

Article

# Intelligent Traffic Signal Phase Distribution System Using Deep Q-Network

Hyunjin Joo  and Yujin Lim \* 

Department of IT Engineering, Sookmyung Women's University, Seoul 04310, Korea; hjoo8@sookmyung.ac.kr

\* Correspondence: yujin91@sookmyung.ac.kr

**Abstract:** Traffic congestion is a worsening problem owing to an increase in traffic volume. Traffic congestion increases the driving time and wastes fuel, generating large amounts of fumes and accelerating environmental pollution. Therefore, traffic congestion is an important problem that needs to be addressed. Smart transportation systems manage various traffic problems by utilizing the infrastructure and networks available in smart cities. The traffic signal control system used in smart transportation analyzes and controls traffic flow in real time. Thus, traffic congestion can be effectively alleviated. We conducted preliminary experiments to analyze the effects of throughput, queue length, and waiting time on the system performance according to the signal allocation techniques. Based on the results of the preliminary experiment, the standard deviation of the queue length is interpreted as an important factor in an order allocation technique. A smart traffic signal control system using a deep Q-network, which is a type of reinforcement learning, is proposed. The proposed algorithm determines the optimal order of a green signal. The goal of the proposed algorithm is to maximize the throughput and efficiently distribute the signals by considering the throughput and standard deviation of the queue length as reward parameters.

**Keywords:** intelligent traffic signal control; reinforcement learning; deep Q-network; multi-intersection; throughput



check for updates

**Citation:** Joo, H.; Lim, Y. Intelligent Traffic Signal Phase Distribution System Using Deep Q-Network. *Appl. Sci.* **2022**, *12*, 425. <https://doi.org/10.3390/app12010425>

Academic Editor: Andrea Prati

Received: 30 November 2021

Accepted: 31 December 2021

Published: 3 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Traffic congestion is a common phenomenon that occurs when a signal is shorter than the number of vehicles attempting to pass. Owing to the increasing number of vehicles, traffic congestion is worsening and is one of the main problems that cities are trying to solve. In the era of the 4th Industrial Revolution, research into smart cities using information communication and Internet of Things (IoT) technologies is being actively conducted. The urban management system of a smart city includes various fields, such as medical, transportation, safety, and energy. The platform that deals with traffic-related problems in smart cities is an intelligent transportation management system in which traffic flow is controlled to reduce traffic congestion. The current traffic signal control is a fixed signal control.

In a traditional fixed signaling system, a green signal is assigned in the order in which it is input in advance. The system also distributes green signals of the same length, and the process is repeated even under various traffic conditions. However, traffic conditions dynamically and unexpectedly change over the time and spatial domains. There are times when more vehicles are moving, such as during rush-hour or the weekend, and times when relatively fewer vehicles are moving. As such, traffic signal systems have complex nonlinear stochastic systems because they are affected by various factors, such as traffic infrastructure and traffic load. Therefore, traffic signal systems have been previously studied for adaptive traffic signal control.

Among the various studies on the traffic signal control system, there are studies using fuzzy technology [1,2] as a type of CI technology, green wave technology [3], and traffic signal control using particle swarm optimization [4] as a heuristic technology. However,

these systems require many resources, such as time or computing power, to calculate the optimal signaling strategy. Because of these resource constraints, the system can only be used on environments and rules that have already been calculated, and there is a limit to optimally control the traffic signal in a dynamic environment that changes in real time. The traffic control system has to make optimal decisions by adapting to the dynamic environment while recognizing the surrounding intersections. Therefore, research on controlling traffic signals using reinforcement learning (RL) has recently attracted attention [5,6].

In this paper, we propose a traffic signal system using a deep Q-network (DQN), which is a type of reinforcement learning method. The proposed model recognizes traffic conditions at intersections and determines the phase order, that is, the order in which the green traffic light turns on. The purpose of the proposed model is to maximize the throughput of the intersection and distribute the signals fairly. Throughput refers to the number of vehicles passing through an intersection over time. The reason for the fair distribution of traffic signals is to avoid the risk of some vehicles waiting too long to increase the throughput.

The remainder of this paper is organized as follows. Section 2 introduces related studies. Section 3 proposes a traffic signal control model using a DQN. Section 4 presents and discusses the experiment results. Finally, Section 5 provides some concluding remarks regarding this research.

## 2. Related Works

Recently, reinforcement learning method is widely applied to traffic signal control and optimization to model and adapt to the dynamic traffic environment. Traffic conditions are unpredictable. Therefore, the traffic signal control model should be able to improve itself through learning. Reinforcement learning proceeds with learning based on past experiences and selects the optimal behavior to achieve its goal. Applying reinforcement learning to traffic signal control research enables flexible traffic flow management in unpredictable and dynamic environments.

There are two main ways through which a traffic signaling system controls the flow of vehicles at an intersection: determining the direction in which the vehicles will proceed [7–10] or determining the duration of the green light [11–15]. Traditional traffic signals have a predetermined phase for each intersection. Here, the phase indicates the direction in which the green signal is assigned. In the method of determining the preceding signal, the phase applied to receive the green signal is flexibly determined. Depending on the situation, the vehicle in the most necessary phase direction passes through the intersection. Meanwhile, another method determines the optimal duration of a green light phase for receiving the next green signal in a predetermined phase order. In [12], there are only two phases, and thus it is determined whether to continue or change the signal. This enables an efficient signal time distribution.

