applied sciences

MDPI

Article

Traffic Signal Optimization for Multiple Intersections Based on Reinforcement Learning

Jaun Gu [1], Minhyuck Lee [1], Chulmin Jun [1,*], Yohee Han [2], Youngchan Kim [2] and Junwon Kim [2]

[1] Department of Geoinformatics, University of Seoul, Seoul 02504, Korea; umseakind2@uos.ac.kr (J.G.); lmhll123@uos.ac.kr (M.L.)
[2] Department of Transportation Engineering, University of Seoul, Seoul 02504, Korea; yeohee@gmail.com (Y.H.); yckimm@uos.ac.kr (Y.K.); mirageno21@uos.ac.kr (J.K.)
* Correspondence: cmjun@uos.ac.kr

Abstract: In order to deal with dynamic traffic flow, adaptive traffic signal controls using reinforcement learning are being studied. However, most of the related studies are difficult to apply to the real field considering only mathematical optimization. In this study, we propose a reinforcement learning-based signal optimization model with constraints. The proposed model maintains the sequence of typical signal phases and considers the minimum green time. The model was trained using Simulation of Urban MObility (SUMO), a microscopic traffic simulator. The model was evaluated in the virtual environment similar to a real road with multiple intersections connected. The performance of the proposed model was analyzed by comparing the delay and number of stops with a reinforcement learning model that did not consider constraints and a fixed-time model. In a peak hour, the proposed model reduced the delay from 3 min 15 s to 2 min 15 s and the number of stops from 11 to 4.7 compared to the fixed-time model.

Keywords: traffic signal optimization; reinforcement learning; adaptive traffic signal control; multiple intersections; Deep Q-network

Citation: Gu, J.; Lee, M.; Jun, C.; Han, Y.; Kim, Y.; Kim, J. Traffic Signal Optimization for Multiple Intersections Based on Reinforcement Learning. *Appl. Sci.* **2021**, *11*, 10688. https://doi.org/10.3390/app112210688

Academic Editors: Nikos D. Lagaros, Vagelis Plevris and Paola Pellegrini

Received: 20 August 2021
Accepted: 10 November 2021
Published: 12 November 2021

1. Introduction

Traffic signal control plays an essential role in city management because traffic congestion brings economic, environmental, and social disadvantages. Traffic signal control aims to minimize congestion by determining the optimal values of parameters such as the cycle length and phases duration [1,2]. In many areas, the traffic signal control systems based on a fixed-time model are still in use [3–5]. While these systems are easy to implement, they cannot respond flexibly to dynamic traffic flows [6,7].

To quickly respond to variety in the traffic environment, signal control systems should be able to choose their own actions without waiting for instructions from a central computer [8]. Therefore, reinforcement learning models are being studied that allow the traffic signal controller to receive realtime data around the intersection, such as traffic volume and vehicle speed, and change signal appropriately for the given traffic situation [9,10]. If the above sentence is expressed in reinforcement learning terms, the controller is the agent, the data input to the controller is the state, the controller's decision is the action, and the benefit provided to the agent according to the action is called reward. The goal of reinforcement learning is to maximize the future reward that an agent can obtain [11].

2. Literature Review

Reinforcement learning is the most recently used algorithm in the field of signal control research. However, most studies have not considered the constraints applied to a real-world intersection or tested in a local area such as a single intersection. Touhbi et al. (2017) analyzed the possibility of using the Q-Learning algorithm for adaptive traffic signal control [12]. The Q-Learning algorithm was helpful in resolving traffic congestion

compared to the fixed-time model, but different results were obtained depending on the definition of reward and various traffic volumes. Liang et al. (2019) proposed a DQN-based signal control model [13]. The state was defined as a grid-type location and the speed of vehicles around the intersection. The reward was the difference in accumulated waiting time between learning cycles, and the action was to select one of phases. Wang et al. (2019) proposed a model based on the assumption that data are collected by a loop detector [14]. Since the state used as input to the model is not a data format that can be obtained, a method for converting the data acquired through the detector into data useful for the model was presented. Gong et al. (2019) proposed a cooperative learning method in which signal controllers at adjacent intersections share a state they can observe with each other [15]. Using this, a traffic signal optimization model for multiple intersections was proposed.

Chu et al. (2019) pointed out the limitations of the centralized reinforcement learning model and suggested a way to optimize a large-scale road network by placing the model at each intersection [16]. As the algorithm of the model, A2C (Advantage Actor Critic) was proposed, and the state of different scales was delivered to the model with city-unit traffic volume data and actual observable traffic flow data. Egea et al. (2020) pointed out the limitations of the realtime response of the existing adaptive signal control method and suggested a reinforcement learning-based signal control model as an alternative [17]. Efficiency was evaluated through various indices for the compensation that is judged to have the greatest impact on the model's performance. Rasheed et al. (2020) introduced a multiagent-based reinforcement learning algorithm [10]. The model was designed to solve high-level problems such as dynamic traffic volume through cooperation between agents. As a result of the simulation, it was shown that the travel time was reduced through the proposed model.

In general, traffic signal control has constraints such as the sequence of phases is fixed, and the minimum green time is given. However, in related studies, the constraints are ignored in consideration of only mathematical optimization such as delay minimization [18–21]. If the phase sequence is random or does not give a minimum green time, this can cause over-waiting for vehicles and confuse drivers [16,22]. In this study, we propose a model that can compensate for the problems that arise when the reinforcement learning-based signal control models proposed in related studies are applied to actual road networks. The proposed model maintains the same signal order as the fixed-time model and gives the minimum green time. Therefore, it can be applied to the actual signal controller. The performance of the proposed model was evaluated in a simulation environment depicting real roads connected with multiple intersections. In order to evaluate the effect of constraints, a comparative simulation between the proposed model and the reinforcement learning model that did not consider the constraints was performed.

3. Methods

3.1. Learning Process

Figure 1 shows the learning process of the reinforcement learning-based traffic signal optimization model. The microscopic simulation environment was implemented by Simulation of Urban MObility (SUMO). The model received the realtime traffic flow of the intersection as a state. Based on this, the action determined by the model was implemented at the intersection by SUMO's signal controller. Traffic flow changed due to signal control was delivered to the model as a reward. By repeating this process, the model learned an optimized signal pattern that minimized traffic congestion, such as the vehicles' delay and the number of vehicles stopped. The reinforcement learning model was designed based on Deep Q-network (DQN). Since the given problem was a classification that selected an action appropriate for the situation, the SoftMax Function was applied to the activation function of the output layer.

