

Integrating Multi-Graph Convolutional Networks and Temporal-Aware Multi-Head Attention for Lane-Level Traffic Flow Prediction in Urban Networks

Sun, Fengmei; Zhu, Hong; Tang, Keshuang; Xiong, Yingchang; Tan, Chaopeng; Tang, Zhixian

DOI

[10.1109/ITSC58415.2024.10920109](https://doi.org/10.1109/ITSC58415.2024.10920109)

Publication date

2025

Document Version

Final published version

Published in

Proceedings of the IEEE 27th International Conference on Intelligent Transportation Systems (ITSC 2024)

Citation (APA)

Sun, F., Zhu, H., Tang, K., Xiong, Y., Tan, C., & Tang, Z. (2025). Integrating Multi-Graph Convolutional Networks and Temporal-Aware Multi-Head Attention for Lane-Level Traffic Flow Prediction in Urban Networks. In *Proceedings of the IEEE 27th International Conference on Intelligent Transportation Systems (ITSC 2024)* (pp. 1878-1884). (IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC). IEEE. <https://doi.org/10.1109/ITSC58415.2024.10920109>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Integrating Multi-Graph Convolutional Networks and Temporal-Aware Multi-Head Attention for Lane-Level Traffic Flow Prediction in Urban Networks

Fengmei SUN, Hong ZHU*, Keshuang TANG, *Member, IEEE*, Yingchang XIONG, Chaopeng TAN, Zhixian TANG

Abstract—The urban signalized road network, characterized by its dynamic and complex nature due to frequent signal control adjustments and unpredictable demand fluctuations, presents significant challenges for predicting lane-level traffic flow. This study introduces the innovative MGCN-TAMA model, which addresses these challenges by integrating multi-graph convolutional networks with a temporal-aware multi-head attention mechanism. The proposed model employs three types of adjacency matrices—a geographical matrix, a signal matrix, and an attention matrix—to capture the complex spatial dependencies among various traffic approaches. Additionally, the model utilizes temporal-aware multi-head attention to discern the non-linear correlations in traffic variations over time. Tested on a real-world dataset from Tongxiang City, the MGCN-TAMA model significantly outperforms traditional models. Notably, in the first 30-minute prediction interval, our model achieves the lowest Mean Absolute Error, with 2.5649 vehicles per 5-minute span. These results underscore the effectiveness of combining graph-based methods with advanced attention mechanisms to enhance the accuracy of predicting lane-level traffic volumes in urban networks.

I. INTRODUCTION

Accurate traffic volume prediction (TFP) is essential for managing congestion, optimizing traffic operations, and enhancing urban mobility. Urban transportation systems are becoming more complex due to rapid urbanization and increased vehicle ownership, necessitating advanced TFP technologies based on new data source.

The recent surge in interest for data-driven TFP on networks can be classified into three distinct levels [1]. Initially, the regional level examines entire urban areas, delving into overarching traffic patterns and behaviors [2]. Moving to the road level, this approach primarily utilizes

traffic sensor data, such as from the PEMS dataset, to analyze the traffic flow dynamics through specific road segments and intersections [3], [4]. The third, the station level, targets transportation hubs like subway stations, bus stops, and parking lots, focusing on the micro-dynamics of individual nodes within the transport network [5].

With the increasing coverage of loop detectors and the development of automatic vehicle identification (AVI) or connected vehicles, achieving higher accuracy of TFP with finer granularity in urban areas becomes more applicable. Regarding the road level TFP at signalized intersections within urban road networks, this study field involves predictions on traffic flow [6], movement [7], [8], and queue length [9], [10], which is characterized by higher data intensity. Most current studies concentrate on isolated intersections, with only a few considering the temporal and spatial correlations among multiple intersections. Even more scarce are studies that have attempted network-wide traffic flow predictions at the lane level.

The emergence of this research gap is due to dynamic, complicated and fluctuating characteristics typical of interrupted flow within signalized networks. Meanwhile, the impact of signal indication and stochastic demand poses significant challenges for TFP modeling in urban areas. To address this issue, this study proposes a hybrid deep learning model based on multi-graph convolutional neural networks and temporal aware multi-head attention mechanism, named as MGCN-TAMA.