The model proposed in [7], which determines the phase order, minimizes the aggregation of the total time spent in all vehicle queues at a single intersection using a transfer RL. In [8], a vehicle is controlled at multiple intersections with the aim of minimizing the queue length. A Q-value transfer strategy using the DQN was applied to reflect on the status of the surrounding intersections. The Q-value of the surrounding intersection is reflected in the Q-value update of the target intersection to recognize the status of the surrounding intersection. The studies in both [9,10] proposed traffic signal control methods to minimize the wait time at multiple intersections. In [9], the road network features are extracted using GCNN (graph convolutional neural network), and the model is optimized by using a neural fitted Q-iteration. The total travel time is considered as a reward parameter. In [10], the proposed DQN model can be extended from a single agent to multiple agents through transfer learning. Various factors such as travel time, waiting time, delay, and emergency stop are considered reward parameters. Therefore, it is difficult to apply in a real environment. In [11], to determine the duration of a green signal, the traffic signal control system is proposed by using a deep recurrent Q-network technique that combines a recurrent neural

network (RNN) and a DQN. In [11], the action of the model determines whether to extend the duration of the current green phase or not. The total number of the halting vehicle is considered as a reward parameter. The aim is to minimize the total number of waited vehicles by establishing a Q-network with long short-term memory, which is a type of RNN. However, because it is based on a single intersection environment, it is necessary to expand to a multi-intersection environment for optimized signal control. In [12], in order to accelerate the learning, a simple RL model was proposed by using the average length of the queue as a reward parameter.

In [13], the authors controlled the signal to minimize the average wait time of the vehicles using RL. It provides the Q-value of the target agent to the surrounding agents. There are two traffic signals consisting of a green signal and a red signal. When the lane is allocated a green signal, the opposite lane is allocated a red signal. Accordingly, the traffic signal determines the duration of the green and red signals. In [14,15], the authors proposed a traffic signal that determines the duration of a green signal. In [14], the wait time was considered as a reward parameter to reduce the travel time and increase the speed of the vehicles. In [15], the queue length was considered as a reward parameter to optimize the traffic flow. In addition, an excessive allocation of the green time in a specific moving direction can cause disadvantages for the waiting vehicles in different directions.

We propose a traffic signal control technique that determines the order of the preceding signals. In addition, the situations of the surrounding intersections are considered in a multi-intersection environment. The proposed model aims to maximize the throughput and fairly allocate the green signals.

### 3. Proposed Model

To identify the performance of the traffic signal system, there are three main factors at the intersection: throughput, queue length, and waiting time. Throughput represents the number of vehicles exiting from the intersection over a period of time. The traffic signal system needs to be optimized to process as many vehicles as possible during a period of time. The queue length refers to the number of vehicles waiting at an intersection. Finally, the waiting time represents the time taken to exit the intersection after the vehicle arrives at the intersection.

Usually, the average of queue length or waiting time are used to determine the situation of the system. However, the average value can make it difficult to accurately analyze the situation at an intersection. This is because the average is the intermediate between the best and worst values. For example, if we focus only on reducing the length of the longest queue to lower the average value of the queue length at the intersection, the progress of vehicles in a short queue may be restricted. However, the standard deviation is a measure of the degree of variance of data to the mean. The standard deviation of the queue length represents information about the deviation between the lane where many vehicles wait within the intersection and the lane where fewer vehicles wait. Therefore, it is effective to compare and control the situation of each lane.

The queue length and waiting time indicate the intersection. A long queue length indicates that many vehicles are waiting at the intersection. In addition, a large waiting time means that there are vehicles that wait for a long time. The queue length refers to the quantitative information of the vehicles, and the wait time is the temporal information of the vehicles. Therefore, depending on the signaling method, more meaningful factors differ. Thus, an experiment was conducted to determine which factors are important according to the signaling technique. A single intersection was assumed in the experimental environment. The experiments were conducted with the following reward parameters: throughput (tp), the combination of throughput and average queue length (tp + avgql), the combination of throughput and standard deviation of the queue length (tp + qlstd), the combination of throughput and average wait time (tp + avgwt), and the combination of throughput and standard deviation of the wait time (tp + wtstd).

Figures 1 and 2 show the experiment results of the phase order allocation techniques. As shown in Figure 1, the technique of combining throughput and queue length-related

elements, which are average queue length or standard deviation of queue length, has a smaller average queue length performance value than other techniques. In step 3, the technique of combining throughput and standard deviation of queue length shows about 20% shorter queue length than the combining technique of throughput and the standard deviation of wait time. In Figure 2, the average wait time is about 35% lower with the technique of combining throughput and queue length-related factors than when considering only throughput. It is also about 22% smaller than the performance of the wait time-related factor combination technique.

