A Review on Deep Q-Network-Based Traffic Signal Control

Romaissa Bendaikha[1] and Djalal Hedjazi[2]

LaSTIC laboratory, Department of Computer Science, University of Batna 2, Batna 05078, Algeria
romaissa.bendaikha@univ-batna2.dz, d.hedjazi@univ-batna2.dz

Abstract. Managing traffic signals is a significant challenge for modern transportation systems, contributing to congestion and environmental degradation. Artificial intelligence (AI) techniques, including Reinforcement Learning (RL), and, more specifically, Deep Q-Networks (DQN), have demonstrated considerable potential in addressing these issues by enabling efficient adaptive traffic signal control (ATSC). This paper comprehensively reviews DQN-based approaches applied to ATSC, focusing on key performance metrics such as vehicle waiting time, queue lengths, and traffic throughput. We explore how DQN addresses these challenges, offering insights into its effectiveness in optimizing traffic signal management at intersections.

Keywords: Intelligent Transportation Systems · Traffic Signal Control · Reinforcement Learning · Deep Q-Network · Optimization.

1 Introduction

In today's metropolitan environments, it is critical to establish an effective Traffic Signal Control (TSC) system that can respond to the numerous changes that occur throughout the day, particularly during periods of heavy congestion [1]. An effective traffic signal controller must integrate a range of advanced technologies—such as sensors, inductive loops, surveillance cameras, and intelligent agents—to deal with the issues associated with TSC.

Traditional TSC methods often produce suboptimal performances due to their reliance on static timing plans. In contrast, the integration of Artificial Intelligence (AI)—especially Machine Learning (ML)—with the Internet of Things (IoT) has paved the way for the development of Intelligent Transportation Systems (ITS) [2, 3]. These systems enable dynamic adjustment of signal timings in real time, thereby enhancing the overall efficiency of urban traffic networks.

Reinforcement Learning (RL) is a ML model where an agent learns to make decisions by interacting with an environment to maximize cumulative rewards [4]. One of the foundational algorithms in RL is Q-learning, which estimates the value of taking a particular action in a given state and updates these estimates iteratively using the Bellman equation [5] . Although Q-learning has proven effective in various discrete and small-scale problems, it struggles to scale to large or continuous state spaces due to its dependence on tabular representations.

To overcome these limitations, Deep Q-Network (DQN) were introduced [6], combining Q-learning with Deep Neural Networks (DNN) to approximate the action-value function. It incorporates techniques such as experience replay and target network to stabilize training and improve convergence. These advancements have led to significant revolutions in areas needing complex decision-making, including TSC, where DQN has been applied to develop adaptive and efficient control policies.

As shown in Fig. 1, this review is based on primary literature retrieved from reputable databases, including *IEEE Xplore, Elsevier* , and *SpringerLink*. These databases were selected due to their wide coverage of high-quality peer-reviewed research in the fields of artificial intelligence, transportation engineering, and intelligent systems. Search focused on the period from 2019 to 2023, during which significant advancements in DQN-based TSC were published.

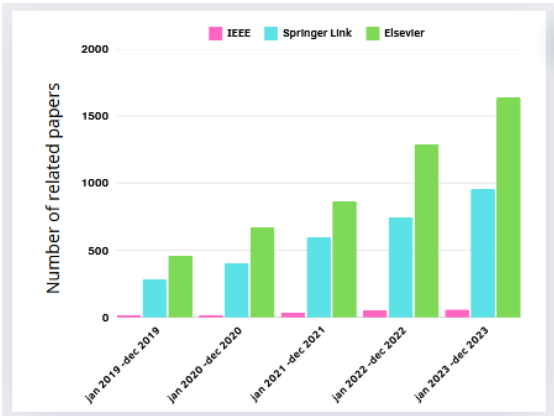

Fig. 1. Indexed DQN-based TSC studies (2019–2023).

The selected sources include *journal articles*, *conference proceedings*, and *book chapters*, offering a diverse and rigorous foundation for analysis.

The present review explores several key research questions: How does applying DQN improve traffic conditions? What are the observed effects of DQN-based TSC on overall traffic efficiency? How does the design of the state, action, and reward functions in DQN influence its effectiveness in TSC? What challenges and limitations are associated with implementing DQN in real-world TSC systems?

The remainder of this paper is organized as follows: Section 2 presents an overview of DRL, while Section 3 discusses the key challenges and requirements of TSC. Section 4 provides a detailed review of DQN-based approaches for TSC. Section 5 outlines potential future research directions, and Section 6 concludes the study by summarizing the main findings and their implications.

2 Overview of DRL

2.1 From RL to Deep RL

RL is a branch of ML in which an agent acquires decision-making skills through interaction with an environment to achieve specific goals [4]. In contrast to supervised learning, which depends on pre-labeled datasets, RL allows an agent to discover effective behaviors by interacting with its environment and receiving evaluative feedback through rewards or penalties [4].

