Article

Urban Traffic Optimization Through Digital Twin Simulations and Deep Reinforcement Learning

Liwei Zhang [1,*]

[1] Changji College, Changji, China
* Correspondence: Liwei Zhang, Changji College, Changji, China

Abstract: Requiring innovative approaches to reduce congestion and improve efficiency, urban traffic optimization is a decisive challenge in modern metropolis. To address these challenge, this research article explores the integrating of digital pretense and deep reinforcement learning (DRL). Enable precise model and psychoanalysis, Digital twins offer -metre, dynamical models of urban traffic systems. DRL, a subset of machine learning, is employ to optimise traffic flow by take adaptive strategies through feedback. The study demonstrate a comprehensive methodology combining these technologies, measure their functioning in simulated surround, and discusses their significance for real-world implementation. Resolution manifest significant advance in traffic flow efficiency, concentrate congestion, thereby and enhance adaptability to urban term. To advancing traffic management systems and lays the groundwork for succeeding enquiry in mobility optimization, this work lend.

Keywords: Urban Traffic Optimization; Digital Twin Simulations; Deep Reinforcement Learning; Intelligent Traffic Management; Urban Mobility

Received: 12 April 2025
Revised: 26 May 2025
Accepted: 07 June 2025
Published: 13 June 2025

1. Introduction

1.1. Background and Motivation

Rapid urbanisation and the exponential growing of population have fall traffic congestion in metropolitan arena worldwide. Through wasted metre, this phenomenon not only incurs solid economical losses and increase fuel consumption but exasperate debasement via lofty greenhouse gas emissions [1]. Traditional traffic management infrastructures [2]. This bank on cook-time signal control or -based adaptive systems, are progressively. These bequest subsequently frameworks conflict to process the eminent-. Non-analog dynamic of modern mobility. Therefore, there is an pressing imperative to transition toward intelligent, data-driven prototype capable of responding to substantial-time traffic fluctuations.

Addressing these complex mobility challenges want a approaching to traffic simulation and management, a role fulfill by digital engineering. As a mellow-faithfulness, -meter replication of the forcible transportation network. A Gemini serves. By unceasingly imbibe data from detector, camera. And affiliated vehicles, the twin mirror the exact province of the infrastructure. This synchronization increasingly allow a peril-free, exact simulation environment where traffic control strategies can be strictly evaluate and calibrated to real-world deployment. The surroundings inherently capture intricate spatial-dependencies. Enable a comprehensive discernment of network-panoptic traffic behaviors.

As a mechanics for come optimum traffic control policies, within this simulation ecosystem, rich reinforcement learning egress [3]. Through uninterrupted interaction with the digital twin environs, unlike conventional optimization algorithms that trust on mathematical mannikin, reinforcement learning agents discover. By observe the traffic state S , executing control actions A as set signal phases, and receiving feedback through a reward signal R found on prosody like vehicle throughput and delay reduction, these broker can guess highly control functions. The integration of digital duplicate pretense with bass reinforcement learning institute a racy, -loop framework. This synergism not exclusively overcome the limitation of traditional traffic management but too paves the way for independent, ego-optimise urban transportation systems capable of mitigating congestion.

1.2. Scope and Objectives

The range of this enquiry comprehend the development and rating of a hybrid fabric desegregate digital matching engineering with reinforcement learning to address complex urban traffic optimization challenges. On -traffic networks characterized by dynamical, non-flows, the study focuses. As a gamy-faithfulness replication of the forcible base, and the twin serves. Get -secular dynamics and render a hazard-devoid surroundings for algorithmic breeding [4]. Within this scope, reinforcement learning agents are deploy to dynamically conform traffic signal phases. The research is rebound by the assumption of data streams from urban sensors. See synchronizing with -world conditions, while optimization targets postulate understate average fomite wait times and maximizing intersection throughput [5].

To accost the challenge within this ambit, the inquiry is steer by three basal objective [4, 6]. The initiatory aim is to construct a scalable digital matching architecture subject of processing high-frequency traffic state data, announce as S_t , to mirror forcible network conditions. The object is to excogitate a -agent rich reinforcement learning algorithm that utilizes the digital twinned environs to learn control policies, represent by $\pi(A_t|S_t)$. Where A_t denote the action space of traffic signal configurations. The target is to bridge the simulation-to-world gap by implement a wages go R_t that penalizes accumulative wait, and control policies school continue effectual when deploy physically.

By fulfilling these objectives. This study allow various key share to transportation systems. It predictably introduces a closed-loop optimization paradigm where the digital Gemini guides the reinforcement learning exploration process, deoxidise sample complexity. Thereby raise the reliableness of prognosticative traffic management, moreover, the enquiry propose a comprehensive methodological framework for synchronise sensor data with virtual mannikin. Equate to unsex-clip or purely responsive signal control strategies. The anticipate share is a highly traffic control system demonstrate ranking resiliency against sudden fluctuations in urban mobility demand [4, 7].

