Sustainable Mobility Through Intelligent Traffic Signals: A Reinforcement Learning Approach to Emission Reduction and Vehicle Prioritization

Hussaini Aliyu Idris
Department of Physics, Informatics and Mathematics
University of Modena and Reggio Emilia
41125 Modena, Italy
hussaini.idris@unimore.it

Giacomo Cabri
Department of Physics, Informatics and Mathematics
University of Modena and Reggio Emilia
41125 Modena, Italy
giacomo.cabri@unimore.it

***Abstract*—Traffic congestion and vehicular emissions remain critical challenges in urban mobility. While reinforcement learning (RL) has shown promise in adaptive traffic signal control, conventional models may inadvertently encourage private vehicle use by merely reducing delay. In this study, we present a Q-learning-based traffic signal control framework enhanced with a vehicle prioritization mechanism for public transport and emergency vehicles. Implemented using the Simulation of Urban Mobility (SUMO), our approach is evaluated on a four-arm intersection scenario. Compared to fixed-time control, the standard Q-learning model achieves an 80% reduction in average vehicle delay and over 80% decrease in CO_2 emissions. The prioritized Q-learning variant further improves delay and emissions metrics while providing preferential treatment to high-impact vehicle categories. Crucially, this prioritization strategy helps incentivize public transport usage, mitigating the risk of increased private car dependence that often follows general congestion reduction efforts. Our results demonstrate that integrating vehicle prioritization into RL-based traffic control supports both sustainability and modal shift goals in intelligent transportation systems.**

***Index Terms*—Reinforcement learning, smart city, traffic signal control, sustainable mobility, intelligent transportation system**

I. Introduction

Traffic congestion and vehicular emissions are two of the most pressing challenges facing modern cities. As urban populations expand and private vehicle ownership continues to increase, traditional traffic signal systems, often based on fixed-time or actuated logic, struggle to adapt to the dynamic and heterogeneous nature of urban traffic flows. The result is increased travel time, air pollution, and inefficient use of road infrastructure [1], [2].

Sustainable urban mobility requires not only reducing congestion but also minimizing environmental impact and promoting equitable access to transport services. A key component of this is the ability to dynamically adapt traffic signal control to current traffic conditions while supporting broader sustainability goals, such as emission reduction and incentivizing high-occupancy vehicles (HOVs) and emergency services [3].

This work was supported by the European Marie Curie (MSCA) COFUND project FutureData4EU (Grant. Agreement n. 101126733) https://site.unibo.it/futuredata4eu/en

Reinforcement Learning (RL) has recently emerged as a powerful paradigm for intelligent traffic signal control [4]. Unlike rule-based systems, RL agents can learn optimal control policies through trial-and-error interactions with the environment, adapting to complex traffic dynamics [5]. However, most RL-based traffic signal control systems focus solely on minimizing vehicle delays or maximizing throughput, with little or no consideration for environmental impact or vehicle prioritization.

Furthermore, despite the societal importance of promoting public transport and ensuring rapid emergency response, few traffic control systems explicitly prioritize these high-importance vehicles in their learning objectives. This represents a significant gap, as the ability to prioritize public buses and emergency vehicles could not only improve service reliability but also encourage a modal shift away from private vehicles.

In this paper, we propose a Q-learning based adaptive traffic signal control framework that promotes sustainable mobility in two ways: First, it introduces a reward-shaping mechanism that prioritizes high-importance vehicles (specifically public buses and emergency vehicles) to improve their flow without deteriorating overall traffic performance, and second, it extends typical RL applications by incorporating emissions reduction as a direct evaluation criterion, unlike previous works that only consider traffic efficiency. Finally, unlike previous studies, our method explicitly addresses the potential rebound effect of congestion reduction by penalizing excessive reliance on private cars and encouraging the use of high-occupancy vehicles (HOVs).

The SUMO emission model, which tracks CO_2, CO, NO_x, HC, and PM_x based on vehicle class and behavior, was used to calculate emission outputs. Our RL agents aim to reduce pollutants indirectly by reducing vehicle idle times and congestion, achieving both environmental and operational goals.

