A DRL Approach for Optimizing the Vehicles Motorway Entry in Congested Traffic Scenarios

Antonio Salcuni
Politecnico di Bari
Bari, Italy
antonio.salcuni@poliba.it

Gaetano Volpe
Politecnico di Bari
Bari, Italy
gaetano.volpe@poliba.it

Agostino Marcello Mangini
Politecnico di Bari
Bari, Italy
agostinomarcello.mangini@poliba.it

Maria Pia Fanti
Politecnico di Bari
Bari, Italy
mariapia.fanti@poliba.it

***Abstract*—This work addresses the problem of vehicle merging at highway ramps and intersections, formulated as an optimization task to minimize sudden braking, waiting times, and collision risks. The problem is modeled as a Markov Decision Process to capture traffic dynamics. A Deep Reinforcement Learning approach based on the Actor-Critic framework is proposed, leveraging state-of-the-art techniques to improve decision-making efficiency. The reward function is designed to dynamically balance safety and efficiency by reducing braking events and congestion. Simulations focus on a high-density merging scenario located in northern Italy, between the "A7 Giovi" Highway and Milan's Western Ring Road. Results demonstrate significant improvements in traffic flow and safety with optimized merging policies outperforming traditional methods.**

I. Introduction

The evolution of autonomous driving technologies has significantly transformed the transportation landscape, offering promising solutions to enhance road safety, traffic efficiency, and reduce environmental impacts. In this context, autonomous driving systems typically adopt two main approaches: the *decomposed scheme*, which divides the decision-making process into distinct modules such as perception, prediction, planning, and control; and the *end-to-end scheme*, which leverages deep neural networks combined with Reinforcement Learning (RL) to directly learn decision policies from observed data [1]. While the decomposed approach benefits from modular design and well-established methods, its complexity and inability to handle all possible scenarios limit its scalability. The end-to-end approach, on the other hand, simplifies the overall process and has shown great potential in adapting to complex environments. Among these advanced techniques, *Deep Reinforcement Learning* (DRL) has emerged as a powerful framework for addressing dynamic decision-making in uncertain and complex settings. DRL, based on *Markov Decision Processes* (MDP), enables agents to iteratively interact with the environment, balancing exploration and exploitation to maximize cumulative rewards [2]. This framework has demonstrated exceptional performance in various domains, such as dynamic resource allocation and traffic management [3].

Moreover, DRL has also been applied to autonomous navigation in complex environments. Traditional navigation methods often struggle with uncertainties, whereas DRL enables systems to learn navigation policies through trial and error, leveraging sensor data and kinematic information to handle challenging scenarios [4]. These approaches have proven effective in domains such as service robots, unmanned aerial vehicles (UAVs), and underwater navigation systems.

In addition to DRL, Particle Swarm Optimization (PSO) has also been explored as a complementary method, especially for traffic signal optimization. PSO dynamically adjusts traffic light timings to suit real-time conditions, outperforming traditional fixed-timing controllers and other evolutionary algorithms, such as Genetic Algorithms (GA) and Ant Colony Optimization [5]. This method highlights the potential of swarm intelligence in solving complex traffic control problems.

The development of innovative control algorithms, such as Sample-Observed Soft Actor-Critic (SOSAC), has shown significant improvements in adaptability, cumulative rewards, and control precision in challenging underwater environments [7].

Simulation platforms such as Simulation of Urban Mobility (SUMO) play a crucial role in testing and validating these solutions. Specifically, SUMO provides realistic traffic environments that support the development of AI-based traffic management frameworks by generating high-fidelity traffic scenarios [8]. These simulations enable the exploration of advanced methods for optimizing traffic flow and reducing congestion.

Efficient traffic management in high-density urban areas is a critical challenge, specifically in highway entry ramps and merging points, where vehicles must integrate smoothly into high-speed traffic flows. The problem of merging optimization involves balancing traffic throughput, minimizing sudden braking events, and reducing the risk of collisions. Traditional traffic control systems struggle to adapt to real-time conditions, necessitating advanced learning-based solutions.

