Decision-Making and Control for Freeway On-Ramp Merging Using Deep Reinforcement Learning

Yuan Lin\(^1\), John McPhee\(^2\), and Nasser L. Azad\(^3\)

Abstract—This work studies high-speed on-ramp merging decision-making and control for an automated vehicle using deep reinforcement learning (DRL). We consider no vehicle-to-everything (V2X) wireless communication and the merging vehicle relies on its own sensors to obtain the states of other vehicles and the road information to merge from on-ramp to the main road. We consider the states of five vehicles as the environment state of the reinforcement learning training framework: the merging vehicle, and two preceding and two following vehicles within the sensing range of the merging vehicle when it is or is projected on the main road. The control action of the reinforcement learning training framework is the acceleration command for the merging vehicle. We use Deep Deterministic Policy Gradient (DDPG) as the DRL algorithm for continuous control, assuming there exists an optimal deterministic policy to match the state to the action. The DRL rewards encourage merging midway between two main-road vehicles with the same speed as the first preceding one, and penalize hard braking, stops, and collisions. When testing the trained policy, we observed 1 collision out of 16975 testing episodes (0.006\% collision rate). By analyzing the merging vehicle’s behaviors, we found that it learned human-like behaviors such as slowing down to merge behind or speeding up to merge ahead a main-road vehicle. For the only collision case, we found that the merging vehicle kept shifting between slowing down and speeding up, suggesting that it might be trapped at a bifurcation state.

I. INTRODUCTION

The development of self-driving vehicles is exciting and important work to improve transportation safety and mobility. There are commercially available Advanced Driver Assistance Systems such as Adaptive Cruise Control, Lane Keeping Assistance, Blind Spot Warning, and Driver Drowsiness Detection [1]. There are also startup and automobile companies working on highly automated vehicles that can perform complex automated driving tasks such as merging, intersection traversing, and roundabout maneuvers. These intelligent vehicles and their functions assist human drivers but do not guarantee full safety. In the future, there will be V2X wireless communications, which could further enhance safety [2]. But fully autonomous vehicles and 100 \% penetration of V2X communications for every road network in the world are still in the future.

1Dr. Yuan Lin is a Postdoctoral Fellow in the Systems Design Engineering Department at University of Waterloo, Ontario, Canada N2L 3G1. y428lin@uwaterloo.ca

2Dr. John McPhee is a Professor and Canada Research Chair in the Systems Design Engineering Department at University of Waterloo, Ontario, Canada N2L 3G1. mcphee@uwaterloo.ca

3Dr. Nasser L. Azad is an Associate Professor in the Systems Design Engineering Department at University of Waterloo, Ontario, Canada N2L 3G1. nlashgarianazad@uwaterloo.ca

Merging is a challenging task for both human drivers and automated vehicles. According to the US Department of Transportation, nearly 300,000 merging accidents happen every year with 50,000 being fatal [3]. The prototype vehicles of the leading self-driving car company, Waymo, have reportedly been seen unable to merge autonomously [4]. There are many variations regarding merging. (1) There are low-speed merging in urban driving environments and also high-speed freeway on-ramp merging. Intuitively, high-speed merging decision-making and control seem to be riskier since there is less time to react given the limits of vehicle dynamics. (2) For freeway on-ramps, there are parallel-type ramps wherein a part of the ramp is connected and parallel to the main road and the merging vehicle could switch to the main road at any point of the parallel portion of the ramp. There are also taper-type ramps wherein the ramp is only connected to the main road at the ramp’s end. Parallel-type ramps are preferred by the highway administration. However, there is a significant presence of taper-type ramps in the world [5]. (3) There is also merging into main roads with light or moderate traffic densities and with dense traffic. Drivers’ or vehicles’ interactions may be more important for dense-traffic merging because they interact more often due to close distances. (4) Also, there are centralized and decentralized merging. Centralized merging relies on roadside centralized controllers to coordinate the vehicles to some extent with V2X communications, while in decentralized merging, the merging vehicle is completely on its own. (5) Last but not least, while designing automated merging systems, there are levels of conservativeness or aggressiveness. For conservative merging systems, the merging vehicle may stop from merging, resulting in the freezing robot problem [6]. In this work, we consider high-speed decentralized non-stop merging into freeways with moderate traffic via taper-type ramps.

