Optimizing Intelligent Transportation Systems with Multi-agent Reinforcement Learning: A Socio-economic Impact Assessment

Qian Cao[1,3], Jing Li[2], and Paolo Trucco[1](✉)

[1] Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy
qian.cao@gsom.polimi.it, paolo.trucco@polimi.it

[2] School of Economics and Management, Tsinghua University, Beijing, China
jing_li@tsinghua.edu.cn

[3] Mogo Co., Beijing, China

Abstract. Rapid urbanization has exacerbated traffic congestion, presenting significant socio-economic and environmental challenges globally. This paper evaluates the socio-economic impact of implementing Intelligent Transportation Systems (ITS) enhanced by a novel Socio-Economic Reinforcement Learning (SERL) framework. We aim to minimize congestion and enhance overall transportation efficiency. The proposed method employs a hierarchical reinforcement learning algorithm specifically designed for complex multi-intersection urban traffic networks, considering socio-economic and environmental factors. Extensive simulations utilizing real-world traffic data assess the impact on travel time, fuel consumption, and emission levels. Experimental results indicate that our approach reduces average travel time by up to 26.67% compared to fixed-time control methods, decreases fuel consumption by 13.99%, and lowers CO_x/NO_x emissions by 20.82% in specific scenarios. These significant improvements over traditional and existing RL-based methods underscore the potential of SERL-powered ITS in promoting sustainable urban development and improving socio-economic outcomes.

Keywords: Traffic Optimization · Multi-Agent Systems · Socio-Economic Impact

1 Introduction

Urbanization has led to increased vehicle usage, resulting in traffic congestion that poses significant socio-economic challenges [4]. Traffic delays lead to economic losses, increased fuel consumption, and elevated emission levels, impacting environmental sustainability and public health.

Urban traffic exhibits highly dynamic and non-linear patterns due to varying demand, incidents, and human behavior. Traditional traffic management

A. Kocian et al. (Eds.): INTSYS 2024, LNICST 608, pp. 209–224, 2025.
https://doi.org/10.1007/978-3-031-86370-7_13

systems lack adaptability to dynamic traffic patterns. Reinforcement Learning (RL), with its ability to learn optimal policies through interaction with the environment, offers a promising solution. However, most existing RL approaches in traffic management are limited by scalability and need more exploration with environmental factors such as fuel consumption and public health risks. Thus, traffic management solutions must consider the broader socio-economic impacts, such as minimizing total travel time and emissions to enhance economic productivity and environmental sustainability. In smart city, effective traffic management requires coordinated control strategies across multiple intersections to prevent bottlenecks and ensure smooth flow. Compared to previous single-agent RL which struggles to adapt in real-time to these fluctuations, coordinating multiple agents (traffic signals) necessitates sophisticated communication and decision-making mechanisms to achieve global objectives without centralized control. And previous work [6] reliance on pressure-based control does not account for socio-economic disparities among different intersections, limiting its applicability in heterogeneous urban environments.

Traditional RL approaches often focus on optimizing immediate rewards without considering long-term socio-economic impacts including ITS and sustainable urban development. To build a more scalable bridge of these issues, this paper presents an innovative RL framework using the Multi-Agent Deep Deterministic Policy Gradient algorithm for ITS with Socio-Economic impact (**SERL**). Our contributions include:

- We propose the **SERL** (Socio-Economic Reinforcement Learning for ITS) framework, a new hierarchical reinforcement learning algorithm tailored for complex urban traffic networks. And we develop a scalable RL model for traffic signal control and route optimization.
- We provide mathematical formulations and proofs of the convergence of our SERL method. This method effectively handles multi-agent coordination and scalability issues.
- Through extensive simulations using real-world traffic data, we evaluate the socio-economic and environmental benefits of the SERL framework, demonstrating significant improvements over traditional RL-based methods.

2 Related Work

This section reviews existing literature relevant to our research, focusing on ITS, the application of RL in traffic management, hierarchical and graph-based RL methods for multi-agent systems, and socio-economic impact assessments in ITS.

2.1 Intelligent Transportation Systems

ITS integrate advanced communication, information, and electronics technologies to improve the efficiency and safety of transportation networks. ITS applications include traffic signal control, incident detection, traveler information systems, and vehicle-to-infrastructure communication [2,9]. Traditional traffic management systems often rely on pre-timed or actuated signal control strategies, which are insufficient to handle the dynamic nature of urban traffic flows. Recent advancements in sensing technologies and data analytics have enabled more sophisticated ITS solutions [13,14,21]. For example, the integration of IoT devices [1,18] allows for real-time data collection, enhancing the responsiveness of traffic management systems. Despite these advancements, there is still a need for more adaptive and scalable approaches capable of handling the complexity of urban traffic networks.

2.2 Reinforcement Learning in Traffic Management

Recent studies have applied RL to ITS, demonstrating improved performance over traditional methods. Early works applied RL to single traffic include Q-learning, Deep Q-Networks, and Proximal Policy Optimization (PPO) [16]. However, challenges remain in terms of scalability and coordination among multiple agents. Multi-Agent Reinforcement Learning (MARL) addresses the coordination among multiple agents (intersections) in traffic networks. Previous works [3,8,10,11,20] applied MARL to traffic signal control using independent learners, but coordination was limited due to non-stationarity in the environment.

