Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 6669028, 13 pages
https://doi.org/10.1155/2021/6669028

Research Article

A Deep Reinforcement Learning Approach for Ramp Metering Based on Traffic Video Data

Bing Liu,[1] **Yu Tang**,[2] **Yuxiong Ji**,[1] **Yu Shen**,[1] **and Yuchuan Du**[1]

[1]*Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China*
[2]*Tandon School of Engineering, New York University, New York 11201, NY, USA*

Correspondence should be addressed to Yu Shen; yshen@tongji.edu.cn

Received 14 October 2020; Accepted 15 October 2021; Published 31 October 2021

Academic Editor: Victor L. Knoop

Copyright © 2021 Bing Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ramp metering that uses traffic signals to regulate vehicle flows from the on-ramps has been widely implemented to improve vehicle mobility of the freeway. Previous studies generally update signal timings in real-time based on predefined traffic measurements collected by point detectors, such as traffic volumes and occupancies. Comparing with point detectors, traffic cameras—which have been increasingly deployed on road networks—could cover larger areas and provide more detailed traffic information. In this work, we propose a deep reinforcement learning (DRL) method to explore the potential of traffic video data in improving the efficiency of ramp metering. Vehicle locations are extracted from the traffic video frames and are reformed as position matrices. The proposed method takes the preprocessed video data as inputs and learns the optimal control strategies directly from the high-dimensional inputs. A series of simulation experiments based on real-world traffic data are conducted to evaluate the proposed approach. The results demonstrate that, in comparison with a state-of-the-practice method, the proposed DRL method results in (1) lower travel times in the mainline, (2) shorter vehicle queues at the on-ramp, and (3) higher traffic flows downstream of the merging area. The results suggest that the proposed method is able to extract useful information from the video data for better ramp metering controls.

1. Introduction

Ramp metering uses traffic signals to regulate vehicle flows from on-ramps to the mainline of the freeway. It alleviates the negative impacts of "capacity drop" resulting from massive merging behaviors and reduces the total time spent in the traffic system [1, 2]. Several field tests have demonstrated the effectiveness of ramp metering in terms of throughput, vehicle-miles-traveled, vehicle-hours-traveled, and travel time reliability [3–6]. For instance, a large-scale experiment conducted in Minneapolis-Saint Paul found that when the ramp meters were turned off, freeway capacity decreased by 9%, travel time increased by 22%, and crashes increased by 26% [7].

Most of the ramp metering methods are driven by traffic measurements, such as volumes, queue lengths, and occupancies, obtained from point detectors (e.g., inductive loop detectors) [8, 9]. These measurements are predefined and only reflect partial information regarding traffic operations. Recently, traffic cameras have been increasingly deployed on road networks for monitoring traffic operations and detecting illegal driving behaviors. Compared with point detectors, traffic cameras cover larger areas and provide more detailed traffic information, such as vehicle locations, vehicle speeds, and headways between vehicles.

We propose a deep reinforcement learning (DRL) method to explore the potential of traffic video data in improving the efficiency of ramp metering. The proposed method does not rely on traditional traffic measurements from loop detectors. Instead, it uses traffic video frames as inputs and learns the optimal control strategies directly from high-dimensional visual inputs by taking advantage of the special neutral structures in deep learning, which are capable of automatically extracting high-level features from raw data. The effectiveness of the proposed method is demonstrated in comparison with a state-of-the-practice method in a real-world case study.

This paper is organized as follows. A brief review of the related literature is presented first, and then the methodology is described, followed by the demonstration in a real-world case study. The final section summarizes the paper and discusses possible directions for future research.

2. Related Works

The signal timing plans for ramp metering could be fixed or traffic responsive. Wattleworth [10] firstly formulated a linear programming model to optimize a fixed signal timing plan for ramp metering based on historical traffic volume data. Responsive approaches adapt to traffic flow fluctuations by updating signal timings in response to real-time traffic measurements. The responsive approaches can be classified as rule-based, optimization-based, and RL-based.

The rule-based approaches update signal timings in real-time according to specific rules. Masher et al. [11] proposed a feedforward ramp metering algorithm, aiming at keeping the downstream traffic flow below the capacity. Papageorgiou et al. [8] proposed a feedback ramp metering algorithm, which is named ALINEA, based on the occupancy obtained by inductive loop detectors downstream of the merging area. The ALINEA regulates the inflows from the on-ramp in the scheme of closed-loop control so that the detected occupancy would approach a predefined target value. The ALINEA was further extended by considering more traffic measurements, such as queue lengths at the on-ramp and traffic volumes in the mainline [12]. Wang et al. [9] proposed the proportional-integral-based ALINEA (i.e., PI-ALINEA), considering the cases where a bottleneck with a capacity lower than that of the merging area is present further downstream of the merging area. They demonstrated numerically that the PI-ALINEA performs better than the original ALINEA. The ALINEA has been incorporated in several coordinated ramp metering algorithms, such as METALINE and HERO [13, 14].

The optimization-based approaches optimize signal timings based on real-time traffic data. The model predictive control framework, which considers the interactions between ramp metering and future traffic states, is often employed to predict traffic evolution for proactive traffic controls. Nevertheless, the models used to describe traffic dynamics may lead to nonlinear control problems, which brings difficulty in finding the optimal signal timings [15–17]. In addition, the effectiveness of the ramp metering strategy depends on the degree of the fitness of the models to the actual traffic dynamics.

The RL-based approaches search for a policy that determines the signal timings based on the current traffic state. Existing RL-based ramp metering approaches were mainly developed based on value-based RL methods, such as the Q-learning algorithm. The Q-learning algorithm evaluates the action value of each state-action pair and improves the policy with a heuristic search. The actions usually refer to the metering rate or signal phase selection. The states are represented by traffic measurements, such as traffic density and volume. The downstream and upstream traffic flow, ramp queue length, and traffic density are often selected to define the control reward [18–21]. The RL has also been introduced for coordinated ramp metering [22, 23]. Note that existing RL methods for ramp metering were driven by traditional traffic measurements. To the best of our knowledge, no studies have attempted to automatically extract information from traffic videos and learn the optimal control strategies directly from the visual inputs for ramp metering. Some studies have considered traffic video data as inputs in the RL framework for intersection signal controls [24–26]. Nevertheless, the elements in the RL framework, such as the action, reward, and state, for ramp metering and intersection signal controls are different.

Compared with the methods based on traditional measurement, the proposed strategy shows great potential in improving ramp metering performance due to the spatio-temporal information contained in the video data. However, the high-dimension data increase the computational complexity in the training phase. Therefore, a well-designed RL framework and algorithm are necessary to guarantee the effectiveness of the proposed method.

3. Methodology

We propose a DRL method for local ramp metering based on traffic video data. The proposed method adopts a flexible two-phase control scheme, which is illustrated in Figure 1. A policy trained from the DRL determines whether the phase in the next time step is green or red at a fixed time step L based on the current traffic state. The adopted control scheme is conceivably more flexible than the conventional "stop-and-go" scheme with fixed red and green phases flashing alternately. To reduce the computation burden, traffic video frames are used as control inputs. Vehicle locations from raw images are extracted to reconstruct visual representations in the DRL method [24].