Recent research has explored the integration of Artificial Intelligence (AI) in traffic management systems, with the aim of replacing human dispatchers with AI-driven neural networks [11]. In particular, the work of [9] presents a neural network-based traffic dispatching system, implemented using AnyLogic and SUMO simulations.

The rapid increase in vehicular traffic in urban and peri-

This study is part of the MOST - Sustainable Mobility National Research Center, receiving funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 D.D. 1033 17/06/2022, CN0000 0023) and part of the IN2CCAM project that has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101076791.

urban areas has made the management of highway entries and critical intersections increasingly complex. Consequently, traditional rule-based and model-based optimization methods struggle to adapt to highly dynamic and uncertain environments. In response to these challenges, AI-based techniques, such as Deep Reinforcement Learning (DRL), have shown promising results in adaptive traffic management.

This paper deals with the highway merging problem by proposing a DRL approach based on the Actor-Critic framework. The study focuses on optimizing vehicle entry at critical congestion points, particularly at highway on-ramps, where sudden braking and inefficient merging strategies significantly impact traffic flow and safety. Unlike traditional optimization techniques, our method leverages DRL to dynamically adapt control policies based on real-time traffic conditions, improving overall coordination between vehicles and infrastructure.

The remaining part of the paper is organized as follows. Section II presents the problem formulation and the basic definitions of the DRL approach. In addition, Section III describes the DRL strategy applied to the considered problem and Section IV discusses the case study. Finally, Section V draws the conclusions.

II. Problem Formulation

The merging problem involves coordinating vehicle interactions in road segments where multiple traffic streams converge, such as highway ramps and intersections. Effective management strategies are needed to minimize congestion, reduce sudden braking, and enhance traffic flow. Traffic control mechanisms integrate predefined routes with intelligent signals to regulate merging operations.

Traffic routes define different vehicle interactions within the merging zone: vehicles entering the main traffic stream ($\mathcal{R}_{\text{in}}$), those changing direction and requiring lane transitions ($\mathcal{R}_{\text{turn}}$), and those traveling without lane changes ($\mathcal{R}_{\text{main}}$), which serves as a reference for evaluating interactions between merging and non-merging vehicles.

Intelligent traffic signals operate through multiple phases to ensure smooth merging. The $\mathcal{G}$ (Green) phase allows movement, while $\mathcal{R}$ (Red) forces vehicles to stop. Intermediate phases such as $\mathcal{GR}$ and $\mathcal{RG}$ regulate merging by allowing traffic from one route while restricting the other.

The system coordinates vehicle entries and merging maneuvers: $\mathcal{S}_{\text{merge}}$ manages merging from a secondary road, alternating between $\mathcal{GR}$ and $\mathcal{RG}$ to prevent conflicts. $\mathcal{S}_{\text{entry}}$ controls vehicle entry at a regulated access point, alternating between $\mathcal{G}$ and $\mathcal{R}$ without stopping the main flow.

By integrating these mechanisms, traffic management dynamically adapts to varying conditions, optimizing throughput while maintaining safety in high-density road networks.

A. Markov Decision Process

The merging problem is modeled as an MDP, defined by the tuple:

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$$

with

- $\mathcal{S}$ (State space): Represents traffic conditions;
- $\mathcal{A}$ (Action space): Represents the set of possible actions controlling traffic signal phases;
- $\mathcal{P}$ (Transition probabilities): Defines the probabilistic evolution of the system states given an action;
- $\mathcal{R}$ (Reward function): Balances safety and efficiency to optimize traffic flow;
- γ (Discount factor): Determines the weight of future rewards in decision-making.

In our context, the class is *model-free*, and therefore, it does not assume a probabilistic distribution $\mathcal{P}$.