In the current literature, there are different rule-based and learning-based methods proposed to tackle the automated merging problem. The traditional and classical rule-based approaches include heuristics and optimal-control-based methods which require modeling the merging problem in a multi-agent system framework [7]. In [8], the authors use an optimal control approach to solve the merging problem in a centralized manner wherein a roadside unit communicates with related vehicles through wireless communications. In [9], the authors use Model Predictive Control to allow an automated vehicle to merge through parallel-type on-ramps. In [10], the authors use the virtual platoon method through centralized control to allow vehicles to merge with a predetermined sequence. In [11], the authors investigate...
merging decision-making using game theory.

In recent years, there are numerous studies that investigate automated merging using DRL, which is a learning-based method. This is because DRL has demonstrated super-human intelligence in playing complex board games [12]. Reinforcement learning finds a state-action mapping policy through trial and error so as to maximize the long-term cumulative reward. DRL utilizes deep (multi-layer) neural nets as the function approximators. In [13], the authors formulate the high-speed (65mph) merging problem within the reinforcement learning framework and use a quadratic function with neural-net weights to approximate the cumulative reward; the merging vehicle is desired to maintain a safe distance between the gap front and gap back vehicles; however, the authors did not show merging decision-making such as gap selection nor did they show merging details. In [14], the authors use the passive actor-critic algorithm for gap-selection and control of the vehicle to merge into dense traffic; the rewards encourage merging midway between vehicles on the main road while keeping the same speed of the preceding vehicle. In [15], the authors take into account the cooperation levels of surrounding vehicles and use Deep Q-Networks for discrete action control of the merging vehicle; the authors consider urban driving environments with dense traffic and the vehicle speeds are around 5m/s. In [16], the authors also consider the cooperation levels of main-road vehicles and develop a multi-agent training algorithm for discrete action control; the vehicle speed of 20m/s is considered and the authors claim zero collision rate in their study.

Our work here also studies merging using DRL. However, there are multiple factors that make our work unique to the published works. Firstly, we use DDPG to train the merging policy. DDPG assumes a deterministic policy and outputs continuous actions for decision-making and control [17]. There is no published paper that utilizes DDPG for merging yet. Secondly, we consider high-speed (65mph) merging decision-making; the speed considered here is 65mph. As stated earlier, high-speed merging could be riskier and DRL may not achieve zero collision. Thirdly, we consider moderate traffic environment wherein we neglect vehicle-to-vehicle interactions and negotiations, which happen more often in dense traffic. This work focuses on applying the DRL framework for high-speed merging and investigating the maximum merging success rate that can be achieved using DDPG. We don’t focus on guaranteeing the safety for merging. Mathematically, the merging problem that we consider is a multi-agent problem with fast-moving agents on the main road and the merging agent seeks to join the main road without causing catastrophic consequences such as collisions.

The rest of the paper is organized as follows: in Section II, the preliminaries of reinforcement learning and DDPG are introduced; in Section III, the merging problem is formulated and fits into the reinforcement learning framework; in Section IV, simulation experiments are presented and results are evaluated; in Section V, we draw conclusions and present possible future work.

