Deep-reinforcement-learning-based Control of Saturated Traffic at Lane-drop Freeway Bottlenecks via Longitudinal and Lateral Trajectory Planning for Limited Connected Vehicles

X. He[a], Y. Zhang[a], K. Feng[a], C. Yu[a,*], W. Ma[a]

[a] Tongji University, Shanghai, China
2151157@tongji.edu.cn, 2153338@tongji.edu.cn, 2150626@tongji.edu.cn,
hughyu90@tongji.edu.cn, mawanjing@tongji.edu.cn
* Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 25, 2024

Keywords: freeway bottleneck; flow arrival pattern control; connected vehicle; trajectory planning; deep reinforcement learning

1 INTRODUCTION

Lane closures on freeways, prompted by accidents or road works, often result in lane-drop bottlenecks and consequent traffic congestion, causing significant delays, increased fuel consumption, and emissions. To alleviate the congestion, variable speed limit (VSL) based traffic control has proven effective in enhancing traffic efficiency and reducing emissions (Liu, *et al.*, 2015). However, the installation and maintenance costs of VSL signs are high.

With the development of connected vehicles (CVs) and connected automated vehicles (CAVs), the concept has been proposed that limited CVs/CAVs are used to harmonize vehicle speeds and regulate the arrival patterns at lane-drop bottlenecks on freeways under saturated traffic. Empirical data have validated its effectiveness (Qi, *et al.*, 2020). However, existing studies primarily focus on formulating analytical models of controlling longitudinal speed profiles of CVs/CAVs assuming no lane changes (Piacentini, *et al.*, 2019). Further, the interactive impacts of multiple CVs/CAVs on traffic flow are overlooked due to the difficulty in analytical modeling.

Inspired by the successful application of deep reinforcement learning (DRL) in vehicle trajectory planning (Ko, *et al.*, 2020; Jiang, *et al.*, 2022), this study proposes a DRL-based approach to regulate the arrival patterns of saturated traffic flow at a lane-drop bottleneck on freeways by trajectory planning for limited vehicles as shown in Figure 1. Considering the higher penetration rates of CVs than CAVs, this study investigates the longitudinal and lateral trajectory planning for limited CVs with discrete speeds to improve safety, efficiency, and fuel economy of traffic flow traveling through the bottleneck. A DRL framework is first designed for single-vehicle control and then extended to multi-vehicle control with the aid of multi-agent DRL.

Figure 1– *Scenario of the lane-drop freeway bottleneck*

2 METHODOLOGY

2.1 Framework

The DRL learning architecture is designed as shown in Figure 2. CVs, as the agents, interact dynamically with the environment to explore optimal driving policies. Initially, agents may exhibit risky and inefficient behaviors. A pretraining process based on an offline RL algorithm is applied with the conservative Q-learning (CQL) algorithm (Kumar, *et al.*, 2020). The learnt policy is then transferred to dynamic environment for online training with the Deep Q-Network (DQN) algorithm (Mnih, *et al.*, 2015). The pretraining allows agents to learn from the empirical policies derived from real-world human driving data to hasten the convergence of online training.