Figure 2 illustrates the ramp metering problem in the general scheme of RL. The ramp meter acts as the agent to interact with the environment, namely, the traffic system including the mainline and the on-ramp. The arrows represent the direction of information interaction in the system. The ramp meter takes certain control action a_t based on current traffic state s_t. Then, the traffic system responds to the control action with the state transition from s_t to s_{t+1}. The ramp meter obtains reward r_t that quantifies the effect of the control action a_t given state s_t. The interaction process is assumed to be a Markov decision process [27]. That is, given the current state s_t and action a_t, the following state s_{t+1} and reward r_t are independent of previous states and actions.

The RL method aims at training the agent (Ramp meter) to find an optimal policy π^* that maximizes the discounted cumulative reward G_t after time step t. To achieve that goal, one approach is to learn the optimal action-value function $Q^*(s, a)$, which is defined by the following:

$$Q^*(s, a) = \max_{\pi} \mathbb{E}_{\pi}\left[G_t | s_t = s, a_t = a\right], \tag{1}$$

where π is a policy mapping traffic states to control actions. Once $Q^*(s, a)$ is found, the action to be taken can be obtained by the following:

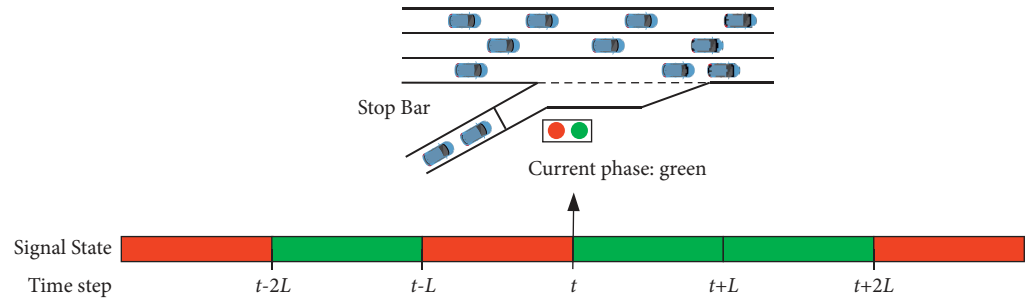

Figure 1: Two-phase control scheme for ramp metering.

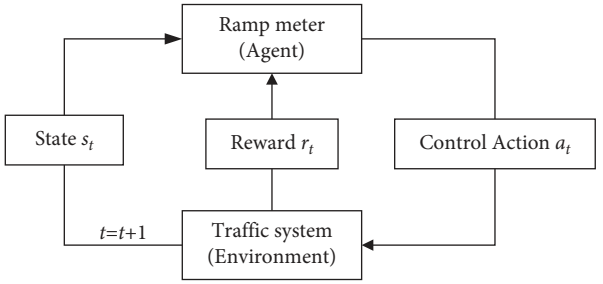

Figure 2: Ramp metering in the general scheme of RL.

$$\pi^*(s) = \arg\max_{a'} Q^*(s, a'). \tag{2}$$

The process of learning policy from the action-value function Q is termed Q-learning. The key step in Q-learning is to estimate $Q^*(s, a)$. To better capture the traffic features in video data. We adopt a deep neural network termed Q-network to approximate $Q^*(s, a)$. The neural network contains several convolutional layers and fully connected layers. The convolutional layers extract traffic features from the video frames, and the fully connected layers estimate the Q value based on the output of convolutional layers.

3.1. State, Action, and Reward Representation

3.1.1. Action. The ramp meter takes two possible actions at fixed decision time step L to determine whether the phase in the next time step is green (G) or red (R), namely, action set $\mathscr{A} = \{G, R\}$.

3.1.2. State. Figure 3 illustrates the process of obtaining visual representations from the raw images to represent traffic states. For simplification purpose, the vehicle is represented by a location point (x_i, y_i) extracted from raw images. Considering a control area of size $X \times Y$ as shown in Figure 3(a), traffic state at time step t is represented by a position matrix $m_t \in \mathbb{R}^{X \times Y}$:

$$m_t(x_i, y_i) = U_V, \quad i \in I_t, \tag{3}$$

where U_V is a constant and I_t refers to the set of vehicles in the control area at time step t.

The position matrix additionally contains the information of the signal states:

$$m_t(x_R, y_R) = \begin{cases} U_G, & \text{if } a_{t-1} = G, \\ U_R, & \text{if } a_{t-1} = R, \end{cases} \tag{4}$$

where (x_R, y_R) is the position of signal light in images and U_G and U_R are constants to represent actions execute in time step $t - 1$.

The other elements in matrix m_t are set to zero:

$$m_t(x, y) = 0, \quad (x, y) \notin \{(x_i, y_i) | i \in I_t\} \cup \{(x_R, y_R)\}. \tag{5}$$

Since one matrix is insufficient to describe vehicle dynamics, we stack the matrices of consecutive N time steps to represent the state s_t:

$$s_t = \{m_{t-(N-1)}, \ldots, m_t\}. \tag{6}$$

For computation efficiency, down-sampling is introduced before the matrix stack to reduce the input dimensions and maintain sufficient information about vehicle positions.

3.1.3. Reward. The reward is critical since it motivates the agent to approach the objective. Vehicle travel times through the system are good measurements of vehicle mobility. However, they are not suitable rewards since they are greatly influenced by previous actions and cannot well quantify the value of the current action. Alternatively, the speed in the merging area and the queue length at the on-ramp are adopted to define the reward. The reward r_t after action a_t is defined by the following:

$$r_t = \mu \bar{v}_t + \omega \bar{q}_t, \quad \mu > 0, \omega < 0, \tag{7}$$

where $\bar{v}_t$ and $\bar{q}_t$ represent the average speed in the merging area and the average queue length at the on-ramp during time step t and $t + 1$, respectively. Positive μ and negative ω denote the reward weights for the speed and the queue length, respectively. The defined rewards balance the priority needs between improving vehicle mobility in the mainline and reducing vehicle delays at the on-ramp.

3.2. Algorithm.

The training problem can be viewed as the regression problem with the loss function defined by the following:

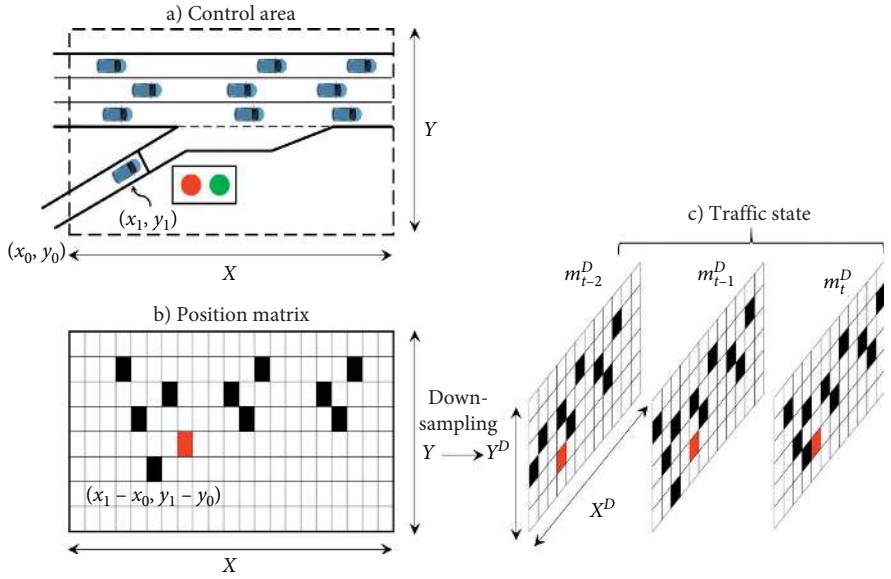

Figure 3: State representation: (a) control area; (b) position matrix m_{t-1}; (c) traffic state s_t.

$$\mathscr{L}(w_i) = \mathbb{E}_{s,a}\left(\left(Q^*(s,a) - Q(s,a,w_i)\right)^2 \right), \tag{8}$$

where w_i denotes the weights in the deep neural network after the ith update and $Q^*(s,a)$ is approximated by the following:

$$Q^*(s,a) \approx r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a', w_i), \quad s_t = s, a_t = a. \tag{9}$$

The gradient-based algorithm Adam [28] is used to update w_i to minimize the loss function.