B. State Space and Action Space

At time t, the state s_t is defined as:

$$s_t = \left\{ n_{\mathcal{R}_{\text{in}}}, n_{\mathcal{R}_{\text{turn}}}, n_{\mathcal{R}_{\text{main}}}, v_{\mathcal{R}_{\text{in}}}, v_{\mathcal{R}_{\text{turn}}}, v_{\mathcal{R}_{\text{main}}} \right\}$$

with

- $n_{\mathcal{R}_{\text{in}}}, n_{\mathcal{R}_{\text{turn}}}, n_{\mathcal{R}_{\text{main}}}$ denote the vehicle counts in each route,
- $v_{\mathcal{R}_{\text{in}}}, v_{\mathcal{R}_{\text{turn}}}, v_{\mathcal{R}_{\text{main}}}$ indicate the average speeds on the respective routes.

The action space is defined as a binary vector representing the state of the traffic signals:

$$S_t = \begin{bmatrix} S_{\text{merge}} \\ S_{\text{entry}} \end{bmatrix}$$

where:

- $S_{\text{merge}} \in \{0, 1\}$ controls the merging traffic light, with:

$$S_{\text{merge}} = \begin{cases} 0, & \text{if the signal is in phase } \mathcal{GR} \\ 1, & \text{if the signal is in phase } \mathcal{RG} \end{cases}$$

- $S_{\text{entry}} \in \{0, 1\}$ controls the entry traffic light, with:

$$S_{\text{entry}} = \begin{cases} 0, & \text{if the signal is in phase } \mathcal{R} \\ 1, & \text{if the signal is in phase } \mathcal{G} \end{cases}$$

Thus, the complete action space is:

$$\mathcal{S} = \{(0,0), (0,1), (1,0), (1,1)\}$$

where each action vector S_t defines the simultaneous state of both traffic lights at time t.

C. Reward Function

The reward function is designed to balance traffic efficiency and safety, incorporating weighted penalties for collisions, sudden decelerations, and waiting times:

$$R(s, a, t) = -w_c n_c(s, a, t) - w_d d(s, a, t) - w_t t(s, a, t) \quad (1)$$

where:

- $n_c(s, a, t)$: Number of collisions at time t.
- $d(s, a, t)$: Cumulative sudden deceleration.
- $t(s, a, t)$: Average waiting time.
- w_c: Weight associated with the severity of collisions.
- w_d: Weight penalizing sudden decelerations, capturing traffic instability.

- w_t: Weight penalizing waiting times, influencing traffic efficiency.

In addition, the cumulative sudden deceleration $d(s, a, t)$ is computed by summing the excessive deceleration events across all vehicles in the system. For each vehicle v, the deceleration is given by:

$$a_v(t) = \frac{v_v(t-1) - v_v(t)}{\Delta t} \tag{2}$$

where $v_v(t)$ is the velocity of vehicle v at time t. If the deceleration exceeds a predefined threshold a_{thr}, it contributes to the total penalty:

$$d(s, a, t) = \sum_{v \in V} \max(0, |a_v(t)| - |a_{\text{thr}}|). \tag{3}$$

This formulation ensures that only significant braking events influence the reward function, penalizing unstable traffic conditions and promoting smoother traffic flow.

III. Deep Reinforcement Learning Approach

The DRL approach adopts the Actor-Critic framework, which integrates value-based and policy-based methods to optimize vehicle merging.

Traditional Q-learning relies on a tabular Q-function representation, where each state-action pair is stored explicitly. However, as the state space grows, the number of entries scales exponentially as $\mathcal{O}(k^n)$, making this approach infeasible for traffic control scenarios with multiple variables. This necessitates function approximation techniques such as Deep Q-Learning (DQN).

While DQN addresses the limitations of tabular Q-learning using neural networks, it remains unsuitable for this problem due to:

- **Continuous action space:** DQN is designed for discrete actions, whereas traffic signal control often requires continuous adjustments;
- **Stability issues:** Traffic conditions evolve dynamically, making DQN harder to stabilize despite experience replay and target networks;
- **Inefficient policy optimization:** DQN indirectly learns a policy via value estimation, leading to suboptimal exploration and control in complex environments.

To overcome these issues, the Actor-Critic framework combines policy optimization with value estimation. The actor network maps states to actions, supporting continuous control, while the critic network stabilizes learning by estimating value functions. This synergy enhances adaptability and efficiency in dynamic traffic management.

