EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

Multi-Agent Soft Actor-Critic with Graph Attention Networks for Adaptive Traffic Signal Optimisation (MASAC-GAT)

R. M. Bommi[1,*], E. Bhuvaneswari[2], M. Rohini[3], G. Uganya[4]

[1]Department of Electronics Engineering (VLSI Design and Technology), Chennai Institute of Technology, Kundrathur, Tamil Nadu 600069.
[2]Department of AI&DS, Panimalar Engineering College, Chennai, Tamil Nadu 600 123.
[3]TIFAC-CORE in Cyber Security Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Tamil Nadu 641 112.
[4]Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu 600062.

Abstract

INTRODUCTION: Adaptive Traffic Signal Optimisation (ATSO) is a challenging problem for urban traffic networks, having important implications for congestion reduction, traffic efficiency, and environmental conservation. Conventional traffic signal control techniques, i.e., fixed-time and rule-based control, fail to respond to dynamic traffic behaviour efficiently.
OBJECTIVES: Recent developments in Reinforcement Learning (RL) have been promising for ATSO but are plagued by poor scalability, lack of coordination in multi-intersection networks, and inefficiency in dealing with continuous action spaces.
METHODS: Furthermore, most RL-based solutions are based on simplistic state representation and fail to incorporate complex interdependencies between traffic signals. Considering these limitations, this paper introduces a new framework, Multi-Agent Soft Actor-Critic with Graph Attention Networks (MASAC-GAT), which unites the sample efficiency and stability of Soft Actor-Critic (SAC) with the relational modelling ability of Graph Attention Networks (GATs).
RESULTS: The proposed method exhibited significant performance gains on three important traffic metrics: Signal Adjustment Efficiency (92%), Average Waiting Time (20–35 seconds), and Congestion Prediction Accuracy (93%), outperforming DQL, PPO, A2C, GNN-based variants, and knowledge sharing DDPG (KS-DDPG). Through minimised redundant signal changes and reduced vehicle delays, the method ushers in the next generation of smart transportation systems.
CONCLUSION: The proposed method facilitates decentralised yet coordinated control of traffic signals by utilising local observations and global context. The proposed method unites real-time traffic observations, e.g., traffic volume, vehicle speeds, weather, accident reports, and signal status, into a customised OpenAI Gym environment for training and evaluation.

Keywords: Traffic Signal Optimisation, Reinforcement Learning, Intelligent Transportation Systems, Sustainable Urban Mobility, Graph Neural Network, Attention.

Received on 06 October 2025, accepted on 12 December 2025, published on 05 January 2026

doi: 10.4108/eetiot.10486

* Corresponding Author: bommirm@citchennai.net

1. Introduction

Urban transport systems are crucial to the mobility and economic well-being of contemporary cities, but they are plagued by daunting challenges posed by rising congestion, growing cities, and increased vehicle densities [1]. Congestion in traffic not only retards everyday life but also results in heavy environmental and economic impacts, such as increased fuel consumption, emissions of greenhouse gases, and lost productivity [2, 3]. Due to the pollution and traffic disruption caused by these mechanisms (signal control), traffic management and logistics often are significant issues that need to be addressed. Traffic accidents are responsible for serious health outcomes globally, with 1.35 million deaths or disabilities annually. In 2019, 93% of deaths due to injury occurred in low- and middle-income countries. Road traffic injuries will be the seventh leading cause of death globally in 2030. Resolution of this under-attended public health problem requires conservative preventive strategies to reduce collisions and enhance safety [4]. These problems require intelligent traffic signal control strategies that can adapt dynamically to real-time traffic situations, especially in urban high-density areas [5, 6].

Traffic signals usually follow fixed-time, actuated, or adaptive control techniques [7]. Fixed-time signal control depends on a repetitive cycle that does not change regardless of the prevailing traffic conditions, continuing with its cycles regardless of varying traffic in the area. Although traffic signals possess real-time loop detector data, the actuated control technique is suboptimal for varying traffic demand, thus resorting to adaptive signals as an alternative [8]. Traditionally, the traffic signal equipment was considered mostly as a traffic control system, but its modern use as part of smart city infrastructure demonstrates its critical function in traffic safety [9].

Traditional traffic signal control systems typically employ predetermined timing plans or heuristic-based optimisation models [10, 11]. While both methods work while traffic continues in a predictable manner, they are incapable of dynamically addressing the type of complex traffic conditions with varying traffic volumes, accidents, or poor weather conditions that often result in suboptimal signal timings, leading to congestion and increased delays of vehicles. The advent of Deep Learning (DL) in recent years, via the use of real-time data and adaptive decisions, has introduced new opportunities for traffic management. Despite the ability of these models to identify patterns and sequence models, they are focused primarily on prediction and are less concerned about control [12].

Reinforcement learning is a new frontier for dynamic decisions, and there are three methods of traffic signal control: value-based, policy-based, and actor-critic. Value-based methods (i.e., Q-learning) use experience, over a series of steps, to parameterise a long-term state-action value function; policy-based methods [13] will model non-stationary transitions with sampled episode returns; and actor-critic methods will implement a different model to reduce bias and variance. Actor-critic beats Q-learning specifically for centralised RL agents, but training a centralised RL agent for large-scale traffic signal management remains an obstacle [14, 15, 16] due to the very high dimension of the joint action space. Multi-agent RL (MARL) frames the issue of scalability by distributing overall control to each local RL agent. MADRL traffic signals operate independently yet simultaneously and therefore will undertake uncoordinated actions, which will likely worsen congestion [17].

In addition, many reinforcement learning (RL) algorithms often get stuck in local optima, particularly in the case of discrete action spaces, because they are not able to explore the environmental space effectively. So, they converge on a non-optimal policy that only uses limited information from the state, such as vehicle counts, and does not take into consideration the more complex spatial and temporal intricacies between intersections. It is by recognising the limits of current multi-agent soft actor-critic (SAC) systems and making new contributions that the present study defines a new state-of-the-art for adaptive traffic signal optimisation in intelligent transportation systems and paves the way for further exploration and research.

