Adaptive Traffic Signal Control using Deep Reinforcement Learning for Public Transport Prioritization in Davao City

Sid Adrian B. Digamon, Jhon Glenn L. Fabul, James Manon-og

ATENEO DE DAVAO UNIVERSITY

SCHOOL OF ARTS AND SCIENCES

COMPUTER STUDIES CLUSTER

DAVAO CITY

April 2025

Adaptive Traffic Signal Control using Deep Reinforcement Learning for Public Transport Prioritization in Davao City

A Thesis Presented to the
Faculty of the
Computer Studies Cluster
Ateneo de Davao University

In Partial Fulfillment
Of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Sid Adrian B. Digamon, Jhon Glenn L. Fabul, James Manon-og

ATENEO DE DAVAO UNIVERSITY
SCHOOL OF ARTS AND SCIENCE
COMPUTER STUDIES CLUSTER

April 2025

Table of Contents

List of Figures

Abstract

This research addresses critical challenges in urban traffic management by developing a passenger-centric deep reinforcement learning (DRL) approach to traffic signal control in Davao City. Traditional traffic signal systems optimize vehicle flow rather than passenger throughput, resulting in inefficient allocation that fails to prioritize high-occupancy vehicles. This study aims to design and implement a DRL-based traffic controller that prioritizes passenger movement over vehicle count, specifically addressing the unique transportation landscape of Davao City with its mix of private vehicles and public transportation options like jeepneys and buses. The methodology incorporates computer vision for vehicle classification, LSTM-enhanced temporal pattern recognition, and multi-agent coordination for network-level optimization. Through simulation and testing, this research contributes to the advancement of intelligent transportation systems that can significantly improve urban mobility by focusing on moving people rather than vehicles, with implications for reducing congestion and travel times across the Philippine urban context.

This research addresses critical challenges in urban traffic management by developing a passenger-centric deep reinforcement learning (DRL) approach to traffic signal control in Davao City. Traditional traffic signal systems optimize vehicle flow rather than passenger throughput, resulting in inefficient allocation that fails to prioritize high-occupancy vehicles. This study designs and implements a DRL-based traffic controller that prioritizes passenger movement over vehicle count, taking into account the unique transportation landscape of Davao City with its mix of private vehicles and public transportation options like jeepneys and buses. The system leverages available data to apply vehicle classification, LSTM-enhanced temporal pattern recognition, and multi-agent coordination for network-level optimization. Through simulation and testing, this research contributes to the advancement of intelligent transportation systems that can significantly improve urban mobility by focusing on moving people rather than vehicles, with implications for reducing congestion and travel times across the Philippine urban context.

Chapter 1: Introduction

1.1 Background of the Study

Rapid urbanization and escalating vehicle ownership have rendered traditional traffic signal systems increasingly ineffective in Davao City due to infrastructure limitations and failing sensor technologies. Traditional inductive loop detectors - used in 92% of Philippine traffic systems suffer 10-40% annual failure rates due to frequent roadworks and tropical climate conditions (Milan, 2023). District Officer Leonardo Pamplona confirms these sensors fail within 3-6 months post-installation due to pavement degradation from ongoing infrastructure projects. Fixed-time controllers, driven by static timing plans, cannot adjust to real-time conditions, while actuated systems extend or shorten green phases based solely on vehicle presence, ignoring vehicle occupancy. Consequently, a fully loaded jeepney carrying multiple passengers receives the same signal priority as a lone private car, underutilizing the roadway's actual people-moving capacity and worsening congestion, delay, and emissions.

Advances in deep reinforcement learning (DRL) offer a powerful alternative. By coupling neural networks with trial-and-error learning, DRL agents can discover adaptive signal-timing policies that respond to complex, time-varying traffic patterns. Yet, existing DRL implementations typically optimize vehicle throughput rather than passenger flow, treating buses and jeepneys identically to vehicles (Cai & Wei, 2024; Wei et al., 2021). Rule-based transit signal priority mechanisms further fall short: simple green-time extensions for high-occupancy vehicles often improve performance at one intersection while creating downstream delays (Bie et al., 2024; Rahman Swapno et al., 2024).

To bridge these gaps, a passenger-centric DRL framework is needed, one that uses computer vision to classify vehicle types, weights them by estimated occupancy, and embeds temporal pattern recognition into the agent's state representation. Such an approach promises higher person-throughput, more equitable multi-objective optimization (Scatto, 2024; Xu et al., 2020), and coordinated "green waves" across intersections (Liu & Park, 2021; Sharma et al., 2023). Tailoring this framework to Davao City's mixed fleet of jeepneys, buses, private cars, and motorcycles can unlock significant improvements in urban mobility by prioritizing the movement of people rather than vehicles.

1.2 Problem Statement

This study aims to address the following research questions:

1. How can a deep reinforcement learning (DRL) traffic controller be designed to optimize for passenger count rather than vehicle count in the context of Davao City's mixed traffic conditions?
2. How can transit signal priority mechanisms be effectively implemented within a DRL framework to extend green time for high-occupancy vehicles such as jeepneys and buses while maintaining overall network efficiency?
3. How can temporal pattern recognition capabilities be integrated into DRL traffic controllers to anticipate and respond to recurring traffic patterns in Davao City's urban environment?
4. How can a multi-agent DRL framework be implemented to coordinate traffic signals across multiple intersections and maximize passenger progression through a corridor?

1.3 Objectives of the Study

The general objective of this study is to develop and evaluate a simulated passenger-centric deep reinforcement learning approach to traffic signal control that optimizes for people movement rather than vehicle counts in the context of Davao City's urban transportation system.

The specific objectives are:

1. To develop a Double-Dueling DQN algorithm with a passenger-centric reward function and integrate YOLO-based vehicle classification detection for jeepneys and other vehicle types, validating its implementation in SUMO by increasing average passenger throughput per cycle and reducing average passenger waiting time by ≥10% versus fixed-time control.
2. To design and implement a transit-signal-priority mechanism that extends green time for high-occupancy vehicles, improving jeepney throughput by ≥15% while limiting overall vehicle delay increase to ≤10%, evaluated across three distinct traffic demand scenarios in a 3-intersection SUMO simulation.

3. To integrate an LSTM-enhanced state encoder into the Double-Dueling DQN agent, achieving at least 80% accuracy in predicting high-occupancy vehicle arrivals one signal cycle in advance over a validation set peak-hour patterns.
4. To extend the single-intersection agent to a multi-agent coordinated DRL system, reducing passenger delay and improving average jeepney travel time by ≥10%, measured over simulated of mixed traffic.

1.4 Significance of the Study

Addressing the identified gaps in Davao City's traffic management system will have significant implications for urban mobility, public transport reliability, and passenger experience. By shifting the optimization paradigm from vehicle-based to passenger-based metrics, the proposed system aims to increase the efficiency of public transportation, reduce overall congestion, and support the city's broader smart mobility and sustainability goals.

1. **Theoretical Contributions**: This study advances the field of deep reinforcement learning by developing novel architectures and reward formulations specifically designed for passenger-centric optimization. The integration of temporal pattern recognition through LSTM networks and the development of multi-agent coordination mechanisms contribute to the broader understanding of DRL applications in complex urban systems.
2. **Practical Applications**: The DRL-based traffic signal control system can theoretically significantly improve the efficiency of Davao City's transportation network by prioritizing the movement of people rather than vehicles. By giving appropriate priority to high-occupancy vehicles like jeepneys and buses, the system can increase the overall passenger throughput of the network, potentially reducing congestion and travel times.
3. **Local Context Relevance**: Unlike general-purpose traffic control solutions, this research is specifically tailored to the unique context of Philippine urban traffic, with its characteristic mix of transportation modes including jeepneys, buses, private vehicles, and motorcycles. The solutions are directly applicable to the specific challenges and opportunities present in Davao City's transportation ecosystem.
4. **Technology Transfer Potential**: The methodologies and frameworks in this research can be adapted and transferred to other Philippine cities facing similar urban traffic challenges, creating potential for broader national impact beyond the immediate scope of Davao City.

1.5 Scope and Limitations of the Study

This study focuses specifically on the application of deep reinforcement learning techniques to passenger-centric traffic signal control in the context of Davao City's urban transportation network. The scope encompasses:

1. **Location:** The study will focus on Davao City's urban core, specifically: the Sandawa intersection, the Ecoland Terminal to Bolton Bridge corridor, and the Ecoland intersection near John Paul College. While the methods developed may be applicable to other urban areas, the implementation and evaluation will be specifically contextualized for Davao City's unique traffic patterns and infrastructure.
2. **Transportation Modes**: The study will consider multiple vehicle types common in Philippine urban settings, including jeepneys, buses, private cars, and motorcycles, with particular emphasis on high-occupancy public transportation vehicles.
3. **Technical Approach**: The research will emphasize deep reinforcement learning techniques, specifically Double-Dueling DQN with LSTM enhancement and multi-agent coordination, as the primary methodological framework.
4. **Evaluation Environment**: All implementations will be tested and evaluated within the SUMO traffic simulation environment, calibrated to represent typical traffic patterns observed in Davao City.

The limitations of the study include:

1. **Simulation-Based Evaluation**: Due to practical constraints, the evaluation will be conducted entirely in simulation rather than through field implementation.
2. **Scope of Implementation**: The study will model a limited network of three intersections rather than a complete urban traffic network, focusing on demonstrating the efficacy of the approach at a manageable scale.
3. **Computational Constraints**: The complexity of DRL models may require simplifications or approximations to ensure tractable computation, potentially limiting the theoretical optimal performance of the approaches.
4. **Pedestrians Affecting Traffic**: Pedestrian flows and their impact on intersection congestion are not modeled or prioritized in this study, even though pedestrians contribute to overall traffic dynamics. This exclusion is due to the focus on maximizing vehicle-based passenger throughput and the lack of reliable pedestrian detection infrastructure in the current system.

Chapter 2: Review of Related Works

2.1 Approaches to Traffic Signal Control

Traffic signal control systems have evolved significantly over the past decades, from simple fixed-time controllers to adaptive systems capable of responding to changing traffic conditions. Traditional approaches to traffic signal control can be categorized into three main types: fixed-time, actuated, and adaptive control strategies.

Fixed-time control operates on predetermined signal timing plans regardless of actual traffic conditions. While simple to implement, these systems are inherently inflexible and perform poorly under variable traffic conditions. As noted by Cai and Wei (2024), fixed-time control cannot distinguish vehicle occupancy, effectively equating all vehicles regardless of passenger load. This fundamental limitation leads to inefficient resource allocation, particularly in contexts with diverse vehicle types carrying significantly different numbers of passengers.