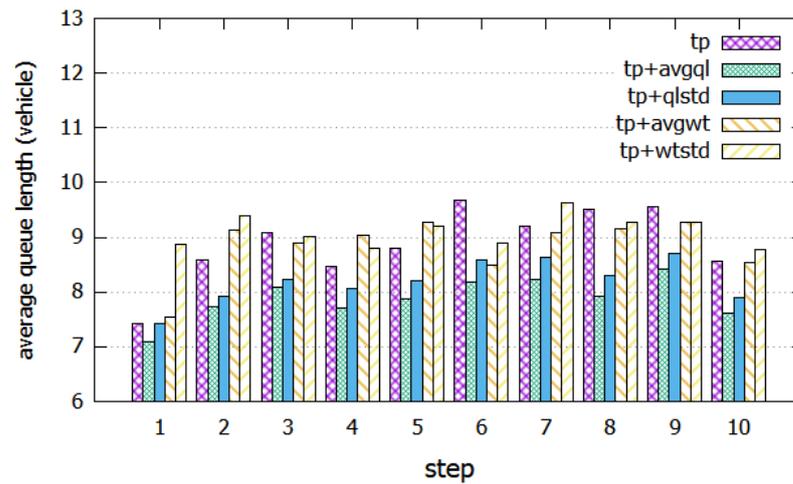


Figure 1. Performance of a technique for determining the phase order in terms of the average queue length.

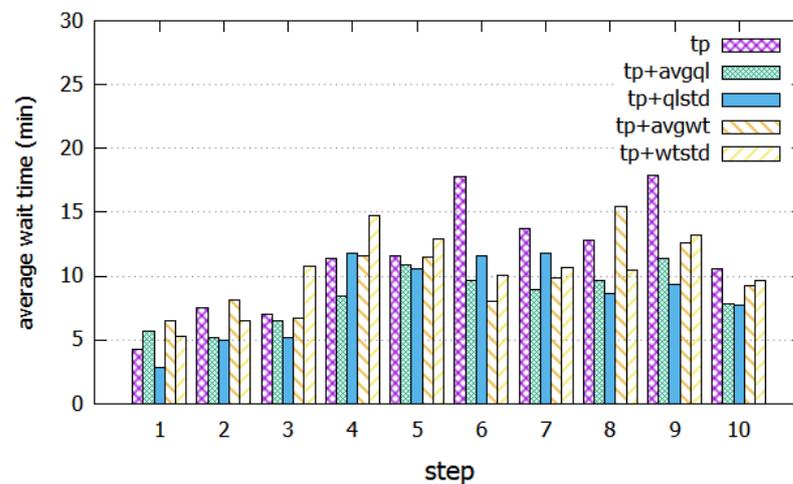


Figure 2. Performance of a technique for determining the phase order in terms of the average wait time.

Figures 3 and 4 show the experiment results of the allocation technique of the green light duration. As shown in Figure 3, the average queue length is similar in all techniques. In Figure 4, by contrast, the average wait time resulted in vehicles waiting less when the wait time-related factor was applied to the reward parameter. On average, the combination of throughput and wait time-related factors has about 35% less wait time than the combination of throughput and queue length-related factors.

In the case of a signal control using an order allocation technique, the direction to which the next green signal is allocated is determined. The queue length factor is more meaningful than the wait time factor in allowing more vehicles to exit during a period of time. Conversely, in the case of signal control using the time duration allocation technique,

the direction that receives the next green signal is already fixed and determines how much time is allocated. Therefore, the wait time factor with the time information of the vehicle is more meaningful. In this paper, a traffic signal control using an order allocation technique is proposed.

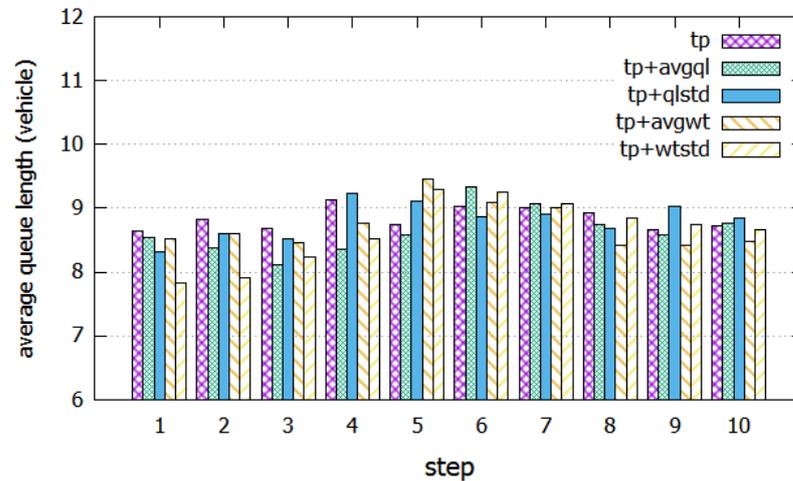


Figure 3. Performance of a technique for determining the green light duration of a phase in terms of the average queue length.

