Check for updates

Wiley
Journal of Advanced Transportation
Volume 2024, Article ID 3035753, 15 pages
https://doi.org/10.1155/2024/3035753

WILEY

Research Article

Improving the Urban Transport System Resilience Through Adaptive Traffic Signal Control Enabled by Decentralised Multiagent Reinforcement Learning

Xiangmin Yang,[1] **Yi Yu**,[2] **Yuxiang Feng**,[1] **and Washington Yotto Ochieng**[1]

[1]*Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK*
[2]*Shanghai AI Laboratory, Shanghai 200232, China*

Correspondence should be addressed to Xiangmin Yang; xiangmin.yang18@imperial.ac.uk

Received 12 March 2024; Revised 7 August 2024; Accepted 16 October 2024

Academic Editor: Kun Wang

Copyright © 2024 Xiangmin Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The principle of system resilience is its ability to withstand disruptions and maintain an equilibrium state. In urban network systems, adaptive traffic signal control (ATSC) has been an effective countermeasure to mitigate traffic flow disturbance and improve resilience. This research has explored the usage of a decentralised advantage actor-critic (a2c) algorithm-based ATSC in mitigating disruptions, particularly nonrecurring congestion caused by car accidents. A reward function has also been proposed, combining deduced resilience metric, safety indicator time to collision (TTC) and system performance. A virtual simulation environment was created using simulation of urban mobility (SUMO) to facilitate the evaluation of the proposed approach. In the grid simulation environment, an overall 5.8% improvement is achieved, exceeding benchmark algorithms in three metrics, especially performance with a margin of over 5.2%. Robustness against different levels of car accidents are proven as well. Further evaluation is also implemented based on a real-world case study and demonstrates an improvement of 20.08%, highlighting the correlation of proposed method's efficiency on the traffic flow rate and road structure.

Keywords: adaptive traffic signal control; multi-agent reinforcement learning; resilience

1. Introduction

Due to the rapid development of globalisation, the world has become more interconnected and thus more vulnerable to disruptions [1]. In response to this increased vulnerability, the term "resilience" has then been raised to describe the ability to tolerate disruptions [2]. Furthermore, four key characteristics of resilience have been identified in the field of critical infrastructure [3]:

I. Robustness, which describes the ability to withstand disruptions while maintaining a certain level of functionality.

II. Redundancy, which describes the capacity for alternative options, choices and substitutions during emergency.

III. Resourcefulness, which describes the ability to allocate resources during emergency.

IV. Rapidity, which describes the speed to restore a new equilibrium state after disruption.

In summary, characteristics I and IV focus on the ability of the system to deal with unexpected disruption from a control perspective, while characteristics II and III aim to increase the capacity of the system to deal with the emergency from a design perspective.

Emergency has been a well-researched area for a long period, but the investigation of unexpected disruption has just emerged in recent years [4–6]. The main difference between emergency and disruption lies in the severity of consequence, while emergency may pose an imminent danger to life or property, disruption will cause some

applications of the system inoperable for an unacceptable period of time [7]. Despite disruptions might not have lethal consequences, they will still lead to severe financial losses and even hazards due to their high frequency and difficulty to be accurately predicted [5]. Following the pace of technical development, studies of resilience progress gradually on the agenda.

In this research, ways of constructing resilient cities have been explored from a traffic control perspective using advanced multiagent reinforcement learning (MARL). The characteristics I and IV of urban traffic network (both proved essential [8–10]) are our main focus, which can be concluded as two objectives, increasing traffic system's ability to withstand disruptions and maintain an equilibrium state.

Utilizing the emerging technology adaptive traffic signal control (ATSC) to improve road network resilience has been proven as an effective countermeasure [3], as the junction controller in ATSC can resolve disruptions by evaluating the traffic state and changing the control policy.

Nonrecurring congestion originated from car accidents has been identified as the major disruption in this research. To simulate the effect of disruptions in traffic flow and facilitate evaluating the performance of the ATSC, an interactive simulation environment can be created using the simulation of urban mobility (SUMO) [11].

In summary, the first objective of resolving urban network disruptions can be achieved through ATSC. The second objective, which is maintaining stability, naturally directed towards the usage of reinforcement learning (RL) due to its robustness to randomness and reaching an equilibrium state [12]. Considering the traffic signals are physically dispersed and a natural multiagent system, a decentralised multiagent actor-critic (a2c) algorithm is used in this research that enables each junction to make separate decisions based on its observation, while still being coordinated to achieve a global goal.

Two simulation experiments have been conducted in this research. The first experiment is a $5 * 5$ grid with 25 junctions. It serves the purpose of validating the necessity and effectiveness of the proposed method in a simulated environment. On the other hand, the second experiment uses a real road network from Xiaoshan district, Hangzhou city, with 31 junctions, which aims to demonstrate the practical applicability and performance of the proposed method in real life with the usage of real data.

In this research, the main contribution includes the following:

I. Propose a MARL-based ATSC method that can increase resilience of urban road networks against disruptions, particularly nonrecurrent congestion caused by car accidents. The proposed method outperforms fixed-time and traditional ATSC strategies, while achieving better results compared to benchmark algorithms in three metrics, especially performance with a margin of over 5.2%.

II. Improve traffic resilience through bespoke reward function containing deduced resilience metric for training. The proposed metric shows a 4.58% improvement against conventional reward function.

III. Present an overall improvement of 5.8% for the proposed method in the grid simulation, with real-world feasibility demonstrated by an improvement of 20.08% in a real network, highlighting the correlation of proposed method's efficiency on the traffic flow rate and varying levels of car accidents.

This paper has been structured into six sections. Section 2 reviews the related work about ATSC. Section 3 summarises the methodologies of this research. Section 4 introduces the proposed approach to simulate disruptions and presents the adopted algorithm. Section 5 presents the detailed settings of both experiments and demonstrates their results. Section 6 summarises our findings and sheds light on future works.

2. Related Work

In urban cities, traffic control is widely applied to improve traffic efficiency and equality. ATSC is one of the essential components of intelligent transportation systems (ITSs). Using a real-time adaptation in response to the changing traffic network dynamics, better traffic control performance can be achieved.

2.1. Traditional ATSC Methods. Famous ATSC techniques such as the split cycle offset optimisation technique (SCOOT) [13] and Sydney coordinated adaptive traffic (SCAT) [14] utilize real-time traffic information collected from detectors to optimize the phase split, cycle length, and offset of the junction. The cycle time is the duration to complete all phases of a signal; phase split is the division of the cycle time into periods of green signal for competing approaches; offset is the time relationship between the start of each phase among adjacent intersections [15].

The controller in the junction evaluates the state in the intersection and then decides whether to extend the current phase or switch to another one to improve the efficiency of the intersection. Two main ATSC methods, Equisaturation policy [16] and P0 policy [17], calculate the red or green time based on flows or delays. Equisaturation policy is the most widely adopted signal-setting method, whose objective is more "fair" than "efficient." In Equisaturation policy, the red time split red_j on lane j is calculated as follows:

$$\mathrm{red}_j = 1 - \frac{x_j t}{s_{aj}}, \quad j \in \{1, 2, \ldots, m\}, \mathrm{red}_j \in [0, 1], \tag{1}$$

where x_j is the flow on edge j, s_{aj} is the saturation flow on lane j, $(t = 1/(x_1/s_{a1}) + (x_2/s_{a2}) + \cdots + (x_j/s_{aj}))$, and m denotes the number of incoming lanes in a junction. Through the above equation, the adaptive controller could rectify the red split to achieve traffic flow equilibrium. Nevertheless, such ATSC methods face challenges for incorporating multiple objective functions and lack of coordination among different junctions despite numerous attempts at upgrades [18–22].

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2. RL and Deep Reinforcement Learning (DRL) for ATSC. Due to distributed and autonomous nature of the traffic signal control [23], self-learning methods can be used to tune its performance. Early in 2003, Adbulhai first introduced model-free reinforcement learning Q-Learning to two-phase intersection with of ATSC [24]. Following that, the authors in [25] have explored the choice of reward function under Q-Learning for ATSC and three have been summarised as follows:

i. Queue length
ii. Cumulative delay
iii. Throughput

As each junction performs its own actions (change traffic light state) and can affect its neighbours (in the form of traffic flow), the urban network system is by nature a multiagent system. As a result, many researchers have employed multiagent approach under Q-Learning framework [26, 27]. However, challenges remained as the state-action space become enormous, especially with a large number of agents [10].

Fortunately, the successful combination of deep learning and RL (DRL) has pushed this area forward [28]. Two popular DRL algorithms (deep Q-Learning and a2c) have been effectively utilised in ATSC to improve junction performance and increase the scalability to train numerous agents simultaneously [3, 10, 29].

Based on DRL training algorithms, additional technical features could be added to ATSC to perform more accurate observation and enable other desirable functionalities. The vision-based observation and prediction of traffic flow have been successfully deployed in a dynamically changing environment [30]. A communication functionality that enables ATSC to exchange observation is also developed by [29], which enables the real-time coordination of junctions in the urban network. In this research, both functionalities have been employed in the experiment environment, as well as the ability to identify traffic accidents [31].

