symmetry

MDPI

Article

Integration of Decentralized Graph-Based Multi-Agent Reinforcement Learning with Digital Twin for Traffic Signal Optimization

Vijayalakshmi K. Kumarasamy [1], Abhilasha Jairam Saroj [2], Yu Liang [1,*], Dalei Wu [1,*], Michael P. Hunter [3], Angshuman Guin [3] and Mina Sartipi [1]

1 Department of Computer Science and Engineering, University of Tennessee at Chattanooga, Chattanooga, TN 37403, USA; lry466@mocs.utc.edu (V.K.K.); mina-sartipi@utc.edu (M.S.)
2 Buildings and Transportation Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA; sarojaj@ornl.gov
3 School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA; michael.hunter@ce.gatech.edu (M.P.H.); angshuman.guin@ce.gatech.edu (A.G.)
* Correspondence: yu-liang@utc.edu (Y.L.); dalei-wu@utc.edu (D.W.)

Abstract: Machine learning (ML) methods, particularly Reinforcement Learning (RL), have gained widespread attention for optimizing traffic signal control in intelligent transportation systems. However, existing ML approaches often exhibit limitations in scalability and adaptability, particularly within large traffic networks. This paper introduces an innovative solution by integrating decentralized graph-based multi-agent reinforcement learning (DGMARL) with a Digital Twin to enhance traffic signal optimization, targeting the reduction of traffic congestion and network-wide fuel consumption associated with vehicle stops and stop delays. In this approach, DGMARL agents are employed to learn traffic state patterns and make informed decisions regarding traffic signal control. The integration with a Digital Twin module further facilitates this process by simulating and replicating the real-time asymmetric traffic behaviors of a complex traffic network. The evaluation of this proposed methodology utilized PTV-Vissim, a traffic simulation software, which also serves as the simulation engine for the Digital Twin. The study focused on the Martin Luther King (MLK) Smart Corridor in Chattanooga, Tennessee, USA, by considering symmetric and asymmetric road layouts and traffic conditions. Comparative analysis against an actuated signal control baseline approach revealed significant improvements. Experiment results demonstrate a remarkable 55.38% reduction in Eco_PI, a developed performance measure capturing the cumulative impact of stops and penalized stop delays on fuel consumption, over a 24 h scenario. In a PM-peak-hour scenario, the average reduction in Eco_PI reached 38.94%, indicating the substantial improvement achieved in optimizing traffic flow and reducing fuel consumption during high-demand periods. These findings underscore the effectiveness of the integrated DGMARL and Digital Twin approach in optimizing traffic signals, contributing to a more sustainable and efficient traffic management system.

Keywords: multi-agent reinforcement learning; Digital Twin; graph neural network; intelligent transportation systems; traffic signal optimization; fuel consumption; traffic congestion; actuated signal control; asymmetric traffic conditions

check for updates

Citation: Kumarasamy, V.K.; Saroj, A.J.; Liang, Y.; Wu, D.; Hunter, M.P.; Guin, A.; Sartipi, M. Integration of Decentralized Graph-Based Multi-Agent Reinforcement Learning with Digital Twin for Traffic Signal Optimization. *Symmetry* **2024**, *16*, 448. https://doi.org/10.3390/sym16040448

Academic Editors: Sahraoui Dhelim, Tao Zhu and Nyothiri Aung

Received: 29 February 2024
Revised: 31 March 2024
Accepted: 3 April 2024
Published: 7 April 2024

1. Introduction

Urban centers worldwide are increasingly adopting Intelligent Transportation System (ITS) technologies to transform conventional corridors into smart and data-driven transportation systems [1–5]. The deployment of smart corridors offers an opportunity to harness high-resolution and high-frequency vehicle and infrastructure data. This, coupled with advancements in machine learning, artificial intelligence, and high-performance computing, presents a promising avenue for addressing safety, mobility, and environmental challenges within transportation systems [6–10].

Efforts to optimize and enhance transportation systems are exploring innovative approaches, and one such solution under investigation is the application of Digital Twin-assisted decentralized multi-agent Reinforcement Learning (RL). This approach involves establishing a seamless connection between the Digital Twin representation of the physical system and the decentralized multi-agent Reinforcement Learning (RL) framework. The primary goal is to demonstrate the successful integration of these components and leverage the resulting application to optimize traffic signal timing.

This study presents an extended version of the work [11], focusing on a real-world case study to demonstrate the practical application of the integrated Digital Twin and decentralized graph-based multi-agent reinforcement learning. Our specific objective is to optimize traffic signal timing, aiming to reduce a performance measure known as Eco_PI, which comprehensively assesses the environmental and efficiency aspects of traffic management by capturing the impact of stops on fuel consumption and delay. The extended methodology introduces red clearance and max green constraints, enhancing simulation accuracy and pedestrian safety while optimizing traffic flow. Additionally, expanded experimental analysis provides insights into traffic patterns, congestion, and overall performance, resulting in notable enhancements in traffic flow optimization and congestion mitigation, particularly during peak hours.

The integration of Digital Twin technology and decentralized graph-based multi-agent reinforcement learning holds significant potential for addressing the complex dynamics of urban traffic systems. By optimizing traffic signal timing through this integrated approach, this study aims to contribute to the broader goal of creating more sustainable, efficient, and intelligent transportation networks. Further insights into the Eco_PI performance measure, including relevant references, can be found in [12–15].

2. Related Work

In the realm of urban traffic management, the optimization of traffic signal control has become increasingly imperative due to the challenges posed by growing populations and urbanization. Traditional traffic control methods often struggle to adapt to dynamic traffic scenarios and efficiently coordinate diverse agents, including vehicles and pedestrians. Recent advancements in Artificial Intelligence (AI) offer promising solutions to address these challenges, with notable applications in domains such as healthcare [16–19], transportation, etc. Deep learning frameworks have demonstrated effectiveness in tasks such as vehicle tracking, visual speed estimation [20], and traffic estimation [21]. In particular, Multi-Agent Reinforcement Learning (MARL) approaches [22,23] show particular potential for intelligently optimizing traffic signals.

Multi-Agent Reinforcement Learning (MARL) involves collaborative decision-making among agents, each relying on local observations and interactions within the environment. Decentralized graph-based MARL models, like Multi-Agent Advantage Actor-Critic (MA2C) [22], have emerged as a breakthrough, effectively distributing control across local agents while coordinating for efficient traffic signal management. MA2C addresses scalability concerns in large-scale networks, enabling independent learning for agents and facilitating quicker policy convergence. Coordinated actions enhance performance, especially in cooperative or competitive scenarios. Advantage Actor-Critic (A2C) scales effectively to larger environments [24], enabling parallelized learning and adaptation to dynamic environments by continuously updating policies based on interactions with other agents. Through A2C, agents refine policies through continual exploration and exploitation of the environment, maximizing cumulative rewards for stable and efficient learning.

The study [25] underscores the critical importance of accurate simulation models in urban transportation planning and optimization, and highlights the necessity of capturing the complexity of city traffic for effective planning by utilizing real-world vehicle speed data and integrating various sources. As the urban landscape evolves, the integration of Digital Twins, inspired by Industry 4.0 principles [26–41], emerges as a transformative element in modernizing systems and processes. Digital Twins offers a promising paradigm to enhance the performance of physical systems through virtual modeling [42–45]. The marriage of Digital Twins with Multi-Agent Reinforcement Learning (MARL) for traffic signal optimization marks a paradigm shift in urban traffic management. Digital Twins serve as virtual representations of physical entities, providing real-time monitoring, analysis, and decision-making capabilities. Studies such as [46] investigate the transformative impact of Digital Twins, Internet of Things (IoT), and machine learning on data utilization, underscoring the potential of Digital Twins in enhancing real-time data utilization for enterprises. Continuously updated in real-time through sensor data and various sources, the Digital Twins provides an accurate depiction of the current and historical state of the physical entity, allowing for improved prediction, control refinement, and operational optimization [47,48]. Diverse applications of Digital Twins, including transportation, modeling techniques, and the benefits of integrating Digital Twins in system design, have been discussed in [49,50].

In contrast to traditional simulation models based on assumptions, Digital Twins rely on actual data, providing a more realistic representation of the physical entity. This attribute proves particularly beneficial in industries like Intelligent Transportation System (ITS) technologies, where the reliability and performance of complex systems are paramount [51].

Within transportation applications, Digital Twins simulations offer a realistic portrayal of real-world transportation systems. Acting as a crucial testbed, Digital Twins facilitate the development of real-time machine learning-based traffic operations applications, providing a safe and economic environment for training and testing artificial intelligence/machine learning (AI/ML) algorithms. Previous studies, exemplified by [52], utilized Digital Twins for transportation systems, leveraging real-time smart corridor data to model current traffic states and provide dynamic updates on traffic performance measures. AI/ML algorithms encompass a range of techniques, including deep learning, reinforcement learning, and other computational methods, to analyze complex transportation data, optimize traffic signal timing, predict traffic congestion, and improve overall transportation efficiency. By integrating Digital Twins with AI/ML algorithms, transportation researchers and practitioners can gain valuable insights into traffic behavior, develop more effective traffic management strategies, and enhance the performance of urban transportation systems.

Digital Twins in transportation systems offer significant advantages, including real-time monitoring, improved coordination, and enhanced traffic efficiency [53]. The integration of Digital Twins with deep learning and reinforcement learning algorithms enhances real-time adaptive, precision-centric, and predictive traffic monitoring [54]. Moreover, Digital Twins assist reinforcement learning algorithms in understanding dynamic traffic states, facilitating better real-time decisions through adaptive signal control [55].

Due to the Digital Twin's ability to behave as a real-time environment with different static and dynamic properties, it can be used to assist deep learning algorithms like training autonomous cars [56], real-time adaptive, precision-centric, predictive traffic monitoring [57], and Reinforcement Learning (RL) algorithms like edge task scheduling [58], intelligent manufacturing systems [59], and in vehicular edge computing [54]. To learn the dynamic traffic flow behavior and make better decisions in real-time through adaptive signal control [55], reinforcement learning algorithms can have better assistance through Digital Twin. The Digital Twin can use data from various transportation components, including vehicle presence time in the detector zone which refers to the elapsed time from when vehicles enter the detector zone until they leave, approaching vehicle counts, and pedestrian recall, to create a comprehensive representation of the transportation system. This enables Reinforcement Learning (RL) agents to learn traffic flow behavior and perform various actions in the digital environment. Moreover, the ability of multiple agents to interact with the same environment and coordinate with each other can lead to better decisions and improved traffic flow. The use of Digital Twin with Reinforcement Learning (RL) agents increases efficiency in decision-making and enables agents to observe their performance for future decisions.

Additionally, broadening signal optimization to encompass all directions, including traffic approaching from various cardinal directions such as east, west, north, and south bounds, and accounting for heterogeneous traffic conditions, enables the development of a more comprehensive traffic management strategy. This approach, as demonstrated in the study by Pandit et al. [60], ensures more efficient service for vehicles on side streets, alleviates congestion on main thoroughfares and secondary routes, and reduces travel times for all road users. While achieving a green wave for main streets is beneficial, addressing diverse traffic demands across all directions is crucial for effective urban traffic signal control. This inclusive approach ensures a more equitable distribution of traffic flow and enhances overall network efficiency. Moreover, the effectiveness of a traffic signal system lies in its ability to dynamically adjust the sequence of signal phases in response to changing traffic conditions [61,62], rather than adhering rigidly to a predefined or fixed sequence. Studies such as [63] dynamically generate phase schemes per cycle based on traffic asymmetry to optimize traffic signal timing. However, being bound to the cycle entails predetermined phase and cycle durations, which do not allow for dynamic adjustments in phase durations based on real-time traffic demand.

Building upon this landscape, our prior work [11] introduces a novel approach by combining Digital Twin assistance with decentralized graph-based multi-agent reinforcement learning (DGMARL) [11] to learn dynamic traffic states. DGMARL agents, strategically distributed at individual intersections, observe traffic state features from multi-directions. DGMARL model considers traffic approaching from various cardinal directions, such as east, west, north, and south bounds, necessitating the consideration of the dynamics of vehicles approaching from each of these directions when optimizing signal timing. Consequently, the model takes into account the diverse flow of traffic when optimizing signal timing. Then the agents exchange information with neighboring agents to optimize intersection signal timing along with dynamic phasing. This innovative method, known as dynamic phasing, allows for dynamic adjustment of signal phases based on real-time traffic conditions, rather than adhering to a predefined sequence. The proposed DGMARL model is designed to handle heterogeneous data, including vehicle presence time in the detector zone, approach-level vehicle count aggregates, pedestrian recall times, and current signal states from current, upstream, and downstream intersections from all directions. The integration leverages a Component Object Model (COM) interface of PTV-Vissim, a traffic simulation software, to control signal timing through Digital Twins.

The proposed model boasts several key technical features that collectively enhance its efficacy in optimizing traffic signal timing:

- Seamless Integration of Digital Twin and Decentralized Graph-based Multi-Agent Reinforcement Learning (DGMARL): The integration of Digital Twins and DGMARL allows for the dynamic optimization of traffic signal timing, leveraging real-time traffic data and simulation capabilities to improve traffic flow and reduce congestion.
- Distributed Multi-Agent Reinforcement Learning: Multi-agent reinforcement learning agents are deployed at individual intersections to observe traffic state features, including vehicle presence time in the detector zone. They exchange this information with neighboring agents to collectively determine an optimal policy for controlling traffic signals. The implementation of actions is validated against rules and constraints, such as minimum green time and pedestrian recall time, ensuring safe mobility for all users. In a coordinated multi-agent environment, the optimal policy is derived through reinforcement learning, where agents interact with the environment, learning from experiences to maximize rewards over time. Through iterative exploration, agents adjust policies to prioritize actions with higher rewards. Furthermore, agents engage in communication and coordination with neighboring agents to enhance decision-making and achieve better outcomes collectively.
- Consideration of All Directions of Traffic Demand: Unlike traditional approaches that may focus solely on specific traffic flows, the DGMARL model considers the traffic demands from all directions, ensuring a comprehensive approach to traffic signal

optimization. This inclusion allows for more efficient management of traffic flow across the entire network.