II. REINFORCEMENT LEARNING

A. Reinforcement Learning

Reinforcement learning is a learning-based method for decision-making and control. In reinforcement learning, an agent takes an action based on the environment state at the current time step, and the environment subsequently moves to another state at the next time step. The agent also receives a reward based on the action taken. The action and reward are based on probabilities. Reinforcement learning algorithms seek to minimize the expected discounted cumulative reward for each episode. Specifically, the discounted cumulative reward for a state-action pair is called the Q-value, i.e. $Q(s_t, a_t) = E[\sum_{t'=t}^{T} \gamma^{t'-t} r(s_{t'}, a_{t'})]$ where $r(s_t, a_t)$ is the reward for the state $s$ and action $a$ at time step $t$, and $\gamma \in [0, 1]$ is the discount factor. The reinforcement learning problem is solved via Bellman’s principle of optimality which means that, if the optimal Q-value for the next step is known, then action for the current time step must be optimal. That is, $Q^*(s_t, a_t) = r(s_t, a_t) + \max_a Q^*(s_{t+1}, a_{t+1})$ for an optimal policy, with $*$ denoting the optimality.

B. Deep Deterministic Policy Gradient

There are different DRL algorithms available, while we use DDPG for continuous control. There are two networks in DDPG: actor and critic networks. The critic network represents the Q-value which is the discounted cumulative reward. The critic network is iteratively updated based on Bellman’s principle of optimality by minimizing the root-mean-squared loss $L_t = r(s_t, a_t) + (1 - I) * Q(s_{t+1}, \mu(s_{t+1} | \theta_\pi)) - Q(s_t, a_t | \theta_\phi)$ using gradient descent where $\theta_\phi$ denotes the critic neural net weights, and $\theta_\pi$ denotes the actor neural net weights. The $I \in \{0, 1\}$ is the indicator for episode termination with $I = 1$ means termination and $I = 0$ means that the episode is not yet terminated. The actor network is the policy network that maps the environment state to action. The actor network is learned by performing a gradient ascent on the Q-value $Q(s_t, \mu(s_t | \theta_\pi))$ with respect to actor network parameters $\theta_\pi$.

Several techniques are necessary to facilitate training and improve training stability. Those include target networks, mini-batch gradient descent, experience replay, and batch normalization. During training, Gaussian noise is added to the action for exploration purpose. Table I shows the DDPG parameter values that are used for the merging problem described next. These DDPG parameters are tuned through trial and error. Both the actor and critic networks have 2 hidden layers with 64 neurons for each layer.

III. MERGING PROBLEM FORMULATION

This section details the merging problem formulation and how it is expressed in the reinforcement learning framework.
We define a control zone for the merging vehicle that is 100 m to the merging point on the on-ramp and 50m from the merging point on the main road, see Fig. 1. The DRL policy is for controlling the merging vehicles only in this control zone. Note that the 100m to the merging point on the on-ramp is smaller than the entire length of an on-ramp. We consider that effective decision-making could be made within 100m to the merging point since the merging vehicle could come to a complete stop within 100m given the speed limit in case that merging is not successful. Additionally, usually the first part of an on-ramp is designed for the merging vehicle to accelerate from low speeds and merging decision may not be made during this initial acceleration phase. The 50m after the merging point is used to evaluate the merging success because collision could occur if the merging vehicle does not merge at appropriate speeds and gaps between vehicles.

We assume that there is no other on-ramp merging vehicle in front of the controlled merging vehicle within the 100m to the merging point. The main road is assumed to be single-lane such that there is no lane change behavior by the main-road vehicles. Each simulation time step is 0.1 seconds.

For each merging episode, the merging vehicle is initially positioned at 100m before the merging point on the on-ramp. The initial velocity of the merging vehicle is randomly distributed in $[20.12, 24.59]$ m/s (45-55mph). This is based on the standard of US Department of Transportation that suggests a merging vehicle’s velocity reach at least 50mph before merging [5]. The initial acceleration of the merging vehicle is zero. When stops, collisions, or successes happen, the corresponding episode is considered finished; the merging vehicle is then deleted and regenerated with the initial conditions mentioned above.