2.3. Advanced Features Powered by DRL for ATSC. Recent advancements in ATSC have focused on incorporating additional objectives and constraints beyond just minimizing traffic delays. Zhang et al. [32] proposed a multiobjective DRL approach that concurrently optimizes safety, efficiency, and decarbonization at intersections. Their duelling double deep Q network (D3QN)-based ATSC algorithm outperformed traditional methods in reducing traffic conflicts and carbon emissions. Incorporating safety considerations into ATSC is another important research direction. Du et al. [33] proposed SafeLight, a RL method that incorporates road safety standards as enforcement to ensure the safety of existing RL methods, aiming towards operating intersections with zero collisions. Their safety-enhanced residual RL method employed multiobjective loss functions and reward shaping for better knowledge integration. Ahmed et al. [34] introduced a control strategy based on intersection capacity, formulated from available space at discharge routes. Their deep Q-Learning agent utilizes density and speed measurements to manage signalized junctions and has demonstrated effectiveness in mitigating operation at a global scale in a simulated real urban traffic network.

The increasing prevalence of connected vehicles (CVs) and crowd sourcing has also opened up new opportunities for ATSC. Tan and Yang [35] developed a privacy-preserving ATSC method using partially CV data. Their approach leverages secure multiparty computation and differential privacy to aggregate individual-level CV data for calculating key traffic parameters while protecting the privacy of CV users. Mishra et al. [36] developed an ATSC system for developing countries using fused parameters derived from crowd-sourcing data. Their method leverages the economical collection and sharing of traffic data enabled by mobile technologies.

The usage of ATSC to increase resilience has been proven practicable. In Shang et al.'s study [3], different ATSC-controlled policies trained by deep Q-Learning have shown resilience against generalised percentage loss of functionality. In this research, out of 3 specific urban network disruptions ATSC has demonstrated capable of mitigating [10]; nonrecurring congestion is studied in detail for retrieving resilience of the system.

3. Methodology

3.1. Resilience. While no consensus has been reached on a unified metric for measuring system's resilience [37], a common way of quantifying the system's resilience is to evaluate its ability to cling to a stable condition under disturbance [9]. One of the commonly used resilience metrics ϕ_{common} [38–40] has been adopted, which can be denoted as follows:

$$\phi_{\text{common}} = \int_{t_e}^{t_r} P(t)dt, \tag{2}$$

where t_e and t_r are the time of disruption and the time of reaching new equilibrium, respectively. $P(t)$ is the performance of a system at given time t [37].

A detailed illustration of equation (2) is demonstrated in Figure 1. The resilience of the system is mainly determined by the loss of performance P through time, and the ability to reduce the loss of performance is considered as resilience characteristic I. On the other hand, as stated in characteristic IV, reaching equilibrium is quite essential so that large gain of performance is also not desirable. Therefore, a high resilience is achieved through maintaining the performance of the system unchanged through time and thus a new resilience metric ϕ can be deduced based on the absolute value of fluctuation of the system performance (so that both gain and loss can be included) as follows:

$$\phi = -\sum_{t=1}^{T} \left| \left(P_{t+1} - P_t \right) \right|, \tag{3}$$

where $|(P_{t+1} - P_t)|$ stands for fluctuation of performance within each time step t and T stands for the period of interest. A minus sign is added so that less fluctuation leads to higher resilience.

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

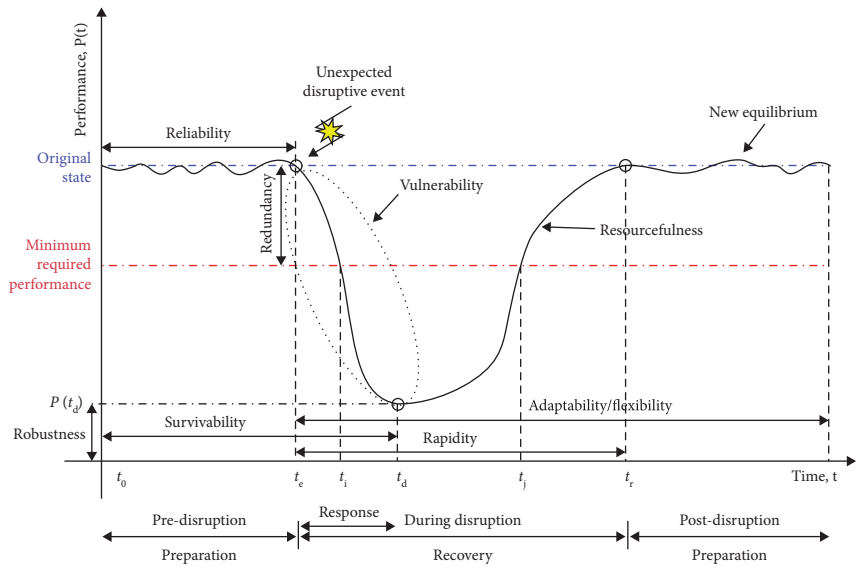


FIGURE 1: Illustration of equation (2) [9].

Moreover, the queue length within each junction has been used as a primary performance metric in urban network systems [41]. It has been proven effective to measure resilience [42] and been commonly used as the reward function for ATSC RL training (as shown in 2). Thus, average queue length $\overline{q}$ at a junction was selected for the performance evaluation P in this research as follows:

$$P = -\overline{q} = -\frac{1}{m}\sum_{j=1}^{m} q_j, \tag{4}$$

where m is the total number of lanes connecting to the intersection, j is the lane index, and q_j measures the number of vehicles waiting in lane j. The minus sign is added as fewer waiting vehicles stand for better performance.

3.2. Safety Index. As the urban network disruption explored in this research is nonrecurring congestion from car accidents, a metric to evaluate the rear-end collision risk, time to collision (TTC), is imported in this research as follows [43]:

$$\mathrm{TTC}_{\mathrm{ego}}(t) = \begin{cases} \min \dfrac{\left|X_{\mathrm{ego}}(t) - X_{\mathrm{su}}(t)\right| - l_{\mathrm{ego}}}{\left|V_{\mathrm{ego}}(t) - V_{\mathrm{su}}(t)\right|}, & \mathrm{if}\left(X_{\mathrm{ego}}(\mathrm{t}) - X_{\mathrm{su}}(\mathrm{t})\right) * \left(\mathrm{V}_{\mathrm{ego}}(\mathrm{t}) - \mathrm{V}_{\mathrm{su}}(\mathrm{t})\right) < 0, \\ \infty, & \text{otherwise}, \end{cases} \tag{5}$$

where ego denotes the vehicle of interest and su represents all surrounding vehicles. X is a two-dimensional vector denoting the location of the corresponding vehicle, while V stands for the velocity. l_{ego} is the length of the vehicle.

The incorporation of this metric can reduce collision risk and limit the secondary consequence of system disruption. As most vehicles choose to change lanes when encountering blockage in their lane (due to car incidents), the collision risk coming from the lateral direction also poses a serious threat. As a result, for each vehicle, TTC in both longitudinal and lateral directions will be calculated and the minimum value will be used as the safety indicator.

3.3. Decentralised MARL. RL is favoured for achieving resilience in urban network systems due to its capacity to interact with the stochastic environment and its high resistance to randomness [12]. While conventional MARL often assumes a centralized learning process with global information available to all agents, decentralized MARL allows each agent to make decisions based on local observations and limited communication with neighbouring agents. This decentralized approach is particularly suitable for traffic signal control, where global information may be impractical to obtain in real-time and local decision-making is crucial for rapid response to changing traffic conditions.

As discussed in Section 2, implementing the MARL in ATSC is more suitable as each agent (in this case traffic signal) will need to perform different actions locally due to the heterogeneously distributed nature of traffic flow. The direct approach to extending single-agent deep RL to a multiagent system is to let each agent learn independently, while considering other agents as part of the environment. Such approaches are usually computationally expensive, and

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

thus an alternative of centralised learning and decentralised execution is utilised, which is called decentralised MARL [44].

Under the decentralisation scenario, each agent will choose the most optimal policy based on its observation, while having a global common goal, and the local observation is often varied by extent. For example, the algorithm proposed by [45] assumes a globally available state, action and reward for each agent but chooses to keep each policy private. Such settings can be made to match the real circumstance for a more accurate simulation.

In the context of traffic signal control, decentralised MARL is particularly advantageous due to the inherently distributed nature of intersections in a road network. Each intersection (agent) makes decisions based on local traffic conditions and limited information from neighbouring intersections, mirroring the real-world constraints of traffic management systems. This approach allows for scalable and adaptive control strategies that can respond quickly to local disruptions while still working towards global traffic optimization.

Decentralised MARL has been increasingly used in networked system control such ass traffic signal control [46] and connected autonomous vehicle control [47]. Different algorithms have also been proposed to demonstrate the usage of decentralised MARL in networked systems [48] but fail to account for the difficulty of gathering global observation from vastly distributed agents in these scenarios. As a result, the NeurComm communication protocol proposed by [29] is used in this research. It only allows communication of observations between neighbouring agents, and the propagation of information through time steps fits perfectly with the nature of traffic flow. The selection of this specific protocol will be further justified in Section 5.

