Hindawi
Journal of Advanced Transportation
Volume 2023, Article ID 1513008, 13 pages
https://doi.org/10.1155/2023/1513008

Research Article

An Integrated Lateral and Longitudinal Decision-Making Model for Autonomous Driving Based on Deep Reinforcement Learning

Jianxun Cui,[1] **Boyuan Zhao**,[1] **and Mingcheng Qu**[2]

[1]*School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China*
[2]*Department of Software, Harbin Institute of Technology, Harbin 150001, China*

Correspondence should be addressed to Jianxun Cui; cuijianxun@hit.edu.cn

Received 12 August 2022; Revised 17 September 2022; Accepted 23 September 2022; Published 13 April 2023

Academic Editor: Rui Zhu

Decision-making is an important component of autonomous driving perception, decision-making, planning, and control pipeline, which undertakes the task of how the ego vehicle makes high-level decision-making behaviors (such as lane change and car following) after sensing the environmental state, and then these high-level decision-making behaviors can be transmitted to the downstream planning and control module for specific low-level action execution. Based on the method of deep reinforcement learning (specifically, Deep Q network (DQN) and its variants), an integrated lateral and longitudinal decision-making model for autonomous driving is proposed in a multilane highway environment with both autonomous driving vehicle (ADV) and manual driving vehicle (MDV). The classic MOBIL and IDM models are used for the lateral and longitudinal decisions of MDV (i.e., lane changing and car following), while the lateral and longitudinal decisions of ADV are dominated by deep reinforcement learning models. In addition, this paper also uses the nonlinear kinematic bicycle model and two-point visual control model to realize the low-level control of both MDV and ADV. By setting a reasonable state, action, and reward function, this paper has carried out a large number of simulation experiments on the proposed autonomous driving decision-making model based on deep reinforcement learning in a three-lane road environment. The results show that under such scenario setting conditions, the deep reinforcement learning-based model proposed in this paper performs well in autonomous driving safety and travel efficiency. At the same time, when compared with the classical rule-based decision-making model (MOBIL&IDM), it is found that the model proposed in this paper can significantly achieve better results in episode rewards after stable training. In addition, through a large number of hyper-parameter tuning experiments, the performance of DQN, DDQN, and dueling DQN models, which are also deep reinforcement learning-based decision-making models, under different hyper-parametric configurations is compared and analyzed, which can provide a valuable reference for the specific scenario application of these models.

1. Introduction

Autonomous driving is hot research and practical issue in the fields of road traffic engineering, vehicle engineering, and artificial intelligence in recent years, which is considered to have great potential in alleviating traffic congestion, reducing environmental pollution, improving traffic safety performance, and even systematically changing the future traffic mobility pattern [1]. In order to realize autonomous driving, a vehicle needs to be able to accurately perceive the state of itself and the surrounding environment, then make corresponding behavioral decisions and consequently generate a safe, efficient trajectory based on perceptual understanding, and finally track the generated trajectory as accurately as possible by controlling the throttle, brake pedal, and steering wheel [2]. This autonomous driving process is usually described as a modular pipeline as shown in Figure 1. After the travel user gives the global information, such as the travel destination and navigation route, the autonomous vehicle will collect the environmental information through its own installed cameras, LIDAR, and other types of sensors at a certain frequency, and then the collected

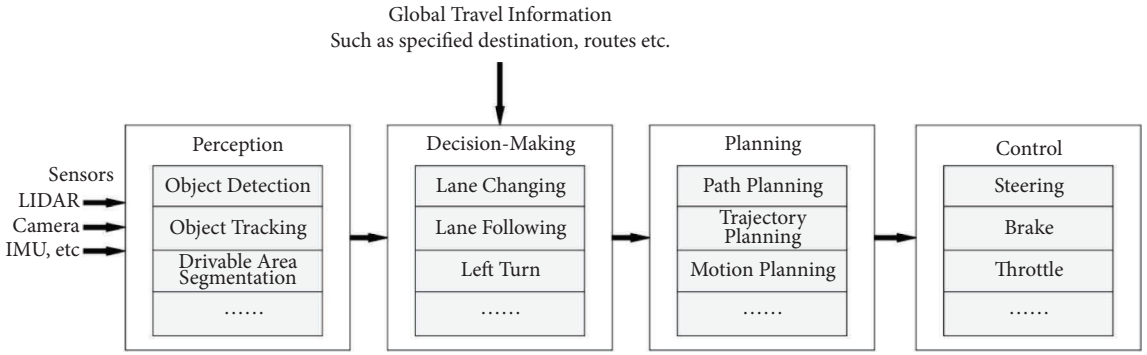

Figure 1: Modular pipeline of autonomous driving.

raw sensor data will be input into the perception module for environmental semantic understanding tasks such as object detection and tracking. Further, on the basis of state perception and the user's global travel information, the autonomous vehicle will make local behavior decisions such as whether to change lanes and further generate behavior instructions to the planning module to generate the optimal trajectory. Finally, the generated trajectory can be tracked by controlling the throttle, brake, steering wheel, and other actors.

Since decision-making is an important part that links perception and trajectory planning and greatly determines the safety and efficiency of autonomous driving, extensive research around this issue can be found in the literature. In general, the typical research method of autonomous driving decision-making can be mainly categorized into 4 classes: rule-based [3–5], classical machine learning-based [6–8], deep reinforcement learning-based [9–12], and deep imitation learning-based [13–15]. Among many research methods, deep reinforcement learning has received great attention in recent years because it does not need a lot of human labeled training data, the learning style is closer to human learning, and the generalization ability is strong. Despite the above advantages, for the application of deep reinforcement learning in automated decision-making modeling, how to construct an effective representation of the environmental state, how to design an effective reward function, and how to compare and analyze the performance differences between the deep reinforcement learning model and the traditional rule-based model are still challenging and needed to be further studied. In view of this, this paper aims to study the modeling of autonomous driving decision-making based on the DQN and its variants under the condition of the mixture of autonomous driving vehicles and manual driving vehicles in a specific scenario of a multi-lane highway. It is hoped that this research can provide effective models for safety, efficiency, etc. for decision-making in multilane autonomous driving scenarios. At the same time, through a large number of hyper-parameter tuning experiments, we will systematically compare the performance of several classical value-based DRL models (i.e., DQN, DDQN, and Dueling DQN) for autonomous driving decision-making, and further evaluate the performance differences between them and other traditional rule-based decision-making models, so as to provide a valuable reference for autonomous driving decision-making modeling in multilanescenarios.

The contributions of this study include the following aspects.

(1) An integrated lateral and longitudinal decision-making model based on deep reinforcement learning is proposed for autonomous driving in a multilane highway with mixed traffic composed of MDVs and ADVs. A large number of simulation experiments are conducted to verify the effectiveness of the proposed model.

(2) Extensive simulations are conducted to compare the model performance between DRL-based models (i.e., DQN, DDQN, and Dueling DQN) and rule-based models (i.e., IDM and MOBIL), results of which show that DRL-based models are significantly superior to rule-based models for autonomous driving decision making.

(3) Performance comparison between DQN and its variants (i.e., DDQN and Dueling DQN) is also conducted, results of which indicate that DDQN and Dueling DQN do improve the performance of DQN model for autonomous driving decision-making by properly estimating Q values and optimizing network structure in terms of training efficiency and reward acquisition.

(4) With different ADV penetration, the training efficiency of DQN-series models for autonomous driving decision-making is compared, according to the rising of ADV penetration, for a single ADV, the environment becomes more uncertain and complex, so the training process of the DQN-series models is more difficult to be stabilized.

The organization of this study is as follows. Section 2 presents a brief literature review of decision-making of autonomous driving. Section 3 introduces our proposed methodology for modeling decision-making of autonomous driving and Section 4 conducts a large number of simulation experiments to verify the proposed models and the results of

which are discussed. Finally, Section 5 concludes this manuscript and briefly discusses future research directions.

