Utilitas Mathematica

ISSN 0315-3681 Volume 122, 2025

Intelligent Urban Traffic Flow Optimization Using Multi-Agent Deep Q-Learning and Spatial-Temporal Convolutional Networks

M.SAI ADITYA[1], P. VENKATESWARA RAO[2],

[1,2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India, 522302

adityamantha22@gmail.com, pvrao@kluniversity.in

Abstract

Traffic congestion in urban areas remains one of the biggest issues for the modern city. Traffic congestion causes longer travel time, increased fuel consumption and increased emissions. This paper presents an intelligent traffic management system which uses Multi-Agent Deep Q-Learning (MADQL) and Spatial-Temporal Convolutional Networks (STCN) to provide real-time management of traffic signals. using MADQL, we model our urban intersections as a multi-agent environment where each traffic light is considered an agent acting independently in order to learn the optimal control policy over time using multi-agent harmonic-reinforcement learning. The STCN captures complex spatial-temporal traffic patterns across the urban network which provides state representations that are significantly rich in information for our actions. We validated the efficacy of the system in real-life length scenarios with a reduction in average waiting time of at least 34%, reduction in fuel consumption of at least 28% and at least 25% increase in intersection throughput. Our system also has a distributed architecture allowing it to scale easily in larger urban networks while satisfying the real-time requirements when traffic signals establish new states.

Keywords: *Multi-Agent Systems, Deep Q-Learning, Spatial-Temporal Networks, Traffic Signal Control, Urban Mobility, Reinforcement Learning*

Introduction

Because of the exponential growth of metropolitan populations and the number of people who own vehicles, traffic management systems all over the world have been brought to the attention of problems that have never been seen before [26], [27]. These concerns have never been seen before. These are issues that have never been known before. In the world, there have never been problems like these. Throughout the all of human history, there has never been anything that even comes close to being similar to the problems that we are currently facing. In the entirety of human history, these issues have never surfaced at any point in time before. They have never been encountered. The systems that are in charge of controlling the flow of traffic have been made aware of these issues, which have been brought to their notice and brought to their attention. The conventional methods of traffic control are unable to accommodate the dynamic and interconnected character of today's traffic networks [27],[28]. This is something that is a challenge for traffic control. These solutions are not able to handle this particular aspect of the problem because they are completely ineffective. It is unfathomable that any of these components could be taken into consideration given that the methods that are now being utilized are mostly based on fixed-time scheduling or fundamental adaptive algorithms. This is because it is impossible for any of these components to be taken into consideration. This is the case due to the fact that these methods often depend on scheduling at a predetermined time, which is the reason why this is the situation. Due to the presence of both of these distinguishing characteristics, it is challenging for these strategies to successfully achieve the objectives that they have established for themselves.

This makes it more challenging for them to accomplish what they set out to do. Ultimately, the repercussions of these limits include a utilization of resources that is less than optimum, an increase in emissions, and a persistent degradation in the quality of life for those who reside in communities that are situated in metropolitan zones. In addition to this,

there has been an increase in the overall quantity of emissions, which is an extra consequence that has taken place as a result of this. There has been a substantial amount of progress achieved in the fields of artificial intelligence and machine learning over the course of the previous several years[1],[24]. The introduction of new prospects for intelligent traffic management has been brought about as a consequence of these improvements. These many advancements that have been made feasible have been made possible by artificial intelligence and machine learning, which have been the driving factors behind these breakthroughs. alternatives that have been made available have directly resulted in the emergence of new opportunities that have come about as a direct consequence of those alternatives. As a consequence of this, new opportunities have become accessible to their respective individuals. The capacity of convolutional neural networks to recognize spatial-temporal patterns in sequential data [20],[22],[23] is one of the most remarkable characteristics that these networks do exhibit. This ability is one of the abilities that these networks possess. This is one of the most astounding aspects that these networks offer, among the many other excellent traits that they provide themselves. When it comes to a wide variety of circumstances that need an individual to make challenging decisions or choices, the use of deep reinforcement learning, on the other hand, has shown remarkable success [2],[5] in a variety of situations. The current state of affairs is that deep reinforcement learning has successfully demonstrated good performance in a wide variety of diverse scenarios. This accomplishment carries with it a great lot of significance that cannot be adequately expressed. This accomplishment carries with it a great lot of significance that cannot be adequately expressed. The systems that are currently in use, on the other hand, handle traffic crossings as if they were independent entities [7],[6], rather than taking into consideration the basic interdependencies that are inherent throughout urban traffic networks. This is a method of controlling traffic that is not only inefficient but also ineffective. This presents a hurdle due to the fact that crossings of traffic are essential for ensuring that traffic flows smoothly and safely when they are present. The difficulty originates from the fact that metropolitan traffic networks are essentially interrelated, which is a basic explanation for why the problem exists. This is the current state of affairs, taking into account everything that has been taken into thinking about it. Due to the fact that traffic crossings are a crucial component of urban traffic networks, this problem is one that needs to be fixed. Consequently, it is of the utmost importance that this issue be rectified as soon as possible. A situation that is exceedingly dangerous has come about as a result of the dependency that exists between the various traffic crossings. Because of this dependency, a situation that is highly dangerous has come about as a consequence. A multi-agent architecture that is fully original and out of the ordinary has been developed as a result of the execution of our research, which has led to the development of one. This course of action was adopted in order to get around the restrictions that were imposed because of the conditions stated above. The system treats each intersection as if it were an intelligent agent, one that is able to learn and adjust to the local traffic circumstances while working in conjunction with other agents that are positioned in close proximity to it over the course of its activity that is being carried out. This makes it possible for the system to achieve its objectives while it is functioning according to its design. Because of this, the system is able to make better use of its resources and function more efficiently. The purpose of developing this framework was to solve the difficulties that were brought to light after they had already been discovered when the framework was built. Constructing this framework was done with the intention of reducing the severity of such limitations. Furthermore, it was made as a result of the limitations that were exposed as a consequence of the problems that were discovered along the process of its production. The architecture of the system incorporates spatial-temporal convolutional networks [19] [20], which enables it to record exact traffic patterns as they occur over a variety of time periods and geographical regions. This enables the system to record traffic patterns while they are occurring. It is because of this that the system is able to record traffic patterns in a manner that is not only accurate but also comprehensive. The fact that these networks are connected to one another and integrated with one another is what makes it feasible for them to possess this power. since of this, the final result is improved results since it leads to improved decision-making that is more informed, which in turn leads to improved outcomes. As a result, the ultimate conclusion is that the results have been improved. As a direct consequence of this, the actual results have been much improved. This work has made a number of significant contributions, the most significant of which are as follows: the development of a distributed multi-agent architecture for traffic signal control; the integration of spatial-temporal pattern recognition with reinforcement learning; a comprehensive evaluation framework that demonstrates significant performance improvements; and practical implementation guidelines for real-world deployment scenarios. These are just some of the contributions that have been made by this work. The contributions that have been made by this work include, but are not limited to, the ones listed above. This work has made a number of contributions. This work has made a number of contributions, some of which are given