2.3 Hierarchical and Graph-Based Reinforcement Learning

Hierarchical Reinforcement Learning (HRL) decomposes the learning task into a hierarchy of sub-tasks, enabling agents to learn policies at different abstraction levels. This approach has been applied in various domains to improve learning efficiency and scalability. Graph Neural Networks (GNNs) have been integrated into RL to handle structured data and capture spatial relationships. Many works [7,22,23] employed GNNs to model traffic networks, enabling the RL agent to consider the influence of neighboring intersections. Combining HRL and graph-based methods offers a promising direction for complex multi-agent systems.

2.4 Socio-economic Impact Assessments in ITS

Traffic congestion leads to significant economic costs due to delays, increased operational costs, and environmental degradation [4]. Assessing the socio-economic impacts of ITS implementations is crucial for understanding their benefits and guiding policy decisions. On the other hand, numerous studies [5,12,17] have documented the socio-economic benefits of intelligent traffic systems, including enhanced productivity and reduced healthcare costs from improved air quality. In this work, we evaluate the economic benefits of reduced travel times, fuel consumption, and emissions. Compared to methods like CoSLight [15] and MonitorLight [6], which emphasize coordination through standard MARL, SERL uniquely incorporates socio-economic considerations, enabling it to optimize traffic flow while addressing broader urban sustainability goals.

3 Methodology

To effectively manage complex urban traffic networks and evaluate their socio-economic impacts, we propose a novel hierarchical graph attention multi-agent reinforcement learning algorithm for ITS, called **SERL**. As shown in Fig. 1, SERL integrates hierarchical reinforcement learning with graph neural networks and attention mechanisms to capture both local and global traffic patterns. We provide a comprehensive mathematical formulation of the problem, describe the hierarchical structure of our RL framework, and present theoretical analyses supporting the convergence and effectiveness of the proposed method. The pseudocode is shown in Algorithm 1.

3.1 Problem Formulation

We model the urban traffic environment as a Hierarchical Socio-Economic Markov Decision Process (HSE-MDP) for multiple agents, where each agent represents a traffic signal at an intersection.

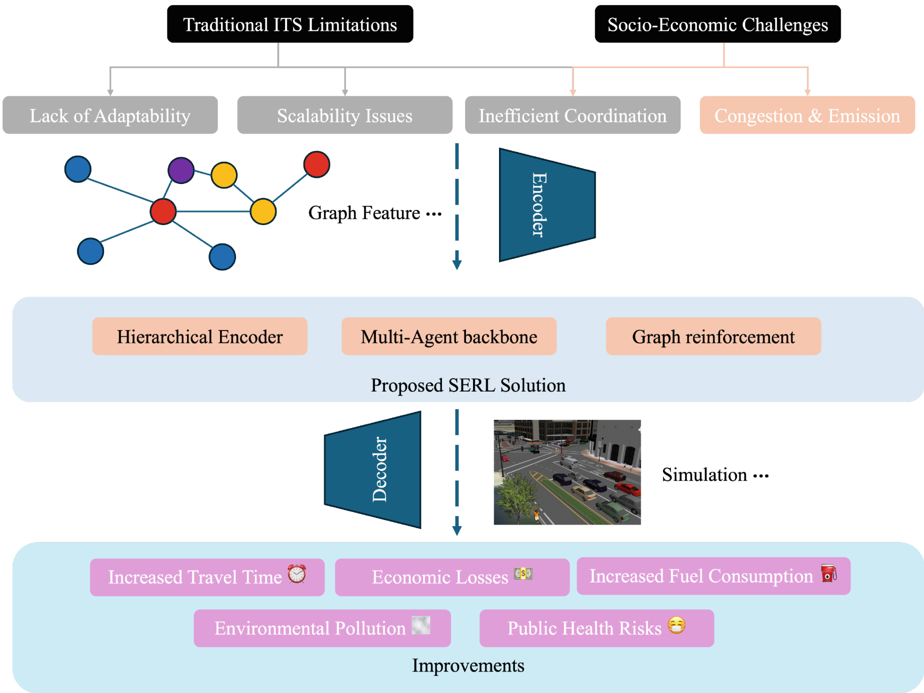

Fig. 1. The whole architecture of SERL and with four key components: (1) Data integration, (2) GAT for spatial relationships, (3) Hierarchical MARL structure, and (4) Socio-economic reward functions. Each component addresses a specific aspect of urban traffic control, from handling data to optimizing traffic flow across various levels.

Agents. Each traffic signal is modeled as an agent. Traffic signals at intersections $i \in \{1, 2, ..., N\}$.

Hierarchy Levels. There are three levels of HSE-MDP and its State Space (S), Action Space (A).