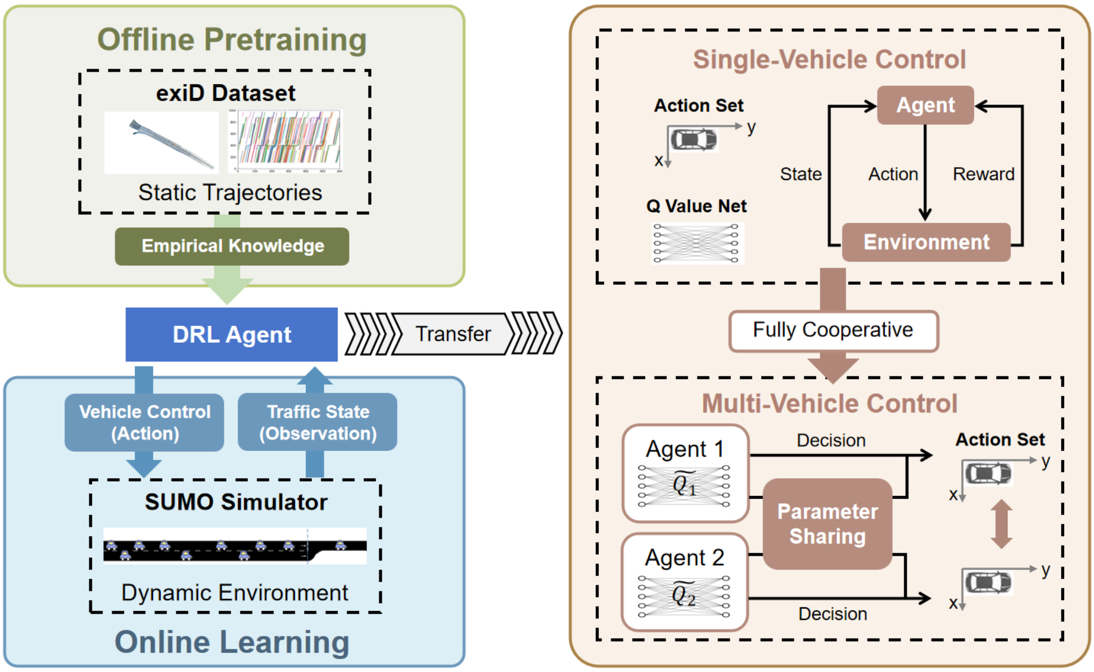

Figure 2 – *DRL Framework*

The fully-cooperative setting for multi-vehicle control (Yadav, *et al.* 2023) is employed in the centralized-training-decentralized-execution (CTDE) way to extend the single-agent DRL framework to the multi-agent one. A portion of the parameters of the topmost layer of Q-networks are shared during training, enabling CVs to learn collaboratively. In decision-making, however, they can make their decisions independently based on their respective networks.

2.2 State and action

Assume vehicle states (i.e., location, speed, and acceleration) can be collected by infrastructure-based detectors (e.g., videos) in real time. The observation space contains the current state of each CV and the vehicles in front and behind it. As two-dimensional discrete actions, longitudinal speed control is comprised of 9 discrete levels, with the step of 0.6 m/s, and the lateral behavior includes left-changing, right-changing, and lane-keeping.

2.3 Reward shaping

To train an altruistic agent, the reward function is designed with three goals: safety, efficiency, and smoothness. Equation (1) indicates the safety risks caused by the agent's unreasonable actions.

$$r_{safe} = -1000 \cdot \delta_0 - 500 \cdot \delta_{ec} - 100 \cdot N_c \tag{1}$$

If the agent makes an invalid action (e.g., change lane to the right in the outermost lane), then $\delta_0 = 1$; otherwise, $\delta_0 = 0$. If the agent collides with others, then $\delta_{ec} = 1$; otherwise, $\delta_{ec} = 0$. Equation (2) is formulated to measure the operational efficiency.

$$E = \frac{3600 \cdot \sum_j v_j}{l} \tag{2}$$

where l is the control range and v_j is the speed of vehicle *j* in the control zone. To cater to the benefits of individual vehicles, vehicles moving at a slow speed (e.g., 5 km/h) should be penalized. The efficiency part is formulated as Equation (3).

$$r_{effi} = \omega_1 \cdot E - \omega_2 \cdot \delta_{v_i<v_0} - \omega_3 \cdot N_{v_j<v_0} \tag{3}$$

where $\delta_{v_i<v_0}$ is a 0-1 variable that indicates whether the agent is traveling at a low speed, and $N_{v_j<v_0}$ is the total number of slow-speed vehicles. Equation (4) indicates the smoothness of traffic flow.

$$r_{smoo} = -\omega_4 \cdot \overline{|a_b|} + \omega_5 \cdot \min(a_i, 0) \tag{4}$$

where $\overline{|a_b|}$ denotes the average absolute value of acceleration of following vehicles, and a_i denotes the acceleration of the agent itself. Since $min(a_i, 0)$ is negative or equals 0, the more intense the deceleration, the greater the agent is penalized. The complete reward function is shown as Equation (5).

$$r = r_{safe} + r_{effi} + r_{smoo} \tag{5}$$

By trial and error, the weights for each component are calibrated as: $\omega_1 = 0.2$, $\omega_2 = 10$, $\omega_3 = 1$, $\omega_4 = 20$, $\omega_5 = 20$.

3 EXPERIMENT RESULTS

3.1 Simulation setup

To validate the effectiveness of this approach, simulation experiments in SUMO are conducted. The proposed reinforcement learning models are tested including the single-vehicle control (SC), the offline-pretrained single-vehicle control (off+SC), and the multi-vehicle control (MC). The typical late merge (LM) control is used as the benchmark. The merging scenario at the two-lane freeway in Figure 1 is used. The demand levels of 1400 veh/h, 1600 veh/h, and 1800 veh/h are tested. The bottleneck capacity is around 1400 veh/h. The exiD dataset (Moers, *et al.*, 2022) is used in the offline pretraining process, which contains real trajectories of human driven vehicles in the scenario of lane-drop freeway bottlenecks.