Firstly, the traffic flow graph is established by using the historical lane level traffic volume data and three type of adjacency matrices. Secondly, graph convolutional neural networks is used to extract dynamic and discontinuous spatial dependencies from multiple subgraphs and then apply gated linear unit to fuse them. Thirdly, an enhanced multi-head attention mechanism, specifically tailored for time series analysis, is proposed to capture the fluctuation of traffic flow across time dimensions. Finally, a comparison experiment with existing models is conducted using the AVI data from a real-world network in Tongxiang, Zhejiang. In addition, an ablation experiment is designed to demonstrate the effect of multi-graph convolutional neural networks and temporal aware multi-head attention modules. The contributions which make our work different from others are as follows:

Firstly, a comprehensive network-wide, lane-level traffic flow graph has been developed. By constructing the traffic

*This study was greatly supported by the National Natural Science Foundation of China Project (Grant No. 52302414 and No. 52372319), the Belt and Road Cooperation Program under the 2023 Shanghai Action Plan for Science, Technology and Innovation (No. 23210750500) and the Fundamental Research Funds for the Central Universities. (Corresponding author: Hong ZHU E-mail: hongzhu1990@tongji.edu.cn)

Fengmei SUN, Hong ZHU, Keshuang TANG and Yingchang XIONG are with the Key Laboratory of Road and Traffic Engineering, Ministry of Education & College of Transportation Engineering, Tongji University, Shanghai, China.

Chaopeng TAN is with the Department of Transport and Planning, Delft University of Technology Gebouw 23, Stevinweg 1, Delft, 2628 CN, Netherlands.

Zhixian TANG is with Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China.

graph with approach as vertices, the model not only simplifies spatial structures but also accommodates the varied lane configurations of each approach.

Secondly, multiple adjacency matrices are introduced to represent the spatial dependencies of the network. This approach addresses the discontinuities and dynamic characteristics of urban traffic flow influenced by traffic signal controls. Additionally, enhancements to the multi-head attention mechanism allow for a more effective capture of the temporal dependencies inherent in time series data.

Finally, to evaluate our model's predictive accuracy, the performance of our model is rigorously tested and compared with other advanced methods based on real-world data from urban road networks and traffic signal systems.

II. RELATED WORKS

Over the last few decades, numerous TFP methods have been proposed. Traditional statistical approaches like ARIMA and its variants, as well as machine learning techniques, have been applied in this field [11], [12]. With the advent of deep learning, several scholars have turned to recurrent neural networks for analyzing time series data, achieving notable success. A significant advancement came when [13] introduced a deep learning framework designed to extract spatiotemporal features, applying it specifically to highway traffic prediction. Following this, the use of Graph Neural Networks (GNN) emerged as a vital technique for analyzing road traffic flow. The innovative graph-convolutional encoder-decoder structure proposed in [14] integrates Graph Convolutional Networks (GCN) and Long Short-Term Memory (LSTM) networks to capture spatiotemporal features effectively. This structure includes a recurrent decoder that projects traffic flow predictions several steps into the future. Capturing the complex spatiotemporal dependencies of traffic flows using a static geographic structure has proven challenging. To address this, [15] introduced a dynamic graph model based on an attention mechanism to more accurately reflect dynamic spatial correlations. However, the correlations of real-time traffic states in urban signal-controlled networks remain underexplored.

More recently, some scholars have already attempted to tackle the TFP issue in signal-controlled road networks. Lee developed a deep learning model for predicting real-time lane-based traffic volumes at three consecutive intersections [7]. In a separate study, [16] exploited traffic volume data from upstream and downstream intersections to model spatiotemporal relationships and predict straight and left-turn movements at a corridor level. Additionally, the introduction of a novel graph-augmented neural model, the Graph-Augmented Neural Network Spatio-Temporal Reasoner (GANNSTER), aims to optimize traffic light systems by forecasting traffic throughput at signalized intersections [4]. Another innovative approach, ScR-DTG, proposed by [15], is a deep learning-based dynamic traffic graph model specifically designed to precisely predict movement-based traffic flow on a cycle-to-cycle basis. However, this model employs a simplified assumption, positing a constant upper limit for the passing flow. This assumption represents an unrealistic constraint that may limit the model's applicability in varying traffic conditions.