Actuated control represents an improvement over fixed-time systems by using sensors to detect vehicle presence and make limited adjustments to signal timing. These systems typically extend or truncate green phases based on detected vehicles but lack the sophistication to optimize for complex objectives or to coordinate effectively across multiple intersections. Bie et al. (2024) observed that actuated systems with simple transit signal priority (TSP) capabilities extend or truncate green phases without balancing downstream impacts, limiting potential passenger-throughput gains.

2.2 Deep Reinforcement Learning for Traffic Signal Control

Recent advances in artificial intelligence, particularly in the domain of deep reinforcement learning (DRL), have created new opportunities for developing more sophisticated traffic control systems. DRL combines deep neural networks with reinforcement learning principles to enable agents to learn optimal decision policies through interaction with complex environments.

The application of reinforcement learning to traffic signal control was pioneered by researchers seeking to overcome the limitations of conventional control methods. Early approaches used tabular Q-learning, but these methods struggled to scale to realistic traffic scenarios due to the curse of dimensionality. The integration of deep neural networks to approximate the Q-function, known as Deep Q-Networks (DQN). Hassan et al. (2023) emphasize that DQN is particularly well-suited for traffic control as it removes the need for extensive environment simulation, allowing it to learn the optimal traffic management policy through direct interaction with real-world traffic conditions.

Several important variants of DQN have been applied to traffic signal control. In his 2024 master’s thesis, Scatto developed a Double Deep Q-Network (DDQN)-based adaptive traffic signal control system to improve urban mobility by balancing efficiency and fairness between vehicles and pedestrians. Using dual neural networks, dynamic experience replay, and real-time data (e.g., queue lengths, waiting times, signal phases), the system makes decisions every 5 seconds. DDQN reduces overestimation bias, achieving up to 32% shorter waiting times and over 40% shorter queues compared to traditional fixed-time and actuated systems. A fairness parameter (β), with an optimal value around 0.5, effectively balances pedestrian and vehicle priorities. DDQN also outperforms other reinforcement learning methods in managing delays and fairness (Scatto, 2024).

Dueling Q-Networks improve traditional DQN by separating the Q-function into two parts: one that estimates the overall quality of the traffic state (value) and one that measures the benefit of each possible signal action (advantage). Chen et al. (2021) showed that this separation is especially useful for traffic signal control, where it's important to judge the best signal phase even in similar traffic conditions. Liu et al. (2023) used this approach for adaptive traffic control, finding that the value stream assesses how good the overall traffic is, while the advantage stream highlights which signal phase is most beneficial. This helps the system prioritize public transport more effectively, even when traffic is heavy.

However, as noted by Li et al. (2022), standard DQN-based controllers treat each traffic snapshot independently, missing the opportunity to learn and exploit recurring patterns such as scheduled bus arrivals and daily rush-hour peaks. This limitation has led researchers to explore architectures that can capture temporal dependencies in traffic patterns.

2.3 Temporal Pattern Recognition in Traffic Control

Traffic patterns exhibit strong temporal dependencies at multiple scales, from short-term fluctuations to daily, weekly, and seasonal patterns. Capturing and exploiting these temporal patterns represents a significant opportunity to improve traffic signal control performance.

Recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have shown promise in capturing these temporal dependencies. Wu et al. (2021) demonstrated that LSTM-based traffic forecasting achieves 23.7% lower mean absolute error than Gaussian-process models by effectively capturing sequential dependencies in traffic data. Similarly, Zai et al. (2023) showed that LSTM architectures can effectively model both short and long-term traffic fluctuations in complex urban networks.

The integration of LSTM networks with DRL approaches for traffic signal control represents a promising direction for improving controller performance. By enabling controllers to learn and exploit recurring traffic patterns, LSTM-enhanced DRL can anticipate traffic state evolution rather than merely reacting to current conditions. This anticipatory capability is particularly valuable for coordinating signals across multiple intersections and for providing priority to scheduled public transportation services.

2.4 Passenger-Centric Optimization Approaches

Traditional traffic signal control systems optimize primarily for vehicle-based metrics such as minimizing vehicle delay or maximizing vehicle throughput. However, this approach inherently treats all vehicles equally, regardless of their passenger occupancy. In urban environments with diverse vehicle types carrying significantly different numbers of passengers, this leads to suboptimal allocation of green time from a passenger movement perspective.

Person-based optimization approaches address this limitation by explicitly accounting for vehicle occupancy in the control objective. Wei et al. (2021) demonstrated through simulation that person-based signal metrics could increase passenger throughput by up to 20% compared to vehicle-based metrics. This significant improvement highlights the potential of passenger-centric approaches, particularly in contexts with high public transportation usage.

To optimize traffic signals with a focus on passengers, it’s important to estimate how full vehicles are, and computer vision provides a way to do this. Among real-time object detection models, YOLOv8 stands out for its speed, accuracy, and simpler, anchor-free design. Elhanashi et al. (2024) explain that YOLOv8 uses computer vision, deep learning, and image processing to achieve excellent object recognition. Sheng et al. (2024) add that its ability to make predictions at multiple scales and its advanced network design help it perform well even in challenging conditions. Building on this, Lin (2022) created a smart traffic monitoring system by combining YOLO with convolutional fuzzy neural networks, showing how blending object detection with deeper learning methods can lead to better traffic analysis. Together, these developments show that YOLOv8 can be a strong foundation for creating more advanced and intelligent traffic monitoring solutions.

2.5 Public Transportation Prioritization

Transit Signal Priority (TSP) systems aim to improve public transportation performance by providing preferential treatment to buses, trams, or other high-occupancy vehicles at traffic signals. Traditional TSP approaches typically use simple rule-based mechanisms such as green time extensions or early green triggers when transit vehicles are detected approaching an intersection.

While these rule-based approaches can improve transit performance at individual intersections, they often fail to optimize network-wide passenger movement and can inadvertently worsen overall traffic flow. Rahman Swapno et al. (2024) observed that rule-based TSP extensions alone have been shown to increase bus delays at adjacent intersections when not coordinated across the network, highlighting the limitations of isolated priority mechanisms.

More sophisticated approaches to transit priority within DRL frameworks offer the potential to balance transit priority with overall network performance. By incorporating transit vehicle detection and passenger load estimation into the state representation and reward function of DRL controllers, these systems can learn to provide appropriate priority based on actual passenger loads and network conditions rather than following fixed rules.

Wang and Chen (2022) conducted a systematic review of public transport priority strategies in intelligent transportation systems, concluding that context-aware, adaptive priority mechanisms outperform static rule-based approaches across a range of urban scenarios. The integration of these insights into DRL frameworks represents a promising direction for improving public transportation performance while maintaining overall network efficiency.

2.6 Multi-Objective Optimization in Traffic Control

Traffic signal control involves inherently competing objectives: minimizing delays for different traffic movements, maximizing throughput, ensuring fairness across modes, and prioritizing certain vehicles based on occupancy or schedule adherence. Balancing these competing goals represents a significant challenge that cannot be adequately addressed by single-objective optimization approaches.

Scatto (2024) demonstrated the value of multi-objective approaches by incorporating a fairness parameter into a DDQN reward function, achieving simultaneous reductions in both pedestrian and vehicle delays. This work illustrates how carefully designed reward functions can balance competing objectives more effectively than simplified single-objective formulations.

Xu et al. (2020) further explored multi-objective reinforcement learning approaches for traffic signal control, demonstrating that explicit multi-objective RL formulations yield Pareto-optimal trade-offs between competing objectives. Their comparative study underscored the need for separate reward components to adequately capture the complexity of traffic control objectives.

In the context of passenger-centric optimization, multi-objective approaches are particularly valuable for balancing the prioritization of high-occupancy vehicles against the need to maintain acceptable service levels for other road users. Morales et al. (2023) developed a multi-objective reinforcement learning framework for balanced urban traffic management that explicitly considers both passenger throughput and vehicle delay objectives, demonstrating superior performance compared to single-objective approaches across a range of traffic conditions.

2.7 Network-Level Coordination for Traffic Signal Control

While optimizing individual intersections represents an important step toward improving traffic signal control, the full potential of intelligent control can only be realized through effective coordination across multiple intersections. This network-level coordination is particularly important for creating "green waves" that allow vehicles, especially high-occupancy transit vehicles, to progress smoothly through multiple intersections without stopping.

Liu and Park (2021) demonstrated that a graph-neural-network approach to multi-intersection control could achieve significant improvements in bus progression through coordinated timing across intersections. By representing the traffic network as a graph and applying message-passing techniques to share state information between intersections, their approach enabled coordinated control decisions that significantly reduced transit vehicle delay.

Similarly, Sharma et al. (2023) showed that coordinated multi-intersection control specifically designed to optimize passenger flow could reduce cumulative passenger delays by over 15% in simulation compared to uncoordinated approaches. This significant improvement highlights the potential benefits of network-level coordination for passenger-centric traffic control.

The development of effective multi-agent reinforcement learning (MARL) approaches for traffic signal control represents a promising direction for achieving this network-level coordination. By enabling multiple intersection controllers to learn cooperative policies that optimize network-wide objectives, MARL approaches can potentially overcome the limitations of isolated control strategies, particularly for creating passenger-focused progression routes through urban corridors.

2.8 Theoretical Framework

2.8.1 OpenCV

OpenCV (Open Source Computer Vision Library) is a robust, open-source collection of programming tools mainly intended for real-time computer vision tasks. Initially created by Intel in 1999, it has grown into a versatile, cross-platform resource widely applied in fields like traffic monitoring and control systems. The library includes a wide range of algorithms and functionalities for processing images and videos, making it highly valuable for traffic-related applications that rely on real-time visual data analysis.

For traffic monitoring and adaptive signal control, OpenCV plays a crucial role by enabling image preprocessing, feature extraction, and object detection. Techniques such as noise filtering, converting images to grayscale, and detecting edges help improve image quality and emphasize important elements like vehicle shapes and license plates. These preprocessing steps are essential for refining traffic camera footage, allowing for more accurate vehicle detection and reliable traffic condition assessments. (Preetha et al., 2025)

2.8.2 YOLOv8

YOLOv8 stands out in real-time object detection due to its high speed, improved accuracy, and anchor-free approach. This model utilizes computer vision, neural networks, deep learning, and image processing techniques to achieve exceptional object recognition performance. Its multi-scale prediction capability and advanced backbone network further enhance detection efficiency. Designed for a variety of detection tasks, YOLOv8 handles everything from basic object identification to more complex challenges like instance segmentation, key point detection, object orientation, and classification.(Elhanashi et al., 2024).

