2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)
October 8-12, 2022, Macau, China

2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC) | 978-1-6654-6880-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ITSC55140.2022.9921995

Learning Transformer-based Cooperation for Networked Traffic Signal Control

Chen Zhao, Xingyuan Dai, Xiao Wang, Lingxi Li, Yisheng Lv†, and Fei-Yue Wang

***Abstract*— Networked traffic signal control (NTSC) is essential for intelligent transportation systems. How to control multiple intersections in a cooperative way based on traffic conditions is critical for the success of NTSC. This paper proposes a Transformer-based cooperation mechanism (TCM) with the consideration of dynamic modeling and scale requirements simultaneously for large-scale traffic network control. Considering the physical constraints in traffic scenarios, a relative position encoding is designed to embed into TCM to characterize traffic conditions better. With the shared TCM module, intersection controllers could adequately exploit spatial-temporal correlations and adaptively capture global traffic dynamics, guiding them to explore collaborative traffic strategies more efficiently. Experimental results on two real-world datasets demonstrate that the suggested strategy greatly outperforms the state-of-the-art methods.**

I. INTRODUCTION

As urban traffic issues become more and more serious, traditional means are no longer adapted to the growing demand. In this regard, smart cities are expected to be an effective paradigm [1]. Researchers and engineers have made great efforts on smart cities construction: traffic big data mining [2], [3], automated driving [4], [5], [6], human-machine hybrid augmented intelligence [7], [8], [9], networked traffic signal control (NTSC) [10], [11], [12], and etc. In this research, we concentrate on NTSC with the goal of improving traffic conditions and shorting travel time. Indeed, the traffic status among intersections are spatio-temporally related. How to facilitate the cooperation of multiple signals is the core issue of NTSC.

Since the introduction of AlphaGo [13], deep reinforcement learning (DRL) has become a popular tool for decision making in complex systems. In particular, DRL algorithms have achieved remarkable results in traffic management for their excellent characteristics of modeling and optimization simultaneously [14]. One way to achieve cooperation is building a centralized DRL model of multiple intersections and optimizing their joint actions [15]. This type of method could model traffic conditions from a global perspective, guiding intersection controllers to make collaborative decisions. However, these centralized methods face the curse of dimensionality with the increasing number of intersections, leading to poor convergence in the large-scale traffic network.

To satisfy the scale requirement in NTSC, researchers had proposed decentralized DRL methods for traffic signal control, which treat each intersection as an isolated agent and train it based on local observations. The cooperation of agents is done by sharing information among neighboring intersections, such as directly concatenating all neighboring observations, adding downstream information into reward [16], and using graph neural networks (GNNs) to aggregate information recursively [17]. However, these methods usually make short-sighted decisions since the localized horizon of intersections, and are prone to "bottleneck phenomenon" when adopting a larger receptive field .

It needs to balance traffic dynamics modeling and network scale requirement to achieve effective cooperation in NTSC. Recent advances of Transformer [18] shed new light on realizing the above idea. Its excellent architecture, namely the self-attention mechanism, has global receptive field and long-term memory, which could free intersection controllers from the locality restrictions like GNNs, and facilitate communication across the traffic system. Transformer can adaptively capture each other's dynamic influences represented by the attention distribution, further making cooperative decisions. In addition, benefiting from the tokenization and decentralized processing, Transformer-based cooperation has an input space linearly increasing with the number of intersections, unlike the quadratic increasing in centralized approaches. Hence, Transformer is applicable to large-scale traffic scenarios, and its learning process could be accelerated through the abundant computational resources of cloud computing.

Although Transformer is desirable for sequence tasks, it cannot be directly applied to handling the structured information in NTSC. It is because there are several immutable constraints in traffic scenarios, which have an impact on the information interaction among intersection controllers. For example, traffic flow is directional, along the lane lines, and intersection connectivity is also a non-negligible factor. These constraints make us face the directional information propagation issues [19], which cannot be ignored in the modeling process.

To address the above problems, we propose a Transformer-based cooperation mechanism by incorporating physical constraints. The main contributions of this paper are as follows.