Although Q-network could handle large state space, it raises the problem of instability in training [29]. Two approaches have been proposed to resolve the problem, illustrated in Figure 4. The experience replay [30] stores the data prepared for training into the replay buffer and samples these data uniformly for training the agent. It contributes to making the training data independently and identically distributed. The target network [29] is introduced for a stable estimation of target value $Q^*(s,a)$. When Q-network is being trained, the target network is freezing. Its parameters w^- are updated by copying w_i in Q-network with a certain frequency. The loss function is redefined as follows:

$$\mathscr{L}_1(w_i) = \mathbb{E}_{s,a}\left(\left(r_{t+1} + \gamma \max_{a'} Q^0(s_{t+1}, a', w^-) - Q^0(s,a,w_i)\right)^2 \right), \tag{10}$$

where $Q^0(s,a,w_i)$ represents the prediction values of the action-value function.

A multitask learning strategy is implemented in the training process to improve learning efficiency. While training the ramp metering Q-network, the mean speed in the merging area and the queue length at the on-ramp are predicted simultaneously. The fully connected layer in the Q-network in multitask learning is extended to two layers to calculate the Q-value and predict the speed and queue length, respectively. Another loss function to measure the prediction accuracy is defined as follows:

$$\mathscr{L}_2(w_i) = \mathbb{E}_{s,a}\left(\left(Q^1(s,a,w_i) - v\right)^2 + \left(Q^2(s,a,w_i) - u\right)^2 \right), \tag{11}$$

where $Q^1(s,a,w_i)$ and $Q^2(s,a,w_i)$ represent the prediction values of the action-value function, mean speed, and queue length, respectively. u and v refer to the ground truth of the speed and queue length, respectively. The total loss function is as follows:

$$\mathscr{L}(w_i) = \mathscr{L}_1(w_i) + \lambda \mathscr{L}_2(w_i). \tag{12}$$

The algorithm for training the ramp metering policy is presented in Algorithm 1. The ramp meter takes new action with the ε-greedy strategy, meaning that the agent does not always adopt the best action derived from current $Q(s,a,w)$. Doing so is helpful to explore unknown policies that may be better.

4. Case Study

4.1. Case Set-Up. The proposed method is evaluated in SUMO, an open-source microscopic traffic simulator [31]. The simulation is performed in the freeway connecting Qingdao and Huangdao through a tunnel in Shandong, China, as shown in Figure 5. The one-lane on-ramp and the upstream three-lane mainline merge. And the freeway gradually reduces to three lanes in the downstream segments.

The simulation considers the road network covering the on-ramp, the upstream and downstream mainlines, and the merging area. The speed limits in the mainline and at the on-ramp are 80 and 40 km/h, respectively. The simulation parameters regarding driving behaviors are calibrated based

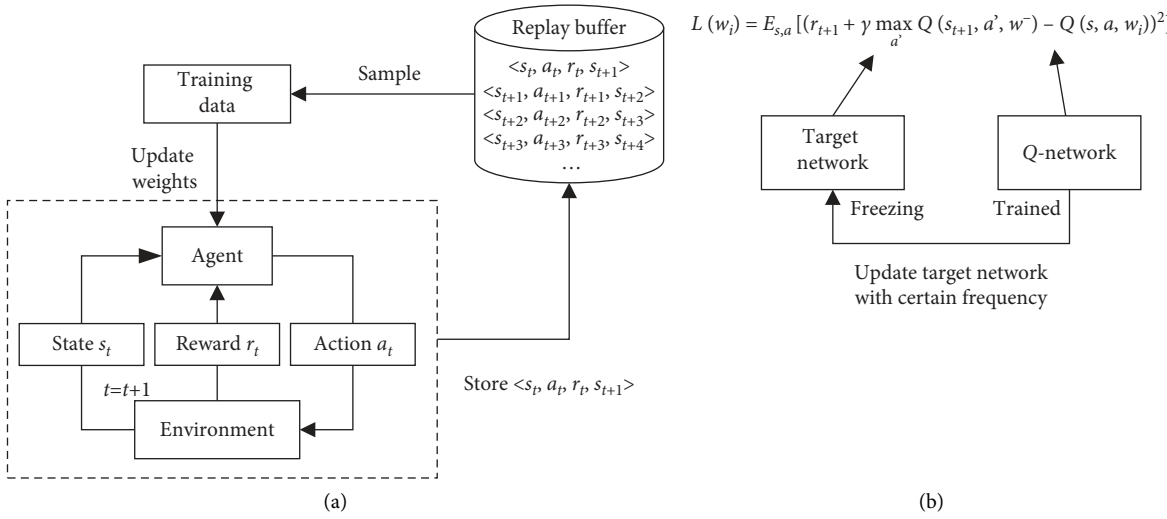

Figure 4: Approaches for stabilizing the training process in deep Q-learning: (a) relay buffer; (b) target network.

Input: Batch size k, learning rate η, decision step size L, exploration rate ε,
Freezing interval F, number of training episode N,
Number of decision time steps T in one episode
Fill the replay buffer randomly
Initialize the action-value function Q with random weights w.
Initialize the target network Q–with weights $w^- = w$
$f \leftarrow 0$
for episode = $\{1, \ldots, N\}$
for $t = \{1, \ldots, T\}$
Observe state s_t and choose action a_t by ε-greedy strategy
Execute action a_t, obtain reward r_t, state s_{t+1}
Store transition <s_t, a_t, r_t, v_t, u_t, s_{t+1}> into replay buffer
Uniformly sample k transitions from replay buffer
Compute the gradient $\nabla_w(\mathscr{L}_1 + \lambda\mathscr{L}_2)$
Update weights w in Q-network by Adam with learning rate η
$f \leftarrow f + 1$
end if
if mod (f, F) = = 0
Update target network Q^- : $w^- \leftarrow w$
end if
end for
end for

Algorithm 1: Deep Q-learning for ramp metering.

on empirical data. The simulation time is 7:50 a.m. to 9:00 a.m. Figure 6 presents the 10-min traffic volumes from 7:50 a.m. to 9:00 a.m. in the mainline and at the on-ramp.