The results show that our method not only reduces emissions and average vehicle delay but also promotes the use of high-occupancy transport modes while discouraging over-

dependence on private vehicles.

The remainder of this paper is organized as follows: Section II reviews related work. Section III details the methodology including state definition, reward design, and simulation setup. Section IV presents the experimental results, analysis and discusses implications and limitations. Section V concludes the paper and outlines future work.

II. Related work

Reinforcement Learning (RL) and Deep Reinforcement Learning (DRL) have emerged as promising solutions to the limitations of traditional static traffic light systems, offering adaptability to dynamic conditions and improvements in congestion and delay reduction [6], [7].

Several studies have explored signal control using DRL. The work in [8] proposed LIT, a lightweight and interpretable DRL model grounded in transportation theory, which demonstrated fast convergence and generalization to multi-intersection networks. Similarly, [9] introduced a Double DQN model with a 3-dimensional action space for phase control and showed superior performance over adaptive and fixed-time baselines.

In [10], a multi-agent RL framework was developed where each intersection operates as an independent agent, optimizing local policies based on traffic densities and signal phases. The study in [11] extended this with a multi-agent DQN approach using centralized learning and decentralized execution, improving throughput in oversaturated networks.

Other studies have incorporated vehicle prioritization or emission awareness. For example, in[12] coordinated signal control between two intersections using Q-matrix sharing, improving flow stability but suffering from traffic imbalance and scalability limitations.

Autonomous vehicle (AV)-centric traffic management is also gaining attention. In [13], a Multi-Agent RL system with communication between traffic lights and AVs improved delay through shared decision-making. Meanwhile, [14] used DDPG to eliminate traffic signals altogether, allowing AVs to make decentralized acceleration decisions. While effective at moderate AV penetration, these models assume ideal AV behavior and lack real-world validation.

Although the reviewed approaches demonstrate substantial improvements in metrics such as waiting time, queue length, and congestion, most remain focused primarily on traffic efficiency. While such gains are beneficial, they may inadvertently promote greater private vehicle use, potentially increasing overall emissions and compromising long-term environmental objectives. In contrast, our work introduces a Q-learning-based framework that integrates a vehicle prioritization mechanism for public buses and emergency services, explicitly evaluating both traffic performance and environmental impact to support broader goals of sustainable urban mobility.

III. Methodology

This section presents the methodology adopted for implementing Q-learning-based traffic signal control with priority handling for public transport and emergency vehicles. The approach is divided into the following stages: traffic network modeling, state and action space definition, reward design, Q-learning algorithm with priority integration, and simulation setup. The block diagram of the methodology adopted in this study is presented in Fig. 1

A. Traffic Network Design

We employed a four-arm signalized intersection as the experimental environment using the Simulation of Urban Mobility (SUMO) platform. The intersection consists of four incoming and four outgoing lanes with a total of 4 lane area detectors (e2_0 to e2_3), each placed on an inbound lane. The traffic light system (TLS) at the intersection is controlled via TraCI and follows a fixed phase sequence:

1) Green for North-South directions
2) Yellow for North-South
3) Green for East-West directions
4) Yellow for East-West

Each green phase lasts for 30 seconds, and each yellow phase for 5 seconds. The TLS operates under the constraint of a minimum green time (`MIN_GREEN_TIME`) to ensure traffic safety.

B. State Representation

The state space is defined as a tuple capturing the current phase index and the queue lengths at each of the four detectors. Queue length is computed as the number of vehicles halted for more than 0.1 seconds with speed below 0.1 m/s. Formally, the state at time t is given by Eqn 1:

$$s_t = [q_0, q_1, q_2, q_3, v_p, P_t] \tag{1}$$

where q_i represents the queue length at detector $e2\_i$, v_p represents the total number of stopped priority vehicles and $P_t \in \{0, 1, 2, 3\}$ denotes the current phase index.

C. Action Space

The action space consists of two discrete actions:

- **0**: Maintain the current traffic signal phase.
- **1**: Switch to the next phase in the predefined sequence.

A phase switch is permitted only if the current phase has elapsed for at least `MIN_GREEN_STEPS` to avoid unsafe and frequent toggling.