The agent explores various actions and receives rewards based on the outcomes, gradually refining its strategy to optimize cumulative rewards over time. The core components of RL include the agent, the environment, states, actions, and rewards—elements that collectively define the learning and decision-making process. This process is modeled as a Markov Decision Process (MDP) [4], defined by defined by a set of states S, a set of actions A, a transition probability function $P(s' \mid s, a)$, and a reward function R that provides immediate feedback for actions taken. It also includes a discount factor $\gamma \in [0, 1]$, which weighs the importance of future rewards, and a policy $\pi(a \mid s)$, which maps each state to a probability distribution over possible actions. The agent attempts to develop an optimal policy (noted π^*) that optimizes the expected cumulative reward (noted G_t) defined as:

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{1}$$

To estimate the quality of state-action pairs under a given policy, RL algorithms define the action-value function $Q(s, a)$, which reflects the expected return of taking action a in state s and following policy π thereafter. In Q-learning, the update rule is based on the Bellman optimality equation [2]:

$$Q(s, a) = (1 - \alpha)Q(s, a) + \alpha[R(s, a) + \gamma \max_{a'} Q(s', a')] \tag{2}$$

Here, α is the learning rate, and s' is the next state, and the term $\max_{a'} Q(s', a')$ represents the highest predicted value achievable from s'. By repeatedly applying this update, the agent progressively refines its Q-values and learns to favor actions that lead to higher cumulative rewards.

These traditional algorithms typically use tabular representations, which become inefficient or infeasible when applied to environments with large-scale or continuous state spaces [7] . This limitation has led to the development of more advanced approaches that employ function approximation techniques—such as DNN—have been developed, leading to the emergence of DRL.

DRL combines the foundational principles of RL with the representational power of DNNs to tackle complex decision-making problems. This synergy enables agents to process and learn from high-dimensional inputs, proving especially effective in scenarios where traditional RL methods are insufficient. Systems based on this paradigm are characterized by three key capabilities: generalization, autonomous learning, and intelligent behavior. [8].

In the domain of TSC, DRL-based approaches utilize the current traffic state at an intersection to determine optimal strategies for selecting signal phases or durations. Existing research in this area varies widely across four main dimensions: state representation, reward function design, action selection strategy, and agent architecture [9].

2.2 Deep Q-Network (DQN)

The DQN is a value-based DRL algorithm that extends classical Q-learning by using DNN. The architecture of DQN comprises several integral components that work together to approximate optimal action-value functions. The DNN's input layer is responsible for processing the agent's current environmental state nu-merically. This is followed by one or more hidden layers, typically composed of fully connected neurons, which extract and transform features from the input into more abstract representations useful for learning. The output layer contains one neuron per possible action, with each output representing the estimated Q-value for taking a specific action in the given state.

To enhance learning stability, DQN employs two crucial mechanisms: *experience replay memory* and a *target network* [6, 10]. Experience replay memory stores past interactions as tuples—consisting of the current state, action, reward, and next state—and samples from this memory during training to break temporal correlations. The target network, a periodically updated copy of the Q-network, generates stable target Q-values during learning to prevent destabilizing feedback loops. A loss function is used to compute the divergence be-tween predicted Q- values and target values derived from the Bellman equation, and this loss is minimized through optimization techniques, typically using Stochastic Gradient Descent (SGD) or its variants to update the network's parameters.

Moreover, DQN leverages deep Convolutional Neural Networks (CNNs) to process raw or structured input, especially in domains where spatial or temporal relationships are critical. CNNs, inspired by the visual cortex in the human brain, are particularly effective in extracting hierarchical features from input data [11]. Given the capabilities of DRL and the advancements brought by architectures such as the DQN, it be-comes essential to examine how these techniques can be applied to real-world problems. One prominent application domain is TSC, where intelligent decision-making is crucial for optimizing traffic flow and reducing congestion.

3 Traffic Signal Control (TSC): Challenges and Requirements.

TSC plays a crucial role in urban mobility management, aiming to optimize traffic flow, minimize congestion, and reduce travel time and CO2 emissions. A traffic network consists of one or more intersections connected via edge nodes, through which vehicles enter and exit the system. Each intersection includes multiple legs and lanes, permitting vehicles to proceed straight, turn left, or turn right [12].

Despite its centrality, the efficient management of traffic remains a challenge due to dynamic and unpredictable conditions, structural complexity, and growing traffic demand. Furthermore, TSC systems must account for a range of interdependent parameters—including signal phase sequencing, cycle length, green phase duration, and offsets—which directly influence queue lengths, waiting times, and throughput.

To better understand these challenges, it is important to distinguish between several TSC optimization domains. Network-level optimization focuses on synchronizing signal operations across intersections to improve overall traffic progression. At the intersection level, optimization efforts target local conditions such as queue lengths and arrival rates, aiming to prevent bottlenecks that can propagate downstream. Roundabouts also present a unique set of optimization problems, particularly in terms of cycle duration and delay minimization. Cycle-level optimization involves adjusting red, yellow, and green phase durations based on local traffic dynamics, a task complicated by unpredictable demand and the difficulty of correctly estimating optimal green times [13].

For instance, Inappropriate phase sequencing and ineffective green time allocation can lead to congestion, cross-blocking, and green idle, especially when timing doesn't reflect real-time traffic demand. Accurate vehicle counts and unexpected incidents can degrade system performance. These challenges are further compounded by the dynamic and high-dimensional nature of traffic environments, where conditions change rapidly and the impact of decisions may be delayed or diffused across the network.

