Research Article

A Novel Dynamic Lane-Changing Trajectory Planning Model for Automated Vehicles Based on Reinforcement Learning

Cenxin Yu, Anning Ni, Jing Luo, Jinghui Wang, Chunqin Zhang, Qin Qin Chen, and Yifeng Tu

1Department of Transportation Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
2School of Civil Engineering and Architecture, Zhejiang Sci-Tech University, Hangzhou, Zhejiang 310018, China
3Zhejiang CICO Expressway Construction Management Co., Ltd, Hangzhou, Zhejiang 310020, China

Correspondence should be addressed to Anning Ni; nianning@sjtu.edu.cn

Received 5 July 2021; Revised 21 February 2022; Accepted 26 April 2022; Published 21 May 2022

1. Introduction

Lane changing behavior has a significant impact on traffic efficiency and may lead to traffic delays or even accidents. It is important to plan a safe and efficient lane-changing trajectory that coordinates with the surrounding environment. Most conventional lane-changing models need to establish and solve constrained optimization models during the whole process, while reinforcement learning can just take the current state as input and directly output actions to vehicles. This study develops a lane-changing model using the deep deterministic policy gradient method, which can simultaneously control the lateral and longitudinal motions of the vehicle. To optimize its performance, a reward function is properly designed by combining safety, efficiency, gap, headway, and comfort features. To avoid collisions, a safety modification model is developed to check and correct acceleration at every time step. The driving trajectory data of 1169 lane-changing scenarios extracted from the Next Generation Simulation (NGSIM) dataset are used to train and test the model. The proposed model can quickly converge in training phase. Testing results show it can complete safe and efficient lane changing in different lane-changing scenarios with both shorter time headway and lane-changing duration than human drivers. Compared with the conventional dynamic lane-changing trajectory planning model, our model can reduce collision risk. It is also evaluated in automated and nonautomated mixed traffic in SUMO. Simulation results show that the proposed model also has a positive effect on the average speed of overall traffic flow.

Lane changing behavior is classified into two categories: mandatory and discretionary [1]. Either may perturb traffic and result in traffic accidents or significant delay [2–4]. Dijkstra and Heijden [5] found that nearly 4% to 10% of accidents were caused by lane changing. In Canada, about 10% of large-scale vehicle accidents were induced by lane-changing motions [6]. It has been found that human factor, such as distraction and insufficient driving experience, is one of the major causes of inappropriate lane changing [7–11].

Automated vehicles (AVs) have attracted increasing interest in recent years. They offer a huge potential to improve the operation and safety of transportation systems [12, 13]. Studies on AVs’ lane changing include two main aspects: lane-changing decision and lane-changing execution. The former determines whether lane changing is performed [14], and the latter addresses how lane changing is performed [15]. Lane-changing execution is more complex because it requires simultaneous control of longitudinal and lateral movements within a period of time to ensure safety and reach the destination. To limit the scope of this study, we will focus on the lane-changing execution problem.

Conventional lane-changing models of AVs are based on predefined rules and clearly designed models [16]. Most of them plan a reference trajectory curve from the current position to the target position and then use a tracking controller to ensure that the vehicles move along the curve. Based on the trajectory planning method, they can be
divided into two categories: static lane-changing trajectory planning (SLTP) and dynamic lane-changing trajectory planning (DLTP) [15]. SLTP plans a trajectory from the current position to the target position before starting to change lanes, and then the vehicle moves along it. DLTP can update lane-changing trajectories in real time in response to the most recent traffic conditions. In reality, the surrounding environment in the lane-changing scenario changes all the time, and the lane-changing vehicle should react to it timely. However, previous research on dynamic-method still have disadvantages such as unrealistic assumptions on velocity and acceleration or ignoring behaviors of proceeding vehicle on the original lane.

Machine learning, especially reinforcement learning (RL), is another method to deal with control problems for AVs. Lane changing is a time-sequential problem, which needs a sequence of actions to achieve the destination, and the performance of the current action will influence the final goal. RL is quite suitable for this kind of problem [17]. Considering that the actions in lane-changing scenarios, such as acceleration and steering angle, are continuous variables, we use a deep deterministic policy gradient (DDPG), as it can be applied in problems with continuous action space.

This study focuses on the lane-changing execution problem for AVs. We proposed a lane-changing model based on the DDPG algorithm. Lane-changing trajectory data from the NGSIM dataset, which provides real lane-changing scenarios, were used to train and test the model. A reward function combining safety, efficiency, gap, headway, and comfort features was properly designed to measure and shape AVs’ lane-changing behaviors. A safety modification model was used to check and correct acceleration to avoid collisions at every time step. The proposed model can simultaneously control the lateral and longitudinal motions of AVs to realize safe and efficient lane changing with smaller headway and shorter lane-changing duration.

The rest of this study is organized as follows: the next section sorts out the previous-related research. The methodology section shows the DDPG-based lane-changing model in detail. The experiments and results analysis section presents the experimental results and discussions. In the final section, we make conclusions and outline the topics of future research.