The agent's observation in DRL covers the simulated road network. Three consecutive frames of the traffic video in the controlled area are needed as the observation input of each decision time step. After zooming and down-sampling, the size of one observation is 4×512 in pixels. Three consecutive observations are stacked to represent the traffic state. To accommodate the input size for ramp metering, the Q-network in Mnih et al. [30] is revised as Figure 7, which consists of three convolutional neural network (CNN) layers and one full connection (FC) layer for each output layer. Thus, there are about 2 million weights in the deep neural network that are stored permanently after training. The hyperparameters in the deep Q-learning for ramp metering are listed in Table 1.

For comparative purposes, two scenarios are also considered in the evaluation. The first scenario does not apply ramp metering, and the second scenario adopts the PI-ALINEA method. The PI-ALINEA method fixes the length of the green phase and determines the length of the red phase in the current signal cycle based on the length of the red phase in the previous cycle and the downstream occupancies in the previous and current cycles. The ramp metering rate r_k for k-th cycle is given by the following:

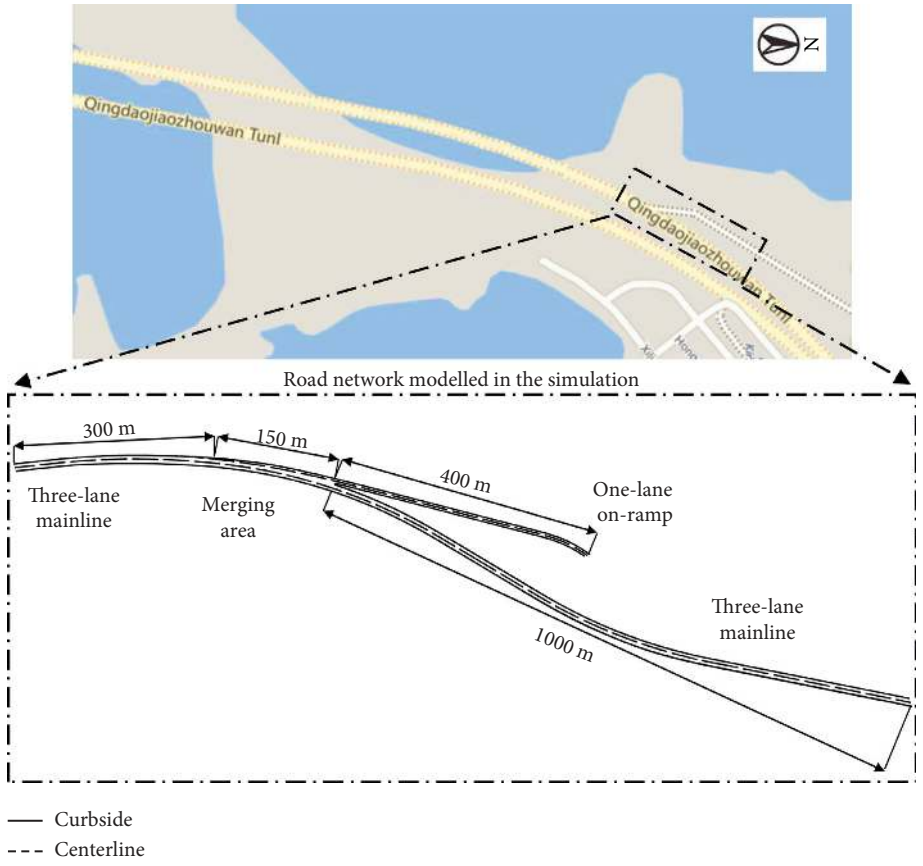

Figure 5: Road network considered in the case study.

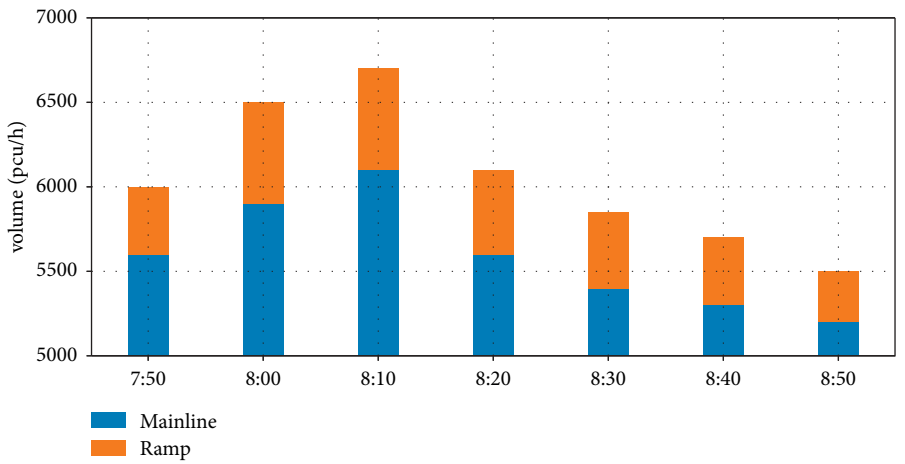

Figure 6: Traffic volumes in the mainline and the on-ramp.

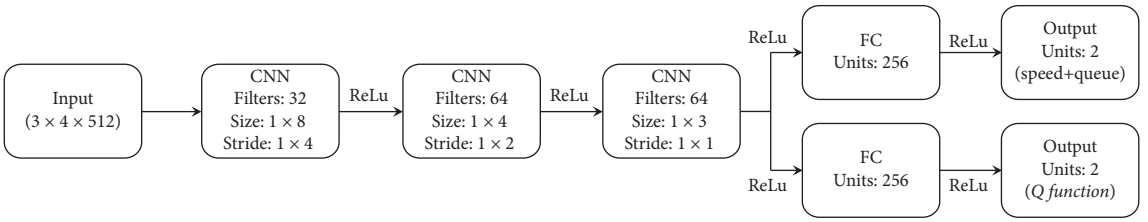

Figure 7: Structure of the deep neural network.

Table 1: Key parameters in deep Q-learning for ramp metering.

Parameters	Value
State matrix dimension	$(3, 4, 512)$
Replay buffer size B	2×10^5
Training steps T	4200
Learning rate η	2.5×10^{-4}
Batch size k	32
Exploration rate ε	0.1
Freeze interval F	10^4
Decision step size L	4 s
Loss weight λ	0.001
Reward weight μ	0.5
Reward weight ω	−0.1

$$r_k = r_{k-1} + K_R[o_k - \widehat{o}] - K_P[o_{k-1} - o_k], \tag{13}$$

where o_k and o_{k-1} represent the mean mainline occupancy during the k-th and $(k+1)$-th cycle, respectively. $\widehat{o}$ refers to the desired value for mainline occupancy, which is set to 16% in this study. $K_R = 90$ and $K_P = 50$ are two regulator parameters in PI-ALINEA, which are fine-tuned to minimize mean travel times in the traffic system. To accommodate the "stop-and-go" scheme, the green time for each cycle is fixed with time step size L and the red time for each cycle is given by $r_k L$.

4.2. Evaluation and Comparison. The Q-network is trained in 10^6 training frames, and each scenario is evaluated in 20 simulation experiments. The training process, presented in Figure 8, can be finished in less than 24 hours on a computing server with NVIDIA Quadro K620 GPU and Intel Xeon CPU E5-1650 (3.6 GHz, 6 cores). The simulations for the training phase and each experiment are initialized with different random seeds. The episode reward and mean travel time are both converged in 400,000 training frames and can maintain a stable value until the end of the training process.

