Design of Intersection Control Algorithm Based on Deep Reinforcement Learning

Yi Zheng[1], Xiaohu Yang[1], Anping Wang[2], Fei Li[3], Chen Li[4*], Yaojun Gui[4]

[1]Yunnan Communications Investment Group Investment Co., Ltd., Kunming 650228, China
[2]Yunnan Xuanhui Expressway Co., Ltd, Xunwei 655400, China
[3]Yunnan Communications Investment Group Public Construction & Bridge Engineering Co., Ltd., Kunming 650228, China
[4]YCIC Broadvision Engineering Cousultants, Kunming 650200, China

*ynkm_leec@foxmail.com

Abstract. Traditional traffic signal control (TSC) schemes rely on fixed timing or rule-based adaptation, which lack real-time responsiveness to stochastic and highly dynamic traffic flows. Deep reinforcement learning (DRL) has become a promising paradigm for adaptive signal control, yet value-based methods such as DQN and DDQN suffer from overestimation bias, unstable training, and low sampling efficiency in complex urban intersection environments. This paper proposes a PPO-GAE control framework, which combines Proximal Policy Optimization (PPO) with Generalized Advantage Estimation (GAE) to achieve stable, efficient, and robust traffic signal timing optimization. PPO constrains policy updates within a trust region to avoid destructive updates and ensure training stability, while GAE effectively reduces gradient variance and improves the accuracy of advantage estimation for continuous traffic states. Experiments are carried out on the SUMO simulation platform using a real intersection in Xi'an. Results show that the proposed PPO-GAE method reduces cumulative delay by 41.2% and average queue length by 33.7% compared with the fixed-time baseline, and outperforms DQN, DDQN-PER, and traditional adaptive methods.

Keywords: Intelligent Traffic Signal Control; Deep Reinforcement Learning; Proximal Policy Optimization; Generalized Advantage Estimation; SUMO; Urban Traffic Efficiency

1 Introduction

The rapid growth of urban vehicle ownership has caused severe traffic congestion, leading to massive economic losses, increased travel time, and environmental pollution. Traditional TSC methods, such as fixed-timing and Webster formula-based control, fail to adapt to real-time fluctuations in traffic demand. With the development of intelligent transportation systems (ITS), data-driven adaptive signal control has become an important research direction.

L. Trajkovic et al. (eds.), *Proceedings of the 2026 2nd International Conference on Data Mining and Project Management (DMPM 2026)*, Advances in Economics, Business and Management Research 390,
https://doi.org/10.2991/978-94-6239-689-0_27

In recent years, deep reinforcement learning (DRL) has been widely applied to traffic signal control because it can learn optimal decision policies through continuous interaction with the environment. Value-based DRL algorithms such as DQN and DDQN have achieved certain effects, but they still have obvious limitations[1-3].

- DQN is prone to **Q-value overestimation**, leading to suboptimal policies.
- DDQN alleviates overestimation but still has low sample efficiency.
- Experience replay sampling is not targeted enough for rare but critical congestion scenarios.
- Training instability increases under high-dimensional and high-uncertainty traffic states.

To overcome these problems, this paper adopts a **policy-gradient-based PPO algorithm** combined with GAE for urban intersection signal control. The contributions are as follows:

- A stable and efficient PPO-GAE signal control framework is designed, suitable for real-time decision-making in complex traffic environments.
- The state space is refined using lane-based vehicle position and speed discretization to support fine-grained control.
- A reward function focusing on cumulative waiting time reduction is constructed to directly optimize intersection traffic efficiency.
- Sufficient experiments on SUMO show that PPO-GAE significantly outperforms fixed-time, DQN, and DDQN-PER methods.

2 Related Work

Early reinforcement learning for TSC mainly uses tabular Q-learning, which is limited by the “curse of dimensionality” and cannot handle high-dimensional state representations. With the rise of deep learning, DQN uses neural networks as function approximators and achieves better perception and decision-making capabilities:

However, DQN uses the max operator to select and evaluate actions simultaneously, which causes **overestimation bias**. DDQN separates action selection from value evaluation to reduce overestimation. PER further improves sample efficiency by prioritizing high-error transitions. Nevertheless, value-based methods are still difficult to maintain stable convergence in highly dynamic traffic scenes.

Policy-gradient algorithms directly optimize the policy function and show better stability in continuous or high-dimensional action spaces. PPO uses a clipped surrogate objective to restrict the update range of the policy, making training more robust and suitable for real-world engineering applications. GAE further improves advantage estimation by weighting multi-step rewards, which reduces variance and accelerates convergence[4-9].

