A Multi-level Hybrid MAS–DRL Model for Distributed Dynamic Control

ABSTRACT

This study presents a hierarchical control architecture integrating a Multi-Agent System (MAS) and Deep Reinforcement Learning (DRL) to address control problems in distributed dynamic systems with large-scale and time-varying state spaces. The proposed model is organized according to a micro–meso–macro structure, enabling a clear separation between state aggregation and decision-making processes. Such a design reduces computational complexity while improving system adaptability and operational efficiency.

The architecture is evaluated through a traffic signal control application at Dong Da intersection in Quy Nhon using the SUMO simulation platform with peak-hour survey data. Simulation results demonstrate that the proposed approach achieves improved performance compared to the conventional fixed-time control method under identical geometric and traffic demand conditions. These findings indicate the potential applicability of the proposed model to distributed intelligent control systems.

Keywords: *Distributed dynamic control, Multi-agent systems, Deep reinforcement learning.*

1. INTRODUCTION

The rapid growth of motorized vehicles in Vietnamese urban areas has placed significant pressure on the existing transportation infrastructure. Traffic congestion frequently occurs at central intersections, particularly during peak hours, leading to increased waiting times, higher fuel consumption, and greater environmental emissions. In this context, optimizing traffic signal control is considered a feasible solution to enhance operational efficiency without requiring modifications to the physical infrastructure.

Traditional control approaches, such as fixed-time control and rule-based adaptive strategies, exhibit limitations in responding effectively to real-time traffic fluctuations. In recent years, Deep Reinforcement Learning (DRL) has been extensively investigated for traffic signal control problems due to its capability to learn optimal control policies through interaction with the environment. However, many existing models adopt a single-agent structure in which the entire intersection state is directly fed into a deep neural network, potentially leading to state noise and imbalanced prioritization among traffic movements.

To address these limitations, a hybrid model integrating a Multi-Agent System (MAS) and DRL is proposed based on a hierarchical micro–meso–macro architecture. This structure enables state aggregation at the directional level before being processed by a macro-level decision agent, thereby improving adaptability and control stability under dynamic traffic conditions.

The proposed model is implemented and evaluated on the Simulation of Urban MObility (SUMO) microscopic simulation platform using real survey data collected at Dong Da intersection in Quy Nhon. Model performance is compared with a conventional fixed-time control strategy using quantitative metrics, including average waiting time, queue length, total delay, and intersection throughput.

Motivated by the limitations of traditional traffic signal control methods and single-agent reinforcement learning models, this study pursues the following objectives:

- To propose a hybrid control architecture integrating a Multi-Agent System (MAS) and Deep Reinforcement Learning (DRL) based on a hierarchical structure, aimed at addressing control problems in distributed dynamic systems with large-scale and time-varying state spaces.

- To develop a generalized modeling framework applicable to various adaptive control systems, in which traffic signal control is employed as an illustrative case study.

- To implement simulation and conduct experimental evaluation at a representative urban intersection in order to validate the feasibility and operational effectiveness of the proposed architecture.

The main contributions of this study are summarized as follows:

- A hierarchical MAS–DRL architecture organized according to a micro–meso–macro structure is proposed, enabling a clear separation between state aggregation and policy learning processes. This design reduces local state noise and improves learning stability.

- A generalized mathematical model describing the coordination mechanism among agents is developed, clarifying the role of the intermediate layer in reducing state-space complexity.

- An experimental evaluation based on real-world survey data collected from an urban intersection is conducted, providing quantitative evidence of performance improvement compared with the conventional fixed-time control strategy.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed model. Section 4 describes the application of the model to the traffic signal control problem. Section 5 analyzes the experimental results. Finally, Section 6 concludes the paper and outlines directions for future research.

2. RELATED WORK

Traditional traffic signal control strategies, particularly fixed-time approaches derived from delay-based analytical formulations [1], have long been applied in urban traffic systems. While these methods are simple and operationally practical, they lack adaptability under dynamic and time-varying traffic demand. Empirical traffic flow studies have shown that congestion patterns exhibit nonlinear and unstable characteristics, especially under high-density conditions [2], highlighting the limitations of static timing schemes.

Reinforcement Learning (RL) was introduced as a model-free learning paradigm capable of optimizing sequential decision-making problems [3]. Early applications of RL in traffic signal control demonstrated the feasibility of learning adaptive signal policies at isolated intersections [4], [5]. However, classical RL methods often struggle with high-dimensional state representations and slow convergence in complex traffic environments.

