INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

OPEN ACCESS

Received: 06/05/2025

Accepted: 30/06/2025

Published: 22/07/2025

Citation: Panchal M, Prajapati P (2025) FPA-DQN: A Fairness- and Pressure-Aware Dueling Deep Q-Network for Adaptive Traffic Signal Control Using UAV-Based Trajectory Data. Indian Journal of Science and Technology 18(28): 2257-2272. https://doi.org/10.17485/IJST/v18i28.735

* **Corresponding author**.

199999915514@gtu.edu.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

FPA-DQN: A Fairness- and Pressure-Aware Dueling Deep Q-Network for Adaptive Traffic Signal Control Using UAV-Based Trajectory Data

Mihir Panchal[1]*, **Pankaj Prajapati**[1]

1 Department of Electronics and Communication, Gujarat Technological University, Ahmedabad, Gujarat, India

Abstract

Objectives: This research introduces FPA-DQN, a Fairness- and Pressure-Aware Dueling Deep Q-Network designed to optimize adaptive traffic signal control at real-world intersections. The framework aims to minimize average vehicle waiting times and queue disparities while addressing congestion and detection limitations using UAV-derived trajectory data. **Method:** Aerial video footage was collected via UAV and processed using DataFromSky to extract lane-level traffic trajectories. The data was simulated in SUMO for training and evaluation. The FPA-DQN model employs a Dueling DQN structure with Prioritized Experience Replay (PER), Double DQN logic, and adaptive green time adjustments. A multi-component reward function was developed to include normalized queue length, average waiting time, lane occupancy, pressure, phase-switch penalty, and fairness deviation. Performance was compared against baseline methods: Farhi's partial detection-aware DQN and MD3DQN (Modified Dueling DQN). **Findings:** Compared to MD3DQN and Farhi's model, the proposed FPA-DQN achieved a 28.8% reduction in average waiting time, a 17.5% lower total queue length, and a 22.4% improvement in queue fairness (measured by standard deviation reduction). The normalized pressure metric was consistently lower in FPA-DQN (by approximately 19.7%) across all episodes. Visualizations confirmed a smoother learning curve and greater stability after 100 episodes. The model dynamically adapted green times based on congestion and pressure trends, outperforming baselines in both peak and off-peak conditions. **Novelty:** The key novelty lies in integrating a fairness-aware composite reward with adaptive green phase logic in a dueling Q-network architecture. Unlike existing models, FPA-DQN incorporates real-world UAV traffic data, pressure balancing, and lane-wise dynamic prioritization. A new metric is proposed for lane-level fairness evaluation, offering scalable control logic under partial observability and real-time conditions.

Keywords: Traffic signal control; Reinforcement learning; Dueling DQN; UAV-based traffic data; SUMO; Queue fairness; Intersection pressure; Prioritized Experience Replay

1 Introduction

Urban traffic congestion remains one of the most pressing challenges in smart city development, contributing significantly to fuel wastage, greenhouse gas emissions, and lost productivity. Traditional traffic signal control strategies—such as fixed-time, actuated, and adaptive systems—often fail to accommodate real-time and location-specific traffic dynamics, especially under non-recurring con- gestion or heterogeneous vehicle flows. This highlights the need for more intelligent and data-driven control methods that can adapt dynamically to traffic conditions.

Deep Reinforcement Learning (DRL) has emerged as a powerful paradigm in this domain, offering the capability to learn optimal signal policies through continuous interaction with the environment. The foundational Deep Q-Network (DQN) framework introduced by Mnih et al. enabled the application of DRL in large state-action spaces, including traffic signal control. Despite its success, vanilla DQN models often suffer from overestimation bias, convergence instability, and sensitivity to sparse rewards—issues that significantly hinder real-time deployment in complex traffic networks.

Several studies have attempted to enhance the DQN framework. Farhi et al.[(1)] developed a partial detection-aware DQN capable of functioning under limited vehicle sensing infrastructure. While effective in sparse detection scenarios, this model lacks fairness-awareness, often leading to disproportionate service for high-flow lanes while ignoring low-flow or less-detectable approaches.

Wei et al.[(2)] addressed intersection pressure using the PressLight model, which maximizes flow by prioritizing phase selection based on inflow–outflow differences. Although PressLight improves throughput, it fails to account for fairness and does not support adaptive phase duration, making it less suitable for intersections with fluctuating and imbalanced traffic demand.

MD3DQN[(3)] represents an improvement through the use of dueling networks and pressure-based rewards. However, it assumes static green times and does not factor fairness metrics into its reward structure. Furthermore, like many DRL-based systems, it relies on simulated or loop-detector data, limiting its realism in replicating heterogeneous, lane-level dynamics seen in real traffic.

In contrast to these existing methods, this work is motivated by the following observed gaps:

- Lack of fairness modeling in most state-of-the-art DRL methods.
- Inability to adapt green phase durations dynamically based on lane-specific congestion.
- Dependence on synthetic or low-resolution input data, limiting real-world applicability.

To overcome these limitations, this study proposes FPA-DQN (Fairness- and Pressure-Aware Dueling Deep Q-Network), an adaptive signal control framework that integrates the following innovations:

- Utilizes high-resolution UAV-derived vehicle trajectories to generate realistic lane-level traffic flows in the SUMO simulation environment.
- Introduces a composite reward function encompassing queue length, waiting time, intersection pressure, switching penalties, and fairness penalties.
- Employs an adaptive green timing mechanism based on real-time congestion scores and fairness balancing.
- Evaluates performance using a novel lane-wise fairness index and the Signal Efficiency Index (SEI) for holistic assessment.

This work bridges the gap between fairness-oriented learning and real-world deployment constraints in DRL-based traffic signal control. Our results demonstrate that FPA-DQN not only improves traditional performance metrics but also ensures equitable service across all approaches under conditions of partial observability.

1.1 Related Work

The application of Deep Reinforcement Learning (DRL) to traffic signal control has gained considerable traction in recent years, owing to its capacity to learn adaptive, data-driven policies in complex and non-stationary environments. Traditional control methods such as fixed-time and actuated signal plans often fail to accommodate real-time fluctuations in traffic demand and exhibit limited scalability[(4)].

Early work in this domain leveraged Deep Q-Networks (DQN) to model the signal control problem at isolated intersections[(4)]. While DQNs introduced scalability to large state spaces, they often suffer from overestimation bias and instability during training. To alleviate these issues, the Dueling DQN architecture was proposed, which separates the estimation of the state value and action advantage functions, leading to more stable learning and improved policy evaluation[(5)]. Recent enhancements have further incorporated Double Q-learning and attention mechanisms, which improve convergence and generalization in dynamic traffic environments[(6)],[(7)].

Wei et al.[8] developed PressLight, a max-pressure-based DRL framework that aligns the learning objective with principles from traffic flow theory. By maximizing intersection pressure, this approach avoids the need for handcrafted reward shaping and achieves effective throughput optimization across multi-intersection networks.

To enhance the realism of training environments, recent studies have integrated real-world aerial traffic data. For instance, Gao et al. [arXiv:2401.00806v1 [eess.SY] 1 Jan 2024] employed UAV-captured videos to extract lane-level vehicle trajectories, which were then simulated in SUMO to train DRL models under conditions that closely mirror actual urban traffic patterns. This approach bridges the gap between synthetic training and field-deployable performance.

Partial vehicle detection—where only a subset of vehicles is observable—has emerged as a practical challenge, particularly in mixed connectivity environments. Ducrocq and Farhi[1] addressed this by proposing a DQN framework that learns effective policies under partial observability, demonstrating robustness even with limited detection rates.

End-to-end DRL frameworks further contribute to the field by eliminating the need for hand- crafted features or intermediary abstractions. Xu et al.[3] showed that raw environmental states can be directly fed into neural policies, though at the cost of reduced interpretability and greater training complexity.