2.8.3 Traditional Q-Learning

Q-Learning is a value-based reinforcement learning algorithm that estimates the optimal action for each state by learning from interactions with the environment (Clifton & Laber, 2020). It has been widely applied in traffic control, particularly for optimizing flow at single intersections and within multi-agent systems. The algorithm updates its knowledge by receiving rewards after actions are taken, gradually improving its decision-making policy (Miletić et al., 2022). Q-learning is limited in complex environments due to its reliance on discrete state and action spaces, making it less suitable for high-dimensional, real-world problems. This has led to the development of more advanced methods like Deep Q-Learning.

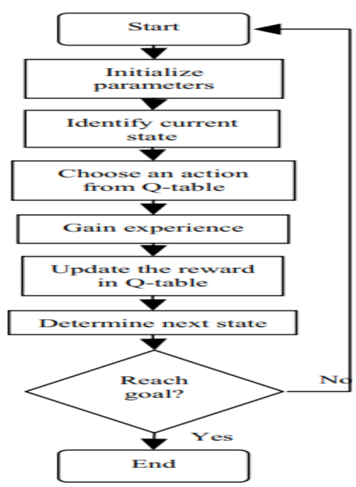


Figure 2.8.3.1 Q-Learning flowchart (Chin et al., 2024)

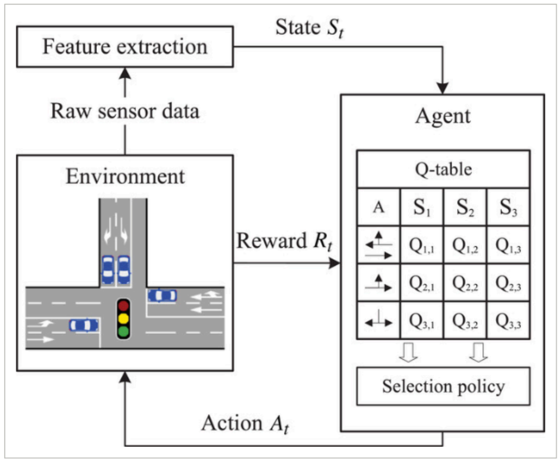


Figure 2.8.3.2 A value-based RL, Q-learning.(Miletić et al., 2022)

2.8.4 Neural networks

Neural networks, or artificial neural networks (ANNs), are computational models inspired by the structure of the human brain. Composed of interconnected nodes or "neurons," these models learn by adjusting the weights of these connections to minimize prediction errors, typically using metrics like Mean Square Error during training (Panagiotou & Dounis, 2022). Their strength lies in handling nonlinear relationships and extracting patterns from complex data, whether in image, text, or tabular form (Abdolrasol et al., 2021).

As the complexity of tasks grew, neural networks evolved into deep neural networks (DNNs)—models with multiple hidden layers that allow for more abstract and powerful feature decision-making tasks. In reinforcement learning, DNNs serve as the backbone of algorithms like Deep Q-Networks (DQNs), empowering them to learn optimal strategies in dynamic environments such as traffic control systems (Mienye & Swart, 2024).

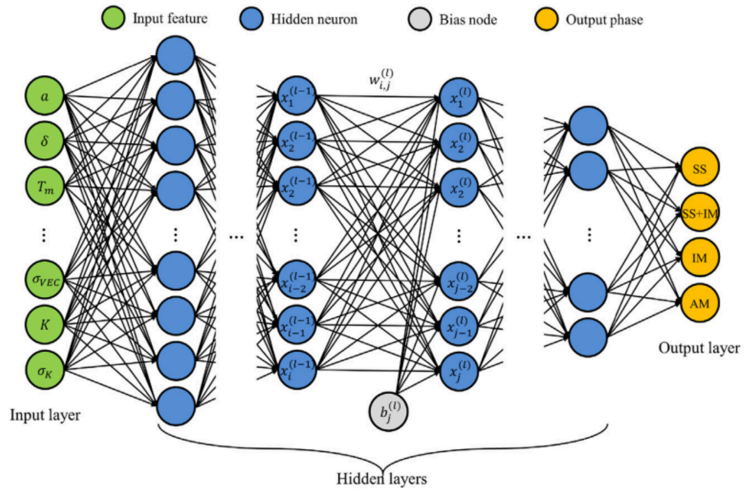


Figure 2.8.5.1 A schematic representation of DNN (Lee et al., 2020)

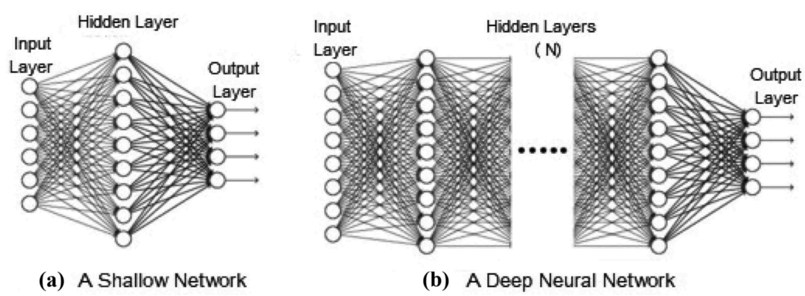


Figure 2.8.5.2 A general architecture of (a) a shallow neural network with a single hidden layer, and (b) a deep neural network containing multiple hidden layers (Sarker, 2022).

2.8.6 Deep Q Network

Deep Q-Learning (DQN) is an advanced reinforcement learning technique that addresses the limitations of traditional Q-learning in complex environments. While Q-learning relies on a tabular Q-function suitable for simple scenarios, DQN uses deep neural networks to approximate action values, making it effective in dynamic, high-dimensional settings (Moreno-Malo et al., 2024). This is especially useful in traffic signal control, where DQN can learn optimal policies directly from real-world data without requiring full environment simulations (Hassan et al., 2023).

Further enhancements to DQN allow it to handle continuous state and discrete action spaces, enabling more scalable and intelligent traffic optimization (Yang et al., 2021). A key feature in traffic applications is the inclusion of vehicle throughput in the reward function—representing the number of vehicles passing through an intersection in a given time. By considering throughput, along with queue length and delays, DQN-based models can effectively improve traffic flow and reduce congestion (Pan, 2023).

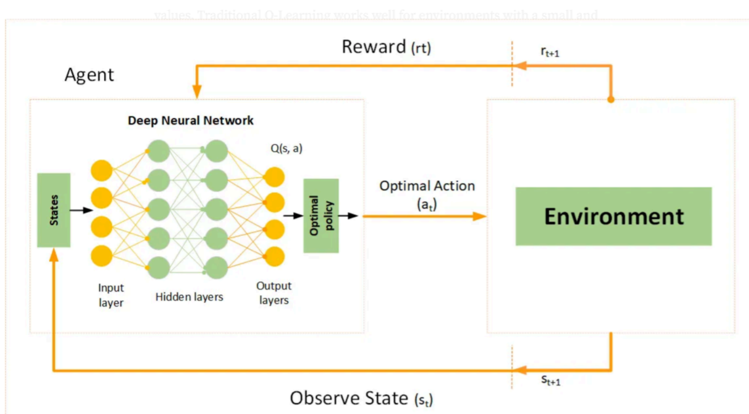


Figure 2.8.6.1 Structure of DQN (Amin, 2020)

2.8.7 Double Deep Q Network

In reinforcement learning, Double DQN (DDQN) enhances standard DQN by addressing its tendency to overestimate action values—an issue that can lead to poor decision-making. In DQN, the same network selects and evaluates actions, which often results in inflated value estimates, especially for rarely chosen actions. This can cause the agent to adopt suboptimal strategies. DDQN resolves this by introducing two separate networks: the main network selects the best action, while a slower-updating target network evaluates it. This separation leads to more accurate value updates, stabilizes learning, and helps prevent "catastrophic forgetting," where previously learned strategies are lost (Tareq et al., 2024).

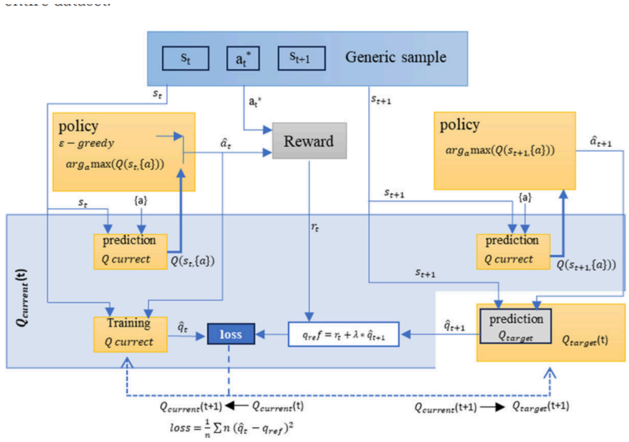
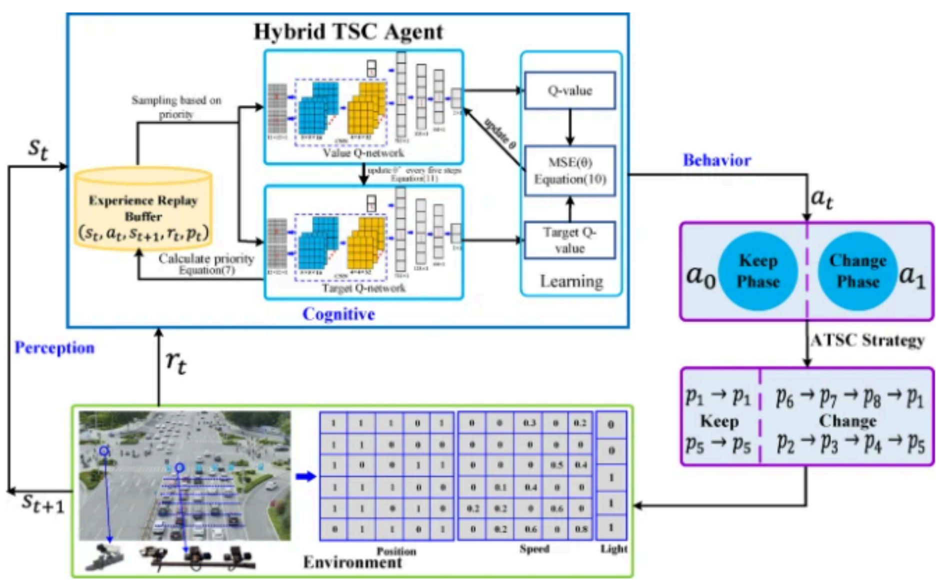


Figure 2.8.7.1 DQN Model: Online Network (Q current) and Target Network (Q target) (Tareq et al., 2024)



Pri-DDQN model

Figure 2.8.7.2 A DDQN Model showing Double DQN being utilized for adaptive traffic signal control (Zheng et al., 2025)

2.8.8 Double Dueling Q Network

In typical deep learning models used in supervised learning, architectures are sequential—each layer connects linearly, with no branches or loops. While DQN and Double DQN maintain this structure, Dueling DQN introduces a more complex design by branching into two sub-networks after the shared layers (such as convolutional layers) (Sewak, 2019).