- We propose a novel Transformer-based cooperation mechanism (TCM) for NTSC. It could model traffic

This work was supported in part by National Natural Science Foundation of China under Grant U1811463 and 62173329.

Chen Zhao, Xingyuan Dai, Xiao Wang, Yisheng Lv and Fei-Yue Wang are with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049, China. They are also with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China.

Lingxi Li is with the Transportation and Autonomous Systems Institute (TASI) and the Department of Electrical and Computer Engineering, Purdue School of Engineering and Technology, Indiana University–Purdue University Indianapolis (IUPUI), Indianapolis, 46202, USA.

†Corresponding Author.

conditions globally while adapting the scalability of the road network. With the shared TCM module, each intersection controller could learn correlations with others and adaptively capture traffic dynamics, further guiding it to make a more cooperative decision.

- Considering the physical constraints in traffic scenarios, we design a relative position encoding and embed it into TCM to model the information propagation. With such design, we are able to make the attention distribution more explanatory and realistic, hence better characterizing spatial relations among intersections.
- The proposed method is validated on two publicly real-world datasets: D_{Hangzhou} and D_{Jinan}. The experimental results demonstrate that TCM outperforms the state-of-the-art approaches in terms of convergence, generalization, and fairness.

The remainder of this paper is organized as follows. The formulation of NTSC is described detailedly in Section II. The implementation details are presented in Section III. Several experiments are conducted in Section IV to verify the performance of TCM in comparison to five baselines. Finally, this paper is summarized in Section V.

II. PROBLEM FORMULATION

In this paper, the NTSC problem is formulated as a MDP $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, which is constructed for multi-intersection scenarios composed of multiple typical intersections like in Fig. 1. The detailed description of corresponding terms are given as follows.

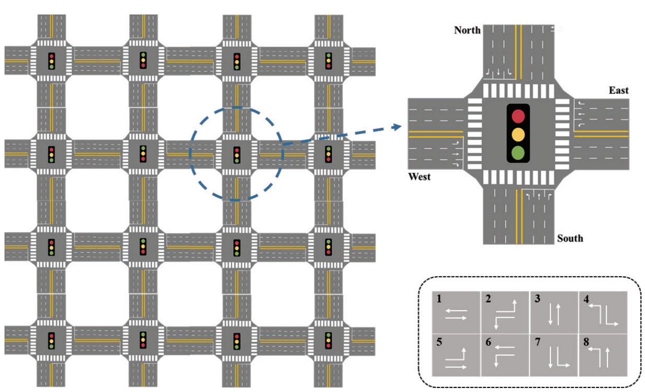


Fig. 1: An example of multi-intersection scenario.

State space $\mathcal{S}$**.** Suppose there are N intersections in a traffic system, each intersection is controlled by an agent. At time t, each agent treats its local observation, namely part of the whole system state $s^t \in \mathcal{S}$, as its state:

$$s_i^t = \left[v_i^t, k_i^t, p_i^t\right], i = 1, 2, ..., N, \tag{1}$$

where v_i^t, k_i^t are the average speed of vehicles and the road occupancy along each incoming lane; p_i^t is the current signal phase.

Action set $\mathcal{A}$**.** Given the current state s_i^t, agent i takes a phase p_i^t from the candidate set $\mathcal{A}_i$ as its action a_i^t based on the control policy π_i, forming a joint action $a^t \in \mathcal{A}$. Each action is executed during a time interval Δt. It should be pointed out that there will be t_y yellow light time between phase switching for a safety guarantee.

Transition probability $\mathcal{P}$**.** Traffic system transits the current state s^t to the next state s^{t+1} based on the probability $\mathcal{P}\left(s^{t+1} \mid s^t, a^t\right) : \mathcal{S} \times \mathcal{A}_i \times ... \times \mathcal{A}_N \to \Omega(\mathcal{S})$.

Reward $\mathcal{R}$**.** At time t, agent i takes an action a_i^t and obtains an immediate feedback r_i^t from environment by a reward function $\mathcal{S} \times \mathcal{A}_i \times ... \times \mathcal{A}_N \to \mathcal{R}$. In this paper, we define the total queue length on all incoming lanes at time t as the reward for agent i, namely $r_i^t = \sum_{l=1}^{l_n} u_{i,l}^t$.