D. Reward Function Design

The reward function is formulated to penalize overall queue lengths and waiting of priority vehicles, encouraging the agent to clear traffic efficiently while prioritizing public transport and emergency vehicles. At each timestep t, the reward R_t is given by Eqn 2:

$$R_t = -\left(\lambda_q \cdot Q + \lambda_p \cdot P_v\right) \tag{2}$$

where:

- Q is the total number of stopped vehicles across all detectors,
- P_v is the number of stopped priority vehicles (buses and emergency vehicles)

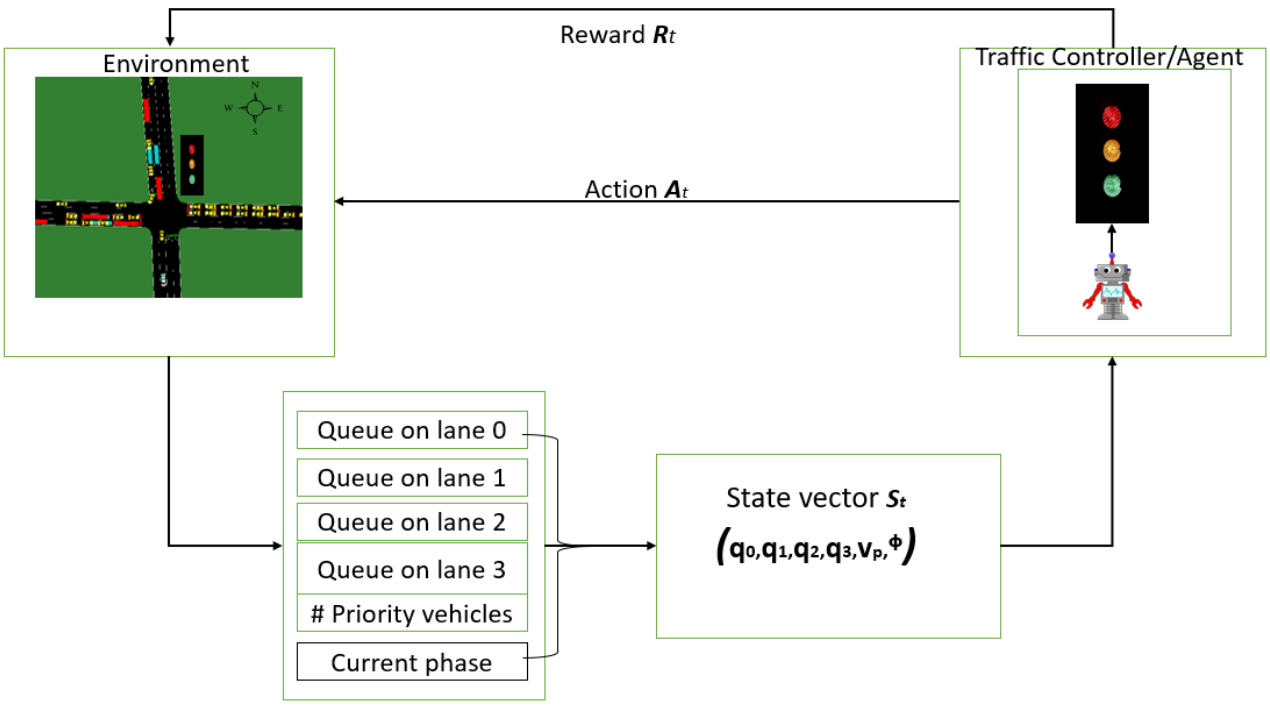

Fig. 1: Methodology Block Diagram

- λ_q = 5.0: penalty weight for stopped non-priority vehicles
- λ_p = 15.0: penalty weight assigned for stopped priority vehicle.

This design ensures that the agent is penalized more heavily for delaying high-priority vehicles, and the penalty weights are chosen empirically to refelect both traffic and policy priorities.

E. Q-Learning with Priority Integration

The agent learns an optimal policy using the Q-learning algorithm. A Q-table is maintained mapping each state-action pair to a Q-value. The update rule for Q-values is given in Eqn 3:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \Big[R_t + \gamma \cdot \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \Big] \tag{3}$$

where α is the learning rate and γ is the discount factor.