To address these complexities, modern TSC systems must incorporate adaptive mechanisms that can react in real time, operate effectively under uncertainty, and scale across multiple intersections. Multi-agent systems offer a decentralized framework for achieving this scalability and coordination, particularly in large urban networks [14].These systems must also meet several functional requirements, including the ability to coordinate signals over large-scale networks, support priority policies for transit or emergency vehicles, and adapt to variable network structures and demand patterns [15]. Signal timing strategies must be synchronized to minimize delays and energy consumption [16], while predictive capabilities become crucial in oversaturated conditions to proactively mitigate congestion [17].

These requirements underline the limitations of traditional fixed-time or rule-based systems and point to the potential of DRL-based methods—as a promising approach to developing intelligent, data-driven traffic controllers.

4 DQN-Based TSC

This section outlines the integration of DQN into TSC systems, provides a review of studies applying this approach, and highlights key challenges that guide current research efforts.

4.1 DQN-Based TSC Model

At the core of this model is an agent that interacts with the environment by observing its current state, se-lecting an action (e.g., changing the signal phase), and receiving feedback in the form of a reward. Over time, the agent learns to optimize its behavior by approximating the optimal action-value function using DNN. The environment in such model typically consists of one or multiple intersections where an agent governs each intersection. The state is a numerical representation of the traffic condition, com-monly including features such as queue lengths, vehicle-waiting times, vehicle positions and speeds, the current signal phase, and the elapsed time since the last phase change [18, 19]. The action space defines the possible signal operations the agent can take, such as switching to the next phase, extending the green light, or maintaining the current phase. Thus, action selection significantly im-pacts both immediate and future rewards, as well as the evolving traffic state [20]. Depending on the problem formulation, the action space may be discrete—such as selecting the next signal phase [21], or continuous, involving fine-tuning signal durations [22]. DQN, as a value-based method, chooses actions that maximize estimated Q-values [23], while policy gradient approaches learn a distribution over actions [24]. Actor-critic methods combine both strategies, with the actor selecting actions and the critic evaluating them based on observed rewards and states [25]. The learned policy continually adapts to improve traffic efficiency at intersections.

The reward function guides learning by encouraging actions that improve traffic flow. Commonly used reward designs aim to minimize average delay, queue length, or vehicle waiting time, and may incorporate penalties [26] for phase switching or congestion. This feedback is used to update the Q-values, which estimate the expected return of taking a given action in a given state. Common reward metrics include average waiting time, travel delay, queue length, and throughput [26]. Each metric influences trade-offs between traffic efficiency and environmental impact—for example, minimizing waiting time can increase vehicle speed and reduce CO 2 emissions, while focusing solely on emissions may raise delays [26]. Positive rewards typically signal improved control [27], whereas negative rewards indicate congestion [28], enabling the agent to refine its decisions through continuous feedback.

The model uses a Deep Learning Network (DNN) with input, hidden, and output layers to approximate the Q-function, stabilize learning with experience replay and periodically updated target network for stable target values.

DQN agent learns adaptive traffic control policies, offering a scalable and data efficient alternative to tra-ditional rule-based or fixed-time signal control strategies in single-agent or multi-agent settings.

As illustrated in Algorithm 1, the training process of a DQN-based TSC model involves iterative interaction between the agent and the traffic environment to improve its decision-making policy.

Algorithm 1. Pseudocode of the DQN-based TSC Algorithm

1. Initialize the Q-network with random parameters θ.
2. Create a target network and set its θ^- equal to θ.
3. Initialize an empty replay memory D.
4. **For** each training episode **do**:
  5. Reset the traffic environment and observe the initial state s_0.
  6. **For** each time step t within the episode **do**:
    7. Select action a_t using the ϵ-greedy strategy.
    8. Apply action a_t in the environment.
    9. Receive reward r_t and next state s_{t+1}.
    10. Store transition (s_t, a_t, r_t, s_{t+1}) in memory D.
    11. Randomly sample a mini-batch from memory D.
    12. For each sampled transition (s, a, r, s'):
      – Compute target Q-value:

$$y = r + \gamma \cdot \max_{a'} Q_{\text{target}}(s', a'; \theta^-)$$

13. Update the Q-network by minimizing the loss function using gradient descent.
14. Every C steps, update $\theta^{-} \leftarrow \theta$.
15. **End for**
16. **End for**

At each time step, the agent observes the current state, selects an action according to ϵ-greedy strategy, and receives a corresponding reward along with the next state. These experiences—consisting of state, action, reward, and next state tuples—are stored in an experience re- play memory to enhance learning stability [29]. During training, mini-batches of stored transitions are sampled randomly from the replay memory to update the DQN's parameters, denoted as θ, using gradient descent. The target Q-values are computed using a separate target network with parameters θ^{-}, which is updated periodically to stabilize learning [30]. The Q-network, parameterized by θ, learns to approximate the action-value function Q(s, a; θ), estimating the expected cumulative reward for each action given the current state. Through successive updates that minimize the temporal-difference loss, θ encodes knowledge of traffic patterns, enabling the agent to improve its decisions over time.

The training process continues until convergence criteria are met indicating the agent has learned an effective traffic control strategy. An overview of the DQN architecture is presented in Fig.2.

