2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI) | 978-1-6654-0898-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICTAI52525.2021.00048

Attn-CommNet: Coordinated Traffic Lights Control On Large-Scale Network Level

Jiashi Gao, Xinming Shi, James J.Q. Yu, *Senior Member, IEEE*
Department of computer science and engineering
Southern University of Science and Technology
Shenzhen, China
12131101@mail.sustech.edu.cn, xxs972@student.bham.ac.uk, yujq3@sustech.edu.cn

***Abstract*—Traffic lights control could be regarded as a multi-agent coordinated problem. A model-free reinforcement learning (RL) approach is a powerful framework for solving such coordinated policy-making problems without prior environmental knowledge. In order to approach a global policy, communication among agents needs to be built. To enable dynamic and scalable communication, we propose a new RL model, CommNet based on Local Attention Mechanism (Attn-CommNet), which uses local selection and attention mechanism between hidden layers to facilitate cooperation. We evaluated the proposed method using synthetic and real word traffic flows under multi-scale road networks. The results demonstrate that the proposed method can get better performance in multi-scale problems, especially large-scale problems compared to the state-of-the-art methods.**

***Index Terms*—reinforcement learning, multi-agent control, coordinated traffic lights control**

I. Introduction

Traffic congestion in urban areas has caused serious problems [1], [2], e.g., long waiting time, high fuel consumption, and increased harmful emissions. One of the effective approaches to solve congestion is to control traffic lights more intelligently. Due to the signal control strategies of intersections are highly interdependent, it is crucial to coordinately control traffic signals in a large region [3].

Reinforcement learning (RL) techniques are effective for coordinately controlling of traffic signals. In early approaches [4]–[6], the intersection agents do not share information and parameters but update their own networks independently. These decentralized methods cause non-stationary for conflicts among agents' independent optimal policies. The centralized RL [3], [7], where the joint states of multiple intersections could be learnt, solves the issue of optimal policy conflicts. However, it cannot learn well due to the curse of the dimensionality in the joint-state space. Another state-of-the-art RL method, CoLight [8], employs local neighboring information to the graph attentional network [9] to participate in the decision-making of the target agent. This network avoids the dimensionality issue, but the neighboring agents for communication is pre-defined and fixed, i.e., cannot be changed considering the traffic dynamics.

This work is supported by the Stable Support Plan Program of Shenzhen Natural Science Fund No. 20200925155105002 and by the General Program of Guangdong Basic and Applied Basic Research Foundation No. 2019A1515011032. James J.Q. Yu is the corresponding author.

In order to improve the performance for coordinated traffic lights control, we propose a novel RL model called Attn-CommNet where communication among agents conducted dynamically. The proposed model is also scalable for large scale road networks. Our work makes the following major efforts:

- We construct a locally connected CommNet to achieve a dynamic and scalable communication scope.
- Attention mechanism is introduced to assign weights for hidden layer states of neighboring agents, where the weights are related to the influence degree of neighboring agents on the target agent.
- We conduct experiments on synthetic and real world datasets to verify the effectiveness and scalability of the proposed method.

The remainder of this paper is organized as follows. Section II reviews the related work on coordinated traffic light control. Section III gives the problem formulation modeling the coordinated traffic light control problem as a Markov decision process. Section IV presents the details about the proposed CommNet based on local attention mechanism. Section V discusses the experiments results. Concluding remarks are described in Section VI.

II. Related Work

The fixed-time traffic lights control [10] is not suitable enough for congestion avoidance under dynamic traffic flows. Recently, RL has been adopted by transforming the coordinate traffic lights control problem as a Markov decision process (MDP). The mainly RL based approaches in coordinated traffic lights control can be divided into two categories: centralized and decentralized methods.

Prashanth *et al.* [7] trained a centralized model to decide the joint-actions for all intersections. Chen *et al.* [3] tackled the problem of multi-intersection traffic signal control for large-scale networks using centralized RL techniques and pressure-based coordination. Kuyer *et al.* [11] estimated the optimal joint action by sending locally optimized messages among connected agents. The centralized RL could obtain global joint-actions meanwhile, but it suffers the curse of dimensionality in large scale road network.