- Dynamic Phasing: The DGMARL model offers dynamic phasing, enabling flexible adjustments to signal timing sequences based on evolving real-time traffic conditions. This flexibility enhances adaptability and responsiveness to changing traffic patterns and congestion levels.
- Handling Heterogeneous Data: The DGMARL model coordinates diverse data types, including vehicle presence time, count aggregates, pedestrian recall times, and current signal states, to optimize traffic flow efficiently. It achieves this through several mechanisms: feature engineering for data preprocessing, message passing and communication among agents, neural network processing to learn complex patterns, reward calculation based on coordinated data inputs, and policy optimization for dynamic signal timing decisions. By integrating these data types into a unified framework, the model performs comprehensive analysis and decision-making, enabling effective traffic management across the transportation network.
- Utilization of Component Object Model (COM) Interface: Leveraging the COM interface of PTV-Vissim, the proposed model can seamlessly take actions and control signal timing through the Digital Twins. This integration streamlines the implementation of optimized signal timing strategies in real-world traffic scenarios.

In this extended version, introduced crucial enhancements to the methodology. Incorporated a red clearance constraint to ensure adequate time for vehicles crossing intersections, thereby enhancing the accuracy of the simulations. Additionally, introduced max green constraints to prioritize pedestrian safety while optimizing traffic flow. These constraints ensure that when a pedestrian recall is enabled in the current green phase, the algorithm switches to another phase with the highest traffic demand, thereby improving the realism and effectiveness of the traffic signal control algorithm.

Furthermore, the experimental analysis expanded to include additional results focusing on vehicles passing through intersections during green signal phases. This comprehensive analysis provides valuable insights into traffic patterns and performance, enriching our understanding of signal optimization effectiveness, traffic congestion, signal efficiency, and overall transportation performance.

These improvements have yielded significant enhancements in traffic flow optimization and congestion mitigation. Notably, both PM-peak hour and 24 h scenarios have experienced notable increases in performance, demonstrating the effectiveness of the extended methodology in addressing complex traffic dynamics and optimizing traffic signal timing for sustainable urban mobility.

3. Digital Twin System for Traffic Network

3.1. Physical Environment and Digital Twin

3.1.1. Digital Twin Architecture

Smart corridor Digital Twins are typically driven using real-time and historic vehicle and infrastructure data from the corridor [52,55,64]. In this study, the Digital Twin is developed using vehicle real-time and historic volume count, turn count, and Signal Phasing and Timing (SPaT) data available from approximately 2.1 miles of Martin Luther King (MLK) Smart Corridor, Chattanooga, TN, USA, consisting of 11 signalized intersections. A smart corridor Digital Twins model architecture typically includes four key components as shown in Figure 1:

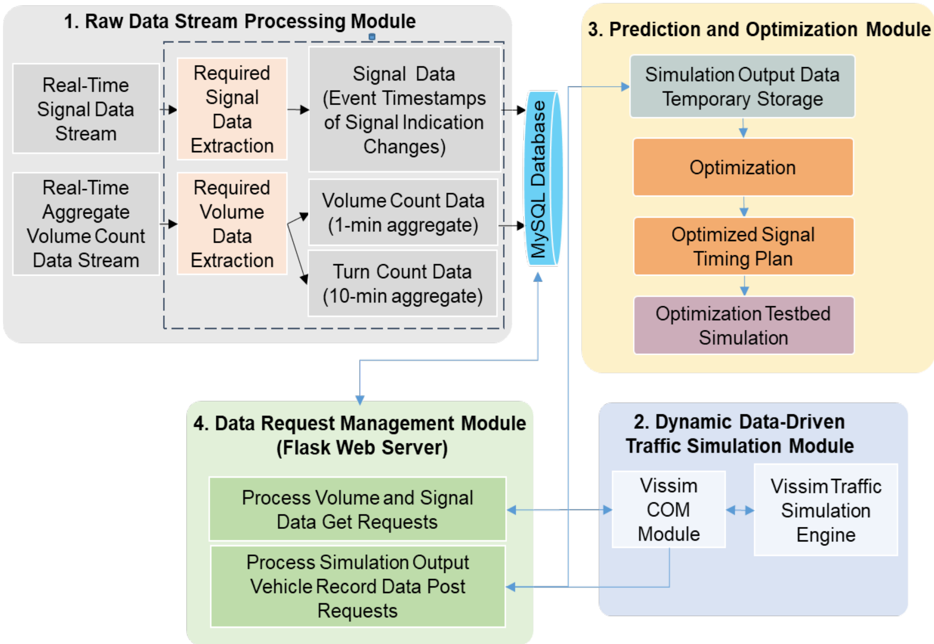

Figure 1. Digital Twin Architecture.

Module 1: Raw Data Stream Processing Module—includes processing of raw data to parse, format, and store the data in a database. From the physical MLK Smart Corridor, using Python 3.8 scripts, 1 min aggregate by turn types (left, through, and right turn vehicle counts) for each approach (Eastbound, Westbound, Northbound, and Southbound) of all the intersections are calculated and stored in a MySQL relational database, to be used in Module 2 dynamically. Furthermore, 10 Hz Signal Phase and Timing (SPaT) data are obtained from the signal controllers in the corridor. SPaT data processing involved two key steps. First, the raw data are parsed to only filter the phase indications corresponding to each phase for each intersection. Next, the high-frequency raw SPaT data from the field is processed to only filter out the records on changes in signal phase events. While the SPaT data are received at high frequency, only the change in phase event data is required. Hence, only that is stored in the database. Turn count aggregate and SPaT data are stored in separate tables in the database that are queried to drive the simulation in Module 2.

Module 2: Dynamic Data-Driven Traffic Simulation Module—includes PTV-Vissim microscopic traffic simulation model of the Smart Corridor, dynamically driven using volume, turn movement ratios, and signal indications data (from Module 1). In this implementation intersection approach level, 1-min aggregate volume counts, 10-min aggregate turn counts data, and signal timing are dynamically driven using PTV-Vissim's Component Object Model (COM) module. Using COM the signal indications can be driven using external SPaT (Signal Phasing and Timing) data or PTV-Vissim's internal Ring Barrier Controller (RBC) module.

Module 3: Prediction and Optimization Module or Simulation Testbed Application Module–consists of tools and algorithms to process simulation outputs based on the requirements of the application. This module contains processes or algorithms that are driven using outputs from the Digital Twins simulation. In this study, the outputs such as vehicle presence time in the detector zone, each direction approach level vehicle count aggregates, vehicle velocity, and current signal state, etc are generated from the PTV-Vissim simulation model in the Dynamic Data Driven Traffic Simulation Module are used as inputs for prediction and optimization for the signal timing plan.

Module 4: Real-Time Data Broker Module—handles real-time dynamic data transactions between modules. This module consists of a Flask-based web service to handle data transactions/communication between the other three modules.

3.1.2. Muti-Tier Incremental Approach for Digital Twin Development

Smart corridor Digital Twins development requires integration and synchronization of multiple components within the Digital Twins architecture described in the previous section. This makes Digital Twins development a time-consuming process susceptible to coding, integration, implementation, data processing, and other errors. To tackle this, in this study a three-tier incremental approach is used that allows a parallel workflow. The Digital Twins development process is broken into three tiers with increasing communication and infrastructure integration complexity. Such an approach enables training and testing of Machine Learning/Reinforcement learning-based (ML/RL) applications early on, on the initial tiers, thus, reducing the wait time required for the development of a fully operational real-time Digital Twins. Such an approach enables the training and testing of ML/RL-based applications such as signal timing optimization developed in this paper, on initial tiers rather than waiting for the completion of the full real-time Digital Twins, thereby expediting the overall study timeline. A detailed description of the incremental approach for Digital Twin development, as shown in Figure 2 is found in [65].

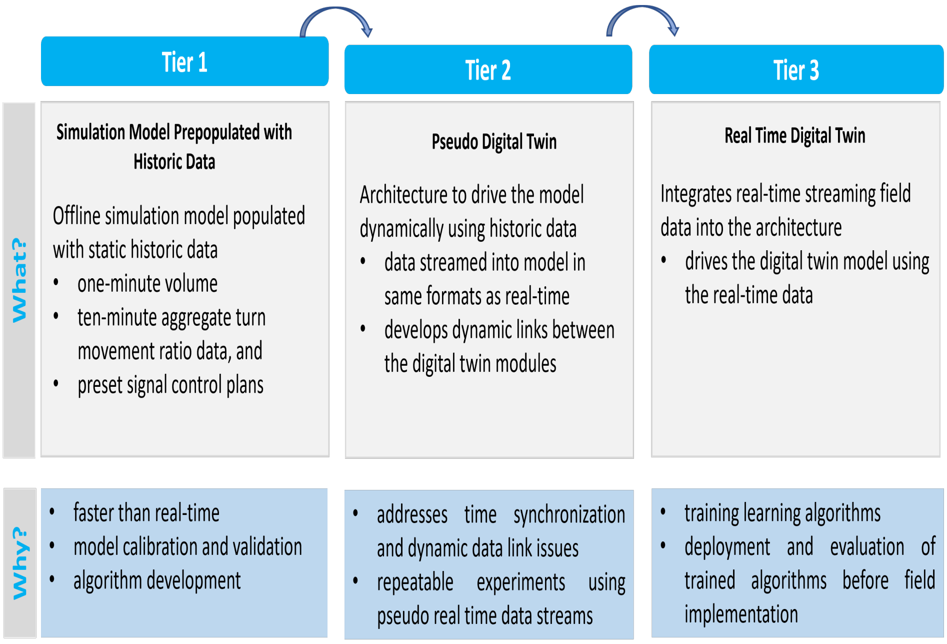

Figure 2. Three-tier Incremental Approach to Digital Twin Development for Application Testing.

The three-tier incremental approach includes the following simulation model versions:

Tier 1—Prepopulated model: traditional simulation model prepopulated with archived data. This version includes automation of raw data extraction and ingestion of extracted data by the PTV-Vissim model. Data extraction and ingestion are partially automated, enabling the ability to efficiently test developed algorithms under different conditions on different days (e.g., weekday vs. weekend, growth scenarios, etc.). Automation of data handling in this level is critical to the overall usability and effectiveness of this version of the model in training and testing the DGMARL model.

Tier 2—Pseudo Digital Twin: simulation model driven dynamically using archived data. In this tier, the data is dynamically fed into the simulation as opposed to prepopulating before the simulation runs in Tier 1. Several parts of the efforts of Tier 1 platform development, such as data investigation, data extraction, and automation of ingestion of input data into simulation are used in Tier 2 platform development. However, a significant advance in Tier 2 is the development of the dynamic links between the modules shown in Figure 1. Further, in this tier the signal indications are controlled using field received SPaT messages, not the internal Vissim Ring Barrier Controller (RBC). Thus, the implemented signal phase times will match the field directly, rather than relying on the accuracy of the simulation signal control emulator. This platform thus provides a test bed to develop the interface that integrates the DGMARL optimization algorithm with data-driven Digital Twins simulation.

Tier 3—Real-time Digital Twin: online simulation model driven dynamically using real-time field data. In this tier, the simulation is driven using real-time data. The Tier 2 platform is modified and updated to stream real-time data. This platform will be used to develop the interface between the physical system represented by the Digital Twins and the optimization development algorithm.

In this study, the interface between the Reinforcement Learning (RL) optimization algorithm and the physical system is initially developed using the Tier-1 platform. The developed Reinforcement Learning (RL) algorithm in the future will also be integrated with the Tier-2 and Tier-3 platforms to test and further improve the algorithm.

4. Digital Twin and Reinforcement Learning (RL)

The seamless integration of the Digital Twins with the Decentralized Graph-based Multi-Agent Reinforcement Learning (DGMARL) model, depicted in Figure 3, constitutes a novel approach for training agents associated with each intersection. This integration enables the agents to learn from the Digital Twins and make informed decisions to optimize signal timings based on real-time observations of the traffic state from multi-direction.

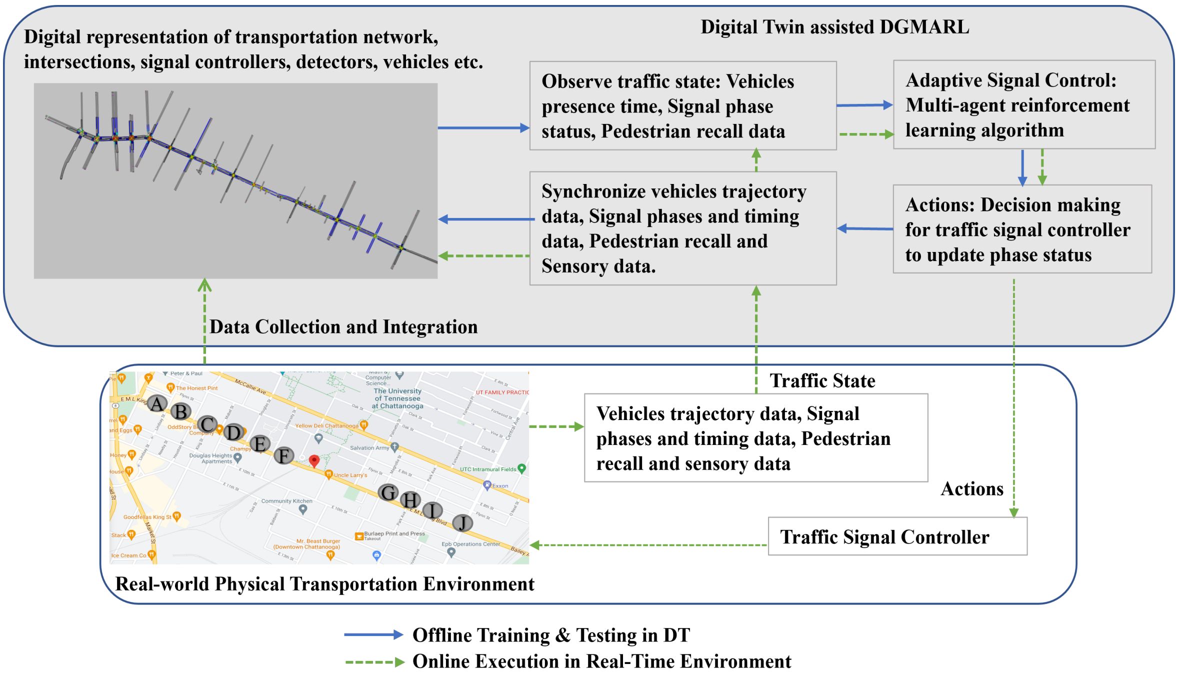

Figure 3. Digital Twin assisted DGMARL Learning.