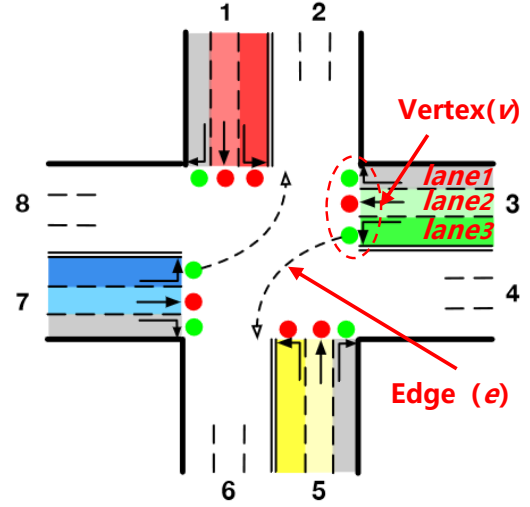


Figure 1. Diagram of the urban road network traffic graph

In conclusion, while deep learning-based methods have shown exceptional performance on highways, research into lane flow prediction in urban networks is still nascent. Urban road networks present uniquely complex and dynamic spatiotemporal traffic characteristics, which are influenced by signal control and stochastic demand generation. This complexity underscores a significant gap in current research. There is a pressing need for further investigation to develop and refine predictive models that are tailored to the intricate and fluctuating conditions of urban traffic systems.

III. METHODOLOGY

A. Preliminary

Graph-based theory is used to model the road network topology to predict the traffic volume of lanes in each approach. The traffic graph consists of vertexes, edges and adjacency matrices, denoted as $\mathcal{G} = \{V, E, A\}$, where V is the set of vertexes, each of which has a feature vector representing the traffic characteristics (e.g., traffic volume, speed, etc.), and E is the set of edges, and A is the adjacency matrix, which represents the connections between vertexes, and it defines how the traffic flow propagates through the graph. Since treating the intersection as a vertex directly will ignore the connection between each inlet lane within the intersection, while treating each lane as a vertex will complicate the road network topology. Considering the research objectives, we treat each approach as a vertex and is represented by x (Fig.1). Each approach consists of left-turn, right-turn, and straight lanes, and therefore can be regarded as a vector $x = [l_1, l_2, \dots, l_i, \dots, l_m] \in \mathbb{R}^M$ where l_i is the traffic volume of the lane i , and M denotes the number of lanes in the approach. The whole urban road network can be represented as $X = [x_1, x_2, \dots, x_j, \dots, x_n] \in \mathbb{R}^{M \times N}$, where N denotes the number of approaches. The T historical flow is denoted as $X_{t-T+1:t} = [X_{t-T+1}, X_{t-T+2}, \dots, X_{t-1}, X_t] \in \mathbb{R}^{m \times n \times T}$.

As in Equation (1), our problem is predicting future network-wide traffic volumes $\hat{Y}$ for the next H time

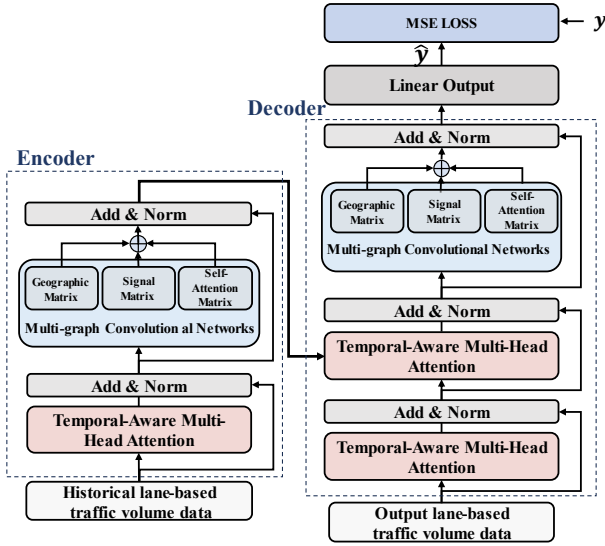


Figure 2. Overview of the proposed model

intervals, based on historical T time intervals $X_{t-T+1:t}$ and the traffic graph $\mathcal{G}$:

$$\hat{Y} = \{y_1, y_2, \dots, y_h\} = f(\mathcal{G}, X_{t-T+1:t}) \quad (1)$$

B. Overview of proposed hybrid model

To address the above problems, a hybrid deep learning model combined multi-graph convolution networks with temporal aware multi-head attention (MGCN-TAMA) is proposed (see Fig.2). The overall architecture of the model includes two parts: encoder and decoder. The structure offers an efficient framework for modeling and predicting time-series data. In this framework, the encoder processes the input sequence and produces a context vector, while the decoder utilizes this vector along with preceding prediction outcomes to generate the output sequence. By predicting multi-step outputs, the model can capture long-term trends and cyclical variations effectively. This approach ensures the continuity of results and mitigates the accumulation of errors over time, thereby enhancing prediction accuracy.