978-1-6654-0898-1/21/$31.00 ©2021 IEEE
DOI 10.1109/ICTAI52525.2021.00048

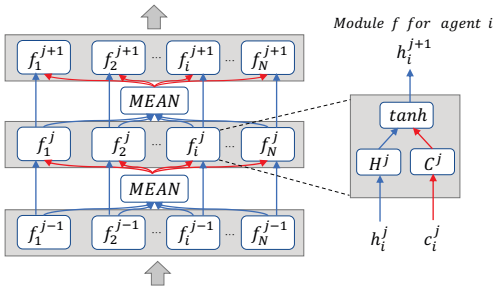

Fig. 1. Basic CommNet Model

Decentralized RL is proposed for coordinated traffic lights control under partial observability. The states of neighboring intersections are commonly used as observations to make policies for one target agent. Nishi *et al.* [12] developed an RL-based traffic signal control method that employs a graph convolutional neural network (GCNN) to extract features among multiple roads. Wei *et al.* [8] proposed a CoLight model, which used graph attention networks to facilitate communication. However, in these methods, the pre-defined and fixed number of neighboring agents might lead to the local optimal.

To address the shortcomings of previous methods, in this paper, we modify the CommNet [13] to realize a dynamic and scalable communication scope for every agent. The basic CommNet structure, as shown in the Fig. 1, is composed of full-connected layers and *mean* aggregation modules, which generate the communication vectors. Each f_i^j takes the hidden state h_i^j and the communication vector c_i^j as input and outputs h_i^{j+1}. Although CommNet is centralized, the corresponding coefficient matrices H^j and C^j are the same in each module f_i^j, preventing the curse of dimensionality. According to the Kinenmatic-wave theory [14] that the upstream intersections have larger influence than downstream intersections, we replace the mean communication mechanism in CommNet by the attention mechanism to assign different weights to multiple neighboring agents. We also apply the local connections to replace the full connections in the hidden layers. When every target agent communicates with its neighboring agents in the hidden layers, it also communicates with the non-neighbors indirectly by attention mechanism among different layers. Therefore, our proposed network will not only guarantee the overall communication ability of CommNet, but also can reduce the number of training parameters effectively.

III. Problem Formulation

In the coordinated traffic lights control problem, the traffic flow is stochastic and can not be modeled accurately, which can be regarded as a random process. Meanwhile, the conditional probability distribution of the future state only depends on the current state. Therefore, we model the coordinated traffic lights control problem as an MDP which is defined by five components $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the finite action space of all feasible actions, $\mathcal{P}$ is the state transition probability, $\mathcal{R}$ is the reward, and γ is the discount

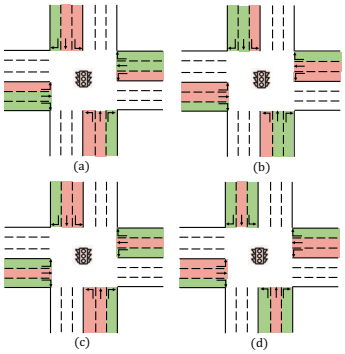

Fig. 2. The illustration of an intersection with four phases. (a) West Straight and East Straight; (b) North Straight and South Straight; (c) West Left and East Left; (d) North Left and South Left.

factor decaying rewards over time. The details are presented as follows:

- State space: We define the state space as $S^t = [s_1^t, s_2^t, ..., s_N^t]$, where N is the number of intersections and s_i^t is the state of i-th intersection at time step t. The intersection state consists of the current signal phase and the number of vehicles on each lane connected with this intersection. Fig. 2 shows a city intersection with four phases.
- Action space: The action space contains all candidate actions. In this paper, the action space of coordinated traffic lights control problem is a full set of all candidate phases, which can be represented as $A^t = \{phase_1, phase_2, ..., phase_m\}$. m is the total number of phases. At each time step, every agent will choose a phase from action space as its action, indicating the next phase of the traffic signal. The action of agent i at time step t is defined as $a_i^t \in A^t$.
- Reward space: We aim at minimizing the travel time for all vehicles in specific region, which is hard to be optimized directly [8]. Hence, we leverage the queue average length on each lane of the i-th intersection as its reward r_i. The reward space is defined as $R^t = \{r_1^t, r_2^t, ..., r_N^t\}$.