A. Actor-Critic Framework

The DRL model utilizes two neural networks:

- Actor network: that maps states to actions using a probabilistic policy $\pi(a|s)$. It outputs the probability of taking each possible action in a given state;
- Critic network: that estimates the value of a state or the advantage of an action using a value function $V(s)$ or $Q(s, a)$.

The *Actor* and *Critic* work synergistically during the learning process. The Actor selects an action based on the current policy, while the Critic evaluates the chosen action by computing an estimate of the advantage or the error. The Critic uses this error to update its parameters and provides feedback to the Actor, which consequently updates its policy to favor actions that lead to higher rewards [10].

The update of the Critic parameters is based on the *Advantage Function* and the *TD-error* (Temporal Difference error). The TD-error measures how much the received reward deviates from the expected reward.

The advantage function $A(s, a)$ of a given action a in state s is defined as the difference between the action-value function and the state-value function:

$$A(s, a) = Q(s, a) - V(s) \tag{4}$$

where:

- $Q(s, a)$ is the action-value function, representing the expected return of taking action a in state s and then following the current policy;
- $V(s)$ is the state-value function, representing the expected return of state s under the current policy.

The advantage function helps determine how much better an action is compared to the average value of the current state. Additionally, the advantage function is closely related to the Temporal Difference (TD) error used in Actor-Critic algorithms to update the Critic and, consequently, improve the Actor's policy:

$$\delta = r + \gamma V(s') - V(s), \tag{5}$$

where r is the reward, γ is the discount factor, s is the current state, and s' is the next state. The architecture of the actor-critic framework is shown in Fig.1.

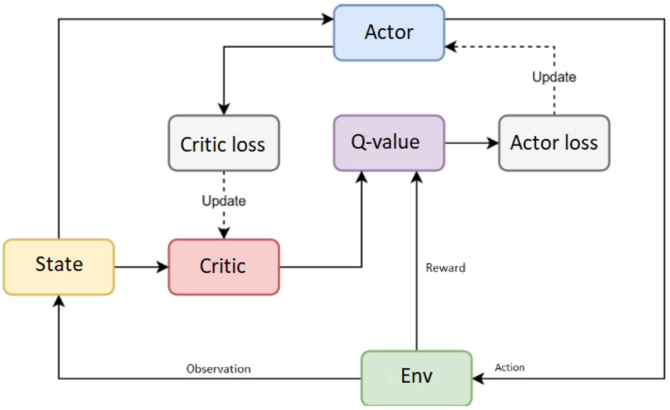


Fig. 1. Schematic of the actor-critic framework. The Actor selects actions based on the current policy, while the Critic evaluates them using the advantage function or Temporal Difference (TD) error. This feedback loop enables policy refinement and optimal decision-making in reinforcement learning.

IV. Case Study

To evaluate the proposed DRL-based traffic control approach, a case study was conducted at the intersection of the A7 Giovi Highway and Milan's Western Ring Road (Tangenziale Ovest), a critical and congested traffic node in northern Italy. This location was selected due to its high vehicle density, short acceleration lanes, and frequent congestion, making it an ideal testbed for assessing the system's adaptability and effectiveness.

Key challenges at this site include:

- **High Traffic Volume**: Thousands of vehicles, including passenger cars and heavy trucks, traverse this intersection daily, with peak-hour congestion amplifying merging difficulties;
- **Structural Constraints**: Short acceleration lanes and closely spaced exits force abrupt lane changes, increasing the risk of delays and conflicts;
- **Safety Concerns**: Frequent lane merging and high vehicle density contribute to elevated accident risks;
- **Real-World Applicability**: The issues at this site mirror challenges in highway intersections worldwide, enhancing the generalization of the findings.

A high-fidelity simulation was developed to replicate real-world traffic dynamics, incorporating:

- **Traffic Flow Data**: Real-world data ensured an accurate representation of typical conditions;
- **Vehicle Heterogeneity**: Various vehicle types, including cars, buses, and trucks, were modeled with distinct movement characteristics;
- **Dynamic Interactions**: The simulation captured merging maneuvers, lane changes, and braking responses;
- **Environmental Factors**: Weather conditions and visibility variations tested the control strategy's robustness.