The NeurComm protocol is applied on the a2c algorithm, which consists of an actor neural network for each agent i, standing for the policy π_{θ_i} parameterized by θ_i, and a critic network, representing state value V_{ω_i} parameterized by ω_i. Both networks are updated by loss functions as follows:

$$\mathscr{L}(\theta_i)_{\text{policy}} = -\log \pi_{\theta_i}\left(a_{i,t} \mid h_{i,t}\right)\left(R_{i,t}^{\pi} - V_{\omega_i}\left(h_{i,t}, a_{\mathscr{N}_i,t}\right)\right), \tag{6}$$

$$\mathscr{L}(\theta_i)_{\text{entropy}} = \pi_{\theta_i}\left(a_i \mid h_{i,t}\right)\log \pi_{\theta_i}\left(a_i \mid h_{i,t}\right), \tag{7}$$

$$\mathscr{L}(\omega_i)_{\text{value}} = \left(R_{i,t}^{\pi} - V_{\omega_i}\left(h_{i,t}, a_{\mathscr{N}_i,t}\right)\right)^2, \tag{8}$$

where N_i stands for the neighbouring agents of agent i and t stands for the step time. $R_{i,t}$ is the sampled action value and estimated state value $V_{\omega_i}(h_{i,t}, a_{\mathscr{N}_i,t})$ takes the input of hidden state $h_{i,t}$ and neighbouring agents' action $a_{\mathscr{N}_i}$.

The hidden state $h_{i,t}$ of agent i is generated based on the observation $\tilde{s}_{i,t}$ and previous hidden state $h_{i,t-1}$ through a LSTM architecture.

$$h_{i,t} = \text{LSTM}\left(h_{i,t-1}, \tilde{s}_{i,t}\right). \tag{9}$$

The observation $\tilde{s}_{i,t}$ includes its own state $s_{i,t}$, together with messages from its neighbouring agents $m_{\mathscr{N}_{i,t}}$.

$$\tilde{s}_{i,t} = \left(s_{i,t}, m_{\mathscr{N}_{i,t}}\right). \tag{10}$$

Messages $m_{i,t}$ sent towards neighbouring agents include its states $s_{i,t-1}$, policies from last step $\pi_{i,t-1}$ and hidden state from last step $h_{i,t-1}$; thus, all its neighbours receive the same message.

$$m_{i,t} = \left(s_{i,t-1}, \pi_{i,t-1}, h_{i,t-1}\right). \tag{11}$$

4. Proposed Method

In most urban network simulation experiments, disruptions are only generalised as a fixed percentage loss of functionality [3, 49]. This approach fails to demonstrate the different consequences brought by specific disruptions. Aslani et al. [10] has demonstrated the ability of RL-based ATSC to mitigate three types of network disruptions: impatient pedestrian, nonrecurring congestion from incidents and noisy sensors.

In this research, nonrecurring congestion originating from car incidents are chosen to be further explored. A fully interactive simulation environment is desirable, in which not only disruptive incidents can be reproduced but also their influence on the system performance can be monitored. As a result, the widely used open source traffic simulator (SUMO) [11] is utilised to create an interactive urban network with ATSC.

4.1. Nonrecurring Congestion Scenario. Nonrecurring congestion has been widely used as the source of disruption to test traffic network's resilience [50] and is generally attributed to car accidents [10]. Unexpected car accidents have a serious impact on the traffic flow [51] and can even lead to gridlock due to the random shocks it is creating [10]. To accurately simulate the nonrecurring congestion in traffic networks, it is essential to reproduce a reasonable amount of car accidents.

In the simulated grid network, the assumption that the number of accidents within a fixed interval N_{accident} follows Poisson distribution [52] has been adopted as follows:

$$N_{\text{accident}} \sim \text{Pois}\left(\lambda_{\text{accident}}\right), \tag{12}$$

where $\lambda_{\text{accident}}$ is the expected value of N_{accident}. Each episode of RL training is taken as the fixed interval and the number of accidents N_{accident} is generated at the beginning of each episode based on equation (12) N_{accident} of time steps are randomly selected from all the time steps of that episode and marked as accident steps.

In the real network experiment, accident steps are decided based on a Monte Carlo method, bootstrapping the representative frequency of car accidents in a chosen region. This could lead to higher accident frequency as the simulated real network is only part of the chosen region. However, as the purpose of this research is to increase the resilience of the system against disruptions, higher accident frequency poses

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

stronger threat towards the system, which can better demonstrate the ability of the proposed method to maintain functionality. In line with this approach, the benchmark value of $\lambda_{\text{accident}}$ in the simulated network is also calculated based on the real data.

At the beginning of each accident step, an accident simulation command f_{accident} will be executed. A vehicle within the scenario is randomly selected to enter the accident state. A setSpeed command is then assigned to this vehicle which will force it to halt immediately to simulate a blockage of the lane caused by an accident.

Depending on the accident location and space for lane changing, some following vehicles will also be forced to halt. This can simulate the severity of accidents, as more vehicles can be affected by the accident. Vehicles behind the blockage point will then take an overtaking action thus reducing the output rate of the traffic flow on this edge. After a certain period, the halting vehicle will resume its journey at the next edge on the route plan, which indicates the accident is resolved.

The benefit of this approach is the consequential effect on traffic flow will entirely depend on the driving mode of surrounding vehicles, which is closer to real-life scenarios. In a multiagent system like an urban network, each vehicle shall perform its action based on its own target and observation. By changing one agent's state, the nonrecurring congestion can be observed in the form of other agents' cumulative behaviour.

In order to increase the cooperation among traffic signals to alleviate secondary effects on neighbouring junctions, a functionality that detects occurrence of traffic accidents is added [31]. An extra accident status $A_{i,t}$ is added to the state information $s_{i,t}$ and encoded in the communication messages $m_{i,t}$. A value of 1 is assigned to accident status if it detects traffic accident(s), while 0 stands for no car accidents on the edge. The number of accidents will not be recorded in the accident status.

$$A_{i,t} = \begin{cases} 1, & \text{if } \exists \text{ accidents,} \\ 0, & \text{if } \nexists \text{ accidents.} \end{cases} \tag{13}$$

4.2. Algorithm. In summary, a decentralised multiagent a2c algorithm is utilised in a SUMO urban network environment with networked ATSC and nonrecurring congestion. The agents are defined to be the traffic signals at the junctions.

The actions of each agent $a_{i,t}$ are determined at the beginning of each time step t based on policy $\pi_{\theta_i}(a_{i,t} \mid h_{i,t})$, i.e., the probability of taking action $a_{i,t}$ given $h_{i,t}$. Under the ATSC scenario, such action would be choosing the traffic signal state for next time step. The exact action space is determined by the scenario setting.

The state $s_{i,t}$ of each agent i includes the arithmetic mean of the traffic flow $f_{j,t}$ and accident status $A_{i,t}$ (as defined in equation (13)) on lanes j it controls. m stands for the number of controlled lanes.

$$s_{i,t} = \left(\frac{1}{m} \sum_{j=1}^{m} f_{j,t},\ A_{i,t} \right). \tag{14}$$

The state $s_{i,t}$ contributes towards the observation $\tilde{s}_{i,t}$ of each agent i, as demonstrated in Section 3.3, combining with the message $m_{N_i,t}$ received from the neighbouring agents N_i. The hidden state $h_{i,t}$ is then determined based on observation $\tilde{s}_{i,t}$.

The local reward $r_{i,t}$ of each agent i is calculated at the end of each training step t. The reward consists of three parts, namely, performance P, resilience ϕ and safety S.

As stated above, the average queue length $-\overline{q_j}$ over lanes each agent controls is the primary performance metric. The resilience ϕ is developed based on the proposed resilience metric and further adjusted to suit MARL training as follows:

$$\phi = \begin{cases} -\left|P_t - P_{t-1}\right|, & \text{if } t > 1, \\ P_t, & \text{if } t = 1, \end{cases} \tag{15}$$

where t is the episode time step.

Similarly, safety S is decided based on TTC. Since TTC could reach as high as infinity, the improvement in average TTC of vehicles within lanes controlled by each agent can be trivial. Another metric is then proposed, which calculates the percentage of vehicles under a certain TTC threshold, which is decided to be 10 in this research, representing less than 10 seconds before potential collision.

In summary, the local reward $r_{i,t}$ of each agent i at time step t is expressed as follows:

$$r_{i,t} = P_i + (-\alpha * S_i) + \beta * \phi_i, \tag{16}$$

where α and β are scaling factors that ensure P_i, S_i and ϕ_i are in the same magnitude. A minus sign is put in front of S_i as it is desirable to decrease the percentage of vehicles below the threshold. The scaling factors α and β are determined through a series of preliminary experiments and sensitivity analyses. It should be acknowledged that these values may need to be adjusted for different urban environments or specific policy priorities.

The cooperation between local agents is facilitated through the NeurComm communication protocol and the use of a global reward r_t. Each agent shares its observations and policies with neighbouring agents, allowing for coordinated decision-making. The global reward $r_t = 1/n\sum_{i=1}^{n} r_{i,t}$, calculated as the arithmetic mean of all local rewards across n agents, encourages agents to act in the best interest of the entire system rather than optimizing only for their local intersection. Global reward r_t is introduced in the calculation of sampled action value $R_{i,t}$ both in equations (6) and (8) as follows:

$$R_{i,t} = r_t + \gamma V_{\omega_i}\left(h_{i,t}, a_{\mathcal{N}_i,t},\right) \tag{17}$$

where γ is the discount factor that controls the trade-off between immediate reward r_t and future reward $V_{\omega_i}(h_{i,t}, a_{\mathcal{N}_i,t})$. As a result, the usage of global reward r_t

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ensures the policy network π_θ is updated towards the best interest of the entire system.

To summarize, the proposed method, both include the simulation of nonrecurring congestion, as well as calculating critical inputs for the a2c algorithm, is presented in Algorithm 1 Pseudocode and visualised in Figure 2.

5. Experiments and Results

Two experiments have been conducted in this research. The first one is a $5 * 5$ grid with 25 junctions controlled by ATSC capable of communicating with its neighbouring junctions.