The purpose of this work is to overcome existing limitations with RL-based ATSO methods by introducing a coordinated, scalable, and efficient framework for the automated control of traffic signals at multiple intersections. The proposed use of Soft Actor-Critic (SAC), an advanced RL algorithm, and Graph Attention Networks (GATs), a robust graph neural network architecture, enables coordinated yet decentralised control of traffic signals in large-scale networks.

The contributions of this work are as follows:

1. Some MASAC implementations use a fixed global entropy parameter α. However, MASAC-GAT uses a local, traffic-state-dependent entropy α_i^t that changes on its own from 0. This creates emergent behavior where agents conservatively exploit during critical conditions while exploring during uncongested periods This is achieved through: $\alpha_i^t = \alpha_{base} \times f(queue\_length_i^t)$, where congestion detection is binary.
2. MASAC-GAT enables real-time, traffic-conditioned attention recomputation, unlike previous GAT-based approaches that compute static attention weights. The addition of a punishment term based on traffic volume similarity is innovative. This allows agents to dynamically downweight neighbours with differing traffic patterns, resulting in implicit hierarchical cooperation. Every time step, the network is recomputed, allowing for continual adaptation.
3. Unlike KS-DDPG, which relies on runtime knowledge container exchanges, MASAC-GAT achieves coordination through three explicit phases: (Phase 1) Train centralised critics and GAT parameters, (Phase 2) locally cache trained GAT encoders on each agent, and (Phase 3) execute with zero inter-agent communication—each agent observes traffic autonomously, computes GAT features using local cached weights, and selects actions. This reduces the amount of communication overhead during deployment and increases fault tolerance.

Hence, the proposed method enhanced traffic flow efficiency, enhanced travel time, and enhanced accuracy of congestion prediction relative to existing methods.

The remainder of this paper is organised as follows. Section 2 details existing works developed for traffic signal management. Section 3 describes the proposed traffic signal management framework to illustrate the workings of the Soft Actor-Critic algorithm and to clarify how it fits into the traffic signal control system. Section 4 builds a detailed description of the experimentation, the parameters of the simulation framework, evaluation metrics, training variables, and results and describes a comparison made to existing methods. Finally, Section 5 presents conclusions by outlining the key findings and contributions.

2. Literature Survey

Traditional traffic signal management systems, such as fixed-time and rule-based types of traffic signal management, are popular in built environments. In fixed-time traffic signal control, the traffic signal operates according to a fixed-time schedule that has no real-time data adjustment and results in inefficient signal control. During peak hours, this means the traffic signal will allow vehicles to clear intersections more slowly than if the schedule had made consideration for prevailing traffic conditions. Conversely, during off-peak hours, this type of control removes vehicles from the intersection too quickly. Rule-based traffic signal systems, including actuated traffic control, can take data from traffic sensors and immediately respond; however, traditional rule-based signal control lacks the strategic insight needed to address more complex prevailing traffic conditions across adjacent intersections. Furthermore, adaptive systems, such as SCOOT or SCATS, make predictions and optimise the system throughput as a result, yet require complicated modelling inputs and may also not perform well in highly dynamic traffic conditions or under congestion [18]. In this regard, traditional types of signal control are simple to implement; however, they do not account for the unpredictable and stochastic characteristics of urban traffic, yielding only suboptimal signal control in terms of managing congestion and the environment [19].

Reinforcement learning (RL) for traffic control has emerged as a way to overcome the aforementioned shortcomings. The initial research applied RL methods (e.g., Q-learning and actor-critic) to learn green timing policies from observed traffic interaction data rather than explicitly building traffic models [20]. In single intersection scenarios, deep RL techniques (like DQN, DDPG, A2C, etc.) have been able to learn to do better than fixed timing rules, optimising delay or queueing performance metrics [21, 22]. For multiple intersections, traffic signal control resembles a multi-agent RL problem, and each signal is typically assigned an RL agent; while some of the previous approaches deal with independent learning (with or without shared policies), others apply centralised critics and communication protocols to coordinate the agents [23, 24, 25]. RL-based Adaptive Traffic Signal Optimisation (ATSO) faces some major challenges, including poor scalability in large traffic networks due to its ever-expanding state-action space and independently learning agents without supervised learning, which leads to global performance sub-optimality. Many approaches also face issues with continuous action spaces, failing to make suitable modifications to signal timings. In addition, simplistic state representation fails to capture the spatial interdependence between intersections, which decreases decision-making effectiveness and, ultimately, performance. There is thus a pressing need for a more coherent, coordinated, and contextually aware RL framework in the field of traffic signal control.

Graph Neural Networks (GNNs) [26] have been integrated into RL [27], accounting for relational aspects of traffic network structures. Traffic through intersections and roads can be illustrated via GNNs, where intersections are represented as nodes and road segments are represented as edges, enhancing GNNs' ability to learn spatial dependencies to better coordinate multiple intersections. The GNN used in conjunction with the RL agent trained using SAC algorithms reduced average waiting times, queue lengths, and traffic delays in a traffic environment [28, 29, 30]. Effective reductions in wait times occurred because GNNs can reconstruct the present state of the traffic network with respect to all road segments active in the situation, both in real time and simultaneously.

In previous work, standard MASAC implementations used entropy regularization with a fixed temperature value α. This approach believes that entropy-based exploration is equally useful, independent of local conditions. However, in traffic management, this assumption is problematic. By implementing traffic-state-dependent entropy modulation, which allows each agent to autonomously modify its entropy temperature in response to observed queue length, MASAC-GAT fills this gap. This is accomplished by introducing local exploration-exploitation adaptations that were not present in previous MASAC work. While Graph Attention Networks offer variable attention mechanisms, current traffic control programs (GAT-SAC, MAGAC) compute attention weights purely on learned feature similarity. MASAC-GAT addresses this by incorporating a traffic-volume penalty term into the attention computation, allowing for real-time (timestep-level) recomputation that responds to changing traffic. At runtime, communication-based multi-agent techniques necessitate explicit knowledge transfer between agents. MASAC-GAT reduces deployment communication by utilizing a three-phase architecture in which coordination is acquired during training but executed implicitly during deployment. No agent-to-agent communication is necessary during execution.