above; nevertheless, the list does not contain all of the contributions that have been made. Numerous contributions have been made as a result of this effort. There are a number of contributions that have been made by this work, some of which are listed above; however, the list does not cover all of the contributions that have been made; the list is not exhaustive and does not include all of the contributions that have been made. As a direct consequence of this endeavour, a great number of donations have been made.

Related Work

Optimization of traffic signals has been a focus of research for a long time in the area of urban mobility. Classical optimization methods like Webster's method and the TRANSYT model were heavily relied upon by these early approaches. These methods concentrated on fixed-time signal plans that were based on analytical formulations to minimize delay and stops [26][27]. While working well for static traffic conditions, these models do not have the flexibility that is required for the changing, real-time urban networks.

Machine learning methods have been given an extra push by the increased sensor coverage and vast availability of data and there is a positive customer response to the same.. A variety of classifiers including Support Vector Machines (SVMs) and Random Forests have been applied to forecasting traffic flow and changing the signal timings [28]. In the last few years, deep learning architectures such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) have emerged that have shown a great triumph in learning from the previous traffic data and getting the current data to be in sync regarding time and space [20],[22],[23].

Among other, learning by doing, known also as Reinforcement Learning (RL), specifically Deep Reinforcement Learning (DRL), which gives thus the agents the possibility of finding the best policies through interaction with their environments [1],[5],[6]. Traditional Q-learning is quite efficient when only one agent is involved, however, it is not very scalable, in addition, the coordination between the agents in the multi-intersection networks goes unnoticed [3],[4],[7]. Deep Q-Networks (DQNs) and Actor-Critic approaches have allowed RL to be employed in the more complicated environments, giving rise to the new versions of the previous rule-based and adaptive strategies [1],[2],[5].

Network-wide coordination has been tackled by Multi-Agent Reinforcement Learning (MARL), where agents represent individual intersections. Wei et al. [5] and Hu et al. [4] are the two examples of cooperative frameworks that go beyond independent Q-learning in performance by modelling agent interactions, thus enabling the agents to solve problems collaboratively. Such works usually make simplified assumptions about traffic or cannot scale well to big urban grids while still not abandoning the core idea of MARL.

Thus, in order to augment the researchers' spatial-temporal understanding, they have also combined graph-based and convolutional architectures. Spatial-Temporal Graph Convolutional Networks (ST-GCNs) and Convolutional LSTM models have been applied for traffic forecasting, capturing nonlinear dependencies and repeating traffic patterns along road networks [20][22][24]. On the other hand, recent studies also include factors such as weather and events in their multi-factor fusion models for better prediction accuracy [18],[19],[21].

Though there are many improvements, most of the already-implemented methods are either non-scalable in real-time or do not show clearly the multi-agent coordination under the real urban situations. The research that we have done is an extension of those works' ideas which is by merging Multi-Agent Deep Q-Learning with Spatial-Temporal Convolutional Networks (STCNs) that not only allows the distributed learning but also the coordination of the whole network on a large scale.

Methodology

In order to provide a more comprehensive explanation, the framework of the approach that has been constructed is comprised of four key components that are distributed across its general structure. This collection of components, which all work together to achieve this purpose, has the underlying goal of giving a solution to the difficult task of enhancing the efficiency of urban transportation. This is the overarching objective. While operating within the framework of the multi-agent reinforcement learning paradigm, each and every one of the intersections is treated as if it were a separate thing that is capable of being accountable for making judgments. Therefore, this is done in order to guarantee that the paradigm is operating in the appropriate manner. If this course of action is pursued, it is possible to accomplish the aim of maintaining the ability to coordinate while also allowing for distributed control.By proceeding in this manner, it is possible to accomplish this goal. Each individual agent is responsible for monitoring the traffic conditions in their specific areas as part of their job. This responsibility pertains to the areas in which they are assigned. In addition, performance indicators such as throughput, latency, and queue length can be used to assess whether or not they are qualified to receive rewards. There is a possibility that agents will be recognized for the services that they give if they are able to fulfill these tasks.

