_{a\in A} p_s(a)\log_{|A|} p_s(a). \tag{22}$$

Differently from [45], to make the probability of entropy converge to 0 at the end of the training, we use the decaying method. The entropy processed by the decaying method can be expressed as

$$H_{de}(s) = H(s)\cdot\left(t_f + (1-t_f)\cdot\exp\left(-\frac{n_r}{\tau}\right)\right) \tag{23}$$

where $H_{de}(s)$ represents the post-decay entropy, n_r is the cumulative number of runs, τ and t_f are constants that denote the post-decay convergence value and the decay speed, respectively. The algorithm for action selection based on the decaying state entropy method is shown in Algorithm 1,

where p_{rule} is the probability of choosing a rule-based algorithm, a_{RULE} is the rule-based action based on IDM model and MOBIL model, and a_{Random} chosen uniformly randomly from $A=\{LC, RC, FD, SD, NONE\}$.

Algorithm 1. Action selection strategies

Input: Value function $Q(s,a)$

1. Compute the probability distribution of action $p_s(a)$ (21)
2. Compute the entropy of state S $H(s)$ (22)
3. Compute the entropy of state S $H_{de}(s)$ (23)
4. Choose an action strategy randomly from the set of action strategies $\left\{\arg\max_a Q(s,a), a_{RULE}, a_{Random}\right\}$ with distribution $\left\{1-H_{de}(s), \frac{H_{de}(s)}{p_{rule}}, H_{de}(s)\frac{p_{rule}-1}{p_{rule}}\right\}$
5. **if** the action strategy is $\arg\max_a Q(s,a)$ **then**
6. $a=\arg\max_a Q(s,a)$
7. **else if** the action strategy is a_{RULE} **then**
8. $a=a_{RULE}$
9. **else**
10. $a=a_{Random}$
11. **end**

Output: action a

Remark 1. The proposed heuristic decaying state entropy exploration method is not limited to using the rules based on IDM model and MOBIL model. It can be equally heuristic by experienced drivers or other driving models.

For the above method, we established the HDSE reinforcement learning algorithm for dueling double DQN, to improve the agent learning efficiency and to predict the value function more accurately.

4. Case Studies

To validate the proposed decision-making algorithm for autonomous vehicles, we provide two different cases to analyze the characteristics of our proposed algorithm from different perspectives. In addition, comparative studies between the proposed vehicle decision-making method and surrounding vehicles decision-making methods in different scenarios are also presented.

4.1. Experiment Setting and Comparative Groups

We adopt Breadth-First Search in the graph description of the road network to generate the shortest path from the initial point to the destination and transfer it to the path tracking model. An MPC controller is used for path tracking whereas the speed tracking is performed by an PI controller.

Both the proposed algorithm and the compared algorithm are trained and evaluated in the highway-env on the python platform [46]. The initial states of the algorithms during training and testing are random, including the initial position and velocity of the self and surrounding vehicles. And the random range of initial velocity is 23-25m/s. In order to compare the performance of different algorithms, we selected the same random seeds for testing to ensure the uniformity of the test results. After generating the random initial state, the behavior of the surrounding vehicles is decided based on the current state. Specifically, the longitudinal and lateral decision models are IDM and MOBIL models, respectively.

To increase the complexity of the case, we set the number of lanes as four, which is more complex than the three lanes and can further verify the reliability of the algorithm. And the initial speed of the self- vehicle is 25m/s. As for the hyperparameters of the dueling double DQN, the learning rate is 0.0005, and the reward discount factor γ is 0.99. The agent only acquires the signal of the vehicle ahead with a range of 180m. For the same scenario, the four following strategies are tested and compared.

1) UHDSE: *HDSE* reinforcement learning algorithm for dueling double DQN considering environment uncertainties. The observation space is the future integrated risk assessment state model (10), and the reinforcement learning model is the HDSE-dueling double DQN.

2) UDDDQN: Dueling double DQN reinforcement learning algorithm considering environment uncertainty. The observation space is the future integrated risk assessment state model (10) in uncertain environments, and the reinforcement learning model is the dueling double DQN.