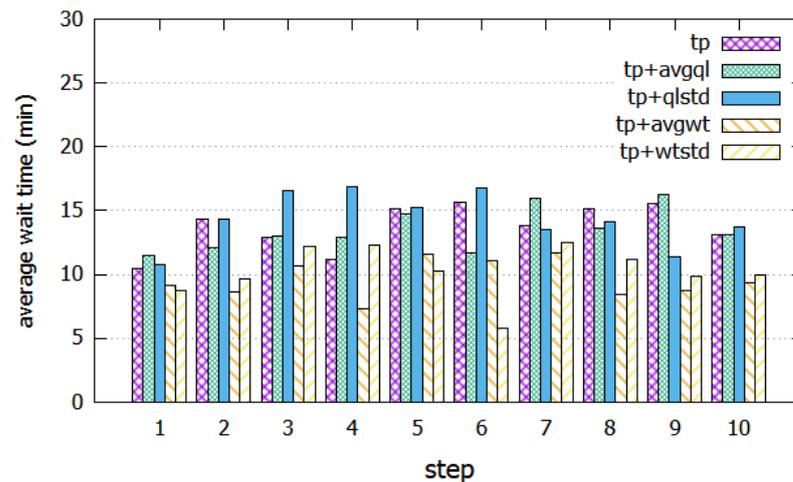


Figure 4. Performance of a technique for determining green light duration of a phase in terms of the average wait time.

When the order allocation technique is used, both in terms of queue length and wait time, it shows good performance when throughput and queue length-related factors are applied as reward parameters in the learning model. We selected the standard deviation of the queue length among the related factors of the queue length. The reason is to reduce the number of vehicles that have to wait for a relatively long time in order to increase the throughput of the intersection. In order to increase the throughput of the intersection, it is advantageous to assign green lights to the lanes with relatively long queues. However, in this case, the drivers of the lanes with a relatively short queue have fewer opportunities to proceed.

We propose an intelligent traffic-signal control system that determines the phase order. It recognizes the situation and determines the lane to receive the green signal depending on the situation. The intersection environment is dynamic, and the traffic volume changes in real-time. Therefore, traffic conditions always change and are difficult to predict. Therefore, we attempt to recognize the traffic condition and learn the optimal policy to determine

directions for receiving green signals using a DQN, which is a reinforcement learning method. The Markov decision process (MDP) of a model with an intersection environment is defined as follows:

3.1. State and Action

The proposed model is a traffic signal control system that controls the vehicles by optimizing the order in which a green light occurs at multiple intersections. Because multiple intersections are involved, it is necessary to recognize the status of the surrounding intersections. According to Figure 5, the lane entering the intersection is defined as  $l_n$ . Therefore, there are a total of eight incoming lanes, from  $l_1$  to  $l_8$ , at a 4-way intersection. There are also eight lanes in the direction leaving the intersection, from  $n_1$  to  $n_8$ .

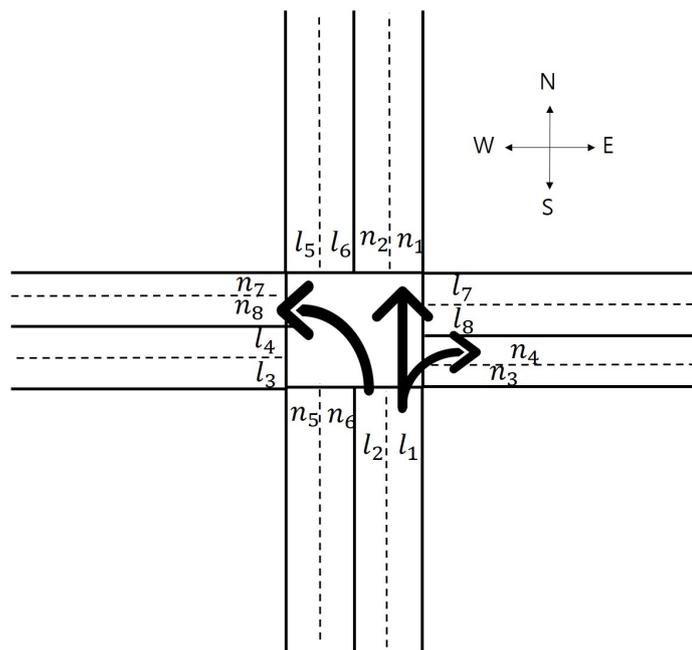


Figure 5. A 4-way target intersection.

The information on the target intersection is determined based on the number of vehicles currently waiting at the intersection. That is, the queue length information ( $ql$ ) in each lane of the target intersection is required. In Figure 5, the queue length of lane  $l_1$  is defined as  $ql_1$ . In addition, the situation of the surrounding intersection is determined by the traffic load in each lane of the surrounding intersection, where the vehicle exits from the target intersection. Traffic load information of the surrounding intersection is required for a vehicle to exit the target intersection. This is because the number of vehicles that exit from the target intersection should be less than the number of vehicles that can be accommodated by the surrounding intersection. The lane in the direction of the exit from the target intersection is the incoming lane of the surrounding intersection, and the traffic load of lane  $n_1$  is defined as  $tl_1$ . As shown in Figure 5, there are two directions on the road: one is a straight lane capable of turning right and the other is a left lane. Thus, state ( $s_t$ ) is defined as in (1):

$$s_t = [ql_1, ql_2, ql_3, ql_4, ql_5, ql_6, ql_7, ql_8, tl_1, tl_2, tl_3, tl_4, tl_5, tl_6, tl_7, tl_8]. \tag{1}$$

The proposed model determines the order of the green signal based on the intersection. The possible movement signal at the 4-way intersection is shown in Figure 6. The action of the MDP indicates the phase of the green signal. The phase of the intersection can vary

depending on the structure of the intersection. At a 4-way intersection, the action ( $a_t$ ) is defined as in (2).

$$a_t = [p_1, p_2, p_3, p_4] \tag{2}$$

$p_1$	$p_2$	$p_3$	$p_4$

Figure 6. The possible movement signal phase in a 4-way intersection.