3 Methodology

3.1 Markov Decision Process Formulation

The intersection signal control problem is modeled as a Markov Decision Process (MDP): (S,A,R,γ):

A. State Space *(S)*: The state is composed of discretized lane states. Each lane is divided into uniform grids. The state vector includes:

- Vehicle occupancy matrix of each lane
- Normalized average speed of vehicles in each section
- Current signal phase and remaining time

B. Action Space (A)**:** Four discrete signal phases are defined:

- North-South straight
- North-South left turn
- East-West straight
- East-West left turn

Each phase switch includes a 3-second yellow light and a 2-second all-red interval to ensure driving safety.

C. Reward Function (*R*): The reward is designed to minimize waiting time and queue length:

$$R_t = -(\lambda_1 \cdot \Delta Wait_t + \lambda_2 \cdot \Delta Queue_t) \tag{1}$$

Where $\Delta Wait_t$ is the change in cumulative waiting time, $\Delta Queue_t$ is the change in average queue length, and λ_1, λ_2 are weighting coefficients.

3.2 PPO-GAE Algorithm

Proximal Policy Optimization (PPO):

PPO optimizes the clipped surrogate objective to ensure stable policy updates:

$$L^{CLIP}(\theta) = E_t\left[min\left(r_t(\theta)\widehat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\widehat{A}_t\right)\right] \tag{2}$$

where $r_t(\theta)$ is the probability ratio between the new and old policy, and ϵ is the clip threshold (usually 0.2).

- **Generalized Advantage Estimation (GAE):** GAE integrates multi-step bootstrapping to compute the advantage function with low variance:

$$\widehat{A}_t^{GAE} = \sum_{l=0}^{\infty} (\gamma\lambda)^l \delta_{t+l} \tag{3}$$

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \tag{4}$$

GAE balances bias and variance and improves training efficiency.

- **Network Structure:**
- Actor network: outputs the probability distribution of four signal phases;
- Critic network: estimates state value;
- Activation: ReLU;
- Optimizer: Adam.

4 Experiments and Analysis

To verify the effectiveness and advancement of the proposed PPO-GAE algorithm for urban traffic signal control, a large number of comparative simulation experiments are carried out on the SUMO platform. The experimental environment is built according to the actual geometric parameters, lane distribution, signal phase sequence and traffic flow characteristics of the intersection of Zhuque Street and Xiaowei West Road in Xi'an, China. The proposed PPO-GAE method is compared with the traditional fixed-time control scheme, the standard DQN algorithm, and the DDQN-PER algorithm in the original paper. The performance differences under different traffic demand levels, training stability, convergence speed and response ability to sudden traffic bursts are fully analyzed.

4.1 Simulation Environment

The simulation environment is shown as Table 1.

Table 1. Simulation Environment.

Type	Description
Platform	SUMO 1.8.0
Intersection	Zhuque Street × Xiaowei West Road, Xi'an
Episode duration	3600s
Warm-up time	300s
Traffic flow	200–800 veh/h/lane
Training episodes	2000

4.2 Evaluation Metrics

The performance of the traffic signal control algorithm is mainly evaluated from two core indicators closely related to traffic operation efficiency:

- **Cumulative Delay:** It refers to the total waiting time of all vehicles in the intersection area during the statistical period, which directly reflects the congestion degree and travel efficiency of the intersection.
- **Average Queue Length:** It refers to the average number of vehicles waiting in each lane during the simulation process, which is used to measure the risk of queue spillback and the service level of the intersection.

In addition, the convergence speed of the model, the stability of the reward curve and the adaptability to sudden traffic flow changes are also used as important auxiliary evaluation indicators.

4.3 Performance Comparison and Result Analysis

Table 2 shows performance improvements compared with the fixed-time baseline.

Table 2. Performance Comparison with Fixed-Time Control.

Method	Cumulative Delay Reduction	Average Queue Length Reduction
DQN	24.3%	19.8%
DDQN-PER	32.8%	26.4%
PPO-GAE (Proposed)	**41.2%**	**33.7%**

Under saturated traffic conditions:

- PPO-GAE reduces delay by approximately 41.2%, effectively alleviating long-time congestion.
- Average queue length is reduced by 33.7%, preventing queue spillback and improving intersection capacity.
- The algorithm converges stably at about 600 episodes, faster than DQN and DDQN-PER.