Figure 9 presents the Q1 (25%), Q2 (50%), and Q3 (75%) of the travel times in the mainline for three scenarios. The results demonstrate that ramp metering not only reduces vehicle travel times but also improves the stability of traffic flows in the mainline. The resulting travel times are close when the demand is low at the beginning (8:00–8:10). With the increase in the demand, the median travel time increases faster in the no-control scenario and reaches the maximum value of 4.25 min at 8 : 28. The maximum values of the median travel times resulting from the PI-ALINEA method and the DRL method are 3.75 min and 3.4 min, respectively, which are 11.7% and 20% lower than those in the no-control scenario. As the demand decreases, the travel times of the three scenarios decrease to similar levels. In addition, all the ramp metering methods narrow the interquartile ranges (Q3-Q1) of the travel times. Overall, the proposed DRL method results in shorter travel times with narrower travel time ranges than that of the fine-tuned PI-ALINEA method.

Table 2 presents mean vehicle travel times in 20 simulation experiments. It is found that the performance of the PI-ALINEA method is not always better than that of the no-control scenario. For example, compared with the no-control scenario, the PI-ALINEA method increases vehicle travel time by 10% in the 6$^{\text{th}}$ experiment. In contrast, the proposed DRL method consistently results in a shorter travel time than that of the no-control scenario, demonstrating the stable performance of the proposed method. On average, the DRL method reduces vehicle travel time by 13% when compared with the no-control scenario.

Figure 10 presents the temporal-spatial distribution of the mean speeds along the mainline. The results further demonstrate the efficiency of the ramp metering and the better performance of the proposed method. In the no-control scenario, the minimum speed in the mainline is 12 m/s (43.2 km/h). The minimum speeds resulting from the PI-ALINEA method and the DRL method are 13.3% and 20% higher than those in the no-control scenario, respectively. The area with a speed lower than 15 m/s (54 km/h) is reduced with the PI-ALINEA method, which is further reduced when the DRL method is implemented.

Figure 11 presents the Q1, Q2, and Q3 of the queue lengths at the on-ramp in three scenarios. The results reveal that ramp metering improves vehicle mobility in the mainline at the cost of delays of vehicles at the on-ramp. Without ramp metering, no vehicle queues are formed at the on-ramp. Both ramp metering methods result in vehicle queues at the on-ramp. Nevertheless, the queues produced by the proposed method are shorter than those produced by the PI-ALINEA method. When the PI-ALINEA method and the proposed DRL method are adopted, the maximum values of the Q3 of the queue lengths are 126 m and 67 m, respectively.

Figure 12 presents the mean queue lengths at the on-ramp and the red ratios (i.e., the proportion of the red phase in the simulation time) in each simulation for the two ramp metering methods. The queue length generally increases with the red ratio. However, compared with the PI-ALINEA method, the queues produced by the DRL method are shorter, suggesting that the proposed method could better utilize the green time.

Figure 13 presents the queue lengths at the on-ramp and the occupancy downstream of the merging area for the two ramp metering methods. Figure 13(a) demonstrates that the queue length and the occupancy are highly correlated when the PI-ALINEA method is adopted. The results are understandable. According to Figure 12, a higher red ratio leads to long queue length. Given high occupancy, the PI-

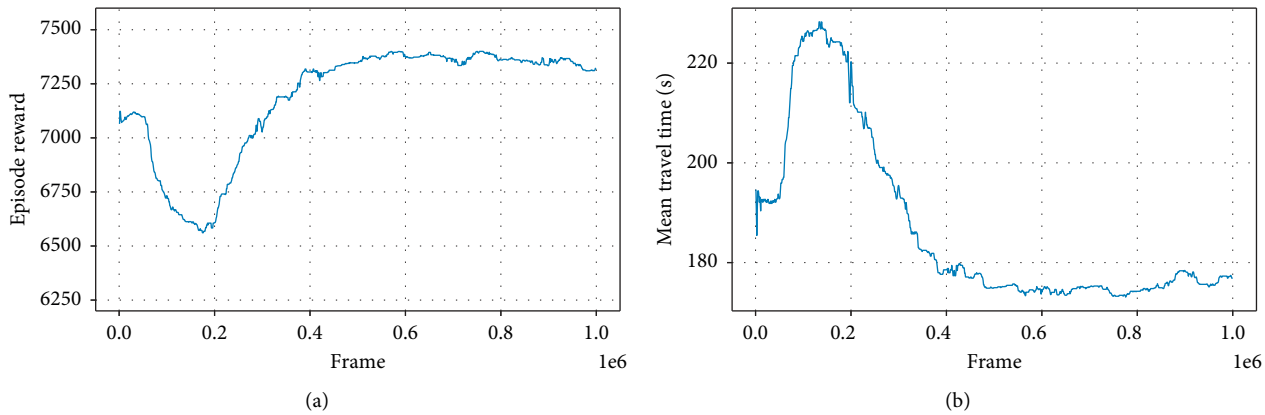

Figure 8: Convergence curves for (a) episode reward and (b) mean travel time.

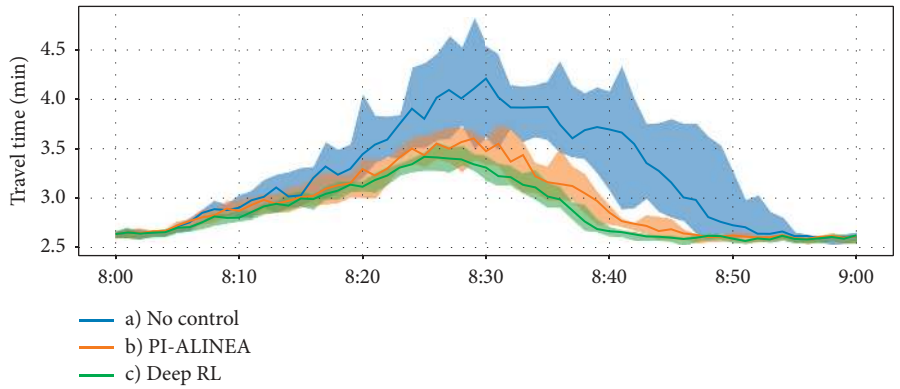

Figure 9: Vehicle travel times in the mainline.

Table 2: Mean travel times in 20 simulation experiments (s).

No.	No controls	PI-ALINEA	DRL
1	186	179 [−4%]	171 [−8%]
2	203	173 [−15%]	171 [−16%]
3	217	175 [−19%]	177 [−18%]
4	227	181 [−20%]	181 [−20%]
5	202	173 [−14%]	177 [−12%]
6	178	**195 [+10%]**	173 [−3%]
7	185	178 [−4%]	174 [−6%]
8	226	180 [−20%]	173 [−23%]
9	175	**214 [+22%]**	169 [−3%]
10	206	183 [−11%]	175 [−15%]
11	188	173 [−8%]	167 [−11%]
12	211	182 [−14%]	170 [−19%]
13	219	175 [−20%]	172 [−21%]
14	170	**172 [+1%]**	170 [−0%]
15	187	182 [−3%]	175 [−6%]
16	216	170 [−21%]	179 [−17%]
17	177	**192 [+8%]**	174 [−2%]
18	228	169 [−26%]	172 [−25%]
19	190	179 [−6%]	175 [−8%]
20	194	180 [−7%]	170 [−12%]

ALINEA method tends to adopt a long red phase, which would lead to long queues at the on-ramp. In contrast, Figure 13(b) shows that the correlation between the queue length and the occupancy is relatively low when the DRL method is implemented. The queue length over time is relatively stable, ranging from 25–50 m. Although the occupancy is low after 8 : 45, the ramp meter is still active. The results indicate that the proposed DRL method not only relies on downstream occupancy but other information extracted from raw data.