An ε-greedy exploration strategy is adopted, where the agent selects a random action with probability ε, and the best-known action otherwise. The value of ε is decayed over time to promote convergence toward optimal policies. Both the standard Q-learning algorithm and the algorithm for the priority-based Q-learning proposed in this study are presented in **Algorithm 1** and **Algorithm 2** respectively

F. Training Procedure

The agent is trained for 10,000 steps per episode across 100 episodes. The SUMO simulation is reset at the beginning of each episode. Each simulation step is synchronized with the Q-learning logic through the TraCI API, allowing real-time control and state observation.

During training, traffic is generated using random routes for passenger cars, buses, and emergency vehicles. The arrival probability and routing of vehicles are defined using route files (`.rou.xml`) with probability of 85%,10% and 5% for private cars, buses and emergency vehicles respectively. Priority vehicles are tagged with specific vehicle types to be recognized during reward computation. Table I shows the simulation parameters utilized in this paper.

TABLE I: Simulation Parameters in SUMO

Parameter	**Value**
Sim. Duration	3600 s
Step Length	1 s
TLS Strategy	Fixed, QL, QL+Priority
Junction Type	4-arm signalized
Phases	4 (G_NS, Y_NS, G_EW, Y_EW)
Phase Durations	30 s (G), 5 s (Y)
Lane Detectors	4 E2,
Veh. Types	Car, Bus, Emergency
Priority Class	Bus, Emergency
State Features	Queue_1-4, VP, Phase
Actions	2 (Keep, Switch)
Routing	Random trips
Emission Log	Enabled (CO_2, CO, NO_x, HC, PM_x)

G. Performance Metrics

To evaluate the system, we use the following key metrics:

- **Average waiting time** for all vehicles

- **Total CO_2 emissions** and other major air pollutants, extracted from SUMO's emission output files

Comparisons are made against a fixed-time baseline controller and a standard Q-learning controller without priority integration.

H. System Implementation

The system was implemented using Python and the SUMO traffic simulator. The TraCI interface enables communication between the Python agent and the simulation. The training and evaluation processes are conducted on Dell Precision 5520 with processor Intel (R) Core(TM) i7-7820HQ CPU @ 290.0GHz, 2901 Mhz 4 Cores,8 Logical Processors and 16GB RAM. All experiments were conducted using SUMO version 1.22.0.

Algorithm 1 Q-Learning for Traffic Signal Control (No Priority)

Initialize $Q(s, a)$ arbitrarily
for each episode **do**
  Reset SUMO environment
  while simulation not done **do**
    Observe state $s \leftarrow$ queue lengths + TLS phase
    Choose action a using ϵ-greedy policy
    Apply a (change TLS phase)
    Simulate for fixed steps
    Observe next state s', compute reward $r \leftarrow -5 \times$ total queue
    Update Q-value:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right]$$

    $s \leftarrow s'$
  end while
end for

IV. Result and Discussion

In this section, we analyze the performance of the proposed reinforcement learning-based traffic signal controllers compared to a traditional fixed-time approach, focusing on vehicle delay and environmental impact. Both Q-learning and Q-learning with priority strategies are evaluated.

Firstly, Fig 2 presents the average vehicle waiting times for the three control strategies. The fixed-time baseline exhibits the highest average delay of 43.56 seconds. By contrast, the Q-learning model reduces this delay to 8.81 seconds, representing a 79.8% decrease. Introducing vehicle prioritization to the Q-learning technique further reduces the delay to 6.54 seconds, resulting in an 85.0% reduction compared to fixed-time control and 25.8% reduction compared to the standard Q-learning technique. These improvements highlight the ability of reinforcement learning agents to optimize signal phasing dynamically, leading to shorter queues and smoother traffic flow. Moreover, the prioritized model demonstrates that integrating real-time preferences for high-importance vehicles does not degrade overall performance, but rather enhances it further.