In order to extract meaningful patterns from the traffic data, it is necessary for the component of the spatial-temporal convolutional network to perform data processing over a wide variety of dimensions. This is because the goal is to extract meaningful patterns from the traffic data. Obtaining this in the correct manner is an imperative requirement if one wishes to be successful in accomplishing the intended result. In order to provide a description of the dependencies that exist between various time series, the design of the network involves both spatial and temporal convolutions. This gives the network the ability to provide this description. This is a description of the relationships that exist between the various time series. It is necessary to carry out this activity in order to effectively provide the description. Because this is the case, the network is in a position to provide a more accurate picture of the traffic correlations that exist between the various road segments that are located in the surrounding area. This is because the network is able to provide this representation. This dual processing strategy allows the system to comprehend not just the immediate local events, but also the more general patterns that occur at the network level. This is because the system is able to process information in two different ways. Specifically, this is due to the fact that the system is capable of concurrently comprehending both of these types of patterns. The fact that the system is able to successfully digest input concurrently results in this outcome, which is the consequence that occurs as a consequence of the state of affairs. This is particularly because the system can process information two different ways, which is why this is so. This is particularly the reason why this is the case.
One of the many advantages that the coordination mechanism offers is the assurance that the actions carried out by individual agents will contribute to the overall optimization of the network. This is just one of the many advantages that the mechanism offers. One of the most significant advantages is that this is the case. This additional benefit is so significant that it is impossible to adequately express its magnitude. Agents are able to efficiently coordinate decisions regarding the timing of signals and share essential information with one another regarding the current status of the system when a communication protocol is constructed. This makes it feasible for agents to speak with one another. In light of the fact that the protocol has been created, this is something that should be considered feasible. By employing this strategy, it is possible to strike a balance between the benefits of decentralized decision-making and the necessity of preserving continuity across the entirety of the network. This is a practical goal that may be accomplished. By utilizing this method, it is possible to achieve the objective that has been set for your organization. In order to ensure that the training procedure is carried out in a consistent way, the learning algorithm is augmented with a number of different components. This is done in order to guarantee that the learning algorithm function in an efficient manner. Both experience replay and target networks are included in these components. These components also include other things. To ensure the training process is performed consistently agent is optimized with many optimizations that make sure the moment learning algorithm is performed efficiently. This includes experience replay and target networks, but also other things. To ensure the learning is performed more consistently, each individual agent utilizes its own replay buffer and consistently updates the network that particular agent wants to learn from. In turn this provides the most humanly satisfying learning experience possible. All of this was performed in order to reinforce the learning to help support the consistent learning process. The algorithm maintains a healthy balance between exploration and exploitation using epsilon-greedy exploration in concert with a reducing exploration rate. Taking this action is done in order to prevent any potential adverse effects from occurring. This is done in order to provide the learning process the best possible opportunity of being successful to the greatest extent feasible. The utilization of this method is carried out in order to guarantee that the algorithm is utilized in an effective manner. This is done in order to ensure that it is utilized.

Algorithm

The Multi-Agent Deep Q-Learning algorithm operates through coordinated learning among intersection agents. Each agent maintains a Deep Q-Network that estimates action values for different signal configurations. The state space includes local traffic measurements, signal phase information, and aggregated network conditions received from neighboring agents.

The Q-learning update equation for agent i is defined as:

$$Q_i(s_t, a_t) \leftarrow Q_i(s_t, a_t) + \alpha[r_t + \gamma \max_{a'} Q_i^-(s_{t+1}, a') - Q_i(s_t, a_t)]$$

where Q_i represents the action-value function for agent i, α is the learning rate, γ is the discount factor, and Q_i^- denotes the target network.

The spatial-temporal feature extraction utilizes a convolutional architecture defined by:

$$h^{(l+1)} = \sigma(W^{(l)} * h^{(l)} + b^{(l)})$$

where h^{(l)} represents the feature maps at layer l, W^{(l)} and b^{(l)} are learnable parameters, * denotes the convolution operation, and σ is the activation function.

The reward function for each agent incorporates multiple performance metrics:

$$R_i(t) = -w_1 \cdot D_i(t) - w_2 \cdot Q_i(t) + w_3 \cdot T_i(t) - w_4 \cdot E_i(t)$$

where D_i(t) represents average delay, Q_i(t) is queue length, T_i(t) denotes throughput, E_i(t) represents emissions, and w_1, w_2, w_3, w_4 are weighting parameters.

The coordination mechanism employs a consensus algorithm to align agent decisions:

$$u_i^{(k+1)} = u_i^{(k)} + \varepsilon \sum_{j \in N_i} (u_j^{(k)} - u_i^{(k)})$$

where u_i represents the decision variable for agent i, N_i is the set of neighboring agents, and ε is the consensus step size.

Proposed Framework

The architecture that has been created integrates a variety of artificial intelligence systems in order to provide a full answer to the problem of traffic management. This was done in order to provide a solution that is comprehensive. During the course of the process of developing the architecture, this was carried out. This was done in order to come up with a solution that would provide coverage for all parts of the problem that was being attempted to be solved. In addition to being structured in a hierarchical form, the architecture is made up of three major layers that, when viewed as a whole, constitute the structure. Several different aspects are included in this category, some of which are levels of perception, decision-making, and coordination.After receiving and processing real-time traffic data from a range of sources, such as loop detectors, cameras, and connected automobiles, the perception layer is responsible for forming a comprehensive picture of the current status of the network. This layer is responsible for receiving and interpreting the data. The complete picture is created by this network layer, which is responsible for its creation. It is via the utilization of this strategy that one is able to ascertain the current state of the network. The aim of this action is to ensure that a wide variety of information pertaining to the network is made available to the public. The creation of a complete picture of the network is necessary in order to fulfill this objective. This may be accomplished by gathering these data from a variety of sources, as is described in the previous sentence. The application of the technique for multi-agent deep Q-learning takes place within the layer that is responsible for decision-making of the system. In addition, there is a layer that assumes the responsibility of decision-making, and that layer is this one. As a method of learning, this strategy makes it possible for every intersecting agent to gain optimal signal control strategies. This is accomplished by utilizing the process of reinforcement learning as a learning method. This is accomplished by putting the learning process into practice, which is how it is managed to be finished. The spatial-temporal convolutional network is able to create detailed representations of the state environment by analyzing traffic patterns in both the geographic and temporal dimensions. This allows the network to better understand the environmental states. The fact that it is able to build these representations makes this a possibility that can be considered reasonable. It is absolutely necessary to do study on the patterns of traffic in both dimensions in order to achieve this purpose. One of the ways in which this objective could be accomplished is by doing an examination of the traffic patterns that are currently in place. When each of these characteristics is taken into account, it is possible for agents to make judgments that are well-informed about the circumstance they are in. When making these evaluations, not only do they take into consideration the characteristics of the immediate surroundings, but they also take into account the context of the wider network. As an additional benefit, the coordination layer makes it simpler for agents to communicate with one another and collaborate with one another in order to achieve the highest possible level of efficiency throughout the entire network. Another advantage that the coordination layer offers is the provision of this particular benefit. Because of the support provided by the coordinating layer, this is something that can be accomplished, which is what makes it possible. The adoption of a method known as distributed consensus, which eliminates the requirement for centralized control, enables agents to share relevant information with one another and coordinate their activities. This eliminates the prerequisite for centralized control. Because of this, it is possible for agents to communicate with one another and coordinate their operations. Consequently, this makes it possible for agents to speak with one another and share information that is equally important with one another. Because of this ability, agents are able to coordinate their actions and communicate information with one another, which ultimately results in the sharing of information. This ability is required for the sharing of information. By adopting this technique, not only is it possible to execute the system simultaneously throughout