### 3.2. Reward

The proposed model aims to maximize the throughput ( $\tau$ ) of an intersection. In addition, the model intends to distribute signals that are not selfish. Therefore, throughput is considered as a reward parameter. In addition, to avoid a selfish signal allocation, the signal is distributed within the intersection in consideration of the standard deviation of the queue length ( $d_{ql}$ ). The throughput and standard deviation of the queue length are adjusted with an adaptive weight factor ( $\alpha$ ). The reward ( $r_t$ ) formula is as indicated in (3):

$$r_t = \alpha \cdot \tau - ((1 - \alpha) \cdot d_{ql}) \tag{3}$$

Figure 7 is a framework of the proposed model. A framework shows the interaction between the environment and the agent. The environment represents multi-intersections. The perceived state, which is the queue length of incoming lanes at the target intersection and the traffic load of the exit lane from the target intersection, is sent from the environment to the agent. Then, the agent learns over the DNN using the received information and determines action in the next state. The action represents the phase order of traffic lights at the intersection. The determined action is sent to the environment. Therefore, according to the designated phase order, the green signal at the intersection is turned on.

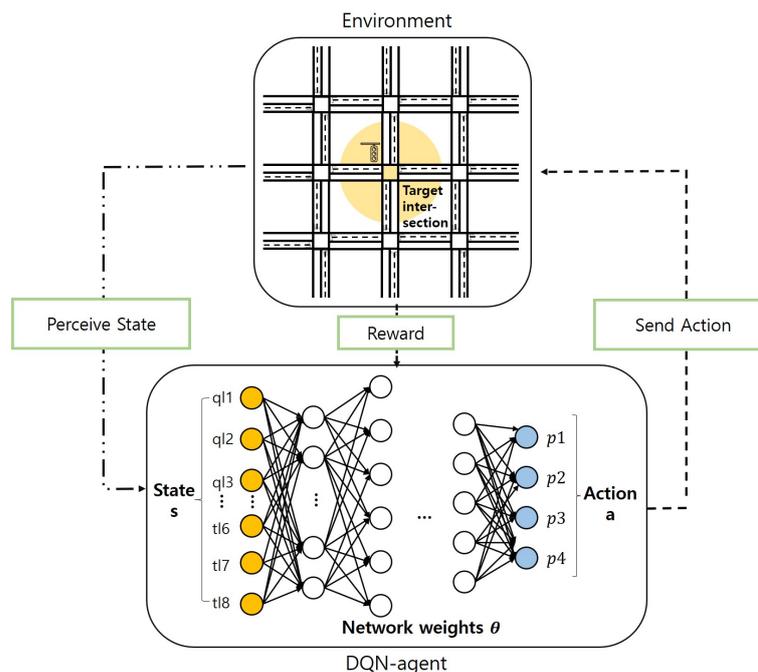
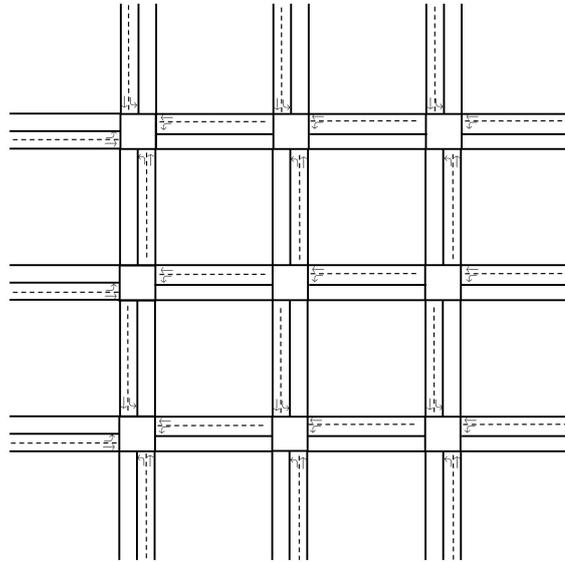


Figure 7. A framework of the proposed model.

#### 4. Performance Evaluation

The performance experiment conducted on the proposed intelligent traffic signal control system was conducted at nine intersections having a 4-way structure, as shown in Figure 8. Vehicles can move in two directions on each road. One can turn left, and the other can continue straight or turn right. The duration of the green signal was set to 30 s. This experiment was conducted to simulate urban mobility (SUMO).