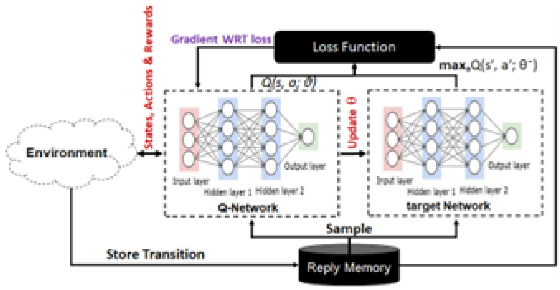

Fig. 2. Overview DQN architecture used for TSC.

4.2 Literature Review

The following subsection reviews prior studies that have employed the DQN model to address TSC challenges.

Since the introduction of the DQN model by [6], numerous studies have investigated its application to the TSC area.researchers. Ge et al. [31], utilized real-time intersection data to reduce average waiting and travel times. Huo et al. [32], proposed a DRL-based method leveraging high-resolution event-based data, outperforming conventional approaches. Wu et al. [33] developed a Double DQN with a dual-agent strategy, enhancing traffic capacity. Katragadda et al.[34], designed a DQN-based system to sequence green signals across intersections, achieving a 30% improve-ment in wait time. Subba Rao et al. [35] integrated real-time GPS data, leading to reduced vehicle wait times compared to fixed-time control. Kim et Sohn [36], introduced a DQN-based green time allocation system that surpassed conventional sequencing systems. Gao et al. [37], used high-quality data to reduce cumulative delay (82%), queue length (66%), and travel time (20%). Pan [38] proposed a reward-aware DQN method that decreased vehicle wait time by up to 100%, queue length by up to 100%, and total travel time by up to 68%. Qi et al. [39] enhanced DQN to achieve a 26.7% reduction in average waiting time and improved queue management. A DQN-based model in [40] demonstrated reduced cumulative vehicle delay, enhancing urban TSC efficiency. Hu et al. [41] introduced a Multi-Agent DDQN that lowered wait times and queues by 40–60% and increased speed by 10–18%. In [42], Deep Q-learning reduced queue lengths by 9.7% and improved overall traffic flow. Shabab et al. [43], proposed a real-time RL method, decreasing wait times by 18–53% and conflicts by 19–25%. Krishnendhu et al [44] applied a Double Dueling DQN (3DQN) to enhance traffic safety and efficiency by 42%. Finally, Shashi et al [45] presented a dynamic DQN method that improved average vehicle delays. Table 1 summarizes the main contributions of these approaches.

Table 1: Key Studies on DQN-Based TSC Solutions.

Ref	**Year**	**Approach**	**Key Results**
[31]	2022	DQN	Reduced AWT and travel time.
[32]	2020	DRL with high-resolution event-based data	Improved efficiency, adaptability, and cost-effectiveness over traditional methods.
[33]	2023	DDQN with a dual-agent architecture	Increased traffic throughput and system robustness.
[34]	2023	DQN-based intelli-gent control	Achieved 30% improvement in AWT.
[35]	2024	DRL (DRQN with RNN)	Reduced AWT versus fixed-time control.
[36]	2022	DQN	Optimized junction capacity and dynamic green alloca-tion.
[37]	2017	DRL	Reduced delay (82%), queue length (66%), and travel time (20%).
[38]	2024	DQN-based real-time TSC	Decreased AWT (up to 100%), AQL (up to 100%), and travel time (up to 68%).
[39]	2022	Improved DQN	Reduced AWT by 26.7%, and improved queue handling.
[40]	2022	DQN	Decreased cumulative delay enhanced overall traffic efficiency.
[41]	2024	Multi-Agent DDQN	Reduced AWT and AQL by 40–60%, increased speed by 10–18%.
[42]	2023	Deep Q-Learning	Reduced AWT and AQL by 9.7%, improved cumulative reward.
[43]	2023	DQN-Based RL	Lowered AWT by 18–53% and traffic conflicts by 19–25%.

[44]	2023	3DQN for ATSC	Improved safety and efficiency by 42% over static timing control.
[45]	2021	DQN	Reduced AWT dynamically.

4.3 Challenges and Limitations of DQN-Based TSC

Despite the growing success of DQN approaches in TSC, several limitations remain. One key challenge is training instability, which arises from correlations in sequential observations and the non-linear approxima-tions of DNNs [6]. Scalability is another concern—while DQN performs well at single intersections, extending it to large-scale networks leads to exponential growth in the state-action space, increasing computational load and training time, thus limiting real-time applicability [1]. Generalization is also problematic. Although DQN agents theoretically learn through interaction with their environment, most TSC research relies on simulators like SUMO, with limited real-world deployment due to safety, infrastructure, and regulatory constraints. As a result, policies trained in simulation may not transfer effectively to real-world settings [46]. Furthermore, designing appropriate reward functions is complex; ill-defined rewards can misguide learning and produce suboptimal behavior. This challenge has motivated the use of Multi-Objective Deep Reinforcement Learning (MODRL), which enables agents to optimize multiple conflicting goals simultaneously [47, 48]. Finally, DQN models often depend on manually tuned hyper-parameters (α, γ and ϵ), whose optimal values are typically determined through costly trial-and-error [49]. Addressing these challenges is essential to advance DQN-based TSC from theoretical frameworks to ro-bust, deployable solutions in ITS.