2. Literature Review

The conventional lane-changing models can be divided into two steps: trajectory planning and trajectory tracking. The trajectory planning generates one or more curves from the current position to the target position. The most commonly used method is the geometric curve method. It is assumed that the lane-changing vehicle will move along a predetermined curve. Typical geometric curves are polynomial curve [15, 18], Bessel curve [19], trapezoidal curve [20], etc. However, most of the above studies are SLTP models. They assume that the velocity of the surrounding vehicles does not change during the entire lane-changing process, which is obviously inconsistent with reality. The lane-changing vehicle cannot respond to the changes in the surrounding environment in time, resulting in lane-changing failures and even collisions. Some researchers try to use DLTP models to make up for the above shortcomings. Luo et al. [18] proposed a time-dependent trajectory planning method. They used a quintic polynomial as the trajectory function and turned the problem into an optimization problem based on lane-changing time and distance. However, their study makes an unrealistic assumption that the final velocity, which is unknown in reality, is the average velocity of the target lane. The dynamic trajectory planning model proposed by Yang et al. [15] improves on this problem, as it can dynamically adjust velocity according to real-time lane-changing environment instead of making any assumption on velocity. However, this research ignores the driving behavior of the preceding vehicle in the original lane during the lane changing.

After trajectory planning, a controller is required to track the vehicle to ensure that it can reach the target lane along the predetermined trajectory curve. The control actions are usually obtained by solving a constrained optimization model. Both Luo et al. [18] and Yang et al. [15] minimize the cost of comfort and efficiency. In addition, model predictive control (MPC) is also often used in the control problems of AVs. Wang et al. [21] proposed an automatic lane-changing system in highway driving with the MPC-based trajectory tracking controller. Yue et al. [22] proposed a lane-changing control system combining a polynomial-based trajectory planning and a robust tube model predictive control (RTMPC). Wang et al. [23] proposed an adaptive control algorithm consisting of a six-order polynomial lane-changing trajectory planning and a model predictive controller based on fuzzy control (FMPC). Larsson et al. [24] proposed a multiobjective model for CAV in mixed traffic based on MPC, to realize prosocial behavior. It can improve efficiency and comfort of the traffic flow with altruistic actions, instead of selfish actions. A constrained optimization model performs well if the situations and model limits are predefined. However, the objection function and constraint conditions need to adjust according to different actual situations. The complex real-world scenarios will also increase the burden of constraint formulation and model solving. Moreover, it may not be able to find a solution to the constrained optimization problem, leading to failed control inputs generation [25].

In recent years, RL is increasingly used in vehicle control problems in ramp metering [26, 27], intersection [28, 29], and freeway work zone [30], in order to improve traffic conditions. RL-based control models for AVs have a positive impact on traffic safety and efficiency [25, 31, 32]. For lane-changing problems of AVs, most RL-based studies focus on the lane-changing decision process [33–38], and the application of RL to lane-changing trajectory planning still needs to be supplemented. Wang et al. [39] proposed a lane-changing model for intelligent vehicles based on a Q-learning algorithm to realize smooth and efficient lane changing. Wang et al. [40] used DDPG to train the continuous lane-changing behavior. Both models take yaw acceleration as the only continuous action, and the
longitudinal motion of the vehicle is controlled by the intelligent driver model (IDM). Shi et al. [17] proposed a deep Q-learning framework to make lane-changing decisions and adjust longitudinal acceleration. The lateral movement is controlled by a pure pursuit controller. Naveed et al. [41] proposed a trajectory planning method that combines hierarchical reinforcement learning (HRL) and proportion-integral-derivative (PID) controller. In the HRL model, the target waypoint and target velocity are determined first, and then the PID controller will plan the trajectory based on the current waypoint, current velocity, target waypoint, and target velocity. However, their reward function mainly consists of fixed constants, which may reduce the learning ability in continuous control problems. Hu et al. [16] used the DDPG model to control both control steering wheel angle and longitudinal acceleration for lane-changing vehicles. However, they only use a security feature in the reward function to avoid collisions. This is just a soft constraint and cannot guarantee complete safety since RL is likely to take unsafe behaviors without a hard constraint [42]. As mentioned above, although there are some studies on RL-based lane-changing control for AVs, disadvantages still exist. Therefore, a novel DDPG-based lane-changing model is proposed for AVs. Zhu et al. [25] proposed a model for car-following velocity control based on RL. Results show it can realize safe, efficient, and comfortable car following and outperform human drivers. They optimized the longitudinal actions of AVs well. Based on this model, this study added the lateral movement to expand their car-following model to the lane-changing model. Unlike the conventional lane-changing models, the DDPG algorithm combines trajectory planning and trajectory tracking in one model. There is no need to establish and solve a constrained optimization model during the lane-changing process. Instead, this model takes the current state as input and directly outputs both lateral and longitudinal actions to the lane-changing vehicle, with a properly designed reward function consisting of safety, efficiency, gap, headway, and comfort features. In addition, a safety modification model, which considers the emergency brake situation, is used to check and correct the acceleration got from the DDPG model at every time step, ensuring security in the whole lane-changing process.