3) DDDQN: Dueling double DQN reinforcement learning algorithm without considering environmental uncertainties. The observation space is the vehicle actual motion state model (9), and the reinforcement learning model is the dueling double DQN.

4) RULE: Rule-based vehicle decision-making method. The longitudinal and lateral decision models are IDM and MOBIL models, respectively.

To show the effectiveness of the risk assessment model proposed in this paper on the vehicle decision problem, we compare the UDDQN algorithm with the DDDQN algorithm. The risk assessment model was used in the UDDQN algorithm, while it is not used in the DDDQN algorithm. Similarly, we show the difference before and after accounting the uncertainty by comparing the performance of UDDQN and DDDQN algorithms.

4.2. Case 1: Low-Density Traffic Flow

For this case, the four methods (UHDSE, UDDDQN, DDDQN and RULE) are trained and tested in a low-density scenario simulating a normal highway driving situation,

which is depicted in Fig. 6. The low-density traffic flow means that the average headway is 28m. The corresponding simulation results are shown in Fig. 7.

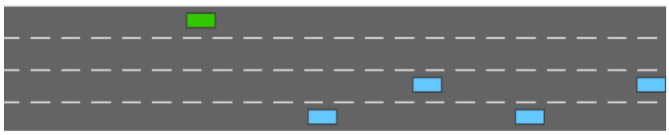

Fig. 6. Diagram of traffic flow for low-density cases: self-vehicle (in green) and surrounding vehicles (in blue).

Table 5. Test data set results in low traffic flow.

Items	UHDSE	UDDDQN	DDDQN	RULE
Average Total Reward	25.14	23.42	22.35	20.16
Crash percentage	0.12	0.24	0.32	0.40
Average vehicle speed	22.33	23.04	21.90	22.13
Lane change maneuver percentage	42.4%	31.5%	37.4%	12.8%
Longitudinal maneuver percentage	57.6%	68.5%	62.6%	81.9%
None action percentage	0.0%	0.0%	0.0%	5.3%

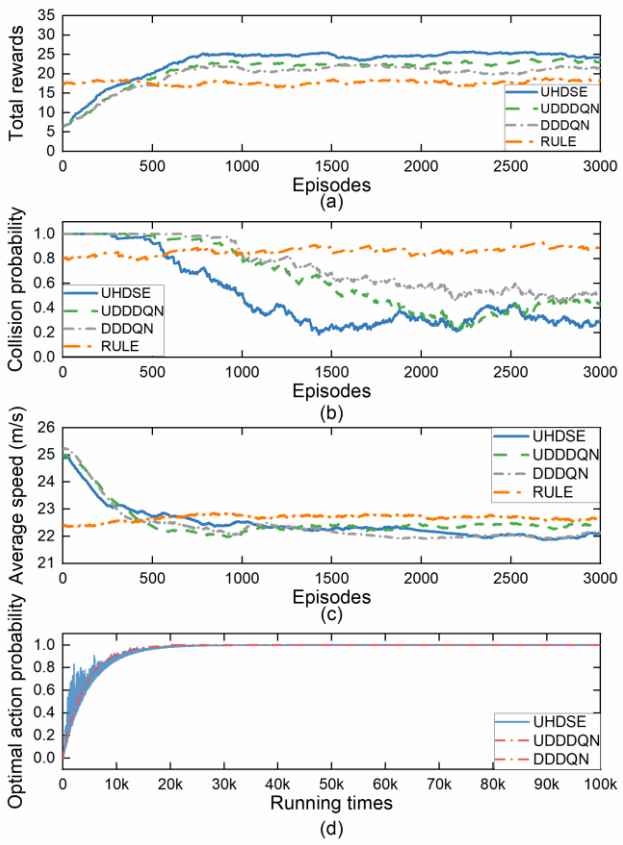

Fig. 7. Results of the low-density traffic case. (a) Total rewards. (b) Collision probability. (c) Average speed of self-vehicle. (d) Optimal action probabilities.