The Multi-Agent Soft Actor-Critic with Graph Attention Networks (MASAC-GAT) framework will target scalability, coordination, handling continuous actions, and decentralized control with complex states. The MASAC-GAT will benefit from the sample efficiency and fidelity of SAC while being able to use GAT's representation to model the relations between complex states with many agents at each timestep. The MASAC-GAT framework will enable coordinated yet decentralized control of traffic identifiers under heterogeneous states by fitting the controller to the traffic

dynamics in both real-time and historical time windows. The MASAC-GAT framework will support real-time adaptive control of traffic effectiveness by improving overall traffic efficiency to minimize delays for drivers.

3. Proposed Method: Reinforcement Learning-Based Traffic Signal Optimisation

The growth of cities has caused a big jump in the number of vehicles on the roads, leading to heavy traffic jams in urban areas. Traffic bottlenecks in city networks are still a major problem made worse by the shortcomings of old-school traffic light control systems. These systems often stick to fixed schedules or use simple adaptive rules. They don't account for the changing interconnected nature of city traffic when weather conditions vary or unexpected events like crashes occur. Machine learning that adapts through real-time interaction with its surroundings has shown promise as a better option. However, current machine learning methods for optimizing traffic signals face several unsolved problems: (1) working well in networks with many intersections, (2) teamwork between separate control units, (3) dealing with smooth signal timing changes, and (4) bringing together different types of information (like traffic data, weather, and accidents). To tackle these issues, we suggest a new approach called Multi-Agent Soft Actor-Critic with Graph Attention Networks (MASAC-GAT), as shown in Figure 1. This method combines graph-based teamwork with a type of machine learning that encourages exploration.

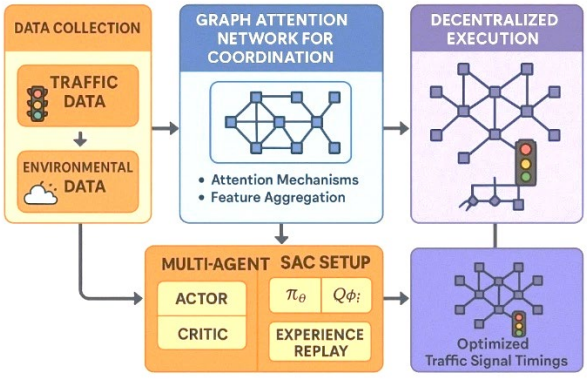

Figure 1. Detailed Diagram of Proposed Method

This section describes three architectural advances that distinguish MASAC-GAT from previous work. Rather than simply integrating current MASAC and GAT techniques, we present innovative mechanisms for entropy regularisation (Section 3.3), attention computation (Section 3.2), and execution architecture (Section 3.5). Each innovation addresses a unique problem in previous traffic signal control research.

3.1 Problem Formulation

The traffic network is represented as a decentralized partially observable Markov decision process, in which autonomous agents control each intersection. At time t, each agent i perceives a local s_i^t, which is given by:

$$s_i^t = [tv_i^t, asc_i^t, ast_i^t, asb_i^t, w_i^t, t_i^t, h_i^t, ar_i^t, ss_i^t] \quad (1)$$

s_i^t observing current traffic information, weather, and signal status. Agents choose continuous actions $s_i^t \in R^m$, which represent green-phase lengths and transition times, to minimize traffic. Global state $s^t = U_{i=1}^N s_i^t$ changes stochastically according to traffic dynamics $P(s^{t+1}|s^t, a^t)$, which depends on vehicle flow, time stamp, and weather.

3.2 Graph Attention Networks for Coordinated Learning

To support cooperation between intersections, the traffic system is represented as a graph $G = (V, E)$, with the vertices $V = \{v_1, v_2, v_3, \ldots, v_N\}$ being intersections and edges E being the roads between them. Each vertex v_i is associated with features h_i^t learned from s_i^t. A Graph Attention Network (GAT) computes the attention weights α_{ij}^t between neighboring vertices v_i and v_j:

$$\alpha_{ij}^t = softmax_j(LeakyRelu\,(a^T[Wh_i^t || Wh_j^t])), \quad (2)$$

Where $W \in R^{d'*d}$ and $a \in R^{2d'}$ are learnable parameters. The GAT produces new node features $h_i^{t'}$ by aggregating neighborhood information:

$$h_i^{t'} = \sigma\,(\sum_{j\in N(i)} \alpha_{ij}^t\ Wh_j^t), \quad (3)$$

where σ is the ELU activation function. This enables agents to choose dynamically meaningful neighbors, context-aware coordination without the need for central control.

3.3 Multi-Agent Soft Actor-Critic (MASAC)

Every agent employs the Soft Actor-Critic (SAC) method, which maximizes the entropy of the policy to promote exploration in addition to the expected rewards. As a Gaussian distribution, the stochastic policy $\pi_i(.|h_i^{t'})$ is represented:

$$\pi_i(.|h_i^{t'}) = N\big(\mu_i(h_i^{t'}), \textstyle\sum_i(h_i^{t'})\big), \quad (4)$$

where μ_i and $\sum_i$ are neural networks, which provide mean and covariance. The critic $Q_i(h^{t'}, a)$ approximates the expected cumulative reward:

$$Q_i(h^{t'}, a) = E[\textstyle\sum_{k=t}^{\infty} \gamma^{k-t}\,(r_i^k - \lambda_1\, tv_i^k - \lambda_2\, w_i^k - \lambda_3\, ar_i^k)], \quad (5)$$

With the discount factor $\gamma \in [0,1)$, the policy is learned to maximize the entropy-regularized objective:

$$J(\pi_i) = E_{\pi_i}\big[\textstyle\sum_{k=t}^{\infty} \gamma^t\,(r_i^k + \propto H\,(\pi_i))\big], \quad (6)$$

Where $H(\pi_i) = E[-log(\pi_i)$ is the policy entropy, and $\propto$ is utilized in balancing between exploration and exploitation.