Discount factor γ**.** To obtain the long-term optimal traffic policy, we set a discount factor $\gamma \in (0, 1]$ to balance exploration and utilization in policy optimization.

NTSC problem: At time t, agent i observes status s_i^t and takes an action a_i^t based on π_i to maximize cumulative reward $G_i^t = \sum_{t=0}^{T} \gamma^t r_i^t$. This paper aims to learn optimal policies $\{\pi_i\}_{i=1}^N$ to coordinate intersection controllers and improve overall traffic conditions.

III. METHODOLOGY

In this section, we present the overall architecture and operation processes of the proposed TCM for NTSC. Then, we detail its internal structure of different modules.

A. Overall Architecture

TCM is made up of three modules that run from the bottom to the top layer, as shown in Fig. 2.The first is the state embedding module which maps the sensor data (e.g., speed, density) to a high-dimensional latent space. The embedding states $h_1...h_N$ contains traffic information that each intersection controller obtains from its local observation.

The second is the Transformer-based communication module which is shared among intersection controllers to pass messages, further to fully exploit the correlations and capture global traffic dynamics for each of them. This module comprises two parts: information exchange and relative position encoding. The former is to exchange the locally obtained information $h_1...h_N$ among intersection controllers to produce a global view of the whole network traffic conditions. The latter is to enforce physical constraints (e.g., distance, direction and connectivity) to information propagation among intersections. Combined with the above two parts, each intersection controller learns the dynamic influence e_{ij} from the rest controllers on itself. Then, the i-th intersection controller can distribute its attention based on the learned importance and obtain the integrated information z_i. It is noted that Transformer could regard $h_1, ..., h_N$ as different tokens and execute decentralized processing of each token, which frees us from the curse of dimensionality. Finally, we adopt a Q-network $Q(s_i^t, a_i^t)$ to perform value-based policy updates.

It is noted that the proposed TCM conforms to the "local simple, remote complex" design principle [20]. The shared Transformer-based communication module could be implemented in a remote traffic operation center (RTOC), which makes guidance for intersection controllers while accelerating its learning process through cloud computing.

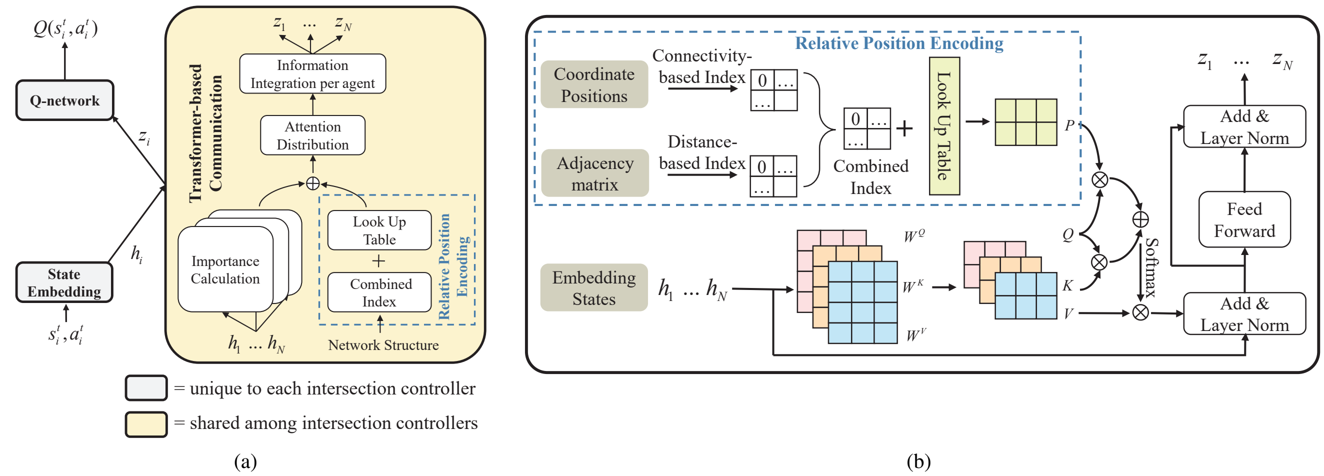


Fig. 2: (a) Architecture and operation processes of the proposed TCM. (b) Internal structure of the Transformer-based communication module.