The parameterized strategy function is defined as follows:

$$\pi_\theta(S, A) = p(A|S, \theta), \tag{1}$$

where p is an action distribution conditioned on state S and parameters θ. The quality of policy π_θ is accessed by the expected reward, which is defined as follows:

$$Q(S^t, A^t; \pi_\theta) = \mathbb{E}_{\pi_\theta}[\textstyle\sum_{k=t}^{T} \gamma^{k-t} R(S^k, A^k)], \tag{2}$$

where T is the total time steps in an episode. In an MDP with an unknown model, the state transition probability $p(S^{t+1}|S^t, A^t)$ is unknown. Hence, we cannot obtain the optimal solution of MDP by Bellman equation directly. Therefore, in order to address this issue, we adopt two model-free Q-networks, Q_ω and $Q_{\hat{\omega}}$, to obtain the optimal policy by interacting with the environment. According to Bellman equation, the target action-value $Q_{\hat{\omega}}$ is defined as:

$$Q_{\hat{\omega}}(S^t, A^t) = R^t + \gamma \max Q_{\hat{\omega}}(S^{t+1}, A^{t+1}). \tag{3}$$

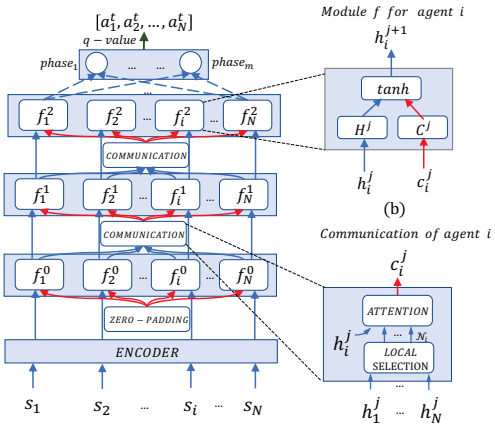

Fig. 3. An overview of our Attn-CommNet Model. (a) full model; (b) module f for agent i; (c) communication module for agent i.

The loss function is to minimize the mean squared error between the target action-value and the predicted action-value, which is defined as follows:

$$L_\omega = \mathbb{E}[(Q_{\hat{\omega}}(S^t, A^t) - Q_\omega(S^t, A^t))^2], \tag{4}$$

where the weight $\hat{\omega}$ is updated with weight ω to increase the learning stability by decorrelating predicted and target action-values.

IV. Proposed Method

In this section, we first introduce the proposed Attn-CommNet for coordinated traffic lights control. Then, the advantages of the proposed model are analyzed.

A. Attn-CommNet

Fig. 3(a) illustrates the overall structure of Attn-CommNet. The first layer of the model is an encoder layer. This layer takes state s_i as the input, and feature vector h_i^0 will be the output. The form of the encoder is problem dependent [13], in our work it is a single layer neural network with sigmoid activation function. The initial communication c_i^0 is set to zero. During the forward propagation, each agent sends its feature vector h_i^j and communication vector c_i^j to f_i^j module. In our work, the module f_i^j is a single linear network with a non-linear activation function tanh, which is depicted in Fig. 3(b). In order to achieve a more effective communication, we introduced a new module to calculate communication vector c_i^j in the hidden layers. As shown in Fig. 3(c), first, the local selection module is used to select neighboring states $\{h_k\}, k \in \mathcal{N}_i$ for each agent i from the concatenated vectors $[h_1, h_2, ..., h_N]$ of all agents. $\mathcal{N}_i$ is the neighboring scope of agent i which is determined by the geographic distance. After local selection, the attention module takes the concatenated vectors of neighbors as input. In order to obtain the impact degree of neighboring intersection on determining the policy for target intersection, in the attention module, we first embed the hidden-layer states of the target agent i and neighboring agents k by parameters W_T and W_N respectively. Then, the impact degree of agent k on determining the policy for agent i could be calculated, of which equation is given by:

Fig. 4. Non-neighboring communication (Black dotted line) between A and C

$$e_{<i,k>} = (h_i W_T)(h_k W_N)^T. \tag{5}$$

After that, softmax is used to normalize the importance index:

$$a_{<i,k>} = \text{softmax}(e_{<i,k>}) = \frac{\exp(e_{<i,k>}/\tau)}{\sum_{k \in \mathcal{N}_i} \exp(e_{<i,k>}/\tau)}, \tag{6}$$

where τ is the temperature factor. The communication vector c is obtained as follows:

$$c_i^j = \textstyle\sum_{k \epsilon \mathcal{N}_i} a_{<i,k>}^j h_k^j. \tag{7}$$

Based on the local attention module mentioned above, the module f_i^j can be formulated as follows:

$$h_i^{j+1} = f_i^j(h_i^j, c_i^j) = \tanh(H^j h_i^j + C^j c_i^j), \tag{8}$$

where H^j and C^j are the corresponding coefficient matrices. In the final layer of the model, a decoder layer with softmax as the activation function is used to output an action-value distribution over the action space.

B. Advantages

Learning in multi-agents environment is complex because all agents may interact with each other potentially. If the agents learn independently, the interactions among multiple agents will reshape the environment and the changes in the policy of one agent will affect the optimal policy of other agents. Thus it will lead to non-stationary [15]. Attn-CommNet improves the non-stationary issue by centralized learning. In addition, Attn-CommNet is suitable for large-scale problems. As shown in the Fig. 4, red and green areas are neighboring scopes for attention module of agents A and B, respectively. The solid line denotes direct-connections and the dashed line denotes indirect-connection. Fig. 4 indicates that the proposed local attention mechanism can construct indirect communication between the target agent A and its non-neighboring agent C by agent B. This dynamical and scalable neighboring scope promotes the scalability of our model.

V. Experiments

In this section, we first introduced the experimental settings. Then, we gave the details about traffic flow datasets, which contains two synthetic datasets and three real-world datasets of New York, Hangzhou and Jinan[1]. The compared methods and evaluation metrics are also presented. Finally, we conduct simulations on multi-scale datasets for coordinated traffic lights control performance evaluation and comparison.

[1]https://traffic-signal-control.github.io/

TABLE I
TRAFFIC FLOW DETAILS

Datasets	Intersections	Arrival Rate (vehicles/300s)			
		Mean	Std	Max	Min
New York	196	1256	264	1441	476
Jinan	12	524	98	672	256
Hangzhou	16	248	40	333	212

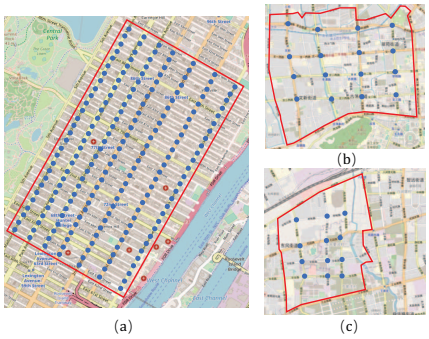

Fig. 5. Road networks for real-world datasets. Red polygons are the selected area and blue dots are intersections we control. (a) 28×7 New York road net; (b) 4×4 Hangzhou road net; (c) 4×3 Jinan road net.

A. Settings

We conduct experiments on CityFlow [16], which is an open source traffic simulator and is friendly to reinforcement learning. The simulator distributes the movement of traffic flow according to given traffic data including road network data, traffic flow data, etc. The simulator can feedback current state to the signal control method and generated traffic lights actions.

B. Datasets

1) Synthetic Data: We conduct experiments on two different synthetic data sets as described in [8], which are uni-directional and bi-directional traffic flow. The road network is a 6×6 grid network. Each intersection in the road network has four directions (West→East, East→West, South→North, North→South) and has 3 lanes (300 meters in length and 3 meters in width) on each direction. In bi-directional traffic flow, traffic flow is allowed on two bi-directions (West⇌East,North⇌South) with 300 vehicles/lane/hour and 90 vehicles/lane/hour, respectively. In the uni-directional traffic flow, traffic flow is only allowed on two uni-directions (West→East, North→South). In the experiments, uni-directional and bi-directional traffic flow are denoted as "6×6-uni" and "6×6-bi", respectively.

