International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Review of Deep Reinforcement Learning for Traffic Signal Control

Mahendralal Prajapati[1], Alok Kumar Upadhyay[2], Dr. Harshali Patil[3], Dr. Jyotshna Dongradive[4]

[1,2]Student, Department of Computer Science, University of Mumbai
[3]Associate Professor, Department of Computer Science, MET Institute of Computer Science
[4]Associate Professor, Department of Computer Science, University of Mumbai

Abstract
Traffic signal control plays a vital role in effectively managing traffic flow and alleviating congestion in urban areas. Traditional methods for controlling traffic signals often rely on fixed timing plans or predefined algorithms, which may not be adaptable to changing traffic conditions. Reinforcement Learning is gaining traction as a favored data-centric method for adapting traffic signal control in intricate urban traffic networks. This article represents a conceptual review of recent studies and techniques that showcase the effectiveness of Deep Reinforcement Learning (DRL) in enhancing the performance of traffic signal control. These improvements include reducing travel time, fuel consumption, and emissions. Additionally, we will delve into different algorithms and learning systems explored in research papers, such as multi-agent reinforcement learning and Deep Q Networks (DQN).

Keywords: Deep reinforcement learning, Deep Q-Network (DQN), Intelligent traffic-control system, Adaptive traffic signal control, multi-agent reinforcement learning, Artificial intelligence.

1. Introduction

Due to the rapid growth in the population, there is an increase in the demand for vehicles and public transportation. The Ongoing traffic light control framework conveys a fixed-time traffic signal change disregarding the continuous traffic conditions. Which results in more traffic jams and fuel wastage at traffic signals. And leads to more road accidents. Also, creates harmful emissions which are not good for the health of humans. During congestion, vehicles decelerate or halt. Which increases the waiting time for vehicles. Traffic jams are a well-established significant issue with basic outcomes at the ecological, Social, and financial levels [2].

Traffic jams have long been a vital factor in urban development, but it has now evolved into a significant issue that requires immediate attention, as the surge in vehicle numbers and transportation demand exacerbates the problem. The reason behind the traffic jams is the number of vehicles which are leaving the traffic signal less than the number of vehicles which are coming towards the traffic signal. Another reason behind the traffic jams is when people are in a hurry to leave the intersection, they cross-block by moving the vehicle to another lane.

To deal with these kinds of problems Deep Reinforcement learning is used. The use of deep learning in RL characterized another field called deep reinforcement learning (DRL) [30]. Deep learning is utilized

for feature extraction and has an altogether further developed execution since this technique can learn with more point-by-point data [32]. In reinforcement learning, the agent starts to learn when the traffic is heaviest, which direction through which vehicles are coming, and how quickly they pass through the intersection. Then they adapted accordingly and continued to test and maximize the reward and minimize the average waiting time [2], [15]. The most important part of RL Which results in fewer traffic jams.

In recent years, DRL has achieved many successes in the world of artificial intelligence. DRL is used where situations are computationally complex. DRL methods have been well applied in the traffic signal control of single- intersections and have shown a better performance than traditional methods [8], [9]. Recent works began to try to apply DRL algorithms, Such as multi-agent Reinforcement learning (MARL) [1], [3], [10], [18], [29], [30], Deep Q-Learning (DQN) [4], [6], [8], [19], [27]. These algorithms have gained popularity. The advantages of Deep Reinforcement Learning (DRL) in Traffic Signal Control are as follows:

- The DRL algorithm can improve traffic signal control by adjusting the timings of signals according to the traffic conditions, in real-time.
- DRL can help to minimize the vehicle's ideal waiting times and unusual stops, by optimizing traffic signal timing. Which leads to reduced fuel emissions and consumption. This can contribute to a more environmentally friendly transportation system.
- Traditional traffic signal control systems often require manual adjustments and maintenance, which are more time-consuming and costly. On the other hand, DRL-based traffic signals are once trained, autonomously optimize signal timing without human intervention, which reduces operational costs and improves resource allocation efficiency.