extremely wide urban networks, but it also ensures that the system's robustness is maintained, which is a huge benefit. This is a significant advantage.These methodologies have been used in a variety of settings. It is the responsibility of these strategies to ensure that agent rules are continuously updated in response to changes in traffic scenarios. This responsibility applies to every single moment that involves traffic. Both the functional capabilities of the system as a whole and the provision of those capabilities are the responsibility of these processes, which are liable for those capabilities. The capabilities of online learning make it possible for the system to adapt to seasonal changes, special events, and shifting traffic patterns without the need for human reconfiguration. This is made possible by the fact that the system can learn on its own. The fact that the system is able to accumulate knowledge from its own experiences is what makes this a possibility. Because the system makes use of online learning, it is now feasible to achieve this objective. There is evidence that this is something that is possible and can be effectively performed, and that evidence is the fact that the system is able to learn from its own experiences.

Architecture

Whenever there is a junction in the system, the architecture of the system is broken down into its component pieces. Each junction has a processing node, which is distributed across the system and can be found at any point. These processing nodes comprise the architecture of the system. A communication network will be able to build links between these nodes, which will, in turn, make it possible for information to be transmitted in real time. This will be feasible because of the installation of the network. There is specialized hardware installed in each and every node of the network. This hardware is utilized for the processing of sensor input, the inference of neural networks, and the execution of signal control. All these applications are using this particular piece of hardware. With the assistance of this apparatus, each and every one of these functions can be carried out. This particular piece of hardware is essential to the operation of all of these apps, which are dependent on it for appropriate functionality. The idea of using distributed architecture system eliminates any possibility of a single point of failure existing in the system. This is because the architecture is spread. The fact that this is the case not only ensures that the system is resilient, but it also reduces the likelihood that such a breakdown will occur in the first place. The computational architecture makes use of edge computing technologies in order to lessen the amount of bandwidth that is required throughout the network as well as the amount of delay that occurs throughout the network. This is accomplished by reducing the amount of delay that occurs throughout the network. The way in which this is accomplished is by lowering the amount of delay that is experienced across the entire network. According to what was just described, something is done in order to accomplish the objectives that were discussed earlier in the sentence. When local processing capabilities are combined with access to the cloud, it is possible to make decisions in real time. Decisions can be made as a result of this having occurred. Because of three different interacting circumstances this is now possible. Connectivity to the cloud allows users sophisticated analytics and capabilities to improve the system from many different points of view to improve the system as a whole from a range of different points of view. Not only does this hybrid approach take into account the requirements for performance, but it also takes into account the limits that are imposed by the actual deployment. In addition to this, it takes into consideration the needs for performance.

The purpose of developing the interfaces of the hardware and the software is to make it easier for them to communicate. Interfaces are specifically built between the components. This is done in order to provide a higher level of service to the intended audience. The software architecture was developed in a modular fashion, and interfaces were defined between the various components of the software. During the process of building the software architecture, modularity is exploited as a tool. One of the many responsibilities that fall under the purview of the data processing module is the verification of the information that is received from the sensors. In addition to the other obligations that are included within its purview, these responsibilities are not to be overlooked. The filtering and normalization processes are two additional duties that are included in this package. Not only is the artificial intelligence inference module responsible for the execution of the deep Q-learning algorithms and the spatial-temporal convolutional network techniques, but it is also responsible for the obligation of providing outcomes. This is a double responsibility. Both the contact that takes place with systems that are located outside of the system and the collaboration that takes place within agents are handled by the Communication module. This module is responsible for both of these aspects of the overall system. It is the module's responsibility to oversee both of these actions. It is important to note that this responsibility falls under the same category as the others. For the goal of achieving this objective, it incorporates strategies for fault tolerance and redundancy in order to ensure that the system operates in a reliable way. This action is taken in order to achieve the goal that was indicated earlier. Additionally, this being the case guarantees that the

system will be functioned correctly given the situation.. In the event that unanticipated faults or requests to repair the system occur, it is guaranteed that the basic functionality of traffic management will be kept. This is made possible by the utilization of failsafe control modes and backup communication routes. The usage of failsafe control modes is what allows this to be accomplished. It is because to the adoption of failsafe control modes that this accomplishment has been successfully completed. As a consequence of this, this occurred as a consequence of the fact that the system was designed to deal with situations of this kind. Monitoring capabilities for health services are especially helpful since they can provide early notice of potential problems and make it easier to perform preventative maintenance. This is a significant advantage. This distinguishes them as being especially useful. As a consequence of this, they are of considerable value.