The emergence of Deep Reinforcement Learning (DRL), particularly value-based and policy-gradient approaches, significantly enhanced representation capacity and scalability. The Deep Q-Network (DQN) framework [6] and its variants such as Dueling DQN [7] improved stability in value estimation, while Actor–Critic methods [8] and Proximal Policy Optimization (PPO) [9] provided more stable policy updates for control tasks. Building upon these advances, DRL-based traffic signal control models such as IntelliLight [10] reported substantial improvements in delay reduction and adaptive phase selection.

To address coordination challenges in multi-intersection networks, Multi-Agent Reinforcement Learning (MARL) frameworks have been proposed. Approaches such as CoLight [11] and PressLight [12] introduced cooperative learning and pressure-based optimization mechanisms. Hierarchical MARL strategies further enhanced scalability by decomposing large-scale control tasks into structured decision levels [13]. More recent studies have explored improved state representation, reward shaping, benchmarking, and graph-based coordination mechanisms within DRL and MARL frameworks [14]–[18], reinforcing their applicability to large-scale intelligent transportation systems.

In parallel, Multi-Agent Systems (MAS) have been widely adopted in traffic and transportation modeling due to their ability to represent heterogeneous agents and decentralized interactions [19]. Integrated MARL-based network control approaches have demonstrated the potential of combining multi-agent coordination with adaptive learning [20]. Nevertheless, most existing studies focus primarily on algorithmic improvements and do not fully integrate structured hierarchical modeling with adaptive deep learning in a unified framework.

The reliability of traffic signal control evaluation also depends on realistic microscopic traffic simulation. The SUMO platform has been extensively used for modeling vehicle dynamics and traffic interactions. Recent works have focused on calibrating car-following and lane-changing models using real-world trajectory and radar data [21]–[24], as well as evaluating environmental and operational impacts under different traffic scenarios [25]. These efforts enhance the credibility of simulation-based validation for traffic control algorithms.

Despite substantial progress in RL, DRL, MARL, and traffic microsimulation, a research gap remains in developing a hybrid framework that integrates MAS-based structural modeling with DRL-based adaptive optimization under a structured micro–meso–macro hierarchy and validates the approach within a realistic simulation environment.

To address this gap, this study proposes a multi-level MAS–DRL framework implemented in the Simulation of Urban MObility (SUMO) environment and evaluated using real-world traffic data collected at Dong Da intersection in Quy Nhon.

3. THE PROPOSED MAS–DRL MODEL

The proposed MAS–DRL model integrates a Multi-Agent System (MAS) with Deep Reinforcement Learning (DRL) to address control challenges in distributed dynamic systems with high-dimensional and time-varying state spaces. Its hierarchical architecture is designed to mitigate computational complexity and improve both learning stability and system scalability.

3.1. Problem Formulation

Consider a distributed dynamic system composed of a set of interacting components $I = \{1,2,\ldots,N\}$.

Each component $i \in I$ is characterized by a local state $x_i \in \mathbb{R}^{d_i}$.

At time step t, the global system state is defined as

$$s_t = \{x_i^t \mid i \in I\}, \qquad (1)$$

where x_i^t denotes the local state of component i at time step t.

A control policy $\pi(a_t \mid s_t)$ determines the probability of selecting action a_t given the current state s_t.

The objective is to learn a policy π that maximizes the expected cumulative discounted reward:

$$J(\pi) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \gamma^t r_t\right], \qquad (2)$$

where r_t is the immediate reward at time step t, $\gamma \in (0,1)$ is the discount factor, and $\mathbb{E}_\pi[\cdot]$ denotes the expectation over trajectories generated under policy π.

In large-scale distributed systems, the global state s_t can be high-dimensional and may contain locally correlated noise. Directly feeding the entire state representation into a deep learning model may lead to instability and slow convergence. Therefore, a structured hierarchical mechanism is introduced to decompose and aggregate state information before policy learning, thereby reducing computational complexity and improving learning stability.

3.2. Hierarchical MAS–DRL Architecture

The proposed control framework adopts a three-level hierarchical architecture consisting of micro, meso, and macro layers, as illustrated in Figure 1. The hierarchical decomposition aims to reduce state dimensionality and structure information flow before policy learning.