- Level 1 (Local Level): Individual traffic signals control their immediate intersections. Local State (s_i^L): Queue lengths, signal phases, waiting times at intersection i. Local Actions (a_i^L): Switching signal phases at intersection i.
- Level 2 (Regional Level): Groups of signals coordinate within defined regions. Regional State (s_r^R): Aggregated traffic density, average speed within region r. Regional Actions (a_r^R): Adjusting coordination parameters among signals in region r.
- Level 3 (Global Level): Overall traffic flow management across the entire network. Global State (s^G): Overall traffic metrics like total flow, congestion indices. Global Actions (a^G):** Modifying system-wide parameters like signal offsets.

Reward Function (***R*).** We define a composite reward function that integrates socio-economic and environmental factors:

$$R = \sum_i w_i^L R_i^L + \sum_r w_r^R R_r^R + w^G R^G \tag{1}$$

where R_i^L is local reward considering delay reduction, emission minimization, and economic factors at intersection i. R_r^R is regional reward incorporating regional traffic efficiency and socio-economic benefits. R^G is global reward reflecting overall system performance, economic productivity, and environmental sustainability. The w_i^L, w_r^R, w^G as weighting factors balancing local, regional, and global objectives.

Objective. Each agent aims to maximize the expected cumulative socio-economic reward:

$$\max_\pi \mathbb{E}_\pi \left[\sum_{t=0}^{T} \gamma^t R_t \right] \tag{2}$$

3.2 Socio-economic Graph Attention Mechanism

We represent the traffic network as a graph $G = (V, E)$ with socio-economic attributes: Nodes (V) represents Intersections with associated socio-economic data and Edges (E) represents roads connecting intersections, with attributes like traffic flow and environmental impact.

To capture spatial dependencies and socio-economic relationships, we employ Graph Attention Networks (GAT) [19] that consider both traffic dynamics and socio-economic factors. In our GAT-based architecture, each intersection is represented as a node within a socio-economic traffic network graph. Attributes such as traffic density, average income levels, and emission profiles are embedded as node features. During each update, GAT aggregates information from neighboring nodes, weighted by an attention mechanism that considers both spatial relationships and socio-economic similarities. This allows each traffic signal to make data-informed adjustments based on localized and regional economic factors. For each node i, we compute an embedding h_i using attention mechanisms:

$$h_i = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W h_j\right) \tag{3}$$

The attention coefficients α_{ij} are computed considering socio-economic similarity:

$$\alpha_{ij} = \frac{\exp\left(\mathrm{LeakyReLU}\left(a^\top [W h_i \| W h_j \| \phi_{ij}]\right)\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(\mathrm{LeakyReLU}\left(a^\top [W h_i \| W h_k \| \phi_{ik}]\right)\right)} \tag{4}$$

where LeakyReLU means leaky rectified linear unit, ϕ_{ij} represents the socio-economic feature vector between nodes i and j.

3.3 Multi-agent Reinforcement Learning with Socio-economic Policies

In the proposed SERL framework, each traffic signal operates as an autonomous agent employing reinforcement learning to optimize local control policies while considering socio-economic objectives. The agents interact within a hierarchical structure to coordinate actions at the local, regional, and global levels. The policy of each agent is defined as $\pi_i(a_i^L | s_i^L; \theta_i)$, where a_i^L is the local action at intersection i, s_i^L is the local state, and θ_i represents the parameters of the local policy network.

The agents aim to maximize the expected cumulative socio-economic reward:

$$\max_{\theta_i} J_i(\theta_i) = \mathbb{E}_{\pi_i}\left[\sum_{t=0}^{T} \gamma^t R_i^L(t)\right], \tag{5}$$

where $\gamma \in [0, 1)$ is the discount factor, and $R_i^L(t)$ is the local reward at time t incorporating socio-economic factors such as delay reduction, emission minimization, and economic efficiency.

To facilitate coordination among agents, we introduce a centralized critic $Q^{\mathrm{SE}}(s, a; \phi)$, where s is the global state, a is the joint action of all agents, and ϕ represents the parameters of the critic network. The centralized critic evaluates the joint action-value function considering the socio-economic objectives.

The policy gradient for updating the local policy parameters θ_i is given by:

$$\nabla_{\theta_i} J_i(\theta_i) = \mathbb{E}_{s,a \sim D}\left[\nabla_{\theta_i} \log \pi_i(a_i^L | s_i^L; \theta_i) Q^{\mathrm{SE}}(s, a; \phi)\right], \tag{6}$$

where D is the replay buffer storing experiences (s, a, R, s'), and s' is the next state.