**Figure 8.** Structure of the experiment environment: nine intersections.

To evaluate the performance of the proposed algorithm, the model classifies the reward parameters into two categories: the combination of throughput and queue length, and the combination of throughput and standard deviation of the queue length. In addition, the performance is evaluated by comparing fixed signal control as a baseline, which is a traditional traffic signal control technique, and comparing two other RL techniques [8,16]. One comparison algorithm is referred to as QT-CDQN [8]. QT-CDQN considers the average queue length as a reward parameter to propose an order-allocation technique, such as the proposed algorithm. The other comparison algorithm is referred to as CRL which also uses an order-allocation technique [16]. CRL considers the difference between the sum of queue length of time  $t$  and  $t + 1$  as a reward parameter

The vehicle distribution data used in the experiment were generated using SUMO. Based on the real data of the Seoul area [17], the SUMO simulator was used to generate vehicle data suitable for our experimental environment. The vehicle route data are set to a repetition rate, period, and arrival rate. Then, based on this, the SUMO simulator determines the time and route of the vehicle entering the road. Accordingly, the vehicle is input into the road network according to a preset start time and route. Figure 9 shows the distribution of the number of vehicles entering an intersection on weekdays. The distribution shows a large amount of traffic during rush hour. The experiments were conducted using vehicle distribution. One step is represented as one hour.

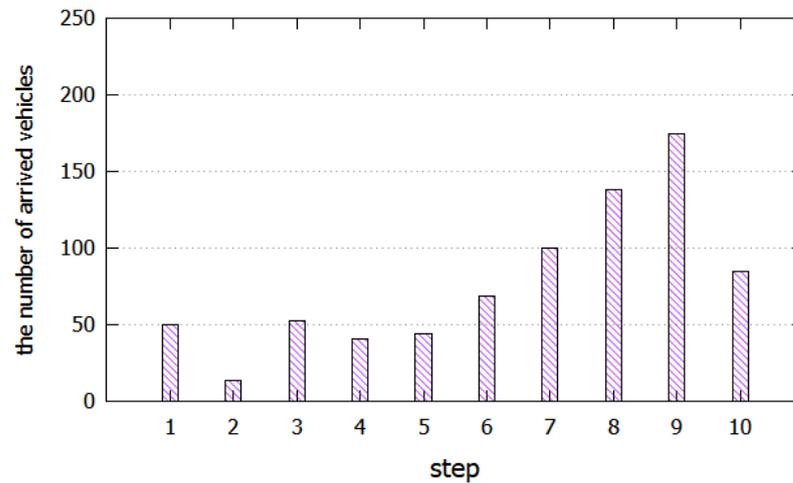


Figure 9. Traffic distribution used in the experiment.

In Figure 10, the performance in terms of average queue length is compared. Step 1 to 5 refer to relatively light traffic load, and step 6 to 10 refer to relatively heavy traffic load. In the case of heavy traffic load,  $tp + qlstd$  has an average queue length about 20% lower than QT-CDQN and about 48% lower than CRL. Moreover,  $tp + qlstd$  is about 30% lower than  $tp + avgql$ . This shows that  $tp + qlstd$  appropriately distributes traffic signals in consideration of the standard deviation of the queue length in the intersection.

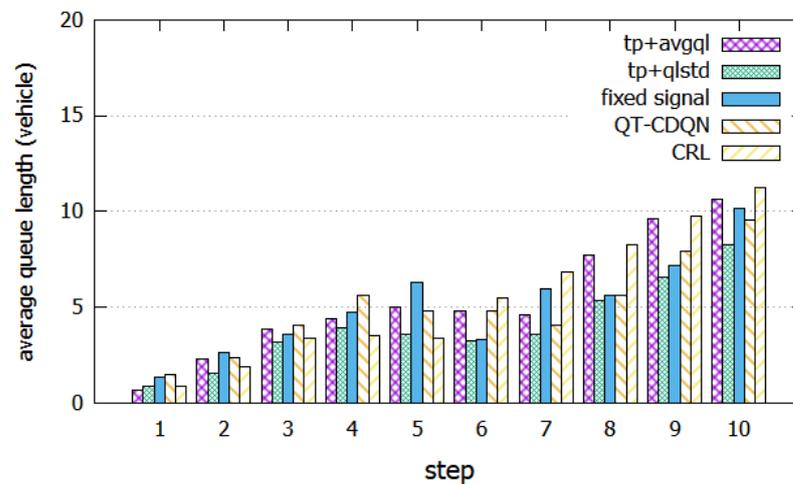


Figure 10. Performance comparison: average queue length.

Figure 11 shows the comparison of the average wait times. For heavy traffic load, QT-CDQN has a similar wait time on average with that of  $tp + qlstd$ . On the other hand,  $tp + qlstd$  has about 20% shorter wait time than that of CRL. As a reward parameter, QT-CDQN applied the average queue length, CRL applied on the sum of queue lengths, and  $tp + qlstd$  applied the standard deviation of the queue length. This shows that the same factor, the queue length, was applied, but the performance varies depending on how it is applied.