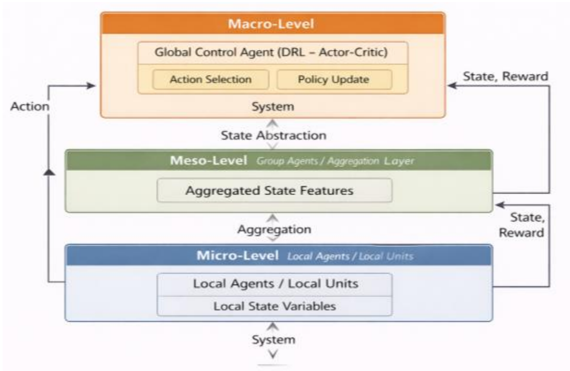

Figure 1. Hierarchical MAS–DRL control architecture.

The proposed hierarchical MAS–DRL framework is illustrated in Figure 1. The architecture consists of micro-level local agents responsible for data acquisition, a meso-level state aggregation mechanism that integrates local traffic features, and a macro-level DRL-based controller that generates global control actions.

Micro-Level

At the micro-level, the distributed system consists of a set of components

$I = \{1,2,\ldots,N\}$.

Each component $i \in I$ observes its local state

$x_i \in \mathbb{R}^{d_i}$.

These local states capture fine-grained system information but may contain local noise and high-dimensional features.

Meso-Level

To mitigate the curse of dimensionality, components are grouped into subsets

$G_j \subset I, j = 1, \dots, M,$
where $M \leq N$.

For each group G_j, a representative aggregated state is defined as

$$\tilde{x}_j = f_j(\{x_i \mid i \in G_j\}) \quad (4)$$

where $f_j(\cdot)$ denotes an aggregation function that transforms local states into a structured group-level representation.

Macro-Level

At the macro-level, the global control agent receives the aggregated system state

$$\tilde{s}_t = \{\tilde{x}_j^t \mid j = 1, \dots, M\} \quad (5)$$

and selects control action $a_t \sim \pi(a_t \mid \tilde{s}_t)$.

3.3. Learning Mechanism and Stability

At the macro level, the global control agent learns a parameterized policy $\pi_\theta(a \mid \tilde{s})$ based on the aggregated state $\tilde{s}_t$ defined in Section 3.2.

Optimization Objective

Consistent with the problem formulation in Section 3.1, the learning objective is to maximize the expected discounted return:

$$J(\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T} \gamma^t r_t\right] \quad (6)$$

where $\gamma \in (0,1)$is the discount factor and the expectation is taken over trajectories generated by policy π_θ.

Policy Update Mechanism

The policy parameters θ are updated using a policy-gradient approach.

The gradient of the objective function is estimated as:

$$\nabla_\theta J(\theta) = \mathbb{E}[\nabla_\theta \log \pi_\theta(a_t \mid \tilde{s}_t) A_t] \quad (7)$$

where A_t denotes the advantage function, which measures the relative benefit of action a_t at state $\tilde{s}_t$.

In practice, the advantage is estimated using temporal-difference (TD) learning:

$$A_t = r_t + \gamma V(\tilde{s}_{t+1}) - V(\tilde{s}_t) \quad (8)$$

with $V(\cdot)$ representing the state-value function.

Role of the Hierarchical Structure in Learning Stability

Unlike conventional DRL approaches that directly utilize the full system state s_t, the proposed model performs state aggregation at the meso level before policy learning. Instead of feeding high-dimensional and potentially noisy raw states into the learning algorithm, the macro-level agent operates on a structured aggregated representation $\tilde{s}_t$.

This hierarchical design contributes to learning stability in several aspects. First, the aggregation mechanism reduces the dimensionality of the input space, which simplifies the optimization process. Second, local fluctuations at the micro level are partially smoothed through group-level representation, thereby limiting abrupt variations in gradient updates. Third, the structured abstraction improves scalability when the number of system components increases.

Through this hierarchical information flow, the model maintains a balance between preserving essential local characteristics and optimizing a global control objective.

4. APPLICATION OF THE PROPOSED MODEL TO TRAFFIC SIGNAL CONTROL

This section presents the application of the multi-level MAS–DRL framework introduced in Section 3 to an urban traffic signal control problem. The case study is used to validate the feasibility of the proposed distributed dynamic control architecture in a real-world traffic context.

4.1. Problem Formulation

The study is conducted at Dong Da Intersection (Intersection No. 3) in Quy Nhon, a central urban junction that frequently experiences congestion during peak hours. Tran Hung Dao Street functions as the primary corridor with significantly higher traffic demand compared to the secondary approaches, leading to an imbalanced flow distribution across directions.