One branch, the Value Network, estimates the value of being in a particular state, **V(s)**, regardless of the action. The other, the Advantage Network, estimates the advantage of taking a specific action in that state, **A(s,a)**. These outputs are then combined to compute the final Q-value.

$$Q(s,a) = V(s) + A(s,a)$$

However, directly summing **V(s)** and **A(s,a)** leads to **unidentifiability**, where multiple combinations of values produce the same Q-value, making learning unstable. To resolve this, the advantage values are normalized—typically by subtracting their average—before being combined. This ensures unique, stable estimates and improves learning performance(Sewak, 2019).

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|A|} \sum_{a'} A(s,a') \right)$$

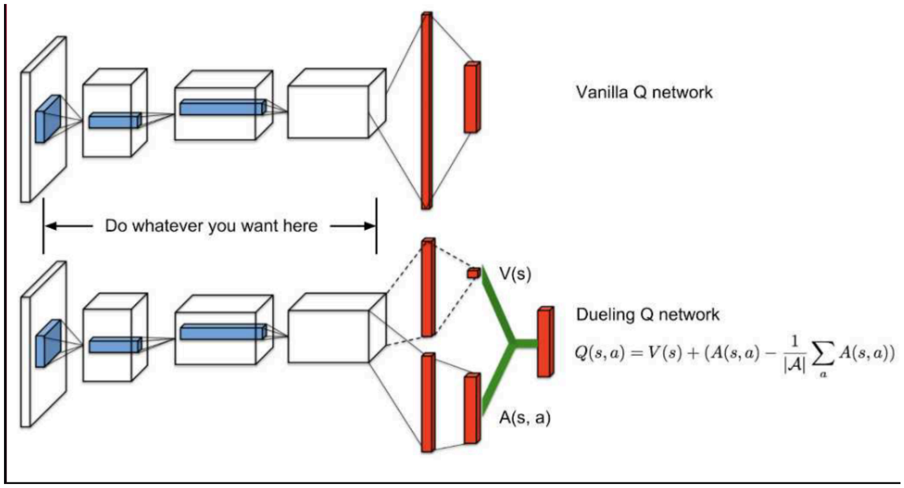


Figure 2.8.8.1 A DQN compared to the Dueling Deep Q network showcasing the value and advantage network. (Pham et al., 2022)

2.8.9 Multi-Agent

Multi-Agent Reinforcement Learning (MARL) is a framework where multiple agents learn to make decisions within a shared environment. Each agent observes its surroundings, takes actions, and receives feedback in the form of rewards. MARL integrates concepts from game theory and machine learning, enabling agents to adapt, coordinate, or compete depending on the task. Unlike single-agent settings, MARL introduces complexities such as non-stationarity, decentralized decision-making, and coordination challenges, making it suitable for dynamic, real-world applications like traffic control and robotics (Huh & Mohapatra, 2024).

In traffic management scenarios, each agent, such as a traffic signal or vehicle, acts independently but often works toward a global goal, like reducing overall congestion. Even without direct communication, agents can learn coordinated behaviors through shared environmental dynamics and a carefully designed reward structure (Azfar & Ke, 2024).

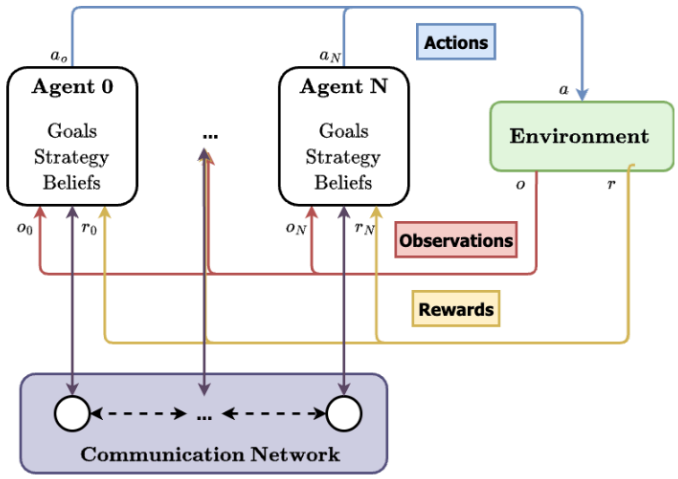


Figure 2.8.9.1 A visualization of a generalized multi-agent system following the iterative control process. (Huh & Mohapatra, 2024).

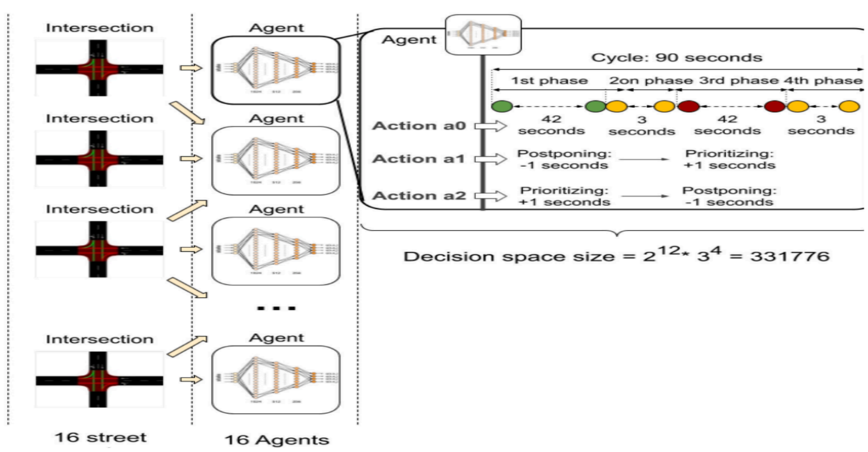


Figure 2.8.9.2 The simulation uses a multi-agent system, each running the proposed DQL algorithm in its study to choose the best action based on current conditions.(Moreno-Malo et al., 2024)

2.8.10 LSTM

Long Short-Term Memory (LSTM) models, a type of recurrent neural network (RNN), have become a significant tool for prediction, particularly in time series analysis. Unlike traditional Artificial Neural Networks (ANNs), LSTM models excel at preserving information from previous time steps, making them highly suitable for capturing temporal dependencies within time series data (Mu et al., 2020). This capability to "remember" past information is crucial for accurately predicting future values in sequences.

2.8.11 Sumo

Simulation of Urban Mobility (SUMO) is a powerful open-source platform for modeling urban traffic in both micro and macroscopic detail. It offers individual-level control over vehicles and infrastructure, such as traffic lights, making it a popular choice for multi-agent reinforcement learning (MARL) experiments (Azfar & Ke, 2024; Clemente, 2022). SUMO’s flexible and extensible design allows seamless integration with external systems, including reinforcement learning agents, computer vision tools, and real-world data sources. As a data-driven simulator, it can use pre-recorded observations—like vehicle counts from CCTV footage—as environmental input, processed through object detection tools such as YOLO and OpenCV. Its real-time interfacing capabilities also support co-simulation with platforms like CARLA, enabling live traffic control scenarios (Azfar & Ke, 2024). This setup enhances realism and allows testing MARL agents under noisy, imperfect sensing conditions, helping bridge the simulation-to-reality gap and validating the feasibility of vision-based traffic control systems (Huh & Mohapatra, 2024).

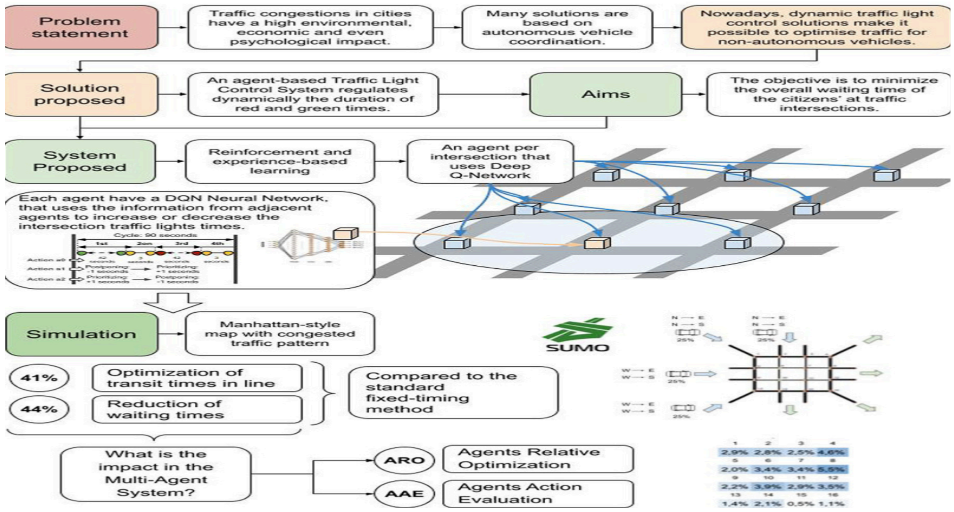


Figure 2.8.11.1 Graphical abstract of an improved traffic light system showcasing the use of SUMO for the simulation. (Moreno-Malo et al., 2024)

Chapter 3: Methodology

3.1 Conceptual Framework

This study's conceptual framework begins with acquiring traffic data using CCTV footage installed at key intersections in Davao City. This footage is processed by a parallel thesis team using a combination of OpenCV and YOLO, producing pre-classified vehicle detection data, which includes vehicle types, positions, and timestamps. The processed data is then structured into traffic state representations, incorporating key features such as vehicle density, estimated passenger count based on vehicle type, lane-level queue lengths, and temporal traffic flow patterns.