2. Literature Review

Decision-making corresponds to a high-level behavior of an automated vehicle, which decides whether the automated vehicle will change lanes, follow or turn et al. Because decision-making represents the response of autonomous vehicles on the environmental state observation and driving goals, and plays a guiding role in the downstream planning and control module, it has attracted a lot of research in the literature.

In general, the research on autonomous driving decision-making can be divided into rule-based, finite state machine-based, and machine learning-based methods. Rule-based methods are based on some predefined parameters that would tune the algorithm for a specific environment, in which the most representative ones are MOBIL [16] for lateral decision-making and IDM [17] for longitudinal decision-making. A common limitation of these approaches is the lack of flexibility under dynamic situations and diverse driving styles [18]. Since both driving contexts and the behaviors available in each context can be modeled as finite sets, a natural approach to automating this decision-making is to model each behavior as a state in a finite state machine with transitions governed by the perceived driving context such as relative position with respect to the planned route and nearby vehicles. In fact, finite state machines were adopted as a mechanism for behavior control by most teams in the DARPA Urban Challenge [19]. However, because the context of open road autonomous driving is highly complex, dynamic, and uncertain, it is intractable to build all possible driving contexts and their corresponding behaviors into finite state machines in essence, which makes the finite state machine destined to be a simplified modeling method for autonomous driving decision-making and difficult to use in real complex scenes [20]. Machine Learning (ML) based methods have a very good generalization ability for unknown scenes when they are properly trained through a large number of data samples, and there is no need to manually specify rules in advance [21]. Vallon et al. [22] proposed a support vector machine (SVM) model to capture the lane change decision behavior of human drivers. After the lane change demand is generated, the maneuver is executed using an MPC. By extracting the features from surrounding vehicles that are relevant to the lane-changing of the subject vehicle, Bi et al. [23] used a randomized forest and back-propagation neural network to model the process of lane-changing in traffic simulation. ML-based methods above for autonomous driving decision-making research fall into the supervised learning paradigm, so it is necessary to collect a great amount of real-world driving behavior data and annotate a large number of manual driving decision-making behaviors, which is usually very time-consuming and labor-intensive. More importantly, it is difficult to pose autonomous driving as a supervised learning problem as it has a strong interaction with the environment including other vehicles, pedestrians, and road networks [10]. In recent years, another machine learning paradigm, reinforcement learning (especially Deep Reinforcement Learning, DRL), which learns the task in a trial-and-error way that does not require explicit human labeling or supervision on each data sample has been widely used in research of autonomous driving decision-making and control. Ngai and Yung [24] adopted a multiple-goal reinforcement learning (RL) framework to model complex vehicle overtaking maneuvers. For lane-keeping assisting decision-making issues, Sallab et al. [10] adopted Deep Q-Network Algorithm (DQN) and Deep Deterministic Actor-Critic Algorithm (DDAC) to model discrete actions category and continuous actions category of autonomous driving respectively. Wang and Chan [25] applied deep reinforcement learning (DRL) techniques to find optimal control policy for automating decision making on a ramp merge. The proposed methods also have the potential to be extended and applied to other autonomous driving scenarios such as driving through a complex intersection or changing lanes under varying traffic flow conditions. Hoel et al. [26] proposed a Deep Q-Network model automatically to generate a decision-making function to handle speed and lane change. For navigation at occluded intersections, Isele et al. [27] used Deep RL methods to provide efficient automated decision-making strategy, which is able to learn policies that surpass the performance of a commonly-used heuristic approach in several metrics including task completion time and goal success rate and have limited ability to generalize. Although great achievements have been made in the research of autonomous driving decision-making using DRL, applying RL to real-world applications is particularly challenging, especially for autonomous driving tasks that involve extensive interactions with other vehicles in a dynamically changing environment. One significant barrier of applying RL to real-world problems is the required definition of the reward function, which is typically unavailable or infeasible to design in practice. Inverse reinforcement learning (IRL) aims to tackle such problems by learning the reward function from expert demonstrations, thus avoiding reward function engineering and making good use of the collected expert data [28, 29]. However, because of the expensive reinforcement learning procedure in the inner loop, it has limited application in problems involving high-dimensional state and action spaces [30]. To overcome the limitation, some state-of-the-art works were conducted, such as generative adversarial imitation learning (GAIL) [30], guided cost learning (GCL) [31], and adversarial inverse reinforcement learning (AIRL) [32]. Although imitation learning theoretically provides a more stable training process, and there is no need to explicitly specify a reward function, it still needs to collect a large number of expert driving data as a demonstration compared with deep reinforcement learning and faces the problem of distribution shift [33].

In view of the learning advantages of DRL in the complex interactive autonomous driving decision-making, this paper attempts to explore a more intelligent decision-making strategy through effective environmental state representation and a fine design of reward function in a specific multilane mixed driving scenario based on DQN and its

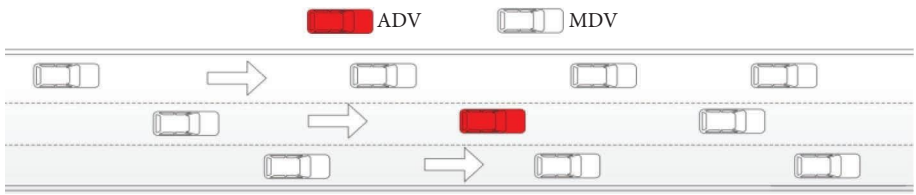

Figure 2: Multilane highway scenario.

variants. Further, combining the proposed DRL-based decision-making models with the low-level effective control model, we will conduct a large number of simulation experiments to determine optimization configuration of various hyper-parameters associated with the decision-making models. In addition, the performance of the proposed decision models will be compared with the traditional rule-based model to validate the efficiency of our models. This research is expected to provide a valuable reference for the application of deep reinforcement learning in autonomous driving decision-making research.

3. Methodology

In this section, we first give the detailed description of the problem that we are addressing in this paper. Next, the rule-based lateral and longitudinal decision-making models of MDV which act as the interacted surrounding traffic of ADV are presented. Then decision-making model of ADV is constructed based on DQN by specifying the state representation, action set, and reward function. Finally, a low-level control model based on a nonlinear kinematic bicycle model combined with two-point visual control is presented to implement the output from the decision-making model of both MDV and ADV.

3.1. Problem Statement. The autonomous driving decision-making scenario concerned in this paper is shown in Figure 2 . This multilane autonomous driving scenario consists of multiple lanes driving in the same direction, in which ADV (in red color) and MDV (in grey color) are in a mixed driving state. The Decision-making of MDV is driven by two rule-based models that are MOBIL and IDM. MOBIL is responsible for lateral decision-making and IDM is responsible for longitudinal decision-making, which will be introduced in detail later. The lateral and longitudinal decision-making of ADV is both achieved by a DRL-based model (i.e., DQN), which is the major research concern of this paper. The output of decision-making models of both MDV and ADV will immediately transmit to the low-level control model which is realized by the nonlinear kinematic bicycle model to generate specific vehicle action execution. The research problem of this paper can be summarized as how to train a safe and effective deep reinforcement learning model by properly representing the environmental state, action set, and reward function of autonomous vehicles in the aforementioned mixed driving scenarios of manual driving and autonomous driving.

