A Gain With No Pain: Exploring Intelligent Traffic Signal Control for Emergency Vehicles

Miaomiao Cao, Victor O. K. Li, *Life Fellow, IEEE*, and Qiqi Shuai

***Abstract*—For the emergency response, every second counts. Intersections are prone to congestion, which greatly hinders the fast response of emergency vehicles. Although emergency vehicles possess the privilege to run a red light, it can be unsafe, and a congested intersection will prevent the exercise of this privilege. When an emergency vehicle arrives, the greedy preemption scheme offers a green signal promptly until it leaves the intersection. This guarantees a fast emergency response in most cases. However, this scheme will lead to an adverse impact on vehicles of conflicting directions and may not work when there are other emergency vehicles traveling from conflicting directions simultaneously. Employing deep reinforcement learning techniques, recent studies have shown promising results for traffic signal control. In this work, we deliver an early attempt to control the traffic signal for emergency vehicles through deep reinforcement learning, which ensures an expeditious emergency response in various scenarios and alleviates the negative influence on the traffic efficiency of conflicting directions. We conduct realistic simulations using traffic data in a real-world network with multiple intersections on different testing parameters. The results verify the feasibility and effectiveness of our model and indicate that our method notably outperforms the other five baseline methods in terms of various performance metrics.**

***Index Terms*—Traffic signal control, emergency vehicles, deep reinforcement learning, intelligent transportation system.**

I. Introduction

PROMPT response of EMergency Vehicles (EMVs), including ambulances, fire engines, police cars, is crucial to saving lives and properties. For example, an average of 700 fatalities is recorded in Ireland due to a delayed ambulance response [1]. A quick medical response can lower the probability of death related to vehicle accidents by one third [2]. In fire accidents, the extent and heat can double every 17 seconds and may go out of control after flashover in seven minutes [3]. Since the beginning of the COVID-19 public health emergency, the number of ambulances required has greatly increased [4]. However, with the continuous expansion of traffic volume, congestion and transportation delay on urban roads become increasingly severe, greatly hindering the rapid response of EMVs. In particular, intersections, as one of the most congestion-prone zones on road networks [5], remain a major challenge to EMVs' fast response.

To save travel time, EMVs (with sirens and flashing lights on) are given the privilege to run a red light at intersections. However, the blocked line of sight [6] may lead to fatal accidents of EMVs. According to a US report [7], roughly 70% of all firetruck accidents occur during emergency response, which is the second leading cause for on-the-job deaths of firefighters. In addition, an EMV cannot reach the stop line at a congested intersection to exercise its privilege.

For the safe and fast passage of EMVs, signal preemption becomes imperative. Existing research indicates that the signal preemption can significantly reduce the travel time for EMVs. Authors in [6] reported that different signal preemption systems can save the travel time of EMVs from 14% to 35%. Greedy signal preemption, a common practice which works on the principle that when an EMV is detected at an intersection, the green signal will be held for its direction until it exits [8]. However, the greedy signal preemption will not work in some scenarios. For example, when two EMVs are passing through an intersection from conflicting directions, the greedy signal preemption will fail to offer the green phase for both of them simultaneously. This is a compelling case when a major emergency, such as a terrorist attack, an accidental explosion, etc., occurs, fleets of EMVs will come from different directions to rescue. Moreover, another drawback is the high negative impact on the delay of vehicles from conflicting directions, thus greatly degrading the overall traffic efficiency. Authors in [9] observed an increase of non-EMV traffic delays from 4% to 58% due to the EMV preemption. In most cases, an EMV will pass through several intersections between its origin and destination, and the greedy signal preemption can lead to accumulative negative impacts on the overall traffic conditions along its path.

Therefore, an intelligent traffic signal control (TSC) to ensure the quick pass of EMVs in various cases is indispensable while alleviating the burden on the traffic of non-EMV directions is also significant. The contributions of this paper are summarized as follows:

1) This work is the first learning-based TSC method specially designed for EMVs, which can not only ensure the quick pass of EMVs in various cases, but also alleviate the congestion of conflicting directions.
2) Our proposed RECAL algorithm guarantees an expeditious emergency response in different scenarios,

Manuscript received 20 February 2021; revised 14 August 2021 and 31 December 2021; accepted 3 March 2022. Date of publication 23 March 2022; date of current version 11 October 2022. This work was supported in part by the Innovation and Technology Fund of Hong Kong under Grant ITS/447/16. The Associate Editor for this article was H. Jula. *(Corresponding author: Miaomiao Cao.)*

Miaomiao Cao and Victor O. K. Li are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: mmcao@eee.hku.hk; vli@eee.hku.hk).

Qiqi Shuai was with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. He is now with Lighthorse Asset Management, Hong Kong (e-mail: qqshuai@eee.hku.hk).

Digital Object Identifier 10.1109/TITS.2022.3159714

especially when more than one EMV is passing through an intersection from conflicting directions simultaneously.

3) We design effective terms and coefficients in the reward and achieve balanced sampling with our proposed concept of current pair. The ablation study verifies that they are all critical components to achieve superior performance in our intelligent scheme.
4) The proposed intelligent scheme follows a novel multi-agent deep reinforcement learning framework with shared parameters so that we can train only one agent for different intersections with all the samples in the traffic network, which enhances the scalability of implementations for our method in practice.
5) We conduct extensive simulations based on the real-world traffic network of a challenging 16 intersections at Gudang Residential District in Hangzhou, China, which demonstrate the outstanding performance of our intelligent scheme compared with five popular and state-of-the-art baseline methods over a multitude of performance metrics, and validate its feasibility and effectiveness in practice.

The remainder of this paper is organized as follows. Section II reviews related literature. In Section III, we present our system model. In Section IV, we propose our method. The experimental results are shown and analyzed in Section V. Finally, we conclude this paper in Section VI.

II. Literature Review

In this section, we introduce the conventional methods and reinforcement learning approaches for traffic signal control and the signal control methods for EMV preemption.

A. Conventional Methods and Reinforcement Learning Approaches for Traffic Signal Control

Fixed time control is a common strategy that can be easily implemented at a low cost [10]. Recently, adaptive signal controls have been proved to effectively reduce traffic congestion. For example, authors in [11] propose algorithms to adjust the sequence of green phases and cycle durations using real-time information. By directly measuring the vehicle speed instead of the queue length, Ren *et al*. [12] propose an adaptive signal control scheme to deal with intersection congestion.

Based on large-scale real-time traffic data, existing research has shown the effectiveness of reinforcement learning algorithms on solving traffic signal control problems [13], [14]. Authors in [15], [16] present comprehensive reviews of reinforcement learning for traffic signal control. When the state space is relatively small, a linear function can be used to estimate the Q value [17]. Recent studies [18] propose to learn a Q-function (Deep Q-networks) through a deep neural network, which extends the application of reinforcement learning to huge state space for traffic signal control. In this way, the state representation contains a multitude of features [19], [20] and to improve the overall performance, authors in [20], [21] consider a combination of measures for the reward design.

B. Traffic Signal Control for Emergency Vehicles

Authors in [8], [22] offer an EMV a green phase when it enters the intersection. This is a typical greedy preemption scheme to ensure the fast pass of EMVs. Based on fuzzy logic-based methods, much research has been done on decreasing the impact on non-EMV traffic, in addition to ensuring the quick pass of EMVs. The key idea is to divide the real-time traffic conditions (i.e., traffic speed, queue length, waiting time, congestion level, etc.) into several cases based on human knowledge and set rules to control the signal accordingly. Miletić *et al*. [23] compare the performance of fuzzy logic-based control and vehicle-tracking-and-queue-length-based control [24] and conclude that the former performs better. In our previous work [25], an emergency vehicle-centered (EMV-centered) scheme, which focuses on EMV performance, is proposed to achieve the fast pass of EMVs while alleviating the negative impact on non-EMV traffic.