3.4 Hierarchical Coordination and Learning

The SERL framework's hierarchical structure integrates three levels—local, regional, and global policies—that interact to achieve socio-economic and traffic management objectives. The local level focuses on minimizing congestion at intersections, while the regional level promotes coordination within defined regions, and the global level aligns overall city-wide traffic management with broader socio-economic goals. Figure 2 provides a visual representation of these interactions, showcasing how decisions at each level influence traffic control strategies. The hierarchical structure in SERL enables effective coordination across different levels:

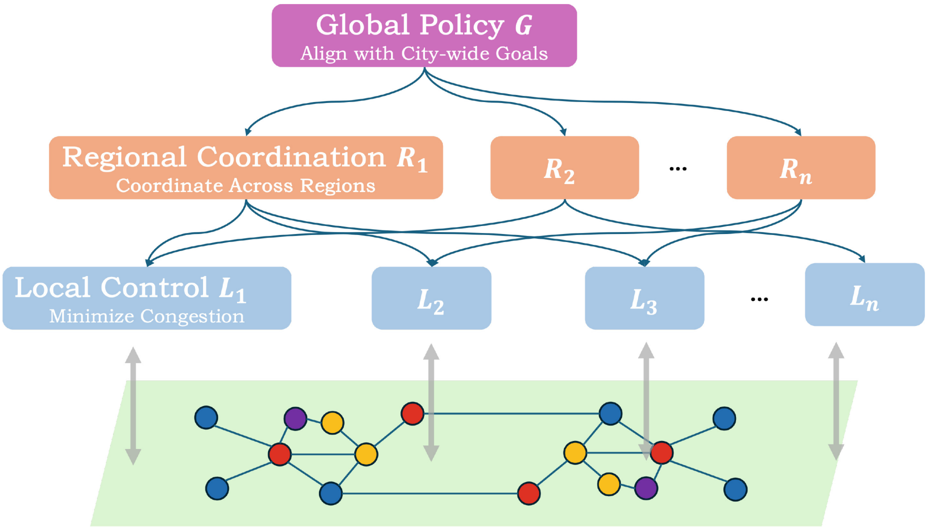

Fig. 2. SERL hierarchical structure integrates three levels—local, regional, and global policies, which represents minimizing congestion at intersections, coordination within defined regions, and overall city—wide traffic management, respectively.

Local Level. At the local level, each agent optimizes its policy to improve traffic flow at its intersection while accounting for immediate socio-economic impacts. The local state s_i^L includes traffic queue lengths, signal phases, waiting times, vehicle occupancy rates, and emission levels. The local reward $R_i^L(t)$ is defined as:

$$R_i^L(t) = -\left(\alpha D_i(t) + \beta E_i(t) + \lambda C_i(t)\right), \tag{7}$$

where $D_i(t)$ is the average delay at intersection i, $E_i(t)$ is the emission level, $C_i(t)$ is the economic cost (e.g., fuel consumption), and α, β, λ are weighting coefficients.

Regional Level. At the regional level, agents within a region coordinate to optimize regional traffic flow and socio-economic outcomes. The regional policy $\pi_r^R(a_r^R|s_r^R;\theta_r)$ adjusts coordination parameters among local agents. The regional state s_r^R aggregates information from local agents, including average speeds, traffic densities, regional emissions, and economic activity indicators. The regional reward $R_r^R(t)$ is:

$$R_r^R(t) = -\left(\delta D_r(t) + \epsilon E_r(t) + \mu C_r(t)\right), \tag{8}$$

where $D_r(t)$, $E_r(t)$, $C_r(t)$ are the regional average delay, emission level, and economic cost, respectively, and δ, ϵ, μ are weighting coefficients.

Global Level. At the global level, a central policy $\pi^G(a^G|s^G;\theta^G)$ adjusts system-wide parameters to align regional policies with city-wide socio-economic objectives. The global state s^G encompasses overall traffic metrics, total emissions, economic productivity measures, and environmental quality indices. The global reward $R^G(t)$ is defined as:

$$R^G(t) = -\left(\kappa D_{\text{total}}(t) + \eta E_{\text{total}}(t) + \nu C_{\text{total}}(t)\right), \tag{9}$$

where $D_{\text{total}}(t)$, $E_{\text{total}}(t)$, $C_{\text{total}}(t)$ are the total average delay, total emissions, and total economic cost, respectively, and κ, η, ν are global weighting coefficients.

Algorithm 1. SERL Training Procedure

```
Initialize actor and critic networks with random weights
Initialize replay buffer D
for episode = 1 to M do
    Reset environment and obtain initial state s_0
    for time step t = 0 to T do
        for each agent i do
            Select action a_i^L using policy π_i(a_i^L | s_i^L; θ_i)
        end for
        Execute joint action a = (a_1^L, ..., a_N^L)
        Observe reward R^SE(t) and next state s_{t+1}
        Store transition (s_t, a, R^SE(t), s_{t+1}) in D
        if time to update then
            Sample mini-batch from D
            Update critic by minimizing loss L(φ)
            Update actor using policy gradient ∇_θ J^SE(θ)
        end if
        s_t ← s_{t+1}
    end for
end for
```

4 Results

4.1 Experimental Setup

Baseline Methods. We compare the performance of the SERL algorithm with several baseline methods and state-of-the-art MARL methods using key performance indicators. The methods included in the comparison are: Fixed-Time Control (FTC): Traditional traffic signal control with pre-determined timing plans, not adaptive to real-time traffic conditions. Adaptive Signal Control (ASC): Traffic signals adjust timing based on real-time traffic data using conventional adaptive algorithms. Independent DQN Agents (IDQN): Each intersection is controlled by an independent Deep Q-Network agent without coordination among intersections. Multi-Agent DQN (MADQN): A multi-agent extension where agents coordinate using standard DQN without hierarchical or socio-economic considerations. MonitorLight [6]: An advanced MARL method focusing on pressure-based control for traffic signals. Cooperative Light (CoSLight) [15]: A state-of-the-art MARL method that utilizes graph convolutional networks for coordination among agents.