Once the traffic states are generated, they are fed into the SUMO simulation environment, which is configured to replicate the actual layout and flow conditions of selected Davao City intersections. The simulation begins with a baseline fixed-time traffic control model to establish a reference point for system performance. After establishing baseline metrics, the reinforcement learning system is initialized using a multi-agent architecture, where each intersection is assigned an autonomous agent responsible for its own signal control policy. Each agent receives localized state inputs and applies an offline-trained Double Dueling Deep Q-Network (DDDQN), enhanced with Long Short-Term Memory (LSTM) layers to capture sequential traffic behavior.

The system operates in episodic cycles, where agents select traffic signal actions based on their state inputs and receive rewards that reflect improvements in passenger throughput, reduced waiting time, and overall intersection efficiency. The SUMO environment processes these actions, updates the network simulation, and returns the new state for the next iteration. Over multiple simulation episodes, each agent’s policy is refined through the reward-based learning mechanism, updating the Q-network weights until a point of convergence is reached, defined by stable performance in key metrics.

Throughout the simulation process, no explicit communication occurs between agents. However, coordination emerges through the shared environment, where the actions of one intersection affect the state observed by others. This indirect influence enables the emergence of cooperative behavior. The final trained system is evaluated against the fixed-time baseline using performance metrics such as average passenger throughput, average vehicle delay, maximum queue length, and public transportation prioritization indicators. These evaluations confirm the system’s ability to learn traffic-optimized policies that align with real-world objectives focused on passenger flow efficiency.

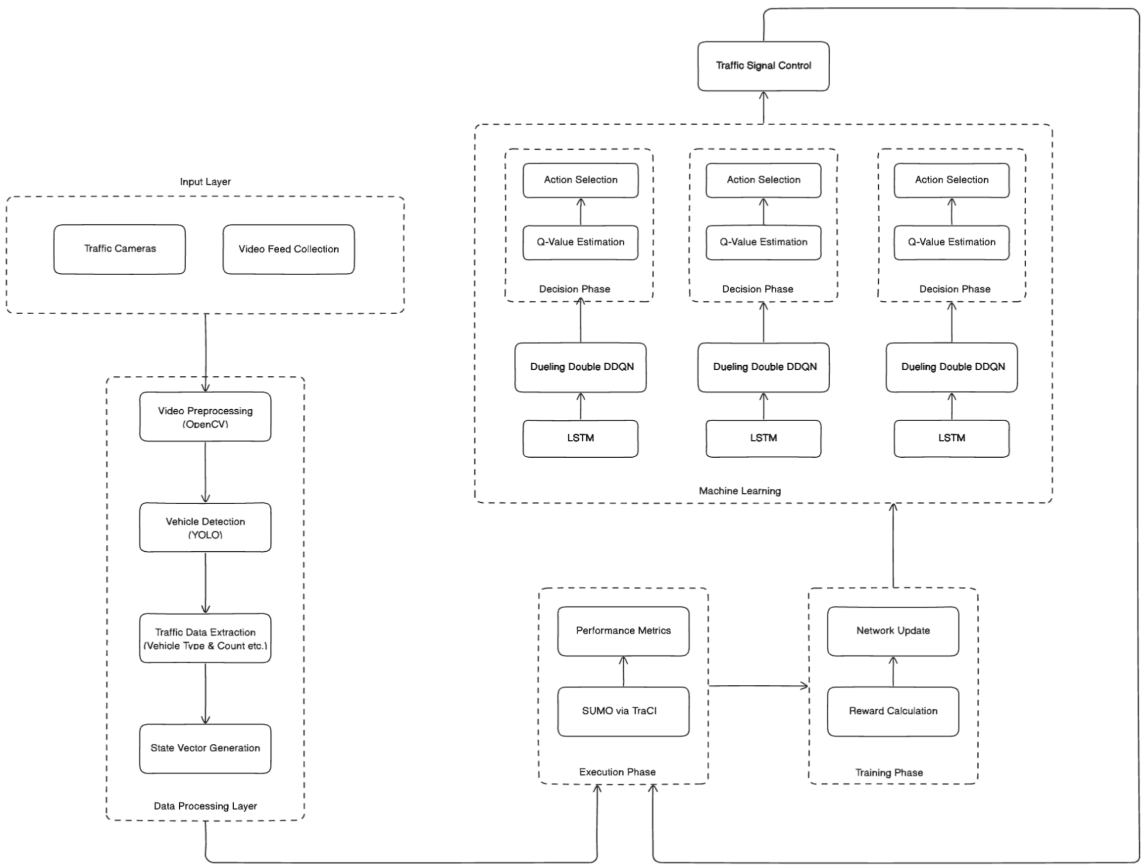


Figure 3.1.1 Conceptual Framework

3.2 Data Collection

The data for this study will be collected from two primary sources to support traffic state construction and simulate intersection behavior within a reinforcement learning framework. These sources include CCTV footage processed by a parallel thesis team using OpenCV and YOLO and simulated traffic flow data generated through the SUMO (Simulation of Urban Mobility) environment. A formal data-sharing arrangement will govern the transfer of the processed vehicle detection outputs, ensuring consistency between the external detection data and the simulation inputs.

3.2.1 Vehicle Detection and Classification Data

Vehicle detection and classification data will be sourced from a parallel thesis team that employs OpenCV for video processing and YOLO for object detection. Their outputs will include timestamped records of vehicle detections, types, positions, and estimated occupancy levels based on vehicle type. These data will serve as the basis for constructing traffic state vectors, which will represent the environment observed by the reinforcement learning agents. Since the object detection system is external to this study, model training and parameter tuning for YOLO are not within this project's scope. Instead, the study will rely on the partner team's validation metrics and accuracy reports for reliability.

3.2.2 Traffic Simulation Data

Before simulation runs begin, a map integration phase will be conducted using SUMO's network generation tools. This process will involve selecting a specific geographic area, Davao City, in this case, through OpenStreetMap (OSM) imports or equivalent sources. The extracted road network will then be customized to reflect actual intersection geometries and traffic conditions, including defining initial traffic volumes, routes, and allowed vehicle types.

To accurately represent the presence of public transportation vehicles such as jeepneys, which are not natively available in SUMO, the study will define a custom vehicle model. SUMO allows users to create vehicle types with specific parameters such as length, acceleration, deceleration, maximum speed, color, and passenger capacity. Using these customization capabilities, jeepneys will be modeled to reflect their physical dimensions and passenger occupancy levels, enabling the system to differentiate their movement and priority within the simulation.

The simulation itself will be prepared using SUMO's network files, route files, and traffic signal logic files, all of which will be stored in a dedicated scenario folder. Once configured, the simulator will execute episodes where traffic flow evolves in response to agent actions and predefined traffic demand conditions. SUMO’s built-in sensors and the TraCI interface will be used to collect performance metrics such as queue lengths, delays, and throughput.

3.2.3 Data Integration and Formatting

Processed visual data and simulation configurations will be integrated through a preprocessing phase that converts vehicle detections into structured traffic state vectors. These vectors will contain lane-level vehicle counts, queue lengths, passenger estimates, and signal phase histories, and will be updated episodically during simulation runs. Python scripts and TraCI commands will handle the data flow between the detection outputs, the reinforcement learning system, and the SUMO environment.

The final integrated dataset will consist of all simulation inputs and outputs required to train and evaluate the multi-agent traffic control system. This dataset will support performance comparisons between baseline fixed-time signal control and the learned policies generated by the reinforcement learning agents.

3.3 Data Analysis

The analysis of system performance will be conducted by evaluating the effectiveness of the proposed reinforcement learning-based adaptive traffic signal control system in comparison to a fixed-time control baseline. This evaluation will rely on simulation outputs generated within the SUMO environment under consistent and controlled traffic scenarios. To ensure the reliability and objectivity of results, multiple simulation episodes will be run and analyzed using established performance indicators.

Passenger-centric metrics will include total and average passenger throughput, as well as average waiting times weighted by estimated occupancy per vehicle. These metrics will be computed based on vehicle classifications and route tracking enabled through SUMO’s lane area detectors and data logging tools. Vehicle-centric metrics will include average vehicle delay, average vehicle throughput per signal cycle, and maximum queue lengths. Maximum queue length will be defined as the highest observed number of queued vehicles in any lane or intersection approach during a simulation episode, highlighting critical congestion conditions not visible through average-based metrics.

It is important to note that most performance indicators retrieved from the SUMO simulation, such as speed, delay, and waiting time, are inherently calculated as average values across vehicles or time intervals. These aggregated averages offer a general view of system performance and will be complemented by additional metrics that capture extreme or edge-case traffic behavior.

Public transportation prioritization metrics will assess improvements in jeepney and bus movement, particularly in terms of delay reductions and service fairness relative to other vehicle types. To assess this, comparisons will be made across vehicle categories, evaluating whether the reinforcement learning system maintains equitable treatment while prioritizing passenger volume.

Each system configuration baseline fixed-time control, and the proposed deep reinforcement learning model using Double Dueling DQN with LSTM will be tested under equivalent conditions. Descriptive statistics will be used to compare the performance of each approach, and where necessary, inferential statistical techniques such as hypothesis testing may be applied to determine the significance of observed differences. The final results will be used to validate whether the proposed system delivers meaningful improvements in traffic flow efficiency and passenger-centric performance.

3.4 Ethical Consideration

This research recognizes the importance of upholding strong ethical standards in the development and potential future deployment of an adaptive traffic signal control system using computer vision and deep reinforcement learning. Throughout the study, care will be taken to prioritize privacy, ensure fairness, promote public safety, and maintain transparency and accountability.

Protecting the privacy of individuals is a central concern in this research. To address this, all vehicle data collected through computer vision will be anonymized immediately, avoiding the storage of personal or identifiable information. Instead of focusing on individual vehicles, only general traffic statistics will be used. Strict protocols will be followed when handling CCTV footage to ensure that no sensitive information, such as license plates or personal appearances, is misused. Secure data storage systems, access controls, and clear data retention policies will be implemented to further protect the data. In cases where real-world data is used for calibration, it will be anonymized and aggregated to prevent the identification of specific individuals or travel patterns.

Fairness will be an important priority in the design and evaluation of the system. The system will aim to benefit all road users fairly, avoiding favoritism toward specific routes, times, or user groups. Efforts will be made to ensure that public transportation users, especially those relying on jeepneys and other high-occupancy vehicles, receive appropriate prioritization without negatively affecting other drivers. Special attention will be given to evaluating how improvements in traffic flow impact different communities and neighborhoods in Davao City, promoting both geographic and social equity across the system. Additionally, system performance will be monitored across different times of day to ensure fairness beyond just rush hours.