3. Methodology

3.1. Preliminaries

3.1.1. Reinforcement Learning. RL is composed of agent, environment, state, action, and reward. After an agent performs an action $a_t$, the environment will change from state $s_t$ to the new state $s_{t+1}$. For the new state $s_{t+1}$, the environment will give a reward signal $r_t$ (positive reward or negative reward) to the agent. Then, with the reward of the new state and environment feedback, the agent performs the new action according to a certain policy $\pi$. It is an interactive process between the agent and the environment through state, action, and reward. RL aims to get the maximum accumulative reward $R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$, where $\gamma$ is the discount factor ranging from 0 to 1 [43]. The agent will finally learn the optimal policy to achieve its goal.

3.1.2. Deep Reinforcement Learning. RL has strong decision-making ability, but it lacks enough representation ability, leading to the inability to solve the perception problem in the actual environment. Deep reinforcement learning (DRL) combines the representation ability of deep learning and the decision-making ability of RL, so that the agent has a stronger learning ability and then can better solve the perception problem of complex systems. To apply the DRL algorithm to the continuous action space, the DDPG algorithm was proposed by Lillicrap et al. [44]. The DDPG algorithm combines a deep neural network with a deterministic policy gradient, and its basic structure is the actor-critic algorithm. As shown in Figure 1, the actor network is the policy network. It is used to update the policy and output the deterministic action value after the input of the current state. The critic network is the value network. It is used to approximate the value function of the state-action pair and provide gradient information. The DDPG algorithm creates two neural networks for the policy network and the value network, respectively. One is the online network, and the other is the target network. After the training of a mini batch of sample data, the parameters of the online network are updated by gradient up or gradient down algorithm, and then the parameters of the target network are updated by soft update algorithm. The dual neural network model structure makes the learning process more stable and faster to converge.

3.2. Scenario and Problem Description. This study studies the lane-changing scenario on the highway without the interference of traffic lights in the mixed traffic flow of human-driven vehicles (HDVs) and AVs. Such a scenario is usually consisting of four vehicles, as shown in Figure 2. SV is the lane-changing vehicle. LVO is the leading vehicle on original lane. LV is the leading vehicle on the target lane. FV is the following vehicle on the target lane. The SV is an AV, while the LVO, LV, and FV may be HDV or AV. During the control process, we assume that there is no communication between vehicles. The AV can use sensors to obtain the position and velocity information of its surrounding vehicles. Besides, information transmission delay and error are not taken into consideration in this study.

3.3. Data Preparation. To train the reinforcement learning model, lane-changing scenarios should be provided. In this study, real lane-changing scenarios data were extracted from the Next Generation Simulation (NGSIM) dataset [45], which can be downloaded from https://catalog.data.gov/dataset/next-generation-simulation-ngsim-vehicle-trajectories-and-supporting-data. As this study studies the lane-changing scenario on the highway, the US-101 dataset and the I-80 dataset were used. The US-101 dataset collected data on a 640-meter section. A total of 45 minutes of data are
available, from 7:50 a.m. to 8:35 a.m. The I-80 dataset collected data on a 503-meter section. There is also a total of 45 minutes of data, including periods of 4:00 p.m. to 4:15 p.m. and 5:00 p.m. to 5:30 p.m. The schematic illustrations of the dataset are shown in Figure 3. After data smoothness, the lane-changing trajectory data were firstly extracted. There have been methods to extract lane-changing trajectory data in previous research [46, 47]. This study uses the method proposed by Thieman et al. [47] to determine the lane-changing process, as it can exclude incomplete lane-changing trajectory data. The lane-changing trajectory data were firstly extracted based on the following criteria:

1. Only discretionary lane-changing scenarios.
2. Single Lane-Changing Scenario. That is to say, the interval between two lane changing of the same vehicle needs to be greater than 5 seconds.
3. The type of lane-changing vehicle is auto.
4. The vehicle has one-way continuous lateral movement, and the starting position and ending position must be on the original lane and the target lane.

Based on the analysis of the lane-changing trajectories, it was found that most lane changing can be completed in 10 seconds. Therefore, the duration of lane-changing scenario was set to be 10 seconds, with 5 seconds on the original lane and 5 seconds on the target lane. Meanwhile, the trajectory data of the LVO, LV, and FV during this period were also extracted. Among them, the lane-changing scenarios near the start or end of the detection section, which means the vehicle ID of LVO, LV, or FV is zero, should be eliminated because of the incomplete data. Finally, 1169 lane-changing scenarios were extracted.

3.4. DDPG-Based Lane-Changing Model

3.4.1. Model Framework. The whole lane-changing model framework is shown in Figure 4. The DDPG model gives actions based on the current state of the SV and corrects the action values through the safety modification model to avoid collisions. The SV reaches the next state after performing safe actions. This process is repeated until the lane changing finishes.