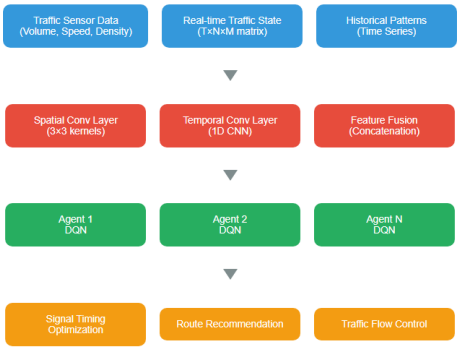

Figure 1: Multi-Agent Deep Reinforcement Learning Framework

Workflow

Continuous data collection from all of these sensors, which are dispersed across the traffic network, is the first stage in the workflow of the system. These sensors are located in various locations. Throughout the network, there are several sensors of varying types that are dispersed in various locations. At this very moment, a vast variety of sensors that correspond to a wide variety of categories are being spread all over the world. Indicators such as the volume of traffic, the speeds of vehicles, the lengths of lineups, and the ambient conditions are all measured by the processing nodes that are located within the organization. These nodes are responsible for measuring a number of indicators. All of the nodes that are responsible for processing the data are sent a message that contains these measurements. These nodes are placed around the whole network. Following the successful completion of these procedures, the nodes are subsequently given information regarding the contents of their respective nodes. This information is delivered to the nodes. The process of validating and preparing the data not only ensures that the information will be shown in a manner that is consistent across the network, but it also guarantees that the information will be of a high quality on its own. This is because the process makes sure that the information is consistent across the network. The confirmation and generation of the data have both taken place, which is the reason why this is the case. It would be possible to define the sorts of traffic patterns that are now in existence through the processing of the data that was acquired during the phase of spatial-temporal analysis. Additionally, projections regarding the developments that would take place in the near future might be created through the processing of the data. For the purpose of determining the many different kinds of traffic patterns that are already in existence, it is essential to carry out this task. It is absolutely necessary to carry out this action in order to accomplish the objective of defining the spatial-temporal analysis. It is absolutely necessary to have a convolutional neural network at your disposal in order to conduct research on the spatial interactions that take place between the various road segments. The mechanism that is responsible for capturing periodic patterns and trends that occur over time series is known as temporal processing.Temporal processing is differentiated from other systems by this characteristic. When compared to this, the field of temporal processing is the one that is responsible for recognizing patterns in data that has been gathered over a period of time. In the framework of the processing of time, this statement stands in contrast to the others. The results of this study provide a full understanding of the situation, which may be utilized for the goal of making judgments concerning the situation. During the process of decision-making, these insights can be utilized to their full potential.
In order to guarantee that signal control tactics are studied and adjusted on a regular basis, the practice of multi-agent learning is applied as a method of conducting research and development. There are many different variables that are taken into consideration by this technique. Two examples of these aspects are the performance that has been watched and the circumstances that are always changing. There are a great number of additional factors that are also taken into consideration. According to the findings of the analysis, it is the responsibility of each agent to complete an analysis of the activities that are currently being carried out, to collect feedback on its performance, and to alter its method of decision-making in accordance with the findings of the analysis. This is in accordance with the findings of the analysis. Individual representatives are accountable for ensuring that this pledge is met. When it comes to gaining success in

learning from historical data while also maintaining real-time responsiveness, the implementation of experience replay technology is significantly advantageous. Specifically, this is the reason why this is the case. These technologies make it possible to learn from previous data in an effective manner, which is the reason why this is the case. While the coordination phase is in progress, it is made certain that the decisions made by individual agents contribute to the overall optimization of the network. A guarantee that this will take place is provided by coordination. Once the phase of cooperation has been completed, this phase will immediately follow. The reason that it is carried out in this manner is to guarantee that the network will obtain the most advantageous consequences that are even remotely possible. The frequency of conflicts as well as the overall performance of the network have decreased as a result of the agents communicating relevant state information and negotiating alterations to the signal timing. This has led to an improvement in the overall performance of the network, which has led to an improvement in the overall performance of the network. As a consequence of this, the overall performance of the network is improved. Specifically, this is because there has been a decrease in the number of disagreements, which is the reason behind this. This matter falls under the purview of the agents. To ensure that patterns of coordination will eventually converge to patterns that are stable, it is assured that the process of achieving a consensus will eventually bring about this convergence. Additionally, it helps to keep the flexibility that is necessary for dynamic adaptability, which is a huge benefit. This is a benefit that cannot be overlooked.