Throughout the entire research process, high standards of research ethics will be followed. Methods and results will be documented clearly and honestly, with limitations and uncertainties openly acknowledged. Data will be handled responsibly to prevent manipulation or r selective reporting. Any potential conflicts of interest will be identified and managed properly. The research will also practice responsible innovation by anticipating possible unintended consequences, considering different future scenarios, and creating ways to adapt if unexpected issues arise. Stakeholder perspectives will be included whenever possible to ensure that diverse voices are considered in system design and evaluation.

3.5 Operational Framework

The methodology described in the study by Azfar et al. outlines a co-simulation framework for traffic signal control that shares key structural similarities with the operational framework developed in this research. Their approach begins by establishing a simulation environment using CARLA and SUMO, where virtual infrastructure cameras are deployed within CARLA to mimic real-world surveillance systems.

Following the simulation setup, virtual camera feeds are generated and processed using YOLO for vehicle detection and counting. The outputs, which include traffic metrics such as lane density and queue length, are then used to construct the state representation for multi-agent reinforcement learning (MARL) agents operating within the SUMO environment. These agents, trained through a Q-learning algorithm, interpret the observed states to determine optimal signal phase transitions, with the objective of improving overall traffic flow efficiency.

A parallel structure is observed in the operational framework proposed in this research. Instead of virtual camera feeds, real-world CCTV footage serves as the primary input. The recorded video data is processed using OpenCV and YOLO to detect vehicles, classify types, and extract critical traffic features, such as vehicle counts and queue lengths. It is important to note that the processed detection outputs are provided by a parallel thesis team specializing in computer vision, and are utilized in this study solely for traffic state generation. These features are then structured to form the state representation within the SUMO simulation environment.

Following state generation, MARL agents, similarly based on reinforcement learning techniques, are tasked with interpreting the current traffic conditions. Using a reward-based learning mechanism, the agents select optimal signal timing actions that are executed within the SUMO environment. The resulting changes in the traffic state are captured and used for further learning cycles.

Although the framework proposed in this study differs from that of Azfar et al. by operating in an offline setting using pre-recorded real-world data instead of real-time simulation, the overall structure remains conceptually aligned. Both approaches follow a consistent flow from visual data acquisition, traffic state generation, reinforcement learning-based decision-making, and feedback-driven control within the SUMO environment. The study conducted by Azfar et al. thus serves as a strong methodological reference supporting the operational framework presented in this research.