3.4.2. State Space. In the lane-changing process, SV’s behavior is closely related to LVO, LV, and FV. The state space includes the SV’s velocity $v$, direction angle $\theta$, the lateral distance to the centerline of the target lane $X_{dt}$, the longitudinal distance from the front of SV to the back of LVS$_{LV}$, the longitudinal distance from the front of the SV to the back of LVS$_{LVO}$, the relative velocity to LV $R_{LV}$, and the relative velocity to LVO $R_{LVO}$, as shown in Figure 5:

$$S = (v, \theta, X_{dt}, S_{LV}, R_{LV}, R_{LVO}).$$

(1)

It should be noted that the influence of FV is not concluded in the state space. In reality, SV will consider FV’s state before lateral position change starts to avoid the collision. Once the lane changing begins, the FV’s behavior may be greatly affected by the SV in a lane-changing process, and
it will automatically adjust its velocity to ensure a safe gap to SV [46]. In our study, FV’s state is just used to decide whether the lane changing can be started. The SV will start to change lane when both the front gap to LV and the rear gap to FV are larger than the threshold.

3.4.3. Action Space. The change of SV’s state is decided by its velocity $v$ and direction angle $\theta$. The SV should learn how to adjust its velocity and direction angle to get its destination, so the action space is defined with acceleration and yaw acceleration, both of which are continuous actions:

$$A = (\text{acc}, a_{yaw}).$$ (2)

3.4.4. State Transition Model. The state transition model presents how the velocity, direction, and position of the SV at time step $t$ change to the new state at time step $t + 1$, with the determined action $(\text{acc}, a_{yaw})$. It can be simplified as a kinematic point-mass model, which has been widely used in many previous studies [18, 25, 32, 48]:

$$v_{t+1} = v_t + \text{acc} \times \Delta t,$$

$$\theta_{t+1} = \theta_t + a_{yaw},$$

$$y_{t+1} = y_t + \frac{v_t \cos \theta_t + v_{t+1} \cos \theta_{t+1}}{2} \times \Delta t,$$

$$x_{t+1} = x_t + \frac{v_t \sin \theta_t + v_{t+1} \sin \theta_{t+1}}{2} \times \Delta t,$$ (3)

where $v_t$, $\theta_t$, $y_t$, and $x_t$ are the velocity, direction angle, longitudinal position, and lateral position of the SV at time step $t$; $v_{t+1}$, $\theta_{t+1}$, $y_{t+1}$, and $x_{t+1}$ are the velocity, direction angle, longitudinal position, and lateral position of the SV at time step $t + 1$; $\Delta t$ is the simulation time interval, which is set as 0.1 s in this study. The acceleration $\text{acc}$ is bound between $-3 \text{ m/s}^2$ and $3 \text{ m/s}^2$, and the yaw acceleration $a_{yaw}$ is bound between $-1^\circ$ and $1^\circ$ every 0.1 second, based on the values of selected lane-changing data.

3.4.5. Reward Function. In the DDPG algorithm, the reward function is the only measure to evaluate the agent’s action. Therefore, designing an appropriate reward function is the most important to shaping the behavior of the SV. In a lane-changing scenario, the SV should be able to get ready to change lane quickly, move to its destination lane safely and comfortably, and maintain a proper headway to the leading vehicle. To get this goal, the reward function is a linear combination of five manually chosen features: comfort, safety, gap, efficiency, and headway.

The comfort depends on the change rate of acceleration and yaw acceleration. The lane-changing process should be smooth and reduce the impact on the comfort of drivers and passengers. The comfort feature is constructed as follows:

$$F_{\text{comfort}} = \frac{(-\text{acc}_t - \text{acc}_{t-1})^2/36 - (a_{\text{yaw}_t} - a_{\text{yaw}_{t-1}})^2/4}{2}.$$ (4)

The base values 36 and 4 were determined by the action bound, as the acceleration is bound between $-3 \text{ m/s}^2$ and $3 \text{ m/s}^2$, and the yaw acceleration is bound between $-1^\circ$ and $1^\circ$ every 0.1 second. The boundary values are set based on the
During the lane-changing process, the SV should also on the whole traffic flow and enlarges road capacity [53].

It was found that the minimum gap from the extracted lane-changing data is about 2.5 m. Therefore, this study sets a constant of 2.5 m as the threshold to consider starting the lane changing. The acceleration is between $-3 \text{ m/s}^2$ and $3 \text{ m/s}^2$, and the yaw acceleration is between $-1^\circ$ and $1^\circ$ every 0.1 second.