### A. Merging Environment

The merging environment is created in Simulation of Urban Mobility (SUMO) [18]. We consider a vehicle seeking to merge into a freeway (main road) via an on-ramp. The merging vehicle is an automated vehicle equipped with a suite of perception sensors such as lidar and radar. We consider the horizontal sensing range of the merging vehicle as a circle with a radius of 100 meters, which is a realistic value regarding sensor capabilities [19]. The automated vehicle is assumed to obtain reliable information of the states of all the vehicles (including itself) within its sensing range. The merging vehicle does not possess wireless communication capability. There is no centralized control among the merging vehicle, main-road vehicles, and infrastructure via wireless communications. The merging vehicle is considered as a point mass and utilizes a kinematic model for state update. That is, there is no vehicle dynamics or delay considered for the merging vehicle.

The main-road vehicles are intelligent such that they can perform car following and collision avoidance based on Krauss’ model [20]. Specially, they could slow down when the merging vehicle enters the small junction area that connects the on-ramp and the main road, see Fig. 1. Main-road vehicles have speeds that are near the speed limit $v_{\text{limit}} = 29.06$ m/s (65mph). Particularly, each main road vehicle has a desired average speed as $v_{\text{limit}} * \gamma$ wherein $\gamma$ is constant for a vehicle and has a value randomly distributed in $[0.8, 1.2]$ with respect to different vehicles; at a time step, the actual speed of the vehicle is $v_{\text{limit}} * \gamma * \beta$ wherein $\beta$ is randomly distributed in $[0.9, 1.1]$ with respect to different time steps. We consider moderate traffic density such that one main road vehicle is generated every 3 seconds at the far bottom of the main road. Thus, the main-road vehicles have an average inter-vehicular distance gap of $29.06 * 3 = 87.18$m. However, due to the speed variations among vehicles and car following parameter variations, the main-road vehicles maintain different inter-vehicular distance gaps case by case. All the main-road vehicles have normal acceleration values in $[-4.5, 2.6]$ m/s$^2$. When abnormal situations happen, such as the merging vehicle merges too closely to a following vehicle, the following vehicle can decelerate further to the minimum -9m/s$^2$ which is the emergency braking deceleration. These definitions for the main-road vehicles remain the same during DRL training and testing. Also, these definitions are the default in SUMO for creating intelligent traffic.

### B. Reinforcement Learning Framework for Merging

Based on the merging environment, we design the reinforcement learning framework as follows so as to train a corresponding merging policy.

The environment state of the reinforcement learning framework includes the states of five vehicles: the merging vehicle ($m$), two preceding ($p_1$, $p_2$) and two following ($f_1$, $f_2$) vehicles when the merging vehicle is or is projected on the main road, see Fig. 1. The state for each vehicle includes its distance to the merging point $d$, its velocity $v$, and its acceleration $a$. The distance to the merging point is positive and negative before and after the merging point, respectively. The distance to the merging point is measured from the front bumper of the vehicle. The environment state is denoted as

$$s = [d_{p2}, v_{p2}, a_{p2}, d_{p1}, v_{p1}, a_{p1}, d_m, v_m, a_m, d_{f1}, v_{f1}, a_{f1}, d_{f2}, v_{f2}, a_{f2}]$$

When there are fewer than two preceding or two following vehicles detected, we assume virtual vehicles at the intersections of the sensing space and the main road to deliberately construct the five-vehicle state vector. This is a conservative approach for safety as there could be vehicles just outside the sensing space in real-world driving. These virtual vehicles have velocity values as the main road speed.
2) When the merging vehicle successfully passes the control zone (50m above the merging point), a reward of 1 is given.
3) When the first following vehicle performs hard braking with acceleration below $-4.5m/s^2$, a penalizing reward is given as:
   $$r_b = -0.05 \times \frac{-4.5 - a_{f1}}{-4.5 + 9} = -0.05 \times \frac{-4.5 - a_{f1}}{4.5}$$
   where 9m/s$^2$ is the maximum deceleration of main-road vehicles for emergency braking. 4) To reduce the jerk of the merging vehicle, we define a penalizing reward as:
   $$r_j = -0.01 \times \frac{|a_{m_{j+1}} - a_{m'_j}|}{2.6 + 4.5} = -0.01 \times \frac{|a_{m_{j+1}} - a_{m'_j}|}{7.1}$$
5) When the merging vehicle stops, a penalizing reward of -0.5 is given and the episode is terminated. 6) When the merging vehicle collides with any vehicle, a penalizing reward of -1 is given and the episode is terminated. A collision is registered when the inter-vehicular distance is less than 2.5m. Note that a larger penalty is given for collisions than for stops because collisions are considered more catastrophic.