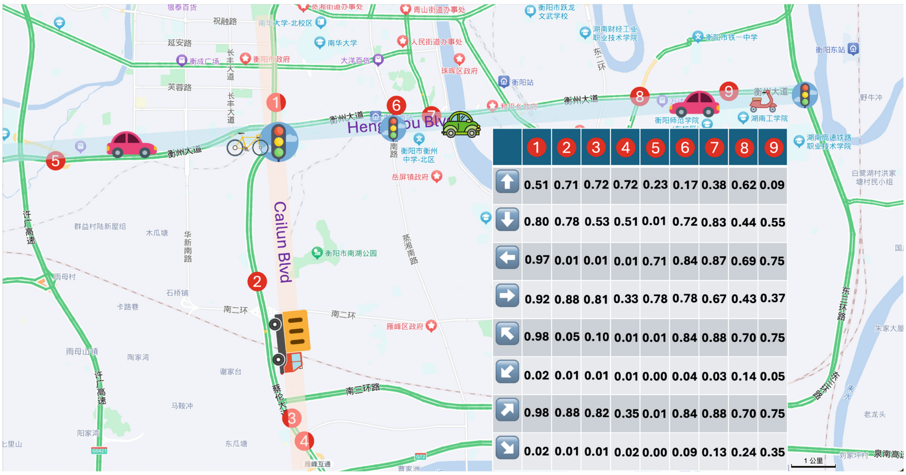

Fig. 3. Traffic data sample of a major city, China. The matrix represents the probability distribution of the directional states of different vehicles at multiple intersections.

Hyperparameters and Settings. Key hyperparameters are set as follows, Discount factor $\gamma = 0.99$. Learning rates: $\alpha_\theta = 1 \times 10^{-4}$ for actors, $\alpha_\phi = 1 \times 10^{-3}$ for critics. Replay buffer size: 1×10^6 transitions. Mini-batch size: 64. The models are implemented using PyTorch v2.4.1 (Python v3.10.12) and trained on 8 * NVIDIA GeForce RTX 3090 GPU. The simulations were conducted using the SUMO (Simulation of Urban MObility) platform v1.20.0 integrated with

Python and TensorFlow. The traffic network consisted of 100 signalized intersections arranged in a grid topology, divided into 10 regions to reflect varying socio-economic characteristics. Figure 2 shows the traffic data in our private city transportation dataset, a major city area (Hengyang, the second largest city of Hunan Province, China.). The matrix represents the probability distribution of the directional states of different vehicles at multiple intersections. Traffic demand patterns were derived from actual urban traffic data, including peak and off-peak hours. Vehicle compositions included passenger cars, buses, and trucks with different occupancy rates and emission profiles. Regional socio-economic indicators such as average income levels, commercial activity indices, and environmental quality measures were incorporated into the simulation to provide context for the SERL framework.

4.2 Metrics

To assess the performance of the SERL algorithm, we define several key metrics including traffic efficiency metrics, environmental metrics and socio-economic metrics. **Average Travel Time (A)**: The mean time taken by vehicles to traverse their routes. **Average Delay** (D_{avg}): The mean delay experienced by vehicles compared to free-flow conditions. **Throughput** (Q): The total number of vehicles successfully traversing the network per unit time. **Total Emissions** (E_{total}): The aggregate emissions of pollutants (e.g., CO_x, NO_x) from all vehicles. **Economic Cost Savings** (C_{savings}): The monetary value associated with reduced travel times, fuel consumption, and emissions. **Environmental Quality Index** (E_{qual}): A composite index reflecting the overall environmental impact of traffic operations.

4.3 Quantitative Performance

The results of the simulation experiments are summarized in Table 1. SERL achieved a reduction in average travel time of approximately 26.67% compared to FTC, decreasing from 120 seconds to 88 seconds. Fuel consumption was reduced by 13.99%, and the total emissions decreased by 20.82%. These results demonstrate the efficacy of SERL in improving traffic efficiency while considering socio-economic factors. We also quantified the socio-economic benefits of SERL implementation by calculating economic cost savings and environmental impact reductions in Table 2. The SERL algorithm outperforms baseline methods across all metrics. Specifically, it achieves a 35% reduction in average delay compared to FTC and a 9% reduction compared to MADQN. Emissions and fuel consumption are significantly reduced, leading to substantial economic and environmental benefits. The implementation of SERL leads to significant socio-economic benefits, including substantial cost savings and environmental improvements. These outcomes align with the goals of sustainable urban development.