With the normal first-come first-served (FCFS) control policy, the studies above can hardly handle EMVs from conflicting directions. Authors in [14], [26], [27] proposed to assign different priorities to vehicles from conflicting directions. For example, authors in [26] formulated a dynamic programming model to optimize the serving sequence for multiple bus priority requests, and they assigned different priorities to buses based on the number of passengers and so on. This is good in that case while it is arduous to decide the priorities of different EMVs in practice. Therefore, in this work, we do not assign different priorities to EMVs from conflicting directions and rely on intelligent TSC to achieve the overall performance for all of them.

In addition, as pointed out by authors in [28], optimization-based methods usually suffer high burden for online computation of the optimal policy. For example, the recent work [29] proposes elastic signal preemption (ESP) to find non-intrusive signal schedules for the fast pass of EMVs through intersections by solving a set of quadratic programming problems. With the method in [29], every few seconds the EMV reports its location to the control center which will trigger the signal preemption optimization and in some cases the model in [29] is not even solvable. Learning-based strategy like the one in this work can deal with the computation issue. This is because, for a well-trained model, it can work like a function so that in each time step, with the states needed as the input, the model can generate the action of TSC for EMVs directly and there is no need to solve optimization problems every few seconds, thus greatly reducing the computational complexity.

III. System Model

We focused on a typical traffic network with many intersections, where congestion is most likely to occur and multiple road side units (RSUs) are located in the traffic network to communicate with EMVs.

Fig. 1 is an illustration of the system model of TSC for EMVs passing through the target traffic network. When an EMV enters the target traffic network or changes its route in this traffic network, it can send its route information to a nearby RSU by vehicle-to-infrastructure (V2I) communications, such as Vehicular ad hoc networks

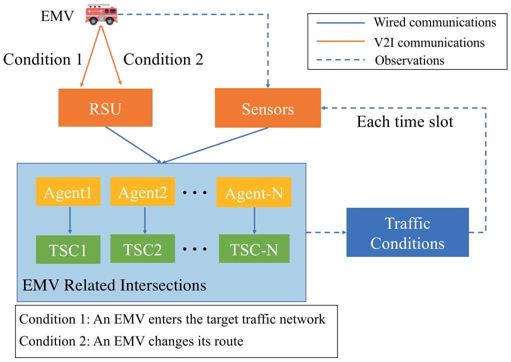

Fig. 1. System model of TSC for EMVs passing through the target traffic network.

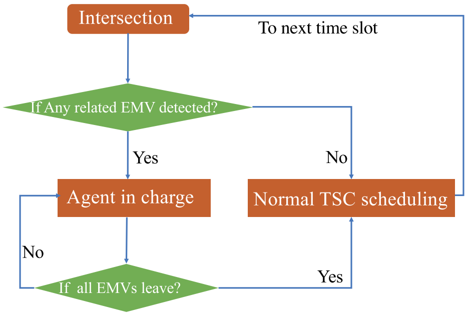

Fig. 2. Workflow of TSC at an intersection in the target traffic network.

(VANET) [30] and Dedicated Short Range Communication (DSRC) [31]. Then the RSU will broadcast the arrival of this EMV to the intersections that the EMV will pass through. When at least one EMV is close to these intersections, they will be controlled by our intelligent scheme and generate TSC actions by the corresponding agents. A time slot is a minimum time step for state updates in our method. In each time slot, sensors will detect the traffic conditions of each intersection and the state of each EMV, and send them to the controller module with wired communications.

Fig. 2 illustrates the workflow of TSC at an intersection in the traffic network. In each time slot, if an intersection receives the information that some EMVs that will pass through it are close to it, the agent will take over the TSC. When all the related EMVs leave this intersection, it will continue to operate on the normal TSC scheduling. Similar to the previous work [29], if any agent fails to give appropriate actions for TSC, the intersection can execute the normal TSC scheduling without any additional control.

For each intersection, we define two green phases for the intersection as shown in Fig. 3. In Phase WEG, the $W-E$ direction can go straight or turn left or right and in Phase SNG, the $N-S$ direction can go straight or turn left or right. When a direction shares the same green phase with an EMV, we call it EMV direction, otherwise it is non-EMV or EMV conflicting directions.

IV. Methodology

A. Deep Reinforcement Learning Framework

Suppose $\mathcal{S}$ is the state space and $\mathcal{A}$ is the action space. As illustrated in Fig. 4, in a typical deep reinforcement learning framework, the environment represents all the traffic conditions at the intersection, which receives an action $a \in \mathcal{A}$ for TSC, and generates state $s \in \mathcal{S}$ and reward r (corresponding to s and a); the agent is our intelligent model, which obtains state s and reward r, and generates action a for TSC.

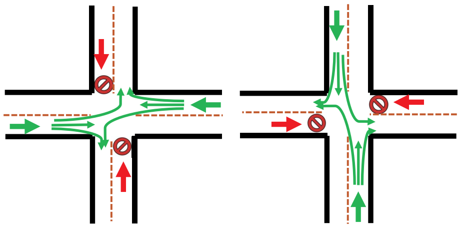

Fig. 3. Two green phases of an intersection.

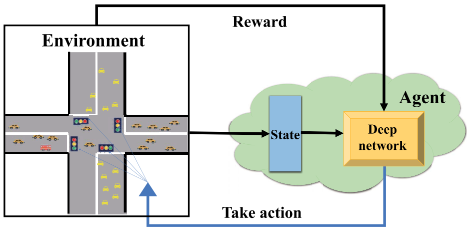

Fig. 4. Deep reinforcement learning framework.

The target of the agent is to learn a policy π^* that can maximize the cumulative expected rewards. When the agent receives state s_t at the start of time step t and generates an action a_t, and receives a series of rewards after time step t as r_m, where $m = t, t+1, t+2\ldots$, then we can obtain the cumulative expected rewards as a Q value, i.e.,

$$\begin{aligned} Q(s_t, a_t) &= \mathbb{E}[\sum_{m=t}^{\infty} \gamma^{m-t} r_m] \\ &= \mathbb{E}[r_t + \gamma\, Q(s_{t+1}, a_{t+1})], \end{aligned} \tag{1}$$

where $\gamma \in [0, 1)$ is the discount factor that calculates the present values of future rewards.

We can calculate the maximum $Q(s_t, a_t)$, which corresponds to the optimal action policy at the time step t, i.e., $\pi_t^* = \arg\max_{\pi_t} Q^{\pi_t}(s_t, a_t)$, recursively, and it can be expressed by the Bellman optimality equation as

$$Q^{\pi_t^*}(s_t, a_t) = \mathbb{E}[r_t(s_t, a_t) + \gamma \max_{a_{t+1} \in \mathcal{A}} Q(s_{t+1}, a_{t+1})]. \tag{2}$$

If the state space is finite and relatively small, we can solve Eq. (2) through dynamic programming with a reasonable computation cost. Otherwise, we can only obtain an approximate value of the maximum Q with a θ function, i.e., $Q(s, a; \theta)$.

B. Model Design

To design our model through deep reinforcement learning, we first define the reward, action and state in our method.