The weights in the rewards are tuned through trial and error for reducing merging collision risk. We don’t consider rewards for car following since car following between the merging and first preceding vehicles demands a certain distance gap which could result in smaller distance and collision risk between the merging and the first following vehicles. Also, we don’t penalize the time it takes the agent to finish an episode since time minimization is incentivized by the reinforcement learning discount factor $\gamma = 0.99$ which is less than 1 [15].

IV. RESULTS

In this section, we show and evaluate the training and testing simulation results, and representative episodes that show the merging vehicle’s learned behaviors during merging.

A. Training

We consider training the merging vehicle for 1 million simulation time steps wherein we observe reasonable convergence of the cumulative episode reward. As shown in Fig. 2 at the initial phase of training, the agent had episode rewards less than -0.5 very often, indicating many stops or collisions. In the end, as the agent learned, such episode rewards less than -0.5 happened much less often. It took 1 hour for the training on a computer with a 16-core (32-thread) AMD processor and a Nvidia GeForce RTX GPU.

B. Testing

We tested the trained policy for another 1 million simulation time steps and recorded the total number of episodes and the numbers of stops, collisions, and successes. We also recorded the emergency braking times wherein the merging vehicle decelerates to -9m/s$^2$. Note that, even though emergency braking happens, if the merging vehicle passes the control zone, it is still considered a successful episode. The average speed of the merging vehicle is calculated as limit and zero acceleration. Note that a vehicle is determined to be preceding or following relative to the merging vehicle position’s projection. A following vehicle may become a preceding vehicle during merging if the merging vehicle slows down and its projections move behind the main-road vehicle’s position.

The action of the reinforcement learning framework is the acceleration control input to the merging vehicle $a_m$. As we don’t consider vehicle dynamics and delays, the acceleration control input is executed instantaneously by the merging vehicle. The velocity and position updates of the merging vehicle follow an Euler forward discretization. The acceleration control input for the merging vehicle is strictly within $[-4.5, 2.6]m/s^2$, which is the same normal acceleration range of a main road vehicle.

The reward at each time step is designed as follows: 1) After the merging point and until the end of the control zone (50m after the merging point), the merging vehicle should be midway between the first preceding and first following vehicles, with the speed of the first preceding vehicle. Thus, the corresponding penalizing reward is defined as
   $$r_m = -0.05 \times (|w| + |v_{p1} - v_m|/20)$$
   where $w$ is the midway ratio defined as the ratio between the difference and sum of the gaps among the merging, its first preceding, and its first following vehicles, and 20m/s is a nominal maximum speed difference that we defined.

   $$w = \frac{|d_{p1} - d_m - l_{p1}| - |d_m - d_{f1} - l_m|}{|d_{p1} - d_{f1} - l_{p1} - l_m|}$$

Fig. 1. Schematic for merging.
the total cumulative travelled distance divided by the total simulation time, which is $16975 \times 150m/100,000s = 25.46m/s$.

From Table II, we see that trained policy resulted in zero stops during testing. Additionally, the average speed is 25.46m/s (57mph), which is relatively high. This means that the trained agent sought to minimize the travel time, resulting from both the stop penalty and the discount factor. The merging vehicle made only 1 collision. However, it caused the first following vehicle to brake emergently 197 times. This is partially because the main road vehicles were relatively conservative based on Krauss’ model and performed hard braking even when the merging vehicle was relatively far away.