Fig. 7(a) shows the total rewards of the four methods under the considered low-density traffic flow. In the initial phase, the RULE strategy provides higher returns than the reinforcement learning-based vehicle decision strategy. However, after 400 training episodes, the reinforcement learning-based strategy leads to significantly higher rewards. Meanwhile, the agent obtains the highest rewards under the UHDSE method. The collision probability of the reinforcement learning-based vehicle decision strategy in Fig. 7(b) rapidly decreases from 1 and converges to 0.1~0.3 with training. In addition, the UHDSE method has the lowest collision rate after 2100 training episodes, which yields a higher total rewards as shown in Fig. 7(a). Fig. 7(c) shows the average speed of the self-vehicle under the four decision-making strategies. The reinforcement learning-based vehicle uses a relatively high speed (25m/s) in the initial phase for higher speed reward $R(s,a)_{speed}$. However, after a training period, the agent learns from the historical data that this is not the optimal choice because it brings a higher collision probability, so the vehicle speed gradually converges to about 22.5km/h. The optimal probabilities in Fig. 7(d) all converge from 0 to 1. In particular, the variation of the optimal action probability based on the UHDSE strategy fluctuates significantly due to the state-entropy based exploration probability. In contrast, the optimal action probabilities of the UDDQN and DDDQN strategies equally and smoothly due to the ε-greedy exploration.

The test results for the four decision-making strategies are summarized in Table 5 It can be seen that the UHDSE-based vehicle decision method leads to the highest rewards and the lowest collision percentage, which is consistent with the training results. The UDDDQN strategy achieves the highest average vehicle speed, and it sacrifices part of the safety for the highest traffic efficiency, which is undesirable. Moreover, the UHDSE-based method achieves a balance between the safety and the traffic efficiency, which improves the vehicle traffic efficiency while ensuring the safety as much as possible. We find that the vehicles performed more lane-

changing behaviors to ensure the traffic efficiency to drive on the appropriate road in the low-density traffic environment. The None action percentage of the reinforcement learning-based vehicle decision strategy is 0 because the agent learned that performing other actions brings higher gains.

4.3. Case 2: High-Density Traffic Flow

To analyze the obstacle avoidance capability of the proposed vehicle decision-making method, we perform a highway test scenario with a high density of vehicles as depicted in Fig. 8. Moreover, the high-density traffic flow means that the average headway is 14m.

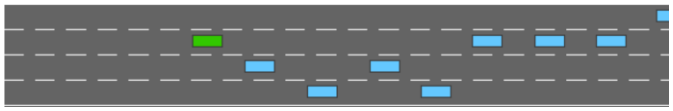

Fig. 8 Diagram of traffic flow for high-density cases: self-vehicle (in green) and surrounding vehicles (in blue).

The considered test scenario is more dangerous than the real-world situations to test the performance of the vehicle decision strategies in extreme environments. The corresponding results are shown in Fig. 9.Fig. 9(a) shows the total rewards of the agent under the four decision-making methods. The agent can obtain the highest reward under the UHDSE method and reaches the convergence at 1300 episodes faster than the UDDQN (1950 episodes) and DDDQN (2150 episodes) methods. In addition, the total rewards of the converged UHDSE method (13.5) are higher than those of the UDDDQN (13.1) and DDDQN (12.3) methods. For the RULE method, the total rewards were better than the other methods when the agent was undertrained, but when the agent was trained for some time (about 800 episodes), the reinforcement learning-based UHDSE, UDDDQN, and DDDQN methods performed significantly better than the RULE-based methods (5.6). From Fig. 9(b), it can be obtained that the collision probability under the UHDSE method is significantly lower than the other three methods. Since safety is the primary concern in the practical application of autonomous driving, the decision model based on the UHDSE method is more promising to be applied in the real driving environment.