3.2. Decision Making of MDV

3.2.1. Longitudinal Decision of MDV. IDM (Intelligent Driver Model) [17] which is a rule-based car following model is employed to model the longitudinal decision making of MDV. IDM was originally proposed in the field of adaptive cruise control (ACC) to generate appropriate acceleration for the ego vehicle based on its relative driving state with the leading on a single lane. The longitudinal decision-making formulas described by IDM are shown in Eq. 1-2.

$$a = a_{\max}\left(1 - \left(\frac{u}{u_d}\right)^{\delta} - \left(\frac{d^*(u,\Delta u)^2}{d}\right)\right), \tag{1}$$

$$d^*(u,\Delta u) = d_{\min} + uT + \frac{u\Delta u}{2\sqrt{ba_{\max}}}, \tag{2}$$

where, a is the instant accelarationof ego vehicle, which is needed to be determined in each decision step; $a_{\max}$ is the maximum acceleration of the ego vehicle; u and u_d is the current and desired speed of the ego vehicle; Δu is the speed difference between the ego vehicle and its leading vehicle; d is the gap between the ego vehicle and its leading vehicle; $d_{\min}$ is the minimum safety gap between the ego vehicle and its leading vehicle; T is safe time headway; b is the desired acceleration of the ego vehicle;

As it is seen in the equation (1) and (2), the original IDM model only restricted the acceleration of the ego vehicle by maximum acceleration $a_{\max}$; however, the minimum deceleration is not indicated. So, a condition depicted by equation (3) is added by us to limit the minimum deceleration of the ego vehicle.

$$a = \begin{cases} a, & a \geq a_{\min} \\ a_{\min}, & \text{otherwise}, \end{cases} \tag{3}$$

where, $a_{\min}$ is the minimum deceleration allowed.

In practice, the MDVs on each single lane execute the IDM longitudinal decision-making model respectively and then generate their own acceleration decisions in each time interval. If there is no leading vehicle in front of an MDV, its Δu and d is set to 0 and $d_{\max}$ (maximum gap for empty lane).

3.2.2. Lateral Decision of MDV. MOBIL (Minimizing Overall Braking Induced by Lane Change) [16] which is a rule-based lane change model is adopted here to make lateral decision of MDV. MOBIL determines whether lane change is safe and accessible according to the relative acceleration

between the ego vehicle and the vehicles on the adjacent lanes. MOBIL's decision-making process is divided into two steps: first, according to the limit of safety standards, the deceleration of new following vehicles should not be too low when lane changing occurs, which is described in (4).

$$\widehat{a}_{\text{new-follower}} > b_{\text{safe}}, \tag{4}$$

where, $\widehat{a}_{\text{new-follower}}$ is the acceleration of new following vehicles after lane change of the ego vehicle, which can be calculated by IDM; b_{safe} is the maximum safe deceleration. Second, if the first condition defined in equation (4) is met, MOBIL will check the second condition defined in equation (5) to make a final decision about whether trigger a lane change of the ego vehicle.

$$\widehat{a}_{\text{ego}} - a_{\text{ego}} + p\left(\widehat{a}_{\text{new-follower}} - a_{\text{new-follower}}\right) + q\left(\widehat{a}_{\text{old-follower}} - a_{\text{old-follower}}\right) > a_{th}, \tag{5}$$

where, $\widehat{a}_{\text{ego}}, a_{\text{ego}}$ are the new acceleration of the ego vehicle calculated by IDM after lane change and the old acceleration before lane change; $\widehat{a}_{\text{new-follower}}, a_{\text{new-follower}}$ are the new and old accelerations respectively of the new follower vehicle when lane change of the ego vehicle occurs; $\widehat{a}_{\text{old-follower}}, a_{\text{old-follower}}$ are the new and old accelerations respectively of the old follower vehicle when lane change of the ego vehicle occurs; p and q are politeness factors respectively of the new and old following vehicles; a_{th} is a predefined threshold value. Equation (5) indicates that only when the collective acceleration gain is greater than a predefined threshold, the lane change behavior of the ego vehicle can be truly triggered.

3.3. Decision Making of ADV. Both lateral and longitudinal decisions of ADV are modeled by the DRL method which here refers to DQN specifically. DQN was originally proposed by Mnih et al. [34] for playing Atari games, which is an effective DRL algorithm for discrete decision problems by combing deep learning and reinforcement learning. Traditionally, the Q value function corresponding to a specific state and action is represented by a table, which is hard to handle the problem with a large space of state variable. DQN overcomes this problem by using a deep neural network to represent the Q value function as $Q(s, a, \theta)$ instead of a table, where θ represents the learnable parameters of the neural network.

(1) **Q value function$Q(s, a, \theta)$of ADV.** Each decision-making action (e.g. left change and right change) of one ADV at the arbitrary time step is realized by choosing the action with the best-expected return according to the strategy of ε -greedy, which needs to establish the Q value function, $Q(s, a, \theta)$ of each state-action pair $(s, a) \cdot (s \in S, a \in A$, where S, A are state and action sets respectively). Here, a fully-connected neural network which takes one specific state as input and the corresponding Q value of each available action as output will be used to represent the Q value function.

(2) **Updating rule of$Q(s, a, \theta)$.** The updating rule of $Q(s, a, \theta)$ is described in Equation .

$$Q_{k+1}(s, a) = Q_k(s, a) + \alpha\left[r(s, a) + \gamma Q_k(s', a') - Q_k(s, a)\right], \tag{6}$$

where, $Q_k(s, a), Q_{k+1}(s, a)$ represents the Q values of kth and $k+1$th step, respectively; $r(s, a)$ is the instant reward received by executing action a under the state s; γ is the discount factor of future return; $\alpha \in [0, 1]$ is the learning rate which is used to trade-off between old and new learned experiences; s' is the state of next step after ADV takes action a under the state s; a' is the adopted action by ADV under state s' according to ε -greedy strategy with the kth Q value function (current or unupdated Q value function).

(3) **Exploration strategy of ADV**. In the process of updating of ADV's observed state, a suitable action must be determined for every step based on the function of the current state and $Q(s, a, \theta)$. If the action of ADV is taken completely according to the past experience; that is, the ADV chooses the action with the largest corresponding Q value, it is possible to be restricted in the existing experience and unable to find out the new action behavior with larger value; on the other hand, if ADV only focuses on exploring new actions, the majority of actions will be worthless, which leads to a very slow learning speed of Q function. Here, ε-greedy strategy which can makes a good balance between experience and exploration [35] is adopted here to select a suitable action under a specific state.

$$\pi(a|s) = \begin{cases} \text{argmax}_{a'} Q(s, a'), & 1 - \varepsilon \\ \text{randomly select an action from } A, & \varepsilon, \end{cases} \tag{7}$$

where, $\pi(a|s)$ is the action exploration function of ADV; ε represents a small probability, usually smaller than 0.05.

(4) **Buffer Replay.** Each update of ADV's Q function requires a lot of state-action pairs and corresponding instant rewards which can be collected only when ADV interacts with the environment. This leads to sample inefficiency which is a usually criticized problem in deep reinforcement learning. Buffer replay originally proposed by Mnih et al. [34] is adopted here to alleviate this problem and improve the performance of the DQN algorithms. A role of the replay buffer is crucial in terms of accessibility to a variety of data from various time steps, which

makes time-independent learning possible, and it allows the DQN algorithm to learn a robust decision policy

(5) **State, Action and Reward of ADV.**

State. Effective state representation directly affects the performance of the deep reinforcement learning algorithm. In the DQN algorithm, the state is the input of the Q network, which represents the ADV's observation of the surrounding environment. For lane change and car following decision making, an ADV should be able to observe its own state (such as speed and position) and the states of other vehicles within a certain range around it. This research uses an ego-centric reference frame as proposed by Bai et al. [36] to represent the states observed by the ego vehicle. Firstly, each lane of the highway is divided into equidistant cells longitudinally, length of each cell is set as the average car length. In each decision step, taking as the cell occupied by ego ADV as the center point, a span of 10 cells in the longitudinal direction is considered as the observable range of this ego ADV. Given there are 3 lanes in the driving direction of ADV, there are total of 30 cells' states should be referred by ADV to make decision. Each cell's state should be described by whether it is occupied by MDV and the current speed of the occupying MDV (if not occupied by a MDV, the speed of the cell will be set to zero). So, at each step, totally 60 variables (totally 30 cells in censoring range and each cell is described by two variables to indicate whether it is occupied and the speed of occupied vehicle) will be used to represent the surrounding environment state observed by ADV.