The intersection is modeled as a discrete-time dynamic system. At time step t, the system state is characterized by traffic indicators observed at each approach, including:

- Queue length;

- Number of waiting vehicles;

- Degree of saturation of each lane group.

The control action corresponds to selecting one of the feasible non-conflicting signal phases.

Consistent with the objective described in Section 3, the goal is to learn a control policy that minimizes congestion-related measures over time, thereby improving overall intersection performance under time-varying traffic demand.

4.2. Hierarchical MAS–DRL Implementation

The micro–meso–macro structure described in Section 3 is implemented for the traffic signal control problem as follows.

Micro Level

At the micro level, each approach (or lane group) is treated as a local component of the distributed system. Each component observes its own traffic state, such as queue length and vehicle density. These observations capture fine-grained traffic dynamics but may contain short-term fluctuations.

Meso Level

At the meso level, local states are aggregated according to signal phases. For each phase, traffic indicators of the approaches served in that phase are combined to form a phase-level state representation, including:

- Total queue length of the phase;

- Average density of the associated lanes.

This aggregation mechanism corresponds to the intermediate layer introduced in Section 3. It reduces the dimensionality of the input space and organizes information according to signal operation logic. Such structuring is particularly important at Dong Da Intersection due to the imbalance between major and minor traffic streams.

Macro Level

At the macro level, a centralized DRL-based agent receives the aggregated phase-level state and determines the signal phase to be activated. The action space consists of all feasible signal phases. Control decisions are executed at fixed observation intervals within the simulation environment.

Through this hierarchical information flow, the model preserves essential local traffic characteristics while enabling stable high-level decision-making, consistent with the stability considerations discussed in Section 3.

4.3. Reward Function and Learning Method

In alignment with the optimization objective defined in Section 3, the reward at time step t is defined as:

$$r_t = -(\alpha D_t + \beta Q_t)$$

where D_t denotes the total intersection delay and Q_t represents the total queue length. The negative form of the reward encourages the agent to reduce congestion and waiting time over time.

The macro-level agent is trained using an Actor–Critic framework based on the Proximal Policy Optimization (PPO) algorithm to ensure stable policy updates in the discrete control setting. The meso-level aggregation supports more stable learning by mitigating the impact of local noise.

4.4. Simulation Setup and Evaluation

The proposed model is implemented using the microscopic traffic simulation platform Simulation of Urban MObility (SUMO). The geometric layout of the intersection is constructed to reflect the actual structure of Dong Da Intersection. Traffic demand is configured according to observed afternoon peak-hour data.

The main simulation settings are as follows:

- Simulation duration: 7,200 seconds.

- Simulation step length: 1 second.

- Control decision interval: 5 seconds.

Two control scenarios are evaluated under identical geometric and traffic conditions:

1. Fixed-time signal control;
2. The proposed hierarchical MAS–DRL model.

Performance is assessed using average waiting time, average queue length, total intersection delay, and throughput. Each scenario is simulated multiple times with different random seeds to reduce stochastic effects.

This case study demonstrates the applicability of the proposed multi-level MAS–DRL framework to a practical urban traffic control problem and provides a basis for the performance analysis presented in the next section.

5. SIMULATION RESULTS AND DISCUSSION

Table 1 presents the average simulation results obtained from 10 independent runs for both the fixed-time control strategy and the proposed hierarchical MAS–DRL model under identical geometric and traffic demand conditions.

Table 1. Average simulation results

Indicator	Fixed-time	MAS–DRL
Avg. waiting time (s/veh)	79.6	69.8
Avg. queue length (veh)	36.2	31.8
Total delay (h)	128.4	112.7
Throughput (veh/2h)	10,954	11,236

Consistent with the hierarchical aggregation mechanism described in Section 3 and implemented for the Dong Da intersection in Section 4, the MAS–DRL model demonstrates measurable performance improvements over the conventional fixed-time strategy. Specifically, the average waiting time decreases from 79.6 s/vehicle to 69.8 s/vehicle, corresponding to a reduction of approximately 12–13%. Similarly, the total intersection delay decreases from 128.4 hours to 112.7 hours, representing an overall reduction of around 12%. The average queue length is reduced from 36.2 vehicles to 31.8 vehicles, equivalent to an improvement of approximately 10–12%. In addition, throughput increases from 10,954 vehicles during the 2-hour simulation period to 11,236 vehicles, corresponding to an increase of approximately 2–3%.