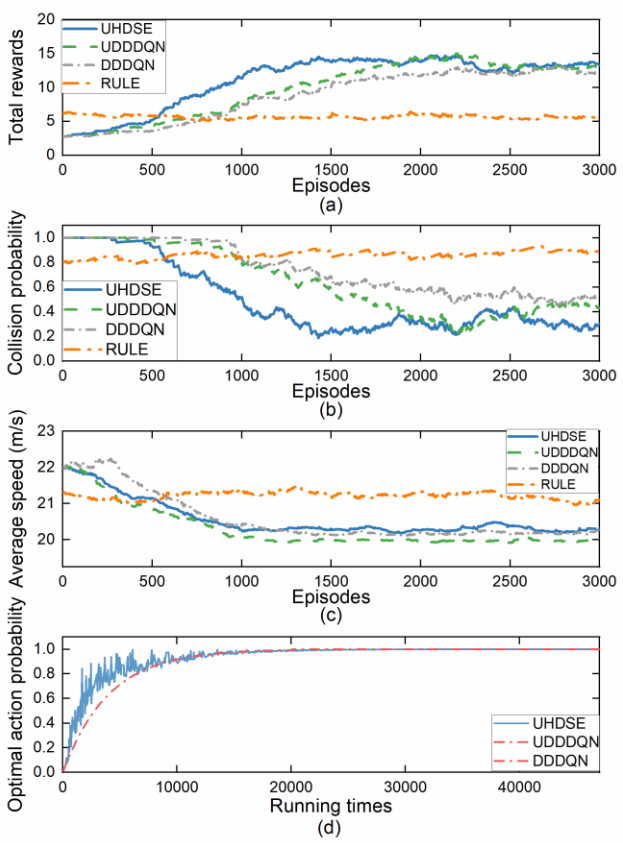

Fig. 9. Results of the high-density traffic case. (a) Total rewards. (b) Collision probability. (c) Average speed of self-vehicle. (d) Optimal action probabilities.

Fig. 9(c) describes the average vehicle speed of the self-vehicle under different methods. We can see that, at the beginning of the training, the reinforcement learning-based method maintains a higher speed than the RULE-based method to obtain a higher speed reward $R(s,a)_{speed}$. In particular, the average speed of the DDDQN-based method tends to increase at the beginning period. However, after a period of training, the reinforcement learning-based agent finds that higher speeds tend to bring negative collision rewards $R(s,a)_{collision}$, so the speed decreases and eventually converges to the maximum reward. Compared with other methods, the UHDSE-based vehicle decision method can guarantee a higher vehicle speed while maintaining the lowest collision probability, improving traffic flow efficiency. Fig. 9(d) shows the variation of optimal action selection probability as the running times increase in the reinforcement learning algorithm. At the initial stage of training in reinforcement learning, due to the lack of historical information, the optimal action probability should be set low since the actual optimal action is not available. The optimal probabilities in Fig. 9(d) all converge from 0 to 1, which is coherent to the requirement of exploration and exploitation balance in reinforcement learning algorithms. Unlike

UDDQN and DDDQN methods, the variation of the optimal action probabilities of the UHDSE-based agent significantly fluctuates. This is due to the state entropy-based exploration probability used by the UHDSE method, which obtains different action probabilities based on the current $Q(s,a)$.

The test results of the four decision strategies are summarized in Table 6 Compared with the DDDQN and RULE strategies, the UHDSE and UDDQN vehicle decision strategies considering environmental uncertainty have a better performance. However, the average vehicle speed of the UDDDQN strategy is the lowest, and a part of the vehicle traffic efficiency is sacrificed to ensure the safety. Moreover, for the specific action percentage, compared with DDDQN, the UHDSE-based vehicle decision method has a higher average speed and a lower collision rate but also a lower lane-changing behavior, indicating that the DDDQN-based method has some invalid lane-changing. That is, lane changing may bring short-term rewards but not long-term rewards, and we will analyze the ineffective lane changing of the DDDQN strategy specifically in Figs. 10 and 11. The lowest percentage of none actions is adopted with the UHDSE method, which indicates that it explores the environment more fully.

Table 6. Test data set results in high traffic flow.