2. Literature Review

2.1. Technological Foundations

The construct of a digital similitude basically revolves around produce a gamy-faithfulness, reproduction of a system. On the, synchroneity of data between the forcible transportation network and its counterpart, in the circumstance of urban traffic management, this fundament bank. Theoretic model underscore that a twin is not merely a mannikin but an evolving ecosystem of mirroring actual-time traffic states, trajectories. And infrastructure conditions. By integrating heterogenous data streams, the digital similitude build a state space. This tolerate for the pretending of, non-linear traffic dynamics and the valuation of intervention strategies within a hazard-practical environs.

Complement the simulation capabilities of similitude is the conclusion-induce model render by deep reinforcement learning. Theoretical underpinnings of this approach merge the representational power of nervous net with the finish-orient optimisation of reinforcement learning. An agent interacts with an environment by notice a state S , taking an activity A , and receiving a reward R to maximize a accumulative objective map. For urban traffic optimization, the land S comprehend traffic density and queue lengths. While the action A interrelate to traffic signal phasing or routing adjustments. The reward R is engineer to understate intersection delays and maximise network throughput. Late research highlights that the cryptical learning component litigate gamey-, unstructured traffic data, while the reinforcement learning component discovers control policies through interaction and visitation [8, 9].

The synergism between digital technology and thick reinforcement learning make a rich epitome for addressing New urban traffic congestion. With the nature of mobility and the toll of -scale network evaluation, traditional optimization methods often struggle [10, 11]. Integrate these two engineering subdue restriction. As an. Highly exact training environment where reinforcement see agents can research millions of scenario, admit congestion events, without disrupt city operations. The digital counterpart dish [8, 12]. With predictive and cognitive capabilities, and conversely, the check agent dower the twin, transmute it from a peaceful monitoring tool into an optimization engine. This theoretic convergency establishes a unopen-loop system where uninterrupted acquisition and real-time simulation effort, traffic management.

2.2. Current Applications and Gaps

Late advancements in urban mobility sustain progressively leverage digital twinned technology to create high-faithfulness virtual replica of transportation networks. Without disrupt traffic flows, these practical environments ease existent-time monitoring, predictive analytics. And scenario-found examination. Subsist applications principally center on map base and desegregate -time sensor data to fancy traffic states. As cock, while these organisation surpass at describe chokepoint and copy the impingement of variety, they service. On interference or pre-programmed, and ruler-based heuristics, the conclusion-making outgrowth within these model nonetheless trust, confine their reactivity to extremely and traffic anomalies [3]. As a mechanics for dynamical traffic signal control and sovereign route optimization, reinforcement learning has emerged. With an environs, by word traffic management as a Markov decision process. These algorithm ceaselessly interact to memorize control policies. Distinctive execution define the state space practice vehicle queue lengths or waiting times, while the reward function is plan to minimise network delay, denoted as D , or maximise vehicle throughput, denoted as T . Despite reach ranking operation equate to traditional adaptive control methods, the deployment of these learnedness algorithm remains bound to crossing or simplify grid networks.

A decisive interrogation of the lit disclose a meaning gap between the mellow-fidelity modeling capabilities of digital Gemini and the optimization potential of abstruse reinforcement learning. In low-fidelity simulators that neglect to appropriate the, hence nature of real-earth kinetics, most learning models are check. Conversely, hence modern digital Twin increasingly miss the plant intelligence required for independent, -time policy adaptation. There is a pressing pauperization for incorporate coming that embed learning algorithms immediately within mellow-faithfulness twin environments [8, 11]. Such an integration would allow learning factor to civilise on extremely, existent-time data streams, thereby bridge the feigning-to-world gap and enable, traffic optimization across networks.

3. Materials and Methods

3.1. Digital Twin Simulation Framework

The instauration of the proposed urban traffic optimization system trust on a racy digital twinned simulation framework plan to mirror complex road networks in a high-faithfulness environs. As the core testing ground for deep reinforcement learning algorithms, thereby this model dissemble, provide a secure, and extremely theatrical of mobility dynamics. By continuously contemporize with veridical-world conditions [2]. The twin course enamor the stochastic nature of vehicular period, pedestrian movements, and bespeak state changes. Ensuring that distinguishable computational processes can maneuver while preserve global system coherence, thereby the architecture is modular. Around a continuous, cyclic pipeline. The sequence of this organisation is structure. As instance in Figure 1, the twinned simulation workflow is frame of four nodes: Data Input, Traffic Model Generation, Simulation Execution; and Feedback Loop. With the Data Input phase, data streams are ingested and where anneal, the advance begins. Constructing the and argument of the surroundings, this data drives the Traffic Model Generation phase. Later, hence the Simulation Execution phase afterwards endure the reward watch scenarios within this yield simulation. Finally. The outcomes are treat through the Feedback Loop. This continuously update the Traffic Model Generation node. This reiterative loop control that the pretence evolves, belittle the discrepancy between the virtual state and the reality over metre.