5 Future Directions

Building on current limitations and promising research trends, future work on DQN-based TSC should ex-plore the following key directions:

5.1 Federated & Hierarchical Multi-Agent Training

Hierarchical Federated Reinforcement Learning (HFRL) [50], combines hierarchical control with federated learning to support scalable, coordinated, and personalized traffic signal control across urban networks. When integrated with DQN-based models, HFRL clusters in-tersections for localized federated training, enabling policy specialization, faster convergence, and greater robustness. Notably, this integration reduces communication overhead and enhances learning stability, making DQN-based TSC more suitable for real-time deployment. Recent studies [49, 51, 52], confirm that HFRL improves scalability, adaptability, and performance—effectively addressing several limitations of traditional centralized or standalone DQN approaches.

5.2 Multi-Objective Optimization via Multi-Objective Deep RL

Multi-Objective DRL (MODRL) enhances traditional DRL by simultaneously optimizing conflicting goals such as safety, efficiency, emissions, and fairness in TSC. Methods like Dueling Double DQN (D3QN) have shown effectiveness in reducing conflicts, delays, and CO2 emissions, supporting sustainable urban mobility [53–55]. Additionally, adaptive weight tuning frameworks offer strong adaptability in mixed-autonomy settings, especially when leveraging connected vehicle data and V2I communication, making MODRL a promising approach for future intelligent traffic systems.

5.3 Graph-Based & Region-Based Coordination

Region-based coordination and graph-based learning are promising strategies for scaling DQN-based TSC in large urban networks. Region-based methods divide traffic networks into sub-regions to enable localized coordination, reduce computational complexity, and facilitate scalable learning—especially when inte-grated with decentralized or federated frameworks. In parallel, graph-based approaches like Graph Atten-tion Networks (GATs) model intersections as nodes and capture spatial dependencies between them. This enables more context-aware control decisions. Integrating these with DQN—such as in CoLight [56] and MGMQ [57], which combines Double DQN with GAT and GraphSAGE—has shown notable reductions in travel time, delay, and queue lengths. Together, these approaches support localized optimization and global coordina-tion, making them well-suited for deployment in ITS. Continued research is essential to harness their full potential in dynamic, multi-modal urban environments. [58].

5.4 Enhancing Sim-to-Real Transfer and Continual Learning

Bridging the gap between simulation and real-world deployment is a key challenge for DQN-based TSC. Policies trained in simulation often degrade in real-world use due to model inaccuracies, sensor noise, and unforeseen events. Incremental update-aware training can enhance robustness in such dynamic conditions [59]. Domain randomization—by varying simulator parameters like traffic flow, vehicle behavior, and sensor input—exposes agents to diverse scenarios, improving generalization. In [60] Müller et Sabatelli showed that combining this with meta-learning [61] boosts transferability across environments. Additionally, Bayesian meta-DQN (BM-DQN), proposed by Zou et Qin [62]), integrates prior knowledge to enable faster adaptation in unfamiliar traffic contexts [62].

5.5 Integration with Connected and Automated Vehicles

The integration of Connected and Automated Vehicles (CAVs) significantly enhances DQN-based TSC by providing enriched state representations through V2I communication. Unlike traditional sensors, CAVs offer real-time data on positions, speeds, and intentions, allowing agents to predict traffic patterns more accurately and optimize signal timing proactively. This reduces observation uncertainty and improves reward shaping, leading to faster convergence and superior DRL policy performance [63]. Additionally, CAV data supports multi-agent DQN coordination across intersections, enabling shared deci-sion-making and reduced network-wide delays [64]. As CAV adoption grows, the scalability and learning efficiency of DQN-based TSC systems are expected to improve further.

5.6 Integration with metaheuristics

Integrating metaheuristic algorithms with DQN and its variants (e.g., DDQN, Dueling DQN) has shown promise in enhancing convergence, stability, and overall performance in complex TSC tasks. Particle Swarm Optimization (PSO) is widely used to fine-tune DQN hyper-parameters, leading to faster conver-gence and improved control, with observed reductions in travel time, queue length, and emissions in SUMO-based simulations [65]. Similarly, Genetic Algorithms (GA) have been applied to evolve interpretable decision trees, such as urgen-cy functions, improving DQN-driven TSC performance [65]. This direction is critical in DQN-based TSC, as it helps overcome issues like slow learning, poor policy exploration, and coordination challenges in multi-agent settings. [66].

6 Conclusion

This paper has presented a comprehensive review of DQN applications in TSC, emphasizing its potential to revolutionize urban traffic management. With the continuous growth of urban areas and the persistent challenge of congestion, DQN-based approaches offer a data-driven, adaptive solution by optimizing signal timings in real time. By effectively handling high-dimensional state spaces and learning traffic dynam-ics, DQN has demonstrated notable improvements in reducing vehicle delays, minimizing queue lengths, and increasing intersection throughput.

The literature reviewed confirms the increasing momentum of DQN-based TSC methods, with success-ful deployments in various simulated and real-world settings. Nevertheless, several limitations persist, par-ticularly in terms of scalability to large, complex networks and integration with legacy infrastructure. Addressing these challenges will require advances in model robustness, the adoption of hybrid frameworks (e.g., metaheuristic-enhanced DQN), and the deployment of MARL systems to manage broader traffic ecosystems.

In summary, DQN represents a significant step forward in the evolution of (ITS, with its adaptability and learning capabilities offering a path toward smarter, more responsive, and sustainable urban mobility. Continued research and innovation in this domain will be key to realizing the full potential of DQN-driven traffic control in real-world environments.