Items	UHDSE	UDDDQN	DDDQN	RULE
Average Total Reward	16.43	15.68	14.03	9.26
Crash percentage	0.28	0.46	0.53	0.81
Average vehicle speed	20.41	19.94	20.33	21.09
Lane change maneuver percentage	18.6%	15.2%	21.4%	8.3%
Longitudinal maneuver percentage	81.2%	80.7%	62.7%	84.0%
None action percentage	0.3%	4.0%	15.9%	7.7%

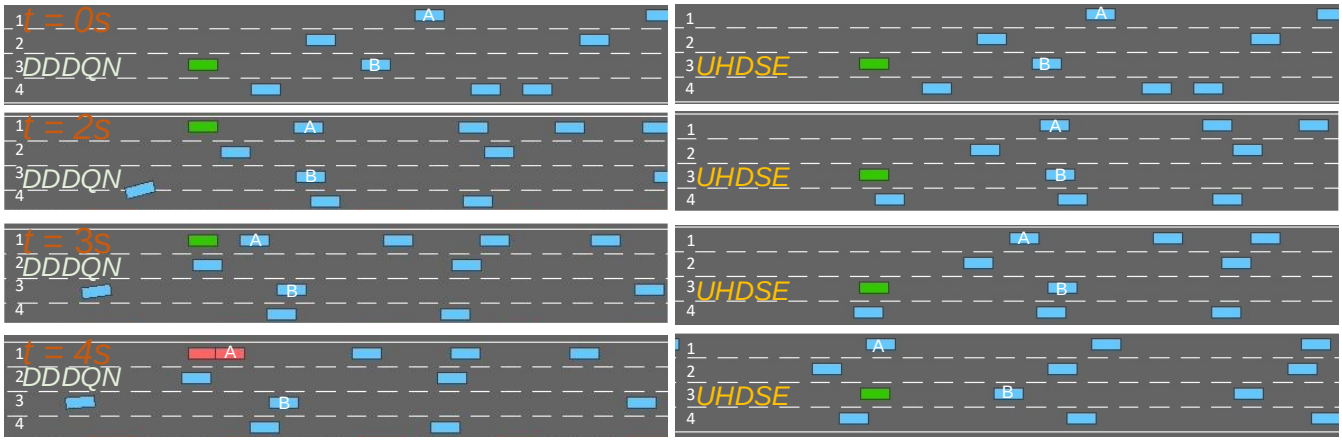

Fig. 10. Vehicle driving conditions under Scenario 1: self-vehicle (in green), surrounding vehicles (in blue), crashed vehicle (in red) and the same markers (A, B, C and D) represent the same vehicles.

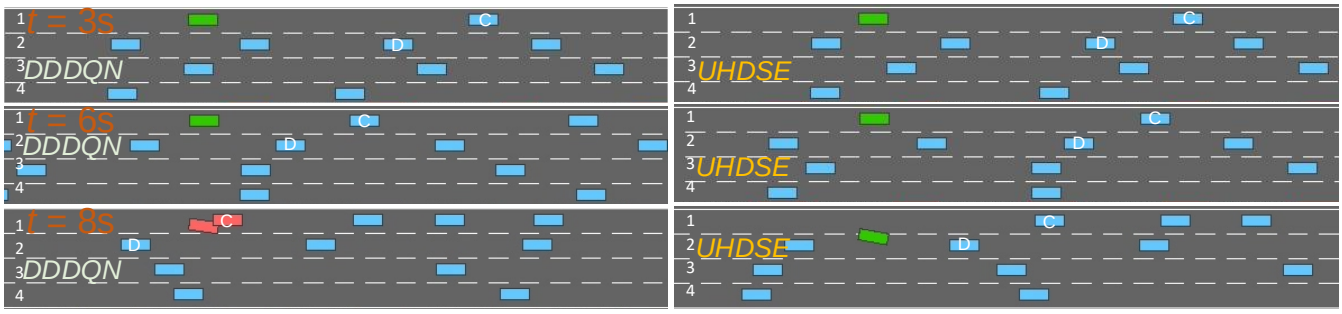

Fig. 11. Vehicle driving conditions under Scenario 2: self-vehicle (in green), surrounding vehicles (in blue), crashed vehicle (in red) and the same markers (A, B, C and D) represent the same vehicles.