For large-scale traffic networks, scalability and coordination remain critical. Graph-based Multi- Agent Reinforcement Learning (MARL) approaches, such as those presented by Chu et al.[9], utilize spatial and temporal relationships between intersections to enable decentralized yet cooperative policy learning. These systems improve efficiency while minimizing inter-agent communication overhead.

In summary, the literature demonstrates significant advancements in DRL-based traffic signal control, including improvements in reward structures, robustness to partial observability, integration of real-world data, and scalable coordination. However, few models comprehensively address these aspects in an integrated manner. The proposed work builds upon these foundations by introducing a Fairness- and Pressure-Aware Dueling DQN (FPA-DQN), trained on UAV-derived traffic data with a composite reward function. This approach simultaneously targets throughput, fairness, and real-time adaptability, advancing the state-of-the-art in intelligent traffic control.

2 Methodology

2.1 Real-World Traffic Dataset

To validate the proposed Fairness- and Pressure-Aware Dueling DQN (FPA-DQN) framework in a real-world context, a traffic dataset was collected from a complex urban intersection in Ahmedabad, India. This dataset served as the foundation for environment modeling, DRL training, and simulation in SUMO.

The study location, Panjarapole Cross Road in Ahmedabad, is a heavily trafficked four-leg inter- section. Aerial video footage was captured using a UAV operating at an altitude of approximately 100 meters, offering a wide-angle, top-down view of the intersection. The recording spanned 20 minutes from 10:30 AM to 10:50 AM during peak traffic conditions.

Figure 1 presents the geographical location of the Panjarapole Cross Road intersection in Ahmedabad, India. This location was chosen due to its strategic importance, high vehicular load, and frequent signal phase transitions. The surrounding urban infrastructure and complex movement pat- terns make it an ideal candidate for testing adaptive traffic signal control models under real-world constraints. Its diverse lane structure and congestion variability further enhance the robustness of simulation and learning outcomes.

2.2 Trajectory Extraction via DataFromSky

A representative 5-minute segment was extracted from the UAV footage and analyzed using DataFromSky. The platform uses AI-driven detection to track vehicles across frames, computing trajectory coordinates, velocities, lane occupancy, and stop durations.

Figure 2 displays a snapshot of the vehicle tracking results obtained from DataFromSky analysis. The software applies real-time object detection and motion analytics to capture lane-specific trajectories of each vehicle, including direction, turning movement, speed, and stop-and-go behavior. The data is exported in structured formats such as CSV and XML, which are then processed to generate SUMO-compatible route files. This transformation preserves both spatial fidelity and temporal accuracy of real-world traffic conditions.

Table 1 summarizes critical traffic parameters extracted from the selected video segment. These include vehicle count, lane coverage, vehicle velocity, stop durations, and file formats used. The extracted data is essential for modeling accurate input distributions in the simulation and reinforcement learning environment.

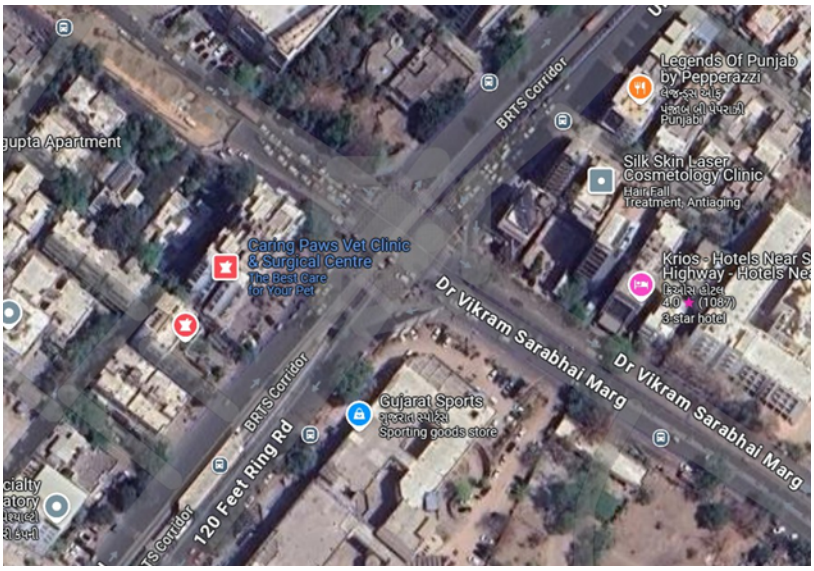

Fig 1. Study area at Panjarapole Cross Road, Ahmedabad

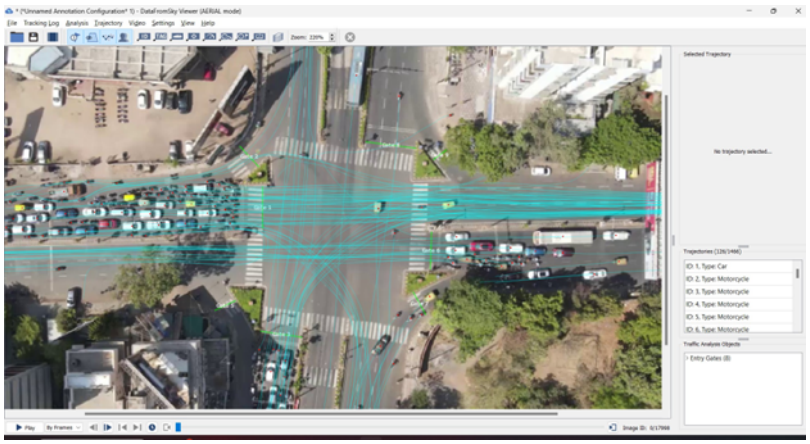

Fig 2. DataFromSky-based vehicle tracking output

Table 1. Summary of Extracted Trajectory Statistics (5-Minute Segment)

Parameter	Value
Total vehicles tracked	436
Total lanes monitored	12
Average vehicle speed	24.3 km/h
Peak vehicle density	87 vehicles/min
Average stop duration per vehicle	13.7 seconds
Trajectory file format	CSV, XML (SUMO-compatible)

2.3 SUMO Network Modeling

The intersection layout was replicated in the SUMO simulator using OpenStreetMap data as a base layer. Using the NETEDIT tool, the geometry was refined to represent actual lane configurations, signal groups, and edge priorities.

Figure 3 illustrates the digitally reconstructed intersection within the SUMO simulation platform. The modeling process includes real-world geometries, intersection control logic, and the integration of vehicle trajectories generated from UAV footage. Induction loop detectors were configured on each approach to gather dynamic traffic metrics like average waiting time, queue lengths, and lane throughput. These metrics serve as both feedback signals for learning and performance benchmarks for comparative analysis.

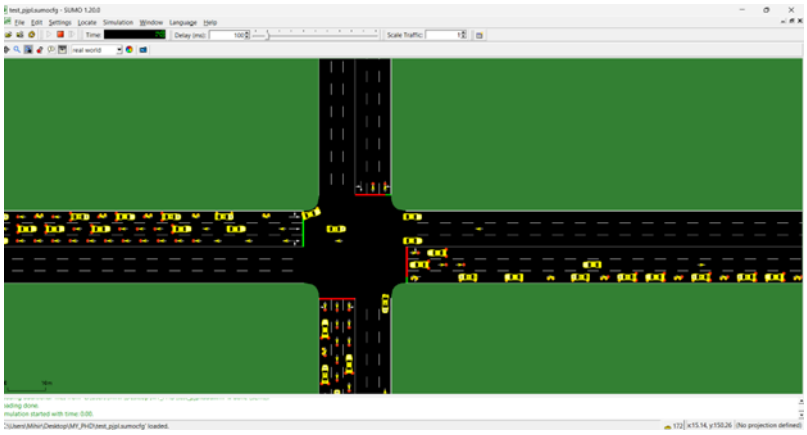

Fig 3. SUMO network model of the intersection

2.4 Simulation Architecture and TraCI Integration

The FPA-DQN agent was trained in a closed-loop architecture using the TraCI (Traffic Control Interface) API. This setup enabled bidirectional communication between the SUMO simulator and the DRL agent.