In the second experiment, we selected the region with the highest accident density in Xiaoshan district, Hangzhou city, as our simulation environment. This region has a high density of intersections (31 intersections) and variable traffic patterns, making it representative of urban environments. The experiments share the same scenario settings as the first one.

The proposed algorithm is developed using TensorFlow. The experiment is conducted on a workstation (CPU: 16xi9-9900K@3.60, GPU: 2xNVIDIA-GeForce-RTX-3090, RAM: 64 GB). ReLU is used as the activation function for the neural network.

During the training simulation, the length of each episode is set to be 3600 s, equivalent to 1 h. The length of each training step is set to be 5 s to prevent the traffic signals from changing too rapidly, resulting in a total of 720 steps in a single episode. Each time a green light turns into a red light, 2 s of yellow light will be inserted in between to alert vehicles to slow down.

5.1. Grid Simulation

5.1.1. Scenario Setting. As shown in Figure 3, the grid consists of 16 identical blocks. Each block is presented as a rectangular with 200 m vertical edges and 150 m horizontal edges. Each edge contains two lanes that can be used to move forward or turn to corresponding directions at the junctions (junction details available at the centre of Figure 3). The speed limit of each lane is set to 20 m/s. During simulation, each junction will only be able to communicate with junctions that are directly connected (as shown in Figure 2).

Every junction, in consequence, is identical to each other and thus same action space is shared among all the junctions. For each edge, three types of green phase are available, including straight-through, left-turn and right-turn. Three combinations of three green phases are selected and added into the action space:

i. straight-through + right-turn + left-turn (yield to straight-through)
ii. left-turn + right-turn
iii. straight + right-turn

which leads to a 6-dimensional action space illustrated in Figure 3. Combinations with only one phase are excluded as it leads to lower efficiency.

For each junction on the sides, a 75 m extension edge is inserted and served as the origin and destination of traffic flow. At every second, a vehicle will be generated at the probability of 15% at each extension edge, equally choosing 25 extension edges as its destination. Therefore, an expected value of 540 vehicles will enter each extension edge during a full episode.

At the beginning of each episode, the number of accidents N_{accident} will be generated according to stats.poisson. The duration of accidents in each episode will be randomly selected from 900, 950, 1000, 1050, 1100, 1150 and 1200 s.

To test the robustness of the proposed method, $\lambda_{\text{accident}} \in (0, 1, 2, 4, 6, 8)$ are evaluated concurrently. The results for $\lambda_{\text{accident}} \in (2, 4, 8)$ are used to compare with fixed time (FT) signal control and one traditional ATSC strategy adopted in SUMO [53].

As stated above, $\lambda_{\text{accident}} = 4$ deducted from the real data is chosen as the benchmark. Under the benchmark, the proposed method is compared with policies trained by other MARL algorithms including multiagent deep deterministic policy gradient (MADDPG) [54] and differentiable inter-agent learning (DIAL) [55] and also the policy trained by the same algorithm used in the proposed method but with only performance as its reward function.

5.1.2. Results. A total of one million training steps have been conducted for each test run, except for the FT signal control and traditional ATSC, which only last for 30 episodes due to no long term improvement. Under the benchmark $\lambda_{\text{accident}} = 4$, the proposed method has reached a 5.85% optimisation for overall reward.

FT control has been widely used in real-life traffic control [56] and thus becomes a perfect benchmark for the proposed method. With FT control, each phase of the traffic signal has a fixed duration. In this experiment, 30 s have been assigned to each phase of the junction shown in Figure 3. Two s of yellow light are inserted when changing from green light to red light.

Similar to SCAT, the ATSC strategy adopted by SUMO uses detectors installed on each lane to decide the length of each phase. If the gap between two successive vehicles last less than 3 s, the current phase is prolonged until longer gap is detected or the maximum duration 60 s is reached.

The comparison of both FT and traditional ATSC with the last 30 episodes of the proposed method is demonstrated in Figure 4. It can be easily deduced that both methods are much easier to be affected by random disruptions, as the standard deviation of all three metrics is much larger than the proposed method under all three scenarios. As the accident frequency increases, both methods tend to pose a higher standard deviation, which indicates severe disruptions.

In terms of the median, the proposed method outperforms both traditional methods except for safety. This should be attributed to the focus of traditional ATSC, which aims to decrease the sudden brake between consecutive vehicles with a small gap, leading to a higher safety index. However, the results still emphasise the inability of

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for t in episode step **do**
  generate accident step;
  select accident duration;
  if t in accident step **then**
    execute f_{accident}
  end
  for i in n agents **do**
    calculate $s_{i,t} = (f_{i,t}, A_{i,t})$;
    send $m_{i,t} = s_{i,t}, h_{i,t-1}, \pi_{i,t-1}$;
    receive $m_{\mathcal{N}_{i,t}}$;
    generate $\widetilde{s}_{i,t} = (s_{i,t}, m_{\mathcal{N}_{i,t}})$;
    generate $h_{i,t} = \text{LSTM}(h_{i,t-1}, \widetilde{s}_{i,t})$;
    generate $\pi_{i,t} \sim h_{i,t}$;
    execute $a_{i,t} \sim \pi_{i,t}$;
    calculate $r_{i,t} = P_i + (-\alpha * S_i) + \beta * \phi_i$
  end
  calculate $r_t = 1/n \sum_{i=1}^{n} r_{i,t}$
end

Algorithm 1: Pseudocode of one simulation episode.

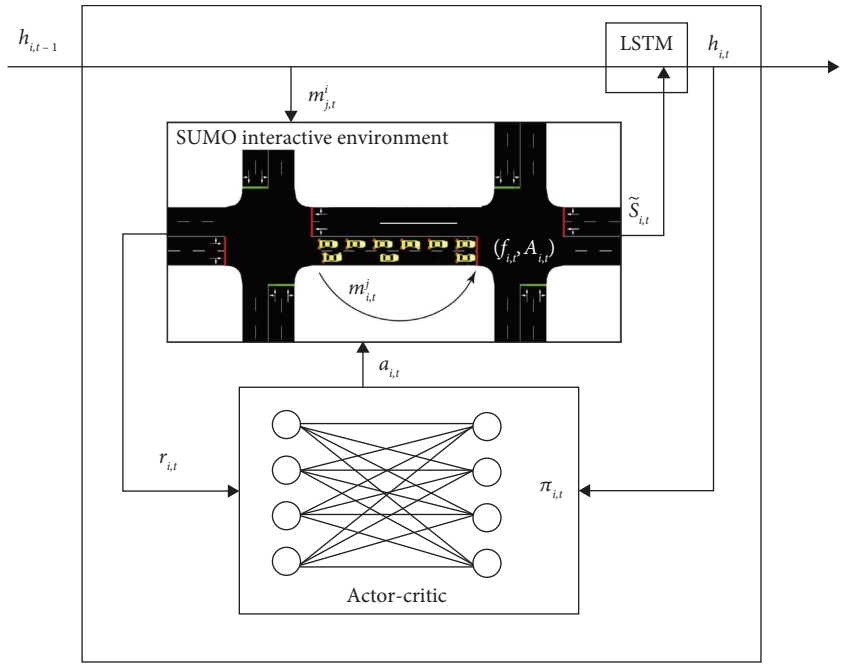


Figure 2: Algorithm architecture.

traditional ATSC to achieve multiple objectives, and the necessity of utilizing the proposed method is thus further verified.

On the other hand, the proposed reward function is tested against the widely used average queue length [25], and the result is presented in Figure 5. A better resilience is achieved with a margin of 4.58%. This validates the adoption of the proposed reward function for achieving better resilience.

Two well-established MARL algorithms that have been commonly used in ATSC training [29, 57] are imported as the benchmark: MADDPG [54], which holds no communication functionality, and DIAL [55], which uses a different communication protocol other than NeurComm [29].

The result is demonstrated in Figure 6, and it is observed that algorithms with communication mechanisms generally outperform the one without. The proposed method outperforms the other two algorithms in all three metrics, especially in performance P, with a margin of 6.04% with MADDPG and 5.21% with DIAL.

Furthermore, to test the robustness of our proposed method against different levels of car accidents, $\lambda_{\text{accident}} \in (0, 1, 2, 4, 6, 8)$ are tested separately. The results are

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

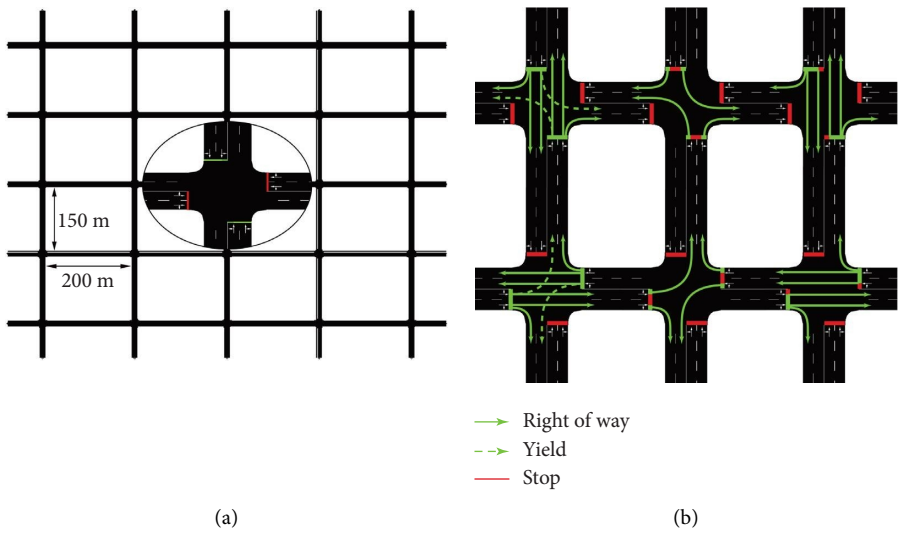


FIGURE 3: Grid scenario setting. (a) $5 * 5$ Traffic grid. (b) Action space of each junction.