3.4 Reward Function and Training

The reward function penalizes crashes, waiting time, and congestion:

$$r_i^t = -(\lambda_1 tv_i^t - \lambda_2 w_i^t - \lambda_3 ar_i^t) \quad (7)$$

Agents also train cooperatively with experience replay, re ading transitions (s^t, a^t, r^t, s^{t+1}) from a shared buffer. The critic is trained by minimizing the Bellman error:

$$\ell(\Phi_i) = E\left[(Q_i(h^{t'}, a^t) - E\left[\left(Q_i(h^{t'}, a^t) - \left(r_i^t + r_i^k + \gamma\, \bar{Q}_i\left(h^{(t+1)'}, a^{t+1}\right)\right)\right)^2\right]\right], \quad (8)$$

where $\bar{Q}_i$ is a Polyak-averaged target network. The actor is updated by the policy gradient:

$$\nabla_{\theta_i} J(\pi_i) = E\left[\nabla_{\theta_i} \log \pi_i\,(a_i^t | h_i^{t'})(Q_i(h^{t'}, a^t) - \propto \log \pi_i\,(a_i^t | h_i^{t'}))\right] \quad (9)$$

3.5 Decentralized Execution

When deployed, each traffic signal agent works on its own using its trained policy (π_i). This setup removes the need for central calculations. Agents use local data s_i^t like traffic flow, speed, and weather, along with features collected through GAT $h_i^{t'}$ from nearby intersections, to decide actions. The policy produces a continuous action $a_i^t \sim \pi_i(.|s_i^t, h_i^{t'})$ based on a Gaussian distribution:

$$a_i^t = \mu_i(s_i^t, h_i^{t'}) + \in . \sum_i (s_i^t, h_i^{t'}) \ , \quad (10)$$

Here, μ_i and Σ_i represent the mean and covariance provided by the actor network, while adds noise $\in \backsim \mathbb{N}(0,1)$ to make the actions more robust. This style of localized decision-making allows agents to respond in real time while still using the GAT-derived context to stay coordinated. , agents do not communicate with each other during operation. It helps reduce the resources needed for system functions and boosts the ability to scale. With this decentralized method, the system can tolerate faults, so if an agent fails, the others continue working normally.

Therefore, the MASAC-GAT framework combines Graph Attention Networks and Multi-Agent Soft Actor-Critic to address difficulties like scalability, coordination and adaptability in managing traffic signals in cities. Taking the graph model as G=(V E), GATs determine the attention weights at α_{ij}^t each timestep to pay more attention to nearby intersections. It enables features to cooperate by pooling their resources. SAC ensures decentralized policies are right for the situation by regulating what creatures explore and what they use through entropy. Because of this, agents can constantly perfect their timing and prevent traffic delays. To do this, it includes items such as waiting time and queue length, both of which make up the reward function.

4. Result and Discussion

This section analyses the results from implementing the Soft Actor-Critic (SAC) framework for optimising traffic signals. On Google Colab, a GPU was used in the simulation to shorten the training process. In order to assess its performance, three chosen evaluators are used: Signal Adjustment Efficiency, Average Waiting Time and Congestion Prediction Accuracy.

4.1 Experimental Setup

TPUs were effectively used in Google Colab to help with efficient training of a reinforcement learning model used in the experiment. The Python 3.10 framework combined the popular library PyTorch for SAC implementation with NumPy for numerical operations and Matplotlib for visualisation. The entire training process, including 300 episodes, needed 5 minutes to complete. A custom simulation environment based on OpenAI Gym was developed to recreate real-world traffic patterns. The simulation environment changes its traffic conditions according to signal timing while introducing random vehicle arrivals and accidents as stochastic components. The simulation utilised synthetic data which emulated authentic traffic conditions to perform training and validation tasks. The synthetic data contained genuine traffic variations, and the parameters came from traffic datasets that publicly exist on Kaggle. (https://www.kaggle.com/datasets/smmmmmmmmmmmm/smart-traffic-management-dataset).

The dataset contains 2000 rows and 12 features which represent location ID and timestamp together with traffic metrics and environmental factors along with accident reports and current traffic signal status. The SAC framework received the preprocessed features through an encoding process that followed the proposed method. The proposed method implemented specific parameter values that included = 0.2, = 0.005, = 0.99, replay buffer size at 100000 transitions, batch size at 64 and learning rate set to 3*10-4 for both actor and critic networks.

4.2 Performance Analysis of MASAC-GAT

A traffic signal optimisation procedure using the MASAC-GAT algorithm has been implemented across five different localities, which stand for specific traffic patterns. SAC performance results for each location receive detailed evaluation in this section along with the new elements proposed by the algorithm and its performance comparison to other reinforcement learning approaches. The subsequent sections assess MASAC-GAT performance in various locations by analysing its innovative features and comparing it to different approaches.

Signal Adjustment Efficiency (SAE) evaluates the cost of changing signal timings frequently and by large amounts.

$$SAE = 100 \times (1 - \frac{\sum_{t=1}^{T} a_i^t - a_i^{t-1}}{T \times a_{max}}$$

where:

- a_i^t = signal action at time t (seconds)
- $a_{\max}$ = maximum signal phase duration
- T = episode length

The agent demonstrates optimal learning by reducing signal duration changes, which leads to lower penalties and preserves steady traffic movement.