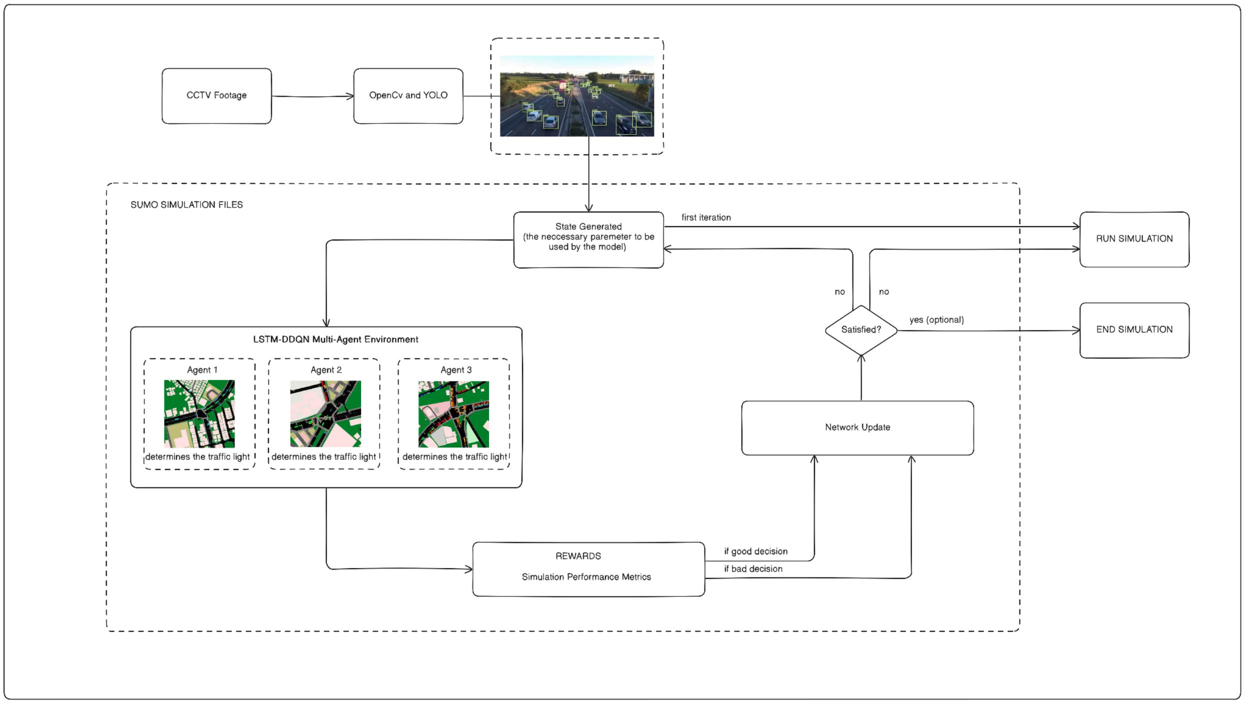


Figure 3.5.1 Operational Framework

3.6 Technical Output Prototype

3.6.1 SUMO Environment

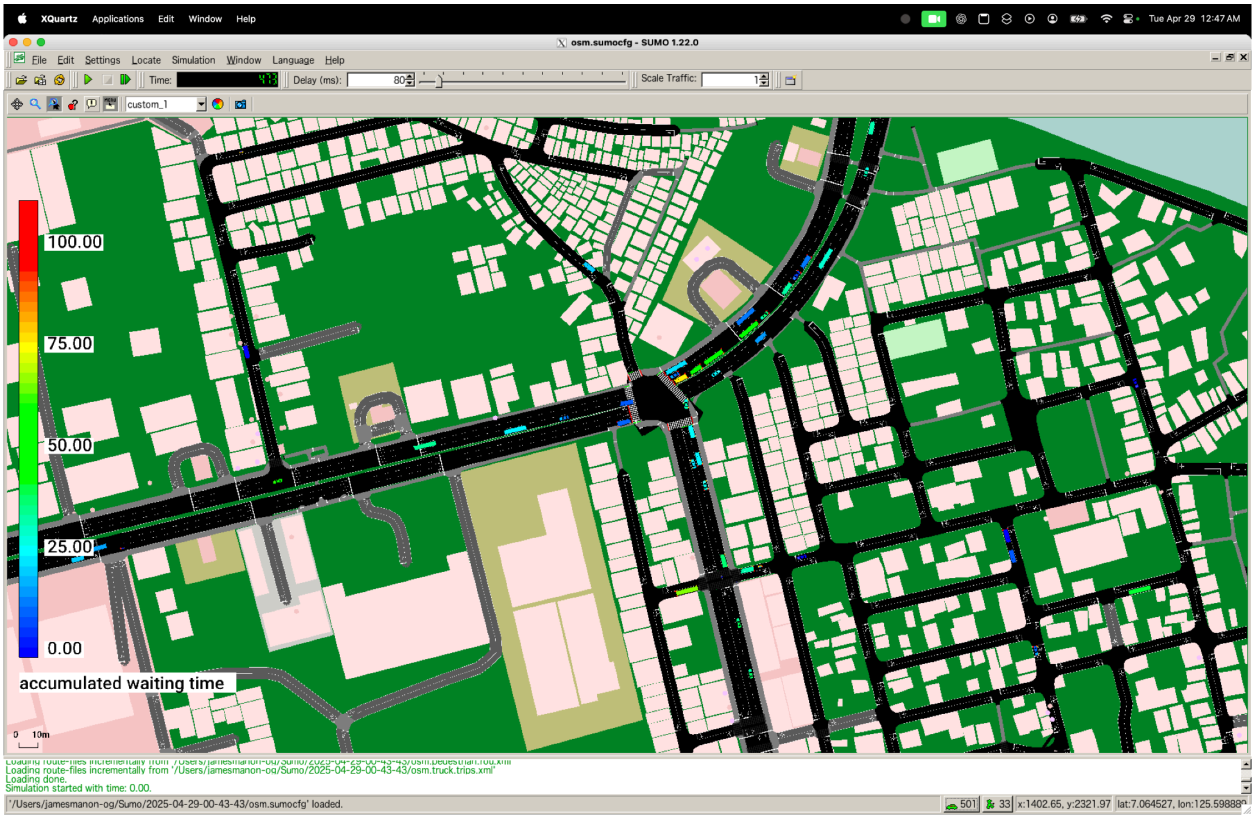


Figure 3.6.1.1 SUMO Home Screen

3.6.2 Vehicle Classification

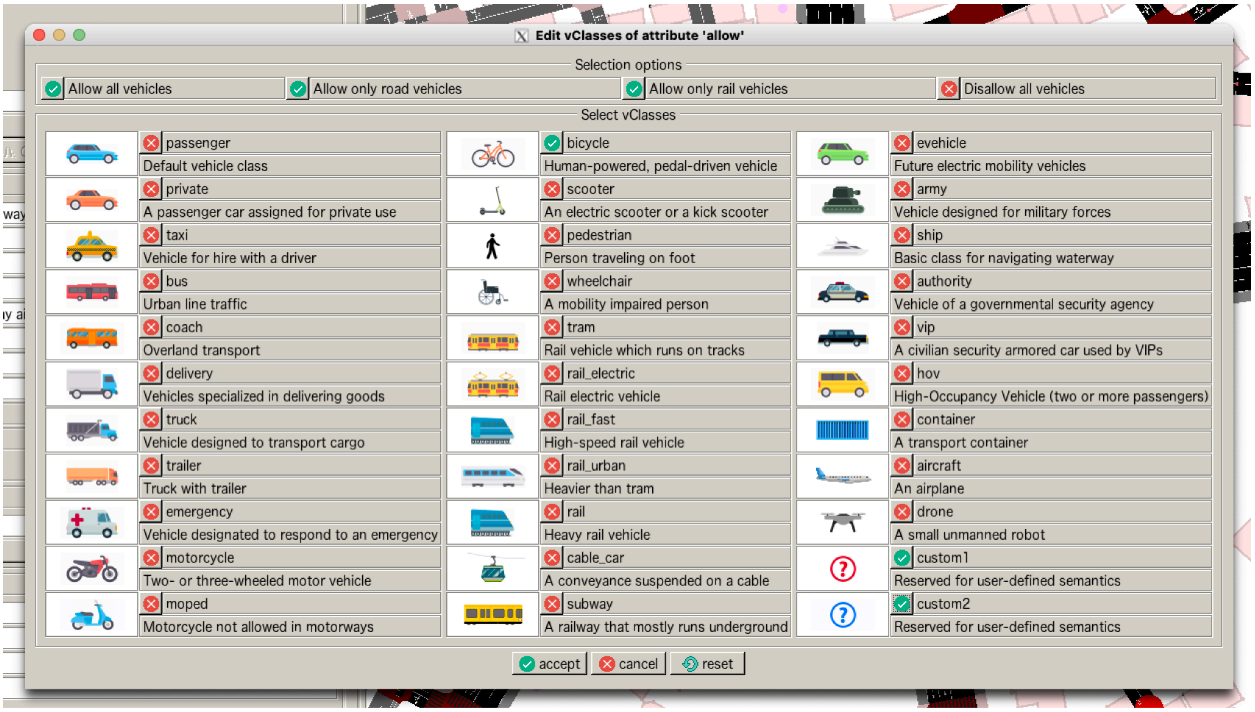
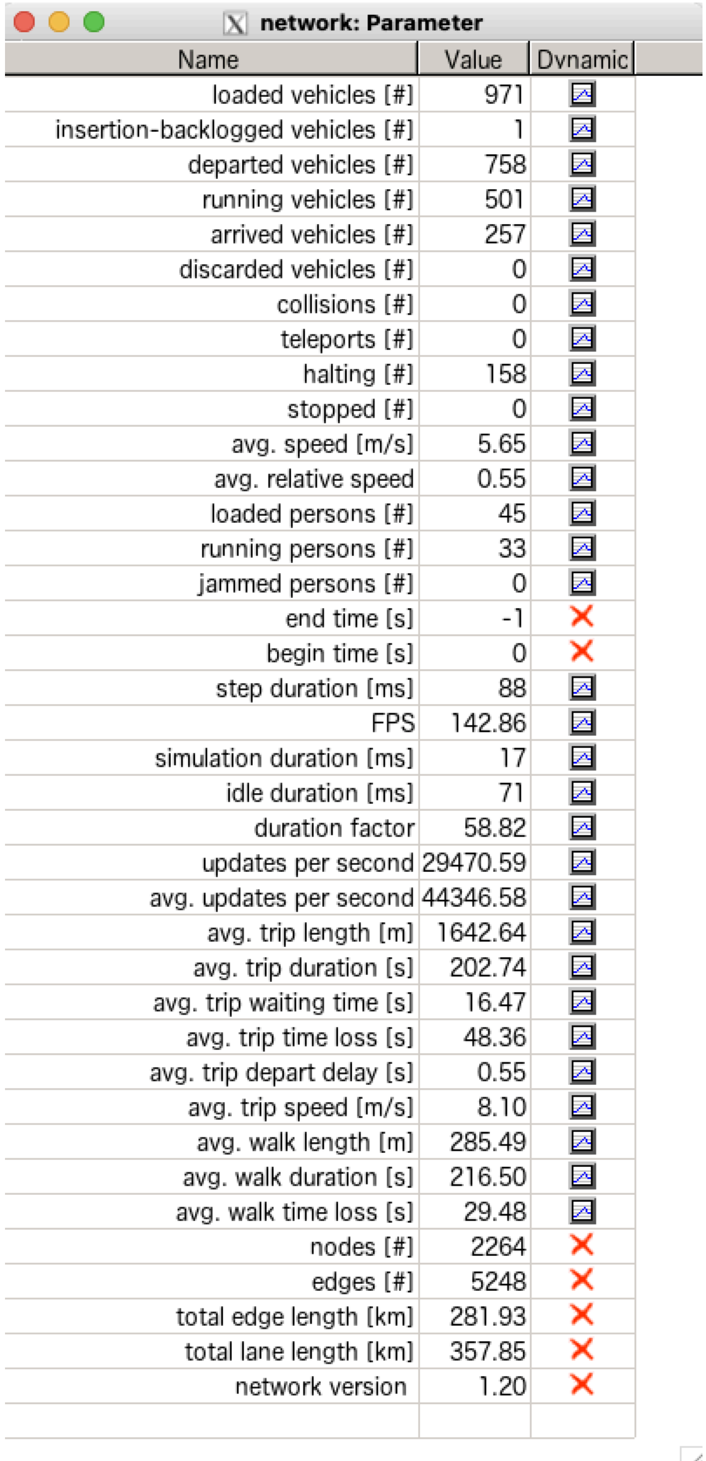


Figure 3.6.2.1 SUMO Vehicle Selection

3.6.3 Network Parameters



network: Parameter

Name	Value	Dynamic
loaded vehicles [#]	971	
insertion-backlogged vehicles [#]	1	
departed vehicles [#]	758	
running vehicles [#]	501	
arrived vehicles [#]	257	
discarded vehicles [#]	0	
collisions [#]	0	
teleports [#]	0	
halting [#]	158	
stopped [#]	0	
avg. speed [m/s]	5.65	
avg. relative speed	0.55	
loaded persons [#]	45	
running persons [#]	33	
jammed persons [#]	0	
end time [s]	-1	×
begin time [s]	0	×
step duration [ms]	88	
FPS	142.86	
simulation duration [ms]	17	
idle duration [ms]	71	
duration factor	58.82	
updates per second	29470.59	
avg. updates per second	44346.58	
avg. trip length [m]	1642.64	
avg. trip duration [s]	202.74	
avg. trip waiting time [s]	16.47	
avg. trip time loss [s]	48.36	
avg. trip depart delay [s]	0.55	
avg. trip speed [m/s]	8.10	
avg. walk length [m]	285.49	
avg. walk duration [s]	216.50	
avg. walk time loss [s]	29.48	
nodes [#]	2264	×
edges [#]	5248	×
total edge length [km]	281.93	×
total lane length [km]	357.85	×
network version	1.20	×