Algorithm 2 Priority-Aware Q-Learning for Traffic Signal Control

Initialize $Q(s, a)$ arbitrarily
for each episode **do**
  Reset SUMO environment
  while simulation not done **do**
    Observe state $s \leftarrow$ queue lengths + TLS phase
    Choose action a using ϵ-greedy policy
    Apply a (change TLS phase)
    Simulate for fixed steps
    Check for halted priority vehicles (speed < 0.01)
    Compute reward:

$$r \leftarrow -\lambda_q \times \text{total queue} - \lambda_p \times \text{priority\_halted}$$

    if total queue < 0.5 **then**
      $r \leftarrow r + 5$ ▷ Bonus for low congestion
    end if
    Update Q-value:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right]$$

    $s \leftarrow s'$
  end while
end for

In addition, Fig 3 illustrates the trend of CO_2 emissions over time across all strategies. The fixed-time control strategy exhibits high and fluctuating emission rates throughout the simulation. In contrast, both RL-based strategies show rapid stabilization and consistently lower emission levels. Specifically, the Q-learning + Priority model displays the most efficient behavior, suggesting that reduced idling and smoother flow lead to fewer emissions.

On the other hand, Table II summarizes the total emissions for CO_2, CO, NO_x, HC, and PM_x. The fixed-time strategy results in the highest overall emissions, including 2.65 million kg of CO_2, 93,483 kg of CO, and 7,407 kg of NO_x. By contrast, the Q-learning model substantially reduces these values to 430,695 kg of CO_2, 10,937 kg of CO, and 1,490 kg of NO_x, respectively. The Q-learning + Priority model delivers the best performance in all pollutants, with further reductions observed: 397,548 kg of CO_2, 9,336 kg of CO and 1,363 kg of NO_x. These improvements are primarily attributed to the reduced vehicle idle times and improved traffic throughput enabled by RL-based optimization.

These findings confirm that the reinforcement learning-based controllers not only enhance traffic flow but also contribute significantly to sustainability objectives by minimizing vehicular emissions. Prioritizing high-importance vehicles (public buses and emergency vehicles) in the Q-learning + Priority model does not degrade performance; instead, it yields additional benefits for both mobility and the environment. Interestingly, the prioritization of high-importance vehicles not

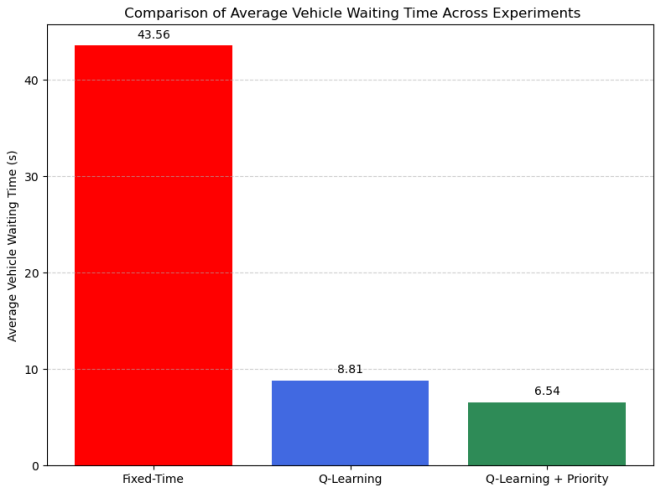

Fig. 2: Average Waiting Time Comparison Per TLS Control method

only benefits those vehicles but also contributes to smoother signal phase transitions and reduced idling across the network. We hypothesize that the learned policies implicitly favor phase selections that align better with natural traffic surges (e.g peak bus periods), thereby improving flow for all. A breakdown of emissions by vehicle type and phase transition analysis is a promising area for future investigation.

TABLE II: Total Emissions and Percentage Reduction Compared to Fixed-Time

Pollutant	**Fixed-Time**	**Q-Learning**	**Q-Learning + Priority**
CO_2 (kg)	2,654,000	430,695(-83.8%)	397,548 (-85.0%)
CO (kg)	93,483	10,937 (-88.3%)	9,336 (-90.0%)
NO_x (kg)	7,407	1,490 (-79.9%)	1,363 (-81.6%)
HC (kg)	1,066	288 (-73.0%)	261 (-75.5%)
PM_x (kg)	237	57 (-75.9%)	52 (-78.1%)

V. Conclusion

A. Summary

In this study, we presented a comparative evaluation of fixed-time, standard Q-learning, and priority-based Q-learning traffic signal controllers using SUMO. The results clearly demonstrate the effectiveness of reinforcement learning in significantly reducing average vehicle waiting times and harmful emissions. Notably, the Q-learning model with prioritization outperforms both fixed-time and standard Q-learning models, offering not only improved traffic efficiency but also meaningful environmental benefits.