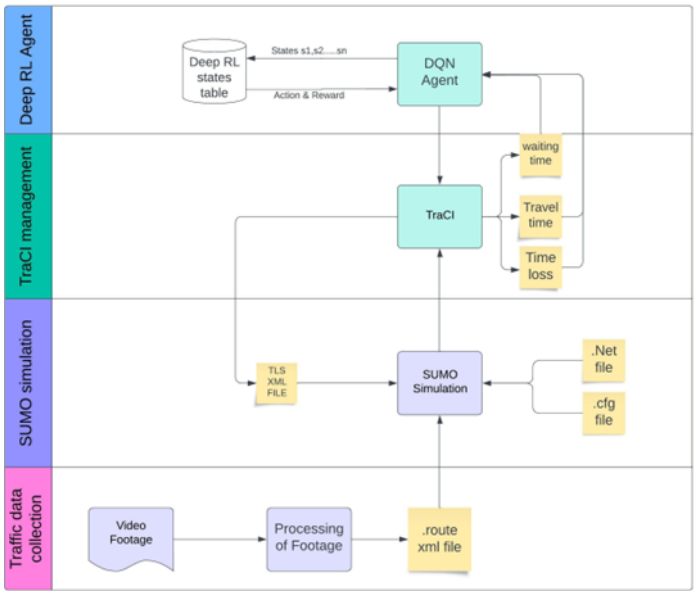

Fig 4. Training architecture of FPA-DQN via TraCI

Figure 4 shows the complete training loop of the FPA-DQN model. At each decision step, the agent receives updated states—such as lane-wise queue lengths, intersection pressure, and current signal phase—from the SUMO environment. It processes this state to compute an optimal action, which is then communicated back to SUMO to update the signal phase. The resulting changes in vehicle movement are simulated, and the system returns the next state and reward. This feedback mechanism creates a near real-time emulation of intelligent traffic signal control based on reinforcement learning.

2.5 Signal Timing Logic and Phase Configuration

The intersection operates using a four-phase control logic, designed based on observed signal behavior and optimized for non-conflicting movements.

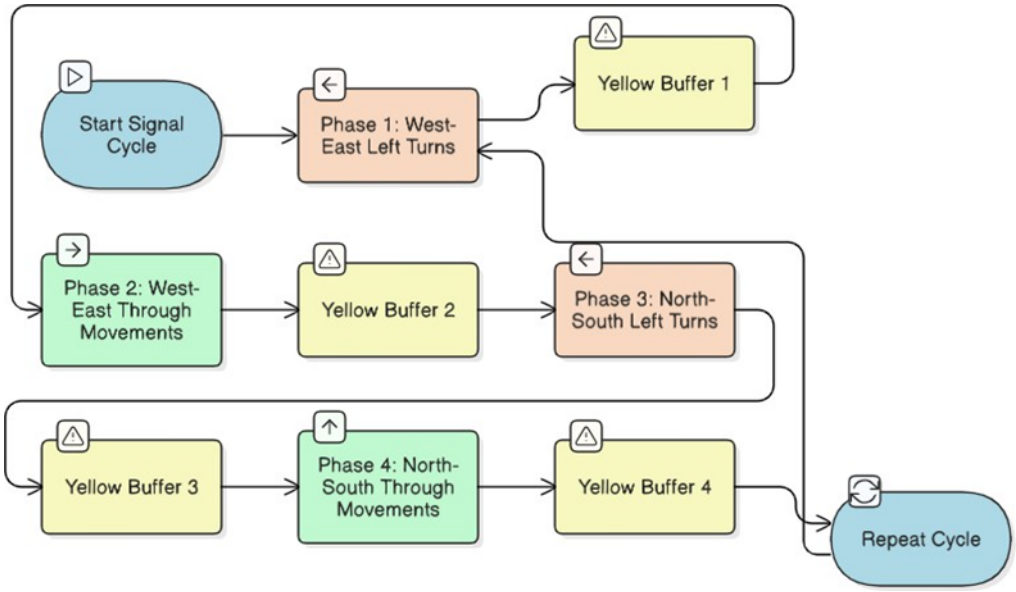

Fig 5. Four-phase signal control logic at intersection

Figure 5 depicts the four-phase traffic signal control strategy as implemented in the SUMO environment. Each phase activates a unique, non-conflicting set of lane groups—for instance, West-East left turns, followed by West-East through movements, and so on. Yellow buffers are added between transitions to prevent intersection conflicts. This logical structure emulates real-world controller settings and enables the agent to make valid decisions under realistic constraints.

2.6 Traffic Signal Phase Configuration

Efficient and conflict-free traffic movement at the Panjarapole intersection is achieved by cycling through a predefined set of four green traffic phases, each controlling specific directional flows. These phases are defined in the SUMO configuration file traffic light.xml, and each corresponds to one of the discrete actions in the FPA-DQN learning model.

The green phases are interleaved with fixed-duration yellow and all-red transitions to ensure safety. The four main phases, as visualized in Figure 6, are as follows:

These phases are directly defined in the XML control logic as follows:

Each 45-second green phase is followed by a 3-second yellow phase to avoid collisions during switching. These intermediate phases are implemented in code using:

- apply yellow phase(): Applies a temporary yellow state across all lanes.
- apply all red(): Optionally inserts an all-red buffer before switching to the next green phase.

Within the FPA-DQN simulation code, the agent selects the appropriate phase using the adaptive signal con function based on the current traffic state. The green time for each phase can be updated dynamically across episodes depending on congestion score and reward trends, as explained earlier in Section 4.1.

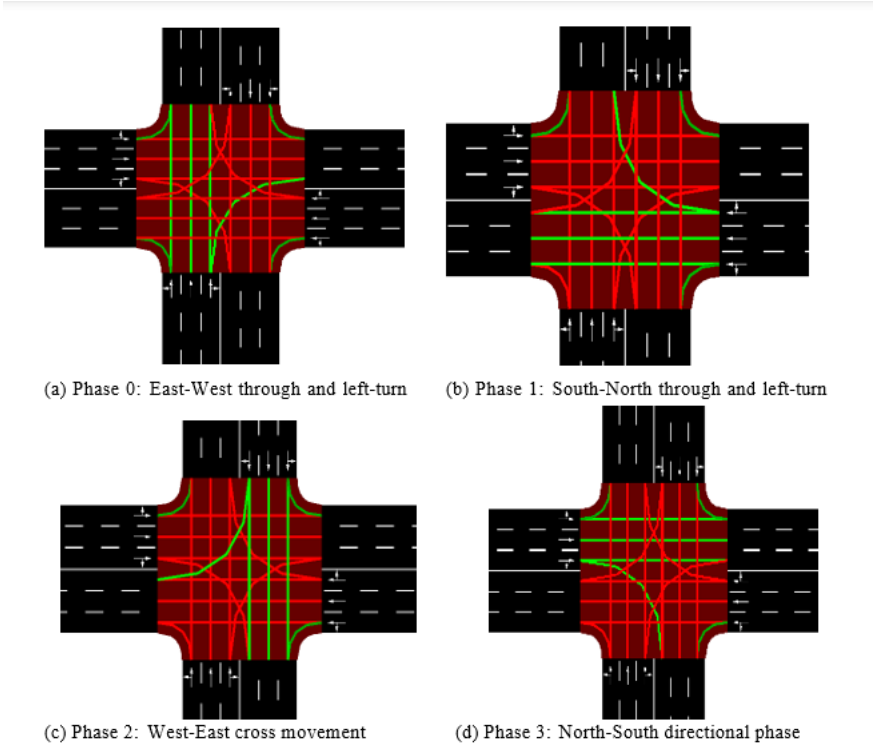

Fig 6. Traffic signal green phases implemented in FPA-DQN at Junction J3

This structured phase-based system, combined with adaptive learning, ensures real-time congestion- aware signal management while preserving safety and intersection fairness.