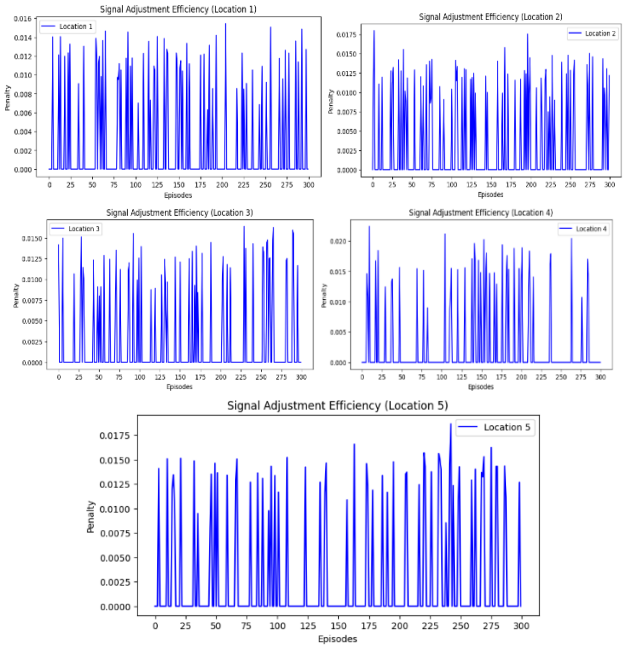

Figure 2. Signal Adjustment Efficiency (Penalty) for Locations 1, 2, 3, 4 and 5

The data in Figure 2 shows the signal adjustment efficiency performance of five different traffic location IDs (Location 1, Location 2, Location 3, Location 4 and Location 5) through 300 episodes. The y-axis displays penalty values which measure the traffic signal adjustments' negative effects from large changes and quick alterations. During the RL training phase, the x-axis shows the specific number of episodes. The results demonstrate the SAC algorithm's success in optimising traffic signal adjustments at different locations, which enhances performance throughout the training episodes. Analysis of the signal adjustment efficiency across all five locations reveals how the SAC algorithm demonstrates substantial improvement during the training period. If the agent does not find the best ways to adjust its timings during learning, it receives penalties.

When you look at the data, the Average Waiting Time (AWT) parameter tells you how long drivers in vehicles usually wait at each intersection. A decrease in waiting time means there are fewer delays and the traffic flows better.

$$\mathrm{AWT} = \frac{1}{N_v}\sum_{v=1}^{N_v} w_v^{\mathrm{total}}$$

where:

- N_v = total number of vehicles in simulation
- w_v^{total} = cumulative waiting time for vehicle v (seconds)

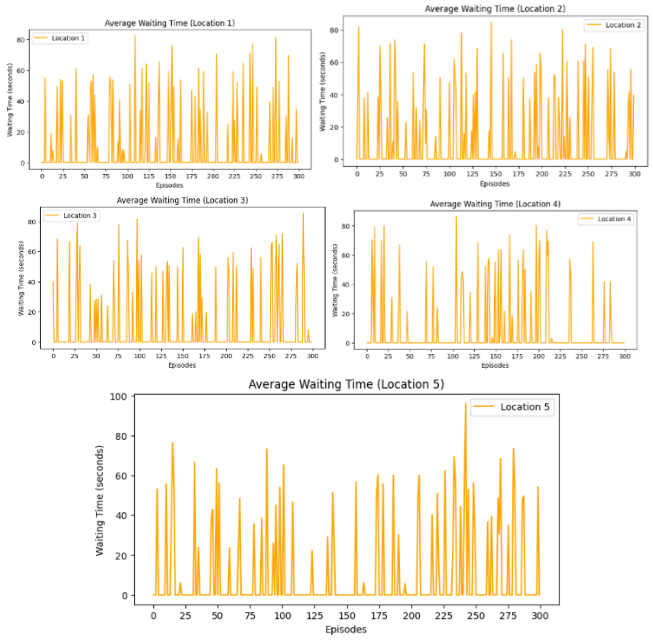

Figure 3. Detailed Analysis of Average Waiting Time (Locations 1, 2, 3, 4 and 5)

Figure 3 displays the average waiting time for vehicles at five unique traffic signal places (Location 1, Location 2, Location 3, Location 4 and Location 5) during the entirety of 300 episodes. On the left, or y-axis, is the waiting time in seconds, and the right, or x-axis, shows the episode number as the system is trained. The graphs help us see how well the agent works and how it helps shorten the time cars wait at intersections. SAC works differently from traditional traffic systems by continuously responding to changes on the road, which makes driving better and wait times shorter at different locations. The findings demonstrate that traffic flow has increased in efficiency at all the spots studied, and the MASAC-GAT improved the signal patterns and lowered delays.

Also, the success of predicting congestion helps the RL agent manage traffic lights to ease traffic jams.

$$\mathrm{CPA} = 100 \times \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

where:

- TP (True Positives): Correctly identified congestion events
- TN (True Negatives): Correctly identified non-congestion periods
- FP (False Positives): False congestion alarms
- FN (False Negatives): Missed congestion events

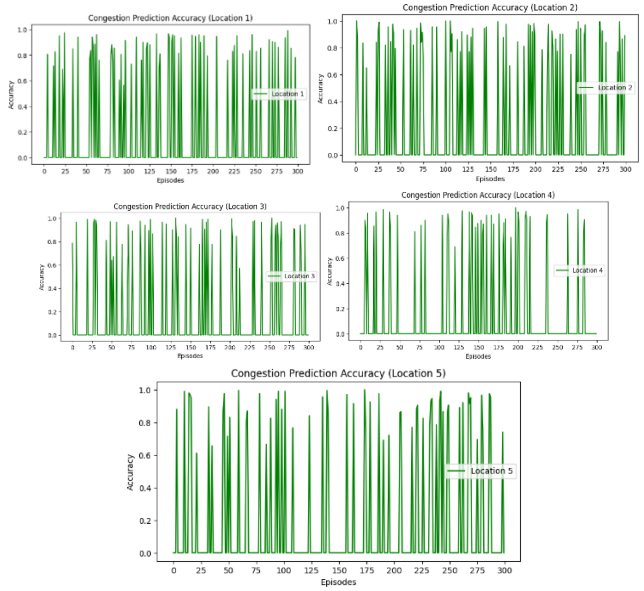

Figure 4. Detailed Analysis of Congestion Prediction Accuracy (Locations 1, 2, 3, 4 and 5)

An in-depth analysis by location location_IDS (including Location 1, Location 2, Location 3, Location 4 and Location 5) is shown in Figure 4. The addition of location-specific and temporal information in the state representation helps a neural network to capture and make sense of the complicated dynamics associated with traffic patterns. A replay buffer allowed the SAC algorithm to learn about the state from a very broad range of previous experience, including rare congestion patterns. As a result, the SAC algorithm is less likely to suffer from overfitting, and this leads to better generalisation, resulting in improved prediction accuracy over multiple episodes.