The integration process between the Digital Twins and the decentralized multi-agent reinforcement learning algorithm is elaborated as follows:

- The decentralized multi-agent reinforcement learning algorithm leverages inputs such as vehicle presence time in the detector zone, current phase state, pedestrian recall time, etc., obtained from the Digital Twins. It then makes decisions, determining whether to maintain the current signal phase or switch to a phase with anticipated high traffic demand following a dynamic phasing approach. This decision is based on the current state of the intersection and the desired objective, which is to minimize the Eco_PI measure.
- The decision made by the decentralized multi-agent reinforcement learning algorithm is fed back to the Digital Twins, prompting an update to its simulation based on the decision. The updated simulation is subsequently utilized to provide new inputs to

the decentralized multi-agent reinforcement learning algorithm. This iterative process continues until the desired optimization objective is achieved.

This integration of the Digital Twins and DGMARL offers a compelling alternative to the traditional method of field training the DGMARL model for signal timing optimization. By learning from the Digital Twins, the model undergoes efficient and safe training and testing, avoiding the challenges associated with tedious field training. The Digital Twins data accumulation capability facilitates efficient visualization and in-depth analysis of the traffic state. The outcome is a validated and reliable output and the offline training defines the DGMARL neural network before deployment, ensuring a robust and optimized system.

5. Implementation of Intelligent Agents to Optimize the Global Transportation

Motivations for AI-Enabled Intelligent Agents in Transportation Networks: AI-enabled intelligent agents represent a transformative force in enhancing transportation networks, offering a myriad of benefits that span efficiency, reliability, safety, and societal well-being. The motivations driving the integration of these agents are multifaceted and include the following.

- Enhanced Efficiency, Reliability, and Safety: Intelligent agents leverage AI algorithms to analyze transportation data [66], enabling them to make optimized decisions. This results in increased efficiency, reliability, and safety within transportation networks [67,68].
- Positive Impacts on Quality of Life, Environment, and Economic Growth: The deployment of intelligent agents has a direct positive impact on people's quality of life by reducing congestion and improving travel experiences. Environmental benefits are realized through eco-friendly transportation practices promoted by optimized traffic flow. Economic growth is fostered as efficient transportation networks contribute to smoother logistics and infrastructure support
- Real-Time Traffic Optimization: Intelligent agents actively monitor real-time traffic conditions, offering dynamic recommendations to drivers. This includes suggesting alternate routes to avoid congestion, adjusting traffic signals for improved flow, and predicting maintenance needs [69].
- Optimized Resource Allocation and Safety Monitoring: Resource allocation is optimized for emergency vehicles, buses, trains, and other vehicles based on real-time demands. Safety is paramount, with intelligent agents detecting potential problems early [70–72]. This proactive approach contributes to a safer transportation environment.

Graph Representation of the Transportation Network: The utilization of graph representation, coupled with intelligent agents, provides a powerful framework for comprehensive situational awareness within transportation networks. This approach leverages graph theory to model the network structure and facilitates advanced decision-making capabilities:

- Comprehensive Situational Awareness: Processing and analyzing the traffic data is highly computationally costly and graph framework provides highly scalable strategies. Graph representation, employing nodes for intersections and edges for routes, provides a holistic view of the entire transportation network. Intelligent agents utilize this graph to track vehicle trajectory [73,74], predict congestion, optimize traffic states, and enhance overall situational awareness by efficiently monitoring the entire network [75–77]. This approach enables precise monitoring and analysis of traffic states within the transportation network, facilitating comprehensive situational awareness. By representing the transportation network as a graph, agents can analyze connectivity between intersections, assess traffic flow patterns, and identify potential bottlenecks or congestion points. This graphical representation empowers agents to make informed decisions regarding traffic signal control, route planning, and overall network management, leading to improved situational awareness and enhanced traffic management strategies.

- Integration of Machine Learning Algorithms: Reinforcement learning algorithms are seamlessly integrated, enabling agents to learn traffic patterns from both historical and real-time data. This integration enhances their adaptability to dynamic traffic conditions. By continuously observing and analyzing traffic data, the algorithm can respond to different traffic conditions such as anomalies, allowing agents to adapt and optimize traffic signal timing dynamically. This adaptability ensures that traffic management strategies remain effective in response to changing traffic patterns and unforeseen events, ultimately leading to improved traffic flow and congestion mitigation.
- Traffic Signal Control Optimization: Intelligent agents interact with local signal controllers, leveraging graph-based insights to optimize traffic signal timings by analyzing sensor data from key locations and communicating with local signal controllers [78,79]. This dynamic control mechanism helps in avoiding congestion and improving traffic flow.

Scalability: Scalability is a pivotal consideration in optimizing traffic signal timings, especially as transportation networks expand in size and complexity.

- Challenges in Single-Agent Architecture: Centralized agents face limitations in processing, communication, and latency as the transportation network grows. Effectiveness in smaller networks may not translate to larger networks due to increasing demands.
- Multi-Agent Architecture for Scalability: Multiple agents, operating independently, optimize traffic signal timings at different intersections within the network. Asynchronous communication protocols, including message passing and attention mechanisms, reduce communication overhead [80]. Distribution of workload and efficient utilization of local data enhance scalability for larger volumes of data and intersections [22,23,81].

Hence integration of AI-enabled intelligent agents, graph representation, and scalable multi-agent architectures presents a holistic approach to transforming transportation networks. By distributing the workload and utilizing local data more efficiently, this approach can handle larger volumes of data and more intersections.

5.1. Graph Neural Network Formulation of Traffic Network

The proposed approach adopts a graph neural network (GNN)-oriented formulation to model the traffic environment as a network, providing a comprehensive representation of the traffic network structure and dynamics. This section discusses the key components of the formulation, including the graph representation, the infrastructure of the Digital Twin-assisted DGMARL system, and the spatio-temporal multi-agent reinforcement learning process.

Multi-agent reinforcement learning is employed to disclose the spatial and temporal patterns of traffic. In this process, agents interact with their environment over both space and time, learning from their actions and experiences to optimize their behavior. Spatial information encompasses various configurations related to signal controllers, pedestrian walk configurations, and intersection-specific timing parameters such as minimum and maximum green times, red clearance times, and yellow times. This information delineates the physical layout of the environment, including the arrangement of intersections and road networks. Meanwhile, temporal information pertains to changes occurring over time, such as fluctuations in traffic flow and congestion levels. By incorporating both spatial and temporal dimensions, this approach enables agents to effectively learn and adapt to dynamic traffic conditions, enhancing traffic management and optimization strategies.

The notations used in this paper are given in Tables 1 and 2.

Table 1. Notation used for the Transportation network.

Notation	Description
$\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$	Bi-directional graph specified by set of intersections/agents (vertices) and set of links (edges)
i, j	ID of intersections/agents/nodes
l	ID of links/edges
$e_{ij} \in \mathcal{E}$	Link connecting intersections i and j
$\mathcal{N}_i$	Set of incoming neighbors of intersection i
δ_i	Traffic flow status intersection i
ϕ_i	Signal phases of intersection i

Table 2. Notations used for Decentralized Graph-based Multi-Agent Reinforcement Learning.

Notation	Description
$\mathcal{S}$, $\mathcal{A}$, p, and r	State space, Action sace, Agent's Policy, and Reward
$s_{i,t} \in S$	State of intersection i at time t
$a_{i,t} \in A$	Action taken by agent i at time t
$r_{i,t} = Eco\_PI$	Eco_PI as reward r at time t
d_l	Average stop delay occurred in the link l
N_l	Number of stops occurred in the link l
K_l	Average stop penalty calculated for link l
V^π	Critic-State Value
Q^π	Actor-Action Value
$s_{i,t} \sim p$	Initial policy p distribution of intersection i at time t
θ_i	Policy network parameters of intersection i
ω_i	Value network parameters of intersection i

Graph Representation of Traffic Network

The traffic environment is represented as a bi-directional graph denoted as $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ represents a set of intersections modeled as agents, and $\mathcal{E}$ represents a set of roads considered links, where $e_{i,j} \in \mathcal{E}$ is a link that connects intersections i and j. Each intersection i has static features such as approach links, signal controllers, signal phases, detectors, the number of lanes, uncontrolled approaching links, and neighboring intersections $\mathcal{N}i \subset \mathcal{V}$. The signal controller at each intersection is associated with signal phases ϕ_i, each having static features like list of signal lights, minimum green serving time, yellow time, red clearance time, pedestrian recall time, and priority phase.

5.2. Infrastructure of DGMARL

Figure 4 shows the architecture of a Digital Twins-assisted multi-agent reinforcement learning empowered traffic environment.

Each intersection of the traffic network was designed as a local agent. The multi-intersection traffic network signal timing optimization problem is addressed with decentralized multi-agent reinforcement learning. The traffic signal control problem is formulated as a Markov Decision Process (MDP): $(\mathcal{S}, \mathcal{A}, p, r)$ where $\mathcal{S}$ denotes the state space, $\mathcal{A}$ represents the action space, and r is the reward that measures the benefit brought about by a specific action. The objective is to learn the optimal policy p that generates the best action for the next step and maximizes the subsequent accumulative discounted rewards produced by the action.

To enhance learning efficiency and inform optimal actions based on approaching traffic, neighboring agents share local observations through communication graphs and knowledge sharing through message passing, enabling agents to communicate and coordinate. This function aggregates the current agent's traffic state and recent policy, along with that of its neighbors. The aggregated state undergoes processing through linear transformation neural network layers and an LSTM (Long Short-Term Memory) layer, adept at capturing long-term dependencies in sequential data. Subsequently, the ReLU (Rectified Linear

Unit) activation function is applied to introduce non-linearity to the network, followed by feeding the resulting output into the actor neural network for decision-making. Linear transformation, a mathematical operation, manipulates input data through linear combinations, with linear transformation layers in neural networks performing these operations to produce output features.

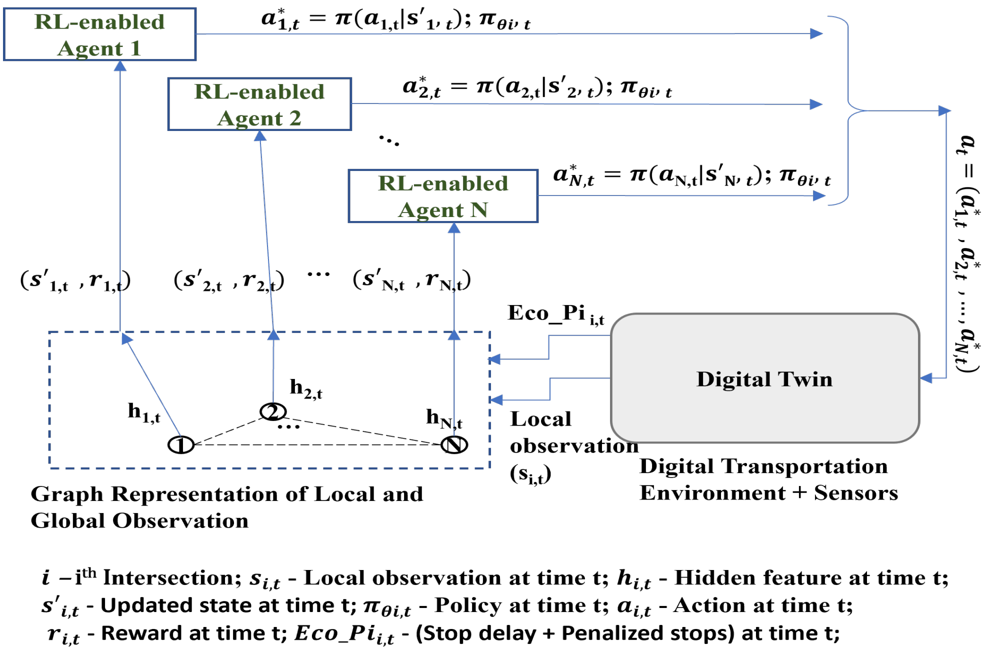

Figure 4. Architecture of Digital Twin assisted DGMARL Learning.

5.2.1. State Space

Each intersection state is derived from the heterogeneous observations of traffic states and traffic signal phase state from multi-direction, and is further refined using a spatial-temporal graph neural network [82]. The state of the global traffic network at time t for the traffic network is defined as

$$S_t = \{s_{i,t}\}_{i=1}^{|\mathcal{V}|}. \quad (1)$$

where $\{s_{i,t}\}$ is the state of the intersection i at time t which is the heterogeneous observation of traffic states and traffic signal phase state from multi-direction.