Safety is the most important during the lane-changing process. In this study, time to collision (TTC), a widely used safety indicator introduced by Hayward [49], was chosen to construct the safety feature. The smaller the TTC value is, the higher the collision risk is. There are different minimum TTC requirements in previous literature, ranging from 1.5 s to 4 s [25, 50–52]. In this study, 3 s is adopted as the TTC threshold. If TTC is less than 3 s, the feature will be negative. As the number approaches zero, the result becomes negative infinity, to represent a penalty for collision risk:

$$F_{\text{safe}} = \begin{cases} \ln \left( \frac{TTC}{3} \right), & 0 < TTC \leq 3 \\ 0, & \text{others} \end{cases}$$

Once the SV has made the decision to change lane and determined the target gap, it should adjust its velocity to get ready to start lane change. The gap feature is constructed by $S_{LV}$. It was found that the minimum gap from the extracted lane-changing data is about 2.5 m. Therefore, this study sets a constant of 2.5 m as the threshold to consider starting the lane changing.

$$F_{\text{gap}} = \begin{cases} e^{\left[2.5-S_{LV}\right]}, & S_{LV} < 2.5m \\ 1, & S_{LV} \geq 2.5m \end{cases}$$

The goal of lane changing is to get the centerline of the target lane. So, the lateral distance to the target centerline is used to construct the efficiency feature. The lane width is 3.6 m. $\left| X_{\text{dis}} \right| \leq 5.4m$ means the forward direction is right, while $\left| X_{\text{dis}} \right| > 5.4m$ represents that the SV moves in the opposite direction beyond the lane bound, and another $-1$ is added as a penalty.

$$F_{\text{efficiency}} = \begin{cases} e^{-\left| X_{\text{dis}} \right|}, & \left| X_{\text{dis}} \right| \leq 5.4 \\ e^{-\left| X_{\text{dis}} \right|} - 1, & \left| X_{\text{dis}} \right| > 5.4 \end{cases}$$

Keeping a safe and short headway has a positive impact on the whole traffic flow and enlarges road capacity [53]. During the lane-changing process, the SV should also maintain an appropriate time headway, as it can reduce the impact of lane changing on the traffic flow. This study extracted the appropriate time headway from lane-changing data. Based on the analysis of the data, it was found that the distribution of the logarithm of time headway could fit the Gaussian distribution well ($R^2 = 0.985$), as shown in Figure 6. The maximum probability appears when $x$ is 0.5625, which means the time headway is 1.75 s. This value was used as the most appropriate time headway in the headway feature:

$$f(x) = 0.0089 e^{-\left(\frac{x - 0.5625}{0.622}\right)^2}, x = \ln(\text{head.way}),$$

$$F_{\text{headway}} = e^{-1.75 - h}, h > 0.$$  

After the definition of five features, the reward function can be constructed as follows:

$$R = w_1 F_{\text{comfort}} + w_2 F_{\text{safe}} + w_3 F_{\text{gap}} + w_4 F_{\text{efficiency}} + w_5 F_{\text{headway}}$$

where $w_1, w_2, w_3, w_4, w_5$ are the weights of the five features, respectively.

3.4.6. Safety Modification Model. Reinforcement learning fails to guarantee complete safety. Although a safety feature is included in the reward function, unsafe actions still happen during the learning process as this feature is just a soft constraint. Therefore, a modification acceleration value should be added to the RL acceleration to avoid a collision, as shown in Figure 7. We considered the worst scenario that when the leading vehicle brakes emergently, the SV can stop with its largest deceleration before collision [54]:

$$\min acc_{SM}^2$$

s.t. $Y_t^{LV} - L_t^{LVO} - Y_t^{SV} \geq \tau v_t^{SV} + \frac{\left( v_t^{SV} \right)^2}{2acc_{max}} - \frac{\left( v_t^{LVO} \right)^2}{2acc_{max}}$  

$Y_t^{LV} - L_t^{LVO} - Y_t^{SV} \geq \tau v_t^{SV} + \frac{\left( v_t^{SV} \right)^2}{2acc_{max}} - \frac{\left( v_t^{LVO} \right)^2}{2acc_{max}}$
4. Experiments and Results Analysis

4.1. Training Experiment. About 75% (878) of the 1169 extracted lane-changing scenarios were selected as training data. Each scenario includes empirical data of four vehicles: SV, LVO, LV, and FV. The LVO and the LV will follow the empirical data in the training process. The data of FV are just used to judge whether to start to change lane and is no longer considered after that. That is because the FV’s behavior will change with the SV’s action after the lane changing has begun. How its action will change is not the focus of this study. The SV is initialized with the empirical data at time step 0, including its position, velocity, and direction angle. At each time step, the state space will be collected, and then the DDPG model will give the acceleration and yaw acceleration to the SV. After that, the reward will be calculated. When one scenario finishes, the environment will be reinitialized with the data of the next scenario. Parameters are listed in Table 1.

For the actor network, a neural network that comprises two hidden layers with 30 neurons in each layer is used. The critic network uses only one hidden layer with 30 neurons. Different numbers of hidden layers and neurons have been tested, but they did not perform better results. The rectified linear unit (ReLU) function is set as the activation function for the hidden layers to accelerate the convergence. The tanh function is used as the activation function for the output layer of the actor network.