Action. The decision-making of ADV includes both lateral and longitudinal actions. The action space of ADV is described in Table 1.

Reward. The design of rewards is crucial to the effectiveness of a reinforcement learning algorithm. In order to encourage high-speed travel and realize complete collision avoidance, reward function should try to balance between travel safety and travel efficiency. Meanwhile, the unconscious violation of the egovehicle during lane change (such as changing from the edge lane to the curb) should also be prohibited. In other words, the criterion for a good decision is that no collision and violation occur. So, totally, reward function proposed in this research is composed of three parts: safety-related reward, efficiency-related reward, and lane change-related reward, which are defined separately in Equation (8)–(11).

Safety-related reward:

$$r_s \begin{cases} 0 & \text{no Collision} \\ -100 & \text{otherwise.} \end{cases} \tag{8}$$

Efficiency-related reward:

Table 1: Action space of ADV.

Action	Description
a_1	Lane change to left
a_2	No lane change (keep current lane)
a_3	Lane change to right
a_4	Acceleration
a_5	Deceleration

$$r_e = l\frac{u_{\text{ego}} - u_{\min}}{u_{\max} - u_{\min}}, \tag{9}$$

where, $u_{\text{ego}}, u_{\max}, u_{\min}$ represents the current speed of ego ADV, the maximum and minimum allowed speed; l is the reward factor.

Lane change-related reward:

$$r_1 = \begin{cases} 1 & \text{lang change success} \\ -1 & \text{otherwise.} \end{cases} \tag{10}$$

Total reward:

$$r_{\text{total}} = \lambda_s r_s + \lambda_e r_e + \lambda_l r_l, \tag{11}$$

where, $\lambda_s, \lambda_e, \lambda_l$ are weight coefficients of different reward components, which can be adjusted to balance between safety and efficiency. Here, $\lambda_s, \lambda_e, \lambda_l$ are set to be 0.5, 0.4, and 0.1, respectively.

3.4. Low-Level Control of MDV and ADV. After receiving the action instruction from the decision-making (car following or lane changing), the low-level controller will control the vehicle accordingly to realize this instruction. Here, nonlinear kinematic bicycle model is used for the simulation of dynamics of both ADV and MDV. The control inputs for the kinematic bicycle model are the front steering angle δ_f and the acceleration a, in which δ_f is calculated by a two-point visual control model of steering [37], and a is calculated by IDM. The description of two-point visual control model can be seen in Figure 3 . The model uses two tangent angles (i.e., θ_n and θ_f in Figure 3) of two reference points in near and far regions to calculate steering angle δ_f, which is described in Equation (12).

$$\delta_f = k_f\theta_f + k_n\theta_n + k_I\int \theta_n \mathrm{d}t, \tag{12}$$

where, k_f, k_n, k_I are the unable parameters of the proportional integration (PI) controller. l_n, l_f are determined by the positions of near and far reference points. When lane change occurs, for empty target lane, l_n, l_f are fixed, while for an occupied target lane, l_n remain fixed but l_f will be the distance between the new leading vehicle and the ego vehicle.

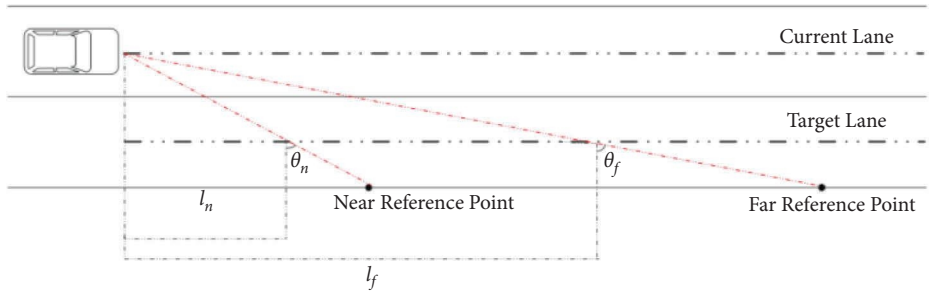

Figure 3: Two-point visual control model.

4. Numerical Experiment and Results

In this section, the proposed DQN-based multilane highway decision-making policy is evaluated by extensive simulation experiments.

4.1. Settings

(1) **Simulation Scenario.** The simulation scenario for evaluating the automatic driving decision-making model based on DQN proposed in this paper is a highway composed of three lanes driving in the same direction, which is shown in Figure 2. The length of this highway is set to 4 km. Once a simulation episode is started, MDV will be continuously generated according to the negative exponential distribution from the leftmost starting point at each lane of the highway, with the average arrival rate of traffic flow λ set as 0.25 veh/s default. By tuning the value of the parameter λ, we can conveniently train and evaluate our proposed decision models under various traffic density conditions. Also, we can assign different values of λ two different lanes, therefore, the imbalance of traffic flow between lanes will be increased, which will potentially trigger more lane-changing needs to better evaluate our proposed model's applicability. The maximum time step of each episode is set to 200 and the time span of each step is set to 1 second, that is within each second, ADV and MDV should make corresponding actions according to the environmental state and their own decision-making models (ADV is driven by DQN-based model while MDV is driven by IDM and MOBIL). One episode will be terminated, and the next episode is started immediately when a collision occurs or the maximum episode duration is arrived.

(2) **Parameters of IDM and MOBIL.** In the simulation experiments, MDVs are driven by IDM and MOBIL for making a longitudinal and lateral decision. The related parameters of IDM and MOBIL are set according to Tables 2 and 3, which are mostly taken from [38].

(3) **Parameters of low-level control model.** In the low-level control layer, the parameters of the two-point visual control model are set according to Table 4.

Table 2: Parameters setting of IDM.

Parameters	Description	Values
$a_{\max}$	Maximum acceleration	0.6m/s^2
$a_{\min}$	Minimum deceleration	−20m/s^2
δ	Acceleration exponent	4
$d_{\min}$	Minimum gap	2 m
T	Safe time headway	1.6 s
b	Desired deceleration	1.7m/s^2
$d_{\max}$	Maximum gap for empty lane	10000 m

Table 3: Parameters setting of MOBIL.

Parameters	Description	Values
b_{safe}	Maximum safe deceleration	−4m/s^2
p	Politeness factor for new following vehicle	1
q	Politeness factor for old following vehicle	0.5
a_{th}	Changing threshold	0.1m/s^2

Table 4: Parameters setting of two-point visual control model.

Parameters	Description	Values
l_n	Distance to near point	5 m
l_f	Distance to far point	100 m
k_f	Proportional gain far point	20
k_n	Proportional gain near point	9
k_I	Integral gain near point	10 s^{-1}

(4) **Hyper-Parameters of DQN-based Decision Model.** We use a fully connected neural network with two hidden layers to realize the Q value function of ADV. The number of neurons in the first and second hidden layer is 128 and 64 and, the number of neurons in the input layer and output layer is 60 and 5, respectively, since each state is represented by a 60-dimension vector and the Q network will output corresponding values for 5 possible actions defined in action sets. The activation functions of hidden layers and output layer are set as RELU and linear, respectively. Also, the best values of other main hyper-parameters are chosen using the tree-structured parzen estimator (TPE) [39] through extensive simulation experiments, results of which are listed in Table 5.

Table 5: Optimal hyper-parameters of DQN-based model.

Hyper-parameters	Best values
Learning rate, α	0.001
Batch size	64
Size of replay buffer	100000
Discount factor, γ	0.9
ε for Greedy exploration	0.01

4.2. DQN-Based Decision Model Performance Analysis. In this section, we show the results about the performance evaluation of the DQN-based decision model of ADV. One metric (i.e., Training Loss) which is used to evaluate the learning performance of the proposed model, and two other metrics (i.e., Average Collision Rate, ACR and Average Episode Reward, AER) which are used to quantify the safety and efficiency of the proposed model are defined as follows:

(1) **Training loss**: The core task of DQN model training is to update the Q-value network according to the Equation (6) step by step with a batch of samples. In Equation (6), the item "$r(s,a) + \gamma Q_k(s',a') - Q_k(s,a)$" reflects the deviation between the estimated Q value and the true Q value. With the increase of training steps, it is expected that this deviation (i.e., Training Loss) to be smaller and smaller, which indicates that the learning of DQN tends to be stable.