1) Reward: As shown in Eq. (4), the reward is defined as a weighted sum of the following penalty factors:

(1) Constraint 1 (C_1): $C_1 = 1$ when an EMV has stayed at the intersection over a certain time period T_{max}; otherwise, $C_1 = 0$. This is a key factor to ensure that the EMV can leave the intersection within a reasonable time period.

(2) Constraint 2 (C_2): $C_2 = 1$ when the distance between an EMV and the stop line is shorter than a given value D_{min} and the green signal for the EMV has not been provided yet; otherwise, $C_2 = 0$. This is a crucial factor to clear the potential waiting vehicles in front of the EMV and allow the EMV to pass the stop line without stops.

(3) Stay time of EMV (T_{EMV}): the time that an EMV has stayed since it entered the intersection.

(4) Number of stops of EMV (SN_{EMV}): the number of stops of an EMV since it has entered an intersection.

(5) Delay of EMV (D_{EMV}): as in [25], we define the delay of the EMV as follows,

$$D_{EMV} = 1 - \frac{EMV\ Speed}{Speed\ Limit}. \tag{3}$$

(6) Queue length of the EMV direction (Q_{EMV}): the total number of waiting vehicles from the EMV direction.

(7) Queue length of non-EMV directions ($Q_{non-EMV}$): the total number of waiting vehicles over all the non-EMV directions.

(8) Waiting time of the EMV direction (W_{EMV}): the total waiting time of vehicles from the EMV direction.

(9) Waiting time of non-EMV directions ($W_{non-EMV}$): the total waiting time of vehicles over all the non-EMV directions.

All the penalty factors above are defined for an EMV, and if there is only one EMV at an intersection, we can obtain the reward with Eq. (4), and all these weights in Eq. (4) are negative to maximize the rewards for optimality.

$$\begin{aligned} Reward = {} & w_1C_1 + w_2C_2 + w_3T_{EMV} + w_4SN_{EMV} \\ & + w_5D_{EMV} + w_6Q_{EMV} + w_7Q_{non-EMV} \\ & + w_8W_{EMV} + w_9W_{non-EMV}. \end{aligned} \tag{4}$$

In practice, more than one EMV may pass through an intersection from the same or conflicting directions and in this general case, we propose Algorithm 1 to calculate the reward of each intersection. This case happens when a major disaster occurs, such as a fire or a terrorist attack, and numerous EMVs are required to travel to the rescue from different directions at the same time. Intuitively, as shown in Algorithm 1, for EMVs that will pass a certain intersection through the same green phase, we only consider the EMV nearest to the centre of the intersection since this EMV stays the longest time and the other EMVs that will pass the intersection through the same green phase can benefit from the TSC for this EMV. Then, for EMVs from conflicting directions, we use the maximum value (i.e., worst performance) of each penalty factor to calculate the reward at an intersection. In this way, the agent can learn to balance the performances of EMVs from conflicting directions and avoid the most unfavourable performance of any EMV.

2) Action: Action space $\mathcal{A}$ consists of all the possible green phases at the intersection, i.e., $\mathcal{A} = \{WEG, SNG\}$. For example, when $a = WEG$, if the current phase is also WEG, we keep the current phase within a minimal signal switch time interval; if the current phase is SNG, we will change the phase to WEG according to the action a.

3) State: State contains all the information that the agent requires to control the traffic signals. It consists of EMV and system states and the factors of them in our model are described in Fig. 5.

4) Deep Q-Network (DQN): Apparently the state space is huge, and it is arduous to calculate the maximum Q value in Eq. (2). With the definitions of reward, action and state, we design a deep Q-network following the popular DQN approach in [18], to approximate the maximum Q value and generate discrete actions for intelligent TSC. As illustrated in Fig. 5, all the factors of EMV and system states are

Algorithm 1 Rewards Calculation (RECAL)

```
In a traffic network with $N$ intersections, the number of green phases at each intersection is at most $M$, $S_{EMV_{mn}}$ is the set of EMVs that will pass Phase $m$ of Intersection $n$, $d_{EMV_{mn}}(T)$ is the distance of $EMV_{mn}$ to the centre of Intersection $n$ at time $T$.
$R_{EMV_n}$ is the set of EMVs used to calculate the reward at Intersection $n$, $PFactors$ is the set of penalty factors in the reward design, $w_{pf}$ is the weight of a certain penalty factor i.e., $pf$, and $pf_{R_{EMV_n}}$ is the set of penalty factors of EMVs in $R_{EMV_n}$, $R_n$ is the reward of Intersection $n$, and $t$ is the length of time slot.
At time $T = t, 2t, 3t, \ldots,$ run RECAL()
function RECAL()
    for $n = 1$ to $N$ do
        $R_{EMV_n} = \emptyset$
        for $m = 1$ to $M$ do
            if $|S_{EMV_{mn}}| > 0$ then
                $R_{EMV_n} = R_{EMV_n} + \arg\min_{EMV_{mn}} d_{EMV_{mn}}(T)$
            end if
        end for
        if $|R_{EMV_n}| > 0$ then
            $R_n = 0$
            for $pf$ in $PFactors$ do
                $R_n = R_n + w_{pf} \max(pf_{R_{EMV_n}})$
            end for
        end if
    end for
end function
```

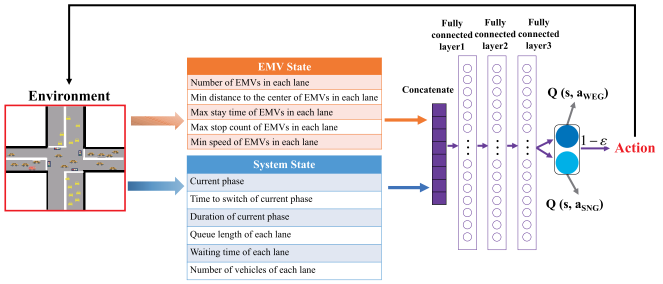
Environment
EMV State
Number of EMVs in each lane
Min distance to the center of EMVs in each lane
Max stay time of EMVs in each lane
Max stop count of EMVs in each lane
Min speed of EMVs in each lane
System State
Current phase
Time to switch of current phase
Duration of current phase
Queue length of each lane
Waiting time of each lane
Number of vehicles of each lane
Concatenate
Fully connected layer1
Fully connected layer2
Fully connected layer3
Q (s, a_WEG)
Q (s, a_SNG)
1 − ε
Action

Fig. 5. Deep Q-network.

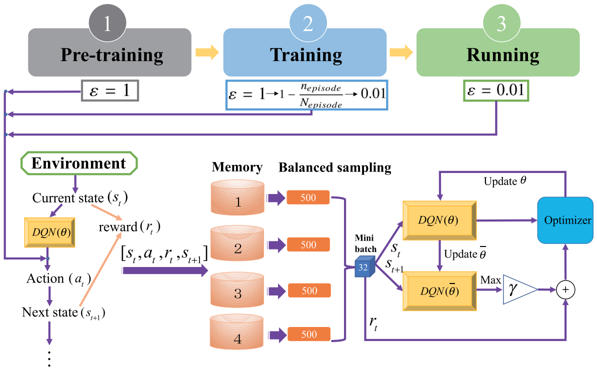

Fig. 6. Training process of our intelligent model.

concatenated and fed into two fully-connected layers of 64 and 16 units, respectively. The final output layer is a fully connected linear layer outputting a vector of maximum Q values, where each element in the vector corresponds to the estimated maximum Q value of each action $a \in \mathcal{A} = \{WEG, SNG\}$. Following the popular ε-greedy method, the agent will select the action with the highest maximum Q value with probability $1 - \varepsilon$ (exploitation) and select a random action with probability ε (exploration). A deep Q-network is designed for one intersection and this work is based on a multi-agent reinforcement learning model with shared parameters. That is, we can train only one agent for different intersections with all the samples in the traffic network, and this process will not be limited by the scale of the traffic network. Therefore, our method can be easily applied in large traffic networks with various intersections, which enhances the scalability of implementations for our method in practice.