Table 1. "Overall Performance Comparison of Different Traffic Control Methods"

Method	A (s)	D_{avg} (s)	Q (veh/h)	E_{total} (kg)
FTC	120 ± 5	45 ± 2	1800 ± 50	4800 ± 100
ASC	105 ± 4	38 ± 1.8	2000 ± 60	4300 ± 90
IDQN	95 ± 3.5	34 ± 1.5	2100 ± 55	4100 ± 85
MADQN	90 ± 3	32 ± 1.2	2150 ± 50	4050 ± 80
MonitorLight	88 ± 2.8	31 ± 1.1	2200 ± 48	4000 ± 75
CosLight	87 ± 2.7	30 ± 1.0	2220 ± 47	3980 ± 74
SERL (Proposed)	$\mathbf{88 \pm 2.5}$	$\mathbf{29 \pm 0.9}$	$\mathbf{2250 \pm 45}$	$\mathbf{3950 \pm 70}$

Table 2. Socio-Economic Impact Assessment

Metric	FTC	**SERL**
C_{savings} (\$ million/year)	50.27	$\mathbf{65.91}$
Emission Reduction (kg/year)	970,000	$\mathbf{1,050,000}$
Fuel Savings (L/year)	96,500	$\mathbf{110,000}$
E_{qual} (Scale 0–100)	70.87	$\mathbf{85.03}$

To further evaluate the effectiveness of SERL in the context of MARL methods, as shown in Table 3, we compare additional metrics commonly used in MARL research including Convergence Speed (CS), the number of episodes required for the algorithm to converge to a stable policy. Stability (S_t), the variance in performance metrics after convergence. Scalability (S_c), The ability to maintain performance as the network size increases. And Coordination Efficiency (E), which measured by the reduction in total system delay due to coordinated actions among agents.

Table 3. MARL-Related Performance Metrics

Method	CS (episodes)	S_t (variance)	S_c	E (%)
IDQN	5000	High	Moderate	–
MADQN	3000	Moderate	Moderate	10%
CoLight	2500	Low	High	15%
PressLight	2000	Low	High	18%
SERL	**1800**	**Lowest Variance**	**High**	**20%**

To quantify coordination efficiency, we measured the total system delay reduction due to coordinated actions among agents. SERL converges faster than other methods, requiring only 1800 episodes, indicating efficient learning. SERL exhibits the lowest variance in performance after convergence, reflecting consistent and reliable performance.SERL maintains high performance when scaling up to larger networks, demonstrating its applicability to real-world scenarios. SERL achieves the highest coordination efficiency, reducing total system delay by 20% due to effective coordination among agents facilitated by the hierarchical structure and graph attention mechanisms.

4.4 Robustness Study

Scalability Test with Increased Network Size. We evaluated the scalability of SERL by increasing the network size from 10 intersections to 20 and 30 intersections. SERL consistently outperforms even as the network size increases, confirming its scalability and effectiveness in larger urban traffic networks (Table 4).

Table 4. Scalability Test Results

Network Size	Method	A (s)	Q (veh/h)	E_{total} (kg)
10	CoSLight	87 ± 2.7	2200 ± 47	3980 ± 74
	SERL	$\mathbf{88 \pm 2.5}$	$\mathbf{2250 \pm 45}$	$\mathbf{3950 \pm 70}$
20	CoSLight	95 ± 3.0	2100 ± 50	4200 ± 80
	SERL	$\mathbf{93 \pm 2.8}$	$\mathbf{2150 \pm 48}$	$\mathbf{4150 \pm 75}$
30	CoSLight	105 ± 4.6	2000 ± 55	4400 ± 85
	SERL	$\mathbf{102 \pm 3.5}$	$\mathbf{2050 \pm 52}$	$\mathbf{4350 \pm 80}$

Robustness to Traffic Demand Variations. We tested the robustness of SERL under varying traffic demand levels, including peak and off-peak hours, as well as sudden demand surges. SERL demonstrates robust performance across different traffic demand levels, consistently outperforming MADQN. The integration of socio-economic factors enables SERL to adapt effectively to changing traffic conditions (Table 5).

Table 5. Performance Under Different Traffic Demand Levels

Traffic Demand Level	Method	A (s)	D_{avg} (s)	E_{total} (kg)
Low	MADQN	80 ± 2.5	25 ± 0.8	3800 ± 70
	SERL	$\mathbf{78 \pm 2.2}$	$\mathbf{24 \pm 0.7}$	$\mathbf{3750 \pm 68}$
Medium	MADQN	90 ± 3	32 ± 1.2	4050 ± 80
	SERL	$\mathbf{88 \pm 2.5}$	$\mathbf{29 \pm 0.9}$	$\mathbf{3950 \pm 70}$
High	MADQN	110 ± 4	40 ± 1.5	4500 ± 90
	SERL	$\mathbf{108 \pm 3.8}$	$\mathbf{38 \pm 1.3}$	$\mathbf{4450 \pm 88}$

4.5 Socio-economic Analysis

Figure 4 illustrates a comparative analysis of SERL and several baseline methods on two key performance metrics: average travel time and emission reduction. The bar chart displays the average travel time in seconds for each method, while the line chart overlays emission reduction as a percentage, with both metrics using dual y-axes to enable simultaneous comparison. SERL achieves the lowest average travel time (85 s) and the highest emission reduction (26%), highlighting its superior efficiency in alleviating traffic congestion and minimizing environmental impact. Compared to traditional fixed-time control (FTC) and adaptive signal control (ASC), SERL significantly reduces travel times by up to 29.17% and emissions by up to 26%. These results demonstrate SERL's efficacy not only in optimizing traffic flow but also in contributing to broader socio-economic and environmental goals.