C. Spatial dependence of urban road network

1) Geographical matrix

The geographic adjacency matrix is first established based on the geometric connectivity of the road network. The geographical connection of the road network usually does not change in the short term. According to the approach i and j are direct upstream and downstream, the geographic adjacency matrix A_{ij}^{geo} takes the value of 1 or 0, where 0 indicates that approach i is not adjacent to approach j , and 1 indicates that approach i is adjacent to approach j . Taking the intersection in Figure 1 as an example, approach 3 is connected to approach 6, so the value of position (3, 6) in the geographic matrix is 1.

2) Signal matrix

Different signal control schemes such as green split and phase sequence have a great impact on the capacity of intersections. In practice, each intersection has a unique signal timing scheme that adjusts dynamically based on the traffic flow at different times of the day to improve accessibility. The

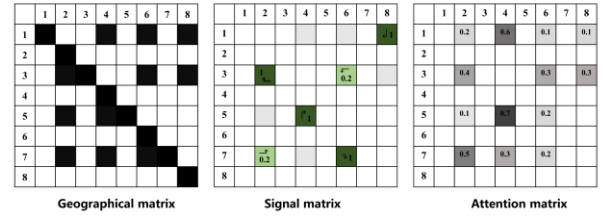


Figure 3. Three types of spatial dependence of urban road network

green split is one of the most important key parameters in traffic signal timing design, that is, the effective green time of a signal phase in a signal cycle. This is also crucial for short-term TVP based on flow direction in urban road networks. However, previous studies related to TVP have neglected the significance of signal control in urban road network scenarios. Although a large number of studies have achieved better prediction performance in highway scenarios using static geographic matrix to describe the spatial structure of roads, this paper proposes to extract the green split in urban road scenarios (g_{ij}) characterizing the dynamic spatial connection between upstream and downstream lane, and the signal matrix is denoted as in Equation (2):

$$A_{ij}^{sig} = g_{ij} = \frac{\text{green time of phase}_{ij}}{\text{cycle time}} \quad (2)$$

As shown in Figure 3, when the signal light of approach 7 turning left into approach 2 is green, the vehicle can pass. The signal matrix with green split as the element represents the passing power between the two approaches.

3) Attention matrix

In addition to the geographic and signal correlation between directly connected upstream and downstream roads, there can also be global semantic correlation between distant roads. To capture the relationship between the more distant roads in the network, this paper adopts an attention mechanism to adaptively assign different weights to the global vertexes to represent the global dynamic relationship between the roads as in Equation (3):

$$A_{ij}^{att} = \text{Attention}_{ij} \quad (3)$$

D. Multi-graph convolution networks

GCN was first proposed in 2017[17]. GCN learn higher-order feature representations by combining the graph's adjacency matrix with vertex feature and performing convolution operations directly on the graph. This allows each vertex to aggregate information from its neighbors, making GCN especially suitable for handling data with complex spatial dependencies, such as in traffic flow prediction for road networks. The formula for GCN is as in Equation (4)-(6):

$$X^{l+1} = f(X^l, A) = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X^{(l)} W^{(l)}) \quad (4)$$

$$\tilde{A} = A + I \quad (5)$$

$$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij} \quad (6)$$

where X^l is the network feature matrix of the l layer, and X^0 is the initial network feature matrix; $W^{(l)}$ is the trainable weight matrix of the l layer; σ is a nonlinear activation

function such as ReLU. A is the adjacency matrix of the graph, and I is the unit matrix, which is used to add self-connections to enhance the representation of the model; $\tilde{D}$ is the degree matrix of $\tilde{A}$.