C. Training Process

The overall workflow of our model can be divided into pre-training, training and running stages as shown in Fig. 6.

At the pre-training stage, the agent randomly selects actions and generates sufficient samples $[s_t, a_t, r_t, s_{t+1}]$. Then at the training stage, we will train our Q-network to be an estimator of the maximum Q value and take the decreasing ε-greedy scheme, in which, as our deep Q-network is trained to become increasingly accurate to estimate the maximum Q value, the value of ε gradually decreases as

$$\varepsilon_n = \max(1 - \frac{n_{episode}}{N_{episode}}, 0.01), \tag{5}$$

where $n_{episode}$ is the current number of episodes, $N_{episode} = 10000$ is the total number of episodes at the training stage. Initially, $\varepsilon = 1$, meaning that the agent exclusively explores; however, as the training progresses, ε gradually decreases and meanwhile the agent increasingly exploits what it has learned. At the end of the training stage, the agent finally learns to get high cumulative rewards by reacting on different traffic scenarios and then similar to previous work [21], we set the exploration probability $\varepsilon = 0.01$. Specifically, each time the model gives an action, we will draw a random number from a uniform distribution from 0 to 1. If the number is smaller than 0.01, we will explore the action; otherwise, we will follow the action generated by the model.

TABLE I
VALUE OF CURRENT PAIR

Current phase	Location of EMV	Current pair
WEG	$W - E$	1
WEG	$S - N$	2
SNG	$W - E$	3
SNG	$S - N$	4

Finally, our well-trained deep-Q networks can be applied in practice to control the traffic signal for EMV preemption in the running stage. Although our agent has experienced multifarious states during the pre-training and training stages, it cannot exhaust the whole state space. Hence, $Q(s_t, a_t; \theta)$ may still not estimate the maximum Q value adeptly for states not experienced. Therefore, in the running stage, the model insists on exploring with probability $\varepsilon = 0.01$, recording fresh samples $[s_t, a_t, r_t, s_{t+1}]$, storing them into the memory, and training regularly. That is, the control policy can be trained offline for an unchanged deployment in a long time, and at the same time, we can adapt the model online with more new samples stored in the memory similar to the previous work [21] that employs deep reinforcement learning for TSC.

As pointed out by authors in [21], traffic flow of different directions can be quite imbalanced in real traffic settings, and consequently the memory may be dominated by the phases and actions that appear most frequently in such imbalanced settings. In this work, we propose the concept of current pair to describe the relationship between the current phase and the location of EMV. The location of EMV is either in the $W - E$ direction or $S - N$ direction. For example, if the current phase is WEG and the location of EMV is in the $W - E$ direction, we define the current pair as 1. Similarly, we can get all the values of the current pair in Table I. Inspired by Memory Palace in [21], we propose to store the samples in different memories according to the current pair of a certain sample, i.e., each of memories 1, 2, 3 and 4 illustrated in Fig. 6 associated with a current pair in Table I. When training the model, we can achieve balanced sampling by taking the same number of samples from different memories, thus enhancing the estimation of the maximum Q value for both frequent and infrequent cases.

V. Performance Evaluation

A. Simulation Setup

1) Real-World Traffic Network: In this part, we will test the performance of different schemes based on a real-world traffic network of 16 intersections in Gudang Residential District, Hangzhou, as illustrated in Fig. 7. Red circle region is the traffic network in the simulations, and yellow dots are the traffic signals controlled by different schemes. This traffic dataset is based on camera data in the road networks, and for more information about this dataset, please refer to [32], [33]. We have conducted necessary augmentation operations to enrich this dataset and extend its time duration. Parameters of an intersection are shown in Table II and our

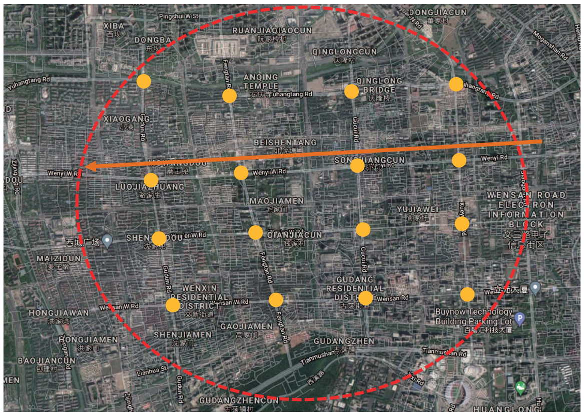

Fig. 7. Real-world traffic network of 16 intersections in Gudang Residential District, Hangzhou, China.

TABLE II

INTERSECTION PARAMETERS

Parameter	Value
No. of lanes on each road segment	3 (left, straight and right)
No. of green phases	2 (WEG and SNG)
Maximal speed limit	50 km/h
Signal between phase switch	3s yellow signal

simulations are conducted on a microscopic traffic simulation package SUMO [34], in which vehicles can be controlled by microscopic traffic models, such as car following models, allowing for vehicle traffic behaviours over road segments with multiple lanes [35] and we can flexibly control traffic signals and replay the trajectories of surrounding vehicles in the traffic network with various TSC schemes. In our simulations, vehicles are controlled by the popular car following model IDM [36], which can intelligently mimic human driving behaviors and is widely applied in previous studies [37], [38].

2) Benchmarks: To evaluate the effectiveness of our method, we compare it with the following five baseline methods:

No preemption scheme. EMVs pass the intersection in the same way as other vehicles. i.e., no priority is given to them. This simple scheme is an extreme case in which EMVs do not influence other vehicles but will lead to adverse EMV performance and is widely used as a benchmark in previous work [24], [27], [39].

Greedy preemption scheme. An EMV receives a green phase when it is detected to arrive at an intersection until it leaves. This is another extreme case that will achieve favorable EMV performance but fail to handle EMVs from conflicting directions and result in a tremendous impact on vehicles of conflicting directions [8].

Fuzzy logic-based scheme. The fuzzy logic-based approach is popular for TSC of EMV preemption [23]. Considering EMV distance, EMV queue length, occupancy level, and conflicting queue length, this scheme designs fuzzy rules to guarantee a fast emergency response while reducing the increase of congestion.

EMV-centered scheme. The EMV performance is always the paramount goal for the EMV-centered scheme [25] and it achieves better EMV performance and suffers slightly longer waiting time and queue length of the conflicting directions than the fuzzy logic-based scheme [25].

TABLE III

PARAMETERS OF DEEP Q-NETWORK

Model parameter	Value
Minimal signal switch time interval	5s
Update frequency of target Q network (M)	5
Gamma (γ)	0.9
Exploration rate (ε)	0.01
Learning rate	0.001
Batch size	32

TABLE IV

REWARD COEFFICIENTS

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9
-5	-5	-1	-0.5	-0.5	-0.5	-0.2	-0.5	-0.2

ESP scheme. The recent work [29] proposes elastic signal preemption (ESP) to obtain the schedules of TSC for the fast pass of EMVs through intersections by formulating quadratic programming problems. Authors in [29] also study the path planning for EMVs, and to fairly compare different methods, we modify this baseline to only control traffic signals for EMVs and ignore the impact of path planning on EMVs.