The state of agent i at time t, $\{s_{i,t}\}$, is formed by the observed traffic flow status Υ^{TF} and the traffic signal phase status Ψ^{TS}. Namely,

$$s_{i,t} = \langle \Upsilon^{TF}_{i,t}, \Psi^{TS}_{i,t} \rangle \quad (2)$$

where

$$\Upsilon^{TF}_{i,t} = \langle \langle \delta^{PT}_{l,i,t}, \ \delta^{\overline{W}}_{l,i,t}, \ \delta^{\overline{D}}_{l,i,t}, \ \overline{V}_{l,i,t} \rangle_{l=1}^{K_{\phi,i}} \rangle_{\phi=1}^{F_i} \quad (3)$$

and

$$\Psi^{TS}_{i,t} = \langle \phi^S_{i,t}, \ \phi^D_{i,t}, \ \phi^{PS}_{i,t}, \ \phi^{MinG}_{i,t}, \ \phi^{MaxG}_{i,t} \rangle. \quad (4)$$

In Equation (3), F_i is the number of phases in the intersection i, and $K_{\phi,i}$ is the number of approaching links at phase ϕ_i; the observations of each approaching link l in each phase ϕ_i include vehicle presence time in the detector zone $\delta^{PT}_{l,i,t}$, average waiting time $\delta^{\overline{W}}_{l,i,t}$, average delay $\delta^{\overline{D}}_{l,i,t}$, and vehicles average speed $\overline{V}_{l,i,t}$. In Equation (4), the variables $\phi^S_{i,t}$, $\phi^D_{i,t}$, and $\phi^{PS}_{i,t}$ correspond to the current instant phase status, current phase duration, and pedestrian serving status, respectively. The variable $\phi^{MinG}_{i,t}$ indicates whether the minimum green time has been fulfilled in the current phase, while $\phi^{MaxG}_{i,t}$ indicates whether the current phase duration has reached the maximum green serving time. The system monitors the maximum green serving time when pedestrian recall is enabled for the current phase ϕ_i. In such cases, vehicles are served until the maximum green time is reached unless the agent decides to switch to another phase based on the ongoing traffic condition, after which pedestrians are served.

$$\phi_{i,t}^{MinG} = \begin{cases} 1, & \text{if } \phi_{i,t}^{D} \geq \max(\phi_{i,t}^{MinGT}, \phi_{i,t}^{PST}) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\phi_{i,t}^{MinGT}$ represents the minimum green serving time, $\phi_{i,t}^{D}$ represents the current phase duration, and $\phi_{i,t}^{PST}$ represent pedestrian serving time which is the sum of walk and flashing don't walk time.

$$\phi_{i,t}^{MaxG} = \begin{cases} 1, & \text{if } \phi_{i,t}^{PR} \ \& \ (\phi_{i,t}^{D} \geq \phi_{i,t}^{MaxGT}) \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $\phi_{i,t}^{PR}$ represents the pedestrian recall flag, $\phi_{i,t}^{D}$ represents the current phase duration, and $\phi_{i,t}^{MaxGT}$ represents the maximum green serving time.

5.2.2. Action Space

The initial action $a_{i,t}$ at intersection i is evaluated against physical constraints, namely the minimum green serving time $\phi_{i,t}^{MinGT}$, the maximum green serving time $\phi_{i,t}^{MaxGT}$ for the current active phase ϕ_i at intersection i, and the pedestrian serving time $\phi_{i,t}^{PS}$ based on the current phase duration $\phi_{i,t}^{D}$. This evaluation ensures the safety of all users within the transportation network. Subsequently, the final decision $a'_{i,t}$ is incorporated into the intersection's signal timing plan.

The decision-making process is expressed by the Equation (7), where $a'_{i,t}$ is determined as follows

$$a'_{i,t} = \begin{cases} a_{i,t}, & \text{if } \phi_{i,t}^{MinG} \\ 1, & \text{if } (\phi_{i,t}^{MaxG} \ \& \ \phi_{i,t}^{PR}) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $\phi_{i,t}^{MinG}$ the flag determines whether minimum green is served for the current phase or not as defined in the Equation (5). In this context, $a'_{i,t} = 0$ signifies that the agent refrains from taking action, while $a'_{i,t} = 1$ implies that the agent will transition the current phase signal to yellow. If the current phase is configured with pedestrian recall $\phi_{i,t}^{PR}$, and if the current phase duration is greater than or equal to maximum green constraint $\phi_{i,t}^{MaxG}$ then the action $a'_{i,t} = 1$ is enforced to switch to another phase. After this transition, the current phase will be ready to serve pedestrian walk plus flashing don't walk time, following the serving of yellow and red clearance timing. Subsequently, the phase $\phi_{i,t}$ with the highest traffic demand will be switched to green. The next phase $\phi_{i,t}$ is determined by the following equation,

$$\phi_{i,t} = \arg\max_{\phi} \langle \Upsilon_{\phi}^{TF} \rangle_{\phi=1}^{F_i} \tag{8}$$

The proposed model follows a dynamic phasing approach, prioritizing the phase with the highest traffic demand to minimize vehicle waiting delays.

Dynamic phasing, allowing dynamic adjustment of signal phases based on real-time traffic conditions, significantly enhances adaptability and responsiveness to changing traffic patterns and congestion levels. This flexibility enables real-time optimization, ensuring traffic signals can adapt immediately to varying traffic conditions. It accommodates the dynamic nature of traffic patterns, including fluctuations due to factors such as time of day, events, accidents, and road construction. By dynamically adapting to traffic conditions, signals with dynamic phasing help reduce congestion, optimize signal timing to minimize delays and queue lengths, and ultimately improve travel times for motorists. Moreover, this adaptability contributes to enhanced safety on the roads by reducing the likelihood of accidents associated with sudden stops and congested traffic conditions. Overall, dynamic phasing plays a crucial role in promoting smoother traffic flow, reducing congestion, shortening travel times, and enhancing safety for all road users.

5.2.3. Reward Based on Eco_PI

In Distributed Multi-Agent Reinforcement Learning, rewards are calculated based on the collective performance of all agents in the environment. In a transportation environment, where each intersection functions as an individual agent. In this study, rewards are calculated locally for each agent. The reward for each agent is determined by a metric called negative Eco_PI, which represents the cumulative impact of stop delays and penalized stops. This metric encapsulates the undesirable effects of traffic congestion and inefficiencies at intersections, allowing agents to optimize their behavior to minimize these negative outcomes. By incorporating stop delays and penalized stops into the reward calculation, agents are incentivized to make decisions that improve traffic flow and reduce congestion, ultimately leading to more efficient and sustainable transportation systems.

The reward function was formulated as Eco_PI by measuring the number of stops and stop delays that occurred in every traffic approach, following an existing fuel consumption model proposed in the study [12,13]. The number of stops a vehicle makes is calculated by counting the number of times the vehicle is stopped in a queue while approaching from all directions in the intersection. The stop delay is calculated as the amount of time a vehicle is stationary in the queue before it reaches the intersection. For example, as shown in Figure 5, at the Cater intersection in MLK Smart Corridor, vehicle stops and stop delays are calculated on the eastbound, southbound, westbound, and northbound approaching links. These metrics are then used to calculate the Eco_PI index, which serves as an indicator of fuel consumption related to stopping. The immediate reward $r_{i,t}$ is calculated for each traffic movement of intersection i as

$$r_{i,t} = \text{Eco\_PI}_i = -(\sum_{l=1}^{L_i} \delta^{SD}_{i,l,t} + (\delta^{SK}_{i,l,t} * \delta^{NS}_{i,l,t})) \tag{9}$$

where $\delta^{SD}_{i,l,t}$ is the stop delays that occurred in link l_i, $\delta^{NS}_{i,l,t}$ is the number of stops, and $\delta^{SK}_{i,l,t}$ is the stop penalty penalized for every stops [14,15]. The policy of each agent i is optimized to maximize the global long-term return $E[R^{\pi}_0]$, where $R^{\pi}_{i,t} = \sum_{\tau}^{T} \gamma^{\tau-t} r_{i,t}$ is the return at time t, with a discount factor γ.

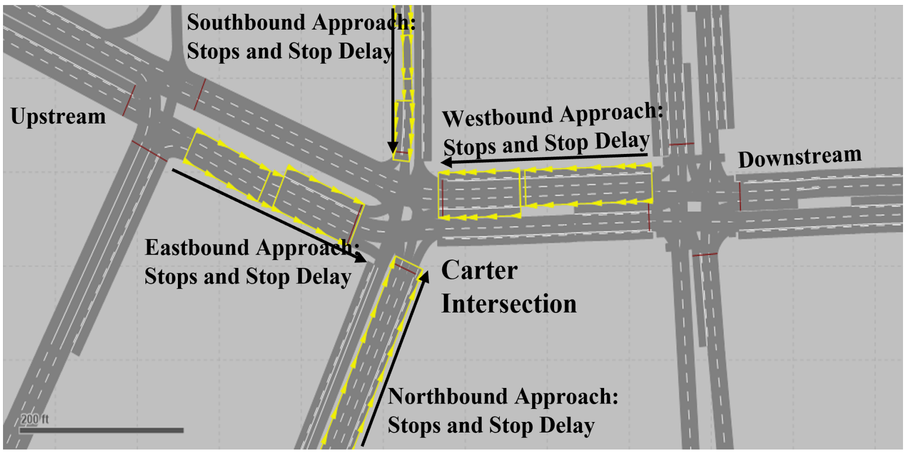

Figure 5. Vehicles Stops and Stop Delays at each approach.

5.2.4. Spatio-Temporal Multi-Agent Reinforcement Learning

Each intersection's behavior is modeled using a decentralized graph network and state and action spaces. Agents use the Multi-Agent Advantage Actor-Critic (MA2C) algorithm, with Actor and Critic designed using a graph neural network. Agents learn spatial and temporal dependencies through asynchronous communication protocols and make decisions based on their current state and policies. Policies are updated based on optimal long-term return values and evaluated and updated based on physical constraints. Each agent's state, action, and reward are communicated to the neighbors through message passing, and the reward is stored to measure the global return for each agent. Therefore, the multi-agent Markov Decision Process (MDP) was updated as (G, S, A, M, p, r, S') where $m_{ji,t} \in M_{ji}$ is the message passed from agent j to agent i including the states, actions, and rewards of the neighboring agent i at time t. $\mathcal{N}_i = \{j \in \mathcal{V} | ij \in \mathcal{E}\}$ represents the set of neighboring agents that are connected to agent i by links $l_{i,j}$ in the communication graph $(\mathcal{V}, \mathcal{E})$. Then the local agent state is updated as ${s'}_{i,t} \in S'$ which is the joint state of the agent's current state and the neighbor's state.

At time t, the state $s_{i,t}$ of intersection i includes traffic state such as volume, vehicle presence time in the detector zone, average waiting time, average delay, and vehicles average speed, as well as traffic signal state such as current phase state, duration, and pedestrian recall time. The states of neighboring agents $\mathcal{N}_i$ are obtained through message passing, including the aggregation of the agent's state and policy.

$$m_{i,t} = g(s_{j,t} \cup h_{j,t-1} \cup \pi_{j,t-1},\ \forall\ j \in \mathcal{N}_i) \tag{10}$$

Then the intersection i state is updated by the linear transformation with a rectified linear function, with the dimensions of the traffic state and traffic signal state input varying for each intersection. The hidden state of temporal traffic information is extracted by the LSTM layer.

$$h'_{i,t} = \xi(s_{i,t} \cup h_{i,t-1} \cup \pi_{i,t-1} \cup m_{i,t}) \tag{11}$$

Then a linear transformation with a rectified linear function is applied to the hidden graphs to identify the optimal policy, π_i. And the softmax function is applied to generate actions a'_i. The policy is evaluated and adjusted by considering mandatory physical constraints.

Advantage Actor-Critic (A2C) with a Graph Neural Network (GNN) stabilizes the learning process and enhances the performance of the proposed model in identifying the optimal policy for maximizing the expected cumulative discounted reward $E[R^{\pi}_{i,0}]$ over time steps for intersection i. The advantage function $A^{\pi}_i(s'_{i,t}, a'_{i,t})$ evaluates the benefit of taking an action $a'_{i,t}$ in a state $s'_{i,t}$ compared to the average value at that state and serves as a reference point for the action-value function $Q^{\pi}_i(s'_{i,t}, a'_{i,t})$. The state-value function $V^{\pi}_i(s'_{i,t})$ defines the predicted cumulative discounted reward from a specific state under a given policy and is calculated as the weighted sum of the action-value function for all possible actions.

The policy distribution approximates the anticipated cumulative discounted reward from taking an action in a state under the policy π_i. The advantage function helps the critic network reinforce the selection of the most suitable action by updating the policy distribution with policy gradients as directed by the critic, which in turn increases the probability of actions proportional to the high expected return $E[R^{\pi}_{i,0}] = \sum_{s'_{i,t} \in S'} p(s'_{i,t}) V^{\pi}_i(s'_{i,t})$.

Learning from experiences: During each time step, the experience replay buffer $\mathcal{D}$ stores the information including the initial state, the updated state with neighbor networks, updated policies and values, the new state after taking action, and the step reward $(s_{i,t}, a'_{i,t}, m_{\mathcal{N}_i,t}, r_{i,t}, s''_{i,t}, v'_{i,t}, \pi_{\theta_{i,t}})$.

In each subsequent time interval, the model learns the temporal dependency by utilizing the batch of experiences $\mathcal{B}$ is $\{(s'_{i,\tau}, m_{N_i,t,\tau}, a'_{i,\tau}, r_{i,\tau}, s''_{i,\tau}, v'_{i,\tau}, \pi'_{\theta_{i,\tau}})\}_{i \in V, \tau \in \mathcal{B}}$ stored in the replay buffer $\mathcal{D}$ and updates the graph neural network parameters based on the calculated losses. Where $\{\pi'_{\theta_i}\}_{i \in V}$ is stationary policy and value $\{V'_{\omega_i}\}_{i \in V}$ were updated after physical constraints evaluation of intersection i. Actor loss incorporates the negative log probability of the action that was sampled under the current policy, and the actor is updated based on the estimated advantage. And the Critic loss, which involves computing the mean squared error between the sampled action-value and the estimated state-value, is updated using the estimated state-value.

5.3. Digital Twin Assisted Method

Widely employed by traffic engineers and researchers, PTV-Vissim [83] is a microscopic road traffic simulator with a user-friendly Graphical User Interface (GUI) for designing road networks and setting up simulations. However, limitations arise when dynamically manipulating objects during simulations. To overcome this challenge, PTV-Vissim provides a solution through a Component Object Model (COM) interface. In this study, we utilize Python 3.8 scripts to develop the COM interface, enabling programmable manipulation of simulator functions and parameters.

To reduce congestion and improve Eco_PI, the PTV-Vissim COM interface was embedded with Digital Twins which represents the decentralized graph-based multi-agent reinforcement learning framework. The Digital Twins serve as a representation of the physical transportation environment, with each intersection mapped to a corresponding reinforcement learning agent. These agents interact with the Digital Twins through the COM interface, ensuring optimal policy maintenance and facilitating efficient decision-making for controlling signal phases within a tolerable time frame, as illustrated in Figure 4.

The Digital Twins-assisted DGMARL algorithm, illustrated in Algorithm 1, maps each intersection in the Digital Twins to a corresponding reinforcement learning agent i. To facilitate seamless scaling and integration of multiple agents, each agent is associated with a unique thread thread$_i$, leveraging multi-threading. This approach enables the agents to learn the global traffic state collectively and make optimal decisions at their respective intersections, thereby improving the Eco_PI.