**TABLE II**

| Testing Results |
|-----------------|
| Total # of episodes | 16975 |
| # (%) of stops by merging vehicle | 0 (0) |
| # (%) of emergency braking maneuvers by first following vehicle | 197 (1.161) |
| # (%) of collisions involving merging vehicle | 1 (0.006) |
| # (%) of successful episodes | 19024 (99.994) |
| Average speed of merging vehicle | 24.22m/s |

**C. Representative Episodes**

To understand how and why the agent merged successfully or failed to merge, we plot the vehicle states and the midway ratio, see Fig. 3. We plotted 4 different episodes wherein the vehicle merges without slowdown, with slowdown, causing emergency braking of the first following vehicle, and causing the only collision, respectively.

For Episode 1, the merging vehicle merged without slowdown. Initially, the merging vehicle’s projection on the main road was a little behind the first preceding vehicle; and the merging vehicle’s speed is smaller than the first preceding vehicle. The merging vehicle accelerated moderately during merging. The midway ratio after merging is around -0.25, which means that it’s closer to the first preceding vehicle than to the first following vehicle.

For Episode 2, the merging vehicle merged with slowdown. Initially, the merging vehicle’s projection on the main road ahead of the first following vehicle. The trained policy enables the merging vehicle to slow down to let the first following vehicle to overtake it to become its first preceding vehicle. Then the merging vehicle accelerated to merge behind the now-called first preceding vehicle. The midway ratio after merging is around -0.15.

For Episode 3, the merging vehicle caused emergency braking of the first following vehicle. Initially, the merging vehicle’s projection on the main road was in the middle between its first preceding and first following vehicle. However, the merging vehicle’s velocity is significantly smaller than the first following vehicle. The merging vehicle decided to accelerate with maximum acceleration limit during the entire merging process. It avoided collisions but was too close to the first following vehicle at some point and caused emergency braking by the first following vehicle. The midway ratio after merging is around 0.55, which means that the merging vehicle is significantly closer to the first following vehicle than to the preceding vehicle.

For Episode 4, the merging vehicle collided with its following vehicle. Initially, the merging vehicle’s projection was closer to its first following vehicle and had a lower speed. During the merging process, the merging vehicle did not slow down significantly to let the first following vehicle pass like in Episode 2, nor did it accelerate significantly to avoid the collision like in Episode 3. This was the only collision among the 16975 testing episodes.

**V. CONCLUSION**

This work solves for a decision-making and control solution via DRL for a merging vehicle to merge from an on-ramp to a freeway at high speeds. We provided a reinforcement learning framework for the merging problem and demonstrated that DDPG could successfully train a merging policy with 99.994% merging success rate or 0.006% collision rate. Note that we also trained the agent with more time steps (1.5 million) and found slightly worse results that are not shown here. This is possibly due to training instability of DDPG or overfitting of the neural nets. It means that more training does not necessarily improve the testing results for DDPG.

When testing the trained policy, we discovered that the trained agent learned human-like decision-making strategies, such as to decelerate to merge behind a vehicle or to accelerate to merge in front of it. A collision also happened during testing wherein the DRL agent collided with its first following vehicle. For this single collision case, the DRL agent neither decelerated nor accelerated persistently at the early phase of merging to avoid the collision. Instead, the DRL agent kept shifting between decelerating and accelerating in the beginning; see the acceleration plot of Episode 4 in Fig. 3. This suggests that the environment states of the collision case might be at the bifurcation threshold that caused the shifting behavior.
Fig. 3. Four representative episodes of the merging vehicle during testing. Episode 1: success without slow down; Episode 2: success with slow down; Episode 3: success while causing emergency braking of the first following vehicle; Episode 4: collision. In each plot on the first row, the thinner dash-dot curve above the thicker ones shows the distance to the merging point $d$ for the second preceding vehicle; the thinner dash-dot curve below the thicker ones shows the distance to the merging point $d$ for the second following vehicle. In the first plot of $d$ for Episode 2, the first following vehicle became the first preceding vehicle at around 1.7s as the merging vehicle slowed down, which causes the sharp change of the $d$ values for the corresponding vehicles.
As safety for the learned policy is not guaranteed, trajectory prediction or envelope control may be needed to exclude unsafe actions in future work. Future work should also include considering vehicle dynamics in training since in our previous work, we discovered that the kinematic-model-trained policy could cause significantly degraded performance in more realistic situations with vehicle dynamics [21]. Another research direction is to consider fuel economy in the reward to train an economic merging policy [22].