3) Deep Q-Network Setup: The parameter setting and reward coefficients for our deep Q-network are shown in Tables III and IV, respectively.

We set the minimal signal switch time interval as 5s to enable our method to flexibly control the traffic signal as in previous work [19], [21]. We investigate the combinations of different coefficients in the reward adopting the grid search method, and finally get the combination of coefficients as shown in Table IV. We set $T_{max} = 45s$ and $D_{min} = 50m$ in the reward design according to the performance of EMVs under the greedy preemption scheme.

In our simulations, all the EMVs randomly depart from the 16 entrances of the traffic network and EMV arrivals are generated by Poisson distribution with a certain average arrival rate, including 1 EMV/min, 1 EMV/2min, 1 EMV/5min, 1 EMV/10min and 1 EMV/30min. The simulations last 24 hours to obtain the statistical results of a large volume of experimental data, and then we not only provide the mean values of different metrics to demonstrate the average performance of different schemes, but also offer their standard deviation values to compare their stability. Our simulations are conducted on a microscopic traffic simulation platform SUMO, in which, we can replay the trajectories of vehicles (except for EMVs) from the real-world traffic networks.

B. Performance Comparison

In this part, we will test the performance of emergency vehicles on different metrics to demonstrate the superiority of our method over five different baseline methods.

1) Overall Performance of Emergency Vehicles: First, we will show the overall EMV performance (including the travel time and speeds) under different schemes. Note that

TABLE V
DELAY PERFORMANCE OF EMERGENCY VEHICLES

Various Schemes	EMV Accumulated waiting time (s)	EMV Stop count	EMV Num. of suffering red light
Intelligent scheme	**3.2** ± **3.29**	**0.98** ± **1.19**	**0.03** ± **0.17**
Greedy preemption scheme	3.41 ± 5.28	1.06 ± 1.23	0.04 ± 0.21
ESP scheme	7.93 ± 13.71	1.54 ± 1.32	0.12 ± 0.23
EMV-centered scheme	14.28 ± 18.21	2.34 ± 1.91	0.51 ± 0.53
Fuzzy logic-based scheme	17.15 ± 20.75	2.87 ± 1.96	0.57 ± 0.55
No preemption scheme	110.2 ± 24.08	5.18 ± 2.66	2.28 ± 1.62

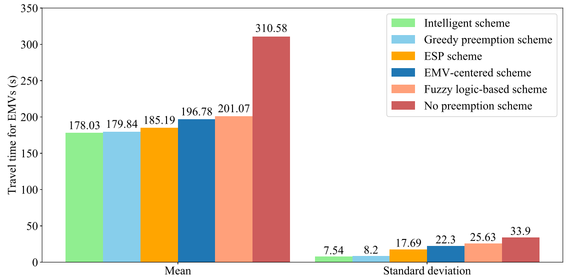

Fig. 8. Mean and standard deviation of the travel time of EMVs in the traffic network.

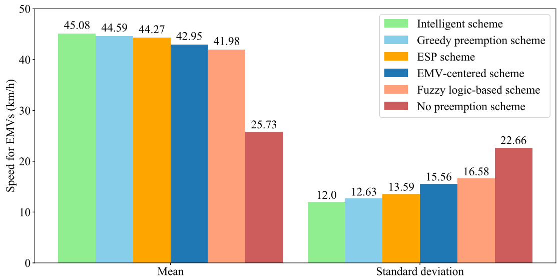

Fig. 9. Mean and standard deviation of the speeds of EMVs in the traffic network.

both the mean and standard deviation of different metrics are significant for the performance evaluation since we aim to ensure the fast and stable pass of all EMVs.

We first compare our scheme with the baselines in terms of the EMV travel time. Travel time of an EMV is the time that an EMV spends in the traffic network through the whole journey. From Fig. 8 we can observe that our Intelligent scheme achieves the smallest mean and standard deviation values of EMV travel time, which indicates that EMVs can enjoy faster and more stable journeys in the traffic network by implementing our scheme. Greedy preemption scheme outperforms the other four baselines since it offers a green signal promptly when an EMV enters the intersection. This often works well most of the time for the EMVs but fails when there are other EMVs simultaneously arriving from conflicting directions, while our schemes can deal with this challenge well. ESP scheme performs slightly worse than the greedy preemption scheme but better than the other three, which manifests the effectiveness of finding non-intrusive signal schedules for the fast pass of EMVs. No preemption scheme performs the worst as it offers no TSC scheduling for the fast pass of EMVs.

Similarly, as shown in Fig. 9, our Intelligent scheme obtains the highest mean speed and its performance is most stable. The mean and standard deviation values of EMV speed in the five baselines are in line with the results shown in Fig. 8, and we can conclude that EMVs can enjoy faster and more stable journeys in the traffic network with our scheme.

2) Delay Performance of Emergency Vehicles: To further evaluate the delay performance of EVMs, we then show more comparisons of EMV metrics, including accumulated waiting time, stop count, and the number of suffering red lights to validate the driving smoothness and offered green wave level of EMVs in the traffic network.

Accumulated waiting time is the total waiting time for an EMV idling in front of red lights in the traffic network, which is a crucial indicator to judge the effectiveness of TSC for EMV preemption. It can be seen from Table V that, our intelligent scheme achieves the shortest accumulated waiting time of EMVs in both mean and standard deviation values even compared with the greedy preemption scheme. In particular, compared with the recent ESP scheme, our method achieves 59.65% drop in the mean value and 76.00% drop in the standard deviation value, demonstrating a significant performance gain in terms of accumulated waiting time. This is because there may be more than one EMV passing the intersection from conflicting directions at the same time, and the greedy preemption scheme has to handle it with FCFS policy, in which one of them has to wait until the other leaves. Even though the two EMVs do not enter the intersection simultaneously, the greedy preemption scheme occupies the green phase of the conflicting direction and can lead to a dreadful traffic condition for the next EMV from the previous conflicting direction.

Stop count is regarded as the number of brakes or sharp decelerations for an EMV during the journey. This is a representative metric to evaluate the experience for both drivers and passengers, especially for the patients in ambulances. From Table V we can observe that, by implementing our intelligent scheme, the average and standard deviation values of stop count have been greatly reduced compared with other five baseline methods. In particular, compared with the recent ESP scheme, our method achieves 36.35% drop in the mean value and 9.85% drop in the standard deviation value, demonstrating a significant performance gain in terms of stop count.

Table V also shows the number of suffering red lights for EMVs in the traffic network under different schemes. In line with the results above, our intelligent scheme suffers the least number of red lights with a mean value of 0.03, demonstrating that almost all of the EMVs enjoy the green waves in the traffic network, resulting in red lights which is comparable and even slightly smaller than that of the greedy

TABLE VI

PERFORMANCE OF CONFLICTING DIRECTIONS

Various Schemes	Queue length	Waiting time (s)
Intelligent scheme	**0.73** ± **1.14**	**9.28** ± **19.99**
Greedy preemption scheme	3.08 ± 4.75	41.65 ± 102.45
ESP scheme	1.54 ± 2.52	26.43 ± 50.71
EMV-centered scheme	2.01 ± 3.68	31.91 ± 83.63
Fuzzy logic-based scheme	1.98 ± 3.63	30.76 ± 81.79
No preemption scheme	1.58 ± 2.63	27.48 ± 52.93

preemption scheme. Compared with the recent ESP scheme, our method achieves 75.00% drop in the mean value and 26.09% drop in the standard deviation value, manifesting a significant performance gain for the number of suffering red lights. EMVs with ESP, EMV-centered, and fuzzy logic-based schemes suffer more red lights than the intelligent and greedy preemption ones, but their mean values are all less than one. While for the no preemption scheme, EMVs suffer more than two red lights on average, which is clearly unacceptable in the real world.