While reviewing other research papers found that there are two common approaches Multi-agent reinforcement learning and Deep Q- Network.

Section I is an introduction. Section II is a literature review and section III is a rules and design consideration of the utilization of DRL for TSC. Section IV will be the conclusion of our review.

TABLE I. ABBREVIATIONS

DRL	Deep Reinforcement Learning
RL	Reinforcement Learning
ATSC	Adaptive Traffic Signal Control
DQN	Deep Q - Network
MARL	Multi-Agent Reinforcement Learning
TSC	Traffic Signal Control
A2C	Advantage-Actor-Critic
IA2C	Independent Advantage-Actor-Critic
MA2C	Multi-Agent Advantage-Actor-Critic
3DQN	Double Dueling Deep Q - Network
PG	Policy Gradient
DDPG	Deep Deterministic Policy Gradient

In this article, we present an overview of RL and DRL. Additionally, we further explain DQN and MARL.

1.1 Reinforcement Learning

Reinforcement learning, a branch of artificial intelligence (AI), differs from supervised and unsupervised learning. It represents a machine-learning technique that holds significant promise in problem-solving. The primary aim of RL is for an agent to acquire an optimal policy in order to interact effectively with a complex environment and ultimately achieve the most rewarding outcome over the long term [11], [15].Reinforcement learning methods to improve urban traffic flows and reduce travel time and fuel emission [13]. RL-based methods are proposed such as DynamicLight [14], DenseLight [16], MetaLight [20], and PressLight [37]. These methods aim to reduce average travel time. The reinforcement learning model is well defined by the tuple (S, A, R, S`) [8], [15],[23],[26] which has the following meaning:

1. S: A feasible group of state space (e.g. queue length, waiting time), s is a specific space (s ϵ S).
2. A: A feasible group of action space, a is an action (a ϵ A).
3. R: A reward is what an agent receives after performing an action.
4. S`: Next state after taking action (a ϵ A) on the state (s ϵ S).

The S2A policy embodies a sequence of indirect measures in reinforcement learning. Its ultimate goal is to uncover an optimal policy that maximizes the cumulative predicted rewards derived from the current state. Typically, the agent progressively moves from an individual state s by taking action a to arrive at state s'. The corresponding reward received, denoted as r, defines the impact of the (s, a, r, s'). Let t represent a t^{th} step in the policy. The cumulative future reward from selecting action a at state s can be expressed by the equation Q(s, a),

$$\mathrm{Q}^{\pi}(\mathrm{s},\mathrm{a}) = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+1} + \cdots | s_t = s, a_t = a, \pi]$$

$$= E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s, a_t = a, \pi\right]. \qquad (1)$$

In the above equation, where γ is the rebate factor, which is typically in $\gamma\epsilon[0,1)$ [7],[8],[26]. It means that the rewards nearest will be worth more than the rewards further away.

The optimal policy π^*can be attained through a recursive approach. When the agent encounters gaining states, the optimal strategy is simply to choose the action that yields the highest reward. Consequently, the optimal $Q(s,a)$ is determined by establishing the optimal Q values for subsequent states. This can be formulated using the Bellman optimality equation $Q^{\pi^*}(s,a)$,

$$Q^{\pi^*}(s,a) = E_{s'}\left[r_t + \gamma \max_{a'} Q^{\pi^*}(s,a) \middle| s,a\right]. \qquad (2)$$

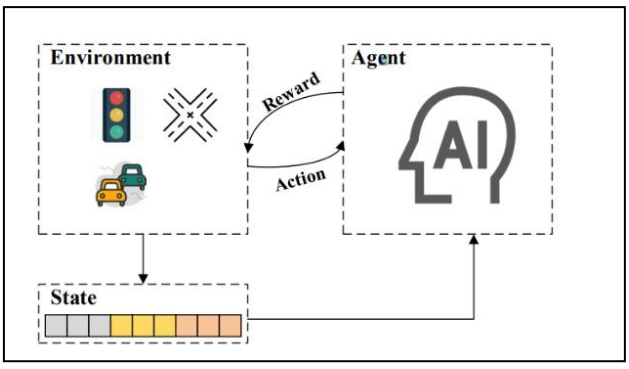

Fig. 1. Basic framework of RL for TSC [22]