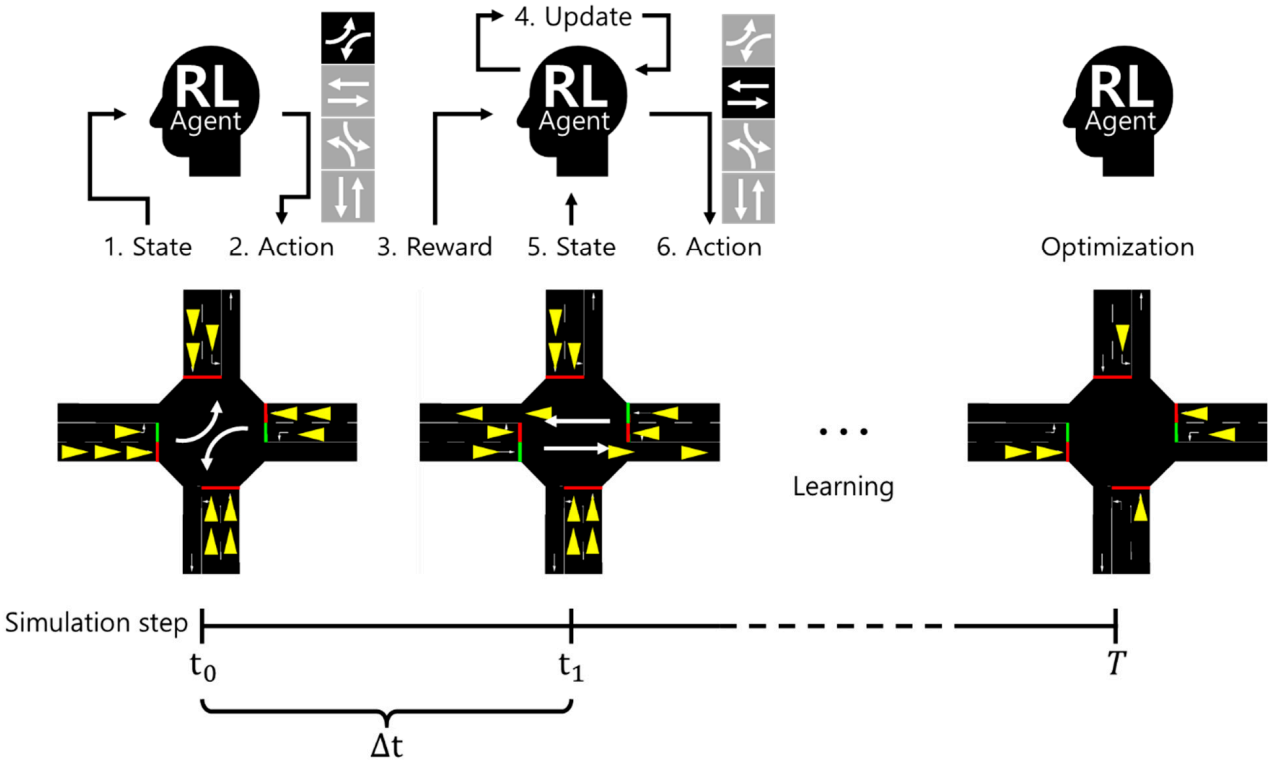


Figure 1. Learning process of the reinforcement learning-based traffic signal control model.

3.2. State

The state is defined as Equation (1). Q_t represents the queue of each lane at time t. At an intersection with k incoming lanes, q_t^i represents the number of vehicles stopped in the incoming lane i at time t. P_t indicates which phase is currently on. When there are n phases, if the jth phase is active, $p_j = 1$, otherwise $p_j = 0$. d_t means the elapsed time of the currently turned-on phase.

$$\begin{aligned} S_t &= [Q_t, P_t, d_t] \\ Q_t &= \left[q_t^1, q_t^2, \ldots, q_t^k\right] \\ P_t &= [p_1, p_2, \ldots, p_n] \end{aligned} \tag{1}$$

Figure 2 shows an example of traffic control at a single intersection. There were 10 incoming lanes into the intersection, and phase 4 was in progress for 10 s. Stopped vehicles are marked in red. Therefore, in the order of the lanes, $Q_t = [0,\ 1,\ 0,\ 0,\ 3,\ 1,\ 2,\ 0,\ 0,\ 0]$ and with phase 4 on, $P_t = [0,\ 0,\ 0,\ 1]$ and accordingly gave a green signal to the 3rd, 4th, 8th, and 9th lanes. Finally, the phase lasted for 10 s, $d_t = [10]$.

3.3. Action

The action was to select whether to keep the current phase ($A_t = 0$) or change to the next phase ($A_t = 1$). Since the proposed model has a constraint that the sequence of the phases is maintained, actions such as returning to the previous phase or skipping the next phase were impossible. In addition, the proposed model included a minimum green time, therefore, every phase must be active once per signal cycle. In Figure 2, the agent decided whether to keep phase 4 or change to phase 1. If the minimum green time for phase 4 was 15 s, the agent could only select to keep phase 4 because d_t is still 10 s.

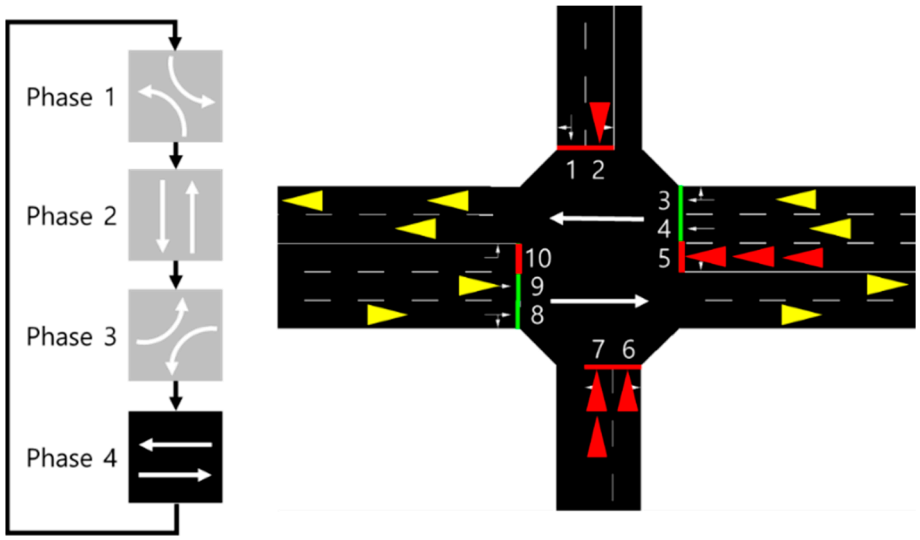


Figure 2. Example of traffic signal control at a single intersection.