References

1. Tarek Amine Haddad, Djalal Hedjazi, and Sofiane Aouag. A deep reinforcement learning-based cooperative approach for multi-intersection traffic signal control. *Engineering Applications of Artificial Intelligence*, 114:105019, September 2022.
2. Rodolfo I. Meneguette, Robson E. De Grande, and Antonio A. F. Loureiro. *Intelligent Transport System in Smart Cities: Aspects and Challenges of Vehicular Networks and Cloud.* Urban Computing. Springer International Publishing, Cham, 2018.
3. Tarek Amine Haddad, Djalal Hedjazi, and Sofiane Aouag. An IoT-Based Adaptive Traffic Light Control Algorithm for Isolated Intersection. In Mustapha Reda Senouci, Mohamed El Yazid Boudaren, Faouzi Sebbak, and M'hamed Mataoui, editors, *Advances in Computing Systems and Applications*, pages 107–117, Cham, 2021. Springer International Publishing.

4. Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction.
5. Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3):279–292, May 1992.
6. Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015. Publisher: Nature Publishing Group.
7. L. P. Kaelbling, M. L. Littman, and A. W. Moore. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237–285, May 1996.
8. Aske Plaat. *Deep Reinforcement Learning, a textbook.* 2022.
9. Hao Dong, Zihan Ding, and Shanghang Zhang, editors. *Deep Reinforcement Learning: Fundamentals, Research and Applications.* Springer Singapore, Singapore, 2020.
10. Hado van Hasselt, Arthur Guez, and David Silver. Deep Reinforcement Learning with Double Q-Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), March 2016. Number: 1.
11. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. Publisher: Nature Publishing Group.
12. Faizan Rasheed, Kok-Lim Alvin Yau, Rafidah Md. Noor, Celimuge Wu, and Yeh-Ching Low. Deep Reinforcement Learning for Traffic Signal Control: A Review. *IEEE Access*, 8:208016–208044, 2020.
13. Palwasha W. Shaikh, Mohammed El-Abd, Mounib Khanafer, and Kaizhou Gao. A Review on Swarm Intelligence and Evolutionary Algorithms for Solving the Traffic Signal Control Problem. *IEEE Transactions on Intelligent Transportation Systems*, 23(1):48–63, January 2022.
14. Djalal Hedjazi, Fateh Layachi, and Djallel Eddine Boubiche. A multi-agent system for distributed maintenance scheduling. *Computers & Electrical Engineering*, 77:1–11, July 2019.
15. Peng Jing, Hao Huang, and Long Chen. An Adaptive Traffic Signal Control in a Connected Vehicle Environment: A Systematic Review. *Information*, 8(3):101, August 2017.
16. Ishu Tomar, S. Indu, and Neeta Pandey. Traffic Signal Control Methods: Current Status, Challenges, and Emerging Trends. In Deepak Gupta, Zdzislaw Polkowski, Ashish Khanna, Siddhartha Bhattacharyya, and Oscar Castillo, editors, *Proceedings of Data Analytics and Management*, volume 90, pages 151–163. Springer Nature Singapore, Singapore, 2022.
17. Marco Wiering, Jelle Veenen, Jilles Vreeken, and Arne Koopman. Intelligent Traffic Light Control. August 2004.
18. Melrose Roderick, J. MacGlashan, and Stefanie Tellex. Implementing the Deep Q-Network. *ArXiv*, November 2017.
19. Duowei Li, Jianping Wu, Ming Xu, Ziheng Wang, and Kezhen Hu. Adaptive Traffic Signal Control Model on Intersections Based on Deep Reinforcement Learning. *Journal of Advanced Transportation*, 2020(1):6505893, 2020.
20. Sangmin Park, Eum Han, Sungho Park, Harim Jeong, and Ilsoo Yun. Deep Q-network-based traffic signal control models. *PLOS ONE*, 16(9):e0256405, September 2021.
21. Xiaoyuan Liang, Xunsheng Du, Guiling Wang, and Zhu Han. Deep Reinforcement Learning for Traffic Light Control in Vehicular Networks. *IEEE Transactions on Vehicular Technology*, 68(2):1243–1253, February 2019.
22. Faizan Rasheed, Kok-Lim Alvin Yau, and Yeh-Ching Low. Deep reinforcement learning for traffic signal control under disturbances: A case study on Sunway city, Malaysia. *Future Generation Computer Systems*, 109:431–445, August 2020.