To further analyze the impact of considering environmental uncertainty on vehicle decision-making, two scenarios from

the test set were selected to analyze the difference in the decision-making of autonomous vehicles under UHDSE and DDDQN strategies. Figs. 10 and11 represent the vehicle positions in the two considered scenarios at different instants. In Fig. 10 at t=0s, the self-vehicle under both methods is moving in lane 3. At t=2s, the self-vehicle under DDDQN strategy changes lanes to lane 1 because vehicle A on lane 1 is further away than vehicle B. After 2s, vehicle A started braking due to the small distance in front of it, so there is a collision with vehicle A, which is an invalid lane change, only considering the short-term traffic efficiency. On the other hand, under the UHDSE strategy, the effect of an uncertain environment is considered, and the future motion state of vehicle A is predicted. Then, it chooses to continue driving in lane 3, which improves the efficiency of the passage with the premise of safety. In Fig. 11, at t = 3s, the self-vehicle moves in lane 1 under both strategies. Since the leading vehicle C in lane 1 is slow, which affects the passing efficiency, the self-vehicle under both strategies decides to perform a lane change operation. However, under the DDDQN strategy, the self-vehicle did not choose the right time to change lanes and started to change lanes before vehicle D. The collision occurred because there was no suitable space for a lane change. On the other hand, under the UHDSE strategy, the self-vehicle slows down in advance and chooses to start a lane change after the vehicle D. There is a sufficient lane change space to ensure the safety of lane change.

5. Conclusions

We have proposed a novel vehicle decision framework based on heuristic deep reinforcement learning for vehicle decision problems in uncertain environments. In view of the environmental uncertainty of highways, the lane change intention and vehicle motion trajectory of vehicles are predicted based on LSTM. Based on this model, a future risk assessment state model is proposed as a part of the reinforcement learning framework. Aiming at the previous literature's exploration and exploitation dilemma of reinforcement learning, we propose a heuristic decay state entropy-based reinforcement learning algorithm based on dueling double DQN, which adopts different exploration weights based on the entropy of current Q-values to improve the exploration efficiency.

The obtained results show that the proposed decision-making framework achieves a good performance in both low-density and high-density traffic flows, reducing the collision rate and improving the traffic efficiency. In the low-density traffic case, the UHDSE algorithm significantly reduces the collision probability while increasing the average speed compared to the DDDQN algorithm, because the UHDSE algorithm considers the future risk and completes the lane change behavior in advance before the risk appears. In addition, compared with the common DDDQN method in high-density traffic cases, the convergence time is reduced about 39.5%, the collision rate is reduced about 25%, and the average total return is improved about 17.1%. However, as shown in Table 6 there is also a 12% probability of collision, which is caused by a mismatch in the dataset. Moreover, the training of the intention recognition is done with the NGSIM real vehicle data, while the training vehicle decisions are done with simulation data. Therefore, building a training environment that better simulates the real world and testing in real vehicles is our future work. Moreover, game theory can be applied to the decision-making framework to represent the interaction between the self-vehicle and surrounding vehicles.