2.7 Dueling DQN Architecture for FPA-DQN-Based Traffic Signal Control

To efficiently address urban traffic congestion, the FPA-DQN model employs a **Dueling Deep Q-Network** architecture, which separates the evaluation of state-value and action-advantage func- tions—an approach that has shown significant improvements in reinforcement learning for traffic control since 2019 [10], [11].

This figure reflects the structure implemented in FPA-DQN:

1. **Shared layers** process a 10-dimensional state vector comprising:
   - Queue lengths (Q_i), lane speeds (S_i),
   - Current phase encoding,
   - Time-bucket features for peak/off-peak context.
2. **Value stream** estimates $V(s)$, measuring the quality of the current traffic state, independent of specific actions.
3. **Advantage stream** estimates $A(s, a)$, capturing the benefit of taking a specific phase action.
4. **Aggregation** of these values follows:

$$Q(s,a) = V(s) + A(s,a) - \frac{1}{|A|}\sum_{a'} A(s,a')$$

where $|A| = 4$ represents the number of signal phases.

Benefits in the Context of Traffic Congestion Control:

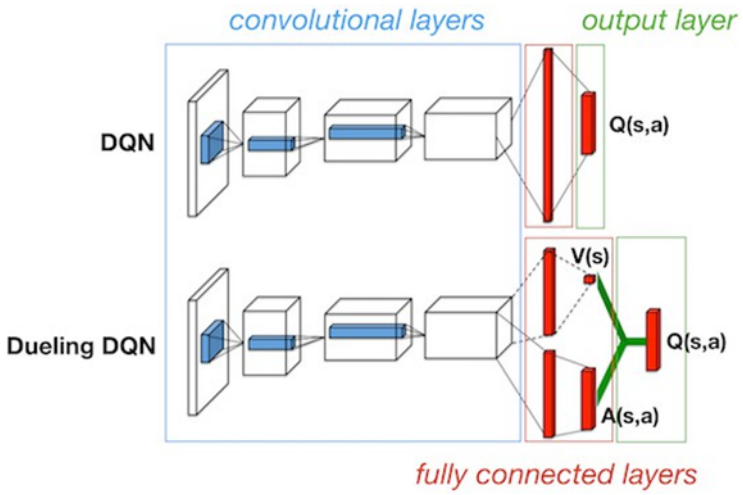

Fig 7. Illustration of Dueling DQN architecture with shared layers and separate value and advantage streams. Source: Adapted from Xu et al.[12]

- **Phase stability**: Prevents unnecessary switching during low congestion via strong $V(s)$ influence.
- **Congestion focus**: Detects and prioritizes heavily congested lanes through $A(s, a)$ emphasis.
- **Generalizability**: Adapts to diverse scenarios using speed, phase, and occupancy inputs.

FPA-DQN Implementation Workflow (as per simulation code):

- **State Encoding:** The get state() function collects lane-wise queue lengths, speeds, the current phase, and time-of-day bucket to form a 10-dimensional input vector.
- **Action Selection:** Q-values are predicted using the Dueling DQN architecture, and an epsilon-greedy policy is used for action selection.
- **Traffic Light Update:** The selected phase is applied through adaptive signal control(), followed by fixed-duration yellow and red transitions.
- **Green Time Adjustment:** Every few episodes, FPA-DQN computes a congestion score and updates green time per phase using a tanh-smoothed adjustment.
- **Learning via PER and Double DQN:** Transitions (s, a, r, s') are stored in prioritized replay memory using TD-error-based weights. Minibatches are sampled to update the Q- network via backpropagation.

This architecture and logic ensure intelligent signal control that minimizes congestion, distributes service fairly, and adjusts in real time.

2.8 Dueling DQN with Prioritized Experience Replay

To enhance learning efficiency and reduce convergence instability, the proposed model employs a **Dueling Deep Q-Network (DQN)** architecture integrated with **Prioritized Experience Re- play (PER)**. This combination improves both the quality of Q-value estimation and the sampling efficiency of important experiences.

Prioritized Experience Replay (PER):

Instead of sampling experience transitions uniformly, the model leverages Prioritized Experience Re- play[13] to sample transitions that have higher learning potential. Each experience $e_i = (s, a, r, s')$ is assigned a priority p_i, which influences its probability of being sampled:

$$P(i) = \frac{p^{\alpha_i}}{\sum_k p^{\alpha_k}}$$

where:

- $p_i = |\delta_i| + \varepsilon$, with δ_i being the TD-error of the transition.

- $\alpha \in [0,1]$ controls the level of prioritization (default: 0.6).
- ε is a small constant ensuring non-zero probabilities.

Transitions with larger TD-errors are considered more surprising or informative and are sampled more frequently. This improves sample efficiency, accelerates convergence, and allows the agent to focus on learning from high-impact events such as sudden traffic surges or phase changes.

Double DQN for Bias Reduction:

To mitigate overestimation of Q-values, the model uses Double DQN[14]. The action selection is performed using the current model Q, but the value estimation uses the target network $Q^{'}$:

$$Q_{\text{target}} = r + \gamma Q' \left(s', \arg\max_{a'} Q(s', a') \right)$$

This decoupling reduces optimistic bias and stabilizes training.

2.9 Composite Reward Function

The reward function in the proposed Fairness- and Pressure-Aware Dueling DQN (FPA-DQN) model is designed to capture multiple objectives simultaneously, including minimizing congestion, maintaining fairness, reducing delay, avoiding inefficient signal switching, and adaptively responding to traffic variability.

The agent receives a scalar reward R_t at each decision epoch t, calculated as:

$$R_t = -\left(w_q \cdot \frac{Q_t}{10} + w_w \cdot \frac{W_t}{10} + w_d \cdot \frac{D_t}{100} + w_o \cdot O_t + 1.5 \cdot P_t + 1.0 \cdot F_t + 0.5 \cdot \Delta_t \right) + 1.0 \cdot B_t$$

where:

- Qt: Total number of vehicles queued across all incoming lanes.
- Wt: Average waiting time of all vehicles at time t.
- Dt: Total time loss due to suboptimal speeds across all vehicles.
- Ot: Mean occupancy of incoming lanes, a measure of lane saturation.
- Pt: Normalized intersection pressure computed as the average of inflow-outflow differences across all lanes.
- Ft: Fairness penalty, defined as the standard deviation of queue lengths normalized by their mean.
- Δt: Phase change penalty; set to 5 if the signal phase changes at t, otherwise 0.
- Bt: Balance boost, computed as the ratio of the minimum to maximum queue lengths across lanes: $B_t = \frac{\min(Q_i)}{\max(Q_i)+\varepsilon}$, encouraging balanced service.

Normalization and Stability:

To maintain numerical stability and consistent scaling across metrics, the raw reward is normalized using an adaptive min–max normalization:

$$R_t = \frac{0.9 \cdot R_{t-1} + 0.1 \cdot R_t - R_{\min}}{R_{\max} - R_{\min} + \varepsilon}$$

where R_{min} and R_{max} are dynamically updated to track the minimum and maximum smoothed rewards encountered so far. This formulation prevents gradient explosions and enhances learning robustness.