2) Real-world Data: The real data was collected in sub-area of three cities: New York, Hangzhou, and Jinan. These data sets are collected through different sources.

- New York traffic flow data are generated by the open-source taxi trip data[2]. Since this taxi trip dataset only gives the origin and destination geo-locations of taxi, the path of each vehicle is established by the shortest path.

[2]http://www.nyc.gov/html/tlc/html/about/trip record data. shtml

TABLE II
PERFORMANCE COMPARISON

Datasets	Average Travel Time (s)/Area Throughout			
	Fixed Time	CoLight	CommNet	Attn-CommNet
6x6-uni	277.6/2160	210.7/2169	212.9/2165	**209.3/2172**
6x6-bi	277.6/4326	216.1/4332	212.7/**4341**	**210.5**/4340
New York	1713.9/7617	1523.9/7973	1427.1/7852	**1367.6/8083**
Jinan	860.5/3788	344.1/5758	349.7/5667	**334.3/5765**
Hangzhou	756.9/2005	389.2/2726	392.0/2729	**384.0/2754**

- Traffic flow data in Hangzhou and Jinan are collected from the roadside surveillance cameras. By analyzing the camera data, identifier, duration time, as well as driving route of vehicles entering and leaving each intersection can be obtained.

The details of traffic flow datasets are shown in the Table I. The road networks are obtained from OpenStreetMap[3] as shown in Fig. 5.

C. Compared Methods

We compare our method to a baseline transportation method and two RL methods:

- Fixed Time [10]: The method gives a fixed cycle length with a predefined green ratio split among all the phases. It is one of the mainstream methods which is applied in practice at present.
- CoLight [8]: This method is an advanced decentralized reinforcement learning method with fixed neighbors for communication. This method is among the state-of-the-art RL methods for traffic lights control.
- CommNet [13]: A basic CommNet is used to compare the performance with the proposed Attn-CommNet in this paper.

D. Evaluation Metric

We use the following two representative measures to evaluate different methods [3]:

- Average travel time: It is commonly used to evaluate traffic lights control methods by calculating the average travel time of from entering to leaving the area.
- Area throughput: It is defined as the number of vehicles that completely pass the area during the simulation time.

E. Results

In this section, we investigate the performance of Attn-CommNet comparing to the baseline methods.

1) Performance Comparison: Table II shows the performance of Attn-CommNet and other compared methods with both synthetic and real-world datasets. Better performance is expected to be observed from a lower average travel time and higher area throughout. From the experimental results, we have the following observations:

- Attn-CommNet outperforms Fixed Time and CoLight in the five datasets, leading to both the best average travel time and the maximum area throughout.

[3]https://www.openstreetmap.org/

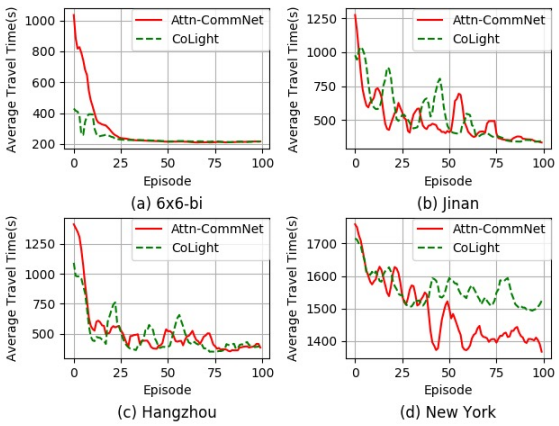

Fig. 6. Convergence performance of Attn-CommNet (red continuous curves) and CoLight (green dashed curves) during training. Curves are smoothed by moving average of five points.

- The advantage of our model is especially significant in large-scale road networks, such as New York with more intersections. Dynamic and scalable communication shows its advantages over fixed neighboring communication under large-scale problems.
- Compared to CoLight, the performance improvement of Attn-CommNet in Hangzhou and Jinan is not as obvious as in New York. This is because there are less number of intersections in Hangzhou and Jinan, incurring a limited scale of dynamic-neighboring communication.
- Compared to basic CommNet with *mean* aggregation for communicating, the performance are similar in synthetic datasets. The advantage of Attn-CommNet is more evident in real-world datasets, where road structures are more complex and traffic flows are more dynamic and uneven.