The study assessed the impact of the DRL-based control system on key metrics such as collision frequency, sudden braking events, and average waiting time, with results analyzed in the following sections.

A. Simulation Environment

The simulation was implemented using SUMO tool. The network topology was extracted from OpenStreetMap (OSM) and converted into a SUMO-compatible format using `netconvert`. The merging scenario was modeled to reflect real-world conditions, including:

- High variability in vehicle speeds and densities, with traffic including both fast-moving cars and slower vehicles, resulting in dynamic interactions;
- Traffic lights at critical points to manage merging conflicts. Two traffic lights $\mathcal{S}_{\text{merge}}$, $\mathcal{S}_{\text{entry}}$ were introduced, alternating between green and red states to regulate traffic flow.

The three primary routes modeled in the simulation include:

- $\mathcal{R}_{\textbf{in}}$: Vehicles coming from the Tangenziale Ovest and merging onto the A7 Highway. This section is prone to sudden braking due to competition for merging space.
- $\mathcal{R}_{\textbf{turn}}$: Vehicles already on the A7 Highway performing a U-turn to switch direction. This involves crossing multiple lanes, increasing complexity.
- $\mathcal{R}_{\textbf{main}}$: Vehicles traveling along the A7 Highway without lane changes. This route is crucial for assessing interactions between merging vehicles and those in transit.

B. Implementation Details

The scenario rappresented in Fig. 2 was implemented with the following key steps:

1) Map extraction: The network layout of the area was exported from OpenStreetMap (OSM) and saved in the `.osm` format.
2) Network conversion: The `netconvert` tool was used to transform the OSM file into a SUMO-compatible `.net.xml` file, capturing road geometries, lanes, and traffic rules.
3) Route definition: Using `netedit`, three main routes $\mathcal{R}_{\textbf{in}}$, $\mathcal{R}_{\textbf{turn}}$, and $\mathcal{R}_{\textbf{main}}$ were defined to simulate different traffic patterns, focusing on critical merging and conflict points.
4) Traffic light configuration: Two traffic lights were introduced to regulate vehicle flows at key junctions, alternating states between green $\mathcal{G}$ and $\mathcal{R}$ based on predefined sequences.

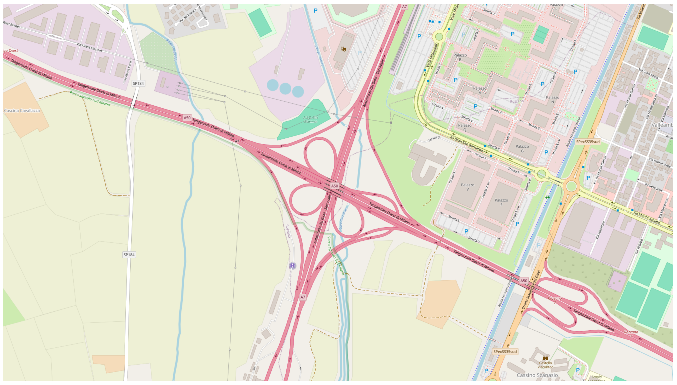


Fig. 2. Geometric layout of the intersection between the A7 Highway and Milan's Western Ring Road. The figure illustrates the merging zones and traffic flow interactions, which serve as the test environment for evaluating the proposed DRL-based traffic control strategy.

C. Performance Metrics

The effectiveness of the DRL-based control strategy was evaluated using the metrics summarized in Table I.

D. Training Phase

The agent was trained over 600 episodes, each consisting of 150 steps. These parameters were chosen to ensure adequate exploration of the environment and allow the agent to learn optimal decision-making strategies in a simulated traffic scenario. The reward function was carefully designed with the following weights to prioritize different objectives:

TABLE I
PERFORMANCE METRICS FOR EVALUATING THE DRL-BASED CONTROL STRATEGY

Metric	Description
Collision frequency	Number of collisions recorded during the simulation, focusing on minimizing incidents in high-conflict zones.
Sudden braking events	Total deceleration events exceeding 4.5 m/s^2, indicating potential risks and traffic inefficiencies.
Average waiting time	Average delay experienced by vehicles at the merging zone, calculated for all three routes.