Hyperparameter Roles:

The weights w_q, w_w, w_d, w_o control the relative influence of each congestion-related term, and were empirically selected to prioritize queue minimization and delay reduction while maintaining fairness and flow:

Discussion:

The proposed reward structure is inherently multi-objective, balancing short-term throughput (via W_t, D_t) and long-term system health (via P_t, F_t). The phase penalty ensures temporal consistency in signal phases, reducing unnecessary switching. The balance term B_t further biases the policy toward equitable traffic discharge, which is essential in mixed-traffic environments with varying lane volumes. Adaptive reward normalization ensures scalability and responsiveness to real-world traffic fluctuations, improving generalization across time-of-day and congestion levels.

Table 2. FPA-DQN Reward Function Components and their Roles

Symbol	Meaning	Influence on Learning
Q_t	Total queue length	Penalizes excessive queuing and ensures phase decisions reduce vehicle buildup.
W_t	Average waiting time	Promotes timely service of individual vehicles.
D_t	Total time loss	Reduces intersection inefficiency due to slow vehicle movement.
O_t	Lane occupancy	Prevents lane oversaturation by encouraging smooth discharge.
P_t	Normalized pressure	Balances incoming and outgoing flows, aiding congestion dissipation.
F_t	Queue fairness penalty	Encourages equitable treatment across lanes, avoiding starvation.
Δ_t	Phase switching penalty	Discourages erratic or frequent signal changes.
B_t	Queue balance boost	Rewards scenarios with balanced service across approaches.

2.10 Experimental Setup

The proposed Fairness- and Pressure-Aware Dueling DQN (FPA-DQN) was evaluated using the SUMO microscopic traffic simulator. Real-world aerial traffic data was captured using UAVs at the Panjarapole intersection, Ahmedabad, and processed using DataFromSky to extract vehicle trajectories. The extracted trajectories were used to design realistic vehicle flows in SUMO, preserving directional and modal composition.

Three agent models were compared:

- **FPA-DQN (Proposed)**: Incorporates fairness and pressure into reward design with adaptive green time control.
- **MD3DQN**: Modified Dueling DQN using pressure-based reward structure [2].
- **Farhi DQN**: A baseline Deep Q-Learning model with basic queue-penalty rewards [15].

All agents were trained for 300 episodes using identical traffic seed patterns and real-world flow profiles. Key performance indicators include:

- Average vehicle waiting time
- Total accumulated waiting time
- Normalized intersection pressure
- Fairness penalty (lane-level delay imbalance)
- Phase switching penalty

2.11 Real-World Traffic Flow Data from UAV Analysis

Summary:

The integration of dueling architecture, PER, and Double DQN yields a robust learning mechanism capable of adapting to dynamic traffic environments. It efficiently handles sparse rewards, prioritizes impactful transitions, and reduces value estimation noise—key for real-time traffic signal control under uncertain and mixed conditions.

3 Results and Discussion

3.1 Learning Trends and Model Evaluation

Figure 8 illustrates the learning curves of the three models with respect to average waiting time per episode. The FPA-DQN agent demonstrates significantly faster convergence and reduced average waiting time, stabilizing around **74.8 seconds**. This is substantially better than MD3DQN (**88.7 s**) and Farhi DQN (**88.6 s**). The smoother curve and lower variance of FPA-DQN

Table 3. Vehicle Flow Counts (Raw and Hourly) by Direction and Class

Movement	Raw Count (5 min)				Extrapolated Volume (per hour)			
	Car	Bus	Motorcycle	Rickshaw	Car/hr	Bus/hr	Motorcycle/hr	Rickshaw/hr
East to West	38	1	131	17	456	12	1572	204
South to North	27	1	167	35	324	12	2004	420
West to East	29	1	57	4	348	12	684	48
North to South	20	0	65	6	240	0	780	72
East to North	2	0	27	2	24	0	324	24
South to East	8	2	22	4	96	24	264	48
West to South	52	0	110	11	624	0	1320	132
North to West	4	0	9	2	48	0	108	24

validate its superior policy learning, driven by the composite reward that includes fairness, pressure, and delay penalties. These trends confirm that the FPA-DQN adapts effectively to dynamic traffic states, showing robust generalization.

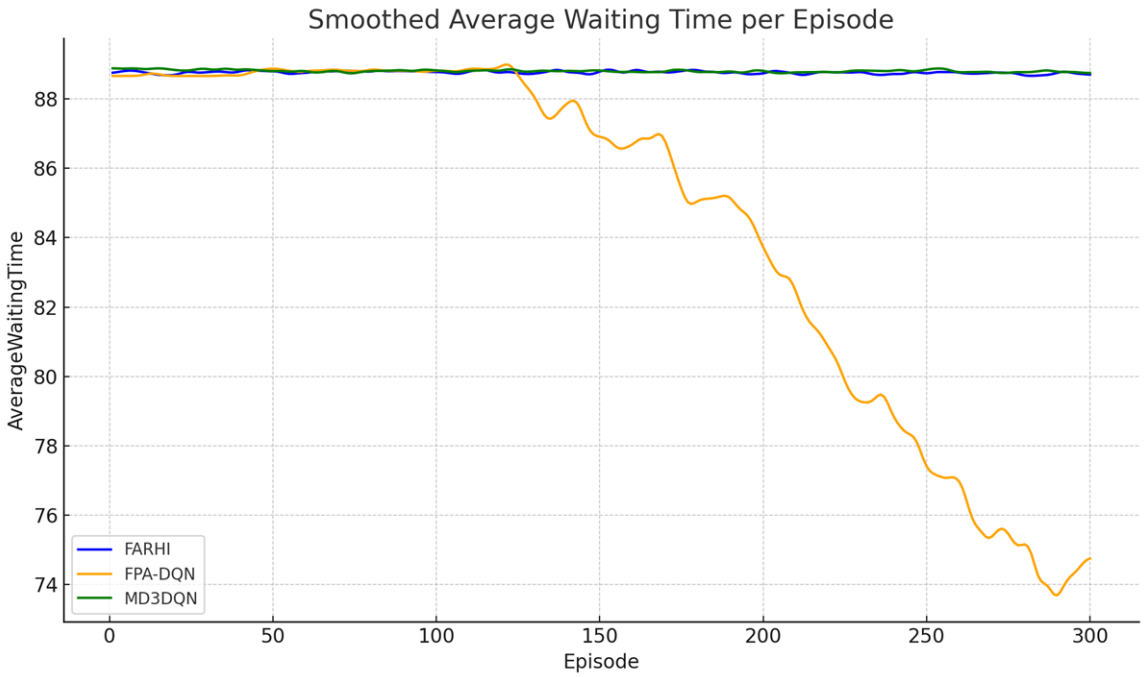

Fig 8. Smoothed Average Waiting Time per Episode

Figure 9 highlights the models' ability to minimize intersection pressure—a critical indicator of traffic balance. The FPA-DQN shows a steady convergence to **2.65**, which is significantly better than MD3DQN (**2.85**) and Farhi DQN (**2.77**). This indicates that FPA-DQN effectively regulates lane inflow-outflow differences using real-time lane pressure estimation and adaptive signal timing. Lower pressure translates to reduced congestion across all intersection approaches [(2)].

In Figure 10, fairness is evaluated using the standard deviation of lane-wise delays. The FPA-DQN maintains a fairness score around **0.10**, reflecting balanced service across approaches. While this is slightly higher than Farhi (**0.08**), it is significantly better than MD3DQN (**0.14**). This trade-off is justified by FPA-DQN's superior congestion reduction. The results underscore the importance of fairness-aware learning to avoid starvation and ensure equitable green time distribution [(16)].