ACKNOWLEDGMENT

The authors would like to thank Toyota, Ontario Centres of Excellence, and the Natural Sciences and Engineering Research Council of Canada for the support of this work.

REFERENCES

[1] A. Eskandarian, Handbook of Intelligent Vehicles. Springer London, 2012.
[2] J. J. Blum, A. Eskandarian, and L. J. Hoffman, “Challenges of intervehicle ad hoc networks,” IEEE Transactions on Intelligent Transportation Systems, vol. 5, no. 4, pp. 347–351, 2004.
[3] “https://www.dulaneylaurathomas.com/blog/merging-into-danger-risks-of-entering-highway-traffic.cfm.”
[4] “https://www.thetruthaboutcars.com/2018/05/thwarted-ramp-waymo-driverless-car-doesnt-feel-urge-merge/”
[5] A. A. of State Highway and T. Officials, “A policy on geometric design of highways and streets,” 2011.
[6] P. Trautman and A. Krause, “Unfreezing the robot: Navigation in dense, interacting crowds,” in 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2010, pp. 797–803.
[7] J. Rios-Torres and A. A. Malikopoulos, “A survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps,” IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 5, pp. 1066–1077, 2016.
[8] ———, “Automated and cooperative vehicle merging at highway on-ramps,” IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 4, pp. 780–789, 2017.
[9] W. Cao, M. Mukai, T. Kawabe, H. Nishira, and N. Fujiki, “Cooperative vehicle path generation during merging using model predictive control with real-time optimization,” Control Engineering Practice, vol. 34, pp. 98–105, 2015.
[10] Z. Wang, B. Kim, H. Kobayashi, G. Wu, and M. J. Burth, “Agent-based modeling and simulation of connected and automated vehicles using game engine: A cooperative on-ramp merging study,” arXiv preprint arXiv:1810.09952, 2018.
[11] K. Kang and H. A. Rakha, “Game theoretical approach to model decision making for merging maneuvers at freeway on-ramps,” Transportation Research Record, vol. 2623, no. 1, pp. 19–28, 2017.
[12] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., “Mastering the game of go with deep neural networks and tree search,” Nature, vol. 529, no. 7587, p. 484, 2016.
[13] P. Wang and C.-Y. Chan, “Autonomous ramp merge maneuver based on reinforcement learning with continuous action space,” arXiv preprint arXiv:1803.09203, 2018.
[14] T. Hu, A. Nakhaei, M. Tomizuka, and K. Fujimura, “Interaction-aware decision making with adaptive strategies under merging scenarios,” arXiv preprint arXiv:1904.06025, 2019.
[15] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” arXiv preprint arXiv:1509.02971, 2015.

[18] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. WieBner, “Microscopic traffic simulation using sumo,” in 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018, pp. 2575–2582.
[19] J. Hecht, “Lidar for self-driving cars,” Optics and Photonics News, vol. 29, no. 1, pp. 26–33, 2018.
[20] S. Krauß, P. Wagner, and C. Gawron, “Metastable states in a microscopic model of traffic flow,” Physical Review E, vol. 55, no. 5, p. 5597, 1997.
[21] Y. Lin, J. McPhee, and N. L. Azad, “Longitudinal dynamic versus kinematic models for car-following control using deep reinforcement learning,” arXiv preprint arXiv:1905.08314, 2019.
[22] M. Vajedi and N. L. Azad, “Ecological adaptive cruise controller for plug-in hybrid electric vehicles using nonlinear model predictive control,” IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 1, pp. 113–122, 2015.