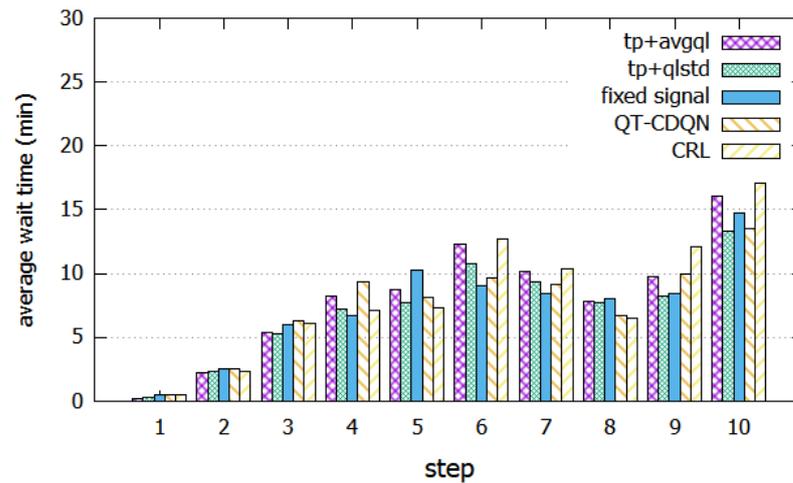


Figure 11. Performance comparison: average wait time.

Figures 12 and 13 show the performance comparison using the standard deviation of performance indicators. In the case of the standard deviation of the queue length, it was the lowest when the standard deviation of the queue length and throughput were considered as the reward parameters. Considering the standard deviation of the queue length, it is interpreted that the queue lengths of the lanes within the intersection are evenly adjusted by efficiently distributing the green signals. By contrast, in QT-CDQN, the standard deviation of the queue length was approximately 35% higher on average than the standard deviation of the combination of throughput and queue length. In addition, in CRL, the standard deviation of the queue length was also about 16% higher on average than that of  $tp + qlstd$ . Because QT-CDQN did not consider the standard deviation of the queue length, it is interpreted that the order of green signals to lanes could not be evenly adjusted.

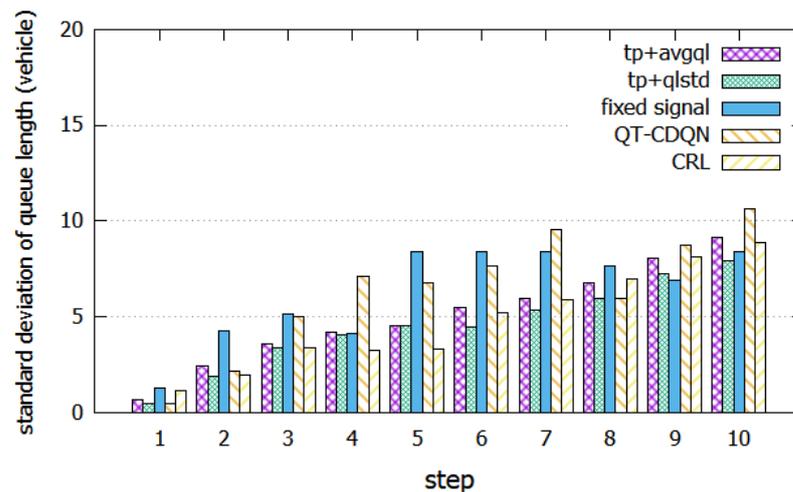
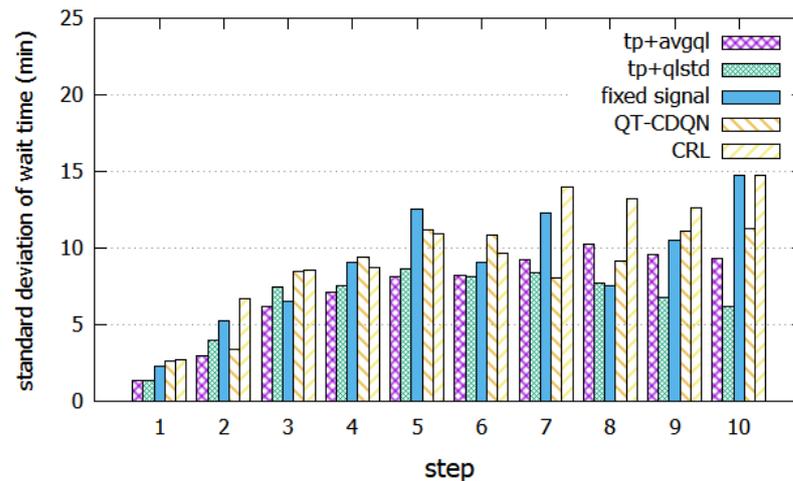


Figure 12. Performance comparison: standard deviation of queue length.

In the standard deviation of the wait time, the performance of the fixed signal is about 40% greater than  $tp + qlstd$ . This denotes that, in the case of a fixed signal, the deviation of vehicles waiting at the intersection is large and signal control is inefficient. In addition, CRL also has a value about 30% greater than  $tp + qlstd$ . The reason is that the difference in the sum of the queue lengths of  $t$  and  $t + 1$  considered by the CRL does not take into account the deviation of the lanes. When the traffic load is heavy,  $tp + avgql$  has a standard deviation of wait time about 20% higher than  $tp + stdql$ . This shows the performance that, considering the standard deviation of the queue length, the deviation of the intersection can be kept small.