B. State Embedding

At time t, each intersection controller obtains the current state s_i^t from the environment, and then map it to the latent space using a multi-layer perceptron:

$$h_i = \sigma(s_i^t W_e + b_e), \tag{2}$$

where $s_i^t \in \mathbb{R}^k$ is the state of agent i at time t and k is the feature dimension of s_i^t. The generated hidden state h_i denotes current traffic information of the i-th intersection.

C. Information Exchange

Importance Calculation. In RTOC, each intersection controller could exchange locally obtained information with others and judge their influence on itself. The calculation process can be given as:

$$e_{ij} = \frac{(h_i W^Q)(h_j W^K)^{\mathrm{T}} + p_{ij}}{\sqrt{d_z}}, \tag{3}$$

where e_{ij} denotes the importance of intersection j (sender intersection) on intersection i (receiver intersection). d_z is a scalar, and p_{ij} is a relative position encoding.

Attention distribution. Then, each intersection controller distributes its attention to capture traffic dynamics based on the normalized importance scores e_{ij}:

$$\alpha_{ij} = \frac{\exp(e_{ij}/\beta)}{\sum_{j \in N} \exp(e_{ij}/\beta)}, \tag{4}$$

where β is a temperature factor.

Information Integration. Finally, each intersection controller can model the global traffic dynamic change for itself based on its attention distribution, and calculate its representation as:

$$z_i = \sigma(W_z(\frac{1}{\mathrm{H}} \sum_{h=1}^{\mathrm{H}} \sum_{j=1}^{\mathrm{N}} \alpha_{ij}(h_j W^K + p_{ij}) + b_z)), \tag{5}$$

where z_i accumulates the key information from the senders for the receiver to make an efficient signal policy. Followed by the above self-attention mechanism, a fully connected feedforward network and layer normalization are added.

D. Relative Position Encoding

Physical constraints need to be embedded into Transformer to model proper information propagation in real-world scenarios. To do this, we index the relative relation between two intersection controllers as an integer r, then look up an learnable encoding table P by index to encode the relative position weights p_{ij}.

Distance-based index. Intuitively, closer senders are more influential to the receiver intersections than further ones. Given the coordinates of intersections, we can encode their relative relation based on Euclidean distances. Meanwhile, considering the directionality of traffic flow, we index the relative position based on horizontal and vertical Euclidean distances separately, and combine them into a 2D index. The formulation is described as:

$$\begin{aligned} p_{ij}^d &= \{P_r | r = r_{ij}^d,\ r_{ij}^d = r_{ij}^x + (2\beta+1) r_{ij}^y\}, \\ r_{ij}^x &= g(x_i - x_j) + \beta, \\ r_{ij}^y &= g(y_i - y_j) + \beta, \end{aligned} \tag{6}$$

where r_{ij}^x is a horizontal index, r_{ij}^y is a vertical index, and r_{ij}^d is a 2D index representing the direction of receiver i to sender j. $g(x) : \mathbb{R} \to \{y \in \mathbb{Z} \mid -\beta \le y \le \beta\}$ is a piecewise index function used in [21]. It is noted that we scale the dimension of distance from 100-meter level to meter level when indexing.

Connectivity-based index. Connectivity is an important factor in traffic stream because traffic flow can only travel along the links. Thus, traffic information propagation need to comply with this rule. Given the adjacency matrix of a road network, we can index it as:

$$\begin{aligned} p_{ij}^c &= \{P_r | r = r_{ij}^c\}, \\ r_{ij}^c &= \begin{cases} 1, & \text{if } i \text{ is adjacent to } j \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{7}$$

, where r_{ij}^c is an index of the connectivity between receiver i and sender j.