1.1.1 Value-based Methods

These approaches rely on a value function to estimate the anticipated reward of every action in a given state. The agent subsequently selects the action that maximizes this value function. Take DQN as an instance, which employs a deep neural network to approximate the Q-function, serving as the value function for the optimal policy. For example, DDQN.

1.1.2 Policy Gradient Methods

PG methods are a class of reinforcement learning algorithms that straightforwardly upgrade the policy, which guides states to actions. In policy gradient methods, the agent gathers directions by interacting with the state and afterward utilizes these directions to refresh the policy parameters. One critical benefit of the policy gradient method is that it can deal with continuous action spaces, as they straightforwardly parameterize the policy. For example, DDPG, MARDDPG, and so on.

1.1.3 Actor-Critic Methods

The actor-critic algorithm in reinforcement learning combines an 'actor,' responsible for action selection based on the state, and a 'critic.' The actor develops a policy correlating states to actions. With the aim of maximizing cumulative reward. The actor executes actions to gain a reward. It updates its policy based on the feedback, using policy gradient optimization. For example, MA2C.

1.2 Deep Reinforcement Learning

Nowadays, it is becoming more popular to use the DRL approach for traffic signal control (TSC) because of its capacity and adaptability toward complex situations. DRL is a merge of Deep learning and Reinforcement learning. In traffic signal control systems, DRL-based approaches and algorithms have resulted in excellent results. For example, the DRL algorithm with experience replay and target network [5], Adaptive traffic signal control system (ATCS) [17], Successor feature [33], and delay-based fairness and throughput-based fairness [35]. However, these methods and algorithms are more effective than fixed-time traffic light control.

1.2.1 Deep Q-Network (DQN)

The Deep Q-Network is widely recognized as the first method based on Deep Reinforcement Learning (DRL) that finds extensive application in traffic signal control (TSC). Unlike a simple tabular representation of values, DQN employs a deep neural network to represent the Q Function. This allows for determining the most favorable action to undertake in any given environment [6]. Through iterative updates of the Q-function, as the agent interacts with the environment, Q-Learning enables the algorithm to handle environments with numerous states and actions, as well as learn from inputs with high dimensionality such as images or sensor data**.**

However, one of the crucial challenges in implementing Deep Q-Learning lies in the Q-function, which often exhibits non-linearity and multiple local minima. Consequently, the neural network may encounter difficulties in converging to the correct Q-function. To address this issue, several approaches have been introduced, including experience replay and target network techniques [5].

DQN has been successfully applied to various domains, including Atari games, robotics, and traffic signal control. It has shown promising results in improving traffic flow efficiency and reducing travel time compared to traditional control methods.

1.2.2 Multi-Agent Reinforcement Learning (MARL)

MARL is characterized by a scenario in which multiple agents interact within a shared environment. However, coordinating these agents is a crucial aspect of the multi-agent framework. In the context of Traffic Signal Control (TSC), MARL commonly employs two approaches. The first approach is known as the centralized approach, wherein a single global agent controls multiple intersections. In a centralized approach, the agent takes all interactions as input and gives the action to all intersections simultaneously. However, it has one limitation which is the dimensionality issue as the state space increments with the number of intersections. The second one is the decentralized approach [31] refers to a system where each traffic signal decides in view of the local conditions. From the above two approaches we have seen that the decentralized approach is more adaptive than the centralized approach.