4.4 Discussion

From the comprehensive experimental results, the PPO-GAE algorithm proposed in this paper has obvious advantages over the fixed-time control, DQN and DDQN-PER algorithms in cumulative delay, average queue length, convergence speed and stability. The fixed-time control relies on the experience design of signal timing, which is lack of adaptability and can not cope with the stochastic traffic flow. The value-based algorithms such as DQN and DDQN-PER have problems such as overestimation bias and unstable training, which limit the further improvement of control performance.

The PPO-GAE algorithm combines the stable optimization of PPO and the high-precision advantage estimation of GAE, and at the same time uses the fine-grained state representation based on lane discretization, so that the model can accurately perceive the traffic state and make stable and efficient decisions. The algorithm performs well in both conventional traffic scenarios and high-saturation and sudden burst scenarios, which proves its high robustness and engineering application potential.

5 Conclusion and Future Work

This paper presents an adaptive traffic signal control algorithm based on Proximal Policy Optimization with Generalized Advantage Estimation (PPO-GAE) for urban isolated intersections. The proposed method addresses the limitations of traditional

fixed-time control and conventional value-based deep reinforcement learning algorithms, including overestimation bias, unstable training, and low efficiency. By formulating the signal timing problem as a Markov Decision Process and adopting a clipped surrogate objective and low-variance advantage estimation, the PPO-GAE model achieves stable and efficient policy optimization. Simulation results on the SUMO platform using a real intersection in Xi'an demonstrate that our method reduces cumulative delay by 41.2% and average queue length by 33.7% compared with the fixed-time baseline, outperforming DQN and DDQN-PER in both control performance and training stability. The fine-grained state representation and fast convergence further enhance its practical potential for real-world deployment.

For future work, we will extend the single-agent framework to multi-agent PPO for coordinated control of adjacent intersections. We also plan to integrate V2I communication and multi-source sensing data to enrich the state representation. Furthermore, field tests on edge computing devices will be conducted to validate real-world performance. Additional objectives such as emission reduction and pedestrian priority will be incorporated into a multi-objective reward design. Finally, we will explore digital twin and mixed traffic flow scenarios to build a more intelligent and self-optimizing urban traffic control system.

Acknowledgments

This study was funded by Yunnan Provincial Department of Transportation Science and Technology Innovation and Demonstration Project: "Research and Demonstration of Precise Monitoring, Early Warning and Operational Risk Prevention Technology for Ice and Snow Disasters on Expressways" (Yunnan Transportation Science and Education Department Document No. [2023]-152) and Yunnan Provincial Broadvision Engineering Consultants Co., Ltd. Independent Technology Project: "Key Technologies and Application Demonstration for the Construction of a Smart Platform in the Field of Highway Engineering Survey and Design" (Project No.: YJSJ-ZL-2023-02).

References

1. Sutton, R.S., Barto, A.G.: Introduction to reinforcement learning. 2nd edn. MIT Press, Cambridge (2018)
2. Abdoos, M., Mozayani, N., Bazzan, A.L.C.: Holonic multi-agent system for traffic signals control. Engineering Applications of Artificial Intelligence 26(5-6), 1575–1587 (2013)
3. Zhang, R., Schmutz, F., Gerard, K.: Increasing traffic flows with DSRC technology: Field trials and performance evaluation. IEEE Spectrum 55(10), 24–29 (2018)
4. Tonguz, O.K.: Red light, green light — no light: Tomorrow's communicative cars could take turns at intersections. IEEE Spectrum 55(10), 24–29 (2018)
5. Li, L., Wen, D., Yao, D.: A survey of traffic control with vehicular communications. IEEE Transactions on Intelligent Transportation Systems 15(1), 425–432 (2014)
6. Shen, J., Wang, Y., Wang, H., et al.: Optimizing urban intersection management in mixed traffic using deep reinforcement learning and genetic algorithms. IEEE Access (2025)

7. Wang, A., Zhang, K., Shao, J., et al.: Deep Reinforcement Learning-based Signal Control for Traffic Risk Reduction and Efficiency Improvement at Urban Large Intersections. IEEE Internet of Things Journal (2025)
8. Saadi, A., Abghour, N., Chiba, Z., et al.: A survey of reinforcement and deep reinforcement learning for coordination in intelligent traffic light control. Journal of Big Data 12(1), 84 (2025)
9. Wang, L., Zhang, W., Yan, Z.: Vehicle–Infrastructure Cooperation Framework for Vehicle Navigation and Traffic Signal Control using Deep Reinforcement Learning. Transportation Research Record 2680(1), 568–583 (2026)