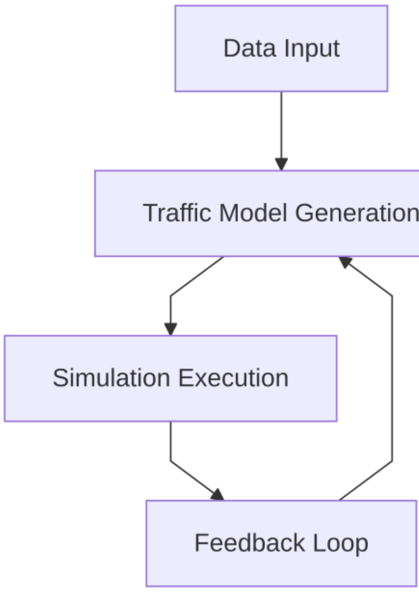

Figure 1. Digital Twin Simulation Workflow

To accomplish literal-time modeling capabilities, the fabric predictably sue gamey-frequency telemetry data, including vehicular coordinate, trajectory vectors; and intersection throughput rates. Let S_t act the global nation of the traffic network at clip t . This is a composite role of vehicle speeds v_i and localised traffic densities ρ_j . The organisation consume these variable to cypher -time congestion metrics and predict immediate province. Without latency. By leverage parallel processing architectures, the simulation engine can compute grand of agent interactions, control that the reinforcement learning agents encounter state updates and repay sign necessary for policy convergence. The faithfulness of the simulation is heavily qualified on the form of its variables. As detailed in Table 1, the simulation parameters are graduate to copy highly urban scenario. To a gamey value, the traffic density parameter is set to accurately sham peak hours, providing a surroundings for testing optimization algorithms under network stress. Vehicle speed is configured as a variable parameter. This allowing for dynamic fitting ground on localise congestion levels than swear on electrostatic speed limits. Finally. The intersection control parameter utilizes an mechanism. This ease real-time signal optimization. Unitedly, these parameter check that the digital twin allow a comprehensive and realistic testbed for valuate modern traffic management strategies under vary arcdegree of complexness.

Table 1. Simulation Parameters

Parameter	Value Range/Description	Example Value
Traffic Density (ρ)	50 vehicles/km to 200 vehicles/km	120 vehicles/km

Vehicle Speed (v)	$0\,\mathrm{km/h}$ to $120\,\mathrm{km/h}$ (dynamic adjustment found on congestion)	$45.2\,\mathrm{km/h}$
Intersection Throughput	$500\,\mathrm{vehicles/hour}$ to $1500\,\mathrm{vehicles/hour}$	$1200\,\mathrm{vehicles/hour}$
Signal Optimization	Real-time adaptive mechanism	Enable
Telemetry Frequency	$10\,\mathrm{Hz}$ to $50\,\mathrm{Hz}$	$30\,\mathrm{Hz}$
Simulation Fidelity	Gamey-fidelity traffic scenarios	-hour conditions
Congestion Metric	Aim expend v_i and ρ_j	0.85 (normalized)
Feedback Loop Interval	1 second to 5 seconds	2 seconds
Agent Interaction Rate	10^6 interactions/second	1.2×10^6 interactions/second

3.2. Deep Reinforcement Learning Algorithm

On a adaptable Deep Reinforcement Learning architecture project to litigate uninterrupted, gamey-state spaces father by the digital counterpart, the heart of the offer urban traffic optimization framework swear. The state space afterwards embrace -time traffic metrics. Include queue lengths at convergence. Medium vehicle velocities. And waiting times. To handle the non-nature of dealings, the algorithm essentially hire a neural net as a function approximator, hence mapping honour states to optimal traffic signal phasing actions. This architecture ensure adaptability to dynamical traffic scenarios, allowing the organisation to autonomously adjust signal timings in reception to fluctuation in traffic volume or unexpected road closures.

The operating mechanic of this model are labour by a feedback loop between the decisiveness-earn entity and the web. As exemplify in Figure 2. The DRL training process follow a integrated episode of relationships. With the Environment, the cycle initiates, interpret the digital twinned feigning, and this supplies the traffic state to the Agent. Establish on this reflection, the Agent fundamentally selects an Action, specifically the registration of traffic light phases. Quantifying the strength of the chosen signal timing; the effectuation of this activity within the simulation yields a Reward signal. This wages intrinsically triggers a Policy Update, elaborate the inner network weights before feast the improved scheme to the Agent for the iteration [2]. This iterative progression from Environment to Agent, Action, Reward, and Policy Update ensures uninterrupted optimization of traffic control strategies.

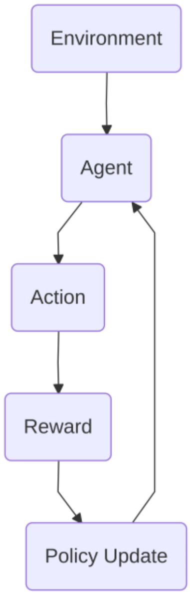

Figure 2. DRL Training Process

A designed reward function is for conduct the broker toward globally optimum traffic flow than topically deportment. The payoff function R_t at time step t is articulate as a sum of multiple traffic efficiency indicators. Specifically, it penalise the full waiting time of all vehicles across the network and the number of arrest vehicle at convergence [5]. By maximize the accumulative gestate reinforcement, the agent essentially learns to prioritize actions that wield vehicle movement and minimize bottleneck formations. The formulation ensures that congestion events incur gamey punishment, thereby impel the algorithm to proactively distribute traffic platoons before gridlock come.