Combined Index. The connectivity-based method can be regarded as a particular case of the distance-based method: there is no distance with the connected intersections and there is an infinite distance with unconnected intersections. The two methods can be combined to consider all factors simultaneously.

$$p_{ij} = \{P_r | r = r_{ij}, r_{ij} = \left(\text{sign}\left(r_{ij}^d\right) - 2\beta^2 - 2\beta\right) \times \left(\beta^2 + \beta\right) \times \left(1 - r_{ij}^c\right) + r_{ij}^d/2\}. \tag{8}$$

E. Q-network

In this module, we adopt a Q-network along with parameters sharing to update the control policy of each intersection controller. The Q-value of each state is calculated as:

$$Q(s_i^t) = z_i W_p + b_p, \tag{9}$$

where s_i^t is the state of agent i at time t, z_i is the output of Transformer-based communication module, and $W_p \in \mathbb{R}^{m\times|\mathcal{A}|}$, $b_p \in \mathbb{R}^{|\mathcal{A}|}$ are learnable parameters. The action of agent i can be selected by maximizing $Q(s_i^t) \in \mathbb{R}^{|\mathcal{A}|}$.

A brief description of the training process is given as follows. At time t, the state transitions of intersections are stored into a replay buffer $\mathcal{D}$. During the training process, , a mini-batch is sampled from $\mathcal{D}$ and the policy is optimized by minimizing the loss defined in Eq.(10):

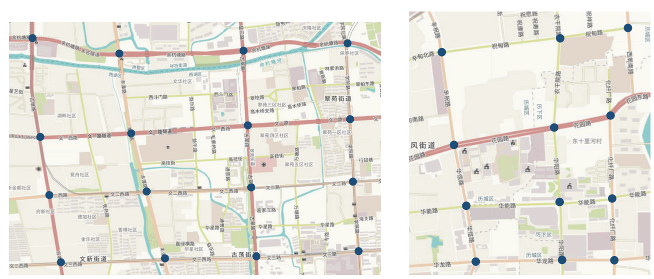
$$\mathcal{L}(\theta) = \mathbb{E}[((r_i^t + \gamma \max_{a'} Q(s_i^{t+1}, a'; \theta') - Q(s_i^t, a_i^t; \theta))^2], \tag{10}$$

where θ denotes all the trainable parameters in TCM, and θ' is a periodically updated copy of θ. Detailed training process for TCM is listed in Algorithm 1.

Algorithm 1 Pseudocode of TCM

Input: The network structure of traffic environment.
Output: The TCM model with parameters θ

```
Initialize all parameters θ and a replay buffer D.
for each epoch do
    Reset the traffic simulation environment.
    for t = 0 to T do
        Observe the state s^t = {s_i^t}_{i=1}^N from environment.
        for agent i = 1 to N do
            Perform traffic state embedding using Eq.(2).
            Encode the relative positions of using Eq.(8).
            Exchange information with ICs using Eq.(3).
            Distribute attention among ICs using Eq.(4).
            Integrate information and model traffic dynamics
              using Eq.(4).
            Compute Q-values using Eq.(9).
            With probability ε pick random action a_i^t, else
              a_i^t = max_{a'} Q(s_i^t).
        end for
        Interact with environment by the joint action a^t =
          {a_i^t}_{i=1}^N and receive reward r^t = {r_i^t}_{i=1}^N and next
          state s^{t+1} = {s_i^{t+1}}_{i=1}^N.
        Store the transition (s^t, a^t, r^t, s^{t+1}) into D.
    end for
    Sample a batch of transitions from D and minimize
      the loss defined in Eq.(10)
end for
```



(a) Gudang district, Hangzhou (b) Dongfeng district, Jinan

Fig. 3: The field testing area in Hangzhou (a) and Jinan (b).