References

1. Dong J, Chen S, Li Y, et al.: Space-weighted information fusion using deep reinforcement learning.: The context of tactical control of lane-changing autonomous vehicles and connectivity range assessment. Transportation Research Part C.: Emerging Technologies. **128**, (2021)
2. Alvaro PK, Burnett NM, Kennedy GA, et al.: Driver education.: Enhancing knowledge of sleep, fatigue and risky behaviour to improve decision making in young drivers. Accident Analysis & Prevention. **112**, 77-83 (2018)
3. Li G, Yang Y, Zhang T, et al.: Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios. Transportation Research Part C.: Emerging Technologies. **122**, (2021)
4. Kiran BR, Sobh I, Talpaert V, et al.: Deep Reinforcement Learning for Autonomous Driving.: A Survey. IEEE Trans Intell Transp Syst. 1-18 (2021)
5. Montemerlo M, Becker J, Bhat S, et al Junior.: The DARPA Urban Challenge.: Autonomous Vehicles in City Traffic. Berlin, Heidelberg.: Springer Berlin Heidelberg. 91-123 (2009)
6. Patz BJ, Papelis Y, Pillat R, et al.: A practical approach to robotic design for the DARPA Urban Challenge. J Field Rob. **25**(8), 528-66 (2008)
7. Jiao Y, Tang X, Qin Z, et al.: Real-world ride-hailing vehicle repositioning using deep reinforcement learning. Transportation Research Part C.: Emerging Technologies. **130**, (2021)
8. Xu X, Zuo L, Li X, et al.: A Reinforcement Learning Approach to Autonomous Decision Making of Intelligent Vehicles on Highways. IEEE Transactions on Systems, Man, and Cybernetics.: Systems. 1-14 (2019)
9. Zhang Y, Gao B, Guo L, et al.: Adaptive Decision-Making for Automated Vehicles Under Roundabout Scenarios Using Optimization Embedded Reinforcement Learning. IEEE Trans Neural Networks Learn Syst. 1-13 (2020)
10. Cao Z, Yang D, Xu S, et al.: Highway Exiting Planner for Automated Vehicles Using Reinforcement Learning. IEEE Trans Intell Transp Syst. **22**(2), 990-1000 (2021)
11. Liu J, Zhao W, Xu C.: An Efficient On-Ramp Merging Strategy for Connected and Automated Vehicles in Multi-Lane Traffic. IEEE Trans Intell Transp Syst. 1-12 (2021)
12. Wang G, Hu J, Li Z, et al.: Harmonious Lane Changing via Deep Reinforcement Learning. IEEE Trans Intell Transp Syst. 1-9 (2021)
13. Chen S, Wang M, Song W, et al.: Stabilization Approaches for Reinforcement Learning-Based End-to-End Autonomous Driving. IEEE Trans Veh Technol. **69**(5), 4740-50 (2020)
14. Li D, Zhao D, Zhang Q, et al.: Reinforcement Learning and Deep Learning Based Lateral Control for Autonomous Driving IEEE Comput Intell Mag. **14**(2), 83-98 (2019)
15. Jaritz M, Charette Rd, Toromanoff M, et al, "End-to-End Race Driving with Deep Reinforcement Learning," Paper presented at.: 2018 IEEE International Conference on Robotics and Automation (ICRA), (2018)