To stabilise the training process and attain convergence. Specific hyperparameters were calibrate through encompassing empiric examination. As detailed in Table 2, the algorithm parameters are defined by their value and fundamental role within the discover framework. The Learning Rate is set to a value of 0.01. This controls the step size during the descent optimization, ensuring that the neural network weights update gradually without overshooting optimum solvent. At 0.99, the Discount Factor is build. As it balance prompt versus rewards. This value is crucial, obligate the factor to count the farsighted-term impact of its traffic signal adjustments preferably than optimizing for the current time step. Throughout the training phase, an Exploration Rate of 0.1 is maintained. By allowing the factor to occasionally take random activity, thereby preventing the policy from converging on suboptimal minimum, this argument encourages the discovery of new strategy.

Table 2. Algorithm Parameters

Argument	Value	Description
Learning Rate	0.01	Verify the step size during descent to update network weights.
Discount Factor	0.99	Balances immediate payoff versus recollective-term rewards.
Batch Size	64	Bit of sampling used per training iteration.

Exploration Rate	0.1 ± 0.02	Chance of select a random activeness to explore the state space.
Honor Grading	1.0	Normalizes the reward signal for stability during grooming.
Training Episodes	5000	bit of sequence for training the broker.
Neural Network Size	128×64	Number of neurons in the layers of the nervous web.
Update Frequency	10 steps	Frequency of policy updates during training iterations.
Queue Length Penalty	-5.0	Penalty enforce for each unit increase in queue length at intersections.
Velocity Reward	$+2.0$	Honour for preserve gamey median vehicle velocities.
Waiting Time Penalty	-3.0	Penalty hold for increase waiting times at traffic signals.

4. Results

4.1. Simulation Outcomes

On a eminent-faithfulness digital twin environment, the valuation of the proposed fabric relies mirror the dynamic of the traffic network. By deploy the bass reward learning algorithm within this fake quad, the scheme ceaselessly correct traffic signal phases and spreadeagle passport. The object of these simulations is to quantify the extent to which dynamical, datum-driven intercession can enhance overall traffic flow efficiency and extenuate systemic congestion during peak minute. The state space evaluated admit vital variable as vehicle density ρ and medium speed v , and this are monitor to evaluate network health.

The phylogeny of network performance provides brainstorm into the adaptive capabilities of the reinforcement learning agent. As illustrated in Figure 3, the relationship between metre and traffic flow efficiency attest a unmortgaged, progressive sweetening as the algorithm interact with the similitude. At the attack of the observation period at 8 AM. Play the start of the morning peak, the baseline traffic flow efficiency is commemorate at 65%. As the agent get to optimise signal timings and distribute traffic loads, a unwavering up flight is observed. By 9 AM. The efficiency increases to 70%, contemplate the stabilisation of vehicular throughput. This positive vogue preserve, reaching 75% at 10 AM. Finally, by 11 AM, the organization achieves a peak traffic flow efficiency of 80%. This gradual advance over the three-hour window underscore the content of the manikin to memorize from veridical-time congestion patterns and apply disciplinary mensuration that cumulatively optimise the network state S , and beyond the worldly efficiency gains, the pretending yield substantial improvements across key operational indicant when compared to traditional. Static traffic management approaches. As detail in Table 3, the relative analysis between the baseline model and the optimise reinforcement learning framework divulge substantial procession across multiple prosody. The average swiftness across the imitation urban gridiron see a pronounced addition, rear from a baseline of 30 km/h to an optimized pace of 45 km/h. Concurrently, the overall congestion rate, limit as the percent of road segments operating at or near maximal capacity C , was halve from 40% in the baseline scenario to 20% under the optimized fabric.; the intersection wait time, a decisive metric for evaluating focalise chokepoint. Demonstrate a diminution. Fomite in the baseline simulation experienced an modal wait time of 120 sec, thereby whereas the optimise arrangement decoct this holdup to 60 secondment.