Beyond technical improvements, a key advantage of the prioritization strategy lies in its potential to influence urban mobility behavior. While reducing congestion alone is beneficial, it can paradoxically lead to increased private vehicle use due to the improved driving conditions. This rebound effect undermines long-term sustainability goals. By contrast, our prioritized RL model explicitly favors public transport and emergency vehicles, creating a system-level incentive structure that discourages over-reliance on private cars.

In conclusion, this work advances the application of reinforcement learning in traffic control by aligning optimization with broader policy objectives. It offers a viable path toward smarter, greener and more equitable urban transport systems that prioritize sustainability alongside operational performance.

B. Limitations and Future Work

Although this work focuses on a single intersection, real-world traffic systems involve complex networks of interdependent signals. Coordination among intersections becomes essential to prevent traffic spillback and ensure network-wide efficiency. Extending our approach to a multi-agent framework where each intersection learns locally but communicates globally could enable scalable and decentralized optimization. Future work will investigate federated and cooperative RL strategies for corridor-wide deployment, enabling both efficiency and sustainability at city scale.

Acknowledgment

This work was supported by the European Marie Curie (MSCA) COFUND. FutureData4EU (Grant Agreement n. 101126733) co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. https://site.unibo.it/futuredata4eu/en.

References

[1] A. G. Sims and K. W. Dobinson, "The sydney coordinated adaptive traffic (scat) system philosophy and benefits," *IEEE Transactions on vehicular technology*, vol. 29, no. 2, pp. 130–137, 1980.

[2] P. Hunt, D. Robertson, R. Bretherton, and M. C. Royle, "The scoot on-line traffic signal optimisation technique," *Traffic Engineering & Control*, vol. 23, no. 4, 1982.

[3] M. Bertogna, P. Burgio, G. Cabri, and N. Capodieci, "Adaptive coordination in autonomous driving: Motivations and perspectives," in *2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, IEEE, 2017, pp. 15–17.

[4] N. Kodama, T. Harada, and K. Miyazaki, "Traffic signal control system using deep reinforcement learning with emphasis on reinforcing successful experiences," *IEEE Access*, vol. 10, pp. 128 943–128 950, 2022.

[5] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1.

[6] M. Miletić, E. Ivanjko, M. Gregurić, and K. Kušić, "A review of reinforcement learning applications in adaptive traffic signal control," *IET Intelligent Transport Systems*, vol. 16, no. 10, pp. 1269–1285, 2022.