16. Fu Y, Li C, Yu FR, et al.: A Decision-Making Strategy for Vehicle Autonomous Braking in Emergency via Deep Reinforcement Learning. IEEE Trans Veh Technol. **69**(6), 5876-88 (2020)
17. Hoel C-J, Driggs-Campbell K, Wolff K, et al.: Combining Planning and Deep Reinforcement Learning in Tactical Decision Making for Autonomous Driving. IEEE Trans Intell Veh. **5**(2), 294-305 (2020)
18. Liao J, Liu T, Tang X, et al.: Decision-Making Strategy on Highway for Autonomous Vehicles Using Deep Reinforcement Learning. IEEE Access. **8**, 177804-14 (2020)
19. Hubmann C, Schulz J, Becker M, et al.: Automated Driving in Uncertain Environments.: Planning With Interaction and Uncertain Maneuver Prediction. IEEE Trans Intell Veh. **3**(1), 5-17 (2018)
20. Pouya P, Madni AM.: Expandable-Partially Observable Markov Decision-Process Framework for Modeling and Analysis of Autonomous Vehicle Behavior. IEEE Syst J. 1-12 (2020)
21. Galceran E, Cunningham AG, Eustice RM, et al.: Multipolicy decision-making for autonomous driving via changepoint-based behavior prediction.: Theory and experiment. Autonomous Robots. **41**(6), 1367-82 (2017)
22. Mehta D, Ferrer G, Olson E, "Fast discovery of influential outcomes for risk-aware MPDM," Paper presented at.: 2017 IEEE International Conference on Robotics and Automation (ICRA), (2017)
23. Ye N, Somani A, Hsu D, et al.: DESPOT.: online POMDP planning with regularization. J Artif Int Res. **58**(1), 231–66 (2017)
24. Zhang L, Ding W, Chen J, et al, "Efficient Uncertainty-aware Decision-making for Automated Driving Using Guided Branching," Paper presented at.: 2020 IEEE International Conference on Robotics and Automation (ICRA), (2020)
25. Okumura B, James MR, Kanzawa Y, et al.: Challenges in Perception and Decision Making for Intelligent Automotive Vehicles.: A Case Study. IEEE Trans Intell Veh. **1**(1), 20-32 (2016)
26. Aradi S.: Survey of Deep Reinforcement Learning for Motion Planning of Autonomous Vehicles. IEEE Trans Intell Transp Syst. 1-20 (2020)
27. Sledge IJ, Príncipe JC, "Balancing exploration and exploitation in reinforcement learning using a value of information criterion," Paper presented at.: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2017)
28. Alizadeh A, Moghadam M, Bicer Y, et al, "Automated Lane Change Decision Making using Deep Reinforcement Learning in Dynamic and Uncertain Highway Environment," Paper presented at.: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), (2019)
29. Ye Y, Zhang X, Sun J.: Automated vehicle's behavior decision making using deep reinforcement learning and high-fidelity simulation environment. Transportation Research Part C.: Emerging Technologies. **107**, 155-70 (2019)
30. Hoel C, Wolff K, Laine L, "Automated Speed and Lane Change Decision Making using Deep Reinforcement Learning," Paper presented at.: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), (2018)
31. Wolf P, Kurzer K, Wingert T, et al, "Adaptive Behavior Generation for Autonomous Driving using Deep Reinforcement Learning with Compact Semantic States," Paper presented at.: 2018 IEEE Intelligent Vehicles Symposium (IV), (2018)
32. Nageshrao S, Tseng HE, Filev D, "Autonomous Highway Driving using Deep Reinforcement Learning," Paper presented at.: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), (2019)
33. Aradi S, Becsi T, Gaspar P, "Policy Gradient Based Reinforcement Learning Approach for Autonomous Highway Driving," Paper presented at.: 2018 IEEE Conference on Control Technology and Applications (CCTA), (2018)
34. Yu C, Wang X, Xu X, et al.: Distributed Multiagent Coordinated Learning for Autonomous Driving in Highways Based on Dynamic Coordination Graphs. IEEE Trans Intell Transp Syst. **21**(2), 735-48 (2020)
35. Mo S, Pei X, Wu C.: Safe Reinforcement Learning for Autonomous Vehicle Using Monte Carlo Tree Search. IEEE Trans Intell Transp Syst. 1-8 (2021)
36. Treiber M, Hennecke A, Helbing D.: Congested traffic states in empirical observations and microscopic simulations. Phys Rev E. **62**(2), 1805-24 (2000)
37. Kesting A, Treiber M, Helbing D.: General Lane-Changing Model MOBIL for Car-Following Models. Transp Res Rec. **1999**(1), 86-94 (2007)
38. Li L, Li P, "Analysis of Driver's Steering Behavior for Lane Change Prediction," Paper presented at.: 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), (2019)
39. V. Alexiadis, J. Colyar, J. Halkias, et al.: The next generation simulation program, ITE J. Inst. Transp. Eng. **74**(8), 22 (2004)
40. Dahl J, De Campos GR, Olsson C, et al.: Collision Avoidance.: A Literature Review on Threat-Assessment Techniques. IEEE Trans Intell Veh. **4**(1), 101-13 (2019)
41. Xu C, Zhao W, Wang C.: An Integrated Threat Assessment Algorithm for Decision-Making of Autonomous Driving Vehicles. IEEE Trans Intell Transp Syst. **21**(6), 2510-21 (2020)
42. Elefteriadou LA.: The Highway Capacity Manual, 6th edition.: A guide for multimodal mobility analysis. TR News. Washington, D.C..: Transportation Research Board; (2016)
43. Arulkumaran K, Deisenroth MP, Brundage M, et al.: Deep Reinforcement Learning.: A Brief Survey. IEEE Signal Process Mag. **34**(6), 26-38 (2017)
44. Mnih V, Kavukcuoglu K, Silver D, et al.: Playing atari with deep reinforcement learning. arXiv preprint arXiv.:13125602. (2013)
45. Usama M, Chang DE.: Learning-Driven Exploration for Reinforcement Learning. (2020)
46. Leurent E.: An Environment for Autonomous Driving Decision-Making. https.://github.com/eleurent/highway-env(2022). Accessed 05 June