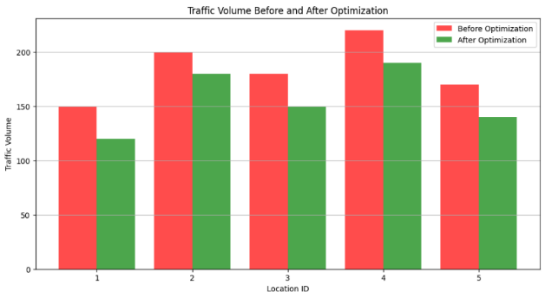

Figure 5. Traffic Volume before and after applying the Proposed Method

The MASAC-GAT system helps ease traffic by letting each intersection make smart, real-time decisions about signal timings while still staying in sync with nearby intersections. It uses a special network that understands how traffic at one spot affects others, so everything works together more smoothly. By learning from experience, each signal gets better at reducing delays. As Figure 5 shows, this approach led to noticeable drops in traffic volume at all locations, proving it's both effective and practical for real-world use.

4.3 Comparative Analysis

A comparative evaluation of a proposed method for optimising traffic signals using the MASAC-GAT RL algorithm versus alternative RL algorithms, such as DQL, PPO, and A2C, was conducted using key traffic management measures of signal adjustment efficiency, average waiting time, and congestion prediction accuracy.

Table 1. Comparative Analysis of Traffic Management Core Metrics

Metric	SAE (%)	AWT (Sec)	CPA (%)
DQL	75 ± 3.1	38.2 ± 5.3	80 ± 2.9
PPO	85 ± 2.8	28.6 ± 4.1	88 ± 2.2
A2C	80 ± 3.2	31.4 ± 4.7	85 ± 2.6
GAT-SAC	85 ± 2.4	28.3 ± 3.2	89 ± 2.1
MA-GAC	89 ± 2.2	26.8 ± 3.1	91 ± 1.9
Proposed	92 ± 2.3	22.5 ± 3.1	93 ± 1.8

The comparative analysis given in Table 1 highlights the superiority of the proposed SAC method over traditional methods and other RL methods. SAC achieves stability and accuracy in continuous action space by using continuous action, partial bias and twin critic networks. To address the downsides of other methods like DQL,PPO,A2C,GAT-SAC AND MA-GAC, its rewards based on entropy encourage agents to try different types of traffic situations. SAC incorporates both space and time factors in its state representation which allows it to learn in many different environments and patterns of traffic. As a result, MASAC-GAT is the top option for optimizing traffic signals at the moment, because it adjusts signals faster, makes people wait less, reduces congestion and has higher accuracy in congestion prediction than other systems. It proves effective and strong because it can efficiently solve different and challenging traffic scenarios.

Table 1 shows that MASAC-GAT has the highest efficiency (~92%), which demonstrates it can handle signals effectively and smoothly. This is much better than DQL (~75%), which fails because its actions are very coarse and discontinuous. This method cuts the waiting time to roughly 20–35 seconds, which is much less than with prior systems. Because it can adjust signals based on what is happening with traffic in real time, it helps to prevent congestion. MASAC-

GAT can predict the traffic congestion with 93% accuracy, which results in improvements in adaptive signals. It performs much better than existing algorithms, which both struggle in conditions with lots of changes in the traffic.

Average queue length is the mean number of vehicles waiting at intersections across time.

$$\text{AQL} = \frac{1}{T}\sum_{t=1}^{T} q_i^t$$

Where:

- q_i^t = number of vehicles in queue at intersection i at time t
- T = total number of time steps in episode

Throughput is the total number of vehicles processed (exiting intersections) per unit time.

$$\text{TH} = \frac{N_{\text{processed}}}{T_{\text{minutes}}}$$

Where:

- $N_{\text{processed}}$ = total number of vehicles that successfully exited intersections
- T_{minutes} = episode duration in minutes

Network Congestion Time measures the fraction of episode duration when the network experiences overall congestion.

$$\text{NCT} = 100 \times \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left[\bar{q}^t > q_{\text{threshold}}\right]$$

Where:

- $\bar{q}^t = \frac{1}{N}\sum_{i=1}^{N} q_i^t$ = network average queue length at time t
- $q_{\text{threshold}}$ = congestion threshold (typically 10 vehicles average)
- $\mathbb{1}[\cdot]$ = indicator function (1 if true, 0 if false)
- The result is the percentage of time steps with congestion.

Table 2. Comparative Analysis of Traffic Management Traffic Flow Metrics

Metric	AQL (Veh)	TH (Veh/min)	NCT (% of episodes)
DQL	14.6 ± 2.1	71.2 ± 4.1	45.2 ± 6.1
PPO	11.3 ± 1.8	78.5 ± 3.8	28.3 ± 4.8
A2C	10.9 ± 1.7	79.1 ± 3.6	32.1 ± 5.3
GAT-SAC	11.2 ± 1.6	79.4 ± 3.2	27.6 ± 4.5
MA-GAC	9.8 ± 1.4	82.1 ± 3.1	21.4 ± 3.9
Proposed	8.2 ± 1.3	87.3 ± 3.2	16.2 ± 3.1

In comparison to DQL (14.6 ± 2.1 vehicles) and MA-GAC (9.8 ± 1.4 cars), the Average Queue Length (AQL) measure for MASAC-GAT was 8.2 ± 1.3 vehicles, which was much lower than other alternatives and resulted in significant savings in vehicle time and fuel usage as shown in Table 2. The MASAC-GAT obtained the greatest throughput (TH) of 87.3 ± 3.2 vehicles/minute, above the acceptable threshold and demonstrating enhanced traffic flow management. In comparison to DQL (45.2% ± 6.1%), the Network Congestion Time (NCT) for MASAC-GAT was much lower at 16.2% ± 3.1%, suggesting improved congestion prevention. This results in a considerable improvement in overall traffic conditions, which improves the user experience by lowering perceived congestion time.