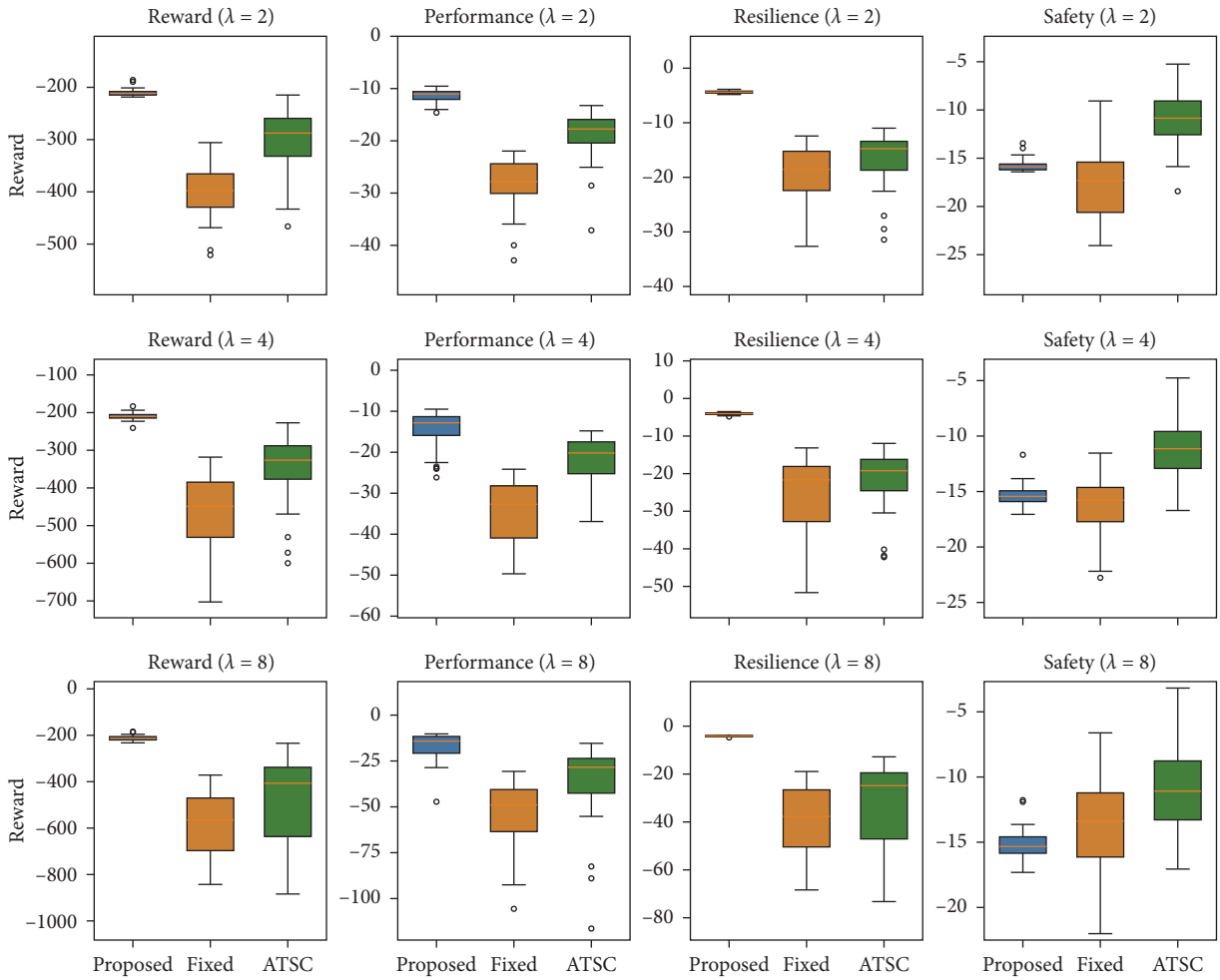


FIGURE 4: Comparison of evaluation metric results with FT and traditional ATSC in grid scenarios.

illustrated in Figure 7, and the proposed method is proved to withhold accident level as high as $\lambda_{\text{accident}} = 8$. It is reasonable to observe that a higher accident level leads to worse performance P, as more blockage of the road results in lower traffic flow. However, it is also noteworthy that the proposed method is capable of achieving almost the same resilience R

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

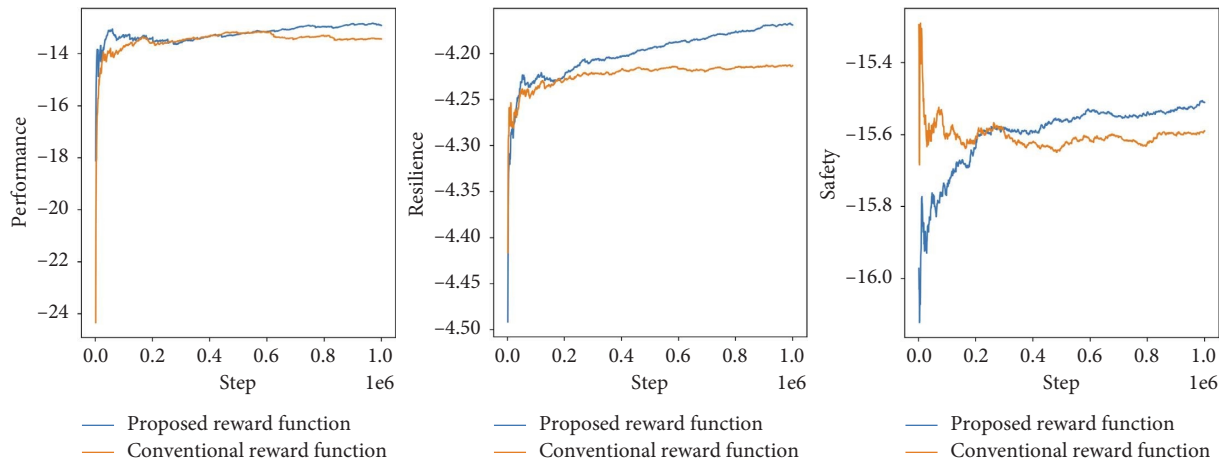


Figure 5: Comparison of evaluation metric results for different reward functions in the grid network scenario.

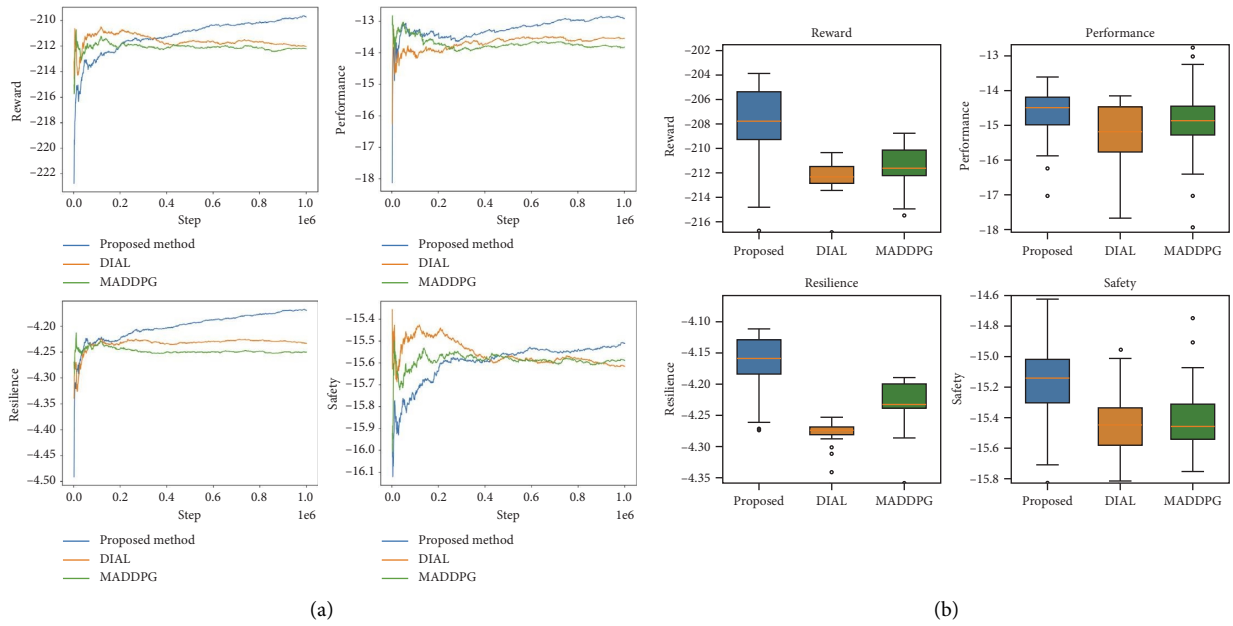


Figure 6: Comparison of evaluation metric results for different algorithms in the grid network scenario. (a) Entire training epoch. (b) Last 30 episodes.

regardless of the accident level, while better safety S are presented under a higher accident level.