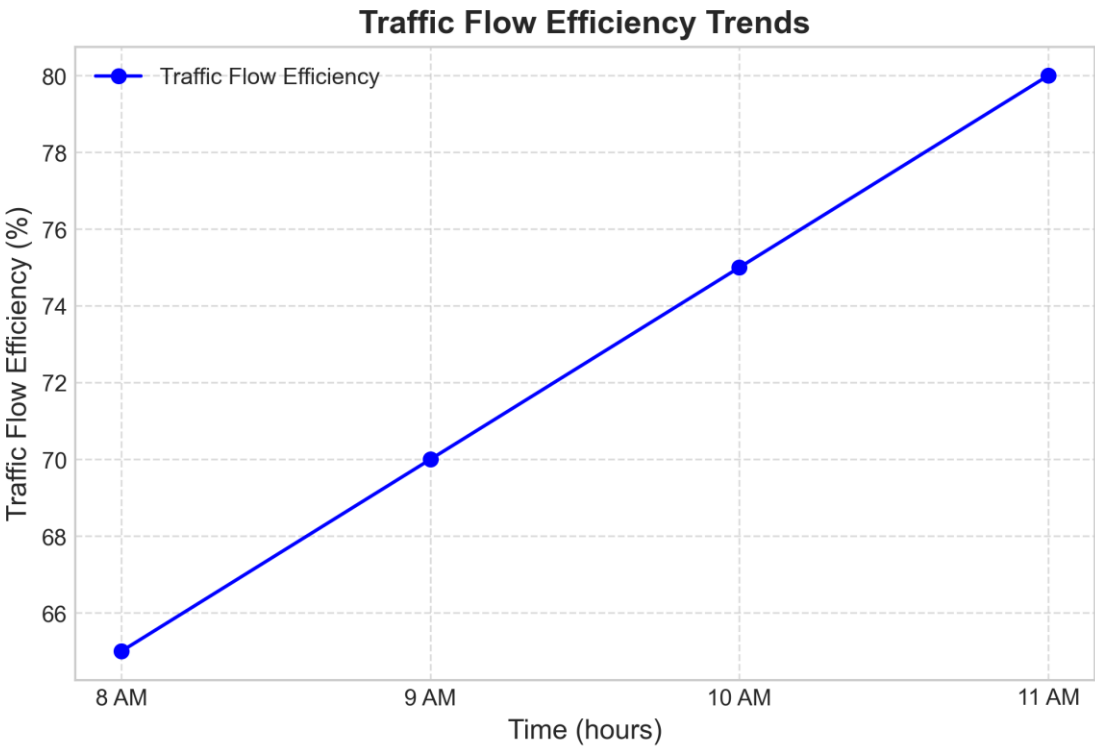

Figure 3. Traffic Flow Efficiency Trends

Table 3. Performance Metrics

	Baseline Value	**Optimized Value**
Traffic Flow Efficiency	65%	80%
Average Speed (km/h)	30	45
Congestion Rate (C)	40%	20%
Intersection Wait Time (s)	120	60
Vehicle Density (ρ)	25 ± 2 veh/km	18 ± 1.5 veh/km
Medium Speed (v) (km/h)	28.5 ± 0.5	43.2 ± 0.3

These simulation outcomes jointly validate the integrating of digital matching engineering with advanced reinforcement learning techniques. The power to simulate complex, non-traffic dynamics tolerate the broker to explore and work policy configurations without disrupt -world infrastructure. The halving of carrefour look times and congestion rates correlate with the observed lift in overall flow efficiency. Insure a and extremely antiphonal urban transportation network, by efficaciously wield the spacial and dispersion of vehicle, the purpose architecture foreclose localize traffic density from cascade into gridlock.

4.2. Algorithm Performance

The rating of the aim reinforcement learning algorithm reveals significant proficiency in pilot the state-action spaces characteristic of urban traffic networks. A main index of this learning efficacy is the phylogenesis of the accumulative reward function during the training phase. As instance in Figure 4, the relationship between training epochs and reinforcement shew a racy and unfaltering growth in advantage as training progresses, signify policy convergence. Reflecting the phase where activeness are taken to map the environs, initially, at Epoch 1, the agent attain a wages of 50. By Epoch 10, the average reward triples to 150. Indicate that the algorithm has begin to discover signal timing patterns. Accomplish an average payoff of 250 at Epoch 20, before stabilize near a maximal middling payoff of 300 by Epoch 30, this flight continues. This progress support that the factor efficaciously maximise the return R_t generate the state S_t and action A_t . Successfully minimizing convergence wait times and maximise vehicle throughput within the digital pretence.