(2) **Average Collision Rate (ACR).** ACR is equal to the number of collisions in each episode divided by the total number of decisions made by ADV. The collision counts both rear-end collisions and side-impact collisions. ACR reflects the safety performance of the autonomous driving decision-making model.

(3) **Average Episode Reward (AER).** AER is the total reward obtained in each episode divided by the number of decisions made. AER reflects the comprehensive performance of the autonomous driving decision-making model with respect to safety, efficiency, and lane change success rate.

The loss of the DQN model under 4000 and 65000 training steps are depicted in Figures 4(a) and 4(b) respectively. In Figure 4(a), no significant loss decrease is found, while in Figure 4(b), loss shows a trend of increasing first, then decreasing, and finally stabilizing. This reveals that when the number of training steps reaches enough, the DQN-based decision-making model proposed in this paper can achieve very good training performance.

Further, in order to evaluate the safety and efficiency of our proposed model, the changing curve of ACR and AER with respect to episodes is also depicted in Figures 5(a) and 5(b). It is obvious that although both ACR and ACR show a certain degree of oscillation, their average values tend to decrease and increase steadily. The results show that with the increasing of training steps, the decision-making model based on DQN proposed in this paper can achieve very good results in terms of driving safety and efficiency.

In order to make the changing trend of ACR and AER more clearly to be seen, we used a simple differential filtering method to process the time series values of them, and the results are shown in Figures 6(a) and 6(b).

4.3. Comparative Analysis between DQN and MOBIL&IDM. In this section, in order to further verify the efficiency of our proposed DRL-based model, we conduct simulation experiments to compare the safety and efficiency of our proposed DQN model with rule-based models (i.e., IDM and MOBIL). For ADV, we use DQN-based decision-making model and IDM combining with MOBIL to drive them for extensive simulation experiments separately, the AERs recorded are shown in Figure 7. It can be seen that MOBIL has a high average reward in the initial stage of the experiment, but with the increasing of training steps, DQN reaches an average reward higher than MOBIL by about 10% after full convergence, which means that DQN can do better in this multilane highway environment where exists dynamic and complex interactions between ADV and MDVs.

4.4. Other Variants of DQN-Based Decision Model. In this section, we further try other variants of the DQN model (i.e., DDQN (Double DQN) and Dueling DQN) to depict ADV's decision-making behavior. DDQN was proposed as a specific adaptation to the DQN algorithm to reduce the observed overestimations [40], while Dueling DQN uses a different network architecture with what is used in DQN to separate the estimation of the state value function and the state-dependent action advantage function [41]. Both DDQN and Dueling DQN are considered could improve the performance of DQN in some extent, so they are attempted to model the decision making of ADV, and the performance comparison between them and DQN are conducted separately.

4.4.1. DQN vs. DDQN. We systematically compare the model performance between DQN and DDQN with respect to ACR, AER, loss and Q value, results are shown in Figures 8(a)–8(d) respectively.

Figure 8 shows that in respect to the decision accuracy and the number of convergence episodes, the two algorithms show relatively similar learning efficiency (DDQN is faster in the early stage and DQN catches up in the later stage), and DDQN has more stable oscillation than DQN in terms of AER, ACR, and network loss, while the network loss and ACR are somewhat lower than DQN.

The Q value of DDQN is significantly lower than that of DQN, and it can be seen that after optimization by DDQN, the decision of the agent tends to be more conservative, which can theoretically have a higher decision accuracy in the application process.

4.4.2. DQN vs. Dueling DQN. The second improvement of DQN is the modification of its network structure. Both DQN and DDQN are single-branch network structures, and the improved Dueling DQN is a dyadic network structure. With

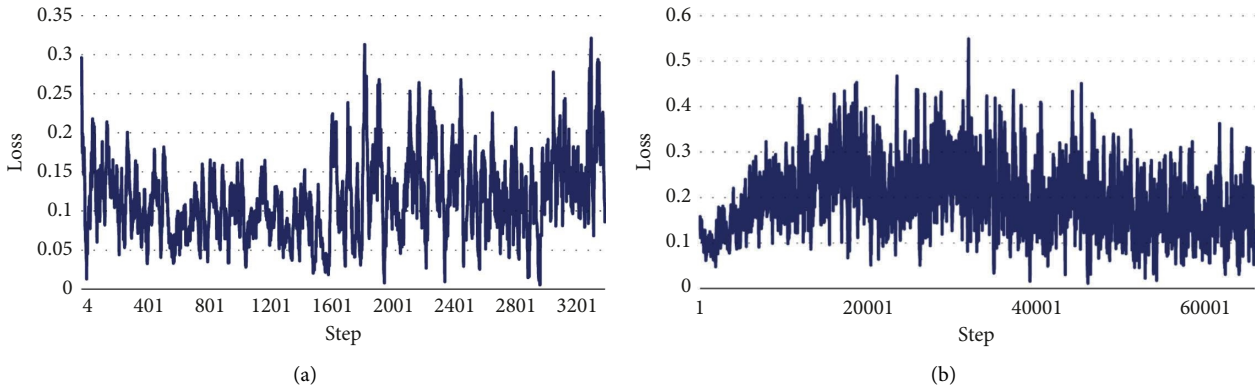

Figure 4: Training loss of DQN-based decision model of ADV.(a) 4000 steps. (b) 65000 steps.

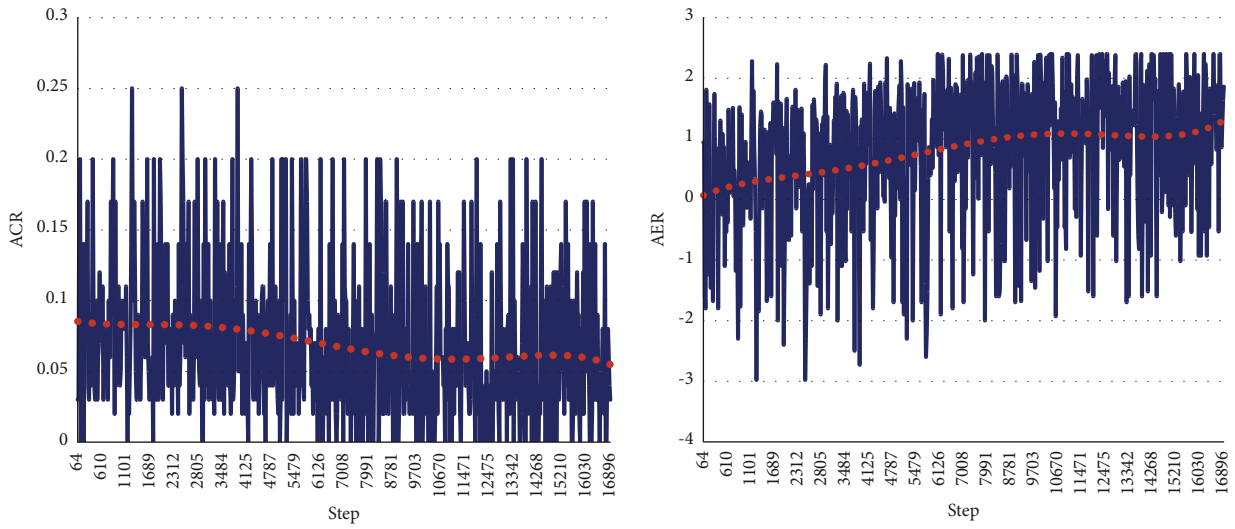

Figure 5: ACR and AER with respect to training steps.(a) ACR. (b) AER.