3.4. Reward

The traffic flow changed during the time interval Δt by the action A_t was given as a reward to the model. Reward $R_{t+\Delta t}$ is defined as Equation (2). $q^i_{t+\Delta t}$ is defined the number of vehicles stopped in the incoming lane i at time step $t + \Delta t$ and $f^o_{t,t+\Delta t}$ is defined as the number of vehicles located in the outgoing lane among the vehicles passing through the intersection between time step t and time step $t + \Delta t$. Accordingly, the reward $R_{t+\Delta t}$ was defined as the number of passing vehicles compared to stopped vehicles. As the number of stopped vehicles decreased and the number of passing vehicles increased, the reward was increased.

$$R_{t+\Delta t} = \sum_{l=1} \left(f^o_{t,t+\Delta t}\right) / \sum_{i=1} \left(q^i_{t+\Delta t}\right) \quad (2)$$

4. Simulation

The performance of the proposed model, the reinforcement learning-based comparison model excluding constraints, and the fixed-time model, PASSER II, were analyzed for two scenarios. The training of the proposed model and the comparison model was carried out by generating random traffic under the same conditions. When the accumulated waiting time or reward no longer decreased and converged, it was judged that learning was complete. Table 1 shows the number of repeated episodes until learning was complete. The simulation time of scenario 2 was based on the period of the acquired traffic data. For the proposed model to respond flexibly to realtime traffic flow, the time interval Δt was set to 3 s. Agents meaning signal controllers are equal to the number of intersections. In this simulation, it was assumed that the signal controller could obtain the traffic situation in realtime using the vehicle detectors.

Table 1. Parameters used in the model.

Parameters	Scenario 1	Scenario 2
Number of episodes	100	160
Simulation time T of one episode (second)	3600	18,000
Time interval Δt (second)	3	3
Learning rate	0.0001	0.0001
Number of intersections	2	6

4.1. Scenario 1

In Scenario 1, performance evaluation of the proposed model and the fixed-time model was performed before comparison with the model without constraints. Figure 3 is a simple road structure used in the evaluation. Two intersections were connected, and the traffic volume is indicated on each lane. The same phase sequence was applied to both intersections. In the fixed-time model, the duration of each phase was calculated sequentially as 14, 56, 13, and 27 s. Therefore, the cycle length was 110 s.

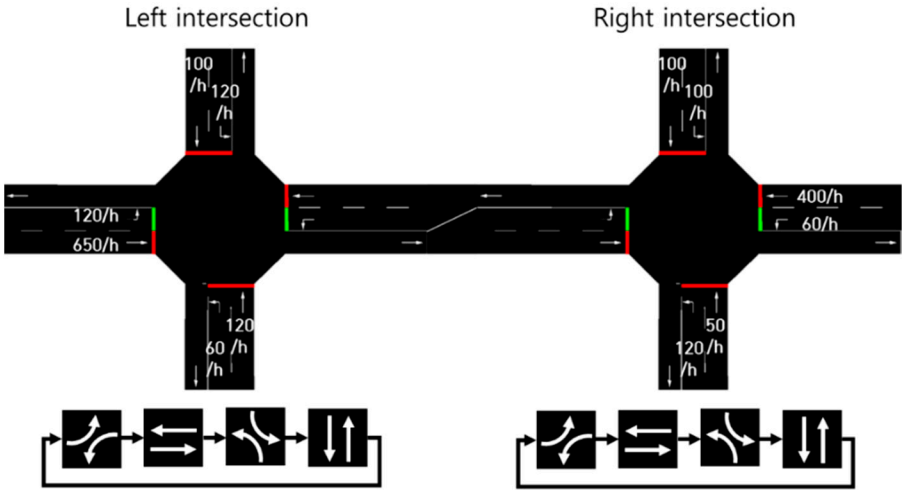


Figure 3. Simple network with two intersections.

Figure 4 shows the duration ratio for each phase of the proposed model and fixed-time model. Since the proposed model dynamically responded to realtime traffic flow, the duration of the phase changed with each signal cycle. The average phase duration ratio appeared similar to the optimization result of the fixed-time model. Compared to the fixed-time model, the proposed model reduced the average delay per vehicle from 40 s to 30 s and the average number of stops per vehicle from 2.5 to 2 times.

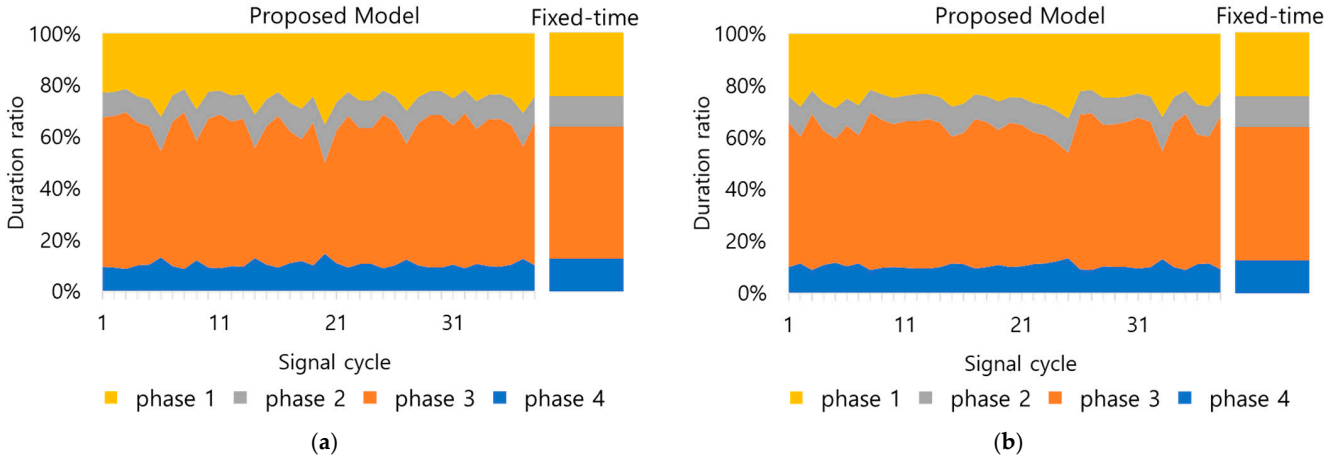


Figure 4. Phase duration ratio of scenario 1: (**a**) Left intersection; (**b**) Right intersection.