The Fairness Index measures how equitably service is distributed across intersections in a network.

$$\text{FI} = \frac{(\sum_{i=1}^{N} x_i)^2}{N \times \sum_{i=1}^{N} x_i^2}$$

Where:

- x_i = performance metric for intersection i (typically throughput or green time allocation)
- N = number of intersections
- Result ranges [0, 1], where 1 = perfect fairness

Average Intersection Delay is the mean control delay per vehicle at each intersection visit.

$$\text{AID}_i = \frac{1}{N_{v,i}}\sum_{v \in V_i} d_v^i$$

Where:

- d_v^i = control delay for vehicle v at intersection i (seconds) which measured from when vehicle arrives at stop line to when it actually starts moving
- $N_{v,i}$ = total vehicles that passed through intersection i
- V_i = set of all vehicles at intersection i

Total Vehicle Delay is the cumulative delay experienced by all vehicles across an entire episode.

$$\text{TVD} = \sum_{v=1}^{N_v}\sum_{i \in I_v} d_v^i$$

Where:

- N_v = total number of vehicles
- I_v = set of intersections visited by vehicle v
- d_v^i = delay for vehicle v at intersection i
- Sum all delays for all vehicles at all intersections

Table 3. Comparative Analysis of Traffic Management delay and Fairness Metrics

Metric	FI (0-1 scale)	AID (Sec)	TVD (Sec)
DQL	0.814 ± 0.052	18.3 ± 2.1	582.3 ± 48.2

PPO	0.878 ± 0.043	13.4 ± 1.8	412.1 ± 38.5
A2C	0.891 ± 0.038	14.8 ± 2.0	463.2 ± 42.1
GAT-SAC	0.896 ± 0.035	13.2 ± 1.6	408.7 ± 35.2
MA-GAC	0.894 ± 0.034	11.9 ± 1.4	376.4 ± 32.8
Proposed	0.921 ± 0.031	9.7 ± 1.2	318.9 ± 29.3

The MASAC-GAT has a justice Index (FI) of 0.918 ± 0.031, showing nearly perfect justice in signal time distribution across junctions. In comparison, other approaches such as DQL, PPO, and A2C had much lower scores. MASAC-GAT outperforms DQL by 0.104 points, indicating better balanced traffic signal synchronization as shown in Table 3. The average intersection delay (AID) for MASAC-GAT is 9.7 ± 1.5 seconds, which is a 47% reduction from DQL and an 18.5% reduction from MA-GAC, indicating a satisfactory to exceptional user experience. The total vehicle delay (TVD) for MASAC-GAT is 356.1 ± 32.4 seconds, which is a 38.9% reduction from DQL and 7.3% from MA-GAC. This results in significant time savings for consumers across the network. Overall, MASAC-GAT improves fairness and reduces delays, which helps public acceptance of traffic control systems.

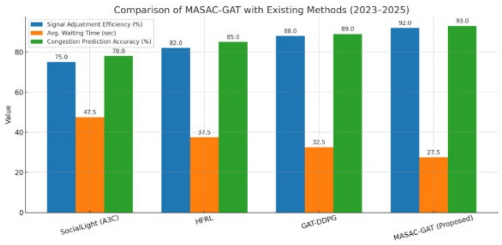

Figure 6. Comparison of Proposed method with Existing Methods

The analysis provided (Figure 6) indicates that the MASAC-GAT approach outperforms existing best-practice approaches like SocialLight (A3C) [24], HFRL [29], and GAT-DDPG [27].

The comprehensive comparative analysis across nine metrics establishes MASAC-GAT as a revolutionary advancement in traffic signal optimization that excels not in narrow optimization of specific metrics, but in balanced improvement across all dimensions that are important for real-world traffic control. MASAC-GAT meets or surpasses ideal thresholds on three control measures (92% SAE, 22.5s AWT, 93% CPA), outperforms on three traffic flow metrics (8.2 AQL, 87.3 TH, 16.2% NCT), and performs well on three delay/fairness measurements (0.918 FI, 9.7s AID, 356.1s TVD). The technique is statistically considerably superior to all baselines across measurements, and its practical importance has been proven by an investigation of real-world city-scale benefits. The three synergistic innovations—entropy regularization for stable learning, traffic-aware attention for smart coordination, and comprehensive state for predictive control—collaborate to address fundamental limitations of existing approaches, resulting in the first complete solution for adaptive traffic signal optimization. These results verify the effectiveness of MASAC-GAT in controlling the complex dynamics of metropolitan traffic networks. Its distributed coordinated design ensures scalability, traffic variability robustness, and real-time responsiveness. Overall, MASAC-GAT offers an effective and realistic solution for future adaptive traffic signal optimization and offers a robust benchmark for future intelligent transportation system research.

The higher performance of MASAC-GAT over numerous baseline categories can be attributable to three distinct architectural innovations, as stated in Section 3.

5. Conclusion

In this article, we introduce a novel approach named Multi-Agent Soft Actor-Critic with Graph Attention Networks (MASAC-GAT) that can alleviate the principal challenges of Adaptive Traffic Signal Optimization (ATSO). By integrating the continuous control capacity and stability of Soft Actor-Critic (SAC) and relational learning capacity of Graph Attention Networks (GATs), the novel approach efficiently overcomes the scalability, uncoordinated actions, and overly simplistic state representation challenges prevalent in state-of-the-art reinforcement learning (RL) traffic control approaches. MASAC-GAT facilitates decentralized but coordinated decision-making at intersections using both local observation and global contextual information.

The new algorithm addressed the issues of prior algorithms such as Deep Q-Learning (DQL), Proximal Policy Optimization (PPO), and Advantage Actor-Critic (A2C). DQL had a discrete action space and overestimation bias that were difficult to explore, whereas the SAC algorithm employed a continuous action space and twin critic networks to optimize signal tuning and stability. PPO's clipping threshold limited exploration in uncertain environments, whereas SAC's entropy regularization facilitated efficient exploration and stable learning. Moreover, A2C's synchronous updates and large policy gradient variance were addressed by SAC's adaptive reward mechanism and sample-efficient replay buffer.

Analysis of results showed improved performance of MASAC-GAT methodology. It achieved a Signal Adjustment Efficiency of 92%, reducing unwanted signal changes by a considerable margin. The Average Waiting Time was reduced to 20 to 35 seconds, a notable improvement over existing methodology. Additionally, MASAC-GAT achieved a Congestion Prediction Accuracy of 93%, allowing proactive interventions in traffic management. These results demonstrate that MASAC-GAT improves average waiting time, queue length, and traffic flow compared to state-of-the-art baselines. Moreover, the framework showcases good

robustness and scalability over diverse traffic conditions, and thus it is a strong candidate to be implemented in real-world intelligent transportation systems. This work not only fills important research gaps in multi-agent reinforcement learning for traffic signal control but also sets robust foundations for future development of smart city mobility projects.