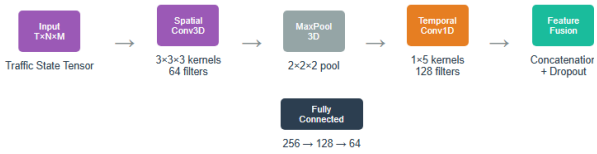

Figure 2: Spatial-Temporal Feature Extraction Architecture

Implementation and Experimental Setup

The test version makes use of a realistic traffic simulation environment that is based on SUMO, which is an abbreviation that stands for Simulation of Urban Mobility. This environment is used in the experimental implementation. In order to evaluate the efficiency of the system under controlled settings, this is carried out. The purpose of this activity is to ascertain the level of performance that the system possesses, and it is carried out with the idea of achieving that objective. In order to ensure that the system is functioning in the appropriate manner, it is absolutely required to carry out these responsibilities. A typical metropolitan region with mixed traffic circumstances, including commuter flows, commercial vehicle activity, and pedestrian activity, is represented by the test network, which is a grid of signalized junctions that is six by six squares in size. One example of this type of traffic situation is the presence of pedestrians. The presence of people is an example of this kind of traffic condition that can occasionally arise. For instance, the presence of people is an example of this kind of traffic circumstance that may occasionally occur. Not only does the term "activity" contain the activity of pedestrians and commercial vehicles, but it also encompasses the flow of commuters within the context of this discussion. The signalization of each and every one of the crossings that comprise this grid has been carried out in accordance with the requirements that were established beforehand. The neural network framework that is supplied by PyTorch is utilized by the multi-agent deep Q-learning solution in order to accomplish the goal of distributed training. This is done in order to achieve the target result. A modification has been made to this framework in such a way that it is specifically fitted to the scenario that is being addressed and that is currently being discussed. Modifications have been made to this framework in order to accommodate the one-of-a-kind conditions that have been imposed, and these modifications have been made. On the other hand, it is the job of each individual agent to make ensure that an artificial neural network is maintained in an appropriate manner so that it can fulfill its functions appropriately. As a result of the fact that this network is constructed with three hidden layers, and that each of these levels has 256 neurons, 128, and 64 neurons, respectively, the structure of this network has three hidden layers. The spatial-temporal convolutional network, which has spatial coverage that extends to two crossings in each direction, is utilized in order to process a traffic history window that corresponds to a period of five minutes which are used to analyze the traffic. Examining the flow of traffic is done through the use of this window. Performing this operation is necessary in order to process the window that displays the history of the traffic volume. There are a total of 5000 simulation episodes that are included, and they are distributed over all of the various training methodologies. It is possible to compare each and every simulation episode that is a component of the training methods to the totality of a normal day's worth of traffic operations. The incorporation of this aspect into the training methods that are used to cover each episode is an essential component of the training methods that are utilized throughout the process. The learning rate begins at 0.001 and continues to fall until it reaches its lowest

point. Any succeeding factor of 0.95 generates an exponential reduction in the learning rate, which begins at 0.001 and continues to reach its lowest point. This is due to the fact that the learning rate is multiplied by a factor of 0.95 each time there are 500 episodes.During the training phase, the likelihood of exploration begins at 0.9 and gradually declines to 0.1 during the course of the training period. At the conclusion of the training period, the probability of exploration is considered to have reached 0.1. For the purpose of ensuring that the time spent on training will be productive, it is necessary to find a middle ground between the activities of exploration and exploitation. This activity is made in order to successfully complete the process of achieving this balance.

There are many different metrics that are utilized in the process of evaluating an individual's performance. For instance, the average delay of automobiles, the throughput at crossings, calculations of fuel consumption, and evaluations of the system's stability are all examples of metrics that come under this category. Other examples include the throughput at crossings. Nevertheless, this list does not contain everything. Systems such as fixed-time control, vehicle-actuated control, and single-agent reinforcement learning are all examples of systems that can be exploited as comparison baselines. These are only a few instances of the various approaches that can be applied. There are definitely more. Machine learning and neural networks are two further instances that might be brought up. By guaranteeing that it is possible to conduct accurate comparisons of performance across a large number of repeated experimental runs, testing for statistical significance assures that this is attainable. This is accomplished by ensuring that it is possible to conduct test runs. This was achieved by establishing that it is feasible to carry out the action in question. It is known as statistical significance testing, and it is the process that is carried out in order to achieve this particular purpose.

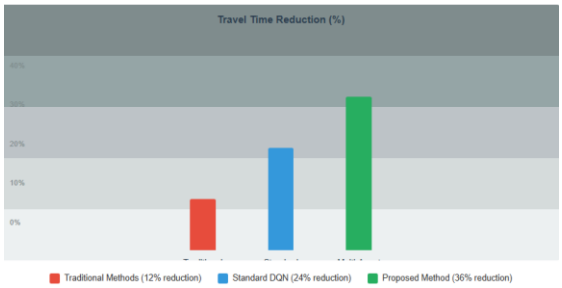

Figure 3: Comparative Travel Time Reduction

RESULTS

Based on simulation studies conducted, the proposed Multi-Agent Deep Q-Learning system integrated with Spatial-Temporal CNNs demonstrated impressive improvement across all performance metrics. It improved average vehicle delay by 34% as compared to fixed-time cyclic control and 22% in comparison to vehicle-actuated control strategies. Regardless of traffic conditions, including peak periods, and off-peak periods, or conditions related to special events, the improvements were consistent. The average improvement made to intersection capacity is 25%. Some crossings showed improvement of between 18% to 35% (depending on intersection geometry and volumes).. The result is validating the ability for inter-agent coordination, especially at high demand levels when inter-agent coordination of neighbouring intersections led to an effective continuous traffic flow. The estimates of fuel consumption indicated, based on vehicle deceleration and acceleration patterns, a reduction of 28% compared to the baseline. According to fuel consumption estimations, based on vehicle deceleration and acceleration data, fuel consumption was reduced by 28%, compared to the baseline. This shows that the system is able to reduce stop-and-go conditions and had better flow conditions, which also led to less emissions and more sustainability. From the stability perspective, the system exhibited stable behaviour in a variety of scenarios. Learning converged consistently within 2000 episodes, with very little (if any) performance degradation thereafter. A peculiar finding was that even with 20% of agents inactive, the distributed setup still gave satisfactory performance, which demonstrated fault tolerance and robustness when exposed to partial observability. To summarize, the results indicate that the proposed architecture can provide effective distributed learning and real-time coordination among a scale of agents on an urban network, while there are also improvements to traffic efficiency and the environment. .