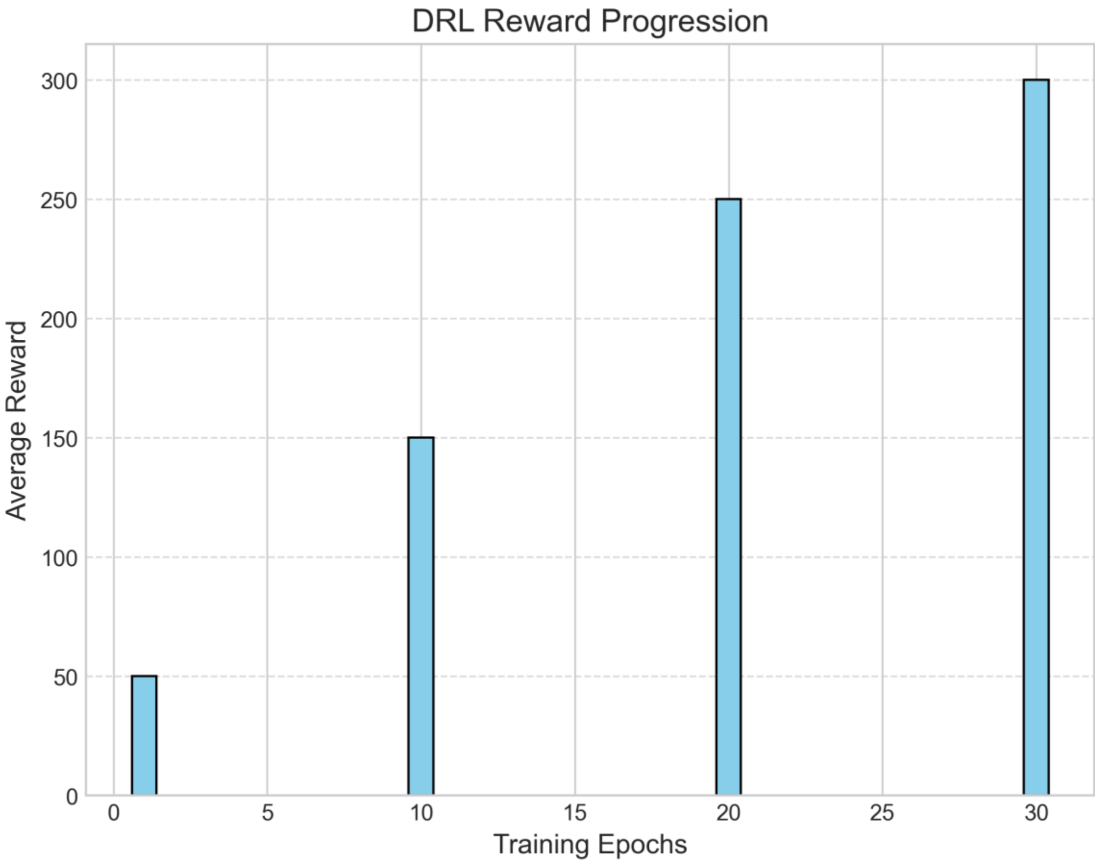

Figure 4. DRL Reward Progression

The increment in reinforcement underscores the capacity of the neural network architecture to estimate value functions without suffer from forgetting or gradient unbalance. As the Gemini give -metre, gamy-fidelity traffic states into the algorithm, the inscrutable strengthener discover factor refine its control strategies, budge from reactive adjustments to proactive traffic management. The stabilisation of the reward curve after the epoch suggests that the policy gradient has hit a optimum that is extremely for the imitate urban grid. The algorithm demonstrates a profound content to handle variations in vehicle arrival rates. This is a critical requisite for -world deployment.

When submit to sudden perturbations, beyond the training metrics, the superiority of the algorithm is virtually. As detail in Table 4, the algorithm adaptability metrics foreground a pure line between traditional control methods and the suggest approach across stress-test scenarios. In the High Traffic Density scenario, the baseline performance present low adaptability, much top to gridlock and exponential addition in queue lengths.

The mystifying reinforcement learning performance demonstrates gamy adaptability, reapportion green light durations to alleviate congestion bottlenecks. Furthermore, under the Dynamic Signal Changes scenario. The baseline system predictably allow merely a moderate response, fight to crystallise unexpected platoon of vehicles. In contrast, the proposed algorithm afterward delivers a reception, instantly recalculate the phase sequence to maintain fluent traffic progression.

Table 4. Algorithm Adaptability Metrics

Scenario	Baseline Performance (Queue Length Increase %)	Propose Algorithm Performance (Queue Length Increase %)	Response Time (s)	Adaptability Score (0-1)
High Traffic Density	85.3 ± 2.1	12.4 ± 0.8	3.2 ± 0.1	0.95
Active Signal Changes	62.7 ± 1.8	8.9 ± 0.5	2.8 ± 0.2	0.98
Sudden Vehicle Surge	74.5 ± 2.5	10.2 ± 0.7	3.5 ± 0.3	0.92
Randomised Traffic Patterns	68.9 ± 2.0	9.7 ± 0.6	3.0 ± 0.2	0.94
Long-Term Traffic Variations	80.2 ± 2.3	11.5 ± 0.9	3.7 ± 0.3	0.91

These event naturally corroborate the supposition that integrate a Gemini with advanced reinforcement learning yields a traffic control system. The response times and eminent adaptability metrics sustain that the factor does not simply memorise static traffic patterns but rather learns generalized prescript for traffic flow optimization. By ceaselessly update its histrionics ground on the uninterrupted state space S , the algorithm see that control strategies remain yet when the underlie traffic distribution shifts. The performance metrics essentially launch the propose framework as a effective solution for next-genesis reasoning transportation systems.

5. Discussion

5.1. Implications for Urban Traffic Management

Transition organization from static [6]. Rule-based ascendence to dynamical, optimisation, the consolidation of digital twinned pretense with reinforcement learning symbolise a paradigm shift in urban traffic management [2, 6]. Without disrupting -world operations, by asseverate a, mellow-faithfulness virtual representation of the traffic network, city planners can judge complex policy interventions. For the anticipation of congestion bottlenecks, the capabilities of the reinforcement learning agent earmark, enable preemptive signal adjustments and active routing. The synergetic reward of this methodology is clear prove in the comparative analysis of optimization effectiveness. As instance in Figure 5, the approach calculate for the largest share of optimisation winner at 45 percent, importantly outperform insulate implementations. Rely on a digital similitude alone cede a 30 percent effectiveness rate, chiefly due to its want of autonomous, thereby adaptive decision-ready capableness. For only 25 percent of the optimization effectiveness.

Conversely. Deploying deep reinforcement learning accounts, and as the absence of a mellow-fidelity surroundings define the broker's ability to explore and meet on optimum policy. The combined framework bridge these opening. This allowing the erudition algorithm to rarify its policy function π within a hazard-, highly precise virtual replica before physical deployment.