**Figure 13.** Performance comparison: standard deviation of wait time.

## 5. Conclusions

We proposed a DQN-based intelligent signal control system to determine the order of green signals at multiple intersections. The purpose of this study is to distribute signals efficiently while maximizing the throughput at the intersections. Through pre-experiments, it was confirmed that the important factors affecting the performance differ depending on the traffic signaling method. By applying various parameters to the proposed model, the performance was compared with that of other signal control methods. This study classified two reward parameters, average queue length and standard deviation of the queue length, to evaluate the performance and compared them with the fixed-time signal, QT-CDQN [8] and CRL [16]. In terms of the wait time, the proposed model performs approximately 30% better when applying the standard deviation of the queue length than when applying the average queue length. In addition, considering the standard deviation of the queue length, it is interpreted that the length of the lanes within the intersection is similarly adjusted by efficiently distributing the green signals.

Because we proposed a traffic signal control method applied at multiple intersections, large-scale signal control studies, including in urban areas by expanding the range, also need to be conducted. In addition, it will be necessary to expand the research area by applying federated learning methods for faster learning with more diverse data [18].

**Author Contributions:** Conceptualization, H.J. and Y.L.; Methodology, H.J.; Software, H.J.; Writing—Review & Editing, H.J. and Y.L.; Supervision, Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2021R1F1A1047113).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing not applicable.

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

## References

1. Lin, Q.; Kwan, B.; Tung, L. Traffic signal control using fuzzy logic. In Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, Orlando, FL, USA, 12–15 October 1997; Volume 2, pp. 1644–1649.
2. Hu, Y.; Thomas, P.; Stonier, R.J. Traffic signal control using fuzzy logic and evolutionary algorithms. In Proceedings of the 2007 IEEE Congress on Evolutionary Computation, Singapore, 25–28 September 2007; pp. 1785–1792.
3. Wu, X.; Deng, S.; Du, X.; Ma, J. Green-wave traffic theory optimization and analysis. *World J. Eng. Technol.* **2014**, *2*, 14–19. [[CrossRef](#)]

4. Celtek, S.A.; Durdu, A.; Alı, M.E.M. Real-time traffic signal control with swarm optimization methods. *Measurement* **2020**, *166*, 108206. [CrossRef]
5. Wei, H.; Zheng, G.; Gayah, V.; Li, Z. A survey on traffic signal control methods. *arXiv* **2019**, arXiv:1904.08117.
6. Yau, K.L.A.; Qadir, J.; Khoo, H.L.; Ling, M.H.; Komisarczuk, P. A survey on reinforcement learning models and algorithms for traffic signal control. *ACM Comput. Surv. (CSUR)* **2017**, *50*, 1–38. [CrossRef]
7. Shabestary, S.M.A.; Abdulhai, B. Deep learning vs. discrete reinforcement learning for adaptive traffic signal control. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 286–293.
8. 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]
9. Nishi, T.; Otaki, K.; Hayakawa, K.; Yoshimura, T. Traffic signal control based on reinforcement learning with graph convolutional neural nets. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 877–883.
10. Van der Pol, E.; Oliehoek, F.A. Coordinated deep reinforcement learners for traffic light control. In Proceedings of the Learning, Inference and Control of Multi-Agent Systems (at NIPS 2016), Barcelona, Spain, 5–10 December 2016.
11. Zeng, J.; Hu, J.; Zhang, Y. Adaptive traffic signal control with deep recurrent Q-learning. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1215–1220.
12. Zheng, G.; Zang, X.; Xu, N.; Wei, H.; Yu, Z.; Gayah, V.; Xu, K.; Li, Z. Diagnosing reinforcement learning for traffic signal control. *arXiv* **2019**, arXiv:1905.04716.
13. Su, S.; Tham, C.K. SensorGrid for real-time traffic management. In Proceedings of the 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, Melbourne, VIC, Australia, 3–6 December 2007; pp. 443–448.
14. Balaji, P.; German, X.; Srinivasan, D. Urban traffic signal control using reinforcement learning agents. *IET Intell. Transp. Syst.* **2010**, *4*, 177–188. [CrossRef]
15. Chin, Y.K.; Kow, W.Y.; Khong, W.L.; Tan, M.K.; Teo, K.T.K. Q-learning traffic signal optimization within multiple intersections traffic network. In Proceedings of the 2012 Sixth UKSim/AMSS European Symposium on Computer Modeling and Simulation, Malta, Malta, 14–16 November 2012; pp. 343–348.
16. Guo, M.; Wang, P.; Chan, C.Y.; Askary, S. A reinforcement learning approach for intelligent traffic signal control at urban intersections. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 4242–4247.
17. Public Data Portal. 2016. Available online: <http://data.seoul.go.kr/dataList/datasetView.do?infId=OA-13314&srvType=A&serviceKind=1&currentPageNo=1/> (accessed on 28 October 2016).
18. Wang, T.; Liang, T.; Li, J.; Zhang, W.; Zhang, Y.; Lin, Y. Adaptive Traffic Signal Control Using Distributed MARL and Federated Learning. In Proceedings of the 2020 IEEE 20th International Conference on Communication Technology (ICCT), Nanning, China, 28–31 October 2020; pp. 1242–1248.