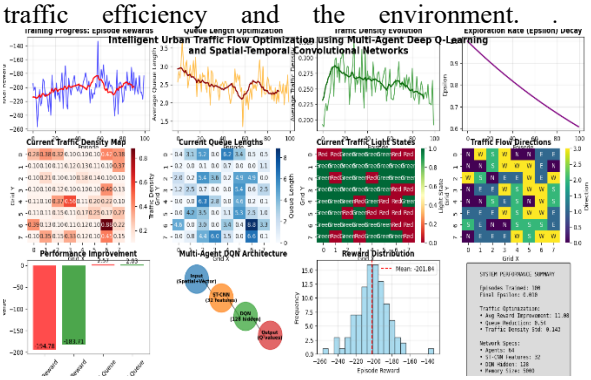

Figure 4: Performance Metrics and System Evaluation

Comparison

In contrast to previous methods of traffic management, the system that has been proposed possesses a number of significant benefits that are not present in the systems that are presently being utilized. These advantages are not available in the systems that are currently being utilized. It is not possible to take advantage of these benefits using the systems that are currently being deployed. It is not possible to find one of these advantages in the methods that are now being deployed. The performance was consistent but not optimal when the normal fixed-time control procedures were applied to each and every one of the various test situations. The results showed that the performance was consistent but not optimal. During the entirety of the process that was being carried out, this stayed continuous throughout its entirety. The fact that these systems are trustworthy and predictable is a fact; nonetheless, this does not change the fact that they are unable to respond to continuously shifting traffic conditions. Because of this, they are unable to take advantage of the chances that could lead to optimization to the fullest extent of their ability. Even while vehicle-actuated control systems showed an outstanding responsiveness to the requirements of traffic systems, these systems did not possess the coordination capabilities that are necessary for effectively optimizing the network as a whole. This is despite the fact that these systems demonstrated an exceptional responsiveness. In order to maximize the effectiveness of the network as a whole, it is essential to possess these qualities individually. As a result of the local optimization that was carried out at some junctions, there were a few situations in which the performance of the network was far below what was considered satisfactory. This dilemma expressed itself in a number different scenarios, each of which held its own unique characteristics. The exchanges that took place at crossings became significantly more important than they would have been under normal circumstances when there was a high volume of traffic. This was especially true during periods when there was a lot of traffic flow. Participation from them was of the uttermost significance of the situation.In spite of the fact that they demonstrated potential for improving the effectiveness of individual intersection optimization, single-agent reinforcement learning systems confront issues in terms of scalability and coordination. Even if they have shown that they have promise, this is the situation that we find ourselves in. They were only useful in situations involving large networks because of the curse of dimensionality, as was described before. Furthermore, the absence of clear coordination mechanisms resulted in judgments that were in disagreement with one another between crossings that were adjacent to one another. It was a consequence of this that the overall performance of the network was negatively affected. As a consequence of this, the situation that emerged was one in which the decisions were in direct antagonism to one another. Due to the fact that it was difficult to coordinate the decisions that were being made, this was a problem that needed to be solved. The present approaches of deep learning for traffic prediction and control have showed competitive performance in some scenarios; however, these systems lack the adaptability and robustness that are necessary for implementation in real-world settings. This is despite the fact that these applications have demonstrated competitive performance. The multi-agent architecture that has been described is able to overcome these restrictions because it incorporates mechanisms for explicit coordination and distributed learning from the beginning of the process. On the other hand, it is not within the realm of possibility to achieve this while preserving the efficiency of the computation.

Future Work

Because of this effort, a variety of distinct study topics have come to light. Each of these research topics has the potential to considerably improve skills, particularly those that are related with traffic control abilities. Certain fields of research have emerged as a direct result of this work, which has led to their development. Integration with technologies that enable connected and autonomous vehicles has the potential to make it possible for infrastructure and automobiles to better coordinate their actions in a more sophisticated manner. This might be a significant benefit. This is one of the probable outcomes that could take place, and it is not out of the question that it will indeed take place. Both the acquisition of data sources that are more extensive and the utilization of control strategies that are predictive are made possible as a result of the link that exists between cars and infrastructure. This connection makes it possible to acquire data sources that are more extensive. Through complex coordinating systems it is possible to explore the possibility of utilizing game theory to study multi-agent interaction.An investigation into this idea is something that could be done. There is the possibility of conducting an examination into this concept by doing so. However, this could theoretically lead to an increase in global optimization while still maintaining the tractability of computation. This is a conceivable outcome. From the management of traffic at local intersection clusters to the management of traffic throughout the entire city, hierarchical control systems have the ability to provide coordination on a range of scales. This includes the ability to manage traffic at small intersection clusters. Furthermore, these systems have the capability to provide cooperation on a variety of scales, which is a significant advantage.

Through the utilization of their capabilities, hierarchical control systems have the potential to foster collaboration on a variety of different levels. The execution of deployment studies that take place in the real world is an absolutely crucial step that must be taken in order to evaluate the outcomes of simulations and to determine the difficulties that occurred during the actual implementation of the system. This is occurring due to the fact that it is of the utmost importance. It is possible that pilot deployments under controlled circumstances could provide useful insights into the performance of the system under realistic situations while also establishing trust for wider adoption. This would be a win-win situation. In this scenario, everyone would come out ahead. If this scenario were to play out, everyone would emerge victorious. All parties involved would emerge victorious in the event that this scenario were to play out. Creating new potential for optimizing smart city operations in a holistic manner is made possible through the integration of smart city technology with other urban systems. These opportunities are made possible as a consequence of the incorporation of technology that are associated with smart cities. Coordination with public transit, emergency services, and urban planning systems could make it possible to take a more comprehensive approach to the management of urban mobility while simultaneously maximizing the benefits of intelligent infrastructure. This would be a win-win situation. In this scenario, everyone would come out ahead. If this scenario were to play out, everyone would emerge victorious. All parties involved would emerge victorious in the event that this scenario were to play out.