2. Literature Review

Past work has been finished to dynamically control adaptive traffic lights. In any case, because of the restricted computational resources and simulation tools, early examinations focused on tackling the issue by fuzzy logic [15], etc. In this work, Street traffic is demonstrated by limited data, which can't be applied on a huge scope. DRL was applied in traffic light control. However, The DRL-based methods [1, 3, 4, 5, 8, 10, 16] have some limitations as follows:

- It requires a lot of training data and computational resources to accomplish good performance, especially for large-scale traffic networks.
- These methods and algorithms assume that the traffic demand is stationary and does not change over time. However, traffic demand may vary due to different factors such as weather, events, or accidents.
- It does not consider the communication and coordination among agents explicitly.

Further research can be undertaken to address the safety concerns associated with Deep Q-network (DQN) in Traffic Signal Control (TSC). Since DRL (Deep Reinforcement Learning) models learn through trial and error, there is a potential risk of accidents occurring during the decision-making process. Each potential action within a given state carries a certain degree of risk. To mitigate this, it is possible to establish a set of rules to ignore highly risky actions from the pool of possible choices. By excluding actions associated with higher risk factors, the safety of DQN-based TSC systems can be enhanced, ensuring a more secure and reliable control mechanism.

Currently, most traffic simulators primarily focus on the movement characteristics of individual vehicles, following a microscopic approach. However, there is a lack of research utilizing various types of Traffic Signal Controllers (TSCs), such as the DQN-based TSCs, to effectively manage overall traffic flow on a macroscopic level. This approach takes into consideration factors like general traffic density and vehicle distribution. To enhance the accuracy of results, future studies can be conducted to explore the incorporation of macroscopic attributes alongside the microscopic approach.

In the field of Multi-Agent Reinforcement Learning (MARL), the number of agents is progressively growing, consequently leading to an expansion in the state-action space. Additionally, the traffic control system (TSC) operates within a non-stationary environment due to these factors. Consequently, it becomes imperative to establish effective communication channels among the agents to collectively determine the optimal decision.

There are several papers that propose the utilization of independent A2C, MA2C, and multi-intersection V2X-based traffic systems to address the challenges posed by end-to-end TSC and multi-intersection traffic congestion[1][24][40]. However, these papers have limitations in terms of the environment they

consider. Further research is required to explore interface time, memory consumption, and the individual action space of each agent to find optimal solutions. These factors can significantly impact the efficiency of the system. Additionally, it is important to consider the areas of data collection for training and to train MARL models for each intersection.

3. Rules and Design Considerations for the Utilization of DRL for TSC

In This part, we will explore the rules and design considerations for the utilization of DRL to TSC. Which will help distinguish the optimal DRL solution for various kinds of TSC issues. In a previous study, a deep reinforcement learning model with the 3DQN design, and a value-based approach was applied to Traffic signal control. This model targets the minimization of vehicles' mean journey time in TSC scenarios [8, 19].

Furthermore, we will define the representations of state, action, and reward, as well as discuss the strategy selection for DRL applications in TSC. By understanding these components, we can optimize the performance of DRL applications in TSC.

3.1 Defining State

It is defined as the dynamic factor that an agent discovers to discern distinctive elements from the operative state. In [8], the objective is to enhance the combined reward while reducing the wait time of vehicles at intersections. Consequently, the agent interprets the state by considering two data snippets of the vehicle, enabling it to determine an appropriate course of action. In [8], the input layer is configured as a 60 × 60 × 2 grid, representing the position and speed of the vehicle.

3.2 Defining Action

In the context of maximizing rewards, an agent strives to identify the most optimal course of action. Moreover, when aiming to minimize average waiting time for vehicles, it is crucial for the agent to appropriately select the time for each traffic stage. This decision involves choosing between maintaining the present stage or transitioning to the next stage within a specified order. Such a choice addresses the issue of inefficient traffic phase allocation. In the specific scenario mentioned in [8, 19], the output layer comprises 9 neurons, each representing the best feasible action.

3.3 Defining Reward

An agent is awarded based on its past actions. In the presented scenario [8, 19], the reward is based on the alteration in the mean wait time for vehicles at a crossroad. The agent improves system performance and achieves its objectives by raising the reward.