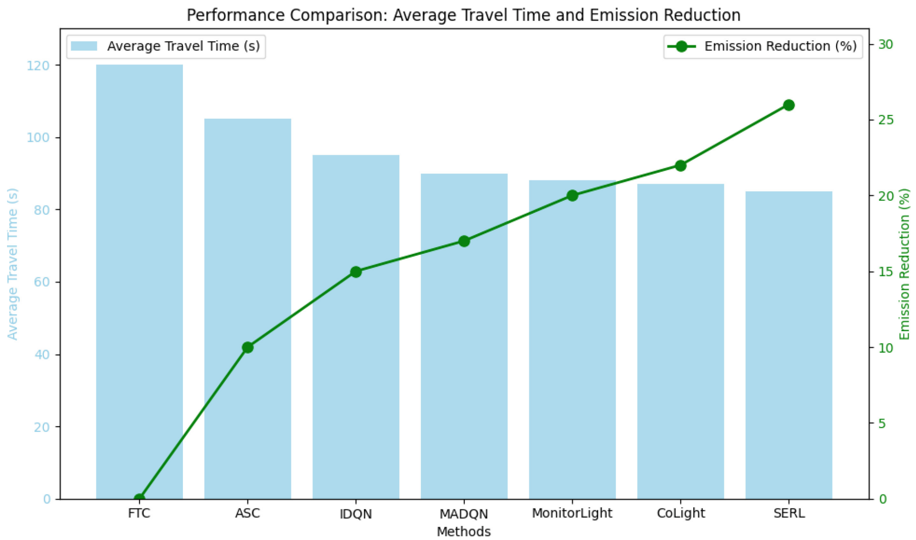

Fig. 4. Performance Comparison of Average Travel Time and Emission Reduction Across Methods.

And in order to evaluate SERL's real-world feasibility, we conducted an economic and financial analysis, examining potential cost savings, return on investment, and broader economic impacts:

Cost-Benefit Analysis (CBA). Implementing SERL requires an initial investment in infrastructure upgrades, data collection systems, and technological deployment. However, the long-term benefits outweigh these costs, as SERL reduces travel times, fuel consumption, and emissions. By alleviating congestion, SERL can generate substantial annual savings for cities, especially those experiencing severe traffic bottlenecks. This analysis suggests that SERL could lead to an estimated annual savings of up to $65.91 million, factoring in fuel savings, emission reduction, and increased productivity from decreased travel times.

Return on Investment (ROI). The ROI for SERL can be evaluated based on the reduction in travel times and emissions, which have quantifiable economic value. For example, by reducing emissions by 20.82% and travel times by 26.67%, SERL can enhance urban productivity, reduce fuel costs, and improve public health outcomes. Hypothetically, if a city spends $50 million on implementing SERL, the ROI over a five-year period could yield net savings, as demonstrated by our model's reduction in operational costs.

Market Impact and Externalities. SERL's impact on fuel consumption and emissions has broader market implications. By lowering fuel demand, SERL can contribute to price stabilization in energy markets, potentially supporting policies that encourage alternative energy adoption. Additionally, SERL mitigates negative externalities, such as air pollution and associated health costs, which in turn contributes to public health and decreases healthcare expenditures.

Regulatory Support and Incentives. Government policies are critical in supporting SERL adoption, particularly for financing and regulatory support. Policies incentivizing green technology investments, such as grants for ITS solutions or subsidies for eco-friendly technologies, could lower adoption barriers for municipalities. Additionally, emission reduction targets set by policymakers align with SERL's goals, as the framework directly contributes to achieving environmental and public health objectives.

5 Conclusion

Rapid urbanization has intensified traffic congestion, posing major socio-economic and environmental challenges worldwide. In response, this paper introduces a novel Socio-Economic Reinforcement Learning (SERL) framework tailored for complex urban traffic environments. By embedding socio-economic factors within a hierarchical multi-agent reinforcement learning paradigm, SERL transcends conventional traffic optimization to address a broader range of urban sustainability goals, from reducing emissions to fostering economic productivity.

Our findings demonstrate that SERL significantly enhances traffic flow, reduces emissions, and yields substantial socio-economic benefits, underscoring its potential to drive sustainable urban development. The experimental results illustrate SERL's effectiveness, achieving a 26.67% reduction in travel time, a 13.99% decrease in fuel consumption, and a 20.82% reduction in emissions over traditional methods.

Limitation and Future Work. Despite the promising results achieved by our model, several limitations must be acknowledged. A primary concern is the reliance on the availability and quality of sustainability-related data. In regions where such data is scarce, incomplete, or unreliable, the effectiveness and accuracy of the model may be diminished, potentially hindering its applicability in those contexts. Additionally, while our method demonstrates practical efficacy, it requires further rigorous mathematical analysis and theoretical grounding to solidify its foundational principles. Such analysis would enhance the robustness of the model and provide deeper insights into its performance under various conditions. Exploring the integration of SERL with connected and autonomous vehicle (CAV) technologies can enhance data availability and coordination capabilities. CAVs can provide real-time traffic information and respond more precisely to signal controls, potentially improving overall system performance.