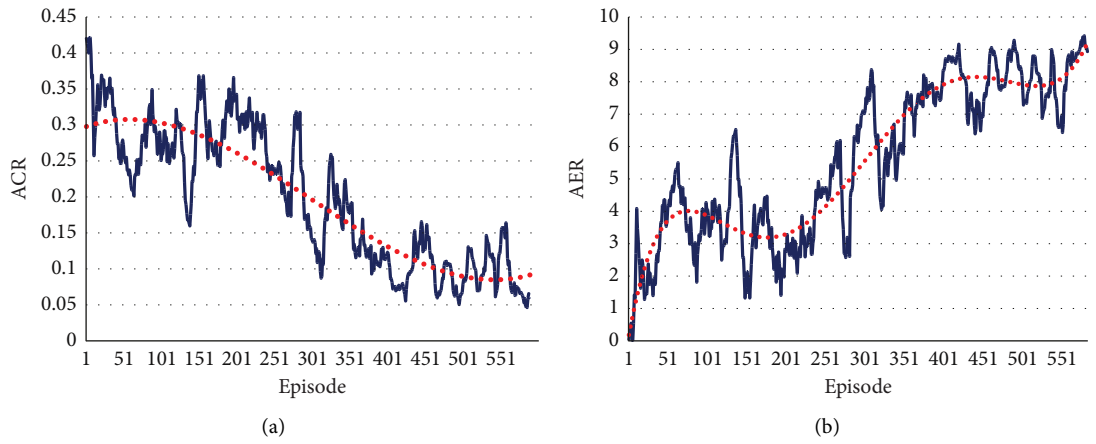

Figure 6: ACR and AER with respect to episodes after processing with differential filtering.(a) ACR. (b) AER.

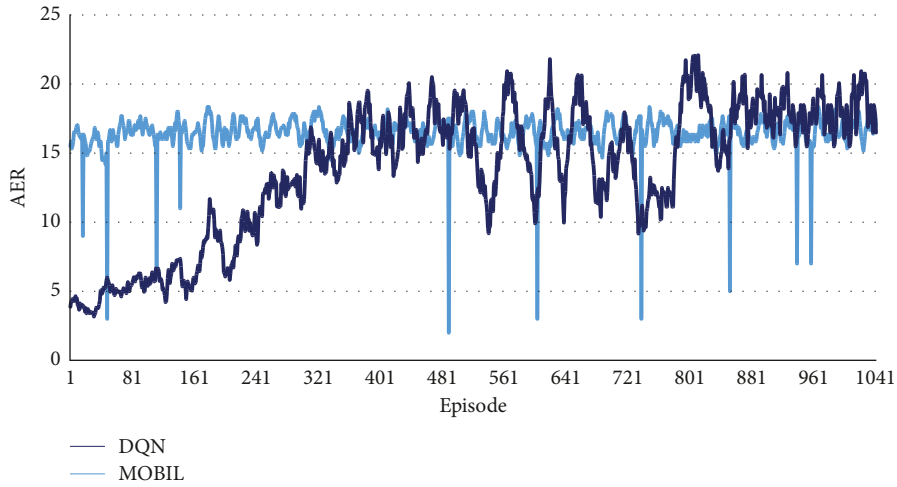

Figure 7: Comparison of AER between DQN and MOBIL&IDM.

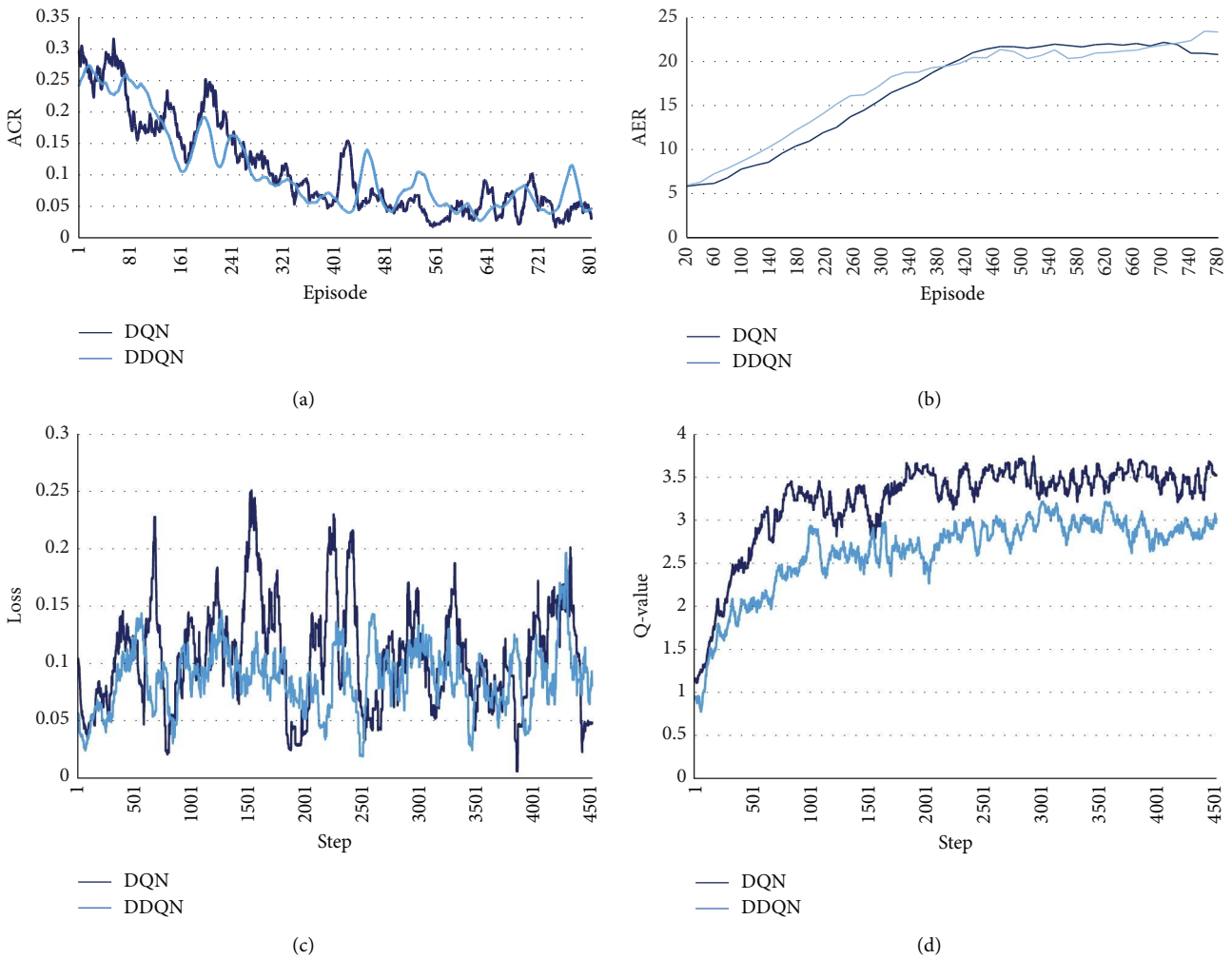

Figure 8: Performance comparison between DQN and DDQN. (a) ACR. (b) AER.(c) Loss. (d) Q value.