- **Collisions**: $w_c = 100$ was assigned to collisions to reflect their critical severity;
- **Waiting times**: $w_t = 10$ for the highway ramp, 5 for route $\mathcal{R}_{\textbf{turn}}$, and 1 for route $\mathcal{R}_{\textbf{main}}$;
- **Sudden braking events**: $w_d = 1$ was assigned to penalize abrupt decelerations.

During the initial training phases, a temporary deterioration in performance was observed, as the agent performed sub-optimal exploratory actions to gather information about the environment.

However, as training progressed, the agent transitioned into the exploitation phase, leading to notable improvements in key metrics. This trend is reflected in the reward progression shown in Fig. 3.

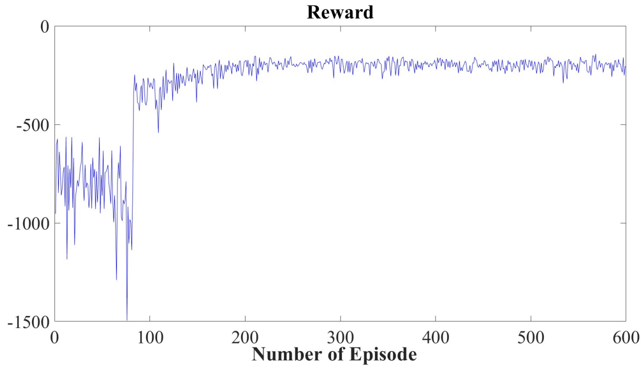


Fig. 3. Evolution of the cumulative reward during the training phase. The plot illustrates the learning progress of the DRL agent. A steady increase in reward values indicates improved decision-making and better adaptation to traffic conditions over time.

E. Testing Phase

After completing the training, the system was tested in a controlled but distinct simulation environment to evaluate the robustness of the learned policies. During this phase, the optimized weights of the Actor and Critic networks were loaded, and the exploration mechanism (ϵ-greedy) was disabled.

This deterministic approach ensured that the agent applied the acquired policies directly.

The key metrics analyzed during the testing phase included:

- **Average waiting times**: The average waiting time on the highway ramp and routes was significantly reduced compared to baseline scenarios:
  - **Scenario 1**: Traffic regulated using SUMO's default management system, where traffic lights operate with static cycles of 30 seconds without optimization.
  - **Scenario 2**: DRL-based traffic control with dynamic phase adjustments to optimize vehicle flow and minimize congestion.
- **Vehicle count**: The number of vehicles present on different routes was monitored to assess traffic density.
- **Cumulative deceleration**: Total amount of deceleration experienced by vehicles

These metrics allow us to compare the results of the testing phase with those obtained during training, identifying potential issues such as overfitting or the model's inability to adapt to new configurations as shown in Fig. 4, 5 and IV-E.

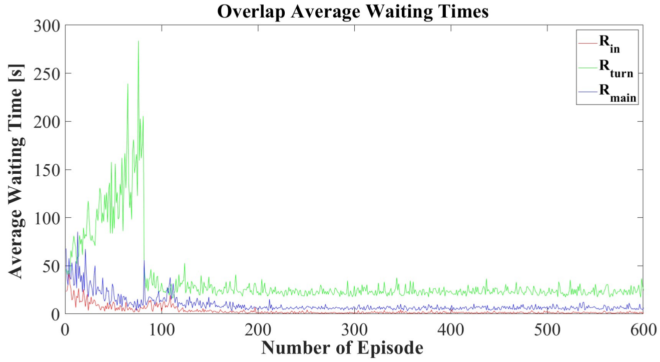


Fig. 4. Average waiting time evolution for different routes ($\mathcal{R}_{\text{in}}$, $\mathcal{R}_{\text{turn}}$, $\mathcal{R}_{\text{main}}$) during the training process. This metric reflects the efficiency of the traffic control strategy, where a reduction in waiting time indicates an improvement in traffic flow.