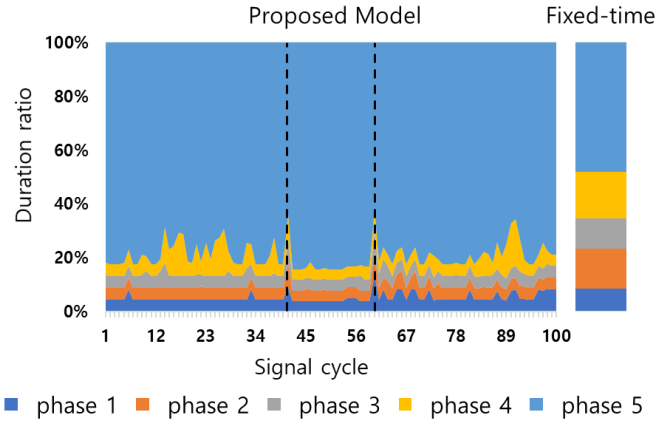
4.2. Scenario 2

In Scenario 2, a comparison model without constraints was added to analyze the performance of the proposed model. Figure 5 show a road network with six continuous intersections in the real world. The total length of the main road with 6 intersections is 2.5 km, and the distance between intersections is 650 m at the maximum and at least 70 m,

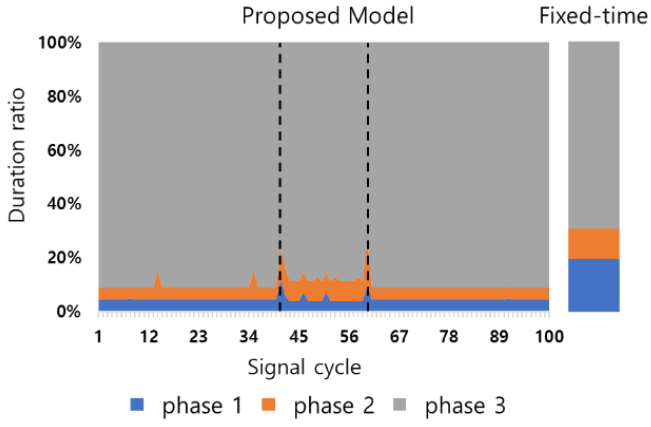
with an average of 400 m. The model was evaluated by applying the collected traffic data. From 4 to 6 pm, 8200 vehicles were created, from 6 to 7 pm, 6200 vehicles were created, and from 7 to 9 pm, 6400 vehicles were created. The least traffic volume was at intersection 3, with 2000 vehicles per hour. On the other hand, the highest traffic volume was intersection 6, with 3800 vehicles per hour. The average speed was 70 km/h. As measure of effectiveness, the cumulative delay, and the cumulative number of stops at each intersection, and the average delay and the average number of stops per vehicle were considered. In addition, the proposed model set the cycle length equal to the fixed model to maintain the set offset at each intersection.

Figure 5. Complex network with six intersections.

Figure 6 is the same as Figure 4, it shows the phase duration ratio of the proposed model and fixed-time model. The comparison model could not calculate the duration ratio, because the phase sequence was not constant. The cycle length was set to 160 s. From the 41st cycle to the 59th cycle, it was set to 180 s because there was heavy traffic. Unlike Scenario 1, the average duration ratio of the proposed model was different from that of fixed-time model. The proposed model showed a tendency to give longer green time to the main road with a lot of traffic.



(**a**)



(**b**)

Figure 6. *Cont.*

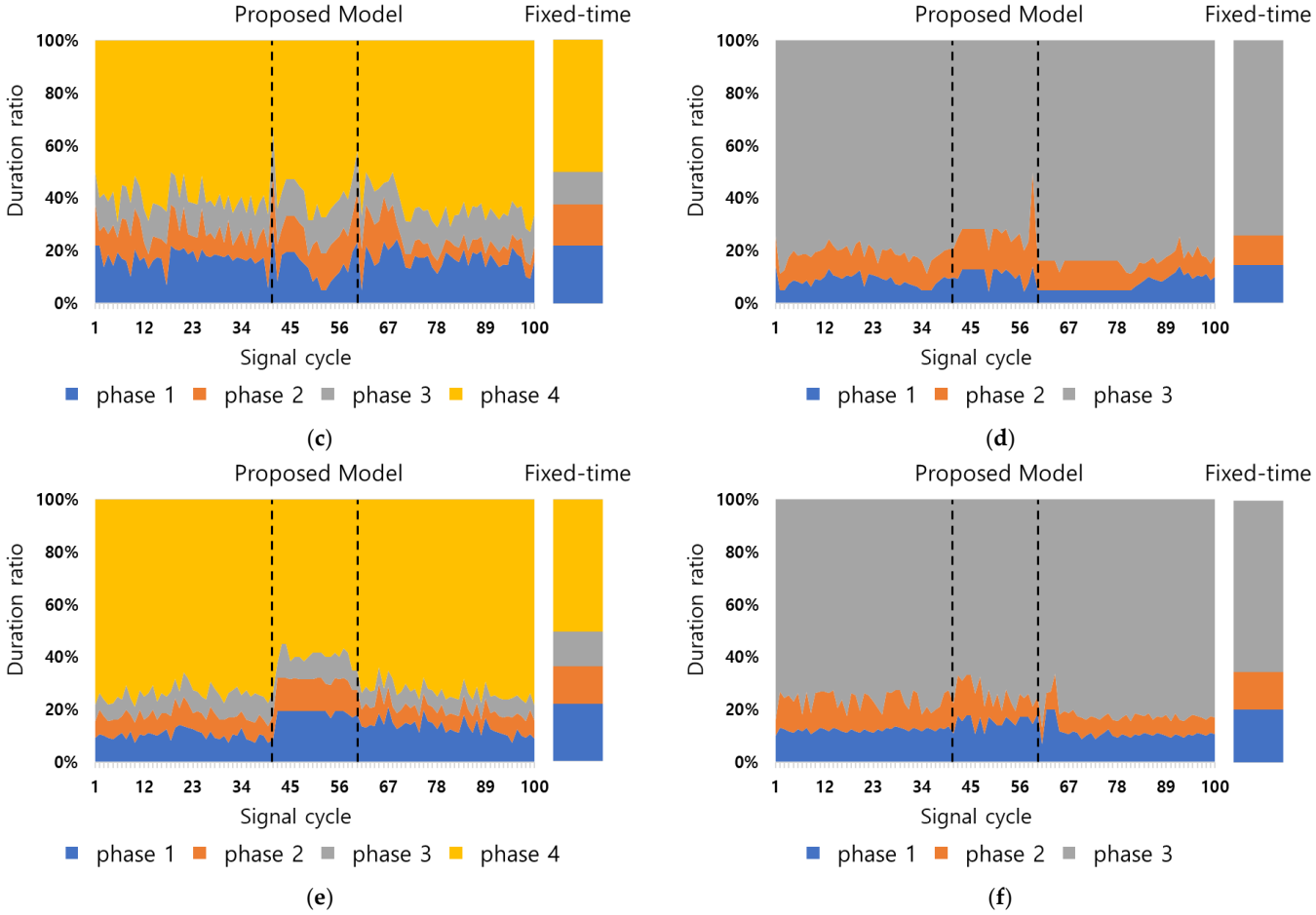
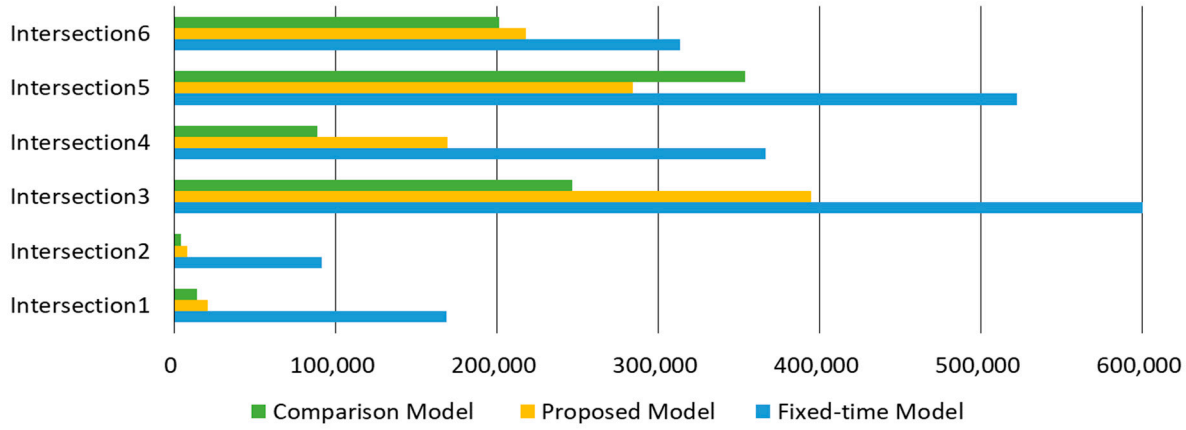