Figure 8(a) shows the change of the rolling mean episode reward during the training process. The rolling mean episode reward is the average of the mean episode reward based on a rolling window size of 100. One episode is one lane-changing scenario. It can be seen that after training for about 400 episodes, the reward curve starts to reach a higher score (2.0). After that, although the reward has fluctuations, it always maintains a range of 1.8 to 2.0.

Figure 8(b) shows the scores for each of the five features that compose the reward function regardless of the weights. It should be noticed that the fluctuation of the reward curve after 400 episodes is mainly caused by the scores of the gap feature and the efficiency feature, which are related to the initial position of all vehicles in the scenario. Typically, the gap between the SV and the LV or the FV may not be large enough to begin a safe lane changing. As a result, the SV needs relatively more time to reach the lane-changing start condition, which leads the two scores to decline. In contrast, if the target gap is large enough at the beginning, the lane changing will start immediately, and then the two scores will rise accordingly. On the whole, the model can converge after 400 episodes of training.

4.2. Testing Experiment

4.2.1. Testing Results. The rest 25% (291) of the extracted lane-changing scenarios were used as testing data to verify the effectiveness of the trained DDPG lane-changing model. No collision was observed during the testing process. Figure 9 shows the distribution of the lane-changing duration (the time from the start to the end of lane changing) for the DDPG model and real data at different velocities. Figure 10 shows the distribution of the longitudinal distance of the SV controlled by the DDPG model and its actual value in the
same scenario duration (10 s). Table 2 lists the detailed testing results.

Lane-changing safety is evaluated by the time-integrated time-to-collision (TIT), a modified surrogate safety measure derived from TTC [55, 56]. The TIT not only considers the severity of the collision risk but also reflects the duration of the risk. It is defined as follows:

\[
\text{TIT} = \sum_{i=1}^{N} \int_{0}^{T} \max(\text{TTC}^* - \text{TTC}_i(t), 0) \, dt,
\]

\[\forall 0 \leq \text{TTC}_i(t) \leq \text{TTC}^*.\] (14)

Here, \(N\) is the total number of vehicles, \(\text{TTC}^*\) is the threshold value of TTC, and \(\text{TTC}_i(t)\) is the TTC of vehicle \(i\) at time \(t\).

As can be seen, the TIT of the DDPG model (10.87) is much smaller than that of the human drivers (73.09). Correspondingly, the average score of the safety feature has
increased. It shows that the proposed DDPG model can reduce collision risk during the lane-changing process.

Under the control of the DDPG model, the lane-changing duration can be reduced. The average lane-changing duration is 3.2 seconds. As shown in Figure 9, the improvement effect varies with the SV's velocity during the lane-changing process. With the vehicle velocity increases, the reduction will decrease from 53% to 35%. The reason may be as follows: our model can eliminate the influence of human factors during the lane-changing process. At lower velocity, human factors have a greater impact on lane-changing behavior, as human drivers are more cautious when in congestion. As the velocity increases and driving conditions improve, the drivers may be more decisive when changing lanes, reducing the impact of human factors on lane changes. Correspondingly, the score of the efficiency feature has significantly improved. Its average has increased from 0.31 to 0.57. This means that the AVs controlled by the DDPG model can complete the lane-changing motion faster and more efficiently.

Meanwhile, the average time headways of the DDPG model and human driver are 1.89 s and 2.11 s, respectively. The AVs controlled by the DDPG model can maintain a smaller headway during the lane-changing process while ensuring safety. The longitudinal distance of the lane-changing vehicle increases, with an average value rising from 98.9 meters to 103.2 meters. That is to say, the SV can still keep a relatively higher velocity. This improvement can reduce the impact of lane-changing behavior on the following vehicles, which in turn is beneficial to the overall traffic flow.

The average score of the comfort feature has declined. However, the importance of comfort is relatively low compared to the features such as safety and efficiency, and from a numerical point of view, it is still very close to the upper limit of this feature (0). Therefore, its value is still within an acceptable range.

4.2.2. Demonstrations with Samples. We chose three lane-changing scenarios as samples to give illustrations of the DDPG model. Figures 11–13 respectively display them. There are two grey vertical lines in each picture, in which the former represents the start time of lane changing while the latter represents the end time.

In the first sample, the gap between the SV and the LV is insufficient at the beginning, and the lane changing cannot be performed until there is enough space between the two vehicles, as shown in Figures 11(a) and 11(b). Figure 11(c) shows the velocity adjustment process of the SV. It can be seen that the changing trend of its velocity is always close to that of the preceding car. Figure 11(d) shows that the SV successfully completed the lane changing.

For the second sample, Figures 12(a) and 12(b) show that the gap between the SV and the LV is insufficient, and the lane changing cannot be started until there is a sufficient gap between the two vehicles. Figure 12(c) shows the velocity change diagram of the SV. Its velocity change trend is also similar to the velocity change of the preceding vehicle. When the SV first enters the target lane, the relative velocity to the LV is also greater due to the larger longitudinal distance from the LV. Figure 12(d) shows that the SV successfully completed the lateral movement from the original lane to the current lane.