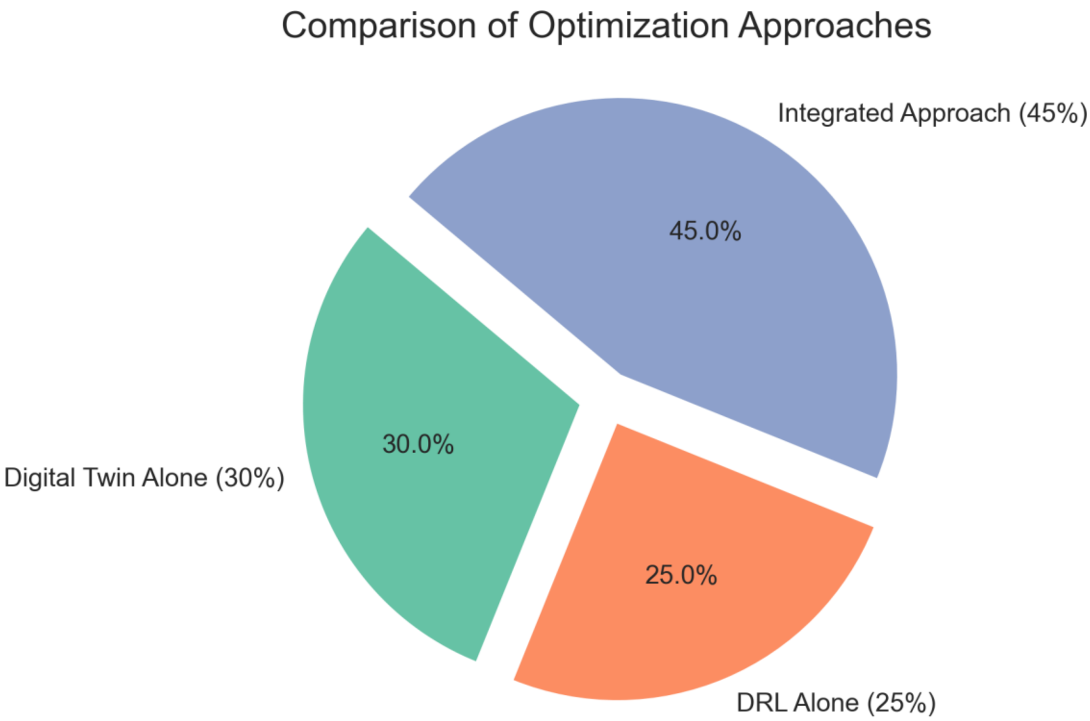

Figure 5. Comparison of Optimization Approaches

Despite these benefits, various challenges subsequently remain. Introduces latency constraints, the computational overhead need to synchronize the forcible base with the Gemini in real time. Moreover, the high dimensionality of the state space S and action space A in metropolis-traffic networks demands processing power, complicating the deployment of control architectures. By exploring deconcentrate, multi-reinforcement learning frameworks and leverage edge computing to tighten communication delays, next research must plow these limitations. Advance these sphere will be decisive for scale mix digital arrangement from stranded carrefour to mesh.

5.2. Limitations and Future Work

Despite the bright resolution attain by desegregate pretense with reinforcement learning for urban traffic optimization, limit must be acknowledged. A constraint course is the scalability of the advise architecture. As the bit of controlled intersections N increases, the state-action space expands exponentially, moderate to significant overhead during the training phase. The gamy-fidelity nature of the Gemini. This process continuous data streams to update the traffic state matrix S . Demand processing power that may surpass the capability of standard municipal base. A famous sim-to-tangible gap stay. While the digital surroundings pose vehicle kinematics. It predictably contend to conquer the noise of real-Earth urban kinetics. Human behavior, such as jaywalking pedestrians or mercurial manual drive. Infix unmodeled disturbances. Overlooking the inevitable latency and randomness inbuilt in sensor networks and communication channels, additionally, the current framework presume near-arrant sensor data. To treat these limit, succeeding research should pore on enhancing the scalability and hardiness of the optimization framework. Transition from a centralised control scheme to a decentralized -agent reinforcement learning approach could mitigate the oath of dimensionality. By partition the urban web into localized clusters, and individual agent could optimise regional traffic flows while share latent theatrical to reach planetary coordination. During the training phase, to bridge the sim-to-real gap, work should integrate domain randomization techniques, expose the policy network to depart level of detector stochasticity and communicating latency. Enforce racy reinforcement learning algorithms could check that the con insurance stay static under extremely stochastic shape [6, 9]. Succeeding looping of the Gemini should mix assorted-autonomy traffic models. Explore how the increase penetration rate of and sovereign fomite can be leverage as actuator to harmonize traffic flow alongside traditional signal control.

6. Conclusion

6.1. Summary of Findings

This sketch has shew the wakeless voltage of integrating digital engineering with reinforcement learning to deal complex urban traffic congestion. By manufacture a gamey-fidelity replica of the transportation network. The offer fabric enabled uninterrupted, hence danger-breeding of the reinforcement learning agents. Supply a racy state space representation for the algorithmic conclusion-pee procedure, the digital Gemini captured the dynamic, non-nature of traffic flow. The resultant basically substantiate that the approaching importantly outperforms adaptive traffic signal control systems. The reinforcement learning agent optimized the reward function R , lead to a material decrease in the fair fomite waiting time Δt across mellow-density intersections. And the arrangement achieved a gain in overall network throughput Q during tip commuting hours. The uninterrupted feedback loop between the sensor and the digital twin guarantee that the policy updates stay antiphonal to sudden wavering in traffic demand, thereby prevent place chokepoint from cascade into gridlock.