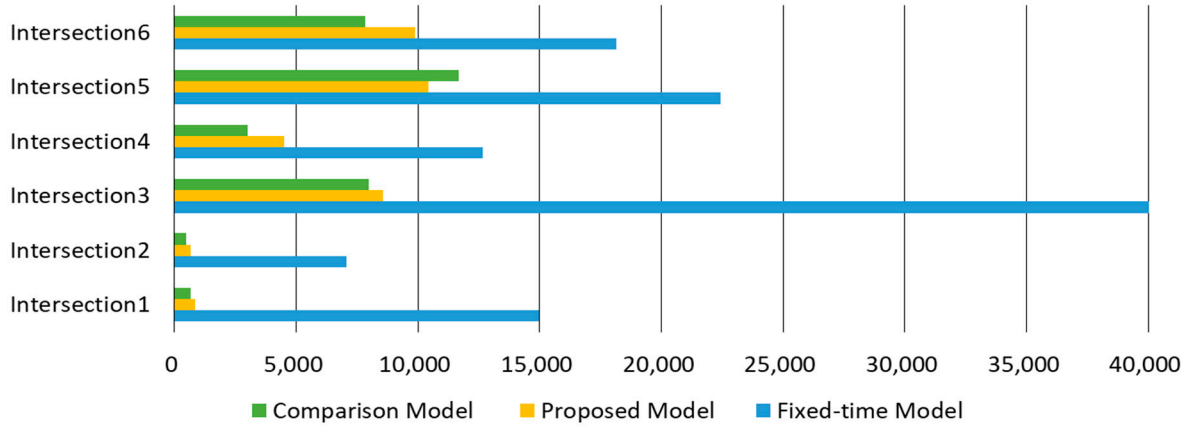
Figure 6. Phase duration ratio of scenario 2: (**a**) Intersection 1; (**b**) Intersection 2; (**c**) Intersection 3; (**d**) Intersection 4; (**e**) Intersection 5; and (**f**) Intersection 6. The time point at which the cycle length changes is indicated by a dotted line. The first cycle length is 160 s and changes to 180 s after the first dotted line, and then changes back to 160 s after the second dotted line.

Figure 7 shows the cumulative delay and number of stops for each intersection during the simulation. At all intersections except for intersection 5, the performance of the comparison model without constraints was the best. However, the number of stops of the proposed model and the comparison model was similar. The proposed model reduced delay by up to 88% to at least 31%, and the number of stops by up to 95% to at least 46% compared to the fixed-time model. The fixed-time model had the worst congestion at intersection 3, and the comparative model had the worst congestion at intersection 5. The proposed model had the longest delay at intersection 3, and the highest number of stops at intersection 5.

Figure 8 compares the average delay and the average number of stops per vehicle of each model by time period. Compared with the fixed-time model, the reinforcement learning-based models showed excellent performance. The delay decreased by 48% for the proposed model and 55% for the comparison model compared to the fixed model. The number of stops decreased by 67% for the proposed model and 73% for the comparison model compared to the fixed model.



(**a**)



(**b**)

Figure 7. Measure of effectiveness by intersections: (**a**) Cumulative delay; (**b**) Cumulative number of stops.

From 6–7 pm, the traffic volume increased by 33% compared to other time periods. Accordingly, the delay and the number of stops also increased. In the peak hour, the fixed model waited for an average of 3 min and 15 s and the number of stops was 11 times, whereas in the proposed model, the delay was reduced to 2 min and 15 s, and the number of stops was reduced by more than half to 4.7. In addition, the number of stops of the proposed model was 0.7 times from 4 to 6 pm when the traffic volume was low. It was an ideal result that the number of stops was less than 1 when passing 6 intersections on a 2.5 km road.