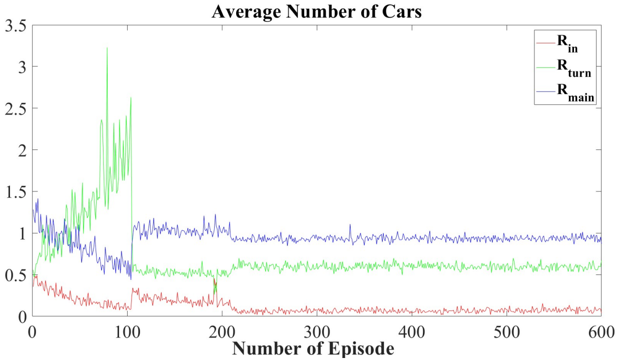
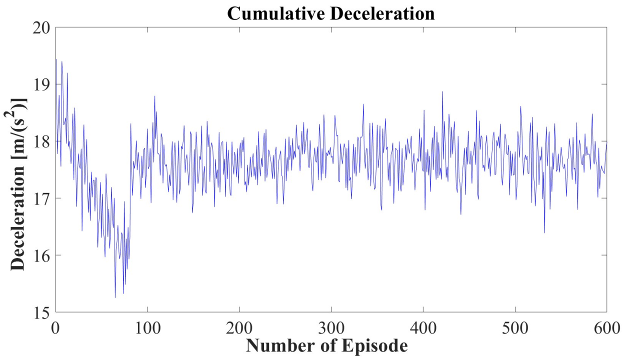


Fig. 5. Average number of vehicles on different routes ($\mathcal{R}_{\text{in}}$, $\mathcal{R}_{\text{turn}}$, $\mathcal{R}_{\text{main}}$) throughout the simulation. This metric provides insights into traffic density and the agent's ability to manage vehicle flow at merging points.



The DRL-based approach aims to minimize these events, ensuring smoother vehicle transitions.

Fig. 6. Cumulative deceleration recorded during the simulation. A high cumulative deceleration value indicates frequent sudden braking, which can be a sign of inefficient traffic merging and potential safety hazards.
The DRL-based approach aims to minimize these events, ensuring smoother vehicle transitions.

F. Discussion of Results

The proposed DRL framework was tested and validated through a series of simulations using the SUMO (Simulation of Urban Mobility) platform. The experiments were designed to evaluate the agent's ability to optimize traffic flow and improve safety metrics in a complex highway merging scenario. A comparative analysis was conducted on a single episode consisting of 400 steps to highlight the benefits of the intelligent traffic light control introduced by the DRL framework. The two scenarios discussed in the testing phase were evaluated. The results of this comparison indicated that the intelligent traffic lights managed by the DRL agent significantly improved overall network efficiency. Moreover, adaptive control reduced the number of cars as shown in Fig. 7.

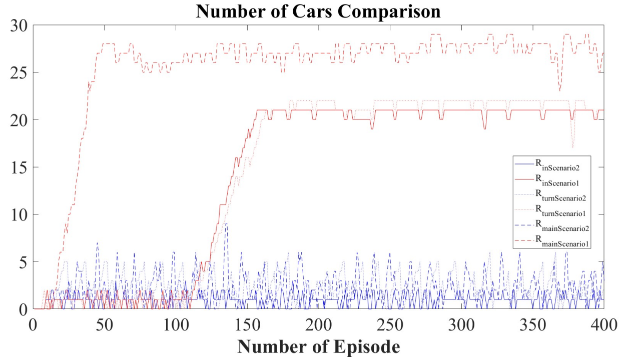


Fig. 7. Comparison of vehicle count on different routes between the proposed DRL framework and the baseline strategy. The analysis reveals an improved distribution of traffic, leading to reduced congestion in critical merging areas.

V. Conclusion

This study introduces a Deep Reinforcement Learning (DRL) framework to optimize vehicle merging at highway ramps and intersections, leveraging an Actor-Critic approach to improve traffic safety and efficiency. The proposed model effectively reduces congestion, sudden braking, and collision risks by enabling adaptive decision-making through a customized reward function. Moreover, the framework was validated considering a critical intersection of an highway of Northern Italy and by the simulation the traffic congestion is compared to some baseline scenarios.