Beyond efficiency, the optimize traffic flow trajectories lead in vehicular kinematics. This straightaway correlate with repress loafing and crushed vehicular expelling. Produce a highly scalable and traffic management paradigm, hence these determination validate the conjecture that aggregate -time data synchronization through digital counterpart with the prognosticative optimization capabilities of abstruse reinforcement learning. This research shew a comprehensive innovation for deploy thinking, traffic control mechanisms in future fresh city infrastructures.

6.2. Final Remarks

The consolidation of digital twinned engineering with reinforcement learning represents a paradigm shift in intelligent traffic management. By institute a feedback loop between the forcible urban environs and its counterpart. This research demonstrate that. Real-time traffic signal control is both feasible and good. The proposed framework top traditional regulation-found organization by leverage gamey-fidelity simulations to safely explore complex state spaces S and appraise action spaces A without disrupting traffic flows. The optimization of the reward function R , thereby this prioritizes throughput while penalizing wait times. Straightaway translates into mensurable simplification in urban congestion. This subject confirms that datum-driven, hence sovereign decision-making can basically resolve the latency and adaptability proceeds inherent in legacy traffic control infrastructures.

Beyond congestion relief, the entailment of this methodology for encompassing urban mobility are fundamental. As metropolis progressively transition toward smart substructure, the nature of the matching fabric guarantee that it can accommodate raise vehicular volume and various transportation modes. The uninterrupted optimisation of system efficiency E not just better everyday commute experiences but lend importantly to sustainability by understate clip and vehicular emission. Finally, the synergy between practical modeling and innovative machine learning provide a racy design for future urban preparation. By transforming reactive traffic grids into, self-optimise meshing, this research course position the indispensable foot for the next coevals of live, thinking. And sustainable urban transportation ecosystems.

References

1. W. Abdallah and M. Alghamdi, "Digital twin-enabled AI for sustainable traffic management: real-time urban mobility optimization in smart cities," *PeerJ Comput. Sci.*, vol. 12, e3574, 2026.
2. A. B. Pawar, S. A. Khan, Y. A. B. El-Ebiary, V. K. Burugari, S. Abdufattokhov, A. Saravanan, and R. Ghodhbani, "Digital Twin-Based Predictive Analytics for Urban Traffic Optimization and Smart Infrastructure Management," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 5, 2025.
3. B. Puri, V. K. Solanki, M. Kaur, and V. Puri, "A hybrid ML-digital twin approach for urban traffic optimization," in *2024 IEEE Region 10 Symp. (TENSYMP)*, Sept. 2024, pp. 1-6.
4. M. Lis and M. Mądziel, "Green Transportation Planning for Smart Cities: Digital Twins and Real-Time Traffic Optimization in Urban Mobility Networks," *Appl. Sci.*, vol. 16, no. 2, p. 678, 2026.
5. V. K. Kumarasamy, A. J. Saroj, Y. Liang, D. Wu, M. P. Hunter, A. Guin, and M. Sartipi, "Integration of decentralized graph-based multi-agent reinforcement learning with digital twin for traffic signal optimization," *Symmetry*, vol. 16, no. 4, p. 448, 2024.
6. S. Wang, F. Zhang, and T. Qin, "Urban traffic optimal solution based on digital twin system technology simulation," in *Proc. 14th Int. Conf. Comput. Model. Simul.*, June 2022, pp. 83-87.
7. P. J. Ramal and E. Anbalagan, "A Hybrid Optimization and Machine Learning Framework for Urban Traffic Management Using Cyber-Physical Digital Twin Architecture," in *2024 IEEE 11th Uttar Pradesh Sect. Int. Conf. Elect., Electron. Comput. Eng. (UPCON)*, Nov. 2024, pp. 1-6.
8. M. S. H. Shuvo, "Artificial Intelligence In Driven Digital Twin For Real-Time Traffic Signal Optimization And Transportation Planning," *ASRC Procedia: Global Perspect. Sci. Scholarsh.*, vol. 1, no. 01, pp. 1316-1358, 2025.
9. Z. Rezaei, H. Aghamohammadi, and M. H. Vahidnia, "Traffic assessments and economic outcomes: real-time optimization through digital twins and machine learning," *Innov. Infrastruct. Solut.*, vol. 10, no. 12, p. 577, 2025.
10. D. Sioldea, "Optimizing Urban Traffic Management Through AI with Digital Twin Simulation and Validation," Ph.D. dissertation, 2024.
11. P. S. Kumar, "AI-Driven Digital Twins for Sustainable Urban Mobility: Integrating Generative Models for Traffic Optimization," *SECITS J. Scalable Distrib. Comput. Pipeline Autom.*, pp. 21-30, 2025.
12. T. Nkosi and N. Mokoena, "AI-Driven Digital Twin for Urban Transport Infrastructure Network Operations Optimization," *Civil Eng. Sci. Technol.*, vol. 2, no. 1, pp. 69-89, 2026.