Figure 9 shows the signal pattern calculated for each model as a space-time diagram. It shows the signal patterns of each intersection and the trajectories of vehicles accordingly. In the case of the fixed-time model, the average travel time per vehicle was 250 s, while the comparison model decreased by about 60 s to 190 s. The proposed model decreased by about 30 s to 220 s. However, for the comparison model, the phase duration was irregular and short.

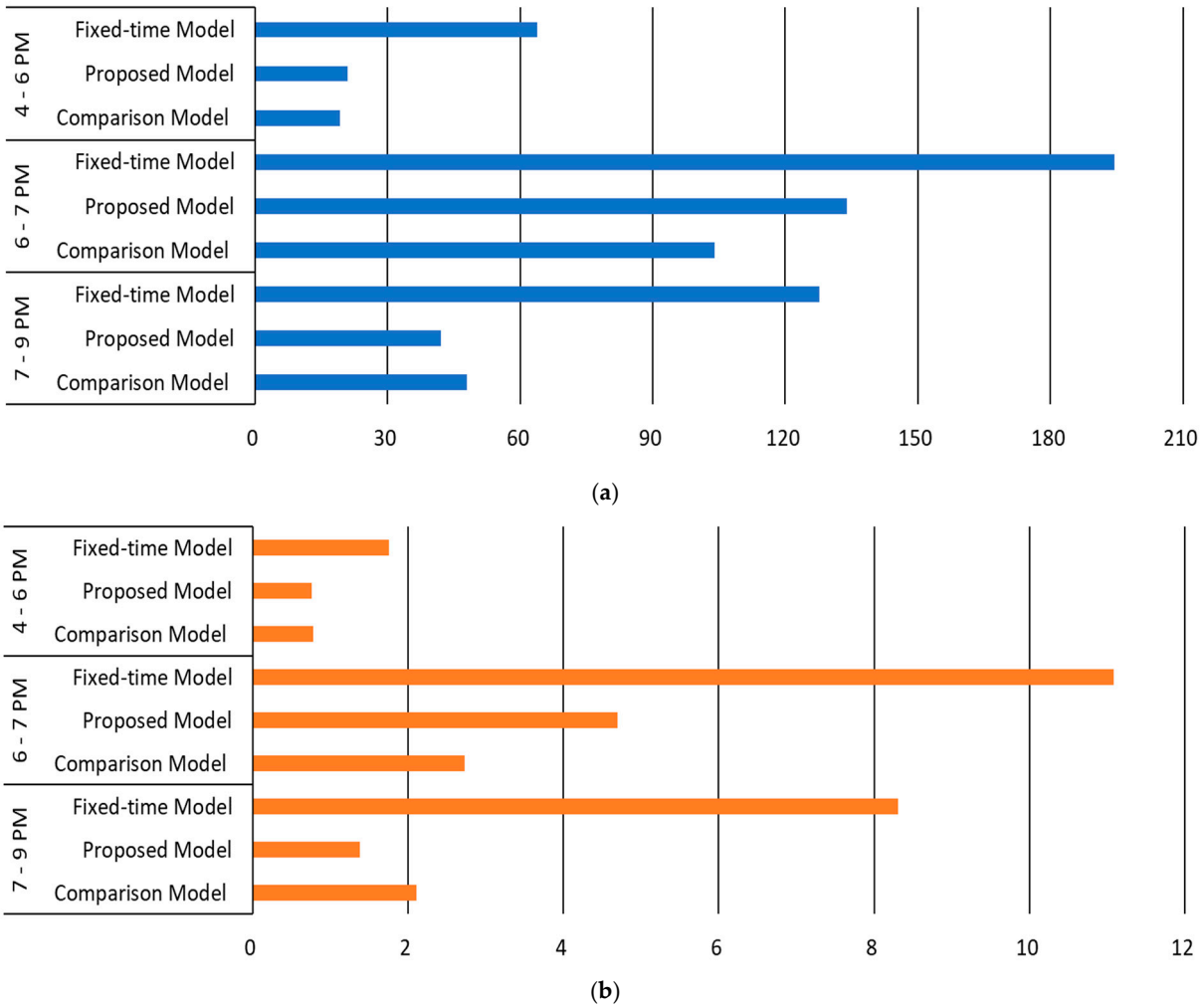
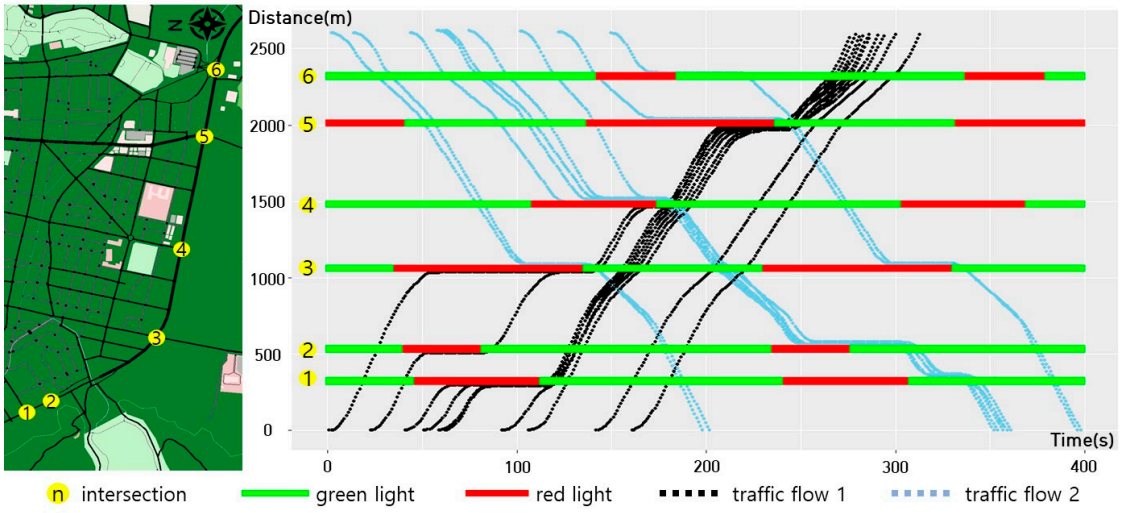