IV. EXPERIMENTS

A. Datasets

In this section, two publicly available real-world datasets, D_{Hangzhou} and D_{Jinan}, are used to evaluate our model. Their road networks are downloaded from OpenStreetMap, the screenshots of which are shown in Fig. 3. Traffic flows are captured from intersection surveillance cameras, the detailed descriptions of which are listed in Table I. Each dataset comprises three configurations, which are collected from different times. They are different in the mean, maximum, minimum, and variance of the arrival rate, which means that there are different patterns of traffic dynamics in the given area. In this paper, config #1 is for training, and the other two are for testing. Experiment were performed on a well-known traffic simulation platform CityFlow [22], which is famous for its efficient computing performance and support for thousands of intersections.

B. Compared Methods

In this section, we list several popular methods in traffic engineering and existing literature as our compared baselines to verify the effectiveness of TCM. They can be categorized into two parts: traditional traffic control methods and RL-based methods. It is noted that none of all RL models including TCM have been pre-trained for a fair comparison.

Traditional traffic control methods:

- **FixedTime** [23]. This approach, which is commonly used in practice, uses a pre-defined signal phase cycle to govern intersections.
- **MaxPressure** [24]. This approach greedily selects the phase with the maximum pressure, which is a cutting-edge traditional method.

RL-based Methods:

- **PressLight** [16]. This method incorporates pressure to coordinate traffic signals in corridors, which shows a superior performance in multi-intersection control problems.

TABLE I: Data statistics of the real-world traffic dataset.

Datasets	# intersections	config	Arrival Rate(vehicles/min)			
			min	mean	max	std
D_{Hangzhou}	16	#1	75	109	134	19.87
		#2	40	50	67	8.24
		#3	39	116	230	63.72
D_{Jinan}	12	#1	50	105	136	19.79
		#2	43	72	101	15.15
		#3	69	92	111	9.51

- **CoLight** [17]. The most recent cutting-edge multi-intersection control method that achieves network-level cooperation via a graph attention network.
- **FRAP** [25]. A recently developed single-intersection traffic control method that captures the phase competition relationship between different traffic movements using a modified network. We apply the method in multi-intersection scenarios via parameters sharing.

C. Evaluation Metric

Following the existing work [16], [17], we compare the performance of various approaches using a regularly used metric in practice, namely average travel time of all vehicles. The travel time (in seconds) of one vehicle is defined as its spent time between origin to destination when completing the trip in a given district. Therefore, the comparative metric in this paper is calculated as:

$$\bar{t} = \frac{1}{n_c}\sum_{i=1}^{n_c}(t_i^d - t_i^o), 0 \le t_i^o \le t_i^d \le T, \tag{11}$$

where n_c is the number of vehicles that complete trips during the time interval $[0, T]$, T is the simulation time, t_i^o, t_i^d are the departure and arrival time for the i-th vehicle, respectively. A shorter travel time $\bar{t}$ means vehicles complete their trip in less time on average, indicating better traffic conditions and model performance.

D. Simulation Settings

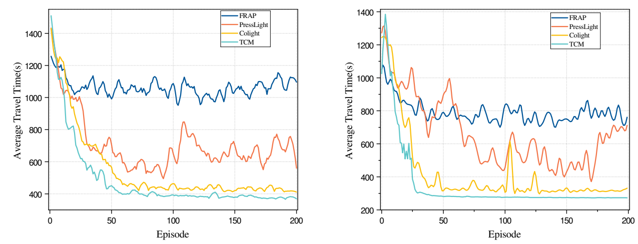
In the experiments, we set the simulation time $T = 3600s$, control period $\Delta t = 10s$ and yellow time $t_y = 3s$. All RL-based method are trained from scratch to compare their performance. The detailed training settings of TCM are shown in Table II.

TABLE II: Parameter settings

Parameters	Values
Learning rate	1e-3
Batch size	32
Number of units in hidden layer (m)	128
Optimization algorithm	Adam
Discount factor (γ)	0.95
ϵ for exploration	1→0.01 (decay: 0.995)
Replay buffer ($\mathcal{D}$) size	5000
Learning start	1000
Target model update interval	200
β in index functions	2

E. Performance Comparison

Convergence. For each dataset, we firstly train all RL-based methods on config #1 to compare their convergence. As described in Fig. 4, TCM outperforms the other three approaches with a faster convergence rate and a smaller convergence value. In particular, the TCM has a much better convergence stability compared to the SOTA algorithm Coligt. It is shown that TCM could coordinate intersections in an efficient way and stabilize their training process.