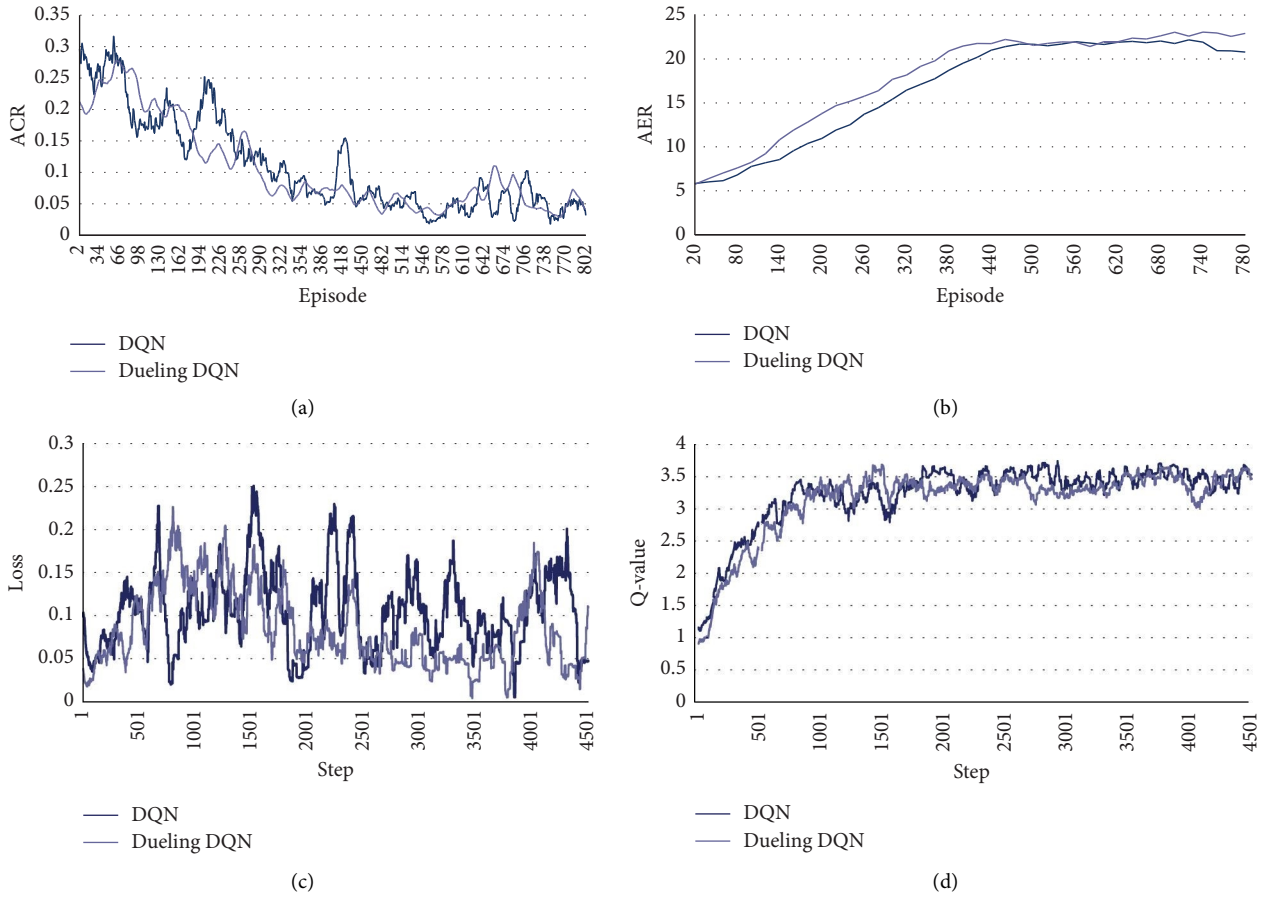

Figure 9: Performance comparison between DQN and Dueling DQN. (a) ACR. (b) AER.(c) Loss. (d) Q-value.

the unchanged input, the output of Dueling DQN will go through two fully connected layer branches corresponding to state values and dominance values, updating the scores of all actions in each iteration, instead of just taking the maximum value as in DQN. The algorithm can increase the convergence speed to some extent under the influence of different network structures. Dueling DQN is also compared with DQN in simulation experiments with respect to ACR, AER, loss and Q value, and experimental results are shown in Figures 9(a)–9(d).

Figure 9 shows that for ACR and AER, Dueling DQN converges almost 10% faster than DQN, can have less oscillation performance in a short time, the loss of Dueling DQN is smaller than DQN, and Q value is comparable to DQN. Overall, under the same hyper-parameter configuration (e.g., learning rate), Dueling DQN can indeed perform better than DQN in terms of ACR, AER, and loss.

4.5. Performance Analysis of DQN-Series Models with Different ADV Penetration. In the previous sections, the considered decision-making scenarios of automatic driving on multilane highway are all mixed travel of a single ADV and multiple MDVs. In this section, we want to investigate the performance of DQN, DDQN and Dueling DQN-based models under different ADV penetration (i.e., the proportion of ADV in all travel vehicles) with respect to episodes needed to converge (i.e., convergence episode). Results about a number of convergence episodes of DQN, DDQN, and Dueling DQN-based models are shown in Table 6.

Table 6: Comparison of convergence episodes with different ADV penetration.

ADV: MDV	DQN	DDQN	Dueling DQN
1 : 100	823	792	691
1 : 50	1767	1743	1531
1 : 30	>2000	>2000	1958
1 : 10	>2000	>2000	>2000

In general, as the number of ADVs increases, the deep reinforcement learning algorithm (i.e., DQN, DDQN, and Dueling DQN) learns and masters the state of the environment more and more difficult, and the average convergence episodes gradually grows, and even fails to converge in finite time (i.e., convergence episode >2000). DQN and DDQN comparing with Dueling DQN converge more slowly. The superior performance of Dueling DQN is attributed to its optimized network structure based on DQN. DDQN optimizes the update logic of DQN and is able to acquire higher Q value, but it does not produce a significant

advantage over DQN in the selection of discrete behaviors, such as vehicle lane change decisions, so the performance improvement is limited. In general, due to the increase of ADV, the state faced by each ADV in the mixed travel environment is more complex, dynamic and in essence nonstationary, it is difficult for ADV to learn a stable policy for decision making and consequently leads to much more convergence episodes needed. Actually, when ADV increasing, multi-agent reinforcement learning [42] can be a good choice to model their collective decision-making behaviors, which may be our research direction to be explored in the future.

5. Conclusions

This paper proposes a reinforcement learning-based decision-making model for autonomous driving on a multilane highway with mixed traffic composed of ADV and MDV. By proper state representation, action set definition and reward function design, DQN, DDQN, and Dueling DQN-based models are developed for automatic making of both lateral and longitudinal decisions. At the same time, in order to construct the simulation environment of mixed traffic, we describe in detail the rule-based decision behavior models (i.e., IDM and MOBIL) which are used to generated decision for MDV vehicles. Further, low-level control of both ADV and MDV is realized by a nonlinear kinematic bicycle model combining with a two-point visual control model.

Through extensive simulation experiments, the safety and efficiency for autonomous driving decision making by DQN, DDQN, and Dueling DQN is verified. Comparing the experimental results of DQN and its variant models with the rule-based decision-making model, it is found that, deep reinforcement learning-based models for decision making of autonomous driving are generally superior to rule-based methods with respect to safety, efficiency, and generalization ability. It is also found, with the increasing of ADV penetration in mixed traffic flow, the training and generalization of DRL-based models becomes more and more difficult, therefore, multi-agent reinforcement learning, through joint consideration of environmental observation and collective decision-making of ADV vehicles, may be an important research direction in the future.

Data Availability

All data and code generated in our study are available at zhaoboyuan825/An-Integrated-Lateral-and-Longitudinal-Decision-Making-Model: code of model and simulation (github.com).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

The authors confirm contribution to the paper as follows: study conception and design: Jianxun Cui and Boyuan Zhao; experiments setup: Jianxun Cui and Boyuan Zhao; analysis of results: Jianxun Cui and Boyuan Zhao; draft manuscript preparation and revision: Jianxun Cui, Boyuan Zhao and Mingcheng Qu.

Acknowledgments

This research was supported by the joint guidance project of Heilongjiang Provincial Natural Science Foundation through Grant #LH2021E074 and the http://dx.doi.org/10.13039/501100012226Fundamental Research Funds for the Central Universities through Grant #HIT.NSRIF202235.