Therefore, from the comparisons above, we can conclude that our intelligent scheme achieves the best EMV performance in terms of various metrics. In particular, our intelligent scheme even achieves better EMV performance than the greedy preemption scheme (an extreme case that is clearly favorable for EMVs) since it performs better in the case that various EMVs arrive from conflicting directions. This validates the effectiveness of RECAL Algorithm in our intelligent scheme to deal with multiple EMVs simultaneously around an intersection.

3) Performance of Conflicting Directions: Then, we will illustrate the performance (including queue length and waiting time) of EMV conflicting directions to investigate the impact under different schemes.

As illustrated in Table VI, the greedy preemption scheme suffers the longest queue length and waiting time of the EMV conflicting directions since it occupies too much green time for the EMV direction. Compared with the no preemption scheme, the fuzzy logic-based and EMV-centered schemes suffer longer queue length and waiting time, while ESP scheme achieves slightly better performance on EMV conflicting directions. Compared with the recent ESP scheme, on average our intelligent scheme achieves 52.60% drop in queue length and 64.89% drop in waiting time, and for the stability performance (standard deviation values), our method achieves 54.76% drop in queue length and 60.58% drop in waiting time. And our intelligent scheme achieves the shortest queue length and waiting time in both mean and standard deviation values. This is because our method not only occupies little green time of vehicles from conflicting directions, but also intelligently distributes green time of the EMV direction to conflicting directions without sacrificing the EMV performance, thus improving the traffic performance of the conflicting directions.

C. Case Study

We first visualize the traces of an EMV with a route of five road segments (its route indicated by the orange arrow in Fig. 7) as a case study, to demonstrate the individual EMV performance under different schemes. As illustrated in Fig. 7, dotted lines represent the positions of four intersections. We can observe that under our intelligent scheme, this EMV encounters no red light and enjoys a green wave before passing the fourth intersection (located at around 2000m from departure), taking about 160s to finish its trip. Under the greedy preemption scheme, this EMV suffers one red light at the second intersection, this is because there is another EMV traveling from the conflicting directions at the second intersection and it arrives the intersection before this EMV. Under the ESP scheme, this EMV encounters one red light at the first intersection and travel smoothly until the end of the journey, taking about 200s. While under the EMV-centered and fuzzy logic-based scheme, this EMV travels with more accumulated waiting time, and finally arrives with about 210s and 230s, respectively. Under the no preemption scheme, this EMV unfortunately endures all the four red lights, a long travel time of about 305s before passing the stop line of the fourth intersection.

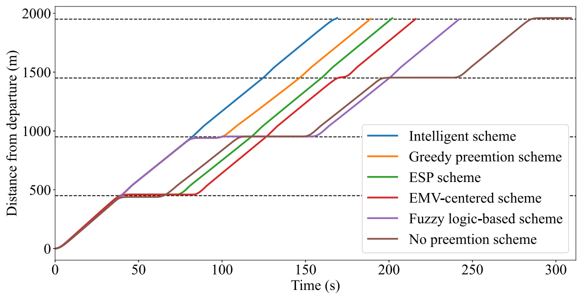

Fig. 10. Case study: Traces of an EMV with a route of five road segments in the traffic network under different schemes.

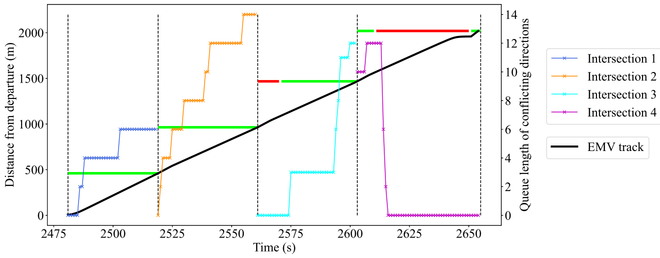

Fig. 11. Case study: Details of traffic light switching and queue length of conflicting directions for an EMV under the greedy preemption scheme.

To further show the performance of intelligent and greedy preemption schemes, we extract and visualize the details of another EMV when it is passing through the traffic network (the route of this EMV is also indicated by the orange arrow in Fig. 7). As shown in Fig. 11, when this EMV enters the traffic network, the traffic signals keep green at Intersections 1 and 2, which is in line with the expectation that the greedy preemption scheme will give priority to this EMV to ensure its fast pass in each intersection. While at Intersections 3 and 4, we can see that the green time for this EMV has been greatly reduced. In particular, the red time accounts for nearly 90% of the whole time when the EMV

TABLE VII
PERFORMANCE OF VARIATIONS OF OUR COMPREHENSIVE METHOD

Methods	EMV Travel time	EMV Stop count	EMV Num. of suffering red light	Non-EMV Queue length	Non-EMV Waiting time
Ours − BS	179.71 ± 12.15	1.14 ± 1.43	0.09 ± 0.21	1.56 ± 2.26	24.83 ± 47.53
Ours − PI	184.23 ± 18.31	2.56 ± 2.78	0.14 ± 0.31	**0.64** ± **1.06**	**8.92** ± **18.14**
Ours − PI + HC	182.94 ± 13.63	2.08 ± 1.82	0.11 ± 0.23	1.18 ± 2.32	17.54 ± 37.36
Ours	**178.03** ± **7.54**	**0.98** ± **1.19**	**0.03** ± **0.17**	0.73 ± 1.14	9.28 ± 19.99

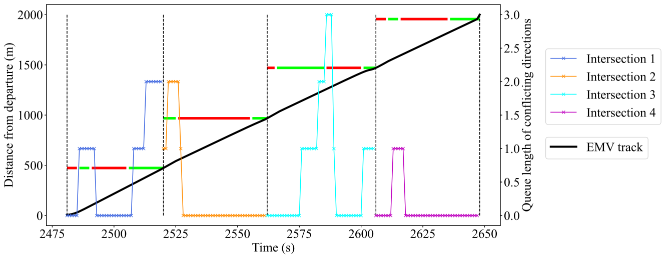

Fig. 12. Case study: Details of traffic light switching and queue length of conflicting directions for an EMV under our intelligent scheme.

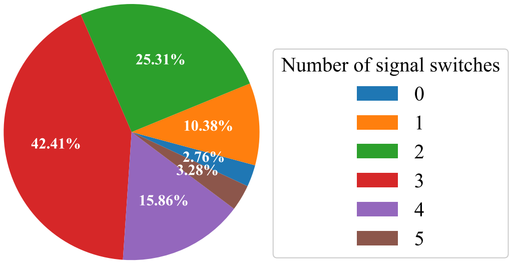

Fig. 13. The percentage of number of signal switches during the stay of an EMV under our intelligent scheme.

is passing through Intersection 4, leading to the waiting of this EMV in front of the red light. This is because there is at least one EMV traveling from conflicting directions and the greedy preemption scheme has to serve the conflicting directions first, resulting in a degraded performance for the EMV we are observing.