5.2. Real Net Experiment

5.2.1. Scenario Setting. In this section, the proposed method is applied to an urban road network with actual traffic data to verify the approach's feasibility in a real-world environment. The traffic flow data collected from the real-time automatic number plate recognition (ANPR) system and accident data recorded by traffic police are imported to create a real-life environment.

The accident data, recorded from 2020/8/1 to 2020/10/31 (91 days in total), were collected in Hangzhou, China. By projecting the place of accidents onto the map (demonstrated as the blue points in Figure 8), one region with the highest density of accidents is selected. The region's exact location within the district is marked as a small rectangular in Figure 8, with a detailed road network shown on the side.

The road network covers approximately 1.46 km^2, consisting of 31 junctions. Each edge contains one or two lanes that can be used for left, right, forward and/or U-turns. Deriving from the real traffic flow data, the vehicle inflow rate is set to be an expected value of 325 vehicles per hour. Unlike the grid simulation, each junction has a unique action space, which has a dimension ranging from 2 to 4, based on its connection feature.

To decide the accident step, a Monte Carlo method is employed here. The accident data are divided into sets of hours, the same length as each episode. At the beginning of each episode, one set of data will be bootstrapped, and the exact minutes of accidents will be transferred into

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

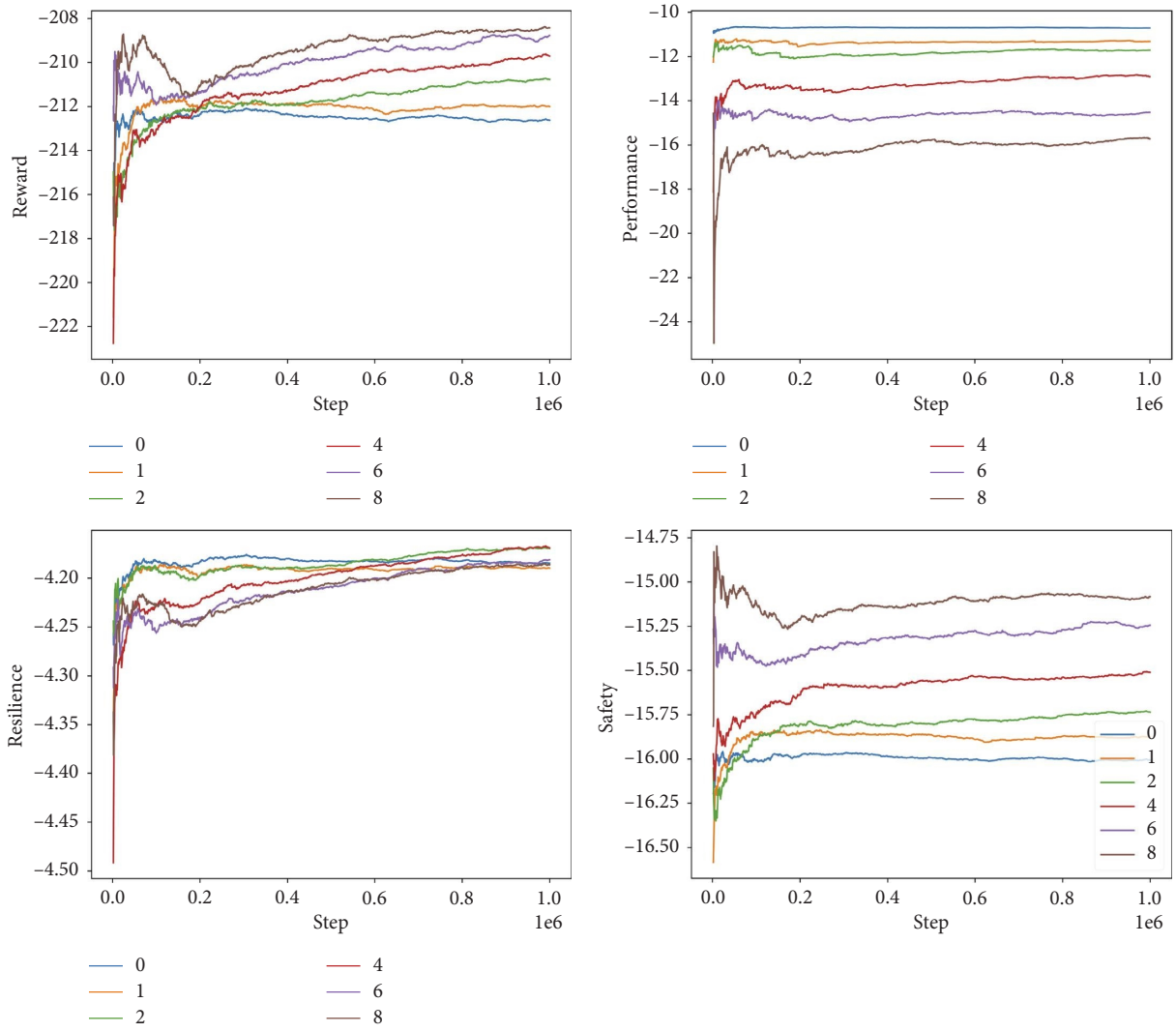


Figure 7: Comparison of evaluation metric results for different $\lambda_{\text{accident}}$ values in the grid network scenario.

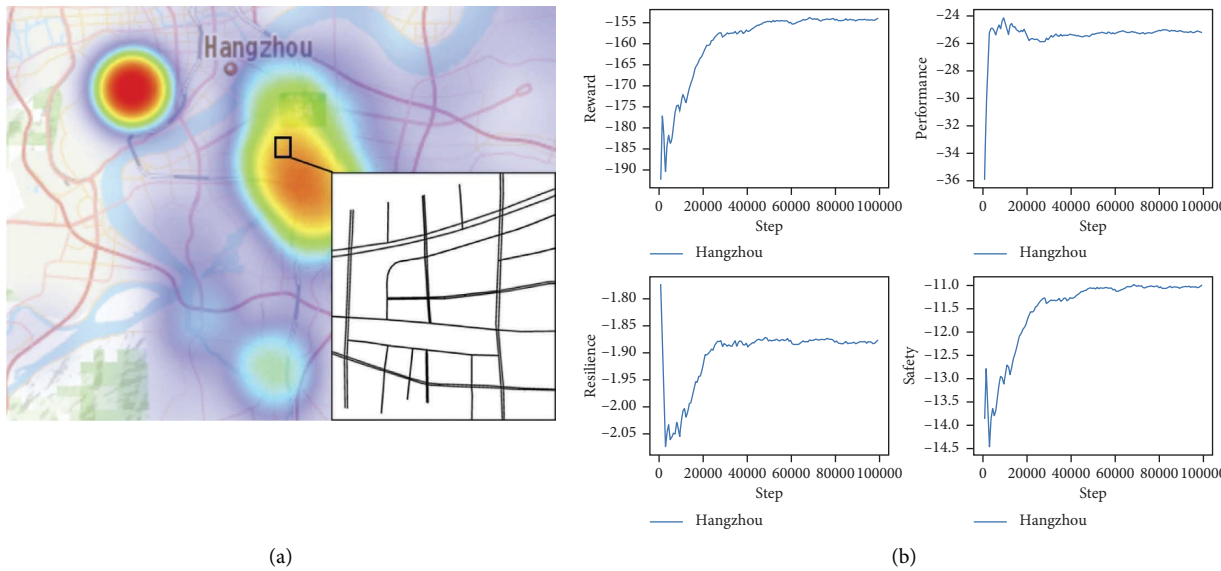


Figure 8: Real net scenario setting. (a) Traffic accident heatmap: road network extracted from areas with high frequency of accidents. (b) Evaluation metric results in the real net simulation.

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

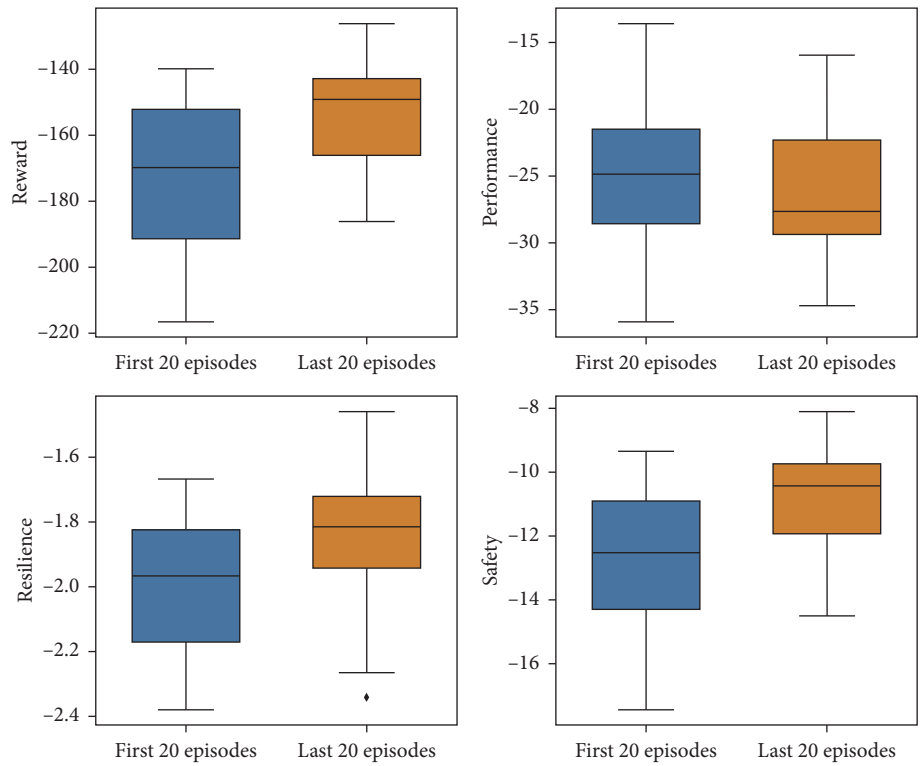


Figure 9: Improvement in the Hangzhou real net experiment.

corresponding episode steps and marked as accident steps (e.g., an accident that occurs at the 23rd minute will mark 276 episode steps as an accident step).