3.2 Results and discussion

Table 1 – *Simulation results: throughput (veh/h) and fuel consumption* ($mg/(veh \cdot s)$)

Demand (veh/h)	LM		SC		off+SC		off+MC	
	Through-put	Fuel Consump-tion	Through-put	Fuel Consump-tion	Through-put	Fuel Consump-tion	Through-put	Fuel Consump-tion
1400	1328	798	1370 (3%↑)	780 (2%↓)	1471 (11%↑)	789 (1%↓)	**1516 (14%↑)**	**769 (4%↓)**
1600	1381	813	1410 (2%↑)	644 (21%↓)	1464 (6%↑)	**631 (22%↓)**	**1611 (17%↑)**	653 (20%↓)
1800	1460	785.1	1520 (2%↑)	633 (19%↓)	1527 (1%↑)	**636 (19%↓)**	**1636 (12%↑)**	649 (17%↓)

Table 1 shows the throughput and fuel consumption of the simulation experiment. In terms of throughput, the improvements by SC are not significant compared to LM at three demand levels. Off+SC shows excellent performance at low demand, suggesting good pretraining results. Although the improvement become less pronounced as demand increases, which is related to the lack of high-flow data in the exiD dataset, the throughput still outperforms SC without pretraining. MC demonstrates the most significant and stable improvement across various demand levels, indicating the model's strong generalization capability. In terms of average fuel consumption, the performance under all three control methods improves considerably. It indicates that whilst the capacity of the bottleneck is enhanced, the overall traffic flow becomes smoother with overall fuel consumption reduced. It is observed that it is challenging for SC and off+SC to achieve significant improvements in both throughput and fuel consumption simultaneously. And off+MC strikes a balance between high efficiency and low energy consumption with the aid of the collaboration between multiple CVs.

4 CONCLUSION

This study proposes a DRL-based approach to regulate the arrival patterns of saturated traffic flow at a lane-drop bottleneck on freeways by longitudinal and lateral trajectory planning for limited CVs. The objective is to enhance safety, efficiency, and fuel economy. The CQL algorithm is used for offline pretraining to speed up the convergence of online training with DQN. The single-agent DRL framework is extended to the multi-agent one by parameter sharing of Q networks. Simulation results validate the advantages of the proposed models.

References

Jiang, L. M., Xie, Y. C., Evans, N. G., Wen, X., Li, T. A. & Chen, D. J., 2022. Reinforcement Learning Based Cooperative Longitudinal Control for Reducing Traffic Oscillations and Improving Platoon Stability. *Transportation Research Part C-Emerging Technologies*, **141**, p. 103744.

Ko, B., Ryu, S., Park, B. B. & Son, S. H., 2020. Speed Harmonisation and Merge Control Using Connected Automated Vehicles on a Highway Lane Closure: A Reinforcement Learning Approach. *IET Intelligent Transport Systems*, **14**, pp. 947-958.

Kumar, A., Zhou, A., Tucker, G. & Levine, S., 2020. Conservative Q-Learning for Offline Reinforcement Learning. *Advances in Neural Information Processing Systems*, **33**, pp. 1179-1191.

Liu, W., Yin, Y. F. & Yang, H., 2015. Effectiveness of Variable Speed Limits Considering Commuters' Long-Term Response. *Transportation Research Part B-Methodological*, **81**, pp. 498-519.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S. & Hassabis, D. 2015. Human-Level Control Through Deep Reinforcement Learning. *Nature*, **518**, pp. 529-533.

Moers, T., Vater, L., Krajewski, R., Bock, J., Zlocki, A. & Eckstein, L., 2022. The exiD Dataset: A Real-World Trajectory Dataset of Highly Interactive Highway Scenarios in Germany. *2022 IEEE Intelligent Vehicles Symposium (IV)*, pp. 958-964.

Piacentini, G., Ferrara, A., Papamichail, I. & Papageorgiou, M., 2019. Highway Traffic Control with Moving Bottlenecks of Connected and Automated Vehicles for Travel Time Reduction. *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 3140-3145.

Qi, X. Z., Sun, T., Ma, W. J., Wang, L. & Yu, C. H., 2020. Evaluation Of a Variable Speed Limit System Using a Hierarchical Framework Under a Connected Vehicles Environment: A Case Study of a Field Test in Seattle. *IET Intelligent Transport Systems*, **14**, pp. 1125-1131.

Yadav, P., Mishra, A. & Kim, S. 2023. A Comprehensive Survey on Multi-Agent Reinforcement Learning for Connected and Automated Vehicles. *Sensors*, **23**(10), p. 4710.