The urban road network has multiple and complex spatial dependencies, this paper comprehensively considers three adjacency matrices represent various sub-graphs and use GCN to extract spatial features of each sub-graph then fuse them. To improve the learning efficiency of the network, this paper use a gating mechanism to help the model to pass useful information, Gated Linear Unit (GLU) is a simplified gating mechanism [18], which controls whether the information should pass through these layers or not. The formula is shown in Equation (7)-(8)

$$X^{l+1} = f(X^l, A_{Geo}) \oplus f(X^l, A_{Sig}) \oplus f(X^l, A_{Att}) \quad (7)$$

where $\oplus$ denotes the vector concatenate operation.

$$X^{l+1} = (X^{l+1} * W + b) \otimes \sigma(X^{l+1} * V + c) \quad (8)$$

where W , V , b , c are learning parameters and σ is the activation function, and $\otimes$ is the corresponding element-wise product.

E. Temporal aware multi-head attention

When dealing with sequence data, the multi-head attention mechanism is used to simultaneously consider the temporal relationships between different time steps. The output of a head's attention $head_i$ can be expressed as in Equation (9):

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (9)$$

where Q , K , V denote the query vector, key vector and value vector, respectively. W_i^Q , W_i^K and W_i^V are the linear transformation matrices of the head i , and $Attention$ is the attention computation function. In multi-head attention, the Self-Attention mechanism is generally used to compute the attention. The formula for Self-Attention is as in Equation (10):

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (10)$$

where d_k is the dimension of the key vector K . The softmax similarity is normalized and the weight of each key vector is calculated, then the weight is multiplied by the value vector and finally the weighted sum is performed to get the attention output.

Ultimately, the calculation of multi-head attention can be expressed as in Equation (11):

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (11)$$

where h denotes the number of heads, and $head_i$ denotes the output of the i_{th} head, and W^O is the output transformation matrix for projecting the result of multi-head attention to the desired dimension.

Although the multi-head attention mechanism has seen success in traffic flow prediction, it often struggles with time series data, particularly in capturing long-term temporal dependencies. This limitation can lead to the underestimation of the time dimension's significance. To address this, this paper proposes a Temporal-aware Multi-head Attention mechanism. By integrating Temporal Convolutional Neural

Network (TCN) to process the inputs Q and K , our approach enhances the model's ability to capture both local and global temporal features in the data.

The input query and key are first temporally convolved to capture the temporal features in the sequence data. Specifically, we can use the TCN [19] to transform the query Q and key K to obtain new query and key representations. Assume that the outputs of TCN are Q' and K' , whose dimensions are the same as the original inputs. Then, the transformed Q' and K' are inputted into the multi-head attention mechanism, respectively, and the attention score matrix of each head is computed. For the first i head, its attention is calculated as follows in Equation (12):

$$Attention(Q, K, V) = softmax(\frac{Q'K'^T}{\sqrt{d_{K'}}})V \quad (12)$$

The final step involves splicing or calculating the weighted sum of the results obtained from all attention heads to generate the ultimate output of temporal aware multi-head attention. By introducing TCN, temporal aware multi-head attention is able to better capture the temporal features in the sequence data, thus improving the prediction performance of the model.

F. Residual connection

Residual connections are crucial in deep learning models. These connections enhance the model by directly adding inputs to outputs, which helps preserve information from the initial inputs throughout the network. This is particularly beneficial in traffic flow prediction, where residual connections aid in capturing long-term trends and cyclical changes, thereby improving the model's accuracy and fit. The basic operation of a residual connection involves combining the input X^l with its corresponding output $f(X^l)$, resulting in the final output y as in Equation (13).

$$y = X^l + f(X^l) \quad (13)$$

where y denotes the final prediction result.

Residual connections enhance model training by allowing direct learning from input changes. These connections add input data directly to outputs, simplifying the learning of residual differences and accelerating training convergence. Additionally, skip connections help prevent gradient vanishing and explosion, boosting model stability and generalization. This approach significantly improves the performance of graph neural network-based models in capturing long-term trends and cyclical changes in traffic flow data, enhancing prediction accuracy.