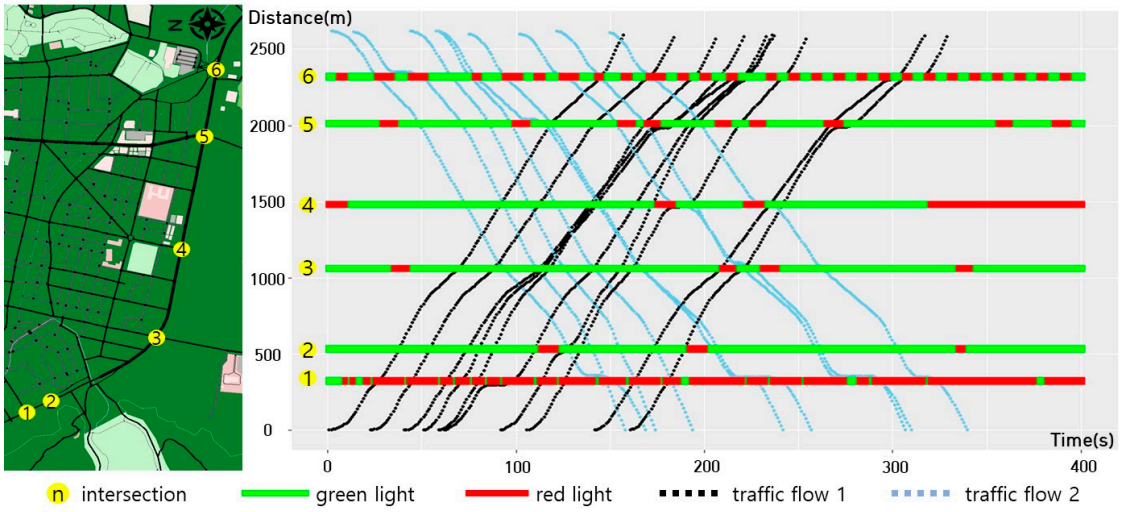
Figure 8. Measure of effectiveness: (**a**) Average delay per vehicle; (**b**) Average number of stops per vehicle.

Figure 10 shows the traffic situation when the proposed model and the comparison model were applied to each intersection. The color of the road indicates the average speed of the vehicles. In the comparison model, traffic congestion occurred at intersection 5. On the other hand, the proposed model can be seen as a solution to the congestion at intersection 5. However, some congested sections occurred between intersection 3 and intersection 4.

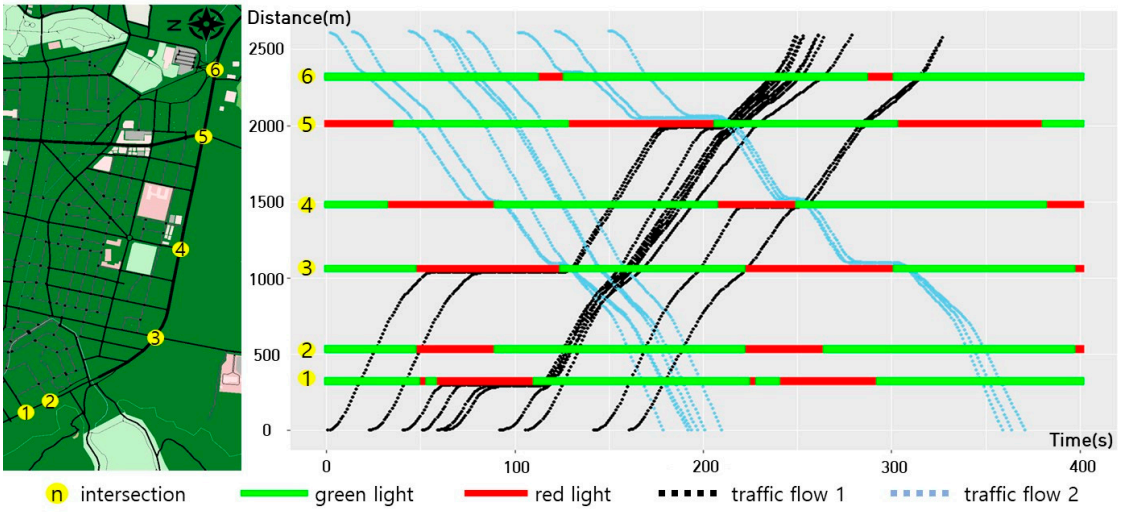
During the entire simulation, the comparison model showed 13% shorter delay and 17% fewer stops than the proposed model. Although the comparison model had the best performance, the proposed model also showed sufficiently ideal results. In addition, the comparative model calculated an irregular signal pattern, while the proposed model calculates a realistic signal pattern. Therefore, the proposed model would be the best in terms of applying it to real-world intersections.



(**a**)

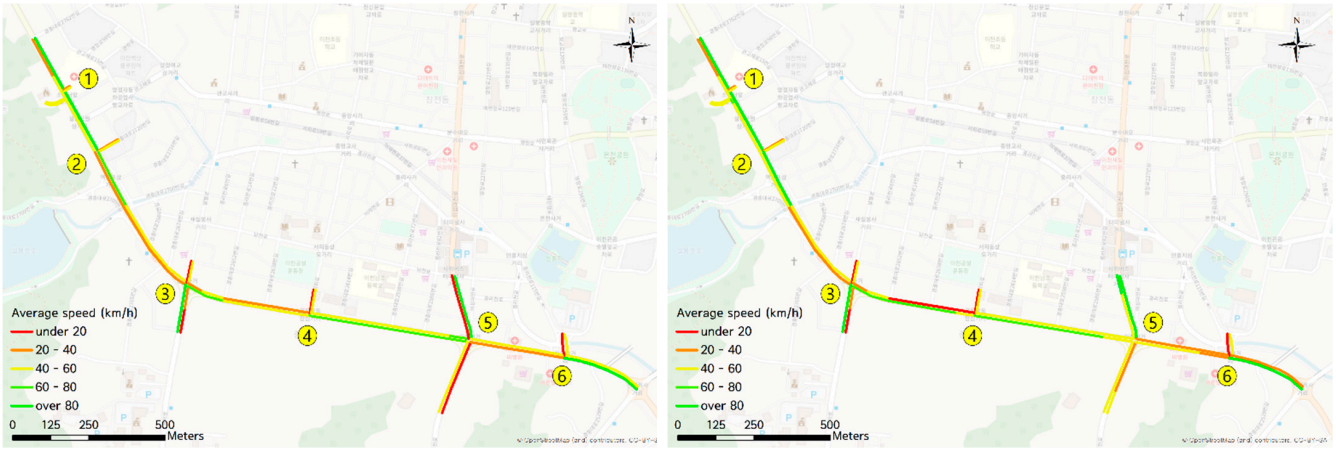


(**b**)



(**c**)

Figure 9. Time-space diagram: (**a**) Fixed-time model; (**b**) Comparison model; and (**c**) Proposed model.