Same as the setting in the grid network, the duration of accidents in each episode will be randomly selected from 900, 950, 1000, 1050, 1100, 1150 and 1200 s. The neighbouring junctions that share information are set as junctions that are directly connected by a single edge.

5.2.2. Results. A total of 100,000 training steps are conducted in the real net experiment and a convergence has been successfully reached with 20.08% improvement (Figure 8).

The training results are presented in detail in Figure 9, with the first 20 episodes compared with the last 20 episodes. It can be observed that the standard deviation has been improved for overall reward and three metrics. Although the median shows a decrease, the overall performance has actually increased by 34.31%. The contradiction is attributed to the rapid improvement, as demonstrated in Figure 8. As most settings in this experiment are retrieved from real data, it is reasonable to conclude that the proposed method is feasible for real-world application.

Compared with the grid simulation, shorter training time (100,000 vs. 1,000,000) as well as better improvement is observed (20.08% vs. 5.8%). This is assumed to originate from the lower flow rate (325 vs. 540) and smaller action space (maximum 4 vs. 6). However, the efficiency of implementing the proposed method in real-world should be further assessed, especially considering more unstable traffic flow and different network structure. These factors could substantially impact the performance of the system and thus pose threat to resilience.

6. Conclusion and Future Work

In this research, a decentralised MARL-based method has been proposed to mitigate random disruptions within urban network systems. The control policy training is based on a proposed reward function including a resilience metric deduced by the authors. Two experiments have then been conducted in SUMO based on the proposed method. The first experiment uses a virtual grid network and the second relies on a real-world urban network. In both experiments, the random effects caused by vehicle accidents are simulated, and an assumption that ATSC is capable of observing the real-time traffic flow and communicating with their neighbouring junctions is adopted.

The results from grid simulation experiments have shown an improvement of 5.8%. Comparative experiments have also been run to validate the adoption of the proposed reward function, which demonstrate that the proposed method outperformed benchmark control policies and present the robustness against different levels of disruptions. Meanwhile, the result from the real net experiment has illustrated the feasibility of utilising the proposed method in real world with an improvement of 20.08%.

While our experiment results in SUMO are promising, we acknowledge the need for further validation in diverse urban environments to fully establish the broad applicability of our method. There are potential challenges such as the communication infrastructure required for our

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

decentralized MARL approach. While modern traffic systems increasingly incorporate smart technologies, ensuring reliable and low-latency communication between intersections may present practical hurdles in some urban environments. Future work should focus on testing the proposed approach in multiple cities with varying traffic patterns, road network structures and other types of disruptions, such as pedestrian crossing, inaccurate traffic flow calculation and noise received from the environment.

In addition, it can be observed that the effectiveness of the proposed method is reliant on characteristics of the urban road network, particularly the accident frequency and traffic flow rate in this research. As a result, comprehending an indicator that reflects correlation between characteristics and effectiveness should the focus of future study, as well as a comprehensive comparison with benchmark methods.

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This work was supported by the National Key R&D Program of China no. 2022ZD0160104. Open Access funding was enabled and organized by JISC GOLD.

References

[1] I. Cardinale, "Vulnerability, Resilience and "Systemic Interest": a Connectivity Approach," *Networks and Spatial Economics* 22, no. 3 (04 2019): 691–707, https://doi.org/10.1007/s11067-019-09462-9.

[2] S. Carpenter, B. Walker, J. M. Anderies, and N. Abel, "From Metaphor to Measurement: Resilience of What to What?" *Ecosystems* 4, no. 8 (2001): 765–781, https://doi.org/10.1007/s10021-001-0045-9.

[3] W.-L. Shang, Y. Chen, X. Li, and W. Y. Ochieng, "Resilience Analysis of Urban Road Networks Based on Adaptive Signal Controls: Day-To-Day Traffic Dynamics with Deep Reinforcement Learning," *Complexity* 2020 (2020): 1–19, https://doi.org/10.1155/2020/8841317.

[4] H.Y. Chen, A. Das, and D. Ivanov, "Building Resilience and Managing Post-Disruption Supply Chain Recovery: Lessons From the Information and Communication Technology Industry," *International Journal of Information Management* 49 (2019): 330–342, https://doi.org/10.1016/j.ijinfomgt.2019.06.002.

[5] J. Fiksel, M. Polyviou, K. Croxton, and T. Pettit, "From Risk to Resilience: Learning to Deal with Disruption," *MIT Sloan Management Review* 56 (2015): 12.

[6] A. Bahadur and D. Dodman, "Disruptive Resilience: An Agenda for the New Normal in Cities of the Global South," (London: IIED), https://www.iied.org/17766iiedTechnical report.

[7] M. Swanson, P. Bowen, A. Phillips, D. Gallup, and D. Lynes, "Contingency Planning Guide for Federal Information Systems," (2010), https://csrc.nist.gov/publications/detail/sp/800-34/rev-1/final.

[8] X. He and H. X. Liu, "Modeling the Day-To-Day Traffic Evolution Process After an Unexpected Network Disruption," *Transportation Research Part B: Methodological* 46, no. 1 (2012): 50–71, https://doi.org/10.1016/j.trb.2011.07.012.

[9] C. Wan, Z. Yang, Di Zhang, X. Yan, and S. Fan, "Resilience in Transportation Systems: A Systematic Review and Future Directions," *Transport Reviews* 38, no. 4 (2018): 479–498, https://doi.org/10.1080/01441647.2017.1383532.

[10] M. Aslani, M. S. Mesgari, and M. Wiering, "Adaptive Traffic Signal Control With Actor-Critic Methods in a Real-World Traffic Network with Different Traffic Disruption Events," *Transportation Research Part C: Emerging Technologies* 85 (2017): 732–752, https://doi.org/10.1016/j.trc.2017.09.020.

[11] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent Development and Applications of Sumo-Simulation of Urban Mobility," *International Journal on Advances in Systems and Measurements* 5, no. 3 & 4 (2012).

[12] J. Choi, K. Park, M. Kim, and S. Seok, "Deep Reinforcement Learning of Navigation in a Complex and Crowded Environment With a Limited Field of View," in *2019 International Conference on Robotics and Automation (ICRA)* (Montreal, Canada, May 2019), 5993–6000.

[13] P. B. Hunt, D. I. Robertson, R. D. Bretherton, and M. C. Royle, "The Scoot On-Line Traffic Signal Optimisation Technique," *Traffic Engineering and Control* 23, no. 4 (1982).

[14] J. Y. K. Luk, "Two Traffic-Responsive Area Traffic Control Methods: Scat and Scoot," *Traffic Engineering and Control* 25, no. 1 (1984).

[15] S. Chiu and S. Chand, "Adaptive Traffic Signal Control Using Fuzzy Logic," in *Proceedings of the First IEEE Regional Conference on Aerospace Control Systems* (1993), 122–126, https://doi.org/10.1109/AEROCS.1993.720907.

[16] F. V. Webster, *Traffic Signal Settings* (H.M. Stationery Office, 1958).

[17] M. J. Smith, "The Existence, Uniqueness and Stability of Traffic Equilibria," *Transportation Research Part B: Methodological* 13, no. 4 (1979): 295–304, https://doi.org/10.1016/0191-2615(79)90022-5.

[18] J. C. Spall and D. C. Chin, "Traffic-Responsive Signal Timing for System-Wide Traffic Control," *Transportation Research Part C: Emerging Technologies* 5, no. 3-4 (1997): 153–163, https://doi.org/10.1016/S0968-090X(97)00012-0.

[19] N. H. Gartner and M. Al-Malik, "Combined Model for Signal Control and Route Choice in Urban Traffic Networks," *Transportation Research Record: Journal of the Transportation Research Board* 1554, no. 1 (1996): 27–35, https://doi.org/10.1177/0361981961554001O4.

[20] S. Yagar and F. Dion, "Distributed Approach to Real-Time Control of Complex Signalized Networks," *Transportation Research Record: Journal of the Transportation Research Board* 1554, no. 1 (1996): 1–8, https://doi.org/10.1177/0361981961554001O1.

[21] M. Smith, "Dynamics of Route Choice and Signal Control in Capacitated Networks," *Journal of Choice Modelling* 4, no. 3 (2011): 30–51, https://doi.org/10.1016/S1755-5345(13)70041-1.

[22] M. Smith and R. Mounce, "A Splitting Rate Model of Traffic Re-Routeing and Traffic Control," *Procedia-Social and Behavioral Sciences* 17 (2011): 316–340, https://doi.org/10.1016/j.sbspro.2011.04.520.

[23] J. L. Adler, G. Satapathy, V. Manikonda, B. Bowles, and V. J. Blue, "A Multi-Agent Approach to Cooperative Traffic Management and Route Guidance," *Transportation Research*

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Part B: Methodological 39, no. 4 (2005): 297–318, https://doi.org/10.1016/j.trb.2004.03.005.