References

[1] Z. Qiao, K. Muelling, J. M. Dolan, P. Palanisamy, and P. Mudalige, "Automatically generated curriculum based reinforcement learning for autonomous vehicles in urban environment," in *Proceedings of the 29th IEEE Intelligent Vehicles Symposium (IV 2018)*, Changshu, China, June 2018.

[2] B. R. Kiran, I. Sobh, V. Talpaert et al., "Deep reinforcement learning for autonomous driving: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 1–18, 2021.

[3] H. Ahn, K. Berntorp, P. Inani, A. J. Ram, and S. Di Cairano, "Reachability-based decision-making for autonomous driving: theory and experiments," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 5, pp. 1–15, 2020.

[4] S. Zhao, Q. Hou, and Y. Zhai, "Decision mechanism of vehicle autonomous lane change based on rough set theory," in *Proceedings of the ICCIR 2021: 2021 International Conference on Control and Intelligent Robotics*, New York NY USA, June 2021.

[5] X. Liu, X. Qu, and X. Ma, "Improving flex-route transit services with modular autonomous vehicles," *Transportation Research Part E Logistics and Transportation Review*, vol. 149, p. 102331, 2021.

[6] Y. Liu, X. Wang, L. Li, S. Cheng, and Z. Chen, "A novel lane change decision-making model of autonomous vehicle based on support vector machine," *IEEE Access*, vol. 7, pp. 26543–26550, 2019.

[7] T. Yin, Y. Li, J. Fan, and Y. A. Shi, "A novel gated recurrent unit network based on svm and moth-flame optimization algorithm for behavior decision-making of autonomous vehicles," *IEEE Access*, vol. 9, p. 99, 2021.

[8] H. Chao, M. Chen, H. Peng, and Y. Xing, "Toward safe and personalized autonomous driving: decision-making and motion control with dpf and cdt techniques," *IEEE*, vol. 26, no. 2, p. 99, 2021.

[9] L. Xin, X. Xin, and Z. Lei, "Reinforcement learning based overtaking decision-making for highway autonomous driving," in *Proceedings of the 6th International Conference on Intelligent Control & Information Processing*, November 2015.

[10] A. E. Sallab, M. Abdou, E. Perot, and S. Yogamani, "Deep reinforcement learning framework for autonomous driving," *Electronic Imaging*, vol. 29, pp. 70–76, 2017.

[11] F. Ye, X. Cheng, P. Wang, C. Y. Chan, and J. Zhang, "Automated lane change strategy using proximal policy optimization-based deep reinforcement learning," in *Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV)*, Las Vegas, NV, USA, February 2020.

[12] H. Wang, H. Gao, S. Yuan et al., "Interpretable decision-making for autonomous vehicles at highway on-ramps with

latent space reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, p. 99, 2021.

[13] M. J. Shin and J. Kim, "Randomized adversarial imitation learning for autonomous driving," 2019, https://arxiv.org/abs/1905.05637.

[14] J. Chen, B. Yuan, and M. Tomizuka, "Deep imitation learning for autonomous driving in generic urban scenarios with enhanced safety," in *Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, November 2019.

[15] A. Jamgochian, E. Buehrle, J. Fischer, and M. J. Kochenderfer, "Safety-aware hierarchical adversarial imitation learning for autonomous driving in urban environments," 2022, https://arxiv.org/pdf/2204.01922.pdf.

[16] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model mobil for car-following models," *Transportation Research Record*, vol. 1999, pp. 86–94, 2007.

[17] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review*, vol. 62, no. 2, pp. 1805–1824, 2000.

[18] P. Wang, C. Y. Chan, and A. Fortelle, "A reinforcement learning based approach for automated lane change maneuvers," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, Changshu, China, June 2018.

[19] M. Buehler, K. Iagnemma, and S. Singh, *The DARPA Urban Challenge: Autonomous Vehicles in City Traffic*, Springer, Berlin, Germany, 2009.

[20] B. Paden, M. Cap, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 33–55, 2016.

[21] X. Ma, H. Zhong, Y. Li, J. Ma, Z. Cui, and Y. Wang, "Forecasting transportation network speed using deep capsule networks with nested LSTM models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 4813–4824, 2021.

[22] C. Vallon, Z. Ercan, A. Carvalho, and F. Borrelli, "A machine learning approach for personalized autonomous lane change initiation and control," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, Los Angeles, CA, USA, June 2017.

[23] H. Bi, T. Mao, Z. Wang, and Z. Deng, "A data-driven model for lane changing in traffic simulation," in *Proceedings of the ACM SIGGRAPH Symposium on Computer Animation*, Goslar, Germany, June 2016.

[24] D. C. K. Ngai and N. H. C. Yung, "Automated vehicle overtaking based on a multiple-goal reinforcement learning framework," in *Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference Seattle*, pp. 818–823, Bellevue, WA, USA, September 2007.

[25] P. Wang and C. Y. Chan, "Formulation of deep reinforcement learning architecture toward autonomous driving for on-ramp merge," in *Proceedings of the IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, Japan, October 2017.

[26] C. J. Hoel, K. Wolff, and L. Laine, "Automated speed and lane change decision making using deep reinforcement learning," in *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, USA, November 2018.

[27] D. Isele, R. Rahimi, A. Cosgun, K. Subramanian, and K. Fujimura, "Navigating occluded intersections with autonomous vehicles using deep reinforcement learning," in *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, Australia, May 2018.

[28] L. Xin, S. E. Li, P. Wang et al., "Accelerated inverse reinforcement learning with randomly pre-sampled policies for autonomous driving reward design," in *Proceedings of the IEEE Intelligent Transportation Systems Conference ITSC*, Auckland, New Zealand, October 2019.

[29] Z. Wu, L. Sun, W. Zhan, C. Yang, and M. Tomizuka, "Efficient sampling-based maximum entropy inverse reinforcement learning with application to autonomous driving," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5355–5362, 2020.

[30] J. Ho and S. Ermon, "Generative adversarial imitation learning," pp. 4565–4573, 2016, https://arxiv.org/abs/1606.03476.

[31] C. Finn, S. Levine, and P. Abbeel, "Guided cost learning: deep inverse optimal control via policy optimization," pp. 49–58, 2016, https://arxiv.org/abs/1603.00448.

[32] J. Fu, K. Luo, and S. Levine, "Learning robust rewards with adversarial inverse reinforcement learning," 2017, https://arxiv.org/abs/1710.11248.

[33] S. Reddy, A. D. Dragan, and S. Levine, "Imitation learning via reinforcement learning with sparse rewards," 2019, https://arxiv.org/abs/1905.11108.

[34] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Playing atari with deep reinforcement learning," 2013, https://arxiv.org/abs/1312.5602.

[35] S. Gu, E. Holly, T. Lillicrap, and S. Levine, "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates," in *Proceedings of the IEEE International Conference on Robotics & Automation*, New York NY USA, May 2017.

[36] Z. Bai, W. Shangguan, B. Cai, and L. Chai, "Deep reinforcement learning based high-level driving behavior decision-making model in heterogeneous traffic," in *Proceedings of the Chinese Control Conference (CCC)*, Guangzhou, China, July 2019.

[37] D. D. Salvucci and R. Gray, "A two-point visual control model of steering," *Perception*, vol. 33, no. 10, pp. 1233–1248, 2004.

[38] A. Alizadeh, M. Moghadam, Y. Bicer, N. K. Ure, U. Yavas, and C. Kurtulus, "Automated lane change decision making using deep reinforcement learning in dynamic and uncertain highway environment," in *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, October 2019.

[39] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kegl, "Algorithms for hyper-parameter optimization," Proceedings of the 24th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, pp. 2546-2554, December 2011.

[40] V. H. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proceedings of the National Conference on Artificial Intelligence AAAI Press*, New York NY USA, February 2016.

[41] Z. Y. Wang, N. Freitas, and M. Lanctot, "Dueling network architectures for deep reinforcement learning," 2015, https://arxiv.org/abs/1511.06581.

[42] L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, no. 2, pp. 156–172, 2008.