In the third sample, the SV has a sufficient safety gap with the LV and the FV, so the lane changing can begin immediately, as shown in Figures 13(a) and 13(b). Figure 13(c) shows the velocity change diagram of the SV. The overall trend is similar to the velocity trend of the preceding vehicle. There is an obvious acceleration process from 2 s to 4 s. This is because after the SV enters the target lane, the longitudinal distance between the SV and the LV is relatively large. As the distance between the two vehicles becomes smaller, the velocity change of the SV decreases, and its relative velocity to the LV decreases. Figure 13(d) shows that the SV successfully finished its lane changing. It can be concluded that the proposed DDPG lane-changing model enables AVs to succeed in completing the lane changing in different scenarios.

4.2.3. Reaction Time Analysis. The reaction time \( \tau \) in the safety modification model has a significant impact on the SV's behavior. Figure 14 shows the position and velocity of the SV when the reaction time is 0.5 s, 0.7 s, and 1.0 s, respectively in a lane-changing scenario. It can be seen that as the reaction time increases from 0.5 s to 1.0 s, the velocity of the SV gradually decreases, and the relative distance to the preceding vehicle correspondingly increases. This is because the increase in response time makes the safety gap to the proceeding vehicle increase. The AV usually has a shorter response time than human drivers. Under the premise of avoiding collisions, it is advisable to apply a shorter response time to achieve larger vehicle velocity, smaller vehicle gap, and greater road capacity.

4.2.4. Comparison with DLTP Model. We compared our DDPG model with the DLTP model, as proposed by Yang et al. [15], and the results are shown in Figure 15. Both models can output lane-changing actions. The lane-changing process in the DLTP model is much gentler, but the SV and LVO collide. During this process, the TIT of DLTP model is 3.3, while that of the DDPG model keeps 0. This is because the DLTP model did not consider the state of the preceding vehicle in the original lane in the process of controlling the behavior of the SV, which resulted in a collision. In fact, the LVO in this scenario is decelerating. If the SV cannot change to another lane in time, then it needs to reduce velocity. In contrast, our model can adjust SV's velocity in real time according to the state of its current leading vehicle through the safety modification model. Results indicate the DDPG model can greatly reduce collision risk with leading vehicles.

4.3. Simulation Experiment. In this set of experiments, we will test the performance of the DDPG lane-changing model in the traffic flow through simulation. Simulation of Urban
Figure 11: #1 sampled lane-changing scenario. (a) Initial and final state. (b) Plot of longitudinal position with respect to time. (c) Plot of longitudinal velocity with respect to time. (d) Lane-changing trajectory.

Figure 12: #2 sampled lane-changing scenario. (a) Initial and final state. (b) Plot of longitudinal position with respect to time. (c) Plot of longitudinal velocity with respect to time. (d) Lane-changing trajectory.
Figure 13: #3 sampled lane-changing scenario. (a) Initial and final state. (b) Plot of longitudinal position with respect to time. (c) Plot of longitudinal velocity with respect to time. (d) Lane-changing trajectory.

Figure 14: Results for different reaction times. (a) Plot of longitudinal position with respect to time. (b) Plot of longitudinal velocity with respect to time.
Mobility (SUMO 1.8.0), an open-source, microscopic, multimodal traffic simulator, was used [57]. Its traffic control interface (TraCI) enables users to get access to the running road traffic simulation. We are allowed to retrieve values of simulated objects and manipulate their behavior.

The simulation network is a two-lane highway basic section without any on-ramps or off-ramps. The lane width is set as 3.6m, and the section has a total length of 5km, as shown in Figure 16. About 30 simulation experiments have been conducted under the conditions of different speed limits, traffic flows, and the penetration rates of AVs. The penetration rates of AVs are 0%, 20%, 40%, 60%, and 80%. To make the traffic flow diverse, two types of HDVs are set in the simulation: fast HDV and slow HDV. In all simulation experiments, the proportion of the slow HDV keeps 10%. The car-following model for all vehicles is the IDM. The lane changing of HDVs is controlled by SL2015, and that of AVs is controlled by our DDPG model. All vehicles will come into the section with the random speed on the random lane. Other parameters are listed in Table 3.

Figure 15 and Table 4 display the results of the 30 simulation experiments. The last column presents the average speed of the road section in the simulation. The case where the penetration rate of AVs is zero is considered as a baseline, and the relative improvement of other cases with different penetration rates against this one is also reported. It can be easily found that the AVs controlled by the DDPG model improve the average speed of traffic flow. Meanwhile, as the penetration rate of AV rises, this improvement will become more obvious. In addition, the road speed limits and the traffic volume also influence the positive effect of AVs on traffic efficiency. As the results show, the improvement effect is the most distinct in the case of a lower speed limit (60 km/h). With the increase of the speed limit, the improvement to traffic efficiency gradually becomes smaller. Moreover, when the

![Figure 15: Comparison between the DLTP model and the DDPG model.](image)