(a) D_{Hangzhou} config #1 (b) D_{Jinan} config #1

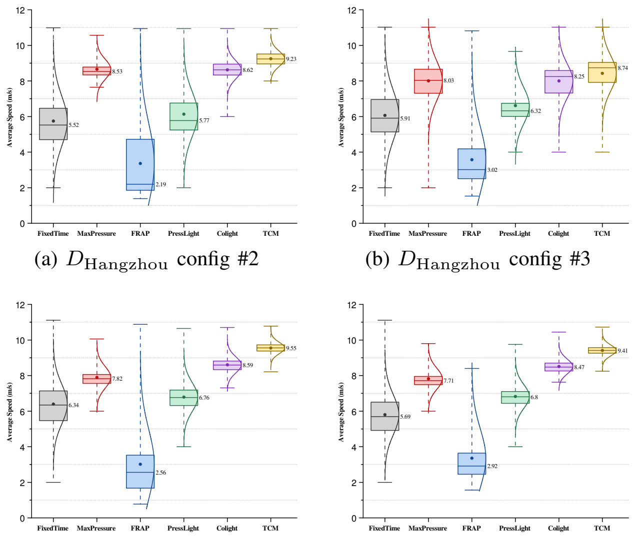
Fig. 4: Training curves of TCM (light blue continue curve) and other three RL-based method on D_{Hangzhou} (a) and D_{Jinan} (b).

Generalization. For each dataset, we have learned the RL-based control model on config #1. Then, we apply the trained model to govern the traffic environments with two other configurations and compare their performance with traditional traffic control methods. The average travel time of all methods is reported in Table III. It is observed that TCM surpasses all other methods in four different traffic scenarios, delivering the shortest travel time. This means that the traffic conditions are significantly better under the control of the TCM model.

TABLE III: Evaluation results on D_{Hangzhou} and D_{Jinan} w.r.t average travel time.

Algorithms	D_{Hangzhou}		D_{Jinan}	
	config #2	config #3	config #2	config #3
FixedTime	547.88	549.92	401.49	425.74
MaxPressure	365.47	412.07	327.34	315.69
FRAP	956.77	735.40	837.92	717.80
PressLight	509.52	483.32	372.97	360.51
Colight	363.70	404.96	299.25	290.65
TCM	**339.69**	**390.08**	**270.21**	**263.43**

Fairness. In order to visualize the driving experience of drivers under different control strategies, we statistically analyze the average speed of road vehicles during time interval $[0, T]$ and describe its distribution as a boxplot. The greater the average speed, the fewer stops drivers have to make during driving, and the better the driving experience. As illustrated in Fig. 5, we can find that TCM's median and mean are better than other methods and its variance is smaller than others. It means that drivers can have a smooth and fast driving experience under the TCM control strategy.



(a) D_{Hangzhou} config #2 (b) D_{Hangzhou} config #3

(c) D_{Jinan} config #2 (d) D_{Jinan} config #3

Fig. 5: Boxplots of vehicles' average speed evaluated on two configurations of D_{Hangzhou} and D_{Jinan}.

V. CONCLUSIONS

This paper proposes a Transformer-based cooperation mechanism (TCM) for networked traffic signal control. The proposed method could model traffic global dynamics for all intersections without numerical restrictions. To better characterize traffic information propagation in real-world scenarios, we design a relative position encoding considering physical constraints in TCM. Combining the above dynamic information modeling and static information restriction, TCM could learn the inter-agent relations among intersections and guide them to make a more cooperative decision. Comparative studies on two real-world datasets reveal that the suggested strategy outperforms state-of-the-art methods significantly. In future, we would perform sim2real and sim2sim traffic control for parallel intelligent transportation systems in a virtual-real interactive way.

REFERENCES

[1] Y. Lv, Y. Chen, J. Jin, Z. Li, P. Ye, and F. Zhu, "Parallel transportation: Virtual-real interaction for intelligent traffic management and control," *Chinese Journal of Intelligent Science and Technology*, vol. 1, no. 1, pp. 21–33, 2019.