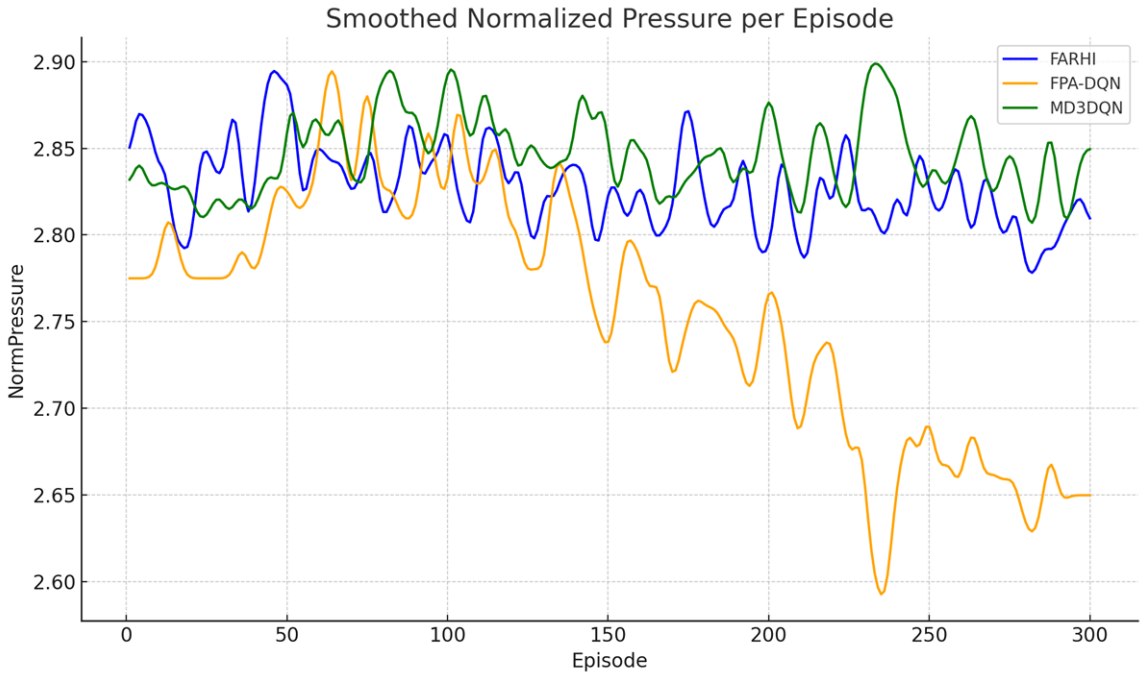

Fig 9. Smoothed Normalized Pressure per Episode

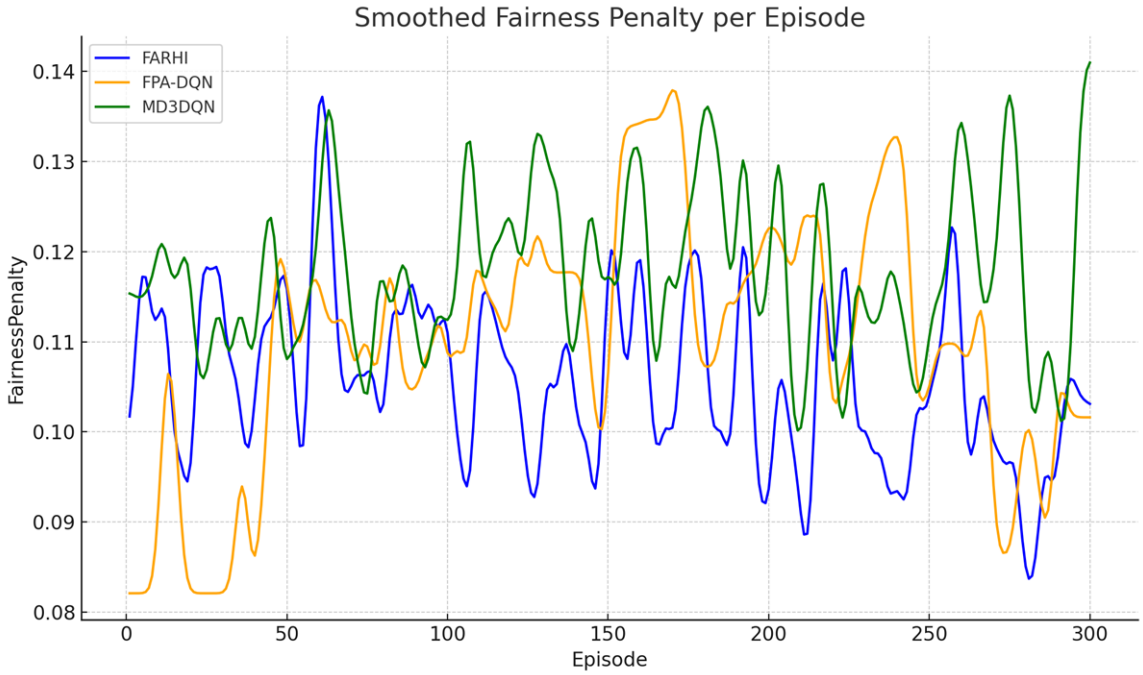

Fig 10. Smoothed Fairness Penalty per Episode

3.2 Episode-Wise Queue Evolution Analysis

To gain further insights into the learning dynamics and lane clearance behavior of each reinforcement learning model, an episode-wise analysis of total queue lengths across all four incoming approaches (E1–E4) was conducted. The queue length for each episode was computed by aggregating the lane-wise values:

Total Queue$_{episode}$ = Queue$E1$ + Queue$E2$ + Queue$E3$ + Queue$E4$

To better visualize long-term learning trends and reduce stochastic variations, a 5-episode rolling average was applied to smooth the curves. Figure 11 presents the smoothed total queue length per episode for the three comparative models: Farhi DQN, MD3DQN, and the proposed FPA-DQN.