![Figure 16: Simulation network.](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>AV</th>
<th>Fast HDV</th>
<th>Slow HDV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penetration rate</td>
<td>90%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Max speed (60 km/h speed limit)</td>
<td>60 km/h</td>
<td>40 km/h</td>
<td></td>
</tr>
<tr>
<td>Max speed (80 km/h speed limit)</td>
<td>80 km/h</td>
<td>60 km/h</td>
<td></td>
</tr>
<tr>
<td>Max speed (110 km/h speed limit)</td>
<td>110 km/h</td>
<td>90 km/h</td>
<td></td>
</tr>
<tr>
<td>Max acceleration</td>
<td>3 m/s²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max deceleration</td>
<td>4.5 m/s²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>2000 vph and 3000 vph</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation time</td>
<td>60 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simulation step</td>
<td>0.1 s</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 17: Overall performance of different penetration rates. (a) Speed limit of 60 km/h. (b) Speed limit of 80 km/h. (c) Speed limit of 110 km/h.

<table>
<thead>
<tr>
<th>Speed limit (km/h)</th>
<th>Traffic flow (vph)</th>
<th>Penetration rate (%)</th>
<th>Average speed (km/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2000</td>
<td></td>
<td>42.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>43.56 (3.33%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>44.64 (5.89%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
<td>45.72 (8.45%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td>46.80 (11.018%)</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>80</td>
<td>39.38</td>
</tr>
<tr>
<td></td>
<td>3000</td>
<td>0</td>
<td>39.60 (0.54%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>39.96 (1.46%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
<td>40.68 (3.29%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td>42.48 (7.86%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
<td></td>
</tr>
</tbody>
</table>
traffic volume is low, the improvement effect is more significant. The simulation results illustrate that the AVs controlled by our DDPG lane-changing model can not only improve AV’s lane-changing efficiency but also have a positive impact on the overall traffic efficiency. Lane changing will affect the motion of surrounding vehicles. Completing lane changing as soon as possible can reduce the impact on traffic efficiency and obtain better driving conditions faster. Vehicles want to get desired speed. When both the traffic density and the overall speed are low, the gap between vehicles is larger and there are more space and opportunities for changing lanes. There is also much more acceleration space after finishing the lane changing. With the increase in traffic density and overall speed, the space available for lane changing is reduced, and the acceleration space on the target lane is also reduced. In the meantime, higher vehicle speed often means higher safety risk, especially when the overall speed is relatively high. Therefore, the speed improvement is not distinct in situations with higher speed limit.

5. Conclusion

This study proposes a lane-changing model for AVs based on the DDPG algorithm. The position, direction, and velocity information of the lane-changing vehicle and its surrounding vehicles are denoted as the state space. The action space consists of acceleration and yaw acceleration. Both the state space and action space are continuous. To measure and shape the lane-changing behaviors, a reward function consisting of five driving features such as safety, efficiency, gap, headway, and comfort is developed. We also use a safety modification model considering an emergency brake situation to check and correct unsafe acceleration got from the DDPG model, to avoid collisions at every time step. After that, the driving trajectory data of 1169 lane-changing scenarios extracted from the NGSIM dataset are used to train and test the model. The proposed model can quickly converge in the training experiment. The testing experiment successfully demonstrates safe and efficient lane changing in different lane-changing scenarios. Compared with human drivers, our model can achieve both shorter time headway and lane-changing duration, which are decreased by 10.4% and 35%–55%, respectively. It also outperforms the human drivers and the conventional DLTP model since the collision risk is greatly reduced. The simulation results also indicate that the AVs controlled by the proposed model can have a positive impact on the average speed of the overall traffic flow under different speed limits and traffic volume, and the maximum improvement can reach nearly 10%.

The future work of the study includes the following aspects: first, different weights of features are tried or more objectives are added to the reward function, such as energy-saving feature. It is also worth trying to set random parameters in the reward function instead of constant and study the performance. Second, the state space and reward function of the model are extended to apply in more complex situations, such as weaving areas and ramps on the highway, which have diverging and weaving scenarios. Third, as communication delay can affect lane-changing performance [58], future research will take this factor into account to improve our lane-changing model.

Table 4: Continued.

<table>
<thead>
<tr>
<th>Speed limit (km/h)</th>
<th>Traffic flow (vph)</th>
<th>Penetration rate (%)</th>
<th>Average speed (km/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>63.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>63.72 (0.34%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>64.80 (2.04%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>65.52 (3.17%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>66.24 (4.30%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>59.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60.48 (1.38%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>60.841 (1.99%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>61.92 (3.80%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>62.28 (4.40%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>91.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>91.80 (0.71%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>92.88 (1.90%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>93.96 (3.08%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>94.32 (3.48%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>87.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>87.84 (0.91%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>88.20 (1.32%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>88.30 (1.45%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>88.92 (2.15%)</td>
<td></td>
</tr>
</tbody>
</table>

Conflicts of Interest

The authors declare that they have no conflicts of interest.
Acknowledgments
This research was supported by the Project of Shanghai Science and Technology Innovation Action Plan (grant no. 20511101800).

References