[24] B. Abdulhai, R. Pringle, and G. J. Karakoulas, "Reinforcement Learning for True Adaptive Traffic Signal Control," *Journal of Transportation Engineering* 129, no. 3 (2003): 278–285, https://doi.org/10.1061/(asce)0733-947x(2003)129:3(278).

[25] S. Touhbi, M. A. Babram, T. Nguyen-Huu, et al., "Adaptive Traffic Signal Control: Exploring Reward Definition for Reinforcement Learning," *Procedia Computer Science* 109 (2017): 513–520, https://doi.org/10.1016/j.procs.2017.05.327.

[26] M. Abdoos, N. Mozayani, and A. L. C. Bazzan, "Holonic Multi-Agent System for Traffic Signals Control," *Engineering Applications of Artificial Intelligence* 26, no. 5-6 (2013): 1575–1587, https://doi.org/10.1016/j.engappai.2013.01.007.

[27] S. Darmoul, S. Elkosantini, A. Louati, and L. Ben Said, "Multi-Agent Immune Networks to Control Interrupted Flow at Signalized Intersections," *Transportation Research Part C: Emerging Technologies* 82 (2017): 290–313, https://doi.org/10.1016/j.trc.2017.07.003.

[28] V. Mnih, A. Puigdomènech Badia, M. Mirza, et al., "Asynchronous Methods for Deep Reinforcement Learning," *CoRR* (2016): 01783, https://arxiv.org/abs/1602.01783.

[29] T. Chu, S. Chinchali, and S. Katti, "Multi-agent Reinforcement Learning for Networked System Control," (2020), https://arxiv.org/abs/2004.01339.

[30] D. Garg, M. Chli, and G. Vogiatzis, "Fully-autonomous, Vision-Based Traffic Signal Control: from Simulation to Reality," in *Proceedings of the 21st International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2022)* (USA: ACM, May 2022), 454–462.

[31] E. Paul Ijjina, D. Chand, S. Gupta, and K. Goutham, "Computer Vision-Based Accident Detection in Traffic Surveillance," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (IEEE, 2019), 1–6, https://doi.org/10.1109/ICCCNT45670.2019.8944469.

[32] G. Zhang, F. Chang, J. Jin, F. Yang, and H. Huang, "Multi-objective Deep Reinforcement Learning Approach for Adaptive Traffic Signal Control System with Concurrent Optimization of Safety, Efficiency, and Decarbonization at Intersections," *Accident Analysis & Prevention* 199 (2024): 107451, https://doi.org/10.1016/j.aap.2023.107451.

[33] W. Du, J. Ye, J. Gu, J. Li, H. Wei, and G. Wang, "Safelight: A Reinforcement Learning Method toward Collision-Free Traffic Signal Control," *Proceedings of the AAAI Conference on Artificial Intelligence* 37, no. 12 (2023): 14801–14810, https://doi.org/10.1609/aaai.v37i12.26729.

[34] M. A. A. Ahmed, H. L. Khoo, and O.-Ee Ng, "Discharge Control Policy Based on Density and Speed for Deep Q-Learning Adaptive Traffic Signal," *Transportmetrica B: Transport Dynamics* 11, no. 1 (2023): 1707–1726, https://doi.org/10.1080/21680566.2023.2243388.

[35] C. Tan and K. Yang, "Privacy-preserving Adaptive Traffic Signal Control in a Connected Vehicle Environment," *Transportation Research Part C: Emerging Technologies* 158 (2024): 104453, https://doi.org/10.1016/j.trc.2023.104453.

[36] S. Mishra, V. Singh, A. Gupta, D. Bhattacharya, and A. Mudgal, "Adaptive Traffic Signal Control for Developing Countries Using Fused Parameters Derived from Crowd-Source Data," *Transportation Letters* 15, no. 4 (2023): 296–307, https://doi.org/10.1080/19427867.2022.2050493.

[37] S. A. Argyroudis, "Resilience Metrics for Transport Networks: a Review and Practical Examples for Bridges," *Proceedings of the Institution of Civil Engineers - Bridge Engineering* 175, no. 3 (2022): 179–192, https://doi.org/10.1680/jbren.21.00075.

[38] G. P. Cimellaro, A. M. Reinhorn, and M. Bruneau, "Framework for Analytical Quantification of Disaster Resilience," *Engineering Structures* 32, no. 11 (2010): 3639–3649, https://doi.org/10.1016/j.engstruct.2010.08.008.

[39] G. Cimellaro, C. Renschler, A. M. Reinhorn, and L. Arendt, "Peoples: A Framework for Evaluating Resilience," *Journal of Structural Engineering* 142, no. 10 (2016): 04016063, https://doi.org/10.1061/(asce)st.1943-541x.0001514.

[40] E. Minaie and F. Moon, "Practical and Simplified Approach for Quantifying Bridge Resilience," *Journal of Infrastructure Systems* 23, no. 4 (2017): 04017016, https://doi.org/10.1061/(asce)is.1943-555x.0000374.

[41] A. Arun, L. Vanajakshi, and A. Sharma, "A Simple Method for Estimation of Queue Length," *International Journal of Innovation Research in Science, Engineering and Technology* 2, no. 1–5 (2013).

[42] E. Balal, G. Valdez, J. Miramontes, and R. L. Cheu, "Comparative Evaluation of Measures for Urban Highway Network Resilience Due to Traffic Incidents," *International Journal of Transportation Science and Technology* 8, no. 3 (05 2019): 304–317, https://doi.org/10.1016/j.ijtst.2019.05.001.

[43] M. Saffarzadeh, N. Nadimi, S. Naseralavi, and A. R. Mamdoohi, "A General Formulation for Time-To-Collision Safety Indicator," *Proceedings of the Institution of Civil Engineers - Transport* 166, no. 5 (10 2013): 294–304, https://doi.org/10.1680/tran.11.00031.

[44] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep Reinforcement Learning for Multi-Agent Systems: A Review of Challenges, Solutions and Applications," *CoRR* (2018): 11794, https://arxiv.org/abs/1812.11794.

[45] K. Li and Q.-S. Jia, "Decentralized Multi-Agent Reinforcement Learning: An Off-Policy Method," *CoRR* (2021): 00438, https://arxiv.org/abs/2111.00438.

[46] X. Wang, L. Ke, Z. Qiao, and X. Chai, "Large-scale Traffic Signal Control Using a Novel Multiagent Reinforcement Learning," *IEEE Transactions on Cybernetics* 51, no. 1 (Jan 2021): 174–187, https://doi.org/10.1109/tcyb.2020.3015811.

[47] J. I. Ge and G. . Orosz, "Dynamics of Connected Vehicle Systems with Delayed Acceleration Feedback," *Transportation Research Part C: Emerging Technologies* 46 (2014): 46–64, https://doi.org/10.1016/j.trc.2014.04.014.

[48] K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Başar, "Fully Decentralized Multi-Agent Reinforcement Learning with Networked Agents," (2018), https://arxiv.org/abs/1802.08757.

[49] K. Jin, W. Wang, X. Li, X. Hua, S. Chen, and S. Qin, "Identifying the Critical Road Combination in Urban Roads Network under Multiple Disruption Scenarios," *Physica A: Statistical Mechanics and Its Applications* 607, no. C (2022): 128192, https://doi.org/10.1016/j.physa.2022.128192.

[50] S. Amini, G. Tilg, and F. Busch, "Evaluating the Impact of Real-Time Traffic Control Measures on the Resilience of Urban Road Networks," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (2018), 519–524.

[51] J. Zeng, Y. Qian, B. Wang, T. Wang, and X. Wei, "The Impact of Traffic Crashes on Urban Network Traffic Flow," *Sustainability* 11, no. 14 (07 2019): 3956, https://doi.org/10.3390/su11143956.

[52] A. Nicholson and Y. D. Wong, "Are Accidents Poisson Distributed? a Statistical Test," *Accident Analysis & Prevention* 25, no. 1 (03 1993): 91–97, https://doi.org/10.1016/0001-4575(93)90100-b.

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[53] German Aerospace Center, "Traffic Lights-SUMO Documentation," (2024), https://sumo.dlr.de/docs/Simulation/Traffic_Lights.html.

[54] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O.A. I. P. Abbeel, and I. Mordatch, "Multi-agent Actor-Critic for Mixed Cooperative-Competitive Environments," in *Advances in Neural Information Processing Systems*, eds. I. Guyon, U. Von Luxburg, S. Bengio, et al., 30 (Curran Associates, Inc, 2017).

[55] J. Foerster, I. Alexandros Assael, N. de Freitas, and S. Whiteson, "Learning to Communicate with Deep Multi-Agent Reinforcement Learning," in *Advances in Neural Information Processing Systems*, eds. D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, 29 (Curran Associates, Inc, 2016).

[56] A. Muralidharan, R. Pedarsani, and P. Varaiya, "Analysis of Fixed-Time Control," *Transportation Research Part B: Methodological* 73 (2015): 81–90, https://doi.org/10.1016/j.trb.2014.12.002.

[57] Z. Li, C. Xu, and G. Zhang, "A Deep Reinforcement Learning Approach for Traffic Signal Control Optimization," (2021), https://arxiv.org/abs/2107.06115.

1409, 2024, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1155/2024/3035753 by Imperial College London, Wiley Online Library on [11/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License