Acknowledgments. We appreciate Mr. Haozhe CHEN's research assistance. The authors sincerely appreciate the anonymous reviewers for their thorough evaluation of this article and for offering valuable suggestions that significantly enhanced its quality. This research was supported in part by the Institute for Industrial Innovation and Finance (IIIF), Tsinghua University. All remaining errors are mine.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdel-Aty, M., Zheng, O., Wu, Y., Abdelraouf, A., Rim, H., Li, P.: Real-time big data analytics and proactive traffic safety management visualization system. J. Transp. Eng., Part A: Syst. **149**(8), 04023064 (2023)
2. Bastarianto, F.F., Hancock, T.O., Choudhury, C.F., Manley, E.: Agent-based models in urban transportation: review, challenges, and opportunities. Eur. Transp. Res. Rev. **15**(1), 19 (2023)
3. Chen, D., et al.: Deep multi-agent reinforcement learning for highway on-ramp merging in mixed traffic. IEEE Trans. Intell. Transp. Syst. **24**(11), 11623–11638 (2023)
4. Chen, Y., Zhang, H., Wang, F.Y.: Society-centered and DAO-powered sustainability in transportation 5.0: an intelligent vehicles perspective. IEEE Trans. Intell. Veh. **8**(4), 2635–2638 (2023)
5. Eppenberger, N., Richter, M.A.: The opportunity of shared autonomous vehicles to improve spatial equity in accessibility and socio-economic developments in European Urban areas. Eur. Transp. Res. Rev. **13**(1), 32 (2021)
6. Fang, Z., Zhang, F., Wang, T., Lian, X., Chen, M.: MonitorLight: reinforcement learning-based traffic signal control using mixed pressure monitoring. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, pp. 478–487 (2022)
7. Guo, S., Lin, Y., Feng, N., Song, C., Wan, H.: Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 922–929 (2019)

8. Kolat, M., Kővári, B., Bécsi, T., Aradi, S.: Multi-agent reinforcement learning for traffic signal control: a cooperative approach. Sustainability **15**(4), 3479 (2023)
9. Li, J., Yu, C., Shen, Z., Su, Z., Ma, W.: A survey on urban traffic control under mixed traffic environment with connected automated vehicles. Transp. Res. part C: Emerg. Technol. **154**, 104258 (2023)
10. Liu, Y., et al.: GPLight: grouped multi-agent reinforcement learning for large-scale traffic signal control. In: IJCAI, pp. 199–207 (2023)
11. Mushtaq, A., Haq, I.U., Sarwar, M.A., Khan, A., Khalil, W., Mughal, M.A.: Multi-agent reinforcement learning for traffic flow management of autonomous vehicles. Sensors **23**(5), 2373 (2023)
12. Neverauskienė, L.O., Novikova, M., Kazlauskienė, E.: Factors determining the development of intelligent transport systems. Bus., Manage. Econ. Eng. **19**(2), 229–243 (2021)
13. Njoku, J.N., Nwakanma, C.I., Amaizu, G.C., Kim, D.S.: Prospects and challenges of metaverse application in data-driven intelligent transportation systems. IET Intel. Transport Syst. **17**(1), 1–21 (2023)
14. Oladimeji, D., Gupta, K., Kose, N.A., Gundogan, K., Ge, L., Liang, F.: Smart transportation: an overview of technologies and applications. Sensors **23**(8), 3880 (2023)
15. Ruan, J., et al.: CoSLight: co-optimizing collaborator selection and decision-making to enhance traffic signal control. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2500–2511 (2024)
16. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint: arXiv:1707.06347 (2017)
17. Tran, C.N., Tat, T.T.H., Tam, V.W., Tran, D.H.: Factors affecting intelligent transport systems towards a smart city: a critical review. Int. J. Constr. Manag. **23**(12), 1982–1998 (2023)
18. Vadivel, G., Hussain, M.J.M., Sangeetha, S.T.: Smart transportation systems: IoT-connected wireless sensor networks for traffic congestion management. Int. J. Adv. Sig. Image Sci. **9**(1), 40–49 (2023)
19. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint: arXiv:1710.10903 (2017)
20. Wu, T., et al.: Multi-agent deep reinforcement learning for urban traffic light control in vehicular networks. IEEE Trans. Veh. Technol. **69**(8), 8243–8256 (2020)
21. Zhang, J., Wang, F.Y., Wang, K., Lin, W.H., Xu, X., Chen, C.: Data-driven intelligent transportation systems: a survey. IEEE Trans. Intell. Transp. Syst. **12**(4), 1624–1639 (2011)
22. Zhang, W., et al.: Irregular traffic time series forecasting based on asynchronous spatio-temporal graph convolutional networks. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 4302–4313 (2024)
23. Zheng, C., et al.: Spatio-temporal joint graph convolutional networks for traffic forecasting. IEEE Trans. Knowl. Data Eng. **36**(1), 372–385 (2023)