Future research will focus on scaling the approach to larger networks, integrating real-time IoT and V2X data, and exploring multi-agent reinforcement learning to enhance coordination and adaptability in traffic management.

References

[1] W. Jin, Q. Xie, W. Wan, Y. Yamaguchi, D. M. Sime, and K. Sheng, "A Comparative Study on DRL-Based Autonomous Driving Under Single-Vehicle and Human-Vehicle Road Coordination," in *2023 2nd International Conference on Automation, Robotics and Computer Engineering (ICARCE)*, Wuhan, China, 2023, pp. 1–5. https://doi.org/10.1109/ICARCE59252.2024.10492530.

[2] C. Zhong, Z. Lu, M. C. Gursoy, and S. Velipasalar, "Actor-Critic Deep Reinforcement Learning for Dynamic Multichannel Access," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Anaheim, CA, USA, 2018, pp. 599–603. https://doi.org/10.1109/GlobalSIP.2018.8646405.

[3] A. Sujatha, M. Bore, A. Deepak, K. Mayuri, A. K. Khan, and R. Raj, "AI-Driven Traffic Control Systems for Smart Cities in Civil Engineering," in *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, Gautam Buddha Nagar, India, 2023, pp. 1589–1594. https://doi.org/10.1109/IC3I59117.2023.10397650.

[4] V. R. Burugadda, N. Jadhav, N. Vyas, and R. Duggar, "Exploring the Potential of Deep Reinforcement Learning for Autonomous Navigation in Complex Environments," in *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, Pune, India, 2023, pp. 1–6. https://doi.org/10.1109/ICCUBEA58933.2023.10392109.

[5] F. Bellotti, L. Lazzaroni, A. Capello, M. Cossu, A. De Gloria, and R. Berta, "Explaining a Deep Reinforcement Learning (DRL)-Based Automated Driving Agent in Highway Simulations," *IEEE Access*, vol. 11, pp. 28522–28550, 2023. https://doi.org/10.1109/ACCESS.2023.3259544.

[6] C. Lu, H. Chen, and S. Grant-Muller, "An Indirect Reinforcement Learning Approach for Ramp Control under Incident-Induced Congestion," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, The Hague, Netherlands, 2013, pp. 979–984. https://doi.org/10.1109/ITSC.2013.6728359.

[7] C. Lu, H. Chen, and S. Grant-Muller, "An Indirect Reinforcement Learning Approach for Ramp Control under Incident-Induced Congestion," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, The Hague, Netherlands, 2013, pp. 979–984. https://doi.org/10.1109/ITSC.2013.6728359.

[8] M. K. Tan, M. R. Ladillah, H. S. E. Chuo, K. G. Lim, R. K. Y. Chin, and K. T. K. Teo, "Optimization of Signalized Traffic Network Using Swarm Intelligence," in *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia, 2021, pp. 1–6. https://doi.org/10.1109/IICAIET51634.2021.9573784.

[9] J. J. Gonzalez-Delicado, J. Gozalvez, J. Mena-Oreja, M. Sepulcre, and B. Coll-Perales, "Alicante-Murcia Freeway Scenario: A High-Accuracy and Large-Scale Traffic Simulation Scenario Generated Using a Novel Traffic Demand Calibration Method in SUMO," *IEEE Access*, vol. 9, pp. 154423–154434, 2021. https://doi.org/10.1109/ACCESS.2021.3126269.

[10] Dutta, D., & Upreti, S. R. (2022). A survey and comparative evaluation of actor-critic methods in process control. *Canadian Journal of Chemical Engineering* https://doi.org/10.1002/cjce.24508

[11] F. Paparella, G. Olivieri, G. Volpe, A.M. Mangini, M.P. Fanti "A Deep Reinforcement Learning Approach for Route Planning of Autonomous Vehicles", *2024 IEEE International Conference on Systems, Man, and Cybernetics*, Oct. 6-10, 2024, Kuching, Sarawak, Malaysia.