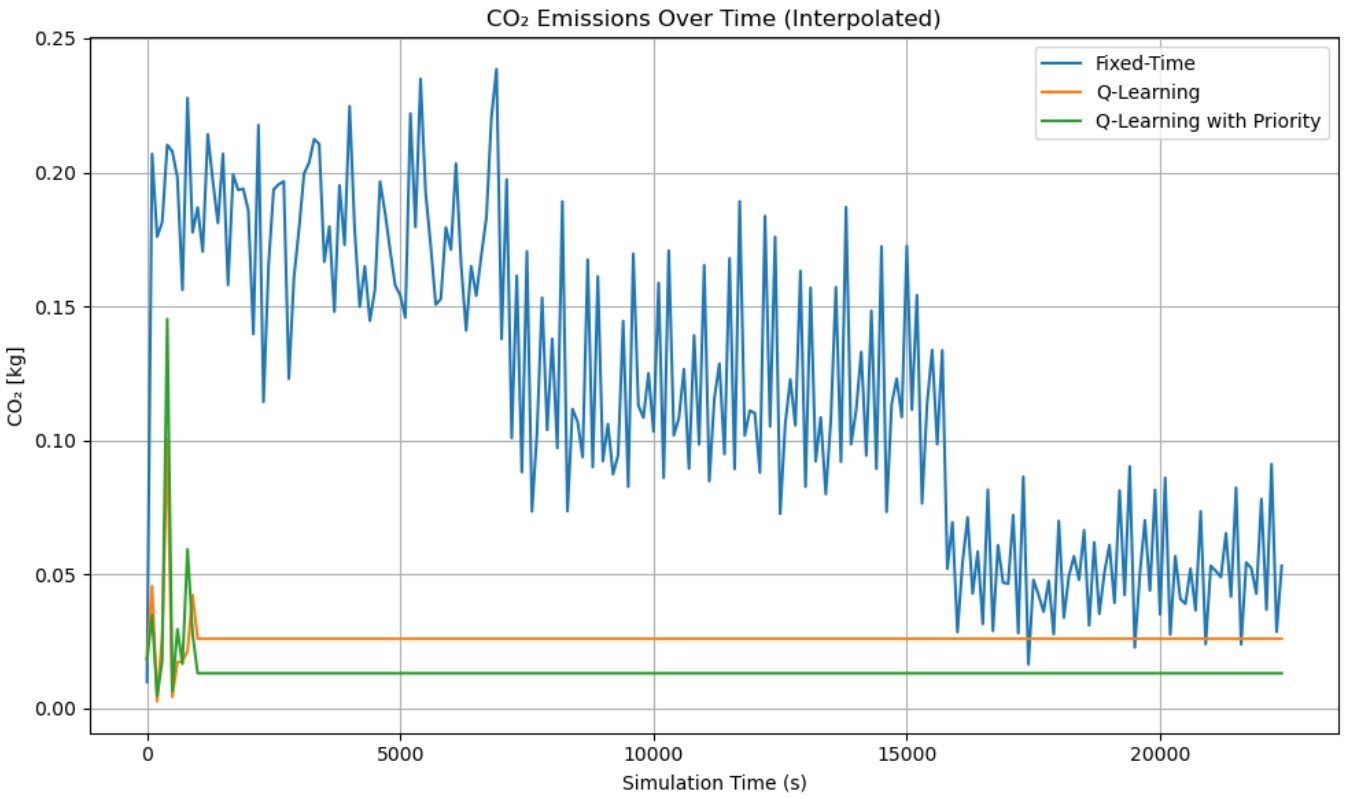

Fig. 3: Average CO2 Emission Over Simulation Steps

[7] H. Wei, G. Zheng, V. Gayah, and Z. Li, "A survey on traffic signal control methods," *arXiv preprint arXiv:1904.08117*, 2019.

[8] G. Zheng, X. Zang, N. Xu, *et al.*, "Diagnosing reinforcement learning for traffic signal control," *arXiv preprint arXiv:1905.04716*, 2019.

[9] Y. Bao-Lin, C. Dong, W. Peng, W. Weimin, L. Li, and C. Bin, "A Traffic Signal Control Method Based on Improved Deep Reinforcement Learning," in *2024 China Automation Congress (CAC)*, IEEE, 2024, pp. 5959–5964.

[10] A. Agafonov and A. Yumaganov, "Agent-based traffic signal control using a reinforcement learning approach," in *2021 International Conference on Information Technology and Nanotechnology (ITNT)*, IEEE, 2021, pp. 1–4.

[11] E. E. Mon, H. Ochiai, and C. Aswakul, "Application of Traffic Light Control in Oversaturated Urban Network Using Multi-Agent Deep Reinforcement Learning," *IEEE Access*, 2024.

[12] J. Reswara, N. Sutisna, I. Syafalni, and T. Adiono, "Q-Learning Algorithm with Double-Agent Reinforcement Learning for Smart Traffic Controller," in *2023 IEEE 66th International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, 2023, pp. 649–653.

[13] Q. Wu, P. Zhi, Y. Wei, *et al.*, "Communicate with traffic lights and vehicles based on multi-agent reinforcement learning," in *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, IEEE, 2021, pp. 843–848.

[14] Z. Hu, X. Wang, and Z. Wang, "Enhancing Urban Intersections Management Through Deep Reinforcement Learning: Superior Control of Autonomous Vehicles in Mixed Traffic Flow," in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2024, pp. 3309–3314.