3.4 Choosing A Method

The objectives relating to the model need a clear definition in a method. To accelerate learning progress, the integration of mechanisms like double Q-learning, Dueling networks, and prioritized experience replay within a unified structure is explored in [8,19]. An increased learning rate diminishes the duration needed to examine every potential state-action combination, thereby facilitating the discovery of the best action. This optimal action determination, like opting to sustain or alter the imminent traffic phase, promotes smoother traffic management for Traffic Signal Controllers(TSCs). The value-based approach evaluates

each state-action duo, facilitating the selection of the optimal action within any particular state. This approach effectively aligns with the goals of TSCs and is thus selected.

3.5 Defining Architecture

To address the challenges posed by Traffic Signal Control (TSC), it is essential to define a suitable Deep Learning (DL) architecture for Deep Reinforcement Learning (DRL). An example of such an architecture is the 3DQN presented in references [8, 19]. This architecture comprises three convolutional and two fully connected layers. Its purpose is to capture the position and speed information of vehicles to address inefficiencies in traffic signal timing, necessitating the integration of convolutional layers within the 3DQN structure due to the reliance on visual inputs. Determining the right number of layers is a challenge in itself. An inadequate number of layers can hinder the model's capacity to learn from the data effectively, whereas too many layers can lead to overfitting, where the model memorizes training data, impairing its performance.

Selecting an ideal quantity of neurons per layer is pivotal. Commonly, the number of neurons in the input layer corresponds to the number of features. For instance, an intersection with 80 cells should be represented by an input layer containing 80 neurons, reflecting queue lengths and vehicle locations. However, the configuration of neuron count in hidden layers is typically empirical. It should be noted that a larger number of neurons generally boosts system efficacy but adds to its complexity.

4. Conclusion

This article presents a conceptual review of the utilization of deep reinforcement learning (DRL) in traffic signal control (TSC). To ensure efficient traffic movement, various methodologies and algorithms are employed to determine the most appropriate traffic phases at intersections. Section 2 comprises a literature review that highlights the deficiencies observed in recent studies. Section 3 provides a set of rules and design considerations for implementing DRL in TSC, aiming to identify the optimal DRL solution for different types of TSC problems. To effectively tackle the challenges of TSC, a well-defined DL architecture for DRL is necessary. The 3DQN architecture with its convolutional and FC layers provides a means to capture vehicle location and velocity in the form of visuals. Moreover, careful consideration of the number of layers and neurons is vital for optimal system performance.

5. References

1. T. Chu, J. Wang, L. Codecá, and Z. Li, “Multi-Agent deep reinforcement learning for Large-Scale traffic signal control,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1086–1095, Mar. 2020, doi: 10.1109/tits.2019.2901791.
2. F. Rodrigues, “Towards robust deep reinforcement learning for traffic signal control: demand surges, incidents and sensor failures,” *arXiv.org*, Apr. 17, 2019. https://arxiv.org/abs/1904.08353.
3. J. Guo, L. Cheng, and S. Wang, “COTV: Cooperative control for traffic light signals and connected autonomous vehicles using deep reinforcement learning,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, pp. 10501–10512, Oct. 2023, doi: 10.1109/tits.2023.3276416.
4. R. Ducrocq and N. Farhi, “Deep Reinforcement Q-Learning for Intelligent Traffic Signal Control with Partial Detection,” *International Journal of Intelligent Transportation Systems Research*, vol. 21, no. 1, pp. 192–206, Feb. 2023, doi: 10.1007/s13177-023-00346-4.