[2] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2014.

[3] Y. Li, X.-Y. Lu, J. Wang, and K. Li, "Pedestrian trajectory prediction combining probabilistic reasoning and sequence learning," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 3, pp. 461–474, 2020.

[4] C. Zu, C. Yang, J. Wang, W. Gao, D. Cao, and F.-Y. Wang, "Simulation and field testing of multiple vehicles collision avoidance algorithms," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 4, pp. 1045–1063, 2020.

[5] W. Wang, X. Na, D. Cao, J. Gong, J. Xi, Y. Xing, and F.-Y. Wang, "Decision-making in driver-automation shared control: A review and perspectives," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 5, pp. 1289–1307, 2020.

[6] W. Qiu, W. ShangGuan, L. Chai, B. Cai, and J. Chen, "Parallel hierarchical control-based efficiency enhancement for large-scale virtual reality traffic simulation," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 4, pp. 145–162, 2022.

[7] F.-Y. Wang, J. Guo, G. Bu, and J. J. Zhang, "Mutually trustworthy human-machine knowledge automation and hybrid augmented intelligence: mechanisms and applications of cognition, management, and control for complex systems," *Frontiers of Information Technology & Electronic Engineering*, pp. 1–16, 2022.

[8] C. Ou and F. Karray, "Enhancing driver distraction recognition using generative adversarial networks," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 3, pp. 385–396, 2020.

[9] X. Li, P. Ye, J. Jin, F. Zhu, and F.-Y. Wang, "Data augmented deep behavioral cloning for urban traffic control operations under a parallel learning framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5128–5137, 2021.

[10] J. Jin, H. Guo, J. Xu, X. Wang, and F.-Y. Wang, "An end-to-end recommendation system for urban traffic controls and management under a parallel learning framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1616–1626, 2021.

[11] J. Jin, D. Rong, Y. Pang, F. Zhu, H. Guo, X. Ma, and F.-Y. Wang, "PRECOM: A parallel recommendation engine for control, operations, and management on congested urban traffic networks," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2021.

[12] M. Abdoos, "A cooperative multiagent system for traffic signal control using game theory and reinforcement learning," *IEEE Intelligent Transportation Systems Magazine*, vol. 13, no. 4, pp. 6–16, 2020.

[13] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, pp. 484–489, 2016.

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

[15] 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.

[16] H. Wei, C. Chen, G. Zheng, K. Wu, V. Gayah, K. Xu, and Z. Li, "Presslight: Learning max pressure control to coordinate traffic signals in arterial network," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1290–1298.

[17] 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*, 2019, pp. 1913–1922.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[19] H. A. Hashim, S. El-Ferik, and F. L. Lewis, "Neuro-adaptive cooperative tracking control with prescribed performance of unknown higher-order nonlinear multi-agent systems," *International Journal of Control*, vol. 92, no. 2, pp. 445–460, 2019.

[20] F.-Y. Wang, "Agent-based control for networked traffic management systems," *IEEE Intelligent Systems*, vol. 20, no. 5, pp. 92–96, 2005.

[21] K. Wu, H. Peng, M. Chen, J. Fu, and H. Chao, "Rethinking and improving relative position encoding for vision transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 033–10 041.

[22] Z. Tang, M. Naphade, M.-Y. Liu, X. Yang, S. Birchfield, S. Wang, R. Kumar, D. Anastasiu, and J.-N. Hwang, "Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8797–8806.

[23] A. Muralidharan, R. Pedarsani, and P. Varaiya, "Analysis of fixed-time control," *Transportation Research Part B: Methodological*, vol. 73, pp. 81–90, 2015.

[24] P. Varaiya, "Max pressure control of a network of signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 177–195, 2013.

[25] G. Zheng, Y. Xiong, X. Zang, J. Feng, H. Wei, H. Zhang, Y. Li, K. Xu, and Z. Li, "Learning phase competition for traffic signal control," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1963–1972.