(**a**) (**b**)

Figure 10. Traffic condition: (**a**) Comparison model; (**b**) Proposed model.

5. Conclusions

In this study, traffic signal control based on a reinforcement learning algorithm was proposed to minimize traffic congestion. Early reinforcement learning-based signal control research focused on mathematical optimization, and when the model was applied to the road, excessively waiting vehicles and confused drivers could have occurred. Therefore, this study proposed a reinforcement learning-based traffic signal control model by applying the constraint that fixed the sequence of the pre-planned phases and provided a minimum green time.

Simulations of the proposed model and the comparison model without constraints and fixed-time model were performed in two scenarios. The scenarios included multiple intersections, and the delay and the number of stops were compared. Compared with the fixed-time model, the reinforcement learning-based models showed excellent performance. Although the comparison model showed the best performance, the proposed model also showed ideal results. Unlike the comparison model, the proposed model will show the best performance when applied to real world intersections, because it calculates realistic signal patterns.

Even if the simulation environment is based on reality, implementation will not be exactly the same. Therefore, it is necessary to test the model on real-world roads in future research. To this end, more constraints and data for safe road driving and simulations in various types of road networks will be required.

Author Contributions: Writing—original draft preparation, J.G.; visualization, M.L.; writing—review and editing, C.J.; methodology, Y.H. and J.K.; project administration, Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Road Traffic Authority grant funded by the Korea government [grant number 1325163906].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Han, Y.; Kim, Y. Spatiotemporal congestion recognition index to evaluate performance under oversaturated conditions. *KSCE J. Civil Eng.* **2019**, *23*, 3714–3723. [CrossRef]
2. Yang, S.; Yang, B.; Wong, H.-S.; Kang, Z. Cooperative traffic signal control using multi-step return and off-policy asynchronous advantage actor-critic graph algorithm. *Knowl. Based Syst.* **2019**, *183*, 104855. [CrossRef]
3. Aslani, M.; Mesgari, M.S.; Seipel, S.; Wiering, M. Developing adaptive traffic signal control by actor–critic and direct exploration methods. *Proc. Inst. Civ. Eng. Transp.* **2019**, *172*, 289–298. [CrossRef]
4. Aslani, M.; Stefan, S.; Marco, W. Continuous residual reinforcement learning for traffic signal control optimization. *Can. J. Civ. Eng.* **2018**, *45*, 690–702. [CrossRef]
5. Mannion, P.; Duggan, J.; Howley, E. An experimental review of reinforcement learning algorithms for adaptive traffic signal control. *Auton. Road Transp. Support Syst.* **2016**, *4*, 47–66. [CrossRef]
6. Li, L.; Lv, Y.; Wang, F.-Y. Traffic signal timing via deep reinforcement learning. *IEEE/CAA J. Autom. Sin.* **2016**, *3*, 247–254.
7. Ge, H.; Song, Y.; Wu, C.; Ren, J.; Tan, G. Cooperative deep q-learning with q-value transfer for multi-intersection signal control. *IEEE Access* **2019**, *7*, 40797–40809. [CrossRef]
8. Al Islam, S.B.; Hajbabaie, A. Distributed coordinated signal timing optimization in connected transportation networks. *Transp. Res. Part C Emerg. Technol.* **2017**, *100*, 272–285. [CrossRef]
9. Mousavi, S.S.; Schukat, M.; Howley, E. Traffic light control using deep policy-gradient and value-function-based reinforcement learning. *IET Intell. Transp. Syst.* **2017**, *11*, 417–423. [CrossRef]
10. Rasheed, F.; Yau, K.L.A.; Low, Y.C. Deep reinforcement learning for traffic signal control under disturbances: A case study on Sunway city, Malaysia. *Future Gener. Comput. Syst.* **2020**, *109*, 431–445. [CrossRef]
11. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
12. Touhbi, S.; Babram, M.A.; Nguyen-Huu, T.; Marilleau, N.; Hbid, M.L.; Cambier, C.; Stinckwich, S. Adaptive traffic signal control: Exploring reward definition for reinforcement learning. *Procedia Comput. Sci.* **2017**, *109*, 513–520. [CrossRef]
13. Liang, X.; Du, X.; Wang, G.; Han, Z. Deep reinforcement learning for traffic light control in vehicular networks. *Mach. Learn.* **2018**, *68*, 1–11. [CrossRef]
14. Wang, S.; Xie, X.; Huang, K.; Zeng, J.; Cai, Z. Deep reinforcement learning-based traffic signal control using high-resolution event-based data. *Entropy* **2019**, *21*, 744. [CrossRef]
15. Gong, Y.; Abdel-Aty, M.; Cai, Q.; Rahman, M.S. Decentralized network level adaptive signal control by multi-agent deep reinforcement learning. *Transp. Res. Interdiscip. Perspect.* **2019**, *9*, 10306–10316. [CrossRef]
16. Chu, T.; Wang, J.; Codecà, L.; Li, Z. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 1086–1095. [CrossRef]
17. Egea, A.C.; Howell, S.; Knutins, M.; Connaughton, C. Assessment of Reward Functions for Reinforcement Learning Traffic Signal Control under Real-World Limitations. In Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Totonto, ON, Canada, 11–14 October 2020; pp. 965–972.
18. Ozan, C.; Baskan, O.; Haldenbilen, S.; Ceylan, H. A modified reinforcement learning algorithm for solving coordinated signalized networks. *Transp. Res. Part C Emerg. Technol.* **2015**, *54*, 40–55. [CrossRef]
19. Kim, D.; Jeong, O. Cooperative traffic signal control with traffic flow prediction in multi-intersection. *Sensors* **2020**, *20*, 137. [CrossRef] [PubMed]
20. Yuan, J.; Abdel-Aty, M.; Gong, Y.; Cai, Q. Real-time crash risk prediction using long short-term memory recurrent neural network. *Transp. Res. Rec.* **2019**, *2673*, 314–326. [CrossRef]
21. Zhao, Y.; Liang, Y.; Hu, J.; Zhang, Z. Traffic Signal Control for Isolated Intersection Based on Coordination Game and Pareto Efficiency. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference, Auckland, New Zealand, 27–30 October 2019; pp. 3508–3513. [CrossRef]
22. Kühnel, N.; Theresa, T.; Kai, N. Implementing an adaptive traffic signal control algorithm in an agent-based transport simulation. *Procedia Comput. Sci.* **2018**, *130*, 894–899. [CrossRef]