At time t the agent i observes various features through Digital Twins components as shown in the algorithm in Table A1, such as the vehicle presence time in the detector zone, each direction approach level vehicle count aggregates, vehicles average speed, and current signal state (line-5). Then collaborates with its neighbors $\mathcal{N}_i$ to share and receive their states through message passing as described in Algorithm 1 line-6. Then the updated state $s'_{i,t}$ of agent i is processed through a graph neural network to derive the optimal policy π_i and select actions to control the signal phase ϕ_i (line-7). Then agent i validates the actions (line-8), against the physical constraints configured in the Digital Twins, the minimum green serving time and pedestrian recall time, to ensure user safety. If the decision is to stay in the current phase in green, then no actions are applied back to the Digital Twins; otherwise, agent i validates the other phase's vehicle presence time in the detector zone and selects the phase ϕ_j that has a higher upcoming traffic demand, then applies the signal phase change action to the signal controller in the Digital Twins (line-9) as shown in the algorithm in Table A2, which updates the simulation. Once the decided action is applied, each agent i estimates the current reward r_i with the new observed traffic state $s_{i,t+1}$ (line-10), and stores the experiences in the replay buffer (line-11). When the buffer size reaches minimum batch size the agent starts to learn from the collection of experiences at every time step to minimize the critic loss $L(\omega_i)$ and actor loss $\hat{J}(\theta)$ (lines 12–14). The agent i repeats the above processes until it achieves the desired objective of identifying optimal policy to choose the best actions for reducing congestion and Eco_PI.

Due to the distributed agent environment, each agent makes different decisions based on their local and neighboring traffic state, so the convergence of an optimal policy is different for each agent and the efficiency of learning is increased. Since agents continue to interact with the real environment through the Digital Twins, the probability of arriving at an optimal policy is faster. Hence, by using a Digital Twins and reinforcement learning, the system can adapt to changing traffic conditions in real-time, leading to more efficient signal control, and it can be further optimized to maximize its benefits.

Algorithm 1 Digital Twin assisted DGMARL Learning

Require α learning rate, β entropy coefficient.
Ensure: Initialize PTV-Vissim objects Vissim, Net, Links, Signal Controllers, and Signal Groups.
Ensure: Initialize graph $G(\mathcal{V}, \mathcal{E})$, agent $i \in \mathcal{V}$, link $l_i \in \mathcal{E}$, physical constraints i_c, policy network parameters θ, and value network parameters ω.

for $e = 1$ to episodes **do**
  for $t = 0$ to $T - 1$ **do**
    for agent $i = 1$ to $\mathcal{V}$ create thread thread$_i$ **do**
      Thread thread$_i$ starts
      Observe state s_i from Digital Twin.
      Update state $s'_{i,t} \approx s_{i,t} \cup \pi_{N_{i,t-1}} \cup h_{N_{i,t-1}}$ through message passing.
      Select policy $\pi_{\theta_{i,t}}$, action $a_{i,t} \approx \pi(a|s'_{i,t})$, and get value $v(s'_{i,t}|\omega, a_{i,t})$.
      Evaluate agent's actions $a'_{i,t} = (a_{i,t}|i_c)$ and update value $v'(s'_{i,t}|\omega, a'_{i,t})$ and policy $\pi'(a'_{i,t}|v'_{i,t}, s'_{i,t})$.
      Take action $a'_{i,t}$ in Digital Twin
      Observe reward $r_{i,t}$ and new state $s'_{i,t+1}$.
      Store the observations in replay buffer $D \leftarrow (s'_{i,t}, \pi'_{\theta_{i,t}}, a'_{i,t}, r_{i,t+1}, s'_{i,t+1}, v'_{\omega,i,t})$.
      if $t >=$ sample batch size B **then**
        Sample random minibatch of B samples $(s^j, a^j, r^j, s'^j j)$ from D $\forall\, j \in 1 \ldots B$.
        Obtain target return $y_i^j = r_i^j + \gamma Q_i^{\pi'}(s'^j, a'_1, \ldots, a'_N)$ where $a'_i = \pi'(s_i'^j)$.
        Update critic by minimizing the loss: $L(\omega_i) = \frac{1}{B}\sum_j [y_i^j j - Q_i^p i(s'^j, a_1^j, \ldots, a_N^j)]^2$ and $\omega_i = \omega_i - \alpha \nabla L(\omega_i)$.
        Update actor using sampled policy gradient descent along with entropy loss: $Adv_i'^j = y_i^j - Q_i^{\pi'}(s'^j, a_1^j, \ldots, a_N^j)$.
        $\hat{J}(\theta) = \frac{1}{B}\sum_j \nabla - \log \pi_{\theta_i}(a_i'^j|s_i'^j) Adv_i'^j + \beta \sum \pi_{\theta_i}(a_i'^j|s_i'^j) \log \pi_{\theta_i}(a_i'^j|s_i'^j)$.
        $\theta_i = \theta_i + \alpha \nabla \hat{J}(\theta)$.
      end if
      Thread thread$_i$ ends
    end for
  end for
end for

---

6. Experiments

This section provides the details of the experiment setup using a real-world dataset and optimization results that show the efficiency of the Digital Twins-assisted DGMARL model.

6.1. Experiment Design

The experiment environment was set up using the real-world dataset collected by the Department of Computer Science and Engineering at the University of Tennessee, Chattanooga, TN, USA [84].

Real-world dataset: The dataset is composed of the corridor that connects 11 intersections on MLK Smart Corridor with bidirectional traffic in East-West, West-East, North-South, and South-North directions and includes data for roadway network geometry, the traffic signal timing plan, camera and zone-detecting device, Signal Phasing and Timing (SPaT), vehicle flow, vehicles speed, and vehicle presence time in the detector zone, etc. The signal timing plan attained from the city for each intersection follows a dual-ring NEMA controller protocol. In the developed DGMARL algorithm an adaptive signal control strategy is adapted where the action has been designed with two decisions: (1) Action 1: Stay in green in the current ongoing phase or (2) Action 2: Switch to the phase that serves the lanes with the highest traffic demand. Before switching to the new phase, the current phase follows the yellow and red clearance times that are specific for each phase at each intersection.

6.2. Digital Twin Setup

The Tier 1 Digital Twins platform is used in this experiment. The simulation model of 11 intersections of the MLK Smart Corridor in PTV-Vissim is developed following network creation guidelines in [85], as shown in Figure 6. The developed model is populated with archived 15 December 2022, one-minute volume counts at network entry edges, 10-min turn percentages at each intersection approach, and signal timing plans received from the city. Two versions of the simulation model are created: (1) the PM peak model that simulates the 15 December 2022, 3:00 p.m.–6:00 p.m. scenario, and (2) the 24 h model that simulates the 15 December 2022, 24 h scenario. This model is prepopulated with data and runs faster than wall clock time.

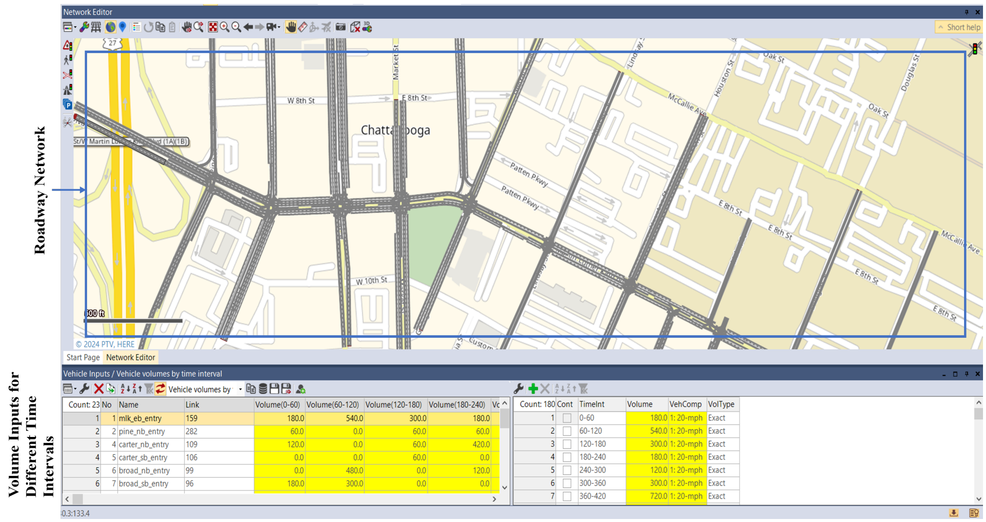

Figure 6. MLK Smart Corridor in PTV-Vissim showing roadway network layout and vehicle volume inputs per time interval.

6.3. Impact of the Application of the Proposed Model

Efficiency of Digital Twins-assisted DGMARL is measured using the number of stops and stop delays at each intersection as the metrics. Also, Eco_PI is calculated to measure the impact of fuel consumption related to stopping. DGMARL starts optimizing signal timing after a non-stationary period of 120 s. At each time step, the graph neural network updates each agent's current state with Relu activation in the message passing layer. Then, the actor and critic neural networks generate the value-assisted action probability. This process continues until the initial batch size of experiences is gathered. Afterward, at each time step, the model learns from experience with random samples and updates the graph neural network parameters to arrive at the optimal policy distribution. The model decay rate is customized based on the current learning episode. To achieve optimal results, the model was trained for 100 episodes using the dataset from the first hour of MLK Smart Corridor on Thursday, 15 December 2022. Each episode's simulation step was 36,000 deciseconds, and the model learned from 240 batch sizes of experience replay at every 240-time step.

6.4. Experiment Results

The developed DGMARL signal timing plan was tested on MLK Smart Corridor for 24 h and PM-peak hour scenarios of 15 December 2022. The performance of DGMARL signal timing plan was compared with the baseline actuated MLK Smart Corridor vehicle actuated signal timing plan. Baseline actuated signal timing is a traditional method for controlling traffic signals at intersections, where signal timings are predetermined based on factors like time of day and traffic volume patterns. The signal phases, including green, yellow, and red intervals, are fixed and not adjusted in real time. Comparing the signal optimization method DGMARL against this baseline provides insights into potential improvements in traffic management strategies.

6.4.1. 24-Hour Scenarios

Figure 7 shows the comparison of the Eco_PI index observed for every second from DGMARL and baseline vehicle actuated signal timing plans. The overall Eco_PI improved by 55.38%, with improvements ranging from 3.17% to 62.14% over the 10 intersections, and 34.77% Eco_PI increased in the Douglas intersection.

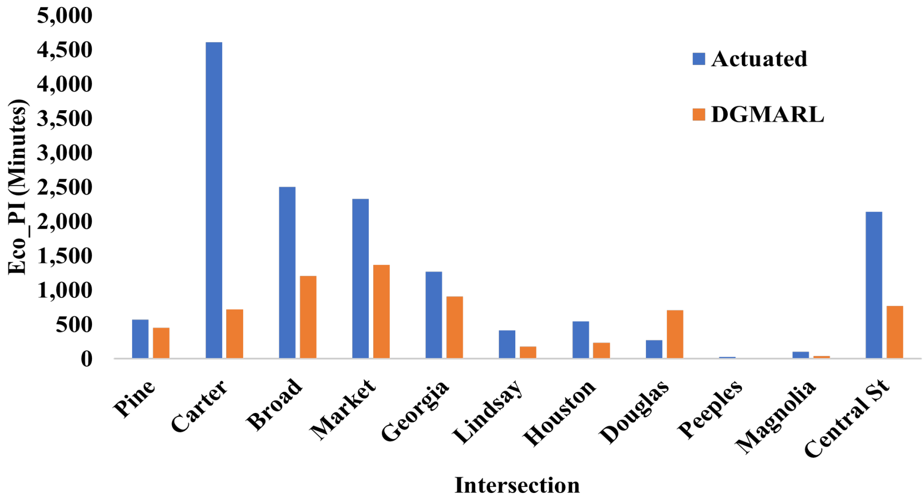

Figure 7. 55.38% improvement of overall Eco_PI in one 24 h test duration.

6.4.2. PM-Peak Hour Scenario

The experiment results were obtained from ten replicate trials, each conducted with different random seeds, during the PM-peak hour. These trials compared the Eco_PI performance observed for every second from the DGMARL approach against the implementation of a vehicle-actuated signal timing plan. Figure 8 illustrates the average Eco_PI over these ten trials, showcasing an improvement of 38.94%. The Eco_PI reduction ranged from 3.17% to 62.14% across the ten intersections analyzed. Notably, the Douglas intersection exhibited a relatively higher Eco_PI than the actuated signal timing plan.

Figure 9 illustrates the average stop delay and the average number of stops observed across ten test runs. On average, stop delays decreased by 42.78%, while the average number of stops increased by 0.82%. In one of the tests using random seed value 32, during the PM-peak hour scenario, a significant reduction in Eco_PI is evident in both stops and stop delay. Specifically, compared to the baseline actuated signal timing scenario, there was a 13% reduction in stops and a 43.29% reduction in stop delay, as illustrated in Figure 10. Among the intersections, the Central St and Market intersections experienced the most substantial improvement, with a 65.04% and 51.80% reduction in average stop delay respectively, while the Pine intersection demonstrated the least improvement at 5.73%. The trend of higher Eco_PI observed for Douglas for DGMARL compared to actuated in Figure 8 above is reflected in stop and delay as well. The stops and stop delays in Douglas intersection are slightly higher for DGMARL compared to the actuated scenario.