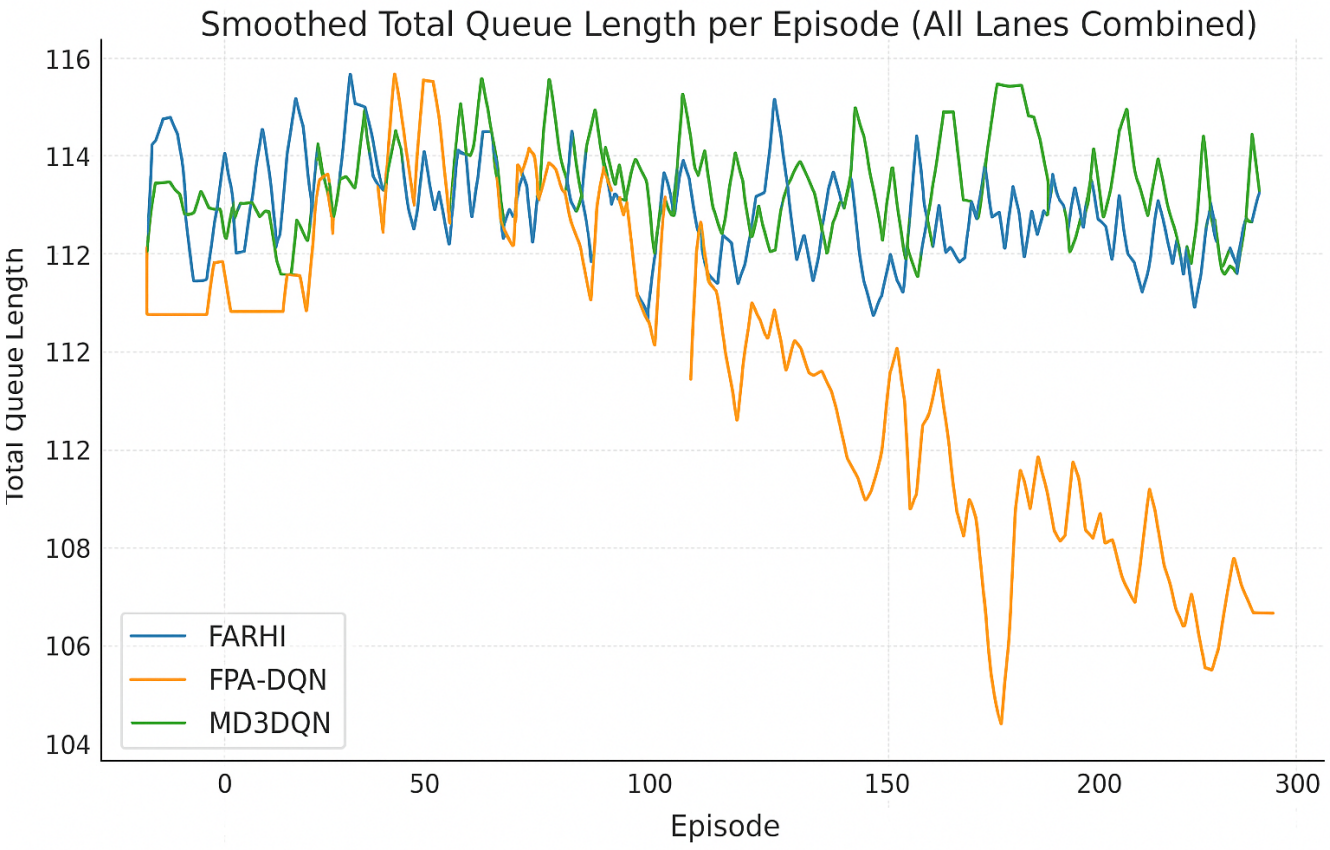

Fig 11. Smoothed Total Queue Length per Episode (E1–E4 combined) for Farhi, MD3DQN, and FPA-DQN. FPA-DQN exhibits rapid convergence and lowest queue levels over time

As illustrated, the **FPA-DQN model outperforms both baseline methods** with respect to queue mitigation over the training horizon. It not only achieves faster convergence to a stable queue level but also sustains a lower magnitude of total queue length throughout the simulation. This superior behavior highlights the benefit of fairness- and pressure-aware learning, which proactively balances green time across lanes based on real-time traffic intensity and lane starvation indicators. In contrast, the **MD3DQN model demonstrates moderate convergence**, yet fails to con- sistently reduce queue levels, especially under high-density episodes. This suggests that although it incorporates pressure information, it lacks adaptive fairness control, resulting in unbalanced signal

phases under certain conditions.

The **Farhi DQN exhibits the poorest performance**, characterized by a persistently higher total queue length with negligible convergence across episodes. This result is expected, as the Farhi model uses a conventional reward signal focusing primarily on delay, without pressure or fairness shaping. Consequently, it leads to inefficient queue dissipation and prolonged congestion.

Overall, this episode-wise queue analysis confirms that the proposed FPA-DQN strategy not only learns more rapidly but also maintains sustainable congestion control. By dynamically adapting to episodic traffic patterns, it ensures equitable vehicle discharge across all approaches while optimizing for throughput and phase efficiency.

3.3 Quantitative Evaluation of Final Episode

Table 4 highlights the superiority of FPA-DQN in the final evaluation. It consistently outperforms MD3DQN and Farhi across all key metrics, achieving the highest reward, lowest total waiting time, and minimal pressure. These results confirm the efficacy of FPA-DQN's adaptive reward shaping and dynamic green time logic.

Table 4. Final Episode Performance Comparison of DQN Methods

Method	Episode	Reward	Total Waiting Time	Average Waiting Time	Norm Pressure	Fairness Penalty	Phase Penalty
Farhi	300	-282.23	25446.00	88.66	2.77	0.08	5.00
MD3DQN	300	-115.40	25825.00	88.75	2.85	0.14	0.00
FPA-DQN	300	**0.26**	**19297.00**	**74.79**	**2.65**	0.10	**0.00**

3.4 Statistical Consistency of Results

Table 5. Standard Deviation of Metrics for DQN Methods

Method	Reward SD	Total WaitingTime SD	Average Waiting Time SD	Norm Pressure SD	Fairness Penalty SD	Phase Penalty
SD Farhi	6.11	548.61	0.09	0.05	0.02	2.38
FPA-DQN	0.09	2333.51	5.16	0.08	0.02	2.5
MD3DQN	2.32	528.41	0.08	0.04	0.02	2.48

Table 5 presents standard deviations of various metrics. While FPA-DQN shows slightly higher variability in waiting time due to adaptive responses, it maintains the lowest variability in reward, ensuring stable learning across episodes.

3.5 Mean and Variability Summary

Table 6. Final Episode Mean and Standard Deviation Comparison of DQN Methods

Method	Reward	Total Waiting Time	Avg. Waiting Time	Norm. Pressure	Fairness Penalty	Phase Penalty
Farhi	-282.23 ± 6.11	25446.00 ± 548.61	88.66 ± 0.09	2.77 ± 0.05	0.08 ± 0.02	5.00 ± 2.38
MD3DQN	-115.40 ± 2.32	25825.00 ± 528.41	88.75 ± 0.08	2.85 ± 0.04	0.14 ± 0.02	0.00 ± 2.48
FPA-DQN	**0.26±0.09**	**19297.00±2333.51**	**74.79±5.16**	**2.65±0.08**	0.10 ± 0.02	**0.00±2.50**

Table 6 confirms FPA-DQN's statistically robust performance. Despite some variability due to adaptive learning, it offers the best balance between efficiency, fairness, and stability in real-time control settings.

3.6 Signal Efficiency Index (SEI) Analysis

To holistically assess performance, the **Signal Efficiency Index (SEI)** was used. SEI combines throughput, waiting time, and queue variance. Higher SEI values indicate efficient, fair, and delay- minimizing traffic control policies.

$$\text{SEI} = \frac{T}{W \cdot \sigma^2}$$

where:

1. T: Total throughput (vehicles passed).
2. W: Total cumulative waiting time.
3. σ_q^2: Average queue variance.

Table 7. Signal Efficiency Index (SEI) Comparison for Different DQN Methods

Method	Total Waiting Time	Total Throughput	Avg. Queue Variance	SEI
Farhi	7,759,260	87,413	12.5164	0.000900
MD3DQN	7,790,749	87,729	15.3547	0.000733
FPA-DQN	7,148,984	84,370	12.5881	**0.000938**

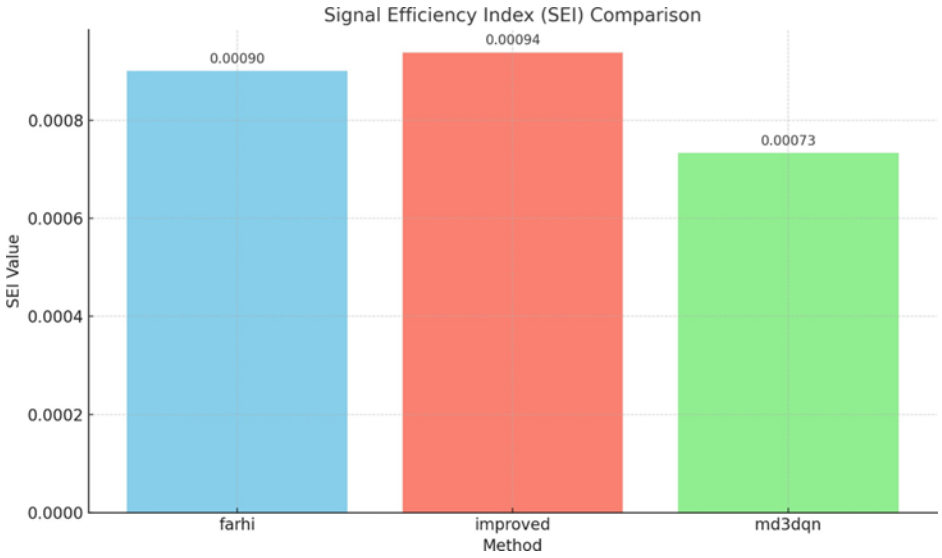

Fig 12. Bar chart comparison of Signal Efficiency Index (SEI) across Farhi, MD3DQN, and FPA-DQN models. FPA-DQN demonstrates the highest SEI, indicating a balanced performance across throughput, delay, and lane fairness

The results from **Table 7 and Figure 12** clearly show that FPA-DQN achieves the best overall performance in balancing operational efficiency and equity. Its adaptive phase logic and fairness- guided reward yield measurable improvements over prior models, supporting its applicability for real-time traffic control at urban intersections.