5. J. Gao, "Adaptive Traffic Signal Control: Deep Reinforcement Learning Algorithm with Experience Replay and Target Network," *arXiv.org*, May 08, 2017. https://arxiv.org/abs/1705.02755.
6. H. Chaudhuri, V. Masti, V. Veerendranath, and S. Natarajan, "A Comparative study of algorithms for Intelligent Traffic Signal Control," in *Smart Innovation, Systems and Technologies*, 2022, pp. 271–287. doi: 10.1007/978-981-16-7996-4_19.
7. W. Genders, "An Open-Source framework for adaptive traffic signal control," *arXiv.org*, Sep. 01, 2019. https://arxiv.org/abs/1909.00395.
8. 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, Feb. 2019, doi: 10.1109/tvt.2018.2890726.
9. H. Wei, "A survey on traffic signal control methods," *arXiv.org*, Apr. 17, 2019. https://arxiv.org/abs/1904.08117.
10. J. Liu, H. Zhang, Z. Fu, and Y. Wang, "Learning scalable multi-agent coordination by spatial differentiation for traffic signal control," *Engineering Applications of Artificial Intelligence*, vol. 100, p. 104165, Apr. 2021, doi: 10.1016/j.engappai.2021.104165.
11. M. Guo, P. Wang, C.-Y. Chan, and S. Askary, "A Reinforcement Learning Approach for Intelligent Traffic Signal Control at Urban Intersections," *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct. 2019, doi: 10.1109/itsc.2019.8917268.
12. J. Chen, B. Yuan, and M. Tomizuka, "Model-free Deep Reinforcement Learning for Urban Autonomous Driving," *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct. 2019, doi: 10.1109/itsc.2019.8917306.
13. S. Kanis, "Back to basics: deep reinforcement learning in traffic signal control," *arXiv.org*, Sep. 15, 2021. https://arxiv.org/abs/2109.07180.
14. L. Zhang, "DynamicLight: Dynamically Tuning Traffic Signal Duration with DRL," *arXiv.org*, Nov. 02, 2022. https://arxiv.org/abs/2211.01025.
15. N. Kumar, S. Rahman, and N. Dhakad, "Fuzzy inference enabled deep reinforcement Learning-Based traffic Light control for intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 4919–4928, Aug. 2021, doi: 10.1109/tits.2020.2984033.
16. J. Lin, Y. Zhu, L. Liu, Y. Liu, G. Li, and L. Lin, "DenseLight: Efficient Control for Large-scale Traffic Signals with Dense Feedback," *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, Aug. 2023, doi: 10.24963/ijcai.2023/672.
17. A. Qu, "Attacking Deep Reinforcement Learning-Based Traffic Signal Control Systems with Colluding Vehicles," *arXiv.org*, Nov. 04, 2021. https://arxiv.org/abs/2111.02845.
18. T. Wu *et al.*, "Multi-Agent deep reinforcement learning for urban traffic light control in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8243–8256, Aug. 2020, doi: 10.1109/tvt.2020.2997896.
19. 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, Feb. 2019, doi: 10.1109/tvt.2018.2890726.
20. X. Zang, H. Yao, G. Zheng, N. Xu, K. Xu, and Z. Li, "MetaLight: Value-Based Meta-Reinforcement Learning for Traffic signal control," *Proceedings of the ... AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 1153–1160, Apr. 2020, doi: 10.1609/aaai.v34i01.5467.