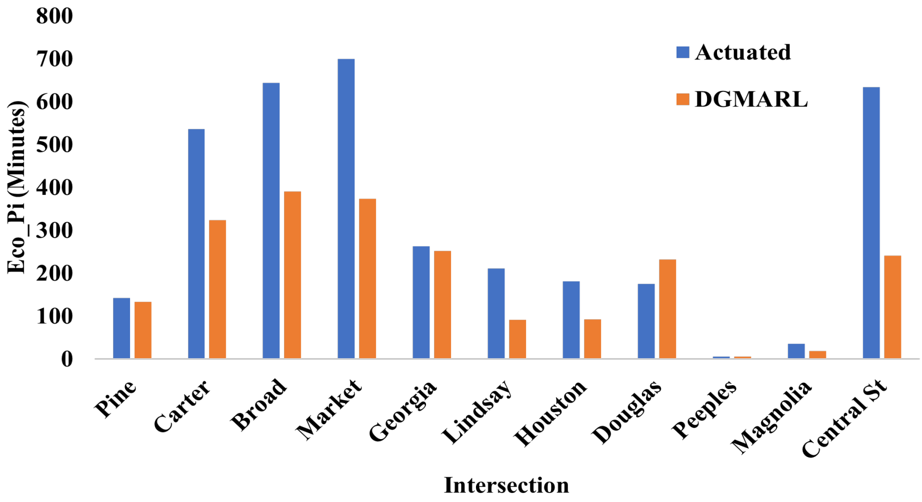

Figure 8. 38.94% improvement in average of Eco_PI during PM-peak hour with 10 runs of tests.

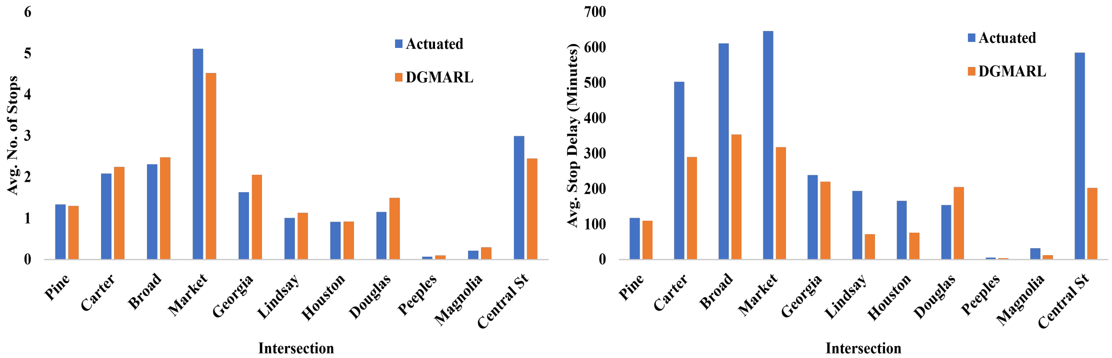

Figure 9. 42.78% reduction in the average of stop delays and 0.82% increase in the average number of stops during Pm-peak hour with 10 runs of tests.

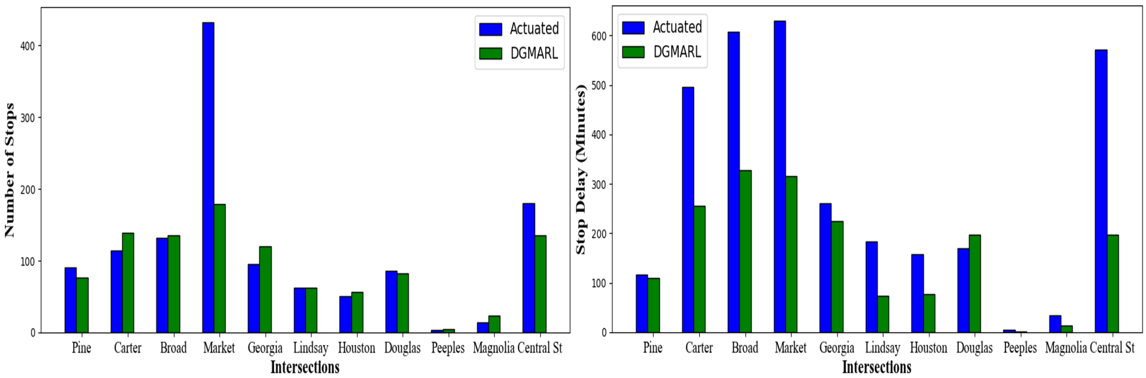

Figure 10. 13% reduction in stops and 43.29% reduction in stop delays during PM-peak hour in one run of the test.

A closer examination of specific intersections, such as Pine, Center, Market, and Douglas, revealed interesting trends in Eco_PI improvements, as depicted in Figures 11 and 12. These plots show the variation of Eco PI during the simulation period for both the actuated and DMARL scenarios for one of the replicate trials.

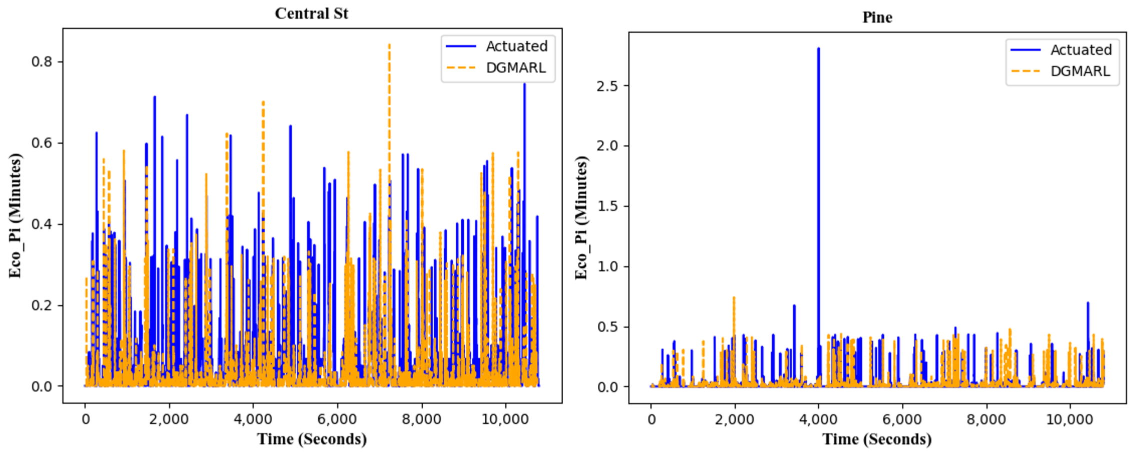

Figure 11. Variation of Eco PI during simulation period for Actuated and DGMARL.

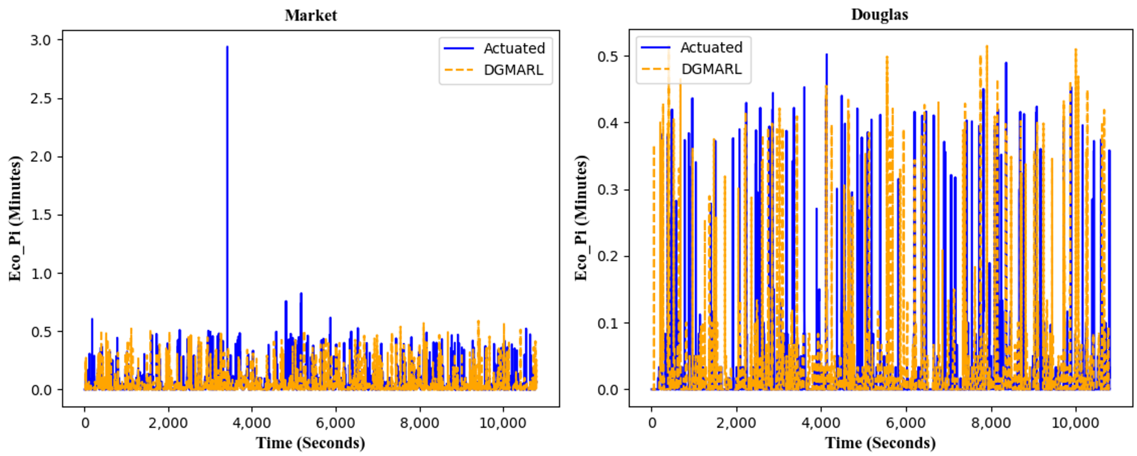

Figure 12. Variation of Eco PI during simulation period for actuated and DGMARL.

As depicted in Figure 11, consistent improvement in Eco_PI at the Central intersection demonstrates the effectiveness of the DGMARL approach in optimizing signal timing. Additionally, notable reductions in Eco_PI at the Pine intersection (Figure 11) and the Market intersection (Figure 12) during specific time intervals highlight the potential for targeted signal control adjustments to alleviate congestion and enhance fuel efficiency. However, the occasional instances of higher Eco_PI at the Douglas intersection (Figure 12) compared to the actuated scenario suggest the need for further investigation into the factors influencing signal optimization outcomes, particularly in complex traffic scenarios.

In this study, both the actuated and DGMARL approaches were subjected to the same number of vehicle inputs, as illustrated in shown in Table A1 in Appendix B. However, DGMARL demonstrated its effectiveness in optimizing traffic signal timing and improving Eco_PI. These findings underscore the dynamic nature of traffic flow and highlight the efficacy of DGMARL in mitigating congestion and promoting eco-friendly transportation practices.

In Figure 13, an analysis of traffic flow patterns reveals insights into the distribution of vehicles crossing each intersection at green. Additionally, detailed information on the cumulative sum of the count of vehicles crossing each intersection at green can be found in the table provided in Table A2.

This analysis highlights specific trends observed at individual intersections. For instance, at the Douglas intersection, there is a notable increase of 17.48% in the percentage of vehicles arriving on green compared to the actuated signal plan. Similarly, at the Georgia, Market, and Pine intersections, the DGMARL scenario shows increases of 7.94%, 4.02%, and 0.77% respectively in the percentage of vehicles arriving on green compared to the actuated scenario. Conversely, at Central St, there is a reduction of 2.76% in vehicles arriving on green compared to the actuated plan. It's important to note that the arrival on green percentage is calculated considering all signal phases and it's all approaches of the intersection.

In Figure 14, a closer examination is made of the traffic flow from approaching links at Central St, Pine, Market, and Georgia intersections to analyze the number of stopped vehicles at each intersection approach. At Central St, there is a notable 12.90% reduction in the number of vehicles stopped compared to the actuated plan, indicating improved traffic flow. Despite Central St having 2.28% fewer moving vehicles compared to the actuated signal timing plan, there is still a reduction in Eco PI, primarily due to the significant decrease in stopped vehicles. Conversely, at the Douglas Intersection, there is a 20.75% increase in the number of stops compared to the actuated plan, leading to a higher Eco_PI, despite a 2.23% increase in the number of moving vehicles during the simulation period. This underscores the importance of considering the distribution and frequency of stopped vehicles at intersection approaches in assessing the impact on Eco_PI.

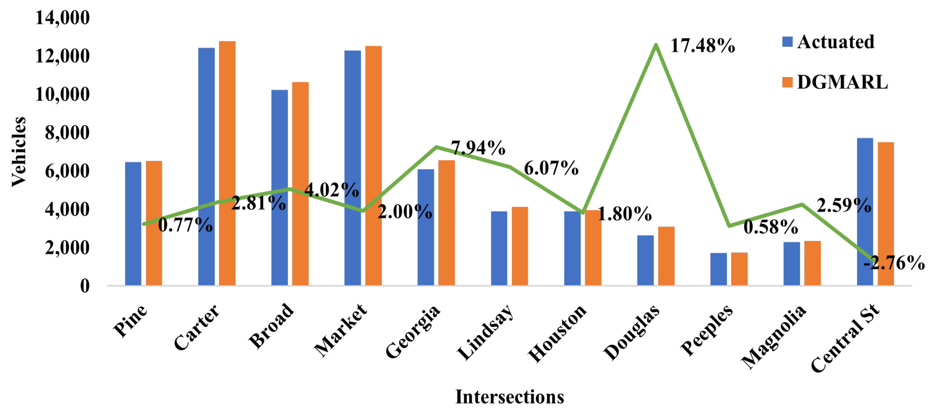

Figure 13. PM-peak hour traffic throughput in the MLK Smart Corridor: Actuated vs DGMARL.

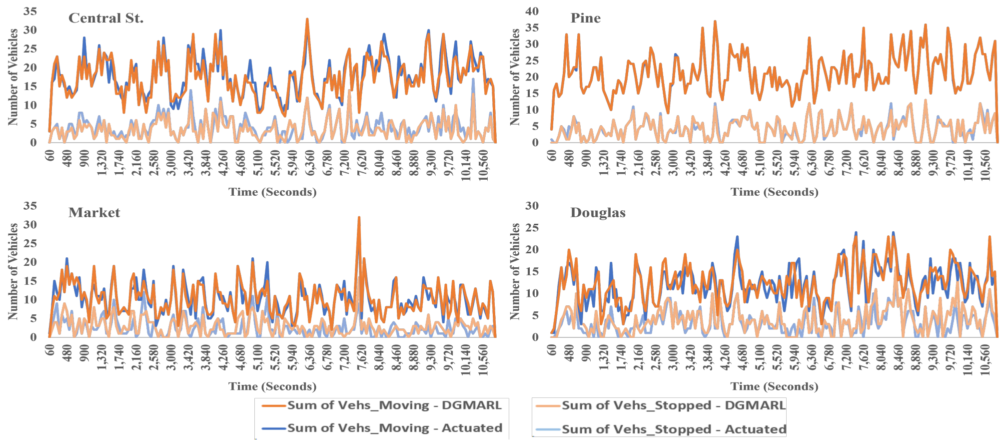

Figure 14. Stopped vs Moving Vehicles in PM-peak hour traffic.

An important consideration with DGMARL is its focus on addressing traffic demands from all directions, including both main streets and side streets, while adhering to mandatory constraints such as minimum green time and pedestrian walk serving time. As a consequence of this approach, the traffic density on certain approaches may increase, reflecting the comprehensive optimization of traffic flow in a multi-directional environment. This heightened traffic density on specific approaches may have implications for the Eco_PI, underscoring the complex interplay between traffic demand patterns and signal optimization strategies.

Overall, the results demonstrate the effectiveness of the DGMARL approach in managing traffic demand across all directions and optimizing traffic signal timing to reduce Eco_PI, while maintaining adherence to mandatory constraints at both corridor and intersection levels.

7. Discussion

The findings from the experiments shed light on the effectiveness of the Digital Twin-assisted graph-based decentralized multi-agent reinforcement learning algorithm in optimizing traffic network signal timing. These results warrant a thorough discussion to interpret their significance in the context of previous studies and the underlying hypotheses, as well as to explore their broader implications and potential future research directions.