As illustrated in Fig. 12, we can observe that when the EMV passes through each intersection under our intelligent scheme, the traffic light switches more frequently, from 2 to 3 times. These flexible red and green changes can ensure an expeditious emergency response for the EMV and alleviate the negative impact on conflicting directions, which can be clearly observed from the queue length of conflicting directions. In Fig. 11, the real-time queue length of conflicting directions in each intersection is much bigger than that in Fig. 12, which is also in line with the results in Table VI. Therefore, by implementing our intelligent scheme, especially in the case that a variety number of EMVs are travelling to a disaster zone for the rescue, all the EMVs from different directions can pass intersections without being delayed due to the queuing vehicles in front of them.

D. Feasibility Analysis

To provide our model with more flexible control over the traffic signal, we let the minimal signal switch time interval be five seconds as shown in Table III. Since we set three-second yellow signal for all directions between any signal switch, the actual time interval between two phases is at least eight seconds. In addition, vehicles will generally slow down when passing the intersection, especially when there is an EMV, therefore the traffic safety under this setting is guaranteed. However, a strategy that frequently switches the traffic signal will lead to terrible experience for drivers. Hence, we will study whether our method would frequently switch the traffic signal during the passage of an EMV.

In Fig. 13, we present the percentage of different times of signal switches under our intelligent scheme. Overall, the number of switches during the pass of an EMV is from zero to five, and the most extreme five-time case happens only in 3.28%. Moreover, four-time signal switch only happens in around 15.86% of all cases. It can also be noted that, there is only fewer than four traffic signal switch among more than 80% passes of EMVs. Therefore, in most cases, the times of signal switch are not more than four, and in practice, the signal switch time interval is several times of the minimal value (e.g., 10s, 15s, 20s...) plus the three-second yellow light, which indicates that our method does not tend to trigger frequent signal switches during the pass of an EMV and is feasible in practice.

E. Ablation Study

We will compare the performance of three variations of our model, i.e., one without Balanced Sampling (BS), one without Punitive Indicators (PI) C_1 and C_2 in Eq. (4), and one replacing Punitive Indicators (PI) with High EMV-related Coefficients in the reward (HC), to validate the effectiveness of the proposed PI terms and coefficients in our reward design. In Table VII, ours represents our comprehensive method.

For the variation without BS, we randomly take samples from the memory when training our model. For the variation without PI, we set w_1 and w_2 in Table IV as zero and retain all other coefficients. For the variation that substitutes HC for PI, we use new reward coefficients as shown in Table VIII. Compared with the original reward coefficients in Table IV, w_1 and w_2 are set to zero, the EMV-related coefficients are doubled and other coefficients are halved, thus further strengthening the EMV performance in the reward design.

From Table VII we can observe that the variation without BS also obtains a positive performance while our comprehensive method is outstanding and stable over all metrics, which validates the effectiveness and importance of BS in our comprehensive method.

TABLE VIII

REWARD COEFFICIENTS WITH HIGH EMV-RELATED WEIGHTS

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9
0	0	-2	-1	-1	-1	-0.1	-1	-0.1

As shown in Table VII, the variation without PI obtains the shortest non-EMV queue length and waiting time (our comprehensive method achieves similar results in terms of the two metrics), but its EMV performance is worse than our comprehensive method, especially for the EMV duration, which is increased by 3.5% on average and 142.8% in the standard deviation. This verifies that the design of PI in the reward is significant to the EMV performance in our comprehensive method.

When we design a reward with HC and without PI, as illustrated in Table VII, compared with the one without PI, EMV performance is better while non-EMV traffic becomes much worse. Over all the metrics, its performance is worse and less stable than our comprehensive method, which further validates the effectiveness of our proposed terms of PI in the reward and the reasonable settings for the reward coefficients in Table IV on ensuring the overall performance.

VI. Conclusion and Future Work

Timely response of EMVs is imperative for saving lives and reducing property damage. Based on deep reinforcement learning, in this work, we propose a novel intelligent signal control for EMV preemption so as to improve the overall traffic efficiency without sacrificing the EMV performance in various cases. The ablation study proves that the design of our critical components, including pivotal factors of rewards and their calculations, and concept of current pair to accomplish balanced sampling, impel the remarkable performance of our model. An extensive set of experiments based on traffic data in a real-world traffic network with multiple intersections has been conducted to demonstrate the feasibility and effectiveness of our method compared with the other five baseline methods over a multitude of performance metrics. Based on the multi-agent deep reinforcement learning model with shared parameters, our intelligent model can be easily applied to large-scale traffic networks with any number of intersections in smart cities.

Neural architecture is significant to the performance of deep learning models, and different designs of neural networks, such as separate neural encodings for EMV and system states, and recurrent neural networks, may achieve performance gains. Much remains to be done for the exploration of different neural networks to further improve the EMV and overall performance in the future.

References

[1] D. Payne, "Poor ambulance response causes 700 deaths annually in Ireland," *Brit. Med. J.*, vol. 321, no. 7270, p. 1176, 2000.

[2] R. Sánchez-Mangas, A. García-Ferrrer, A. de Juan, and A. M. Arroyo, "The probability of death in road traffic accidents. How important is a quick medical response?" *Accident Anal. Prevention*, vol. 42, no. 4, pp. 1048–1056, Jul. 2010.

[3] *Standard for the Organization and Deployment of Fire Suppression Operations, Emergency Medical Operations, and Special Operations to the Public by Career Fire Departments*. National Fire Protection Association, Quincy, MA, USA, 2001.

[4] Y. Chen, Y. Yang, W. Peng, and H. Wang, "Influence and analysis of ambulance on the containment of COVID-19 in China," *Saf. Sci.*, vol. 139, Jul. 2021, Art. no. 105160.

[5] L. Qi, M. Zhou, and W. Luan, "Emergency traffic-light control system design for intersections subject to accidents," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 170–183, Jan. 2016.

[6] J. Paniati and M. Amoni, *Traffic Signal Preemption for Emergency Vehicle: A Cross-Cutting Study*. Washington, DC, USA: US Federal Highway Administration, 2006.

[7] I. Arnold. (Mar. 2018). *Statistics on Emergency Vehicle Accidents in the United States*. [Online]. Available: https://www.fettolawgroup.com/Personal-Injury-Blog/2018/March/Statistic%s-on-Emergency-Vehicle-Accidents-in-the.aspx

[8] C. Suthaputchakun and Y. Cao, "Ambulance-to-traffic light controller communications for rescue mission enhancement: A Thailand use case," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 91–97, Dec. 2019.

[9] E. J. Nelson and D. Bullock, "Impact of emergency vehicle preemption on signalized corridor operation: An evaluation," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1727, no. 1, pp. 1–11, Jan. 2000.

[10] A. J. Miller, "Settings for fixed-cycle traffic signals," *J. Oper. Res. Quaterly*, vol. 14, no. 4, pp. 373–386, Dec. 1963.

[11] O. Younis and N. Moayeri, "Employing cyber-physical systems: Dynamic traffic light control at road intersections," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2286–2296, Dec. 2017.

[12] Y. Ren, Y. Wang, G. Yu, H. Liu, and L. Xiao, "An adaptive signal control scheme to prevent intersection traffic blockage," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1519–1528, Jun. 2016.

[13] M. Guo, P. Wang, C.-Y. Chan, and S. Askary, "A reinforcement learning approach for intelligent traffic signal control at urban intersections," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 4242–4247.

[14] N. Kumar, S. S. Rahman, and N. Dhakad, "Fuzzy inference enabled deep reinforcement learning-based traffic light control for intelligent transportation system," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 4919–4928, Aug. 2021.

[15] F. Rasheed, K.-L. A. Yau, R. M. Noor, C. Wu, and Y.-C. Low, "Deep reinforcement learning for traffic signal control: A review," *IEEE Access*, 2020.