23. Song Wang, Xu Xie, Kedi Huang, Junjie Zeng, and Zimin Cai. Deep Reinforcement Learning-Based Traffic Signal Control Using High-Resolution Event-Based Data. *Entropy*, 21(8):744, July 2019.
24. Jianfeng Gu, Yong Fang, Zhichao Sheng, and Peng Wen. Double Deep Q-Network with a Dual-Agent for Traffic Signal Control. *Applied Sciences*, 10(5):1622, February 2020.
25. Hyunjin Joo and Yujin Lim. Intelligent Traffic Signal Phase Distribution System Using Deep Q-Network. *Applied Sciences*, 12(1):425, January 2022.
26. Tongyu Zhao, Peng Wang, and Songjinag Li. Traffic Signal Control with Deep Reinforcement Learning. In *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, pages 763–767, Chongqing, China, December 2019. IEEE.
27. Hyunjin Joo and Yujin Lim. Traffic Signal Time Optimization Based on Deep Q-Network. *Applied Sciences*, 11(21):9850, October 2021.
28. Wade Genders and Saiedeh Razavi. Using a Deep Reinforcement Learning Agent for Traffic Signal Control, November 2016.
29. Anjukan Kathirgamanathan, Eleni Mangina, and Donal P. Finn. Development of a Soft Actor Critic deep reinforcement learning approach for harnessing energy flexibility in a Large Office building. *Energy and AI*, 5:100101, September 2021.
30. Kun Zhang, Hongfeng Xu, Baofeng Pan, and Qiming Zheng. Explicit coordinated signal control using soft actor–critic for cycle length determination. *IET Intelligent Transport Systems*, 18(8):1396–1407, August 2024.
31. Hongwei Ge, Dongwan Gao, Liang Sun, Yaqing Hou, Chao Yu, Yuxin Wang, and Guozhen Tan. Multi-Agent Transfer Reinforcement Learning With Multi-View Encoder for Adaptive Traffic Signal Control. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):12572–12587, August 2022.
32. Yusen Huo, Qinghua Tao, and Jianming Hu. Cooperative Control for Multi-Intersection Traffic Signal Based on Deep Reinforcement Learning and Imitation Learning. *IEEE Access*, 8:199573–199585, 2020.
33. Chunliang Wu, Inhi Kim, and Zhenliang Ma. Deep Reinforcement Learning Based Traffic Signal Control: A Comparative Analysis. *Procedia Computer Science*, 220:275–282, January 2023.
34. Harshith d, Natarajan B, and R Bhuvaneswari. Enhancing Traffic Efficiency through Real-Time Traffic Signal Control Using Deep Q-Learning. In *2023 Global Conference on Information Technologies and Communications (GCITC)*, pages 1–6, December 2023.
35. P V R Subba Rao, Venkata Ramana Murthy Polisetty, K Krishna Jayanth, N V Sai Manoj, Vankadari Mohith, and R Prasanna Kumar. Deep Adaptive Algorithms for Local Urban Traffic Control: Deep Reinforcement Learning with DQN. In *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pages 592–598, January 2024.
36. Gyeongjun Kim and Keemin Sohn. Area-wide traffic signal control based on a deep graph Q-Network (DGQN) trained in an asynchronous manner. *Applied Soft Computing*, 119:108497, April 2022.
37. Juntao Gao, Yulong Shen, Jia Liu, Minoru Ito, and Norio Shiratori. Adaptive Traffic Signal Control: Deep Reinforcement Learning Algorithm with Experience Replay and Target Network, May 2017.
38. Taoyu Pan. Traffic light control with reinforcement learning. *Applied and Computational Engineering*, 43:26–43, February 2024.
39. Fan Qi, Rui He, Longhao Yan, Junfeng Yao, Ping Wang, and Xiangmo Zhao. Traffic Signal Control with Deep Q-Learning Network (DQN) Algorithm at Isolated Intersection. In *2022 34th Chinese Control and Decision Conference (CCDC)*, pages 616–621, August 2022.

40. R SenthilPrabha, Baskaran J, Darshan Sasikumar, Gowtham Sriram, Kevin Nelson, and Pranavh Harish. Automation of Traffic Control Systems with Deep Q-Learning Network. In *2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, pages 1–6, December 2022.
41. Ta-Yin Hu and Zhuo-Yu Li. A multi-agent deep reinforcement learning approach for traffic signal coordination. *IET Intelligent Transport Systems*, 18(8):1428–1444, 2024.
42. S M Masfequier Rahman Swapno, Gunjan Chhabra, Keshav Kaushik, SM Nuruzzaman Nobel, Md Babul Islam, and Md. Shahiduzzaman. An Adaptive Traffic Signal Management System Incorporating Reinforcement Learning. In *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*, pages 1–6, November 2023.
43. Kazi Redwan Shabab, Syed Mostaquim Ali, and Mohamed H. Zaki. Deep Reinforcement Learning-Based Short-Term Traffic Signal Optimizing Using Disaggregated Vehicle Data. *Data Science for Transportation*, 5(2):13, July 2023.
44. S. P. Krishnendhu, Mainampati Vigneshwari Reddy, Thulunga Basumatary, and Prabu Mohandas. A Reinforcement Learning Based Adaptive Traffic Signal Control for Vehicular Networks. In G. Ranganathan, Robert Bestak, and Xavier Fernando, editors, *Pervasive Computing and Social Networking*, pages 547–561, Singapore, 2023. Springer Nature.
45. Farjana Islam Shashi, Salman Md Sultan, Afroza Khatun, Tangina Sultana, and Tahira Alam. A Study on Deep Reinforcement Learning Based Traffic Signal Control for Mitigating Traffic Congestion. In *2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, pages 288–291, May 2021.
46. Xingshuai Huang, Di Wu, Michael Jenkin, and Benoit Boulet. Sample-Efficient Meta-RL for Traffic Signal Control. In *2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 312–318, August 2024. ISSN: 2576-7046.
47. Thanh Thi Nguyen, Ngoc Duy Nguyen, Peter Vamplew, Saeid Nahavandi, Richard Dazeley, and Chee Peng Lim. A Multi-Objective Deep Reinforcement Learning Framework. *Engineering Applications of Artificial Intelligence*, 96:103915, November 2020. arXiv:1803.02965 [cs].
48. Lei Nie, Dandan Qi, Bingyi Liu, Peng Li, Haizhou Bao, and Heng He. CMRM: Collaborative Multi-agent Reinforcement Learning for Multi-objective Traffic Signal Control. *IEEE Transactions on Consumer Electronics*, pages 1–1, 2025.
49. Yongjie Fu, Lingyun Zhong, Zifan Li, and Xuan Di. Federated Hierarchical Reinforcement Learning for Adaptive Traffic Signal Control, April 2025. arXiv:2504.05553 [cs].
50. Yao Jing, Bin Guo, Nuo Li, Ruonan Xu, and Zhiwen Yu. Federated multi-agent reinforcement learning: A comprehensive survey of methods, applications and challenges. *Expert Systems with Applications*, 293:128729, December 2025.
51. Rongqing Zhang, Jingxin Mao, Hanqiu Wang, Bing Li, Xiang Cheng, and Liuqing Yang. A Survey on Federated Learning in Intelligent Transportation Systems. *IEEE Transactions on Intelligent Vehicles*, pages 1–17, 2024.
52. Mamta Kumari, Zoirov Ulmas, Suseendra R, Janjhyam Venkata Naga Ramesh, and Yousef A. Baker El-Ebiary. Utilizing Federated Learning for Enhanced Real-Time Traffic Prediction in Smart Urban Environments. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15(2), 2024. Number: 2 Publisher: The Science and Information (SAI) Organization Limited.
53. Xin Chu and Xingqin Cao. Adaptive traffic signal control for road networks based on dueling double deep q-network. In *International Conference on Frontiers of Traffic and Transportation Engineering (FTTE 2024)*, volume 13645, pages 167–174. SPIE, June 2025.