2) Model Convergence: In Fig. 6, we compare the performance convergence with respect to the episodes of Attn-CommNet and CoLight. Evaluated with four datasets, Attn-CommNet shows more stable convergence with smaller fluctuations compared to CoLight. In the CoLight with decentralized learning, the changes in the policy of one agent can affect the optimal policy of other agents, causing the non-stationary issue. Different from Colight, Attn-CommNet can improve convergence stability by centralized learning and scalable communication.

VI. Conclusion

In this paper, we proposed a new reinforcement learning model based on CommNet for coordinated traffic lights control. Specifically, our method achieved dynamic and scalable communication among agents. The proposed model also has a more stable convergence and avoids the curse of dimensionality in joint action space through parameter sharing. We conduct experiments to demonstrate the outstanding performance of our proposed method over state-of-the-art methods under synthetic and real-world datasets. Our proposed method has shown its superior performance especially in large-scale problem.

Future work will focus on enhancing the applicability of our model. One possible direction is to select the neighborhood scope considering more realistic factors besides geographic distance. We will also evaluate the practicality of the proposed model in other new ITS scenarios, such as coordinated charging of electric vehicles, etc.

References

[1] D. Zhao, Y. Dai, and Z. Zhang, "Computational intelligence in urban traffic signal control: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 485–494, 2011.

[2] S. Bharadwaj, S. Ballare, M. K. Chandel *et al.*, "Impact of congestion on greenhouse gas emissions for road transport in mumbai metropolitan region," *Transportation Research Procedia*, vol. 25, pp. 3538–3551, 2017.

[3] C. Chen, H. Wei, N. Xu, G. Zheng, M. Yang, Y. Xiong, K. Xu, and Z. Li, "Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control," in *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, New York, 2020, pp. 3414–3421.

[4] I. Arel, C. Liu, T. Urbanik, and A. G. Kohls, "Reinforcement learning-based multi-agent system for network traffic signal control," *IET Intelligent Transport Systems*, vol. 4, no. 2, pp. 128–135, 2010.

[5] B. C. Da Silva, E. W. Basso, F. S. Perotto, A. L. C. Bazzan, and P. M. Engel, "Improving reinforcement learning with context detection," in *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems*, Hakodate, 2006, pp. 810–812.

[6] K. Dresner and P. Stone, "Multiagent traffic management: Opportunities for multiagent learning," in *Proceedings of the 1st International Workshop on Learning and Adaption in Multi-Agent Systems*, Utrecht, 2005, pp. 129–138.

[7] L. Prashanth and S. Bhatnagar, "Reinforcement learning with function approximation for traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 412–421, 2010.

[8] H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, W. Zhang, Y. Zhu, K. Xu, and Z. Li, "Colight: Learning network-level cooperation for traffic signal control," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, Beijing, 2019, pp. 1913–1922.

[9] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[10] P. Koonce and L. Rodegerdts, "Traffic signal timing manual." United States. Federal Highway Administration, Tech. Rep., 2008.

[11] L. Kuyer, S. Whiteson, B. Bakker, and N. Vlassis, "Multiagent reinforcement learning for urban traffic control using coordination graphs," in *Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Antwerp, 2008, pp. 656–671.

[12] T. Nishi, K. Otaki, K. Hayakawa, and T. Yoshimura, "Traffic signal control based on reinforcement learning with graph convolutional neural nets," in *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC)*, Hawaii, 2018, pp. 877–883.

[13] S. Sukhbaatar, A. Szlam, and R. Fergus, "Learning multiagent communication with backpropagation," *arXiv preprint arXiv:1605.07736*, 2016.

[14] W. D. Hayes, "Kinematic wave theory," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 320, no. 1541, pp. 209–226, 1970.

[15] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3826–3839, 2020.

[16] H. Zhang, S. Feng, C. Liu, Y. Ding, Y. Zhu, Z. Zhou, W. Zhang, Y. Yu, H. Jin, and Z. Li, "Cityflow: A multi-agent reinforcement learning environment for large scale city traffic scenario," in *Proceedings of the World Wide Web Conference*, San Francisco, 2019, pp. 3620–3624.