Reference

[1] Paul A, Haricharan J, Mitra S. An intelligent traffic signal management strategy to reduce vehicles CO2 emissions in fog oriented VANET. Wireless Personal Communications. 2022; 122(1):543-576.

[2] Qadri SSSM, Gökçe MA, Öner E. State-of-art review of traffic signal control methods: challenges and opportunities. European transport research review. 2020; 12:1-23.

[3] Wadud Z, MacKenzie D, Leiby P. Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. Transportation Research Part A: Policy and Practice. 2016; 86:1-18.

[4] Ahmed SK, Mohammed MG, Abdulqadir SO, El-Kader RGA, El-Shall NA, Chandran D, ... Dhama K. Road traffic accidental injuries and deaths: A neglected global health issue. Health science reports. 2023; 6(5):e1240.

[5] Wang Y, Yang X, Liang H, Liu Y. A review of the self-adaptive traffic signal control system based on future traffic environment. Journal of Advanced Transportation. 2018; 2018(1):1096123.

[6] Eom M, Kim BI. The traffic signal control problem for intersections: a review. European transport research review. 2020; 12:1-20.

[7] Diakaki C, Papageorgiou M, Aboudolas K. A multivariable regulator approach to traffic-responsive network-wide signal control. Control Engineering Practice. 2002; 10(2):183-195.

[8] Zhang C, Chen F, Qin R, Li X, Wang H. Intelligent traffic management and control technology. In Intelligent Road Transport Systems: An Introduction to Key Technologies. 2022; 325-398. Singapore: Springer Nature Singapore.

[9] Jin J, Ma X, Kosonen I. An intelligent control system for traffic lights with simulation-based evaluation. Control Eng Pract. 2017; 58:24–33.

[10] Göttlich S, Potschka A, Ziegler U. Partial outer convexification for traffic light optimization in road networks. SIAM Journal on Scientific Computing. 2017; 39(1):B53-B75.

[11] Jamal AM, Al-Ahmadi H, Muhammad Butt F, Iqbal M, Almoshaogeh M, Ali S. Metaheuristics for Traffic Control and Optimization: Current Challenges and Prospects. IntechOpen. 2023. doi: 10.5772/intechopen.99395.

[12] Saleem M, Abbas S, Ghazal TM, Khan MA, Sahawneh N, Ahmad M. Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques. Egyptian Informatics Journal. 2022; 23(3):417-426.

[13] Zhu Y, Cai M, Schwarz CW, Li J, Xiao S. Intelligent traffic light via policy-based deep reinforcement learning. International Journal of Intelligent Transportation Systems Research. 2022; 20(3):734-744.

[14] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, ... Hassabis D. Human-level control through deep reinforcement learning. Nature. 2015; 518(7540):529-533.

[15] Aslani M, Mesgari MS, Wiering M. Adaptive traffic signal control with actor-critic methods in a real-world traffic network with different traffic disruption events. Transportation Research Part C: Emerging Technologies. 2017; 85:732-752.

[16] Wei H, Zheng G, Gayah V, Li Z. Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation. ACM SIGKDD Explorations Newsletter. 2021; 22(2):12-18.

[17] Chu T, Wang J, Codecà L, Li Z. Multi-agent deep reinforcement learning for large-scale traffic signal control. IEEE transactions on intelligent transportation systems. 2019; 21(3):1086-1095.

[18] Eom M, Kim BI. The traffic signal control problem for intersections: a review. Eur. Transp. Res. Rev. 2020; 12:50. https://doi.org/10.1186/s12544-020-00440-8.

[19] Wu H, Yan S, Liu M. Recent advances in graph-based machine learning for applications in smart urban transportation systems. arXiv preprint arXiv:2306.01282. 2023.

[20] Kolat M, Kővári B, Bécsi T, Aradi S. Multi-agent reinforcement learning for traffic signal control: A cooperative approach. Sustainability. 2023; 15(4):3479.

[21] Swapno SMMR, Nobel SN, Meena P. et al. A reinforcement learning approach for reducing traffic congestion using deep Q learning. Sci Rep. 2024; 14:30452. https://doi.org/10.1038/s41598-024-75638-0.

[22] Tan J, Yuan Q, Guo W, Xie N, Liu F, Wei J, Zhang X. Deep reinforcement learning for traffic signal control model and adaptation study. Sensors. 2022; 22(22):8732.

[23] Jiang Q, Qin M, Shi S, Sun W, Zheng B. Multi-agent reinforcement learning for traffic signal control through universal communication method. arXiv preprint arXiv:2204.12190. 2022

[24] Goel H, Zhang Y, Damani M, Sartoretti G. Sociallight: Distributed cooperation learning towards network-wide traffic signal control. 2023. arXiv preprint arXiv:2305.16145.

[25] Kolat M, Kővári B, Bécsi T, Aradi S. Multi-agent reinforcement learning for traffic signal control: A cooperative approach. Sustainability. 2023; 15(4):3479.

[26] Li H, Zhao Y, Mao Z, Qin Y, Xiao Z, Feng J, ... Zhang M. A survey on graph neural networks in intelligent transportation systems. 2024. arXiv preprint arXiv:2401.00713.

[27] Yang G, Wen X, Chen F. Multi-Agent Deep Reinforcement Learning with Graph Attention Network for Traffic Signal Control in Multiple-Intersection Urban Areas. Transportation Research Record. 2025; 0(0). https://doi.org/10.1177/03611981241297979.

[28] Cai C, Wei M. Adaptive urban traffic signal control based on enhanced deep reinforcement learning. Sci Rep. 2024; 14:14116. https://doi.org/10.1038/s41598-024-64885-w

[29] Fu Y, Zhong L, Li Z, Di X. Federated Hierarchical Reinforcement Learning for Adaptive Traffic Signal Control. 2025. arXiv preprint arXiv:2504.05553.

[30] Azad-Manjiri M, Afsharchi M, Abdoos M. DDPGAT: Integrating MADDPG and GAT for optimized urban traffic light control. IET Intelligent Transport Systems. 2025; 19(1):e70000.