These improvements can be attributed to the hierarchical control structure. The meso-level aggregation of traffic states by signal phase enables the macro-level agent to allocate green time more flexibly to heavily loaded phases, particularly along the major corridor. As a result, the proposed model is more responsive to imbalanced traffic demand compared to the fixed-time plan, which operates with predetermined green splits.

The observed performance gains are more evident for approaches with high traffic volumes. For low-demand approaches, the difference between the two control strategies is relatively small. Furthermore, when traffic demand approaches saturation conditions, the improvement margin tends to decrease due to geometric capacity constraints of the intersection. This behavior is consistent with practical traffic flow theory, where control strategies have limited influence once infrastructure capacity becomes the dominant restriction.

It should be noted that the reported results correspond to the surveyed afternoon peak-hour condition. Performance may vary under different demand patterns or intersection configurations. The present findings therefore represent an initial validation of the proposed framework in a realistic urban setting. Further parameter tuning and extended testing under multiple traffic scenarios may enhance robustness and generalization capability.

Overall, the results confirm the feasibility of the proposed multi-level MAS–DRL framework for distributed dynamic traffic signal control and support its potential applicability in urban intersections with imbalanced traffic demand.

6. CONCLUSION AND FUTURE WORK

This study proposed a multi-level hybrid MAS–DRL framework for distributed dynamic control and applied it to an urban traffic signal control problem. Motivated by the limitations of centralized reinforcement learning approaches in handling large and imbalanced state spaces (as discussed in Section 1), the proposed architecture integrates a hierarchical micro–meso–macro structure to improve scalability and learning stability.

Section 3 introduced the conceptual design of the hierarchical MAS–DRL model, in which local observations are aggregated at an intermediate level before being processed by a macro-level decision agent. This design aims to reduce state dimensionality while preserving essential system characteristics. In Section 4, the framework was implemented for traffic signal control at Dong Da Intersection in Quy Nhon, where traffic demand is unevenly distributed among approaches.

Simulation results presented in Section 5 indicate that the proposed model achieves consistent improvements over the fixed-time control strategy under peak-hour conditions. Reductions of approximately 10–13% in waiting time, queue length, and total delay were observed, along with a modest increase in throughput. These results suggest that the hierarchical aggregation mechanism enables more flexible allocation of green time under imbalanced traffic demand, while maintaining stable learning performance.

Although the performance gains are moderate, they are obtained without modifying the geometric structure of the intersection. The findings therefore demonstrate the practical feasibility of the proposed framework in a realistic urban traffic context. At the same time, the results confirm that control-based improvements remain constrained by infrastructure capacity when traffic demand approaches saturation levels.

Future research may focus on several directions. First, further parameter tuning and testing under multiple traffic scenarios, including off-peak and stochastic demand variations, may enhance robustness and generalization capability. Second, extending the framework to multi-intersection networks would allow evaluation of coordination performance in larger-scale systems. Third, incorporating additional traffic indicators, such as travel time reliability or fuel consumption, may provide a more comprehensive assessment of control effectiveness.

Overall, this study provides an initial validation of a hierarchical MAS–DRL framework for distributed dynamic traffic signal control and contributes a practical case study for intelligent transportation applications in medium-scale urban environments.

REFERENCES

[1] F. V. Webster, *Traffic Signal Settings*, Road Research Technical Paper No. 39. London, U.K., 1958.

[2] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, 2000.

[3] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3–4, pp. 279–292, 1992.

[4] B. Abdulhai, R. Pringle, and G. Karakoulas, "Reinforcement learning for traffic signal control," *Transportation Research Part C*, vol. 11, no. 3–4, pp. 257–275, 2003.

[5] C. Thorpe and T. Regan, "Q-learning for traffic control," presented at the IEEE Intelligent Transportation Systems Conference, Seattle, USA, 2005, pp. 702–707.

[6] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, 2015.

[7] Z. Wang, T. Schaul, M. Hessel, and H. van Hasselt, "Dueling network architectures for deep reinforcement learning," presented at the 33rd International Conference on Machine Learning (ICML), New York, USA, 2016, pp. 1995–2003.

[8] V. R. Konda and J. N. Tsitsiklis, "Actor–Critic algorithms," presented at Advances in Neural Information Processing Systems (NeurIPS), Denver, USA, 2000, pp. 1008–1014.

[9] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," presented at the 34th International Conference on Machine Learning (ICML), Sydney, Australia, 2017, pp. 2823–2831.

[10] K. Lin *et al.*, "IntelliLight: A reinforcement learning approach for intelligent traffic light control," presented at the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 2018, pp. 2496–2505.