Figure 3.6.3.1 SUMO Monitoring Dashboard

3.6.4 Multi-Agent Environment

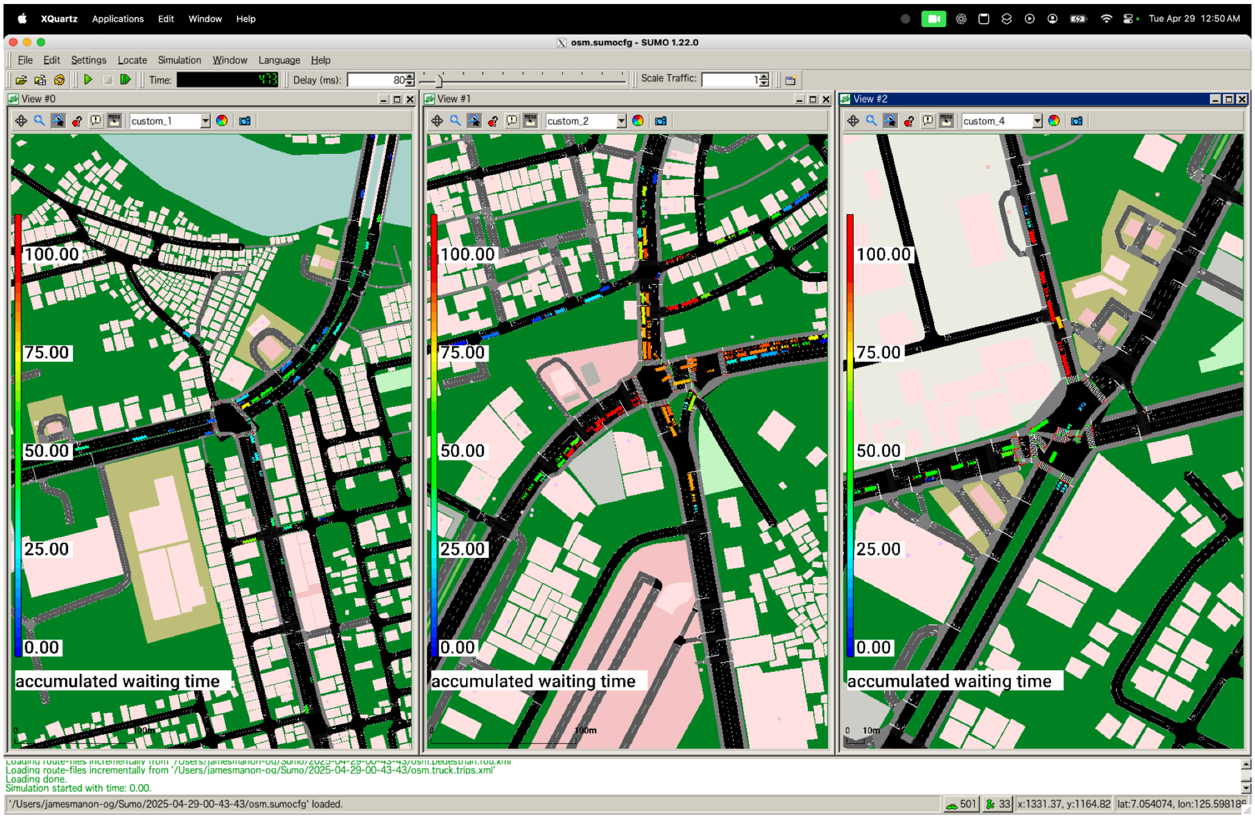


Figure 3.6.4.1 SUMO Traffic Control Across Three Intersections

3.6.5 Traffic Light Parameter

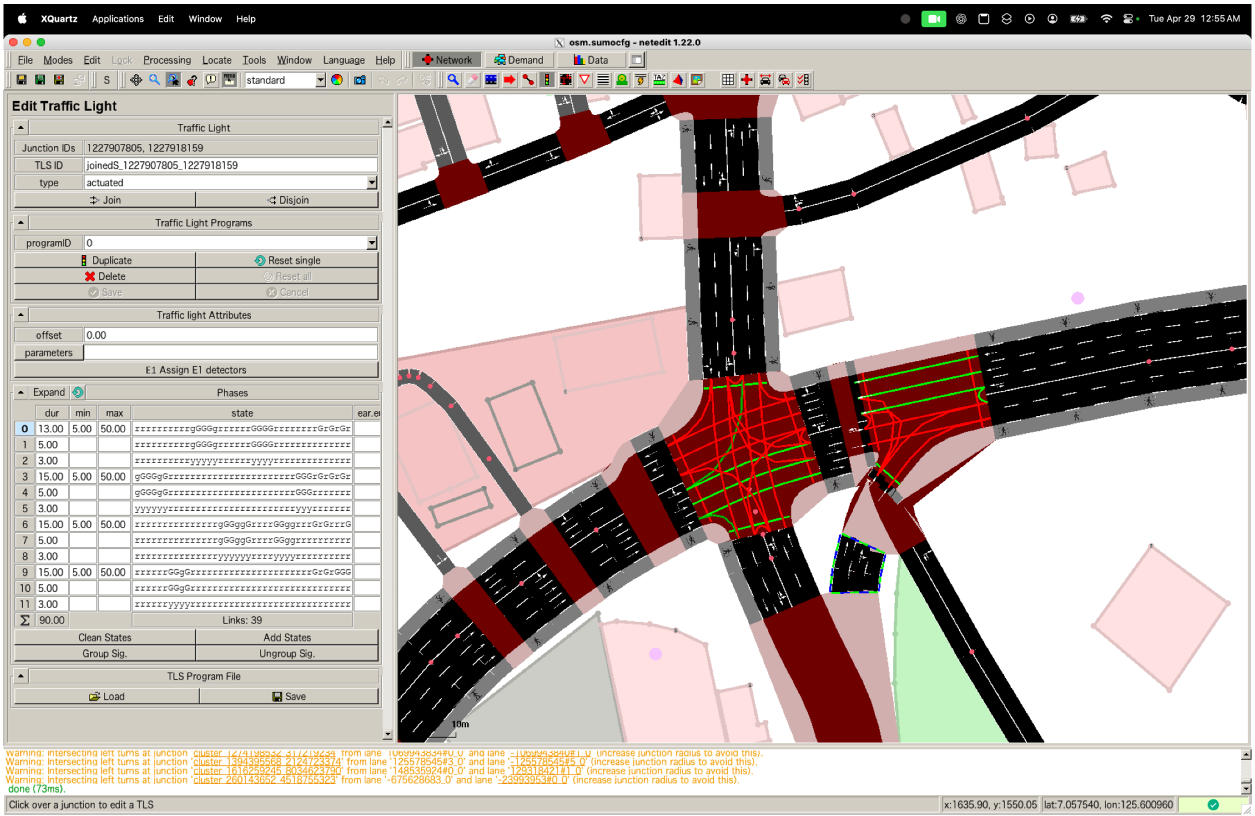


Figure 3.6.5.1 SUMO Traffic Light Phase Settings

Bibliography

Abdolrasol, M. G. M., Hussain, S. M. S., Ustun, T. S., Sarker, M. R., Hannan, M. A., Mohamed, R., Ali, J. A., Mekhilef, S., & Milad, A. (2021). Artificial Neural Networks Based Optimization Techniques: A Review. Electronics, 10(21), 2689. https://doi.org/10.3390/electronics10212689

Amin, S. (2020, May 15). Deep Q-learning (DQN). Medium. https://medium.com/@samina.amin/deep-q-learning-dqn-71c109586bae

Azfar, T., & Ke, R. (2024). Traffic co-simulation framework empowered by infrastructure camera sensing and reinforcement learning. arXiv. https://doi.org/10.48550/arXiv.2412.03925

Bie, Y., Wang, D., & Qu, X. (2024). Adaptive transit signal priority for balanced urban mobility. IEEE Transactions on Intelligent Transportation Systems, 25(1), 601–612.

Cai, C., & Wei, M. (2024). Adaptive urban traffic signal control based on enhanced deep reinforcement learning. Scientific Reports, 14, 14116.

Chen, Y., Zhang, J., & Wu, K. (2021). Dueling DQN-based adaptive traffic signal control for multi-intersection networks. Transportation Research Part C: Emerging Technologies, 125, 103046. https://doi.org/10.1016/j.trc.2021.103046

Chin, Y. K., Kow, W. Y., Khong, W. L., Tan, M. K., & Teo, K. T. K. (2024). Q-Learning Traffic Signal Optimization within Multiple Intersections Traffic Network. Modelling, Simulation & Computing Laboratory, Material & Mineral Research Unit. https://www.researchgate.net/publication/261234900_Q-Learning_Traffic_Signal_Optimization_within_Multiple_Intersections_Traffic_Network

City Transport and Traffic Management Office. (2023). CTTMO identifies 10 most traffic-congested areas in Davao City. City Government of Davao. https://www.sunstar.com.ph/article/1956681/davao/local-news/cttmo-identifies-10-most-traffic-congested-areas-in-davao-city

Clifton, J., & Laber, E. (2020). Q-Learning: Theory and Applications. Annual Review of Statistics and Its Application, 7, 47–69. https://www.annualreviews.org/doi/pdf/10.1146/annurev-statistics-031219-041220

Clemente, M. L. (2022). Building a real-world traffic micro-simulation scenario from scratch with SUMO. Center for Advanced Studies Research and Development in Sardinia. SUMO User Conference Proceedings, 3. https://doi.org/10.52825/scp.v3i.109

Elhanashi, A. E., Dini, P., Saponara, S., & Zheng, Q. (2024). TeleStroke: Real-time stroke detection with federated learning and YOLOv8 on edge devices. Journal of Real-Time Image Processing, 21(121). https://link.springer.com/content/pdf/10.1007/s11554-024-01500-1.pdf

Hassan, A., Jiang, H., Khani, A., & Kim, S. (2023). Deep reinforcement learning for adaptive traffic signal control: Review and application. Transportation Research Part C: Emerging Technologies, 147, 104090. https://arxiv.org/pdf/1911.06294

Huh, D., & Mohapatra, P. (2024). Multi-agent reinforcement learning: A comprehensive survey. arXiv. https://arxiv.org/abs/2312.10256

Lee, S.-Y., Byeon, S., Kim, H., Jin, H., & Lee, S. (2020). Deep learning-based phase prediction of high-entropy alloys: Optimization, generation, and explanation. Materials and Design, 197. https://www.researchgate.net/publication/345137756_Deep_learning-based_phase_prediction_of_high-entropy_alloys_Optimization_generation_and_explanation

Li, X., Chen, T., & Wang, Y. (2022). Temporal pattern recognition in traffic signal control using LSTM-enhanced reinforcement learning. Transportation Research Part C: Emerging Technologies, 142, 103756.

Lin, C. (2022). Intelligent Traffic-Monitoring System Based on YOLO and Convolutional Fuzzy Neural Networks. https://www.researchgate.net/publication/358325100_Intelligent_Traffic-Monitoring_System_Based_on_YOLO_and_Convolutional_Fuzzy_Neural_Networks

Liu, B., Zhao, D., & Li, Y. (2023). Dueling DQN with double state representation for multi-modal traffic signal control. IEEE Transactions on Vehicular Technology, 72(5), 6084–6097. https://www.sciencedirect.com/science/article/pii/S2772586325000048

Liu, Y., & Park, B. (2021). Graph neural network approach for coordinated traffic signal control. Transportation Research Part C: Emerging Technologies, 129, 103275.

Merenda, M., Porcaro, C., & Iero, D. (2020). Edge Machine Learning for AI-Enabled IoT Devices: A Review. Sensors, 20, 2533. https://www.researchgate.net/publication/345137756_Deep_learning-based_phase_prediction_of_high-entropy_alloys_Optimization_generation_and_explanation

Milan, A. (2023, August 23). Davao faces trade-off between traffic, infra dev't. SunStar Davao. https://www.sunstar.com.ph/davao/business/davao-faces-trade-off-between-traffic-infra-devt

Miletić, M., Ivanjko, E., Gregurić, M., & Kušić, K. (2022). A review of reinforcement learning applications in adaptive traffic signal control. IET Intelligent Transport Systems, 16(10), 1269–1285.
https://www.researchgate.net/publication/361207511_A_review_of_reinforcement_learning_applications_in_adaptive_traffic_signal_control

Morales, J., Singh, A., & Thompson, R. (2023). Multi-objective reinforcement learning for balanced urban traffic management. Transportation Research Part C: Emerging Technologies, 148, 104120.

Moreno-Malo, J., Posadas-Yagüe, J.-L., Cano, J. C., Calafate, C. T., Conejero, J. A., & Poza-Lujan, J.-L. (2024). Improving traffic light systems using Deep Q-networks. Expert Systems with Applications, 252(Part B), 124178. https://doi.org/10.1016/j.eswa.2024.124178

Mu, L., Zheng, F., Tao, R., Zhang, Q., & Kapelan, Z. (2020). Hourly and Daily Urban Water Demand Predictions Using a Long Short-Term Memory Based Model. Journal of Water Resources Planning and Management, 146(12), 04020083. https://www.researchgate.net/publication/344027818_Hourly_and_Daily_Urban_Water_Demand_Predictions_Using_a_Long_Short-Term_Memory_Based_Model

Pan, T. (2023). Traffic light control with reinforcement learning. arXiv. https://doi.org/10.48550/arXiv.2308.14295

Panagiotou, D. K., & Dounis, A. I. (2022). Comparison of Hospital Building's Energy Consumption Prediction Using Artificial Neural Networks, ANFIS, and LSTM Network. Energies, 15(17), 6453. https://doi.org/10.3390/en15176453

Patel, R., & Nguyen, H. (2022). Comparative analysis of reward functions in reinforcement learning for multi-modal traffic signal control. Transportation Research Part C: Emerging Technologies, 135, 103538.

Pham, T. N., Vu, D., Bui, D. T., & Vo, A. (2022). Investigating Q-Learning and DQN variants for Traffic Signal Control problem [GitHub repository]. GitHub. https://github.com/thoth-ml/DRL_Traffic-Signal-Control

Preetha, S., Devi, R. S., & Kumar, M. A. (2025). A comprehensive approach to traffic monitoring using OpenCV: From image preprocessing to vehicle identification. International Journal of Computer Vision Applications, 17(2), 214–229. https://ijsret.com/wp-content/uploads/2025/01/IJSRET_V11_issue1_166.pdf

Rahman Swapno, A., Ahmed, S., & Khan, M. (2024). Impact of isolated transit signal priority on network performance. Journal of Transportation Engineering, 150(2), 04023089.

Sarker, I. H. (2022). AI-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Computer Science, 3(1), 158. https://doi.org/10.1007/s42979-022-01043-x

Scatto, M. (2024). Multi-objective deep reinforcement learning for fair urban mobility. Transportation Research Part C: Emerging Technologies, 150, 104150.

Sewak, M. (2019). Deep Q Network (DQN), Double DQN, and Dueling DQN: A step towards general artificial intelligence