21. K. L. Tan, A. Sharma, and S. Sarkar, "Robust deep reinforcement learning for traffic signal control," *Journal of Big Data Analytics in Transportation*, vol. 2, no. 3, pp. 263–274, Dec. 2020, doi: 10.1007/s42421-020-00029-6.
22. D. Ma, B. Zhou, X. Song, and H. Dai, "A deep reinforcement learning approach to traffic signal control with temporal traffic pattern mining," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 11789–11800, Aug. 2022, doi: 10.1109/tits.2021.3107258.
23. Z. Li, "A deep reinforcement learning approach for traffic signal control optimization," *arXiv.org*, Jul. 13, 2021. https://arxiv.org/abs/2107.06115.
24. Y. Huo, Q. Tao, and J. Hu, "Cooperative control for Multi-Intersection traffic signal based on deep reinforcement learning and imitation learning," *IEEE Access*, vol. 8, pp. 199573–199585, Jan. 2020, doi: 10.1109/access.2020.3034419.
25. K.-F. Chu, A. Y. S. Lam, and V. O. K. Li, "Traffic signal control using End-to-End Off-Policy Deep Reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 7184–7195, Jul. 2022, doi: 10.1109/tits.2021.3067057.
26. W. Genders, "Using a deep reinforcement learning agent for traffic signal control," *arXiv.org*, Nov. 03, 2016. https://arxiv.org/abs/1611.01142.
27. C. Choe, S. Baek, B. Woon, and S.-H. Kong, "Deep Q Learning with LSTM for Traffic Light Control," *2018 24th Asia-Pacific Conference on Communications (APCC)*, Nov. 2018, doi: 10.1109/apcc.2018.8633520.
28. R. Huang, J. Hu, Y. Huo, and X. Pei, "Cooperative Multi-Intersection traffic signal control based on deep reinforcement learning," *CICTP 2019*, Jul. 2019, doi: 10.1061/9780784482292.256.
29. D. Zhong and A. Boukerche, "Traffic Signal Control Using Deep Reinforcement Learning with Multiple Resources of Rewards," *Proceedings of the 16th ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, Nov. 2019, doi: 10.1145/3345860.3361522.
30. M. Kolat, B. Kővári, T. Bécsi, and S. Aradi, "Multi-Agent Reinforcement Learning for Traffic Signal Control: a cooperative approach," *Sustainability*, vol. 15, no. 4, p. 3479, Feb. 2023, doi: 10.3390/su15043479.
31. L. Zhu, P. Peng, Z. Lu, and Y. Tian, "MetAVIM: Meta variationally Intrinsic Motivated Reinforcement Learning for decentralized traffic signal control," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–18, Jan. 2023, doi: 10.1109/tkde.2022.3232711.
32. T. Saiki and S. Arai, "Flexible traffic signal control via Multi-Objective Reinforcement Learning," *IEEE Access*, vol. 11, pp. 75875–75883, Jan. 2023, doi: 10.1109/access.2023.3296537.
33. L. Szőke, S. Aradi, and T. Bécsi, "Traffic Signal Control with Successor Feature-Based Deep Reinforcement Learning Agent," *Electronics*, vol. 12, no. 6, p. 1442, Mar. 2023, doi: 10.3390/electronics12061442.
34. M. Coşkun, A. Baggag, and S. Chawla, "Deep Reinforcement Learning for Traffic Light Optimization," *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, Nov. 2018, doi: 10.1109/icdmw.2018.00088.
35. M. Raeis, "A deep reinforcement learning approach for fair traffic signal control," *arXiv.org*, Jul. 21, 2021. https://arxiv.org/abs/2107.10146.

36. D. L. Li, J. Wu, M. Xu, Z. W. Wang, and K. Hu, "Adaptive Traffic Signal Control model on intersections based on deep reinforcement learning," *Journal of Advanced Transportation*, vol. 2020, pp. 1–14, Aug. 2020, doi: 10.1155/2020/6505893.
37. H. Wei *et al.*, "PressLight : Learning Max Pressure Control to Coordinate Traffic Signals in Arterial Network," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Jul. 2019, doi: 10.1145/3292500.3330949.
38. G. Zheng, "Diagnosing reinforcement learning for traffic signal control," *arXiv.org*, May 12, 2019. https://arxiv.org/abs/1905.04716
39. A. Vidali, L. Crociani, G. Vizzari, and S. Bandini, "A deep reinforcement learning approach to adaptive traffic lights management.," *WOA*, pp. 42–50, Jan. 2019, [Online]. Available: http://ceur-ws.org/Vol-2404/paper07.pdf
40. A. Oroojlooyjadid and D. Hajinezhad, "A review of cooperative multi-agent deep reinforcement learning," *Applied Intelligence*, vol. 53, no. 11, pp. 13677–13722, Oct. 2022, doi: 10.1007/s10489-022-04105-y.