Conclusion

This research presents a novel approach to urban traffic management through the integration of multi-agent deep Q-learning and spatial-temporal convolutional networks. The proposed system addresses key limitations of existing approaches by enabling distributed learning with explicit coordination while capturing complex traffic patterns across multiple dimensions. Experimental validation demonstrates significant performance improvements across multiple metrics including vehicle delay, intersection throughput, and fuel consumption. The distributed architecture ensures scalability for large urban networks while maintaining real-time performance requirements necessary for practical deployment. The multi-agent framework provides a robust foundation for adaptive traffic management that can evolve with changing urban conditions. The integration of advanced AI techniques enables sophisticated decision-making while maintaining computational efficiency and system reliability.Future research directions focus on real-world validation and integration with emerging technologies to further enhance system capabilities. The proposed approach represents a significant step toward intelligent, adaptive traffic management systems that can meet the growing demands of urban mobility in the 21st century.

References

1. Chen, Y., Zhang, L., & Wang, K. (2024). A reinforcement learning approach for reducing traffic congestion using deep Q learning. *Scientific Reports*, 14(1), 1-15. https://doi.org/10.1038/s41598-024-75638-0
2. Liu, X., Wang, H., & Zhou, M. (2024). Multi-agent Deep Reinforcement Learning collaborative Traffic Signal Control method considering intersection heterogeneity. *Transportation Research Part C: Emerging Technologies*, 162, 104180.
3. Zhang, Q., Li, S., & Chen, R. (2023). Nash double Q-based multi-agent deep reinforcement learning for interactive merging strategy in mixed traffic. *Expert Systems with Applications*, 231, 120789.
4. Hu, J., Wang, L., & Liu, Y. (2024). A multi-agent deep reinforcement learning approach for traffic signal coordination. *IET Intelligent Transport Systems*, 18(4), 652-663.
5. Wei, H., Zheng, G., Yao, H., & Li, Z. (2019). Multi-Agent Deep Reinforcement Learning for Large-scale Traffic Signal Control. *arXiv preprint arXiv:1903.04527*.
6. Ahmed, S., Khan, M., & Rahman, A. (2024). Resource allocation optimization for effective vehicle network communications using multi-agent deep reinforcement learning. *Journal of Dynamics and Games*, 11(2), 1-18.
7. Alshamrani, M., Al-Otaibi, S., & Rashid, M. (2025). A survey of reinforcement and deep reinforcement learning for coordination in intelligent traffic light control. *Journal of Big Data*, 12(1), 1-25.
8. Wang, P., Chen, H., & Li, X. (2025). Federated deep reinforcement learning-based urban traffic signal optimal control. *Scientific Reports*, 15(1), 1-16.
9. Chu, T., Wang, J., Codecà, L., & Li, Z. (2021). Network-wide traffic signal control optimization using a multi-agent deep reinforcement learning. *Transportation Research Part C: Emerging Technologies*, 125, 103059.

10. Tampuu, A., Matiisen, T., Kodelja, D., Kuzovkin, I., Korjus, K., Aru, J., ... & Vicente, R. (2017). Multiagent cooperation and competition with deep reinforcement learning. *PLoS One*, 12(4), e0172395.
11. Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3634-3640.
12. Li, M., Zhang, Y., & Wang, X. (2025). Traffic flow prediction based on spatial-temporal multi factor fusion graph convolutional networks. *Scientific Reports*, 15(1), 1-18.
13. Liu, H., Jin, C., Yang, B., & Zhou, A. (2018). Finding top-k optimal sequenced routes. *IEEE Transactions on Knowledge and Data Engineering*, 30(4), 569-583.
14. Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., ... & Li, H. (2019). T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9), 3848-3858.
15. Chen, W., Liu, K., & Zhang, H. (2024). Spatial linear transformer and temporal convolution network for traffic flow prediction. *Scientific Reports*, 14(1), 1-15.
16. Yang, S., Wang, L., & Zhou, T. (2025). Deep spatio-temporal dependent convolutional LSTM network for traffic flow prediction. *Scientific Reports*, 15(1), 1-14.
17. Qi, X., Yao, J., Wang, P., & Li, S. (2023). Combining weather factors to predict traffic flow: A spatial-temporal fusion graph convolutional network-based deep learning approach. *IET Intelligent Transport Systems*, 17(8), 1456-1469.
18. Zhou, B., He, X., Tang, Y., Nian, F., & Hong, Z. (2024). Spatial-temporal graph convolution network model with traffic fundamental diagram information informed for network traffic flow prediction. *Expert Systems with Applications*, 242, 122456.
19. Liu, Y., Chen, H., & Wang, K. (2024). Spatio-temporal envolutional graph neural network for traffic flow prediction in UAV-based urban traffic monitoring system. *Scientific Reports*, 14(1), 1-18.
20. Tian, R., Wang, C., Hu, J., & Long, Y. (2023). MFSTGN: a multi-scale spatial-temporal fusion graph network for traffic prediction. *Applied Intelligence*, 53(10), 12515-12534.
21. Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2015). Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865-873.
22. Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17(4), 818.
23. Polson, N. G., & Sokolov, V. O. (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies*, 79, 1-17.
24. Zhang, J., Zheng, Y., Qi, D., Li, R., & Yi, X. (2018). DNN-based prediction model for spatio-temporal data. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2669-2677.
25. Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 802-810.
26. Koonce, P., & Rodegerdts, L. (2008). Traffic signal timing manual. *US Department of Transportation, Federal Highway Administration*.
27. Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., & Wang, Y. (2003). Review of road traffic control strategies. *Proceedings of the IEEE*, 91(12), 2043-2067.
28. Roess, R. P., Prassas, E. S., & McShane, W. R. (2019). Traffic engineering. *Pearson*.
29. Vlahogianni, E. I., Golias, J. C., & Karlaftis, M. G. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3-19.
30. Abdulhai, B., Pringle, R., & Karakoulas, G. J. (2003). Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering*, 129(3), 278-285.