The better performance of the proposed DRL method is further revealed when considering traffic flows downstream of the merge area, which is demonstrated in Figure 14. In the beginning, traffic flows in three scenarios are close. During the peak period between 8 : 10–8 : 35, the flows resulting from the DRL method are higher and more stable, suggesting that the proposed DRL method could better alleviate the negative impacts of "capacity drop" resulting from massive merging behaviors. When traffic demand decreases during 8 : 40–9 : 00, traffic flows in the two ramp metering scenarios also decline. Nevertheless, traffic flows in the no-control scenario are still high during 8 : 45–8 : 55 since more vehicles are held in the system in the previous periods.

A perturbation analysis is conducted to evaluate the impact of information contained in video data. As shown in Figure 15, the control area is divided into eight subareas. The sensors applied in PI-ALINEA are located in subarea 2. The information in the subarea is removed from original data to generate perturbed data for each subarea. The impact of

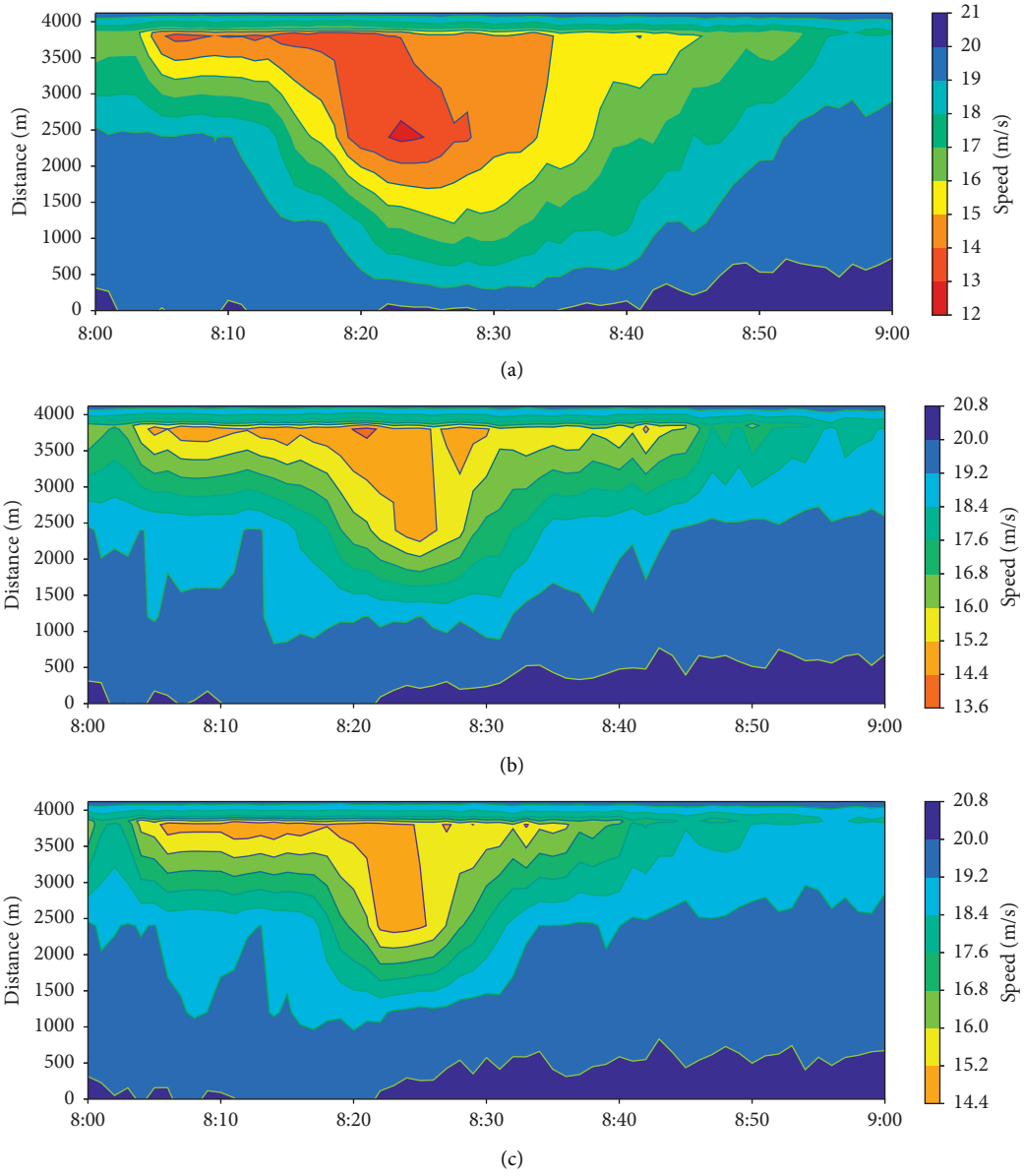

Figure 10: Temporal-spatial distribution of speed in (a) no-control; (b) PI-ALINEA; (c) DRL.

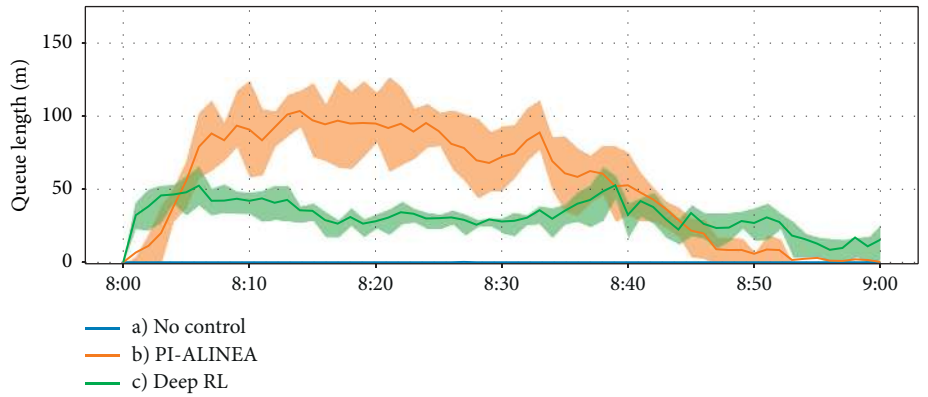

Figure 11: Queue length at the on-ramp.