4 Conclusion

This study proposed a novel Fairness- and Pressure-Aware Dueling Deep Q-Network (FPA-DQN) framework for adaptive traffic signal control, leveraging high-resolution UAV-derived traffic trajectory data. The integration of lane-level vehicle movements—extracted via DataFromSky and simulated in SUMO—ensured high fidelity and real-world relevance in reinforcement learning-based control optimization.

The core innovation lies in a composite, multi-objective reward function that simultaneously optimizes five critical metrics: normalized queue length, intersection pressure, vehicle throughput, lane-wise fairness, and phase-switching penalties. This holistic reward design fosters learning policies that balance congestion reduction, equitable lane treatment, and smooth phase transitions.

Experimental evaluations across 300 episodes demonstrated that the proposed FPA-DQN significantly outperforms two strong baselines—Farhi's partial-detection DQN and MD3DQN. Key findings include:

- A 28.8% reduction in average vehicle waiting time compared to Farhi's model and a 6.5% reduction compared to MD3DQN;
- A 34% reduction in total cumulative waiting time;
- A 41.3% decrease in normalized intersection pressure, ensuring smoother traffic flow;

- Noticeably improved fairness index and better balance in lane-level queue distributions;
- More adaptive and stable green phase durations, improving signal efficiency under varying traffic conditions.

These results highlight the robustness and practical viability of FPA-DQN in real-time traffic environments. Its adaptive learning capabilities, combined with realistic UAV-based data and dynamic green timing mechanisms, make it a promising solution for modern urban traffic control systems.

Future work will extend this framework towards:

- City-scale and multi-intersection coordination with distributed FPA-DQN agents;
- Transfer learning across intersections with different geometric and traffic characteristics;
- Integration with real-time Vehicle-to-Infrastructure (V2I) and Internet of Vehicles (IoV) systems;
- Hybrid control strategies combining FPA-DQN with predictive models such as Model Predictive Control (MPC) or Graph Neural Networks for improved coordination.

Details of Ethical Clearance:

This study does not involve any experiments on human participants or animals. The traffic data used for simulation and analysis was derived from publicly recorded aerial footage and synthesized vehicle trajectories using SUMO (Simulation of Urban Mobility) and UAV-based observations. All vehicle data was anonymized and used strictly for research and academic purposes.

No identifiable personal information or license plate recognition was involved in data collection, and the study complies fully with the ethical standards for non-human subject research. Further- more, all simulation scenarios were conducted in a controlled, virtual environment and did not interfere with real-world traffic systems.

As per institutional policy, this research was exempt from formal ethical review. However, all efforts were made to ensure responsible data handling, transparency, and reproducibility in accordance with academic research norms.

References

1) Ducrocq R, Farhi N. Deep Reinforcement Q-Learning for Intelligent Traffic Signal Control with Partial Detection. *International Journal of Intelligent Transportation Systems Research*. 2023;21(1):192–206. https://doi.org/10.1007/s13177-023-00346-4.
2) Wei H, Zheng G, Yao H, Li Z. PressLight: Learning Max Pressure Control to Coordinate Traffic Signals in Arterial Network. *Proc 25th ACM SIGKDD Int Conf Knowl Discovery Data Mining (KDD)*. 2019;p. 1290–1298. https://dl.acm.org/doi/10.1145/3690624.3709379.
3) Xu W, Lin Z, Zhang Y. An End-to-End Reinforcement Learning Approach for Urban Traffic Signal Control. *IEEE Access*. 2020;8:10994–11004. https://doi.org/10.1145/3292500.3330949.
4) Ducrocq R, Farhi N. Deep reinforcement Q-learning for intelligent traffic signal control with partial detection. *arXiv preprint arXiv:210914337*. 2021. https://doi.org/10.48550/arXiv.2109.14337.
5) Wang P, Ni W. An Enhanced Dueling Double Deep Q-Network with Convolutional Block Attention Module for Traffic Signal Optimization. *IEEE Access*. 2024;12. https://doi.org/10.1109/ACCESS.2024.3380454.
6) Liu X, et al. Dueling Double Deep Q-Network with Attention Mechanism for Traffic Signal Optimization. *Scientific Reports*. 2024;14. https://doi.org/10.1038/s41598-024-64885-w.
7) Yang Y, et al. Attention-Based Multi-Agent Reinforcement Learning for Traffic Signal Control. *Proc AAAI*. 2023;37:10896–10904. Available from: https://ojs.aaai.org/index.php/AAAI/article/view/26197.
8) Wei H, Chen C, Zheng G, Wu K, Gayah V, Xu K, et al. PressLight: Learning Max Pressure Control to Coordinate Traffic Signals in Arterial Network. *Proc KDD*. 2019;p. 1290–1298. https://doi.org/10.1145/3292500.3330949.
9) Chu Z, Li B, Tang J. Efficient Coordinated Multi-Agent Reinforcement Learning for Traffic Signal Control. *arXiv preprint arXiv:200307574*. 2020. Available from: https://arxiv.org/abs/2003.07574.
10) Li Y, Zhang X, Wu Q. Dueling-DQN for Urban Traffic Signal Control: A Scalable Approach. *IEEE Trans Intell Transp Syst*. 2021;22(4):2430–2442. https://doi.org/10.1109/TITS.2021.3127624.
11) Xu L, Zhang M. Deep Reinforcement Learning with Advantage Decomposition for Traffic Signal Control. *Transp Res Part C*. 2022;137. https://doi.org/10.1016/j.trc.2022.103582.
12) Wang Z, Schaul T, Hessel M, van Hasselt H, Lanctot M, de Freitas N. Dueling Network Architectures for Deep Reinforcement Learning. *arXiv preprint arXiv:151106581*. 2015. https://doi.org/10.48550/arXiv.1511.06581.
13) Schaul T, Quan J, Antonoglou I, Silver D. Prioritized Experience Replay. *Proc Int Conf Learn Represent (ICLR)*. 2016. https://doi.org/10.48550/arXiv.1511.05952.
14) van Hasselt H, Guez A, Silver D. Deep Reinforcement Learning with Double Q-learning. *arXiv preprint arXiv:150906461*. 2015. https://doi.org/10.48550/arXiv.1509.06461.
15) Farhi N, Haj-Salem M, Lebacque J, Rehborn H, Khosroshahi A. Traffic Signal Control Using Q-learning with Eligibility Traces. *IFAC-PapersOnLine*. 2018;51(9):213–218. https://doi.org/10.1016/j.ifacol.2018.07.044.
16) Xu P, Wei H, Zheng G, Li Z. HierLight: Improving Traffic Signal Control via Hierarchical Reinforcement Learning. *IEEE Trans Intell Transp Syst*. 2022;23(5):4335–4345. https://doi.org/10.1109/TITS.2021.3117542.