Firstly, the observed improvements in learning efficiency and performance corroborate with prior research that has emphasized the advantages of multi-agent reinforcement learning in dynamic and complex environments. By allowing agents to interact with their surroundings, exchange knowledge with neighboring agents, and explore multiple actions simultaneously, our approach aligns with the principles outlined in previous studies on traffic signal optimization.

Furthermore, the efficacy of the Digital Twin in assisting the algorithm highlights the growing role of Digital Twin technology in optimizing real-world systems. By providing a virtual representation of the physical transportation environment, the Digital Twin enables more accurate observations and simulations, leading to enhanced decision-making capabilities for the reinforcement learning algorithm.

In discussing the implications of these findings, it becomes evident that the proposed approach holds promise for addressing traffic congestion and improving overall transportation efficiency on a larger scale. The ability to optimize traffic signal timing in real-time based on evolving traffic conditions offers significant potential for reducing travel times, minimizing delays, and enhancing the overall commuter experience.

Looking ahead, future research directions should focus on further validating and refining the algorithm through extensive testing in real-road environments. This includes deploying the algorithm in larger traffic networks, incorporating additional functionalities such as adaptive learning mechanisms, and exploring variations in optimization frequencies. Additionally, investigations into the algorithm's robustness under diverse traffic scenarios and its scalability to accommodate growing urban infrastructures are warranted.

In summary, the results of our experiments underscore the promising prospects of leveraging Digital Twin-assisted multi-agent reinforcement learning for traffic signal optimization. By engaging in discussions that contextualize these findings within the existing literature, highlight their implications, and delineate future research avenues, we aim to contribute to the ongoing dialogue on enhancing traffic management systems and mitigating congestion in urban environments.

8. Conclusions

This paper has delved into the application of a Digital Twin-assisted graph-based decentralized multi-agent reinforcement learning algorithm for real-time optimization of traffic network signal timing. Through enabling interactions among multiple agents, facilitating knowledge exchange among neighboring agents, and allowing simultaneous exploration of multiple actions, this approach has exhibited notable enhancements in learning efficiency and performance, all while maintaining lower latency.

The experiment results have underscored the effectiveness of leveraging Digital Twin technology to assist the multi-agent reinforcement learning algorithm in optimizing traffic network signal timing. Through extensive experimentation on the MLK Smart Corridor in Chattanooga, Tennessee, USA, we observed significant improvements in traffic flow and eco-friendly transportation practices compared to traditional vehicle-actuated signal timing plans. Notably, our results demonstrate a substantial reduction in the Eco_PI index, indicating enhanced fuel efficiency and reduced emissions.

However, it's important to acknowledge the disparities between our findings and those reported in previous studies. While our approach showcases promising results in managing traffic demand and optimizing signal timings, variations in traffic patterns, infrastructure layouts, and environmental factors may contribute to differences in outcomes across different contexts. Therefore, further investigation and comparative analysis are warranted to elucidate the factors influencing these disparities and refine our understanding of the DGMARL algorithm's performance under diverse conditions. By engaging in such discussions and continuously evaluating our findings in light of previous research, we can gain deeper insights into the capabilities and limitations of MARL-based approaches for traffic management.

In summary, this study charts a promising course for advancing traffic management systems and alleviating congestion on a broader scale. The integration of Digital Twin technology with multi-agent reinforcement learning provides a robust framework for optimizing complex systems characterized by multiple agents and diverse interactions. By continuing to explore and refine this approach through real-world testing and observation, we can unlock its full potential to revolutionize urban mobility and enhance overall transportation efficiency.

Author Contributions: Conceptualization, V.K.K. and A.J.S.; methodology, V.K.K., A.J.S., Y.L., D.W., M.P.H. and A.G.; software, V.K.K. and A.J.S.; validation, M.P.H. and A.G. and M.S.; formal analysis, V.K.K. and A.J.S.; investigation, V.K.K. and A.J.S.; resources, M.S.; data curation, M.S.; writing—original draft preparation, V.K.K. and A.J.S.; writing—review and editing, V.K.K., A.J.S., Y.L., D.W., M.P.H. and A.G.; visualization, V.K.K. and A.J.S.; supervision, Y.L. and D.W.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partly supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Award Number DE-EE0009208. The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan (https://energy.gov/downloads/doe-public-access-plan, accessed on 5 March 2024).

Data Availability Statement: The data associated with this research study are securely stored and accessible only upon approval from the Center for Urban Informatics and Progress (CUIP) within the Department of Computer Science at the University of Tennessee at Chattanooga, USA. Access to the data requires compliance with institutional policies and regulations governing data access and confidentiality. Researchers interested in accessing the data should contact the CUIP team for approval and further instructions.

Acknowledgments: This material is based upon work partially supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Award Number DE-EE0009208. This manuscript has been partly authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan DOE Public Access Plan (https://energy.gov/downloads/doe-public-access-plan, accessed on 5 March 2024). This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DT	Digital Twin
DGMARL	Decentralized Graph-Based Multi-Agent Reinforcement Learning
AI	Artificial Intelligence
ML	Machine Learning
IoT	Internet of Things
RL	Reinforcement Learning
MARL	Multi-Agent Reinforcement Learning
MDP	Markov Decision Process
A2C	Advantage Actor-Critic
MA2C	Multi-Agent Advantage Actor-Critic
GNN	Graph Neural Network
LSTM	(Long Short-Term Memory)
ReLU	(Rectified Linear Unit)
ITS	Intelligent Transportation System
SPaT	Signal Phasing and Timing
RBC	Ring Barrier Controller
GUI	Graphical User Interface
COM	Component Object Model
MLK	Martin Luther King

Appendix A

Algorithm A1 Observing state s_i from Digital Twin by thread thread$_i$

```
Require PTV-Vissim objects and agents initialized, time t, thread thread_i
for signal group sg_i do
    if sg_i is current phase signal group then
        Observe current phase status φ^S_{i,t}, current phase duration φ^D_{i,t}, pedestrian recall status φ^{PR}_{i,t}
        Validate minimum green is served φ^{MinG}_{i,t}
        if sg_i has pedestrian recall enabled then
            Validate maximum green is served φ^{MaxG}_{i,t}
        end if
    end if
    for links l = 1 to K do
        Observe traffic state δ^{PT}_{l,i,t}, δ^{W̄}_{l,i,t}, δ^{D̄}_{l,i,t}, and V̄_{l,i,t}
    end for
end for
```

Appendix B

Table A1. Vehicles Input.

Input Vehicles Generated	Actuated	DGMARL
Number of Vehicles	8788	8788
Unique Vehicle IDs	[1, 2, 3, ..., 8786, 8787, 8788]	[1, 2, 3, ..., 8786, 8787, 8788]

Algorithm A2 Apply action a_i' in Digital Twin by thread thread$_i$

Require Time t, thread thread$_i$, Current phase ϕ_i, Next phase ϕ_j

```
for signal group sg_i do
  if φ_i-current phase then
    if φ_i is green then
      Set φ_i to yellow
      Continue simulation
    else if φ_i is yellow and yellow served then
      Set φ_i to red clearance phase
      Continue simulation
    else if φ_i is red and red clearance served then
      Set status φ_i red clearance served
    end if
  end if
  if φ_i is red clearance served then
    Set φ_j to green
    Continue simulation
  end if
end for
```

Table A2. Cumulative sum of count of vehicles passed through each intersection at green.

Intersection	Actuated	DGMARL	Increase in % of Vehicles Crossed the Intersection at Green
Pine	6472	6522	0.77%
Carter	12,430	12,779	2.81%
Broad	10,217	10,628	4.02%
Market	12,282	12,528	2.00%
Georgia	6083	6566	7.94%
Lindsay	3886	4122	6.07%
Houston	3886	3956	1.80%
Douglas	2632	3092	17.48%
Peeples	1723	1733	0.58%
Magnolia	2280	2339	2.59%
Central St	7723	7510	−2.76%

References

1. Governments. Smart Corridor City/County Association of Governments. Available online: https://ccag.ca.gov/projects/smart-corridor/ (accessed on 1 June 2023).
2. ARCADIS. Creating an Intelligent Transportation Systems for Atlanta's First Smart Corridor. Available online: https://www.arcadis.com/en-us/projects/north-america/united-states/north-ave-corridor (accessed on 1 June 2023).
3. California, US, R.T.T. I-80 SMART Corridor Project. Available online: https://www.roadtraffic-technology.com/projects/i-80-smart-corridor-project-california/ (accessed on 1 June 2023).
4. San Francisco, CA, P. I-80 SMART Corridor. Available online: https://www.parsons.com/project/80-smart-corridor/ (accessed on 1 June 2023).
5. Journal, A. Tennessee DOT Starts Phase 2 of I-24 SMART Corridor. Available online: https://aashtojournal.org/2022/04/01/tennessee-dot-starts-phase-2-of-i-24-smart-corridor/ (accessed on 1 June 2023).
6. Wu, J.; Wang, X.; Dang, Y.; Lv, Z. Digital Twins and artificial intelligence in transportation infrastructure: Classification, application, and future research directions. *Comput. Electr. Eng.* **2022**, *101*, 107983. [CrossRef]
7. Saroj, A.J.; Guin, A.; Hunter, M. Deep LSTM recurrent neural networks for arterial traffic volume data imputation. *J. Big Data Anal. Transp.* **2021**, *3*, 95–108. [CrossRef]
8. Farazi, N.P.; Zou, B.; Ahamed, T.; Barua, L. Deep reinforcement learning in transportation research: A review. *Transp. Res. Interdiscip. Perspect.* **2021**, *11*, 100425.
9. Chowdhury, M.; Sadek, A.W. Advantages and limitations of artificial intelligence. *Artif. Intell. Appl. Crit. Transp. Issues* **2012**, *6*, 360–375.

10. Machin, M.; Sanguesa, J.A.; Garrido, P.; Martinez, F.J. On the use of artificial intelligence techniques in intelligent transportation systems. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Barcelona, Spain, 15–18 April 2018; pp. 332–337.
11. Kumarasamy, V.K.; Saroj, A.J.; Liang, Y.; Wu, D.; Hunter, M.P.; Guin, A.; Sartipi, M. Traffic Signal Optimization by Integrating Reinforcement Learning and Digital Twins. In Proceedings of the 2023 IEEE Smart World Congress (SWC), Portsmouth, UK, 28–31 August 2023; pp. 1–8.
12. Stevanovic, A.; Shayeb, S.A.; Patra, S.S. Fuel consumption intersection control Performance Index. *Transp. Res. Rec.* **2021**, *2675*, 690–702. [CrossRef]
13. Alshayeb, S.; Stevanovic, A.; Effinger, J.R. Investigating impacts of various operational conditions on fuel consumption and stop penalty at signalized intersections. *Int. J. Transp. Sci. Technol.* **2022**, *11*, 690–710. [CrossRef]
14. Stevanovic, A.; Dobrota, N. Impact of various operating conditions on simulated emissions-based stop penalty at signalized intersections. *Sustainability* **2021**, *13*, 10037. [CrossRef]
15. Alshayeb, S.; Stevanovic, A.; Park, B.B. Field-based prediction models for stop penalty in traffic signal timing optimization. *Energies* **2021**, *14*, 7431. [CrossRef]
16. Nia, N.G.; Amiri, A.; Nasab, A.; Kaplanoglu, E.; Liang, Y. The Power of ANN-Random Forest Algorithm in Human Activities Recognition Using IMU Data. In Proceedings of the 2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Pittsburgh, PA, USA, 15–18 October 2023; pp. 1–7.
17. Nia, N.G.; Kaplanoglu, E.; Nasab, A.; Qin, H. Human Activity Recognition Using Machine Learning Algorithms Based on IMU Data. In Proceedings of the 2023 5th International Conference on Bio-Engineering for Smart Technologies (BioSMART), Paris, France, 7–9 June 2023; pp. 1–8.
18. Nia, N.G.; Nasab, A.; Kaplanoglu, E. Reinforcement Learning-Based Grasp Pattern Control of Upper Limb Prosthetics in an AI Platform. In Proceedings of the 2022 3rd International Informatics and Software Engineering Conference (IISEC), Ankara, Turkey, 15–16 December 2022; pp. 1–4.
19. Sun, C.; Kumarasamy, V.K.; Liang, Y.; Wu, D.; Wang, Y. Using a Layered Ensemble of Physics-Guided Graph Attention Networks to Predict COVID-19 Trends. *Appl. Artif. Intell.* **2022**, *36*, 2055989. [CrossRef]
20. Hassan, Y.; Sartipi, M. ChattSpeed: Toward a New Dataset for Single Camera Visual Speed Estimation for Urban Testbeds. In Proceedings of the 2023 IEEE International Conference on Big Data (BigData), Sorrento, Italy, 15–18 December 2023; pp. 2598–2605. [CrossRef]
21. Hassan, Y.; Zhao, J.; Harris, A.; Sartipi, M. Deep Learning-Based Framework for Traffic Estimation for the MLK Smart Corridor in Downtown Chattanooga, TN. In Proceedings of the 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), Bilbao, Spain, 24–28 September 2023; pp. 4564–4570. [CrossRef]
22. Chu, T.; Chinchali, S.; Katti, S. Multi-agent Reinforcement Learning for Networked System Control. *arXiv* **2020**, arXiv:2004.01339.
23. Wang, Y.; Xu, T.; Niu, X.; Tan, C.; Chen, E.; Xiong, H. STMARL: A spatio-temporal multi-agent reinforcement learning approach for cooperative traffic light control. *IEEE Trans. Mob. Comput.* **2020**, *21*, 2228–2242. [CrossRef]
24. Chu, T.; Wang, J.; Codecà, L.; Li, Z. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 1086–1095. [CrossRef]
25. Khaleghian, S.; Neema, H.; Sartipi, M.; Tran, T.; Sen, R.; Dubey, A. Calibrating Real-World City Traffic Simulation Model Using Vehicle Speed Data. In Proceedings of the 2023 IEEE International Conference on Smart Computing (SMARTCOMP), Nashville, TN, USA, 26–30 June 2023; pp. 303–308. [CrossRef]
26. Gurjanov, A.; Zakoldaev, D.; Shukalov, A.; Zharinov, I. Formation principles of Digital Twins of Cyber-Physical Systems in the smart factories of Industry 4.0. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *483*, 012070. [CrossRef]
27. Leng, J.; Wang, D.; Shen, W.; Li, X.; Liu, Q.; Chen, X. Digital Twins-based smart manufacturing system design in Industry 4.0: A review. *J. Manuf. Syst.* **2021**, *60*, 119–137. [CrossRef]
28. Stavropoulos, P.; Mourtzis, D. Digital Twins in industry 4.0. In *Design and Operation of Production Networks for Mass Personalization in the Era of Cloud Technology*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 277–316.
29. Wagner, R.; Schleich, B.; Haefner, B.; Kuhnle, A.; Wartzack, S.; Lanza, G. Challenges and potentials of Digital Twins and industry 4.0 in product design and production for high performance products. *Procedia CIRP* **2019**, *84*, 88–93. [CrossRef]
30. Schluse, M.; Priggemeyer, M.; Atorf, L.; Rossmann, J. Experimentable Digital Twins—Streamlining simulation-based systems engineering for industry 4.0. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1722–1731. [CrossRef]
31. Cinar, Z.M.; Nuhu, A.A.; Zeeshan, Q.; Korhan, O. Digital Twins for industry 4.0: A review. In *Proceedings of the Industrial Engineering in the Digital Disruption Era: Selected Papers from the Global Joint Conference on Industrial Engineering and Its Application Areas, GJCIE 2019, Gazimagusa, Turkey, 2–3 September 2019*; Springer: Cham, Switzerland, 2020; pp. 193–203.
32. Jiang, Y.; Yin, S.; Li, K.; Luo, H.; Kaynak, O. Industrial applications of Digital Twins. *Philos. Trans. R. Soc. A* **2021**, *379*, 20200360. [CrossRef] [PubMed]
33. Sahal, R.; Alsamhi, S.H.; Brown, K.N.; O'shea, D.; McCarthy, C.; Guizani, M. Blockchain-empowered Digital Twins collaboration: Smart transportation use case. *Machines* **2021**, *9*, 193. [CrossRef]
34. Kosacka-Olejnik, M.; Kostrzewski, M.; Marczewska, M.; Mrówczyńska, B.; Pawlewski, P. How Digital Twin concept supports internal transport systems?—Literature review. *Energies* **2021**, *14*, 4919. [CrossRef]