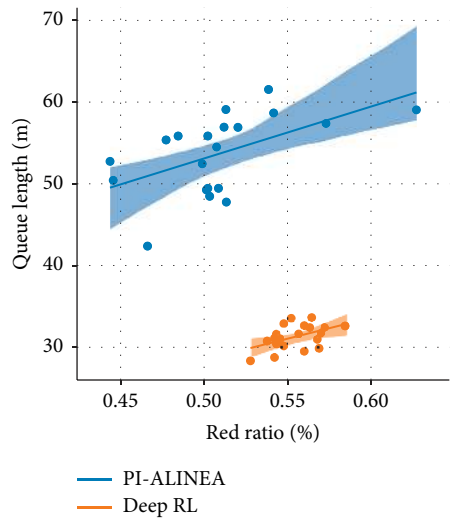

Figure 12: Relationship of queue length and red ratio for (a) PI-ALINEA and (b) DRL.

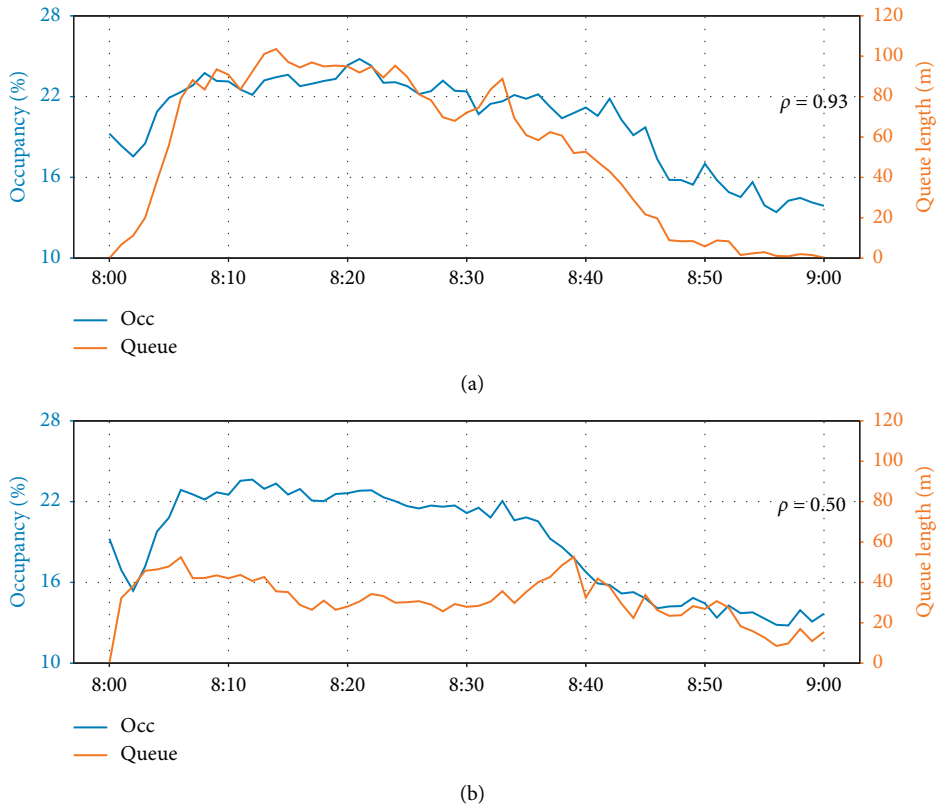

Figure 13: Queue length and occupancy for (a) PI-ALINEA and (b) DRL.

missing information for each subarea is evaluated in 20 simulation experiments with perturbed data.

Figure 16 presents the box plot of travel time when the information of different subareas is removed, and the baseline is the result of experiments that take original data as input. Apart from subarea 2, the information from other subareas such as subarea 3 and 4 also influences the control performance obviously, which proves that the DRL-based controller takes more information from video data into consideration besides the measurements applied in PI-ALINEA.

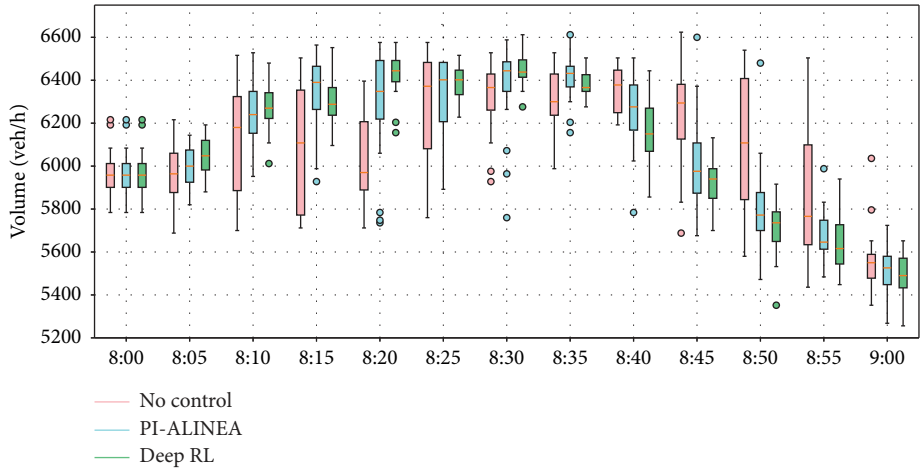

Figure 14: Downstream volume in three scenarios.

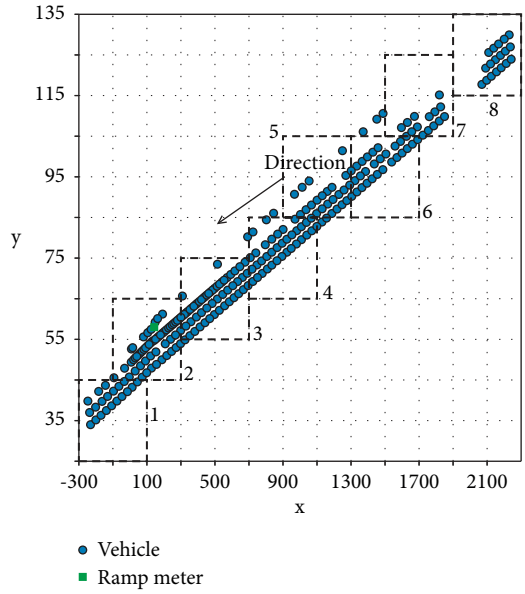

Figure 15: Division subareas.

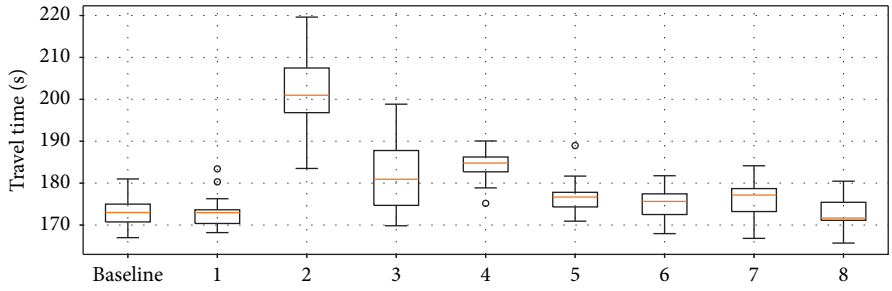

Figure 16: Travel time with perturbed input data.

5. Conclusion and Future Research

This study proposes a DRL method for local ramp metering based on traffic video data. The proposed method learns optimal strategies directly from high-dimensional visual inputs, overcoming the reliance on the traditional traffic measurements. The better performance of the proposed method is demonstrated in a real-world case study. Compared with the traditional PI-ALINEA method, the proposed method results in lower travel times in the mainline and

shorter vehicle queue lengths at the on-ramp and could better alleviate the negative impacts of "capacity drop" resulting from massive merging behaviors. The results suggest that the proposed DRL method is able to extract useful information from the video data for better ramp metering controls.