[11] T. Chu, J. Wang, L. Codeca, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," presented at NeurIPS Workshop on AI for Social Good, Vancouver, Canada, 2019.

[12] H. Wei, G. Zheng, V. Gayah, and Z. Li, "PressLight: Learning max pressure control to coordinate traffic signals in arterial network," presented at the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, USA, 2019, pp. 1290–1298.

[13] X. Ma *et al.*, "Hierarchical multi-agent reinforcement learning for large-scale traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3314–3327, 2020.

[14] S. Bouktif, A. Fiaz, and A. Ouni, "Deep reinforcement learning for traffic signal control with improved state and reward design," *IEEE Access*, vol. 11, pp. 45678–45690, 2023.

[15] M. Skuba, "Benchmarking deep reinforcement learning algorithms for traffic signal control," *Transportation Research Record*, vol. 2677, no. 5, pp. 112–125, 2023.

[16] S. Yang *et al.*, "HG-M2I: Hierarchical graph multi-agent reinforcement learning for traffic signal control," *Transportation Research Part C*, vol. 152, Art. no. 104143, 2023.

[17] Y. Bie *et al.*, "Collaborative V2I traffic signal optimization using multi-agent reinforcement learning," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 1, pp. 210–223, 2024.

[18] Y. Zhao, L. Li, and X. Wang, "A survey on deep reinforcement learning for traffic signal control,"

Transportation Research Part C, vol. 158, Art. no. 104421, 2024.

[19] A. L. C. Bazzan, "Opportunities for multiagent systems and multiagent reinforcement learning in traffic control," *Autonomous Agents and Multi-Agent Systems*, vol. 18, no. 3, pp. 342–375, 2009.

[20] S. El-Tantawy, B. Abdulhai, and H. Abdelgawad, "Integrated network traffic control using multi-agent reinforcement learning," *Transportation Research Part C*, vol. 35, pp. 167–186, 2013.

[21] M. Schrader, M. Al Abdraboh, and J. Bittle, "Comparing measured driver behavior distributions to results from car-following models using SUMO and real-world vehicle trajectories from radar," *SUMO User Conference Proceedings*, vol. 4, pp. 41–54, 2023.

[22] M. Schrader, A. Karnik, A. Hainen, and J. Bittle, "Calibrating car-following models using SUMO-in-the-loop and vehicle trajectories from roadside radar," *SUMO User Conference Proceedings*, Munich, Germany, 2024.

[23] Z. Chen *et al.*, "Investigating autonomous vehicle driving strategies in microsimulation using LC2013 lane-change model in SUMO," *Mathematics*, vol. 12, no. 3, Art. no. 455, 2024.

[24] M. Yavuz and H. Özen, "Calibration of microscopic traffic simulation of urban road using Krauss car-following model and LC2013 lane-change model in SUMO," *SJSU Transportation Science & Technology*, vol. 122, pp. 1–15, 2024.

[25] M. Sadat, A. Rahman, and K. Islam, "Environmental impact of slow moving electric vehicles using SUMO car-following Krauss and SL2015 lane-changing model," *Sustainability*, vol. 16, no. 4, Art. no. 1890, 2024.

Mô hình lai đa tầng MAS-DRL
cho bài toán điều khiển động phân tán

TÓM TẮT

Bài báo đề xuất một kiến trúc điều khiển phân cấp kết hợp giữa Hệ thống Đa Tác Tử (MAS) và Học Tăng cường Sâu (DRL) nhằm giải quyết bài toán điều khiển trong các hệ thống động phân tán có không gian trạng thái lớn và biến động theo thời gian. Mô hình được tổ chức theo cấu trúc micro–meso–macro, cho phép tách biệt quá trình tổng hợp trạng thái và ra quyết định, qua đó góp phần giảm độ phức tạp tính toán và nâng cao khả năng thích ứng của hệ thống.

Nghiên cứu triển khai mô hình đề xuất trên ứng dụng điều khiển tín hiệu giao thông tại nút Đống Đa (Quy Nhơn) trên nền tảng mô phỏng SUMO với dữ liệu khảo sát giờ cao điểm. Kết quả mô phỏng bước đầu cho thấy mô hình đạt mức cải thiện so với phương pháp điều khiển chu kỳ cố định trong cùng điều kiện hình học và lưu lượng.

Từ khóa: *Điều khiển động phân tán, Hệ thống đa tác tử, Học tăng cường sâu*