35. Schwarz, C.; Wang, Z. The role of Digital Twins in connected and automated vehicles. *IEEE Intell. Transp. Syst. Mag.* **2022**, *14*, 41–51. [CrossRef]
36. Samuel, P.; Saini, A.; Poongodi, T.; Nancy, P. Artificial intelligence–driven Digital Twins in Industry 4.0. In *Digital Twin for Smart Manufacturing*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 59–88.
37. Fedorko, G.; Molnar, V.; Vasil', M.; Salai, R. Proposal of Digital Twin for testing and measuring of transport belts for pipe conveyors within the concept Industry 4.0. *Measurement* **2021**, *174*, 108978. [CrossRef]
38. Novák, P.; Vyskočil, J. Digitalized Automation Engineering of Industry 4.0 Production Systems and Their Tight Cooperation with Digital Twins. *Processes* **2022**, *10*, 404. [CrossRef]
39. Aheleroff, S.; Xu, X.; Zhong, R.Y.; Lu, Y. Digital Twin as a Service (DTaaS) in Industry 4.0: An Architecture Reference Model. *Adv. Eng. Inform.* **2021**, *47*, 101225. [CrossRef]
40. Bhatti, G.; Mohan, H.; Raja Singh, R. Towards the future of smart electric vehicles: Digital Twin technology. *Renew. Sustain. Energy Rev.* **2021**, *141*, 110801. [CrossRef]
41. Deryabin, S.A.; Temkin, I.O.; Zykov, S.V. About some issues of developing Digital Twins for the intelligent process control in quarries. In Proceedings of the 24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, KES2020, Virtual, 16–18 September 2020; Procedia Computer Science; Volume 176, pp. 3210–3216. [CrossRef]
42. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y. Digital Twin in industry: State-of-the-art. *IEEE Trans. Ind. Inform.* **2018**, *15*, 2405–2415. [CrossRef]
43. Farsi, M.; Daneshkhah, A.; Hosseinian-Far, A.; Jahankhani, H. (Eds.) *Digital Twin Technologies and Smart Cities*; Springer: Cham, Switzerland, 2020.
44. Jones, D.; Snider, C.; Nassehi, A.; Yon, J.; Hicks, B. Characterising the Digital Twin: A systematic literature review. *CIRP J. Manuf. Sci. Technol.* **2020**, *29*, 36–52. [CrossRef]
45. Liu, M.; Fang, S.; Dong, H.; Xu, C. Review of Digital Twin about concepts, technologies, and industrial applications. *J. Manuf. Syst.* **2021**, *58*, 346–361. [CrossRef]
46. Kaur, M.J.; Mishra, V.P.; Maheshwari, P. The convergence of Digital Twin, IoT, and machine learning: Transforming data into action. In *Digital Twin Technologies and Smart Cities*; Springer: Cham, Switzerland, 2020; pp. 3–17.
47. Tao, F.; Xiao, B.; Qi, Q.; Cheng, J.; Ji, P. Digital Twin modeling. *J. Manuf. Syst.* **2022**, *64*, 372–389. [CrossRef]
48. Haag, S.; Anderl, R. Digital Twin–Proof of concept. *Manuf. Lett.* **2018**, *15*, 64–66. [CrossRef]
49. Rasheed, A.; San, O.; Kvamsdal, T. Digital Twin: Values, challenges and enablers from a modeling perspective. *IEEE Access* **2020**, *8*, 21980–22012. [CrossRef]
50. Boschert, S.; Rosen, R. Digital Twin—the simulation aspect. In *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and Their Designers*; Springer: Cham, Switzerland, 2016; pp. 59–74.
51. Bao, L.; Wang, Q.; Jiang, Y. Review of Digital Twin for intelligent transportation system. In Proceedings of the 2021 International Conference on Information Control, Electrical Engineering and Rail Transit (ICEERT), Lanzhou, China, 30 October–1 November 2021; pp. 309–315.
52. Saroj, A.J.; Roy, S.; Guin, A.; Hunter, M. Development of a connected corridor real-time data-driven traffic Digital Twin simulation model. *J. Transp. Eng. Part A Syst.* **2021**, *147*, 04021096. [CrossRef]
53. Rudskoy, A.; Ilin, I.; Prokhorov, A. Digital Twins in the intelligent transport systems. *Transp. Res. Procedia* **2021**, *54*, 927–935. [CrossRef]
54. Zhang, K.; Cao, J.; Zhang, Y. Adaptive Digital Twin and multiagent deep reinforcement learning for vehicular edge computing and networks. *IEEE Trans. Ind. Inform.* **2021**, *18*, 1405–1413. [CrossRef]
55. Dasgupta, S.; Rahman, M.; Lidbe, A.D.; Lu, W.; Jones, S. A Transportation Digital-Twin Approach for Adaptive Traffic Control Systems. *arXiv* **2021**, arXiv:2109.10863.
56. Chen, D.; Lv, Z. Artificial intelligence enabled Digital Twins for training autonomous cars. *Internet Things Cyber-Phys. Syst.* **2022**, *2*, 31–41. [CrossRef]
57. Kumar, S.A.; Madhumathi, R.; Chelliah, P.R.; Tao, L.; Wang, S. A novel Digital Twin-centric approach for driver intention prediction and traffic congestion avoidance. *J. Reliab. Intell. Environ.* **2018**, *4*, 199–209. [CrossRef]
58. Wang, X.; Ma, L.; Li, H.; Yin, Z.; Luan, T.; Cheng, N. Digital Twin-Assisted Efficient Reinforcement Learning for Edge Task Scheduling. In Proceedings of the 2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring), Helsinki, Finland, 19–22 June 2022; pp. 1–5.
59. Xia, K.; Sacco, C.; Kirkpatrick, M.; Saidy, C.; Nguyen, L.; Kircaliali, A.; Harik, R. A Digital Twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. *J. Manuf. Syst.* **2021**, *58*, 210–230. [CrossRef]
60. Pandit, K.; Ghosal, D.; Zhang, H.M.; Chuah, C.N. Adaptive traffic signal control with vehicular ad hoc networks. *IEEE Trans. Veh. Technol.* **2013**, *62*, 1459–1471. [CrossRef]
61. Cao, M.; Li, V.O.; Shuai, Q. Book Your Green Wave: Exploiting Navigation Information for Intelligent Traffic Signal Control. *IEEE Trans. Veh. Technol.* **2022**, *71*, 8225–8236. [CrossRef]
62. Guo, Q.; Li, L.; Ban, X.J. Urban traffic signal control with connected and automated vehicles: A survey. *Transp. Res. Part C Emerg. Technol.* **2019**, *101*, 313–334. [CrossRef]

63. Jiang, X.; Jin, Y.; Ma, Y. Dynamic phase signal control method for unstable asymmetric traffic flow at intersections. *J. Adv. Transp.* **2021**, *2021*, 8843921. [CrossRef]
64. Maroto, J.; Delso, E.; Felez, J.; Cabanellas, J.M. Real-time traffic simulation with a microscopic model. *IEEE Trans. Intell. Transp. Syst.* **2006**, *7*, 513–527. [CrossRef]
65. Saroj, A.J.; Hunter, M.; Roy, S.; Guin, A. A Three-Tier Incremental Approach for Development of Smart Corridor Digital Twins. In Proceedings of the 2023 IEEE 26th International Symposium on Real-Time Distributed Computing (ISORC), Nashville, TN, USA, 23–25 May 2023; pp. 214–219.
66. Rudowsky, I. Intelligent agents. *Commun. Assoc. Inf. Syst.* **2004**, *14*, 14. [CrossRef]
67. Iyer, L.S. AI enabled applications towards intelligent transportation. *Transp. Eng.* **2021**, *5*, 100083. [CrossRef]
68. Schleiffer, R. Intelligent agents in traffic and transportation. *Transp. Res. Part Emerg. Technol.* **2002**, *10*, 325–329. [CrossRef]
69. Sadek, A.; Basha, N. Self-learning intelligent agents for dynamic traffic routing on transportation networks. In *Unifying Themes in Complex Systems: Proceedings of the Sixth International Conference on Complex Systems, Boston, MA, USA, 25–30 June 2006*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 503–510.
70. Roozemond, D.A. Using intelligent agents for pro-active, real-time urban intersection control. *Eur. J. Oper. Res.* **2001**, *131*, 293–301. [CrossRef]
71. Zhu, L.; Yu, F.R.; Wang, Y.; Ning, B.; Tang, T. Big data analytics in intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 383–398. [CrossRef]
72. Roozemond, D.A. Using autonomous intelligent agents for urban traffic control systems. In Proceedings of the 6th World Congress on Intelligent Transport Systems (Its), Toronto, ON, Canada, 8–12 November 1999.
73. Nguyen, T.T.; Nguyen, H.H.; Sartipi, M.; Fisichella, M. Multi-Vehicle Multi-Camera Tracking with Graph-Based Tracklet Features. *IEEE Trans. Multimed.* **2023**, *26*, 972–983. [CrossRef]
74. Nguyen, T.T.; Nguyen, H.H.; Sartipi, M.; Fisichella, M. Real-time multi-vehicle multi-camera tracking with graph-based tracklet features. *Transp. Res. Rec.* **2024**, *2678*, 296–308. [CrossRef]
75. Li, J.; Ma, H.; Zhang, Z.; Li, J.; Tomizuka, M. Spatio-temporal graph dual-attention network for multi-agent prediction and tracking. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 10556–10569. [CrossRef]
76. Palit, J.R. Application of Machine Learning and Deep Learning Approaches for Traffic Operation and Safety Assessment at Signalized Intersections. Master's Thesis, University of Tennessee at Chattanooga, Chattanooga, TN, USA, 2022.
77. Protogerou, A.; Papadopoulos, S.; Drosou, A.; Tzovaras, D.; Refanidis, I. A graph neural network method for distributed anomaly detection in IoT. *Evol. Syst.* **2021**, *12*, 19–36. [CrossRef]
78. Basmassi, M.A.; Boudaakat, S.; Chentoufi, J.A.; Benameur, L.; Rebbani, A.; Bouattane, O. Evolutionary reinforcement learning multi-agents system for intelligent traffic light control: New approach and case of study. *Int. J. Electr. Comput. Eng.* **2022**, *12*, 5519. [CrossRef]
79. Xu, M.; Wu, J.; Huang, L.; Zhou, R.; Wang, T.; Hu, D. Network-wide traffic signal control based on the discovery of critical nodes and deep reinforcement learning. *J. Intell. Transp. Syst.* **2020**, *24*, 1–10. [CrossRef]
80. Ge, H.; Gao, D.; Sun, L.; Hou, Y.; Yu, C.; Wang, Y.; Tan, G. Multi-agent transfer reinforcement learning with multi-view encoder for adaptive traffic signal control. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 12572–12587. [CrossRef]
81. Zhu, R.; Li, L.; Wu, S.; Lv, P.; Li, Y.; Xu, M. Multi-agent broad reinforcement learning for intelligent traffic light control. *Inf. Sci.* **2023**, *619*, 509–525. [CrossRef]
82. Jin, G.; Wang, M.; Zhang, J.; Sha, H.; Huang, J. STGNN-TTE: Travel time estimation via spatial–temporal graph neural network. *Future Gener. Comput. Syst.* **2022**, *126*, 70–81. [CrossRef]
83. *PTV Vissim 2022*; PTV Group: Karlsruhe, Germany, 2022. Available online: https://www.ptvgroup.com/en/solutionsproducts/ptv-vissim/ (accessed on 1 June 2023).
84. Harris, A.; Stovall, J.; Sartipi, M. Mlk smart corridor: An urban testbed for smart city applications. In Proceedings of the IEEE International Conference on Big Data, Los Angeles, CA, USA, 9–12 December 2019; pp. 3506–3511.
85. GDOT. *VISSIM Simulation Guidance*; FHWA-GA-21-1833, 18-33; Georgia Department of Transportation: Atlanta, Georgia, 2021.