It is convenient to deploy the proposed model in practice since the well-trained model is lightweight, and the input image data are accessible with the latest communication technologies. For future research, additional studies are suggested to conduct to investigate the stability of the proposed method under various traffic and weather circumstances. We assume that vehicle location information from the videos is accurate. Actually, the data quality may be disturbed by factors such as weather, vehicle sizes, and colors. It is valuable to design a ramp metering algorithm that is robust to those disturbances. In addition, abnormal events, such as accidents, are not considered in this study. Efficiently training the ramp metering to adapt to abnormal events remains to be investigated. Finally, extending the proposed method for coordinated ramp metering controls is worth pursuing.

Data Availability

The data used to support the results of this study are available from the corresponding author upon request.

Disclosure

An initial draft of this paper was submitted as a preprint in the following link: https://arxiv.org/abs/2012.12104 [32]. The aim behind posting this version was to collect insightful comments from the research community. This current document is the last update of our study.

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (71671124 and 71901164), the Natural Science Foundation of Shanghai (19ZR1460700), the Shanghai Science and Technology Committee (19DZ1202900), the Shanghai Municipal Science and Technology Major Project (2021SHZDZX0100), and the Fundamental Research Funds for the Central Universities. The corresponding author was sponsored by the Shanghai Pujiang Program (2019PJC107).

References

[1] A. Srivastava and N. Geroliminis, "Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model," *Transportation Research Part C: Emerging Technologies*, vol. 30, pp. 161–177, 2013.

[2] M. Papageorgiou, C. Kiakaki, V. Dinopoulou, A. Kotsialos, and W. Yibing, "Review of road traffic control strategies," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2043–2067, 2003.

[3] D. Levinson and L. Zhang, "Ramp meters on trial: evidence from the Twin Cities metering holiday," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 10, pp. 810–828, 2006.

[4] N. Bhouri, H. Haj-Salem, and J. Kauppila, "Isolated versus coordinated ramp metering: field evaluation results of travel time reliability and traffic impact," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 155–167, 2013.

[5] L. Faulkner, F. Dekker, D. Gyles, I. Papamichail, and M. Papageorgiou, "Evaluation of HERO-coordinated ramp metering installation at M1 and M3 freeways in queensland, Australia," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2470, no. 1, pp. 13–23, 2014.

[6] X. Y. Lu, C. J. Wu, and X. Y. Lu, "Field test implementation of coordinated ramp metering control strategy: a case study on SR-99N," in *Proceedings of the Transportation Research Board Meeting 96th Annual Meeting*, Washington D.C., WA, USA, January 2018.

[7] C. Systematics, *Twin Cities Ramp Meter Evaluation*, Cambridge Systematics, Medford, MA, USA, 2001.

[8] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville, "ALINEA: a local feedback control law for on-ramp metering," *Transportation Research Record*, vol. 1320, no. 1, pp. 58–67, 1991.

[9] Y. Wang, E. B. Kosmatopoulos, M. Papageorgiou, and I. Papamichail, "Local ramp metering in the presence of a distant downstream bottleneck: theoretical analysis and simulation study," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2024–2039, 2014.

[10] J. A. Wattleworth, *Peak-period Analysis and Control of a Freeway System*, Texas Transportation Institute, San Antonio, TX, USA, 1965.

[11] D. P. Masher, D. Ross, P. Wong, P. Tuan, H. M. Zeidler, and S. Petracek, *Guidelines for Design and Operation of Ramp Control Systems*, Stanford Research Institute, Menlo Park, CA, USA, 1975.

[12] E. Smaragdis and M. Papageorgiou, "Series of new local ramp metering strategies," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1856, pp. 74–86, 2004.

[13] M. Papageorgiou, J.-M. Blosseville, and H. Haj-Salem, "Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part II: coordinated on-ramp metering," *Transportation Research Part A: General*, vol. 24, no. 5, pp. 361–370, 1990.

[14] I. Papamichail and M. Papageorgiou, "Traffic-responsive linked ramp-metering control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 111–121, 2008.

[15] M. Papageorgiou and R. Mayr, "Optimal decomposition methods applied to motorway traffic control," *International Journal of Control*, vol. 35, no. 2, pp. 269–280, 1982.

[16] T. Bellemans, B. De Schutter, and B. De Moor, "Model predictive control for ramp metering of motorway traffic: a case study," *Control Engineering Practice*, vol. 14, no. 7, pp. 757–767, 2006.

[17] I. Papamichail, A. Kotsialos, I. Margonis, and M. Papageorgiou, "Coordinated ramp metering for freeway networks-a model-predictive hierarchical control approach," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 3, pp. 311–331, 2010.

[18] M. Davarynejad, A. Hegyi, J. Vrancken, and J. van den Berg, "Motorway ramp-metering control with queuing consideration using Q-learning," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1652–1658, Washington, DC, USA, October 2011.

[19] K. Rezaee, B. Abdulhai, and H. Abdelgawad, "Application of reinforcement learning with continuous state space to ramp metering in real-world conditions," in *Proceedings of the15th International IEEE Conference on Intelligent Transportation Systems*, pp. 1590–1595, Anchorage, AK, USA, September 2012.

[20] A. Fares and W. Gomaa, "Freeway ramp-metering control based on reinforcement learning," in *Proceedings of the 11th IEEE International Conference on Control & Automation (ICCA)*, pp. 1226–1231, Taichung, Taiwan, June 2014.

[21] H. Yang and H. Rakha, "Reinforcement learning ramp metering control for weaving sections in a connected vehicle environment," in *Proceedings of the Transportation Research Board 96th Annual Meeting*, Washington DC, WA, USA, January 2017.

[22] F. Belletti, D. Haziza, G. Gomes, and A. M. Bayen, "Expert level control of ramp metering based on multi-task deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1198–1207, 2018.

[23] F. Deng, J. Jin, Y. Shen, and Y. Du, "Advanced self-improving ramp metering algorithm based on multi-agent deep reinforcement learning," in *Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Auckland, New Zealand, Octom 2019.

[24] J. Gao, Y. Shen, J. Liu, M. Ito, and N. Shiratori, "Adaptive traffic signal control: deep reinforcement learning algorithm with experience replay and target network," 2017, https://arxiv.org/abs/1705.02755.

[25] H. Wei, G. Zheng, H. Yao, and Z. Li, "IntelliLight: a reinforcement learning approach for intelligent traffic light control," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA, July 2018.

[26] X. Liang, X. Du, G. Wang, and Z. Han, "A deep reinforcement learning network for traffic light cycle control," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1243–1253, 2019.

[27] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, USA, 1998.

[28] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.

[29] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

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

[31] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker-Walz, "Recent development and applications of SUMO-Simulation of Urban MObility," *International Journal of Agile Systems and Management*, vol. 5, no. 3&4, 2012.

[32] B. Liu, Y. Tang, Y. Ji, Y. Shen, and Y. Du, "A deep reinforcement learning approach for ramp metering based on traffic video data," 2020, https://arxiv.org/abs/2012.12104.