54. JinFang Sheng, Wang Cai, and Bin Wang. A dueling double deep Q-network with style-based recalibration module for traffic light optimization in deep reinforcement learning. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, page 09544070241310057, January 2025. Publisher: IMECHE.
55. Peng Wang and Wenlong Ni. An Enhanced Dueling Double Deep Q-Network With Convolutional Block Attention Module for Traffic Signal Optimization in Deep Reinforcement Learning. *IEEE Access*, 12:44224–44232, 2024.
56. Hua Wei, Nan Xu, Huichu Zhang, Guanjie Zheng, Xinshi Zang, Chacha Chen, Weinan Zhang, Yanmin Zhu, Kai Xu, and Zhenhui Li. CoLight: Learning Network-level Cooperation for Traffic Signal Control. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, pages 1913–1922, New York, NY, USA, November 2019. Association for Computing Machinery.
57. Tao Wang, Zhipeng Zhu, Jing Zhang, Junfang Tian, and Wenyi Zhang. A large-scale traffic signal control algorithm based on multi-layer graph deep reinforcement learning. *Transportation Research Part C: Emerging Technologies*, 162:104582, May 2024.
58. Hamed Vahdat-Nejad, Tahereh Mohammadi, and Wathiq Mansoor. CATMI: Context-Aware Traffic Management at Autonomous Un-Signalized Intersections. *International Journal of Intelligent Transportation Systems Research*, 21(1):145–158, April 2023. Publisher: Springer Verlag.
59. Gurseerat Kaur Rattan, Zhenyu Mao, Jialong Li, Mingyue Zhang, and Kenji Tei. Incremental Reinforcement Learning with Prioritized Sweeping for Traffic Signal Control. In *2023 IEEE 8th International Conference on Intelligent Transportation Engineering (ICITE)*, pages 21–28, October 2023.
60. Arthur Müller and Matthia Sabatelli. Bridging the Reality Gap of Reinforcement Learning based Traffic Signal Control using Domain Randomization and Meta Learning. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, pages 5271–5278, September 2023. ISSN: 2153-0017.
61. Marcin Korecki. Deep Reinforcement Meta-Learning and Self-Organization in Complex Systems: Applications to Traffic Signal Control. *Entropy*, 25(7):982, July 2023. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
62. Yayi Zou and Zhiwei Qin. Bayesian Meta-reinforcement Learning for Traffic Signal Control, October 2021. arXiv:2010.00163 [cs].
63. Shan Fang, Lan Yang, Wen-long Shang, Xiangmo Zhao, Fengze Li, and Washington Ochieng. Cooperative Control Model Using Reinforcement Learning for Connected and Automated Vehicles and Traffic Signal Light at Signalized Intersections. *IEEE Internet of Things Journal*, pages 1–1, 2025.
64. Guillen-Perez Antonio and Cano Maria-Dolores. Multi-Agent Deep Reinforcement Learning to Manage Connected Autonomous Vehicles at Tomorrow's Intersections. *IEEE Transactions on Vehicular Technology*, 71(7):7033–7043, July 2022.
65. Yanjun Shi, Yuhan Qi, Lingling Lv, and Donglin Liang. A Particle Swarm Optimisation with Linearly Decreasing Weight for Real-Time Traffic Signal Control. *Machines*, 9(11):280, November 2021. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
66. Xiao-Cheng Liao, Yi Mei, and Mengjie Zhang. Learning Traffic Signal Control via Genetic Programming. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 924–932, July 2024. arXiv:2403.17328 [cs].