[16] H. Wei, G. Zheng, V. Gayah, and Z. Li, "Recent advances in reinforcement learning for traffic signal control: A survey of models and evaluation," *ACM SIGKDD Explor. Newslett.*, vol. 22, no. 2, pp. 12–18, Jan. 2021.

[17] P. Palos and A. Huszak, "Comparison of Q-Learning based traffic light control methods and objective functions," in *Proc. Int. Conf. Softw., Telecommun. Comput. Netw. (SoftCOM)*, Sep. 2020, pp. 1–6.

[18] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

[19] L. Li, Y. Lv, and F.-Y. Wang, "Traffic signal timing via deep reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 3, no. 3, pp. 247–254, Jul. 2016.

[20] E. Van der Pol and F. A. Oliehoek, "Coordinated deep reinforcement learners for traffic light control," in *Proc. Learn., Inference Control Multi-Agent Syst. (NIPS)*, 2016, pp. 1–8.

[21] H. Wei, G. Zheng, H. Yao, and Z. Li, "IntelliLight: A reinforcement learning approach for intelligent traffic light control," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 2496–2505.

[22] H. Homaei, S. R. Hejazi, and S. A. M. Dehghan, "A new traffic light controller using fuzzy logic for a full single junction involving emergency vehicle preemption," *J. Uncertain Syst.*, vol. 9, no. 1, pp. 49–61, 2015.

[23] M. Miletic, B. Kapusta, and E. Ivanjko, "Comparison of two approaches for preemptive traffic light control," in *Proc. Int. Symp. ELMAR*, Sep. 2018, pp. 57–62.

[24] B. Kapusta, M. MiletiC, E. Ivanjko, and M. Vujic, "Preemptive traffic light control based on vehicle tracking and queue lengths," in *Proc. Int. Symp. ELMAR*, Sep. 2017, pp. 11–16.

[25] M. Cao, Q. Shuai, and V. O. K. Li, "Emergency vehicle-centered traffic signal control in intelligent transportation systems," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 4525–4531.

[26] W. Ma, Y. Liu, and X. Yang, "A dynamic programming approach for optimal signal priority control upon multiple high-frequency bus requests," *J. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 282–293, 2013.

[27] A. Khan, F. Ullah, Z. Kaleem, S. U. Rahman, H. Anwar, and Y.-Z. Cho, "EVP-STC: Emergency vehicle priority and self-organising traffic control at intersections using Internet-of-Things platform," *IEEE Access*, vol. 6, pp. 68242–68254, 2018.

[28] A. Pozzi, S. Bae, Y. Choi, F. Borrelli, D. M. Raimondo, and S. Moura, "Ecological velocity planning through signalized intersections: A deep reinforcement learning approach," in *Proc. 59th IEEE Conf. Decis. Control (CDC)*, Dec. 2020, pp. 245–252.

[29] W. Min, L. Yu, P. Chen, M. Zhang, Y. Liu, and J. Wang, "On-demand greenwave for emergency vehicles in a time-varying road network with uncertainties," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 7, pp. 3056–3068, Jul. 2020.

[30] E. C. Eze, S. Zhang, and E. Liu, "Vehicular ad hoc networks (VANETs): Current state, challenges, potentials and way forward," in *Proc. 20th Int. Conf. Autom. Comput.*, Sep. 2014, pp. 176–181.

[31] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States," *Proc. IEEE*, vol. 99, no. 7, pp. 1162–1182, Jul. 2017.

[32] H. Wei *et al.*, "CoLight: Learning network-level cooperation for traffic signal control," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1913–1922.

[33] G. Zheng *et al.*, "Learning phase competition for traffic signal control," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 1963–1972.

[34] P. A. Lopez *et al.*, "Microscopic traffic simulation using SUMO," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2575–2582. [Online]. Available: https://elib.dlr.de/124092/

[35] M. J. Khabbaz, W. F. Fawaz, and C. M. Assi, "A simple free-flow traffic model for vehicular intermittently connected networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1312–1326, Sep. 2012.

[36] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, p. 1805, Aug. 2000.

[37] A. S. M. Bakibillah, M. A. S. Kamal, C. P. Tan, T. Hayakawa, and J.-I. Imura, "Event-driven stochastic eco-driving strategy at signalized intersections from self-driving data," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8557–8569, Sep. 2019.

[38] C. Sun, J. Guanetti, F. Borrelli, and S. J. Moura, "Optimal eco-driving control of connected and autonomous vehicles through signalized intersections," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3759–3773, May 2020.

[39] M. Asaduzzaman and K. Vidyasankar, "A priority algorithm to control the traffic signal for emergency vehicles," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–7.

Miaomiao Cao received the B.Eng. degree in electric power engineering and automation from Shanghai Jiao Tong University, China, in 2012, and the M.Sc. and Ph.D. degrees in electrical and electronic engineering from The University of Hong Kong, Hong Kong, China, in 2016 and 2021, respectively. She was a Visiting Ph.D. Student in electrical engineering and computer science with the Massachusetts Institute of Technology from August 2019 to January 2020. Her research interests include the Internet of Things, big data, deep reinforcement learning, and intelligent transportation systems.

Victor O. K. Li (Life Fellow, IEEE) received the S.B., S.M., E.E., and Sc.D. degrees in electrical engineering and computer science from MIT. He is the Chair of information engineering and the Cheng Yu-Tung Professor in sustainable development with the Department of Electrical and Electronic Engineering (EEE), The University of Hong Kong. He is the Director of the HKU-Cambridge Clean Energy and Environment Research Platform and the HKU-Cambridge AI to Advance Well-being and Society Research Platform, which are interdisciplinary collaborations with Cambridge University. He was the Head of EEE, an Associate Dean (Research) of engineering and managing, and the Director of Versitech Ltd. He serves on the board of Sunevision Holdings Ltd., listed on the Hong Kong Stock Exchange and co-founded Fano Labs Ltd., an artificial intelligence (AI) company with his Ph.D. student. Previously, he was a Professor of electrical engineering at the University of Southern California (USC), Los Angeles, California, USA, and the Director of the USC Communication Sciences Institute. He served as a Visiting Professor at the Department of Computer Science and Technology, University of Cambridge, from April to August 2019. His research interests include big data, AI, optimization techniques, and interdisciplinary clean energy and environment studies. He is a fellow of the Hong Kong Academy of Engineering Sciences, IAE, and HKIE. In January 2018, he was awarded a USD 6.3M RGC theme-based research project to develop deep learning techniques for personalized and smart air pollution monitoring and health management. Sought by government, industry, and academic organizations, he has lectured and consulted extensively internationally. He has received numerous awards, including the PRC Ministry of Education Changjiang Chair Professorship at Tsinghua University, the U.K. Royal Academy of Engineering Senior Visiting Fellowship in Communications, the Croucher Foundation Senior Research Fellowship, and the Order of the Bronze Bauhinia Star, Government of the HKSAR.

Qiqi Shuai received the B.Eng. degree in information engineering from Shanghai Jiao Tong University, Shanghai, China, in 2012, and the Ph.D. degree in electrical and electronic engineering from The University of Hong Kong, Hong Kong, China, in 2016. He was a Post-Doctoral Fellow with the Department of Electrical and Electronic Engineering, The University of Hong Kong, from 2016 to 2019. He is currently a Principal Scientist with Lighthorse Asset Management. His research interests include big data analytics and deep reinforcement learning applications in trading and intelligent transportation systems.