IV. NUMERICAL EXPERIMENT

A. Data description and preparation

The proposed model was applied to a real-world urban road network in Tongxiang City, Zhejiang, China. This network, depicted in Fig.4, consists of a 4×4 grid with 16 intersections and 64 approaches. Each intersection is equipped with AVI to monitor vehicle entries per lane, and signal timing data were collected concurrently. Data collection spanned from September 20, 2020, to September 29, 2020, covering a total of ten days. Observations recorded between 0:00 and 5:00 a.m. were excluded from the analysis due to their extremely



Figure 4. Urban road network experiment scene

TABLE I. PREDICTION PERFORMANCE OF THE MODELS

Models	MAE	RMSE	MAPE
HA	3.3889	5.4671	0.2926
Seq2seq [20]	2.7777	4.207	0.2395
DISTN [7]	2.7745	4.1377	0.2369
GCN-LSTM [14]	2.7377	4.1846	0.2237
ScR-DTG [15]	2.7325	3.9728	0.2274
MGCN-TAMA	2.5649	3.9156	0.2262

low traffic volume. The whole data set is divided into periods at 5-minute intervals, and the traffic volume of the lane is calculated by the total number of vehicles passing the stop line every 5 minutes, and the signal timing data corresponding to the lane in each period is processed. After preprocessing, the AVI data yielded a dataset comprising 2,280 time-series entries for further analysis.

B. Experimental setting

70% of the dataset is divided into a training set, 10% into a validation set, and 20% into a test set. The data in the training set is randomly disrupted. Taking into account that the traffic flow in urban road networks is influenced by approximately one hour of historical data, we utilize one hour's worth of traffic flow records (segmented into 12 five-minute intervals) as input for our model to forecast the traffic conditions for the subsequent hour (similarly segmented into 12 five-minute intervals). The model is optimized using the Adam optimizer and utilizes Mean Squared Error (MSE) as the loss function. The evaluation indexes include Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The above indexes are calculated as shown in Equation (14)-(17)

$$MSE = \frac{1}{H * N * M} \sum_{t=1}^H \sum_{n=1}^N \sum_{m=1}^M (y_{tnm} - \hat{y}_{tnm})^2 \quad (14)$$

$$MAE = \frac{1}{H * N * M} \sum_{t=1}^H \sum_{n=1}^N \sum_{m=1}^M (y_{tnm} - \hat{y}_{tnm}) \quad (15)$$

$$RMSE = \sqrt{\frac{1}{H * N * M} \sum_{t=1}^H \sum_{n=1}^N \sum_{m=1}^M (y_{tnm} - \hat{y}_{tnm})^2} \quad (16)$$

$$MAPE = \frac{1}{H * N * M} \sum_{t=1}^H \sum_{n=1}^N \sum_{m=1}^M \left| \frac{(y_{tnm} - \hat{y}_{tnm})}{y_{tnm}} \right| \quad (17)$$

Additionally, several state-of-the-art models are utilized as baseline models to evaluate and compare the performance of the proposed models in this study.

- **HA**: The method is based on the assumption that future traffic flows are similar to past average flows. It provides a basic baseline for forecasting that can be used to compare and evaluate against more sophisticated models.
- **Seq2seq** [20] : The encoder processes the historical traffic data sequences using LSTM, and the decoder also uses LSTM to receive the hidden states of the encoder as initial states and gradually generates a sequence of traffic flow predictions for a future period of time.
- **DISTN** [7] : A structure called Deep Intersection Spatial-Temporal Network. The model takes into account spatial and temporal features by combining CNN and LSTM, respectively.
- **GCN-LSTM** [14]: A proposed data-driven approach based on a deep learning architecture combining GCN and LSTM, to predict hourly traffic flow at the intersection level with multiple time-steps.
- **ScR-DTG** [15]: A proposed deep learning-based signal control refined dynamic traffic graph model that introduces GCN, LSTM and temporal attention.

C. Performance comparison

Table 1 demonstrates the prediction performance of the proposed model compared to baseline models over the ensuing hour. Our model registers a MAE of 2.5649, a RMSE of 3.9156, and a MAPE of 0.2262, marking enhancements of 0.824, 1.5515, and 6.64% respectively over the HA. In the prediction task over the next hour (including 12 periods), the proposed model had the best overall performance. The results show that the hybrid model of multi-graph convolutional networks and temporal aware multi-head attention can better extract complex and dynamic spatio-temporal features to improve the accuracy of lane level traffic prediction. Furthermore, the performance of the GCN-LSTM model shows an improvement over CNN-LSTM, demonstrating that GCNs are more adept at extracting spatial features from road networks than CNN.

Further analysis of the detailed evaluation results for the next 5min-60min, as illustrated in Fig.5-6, reveals that the proposed MGCN-TAMA demonstrates a clear advantage in predicting traffic flow volume within the first 30 minutes, achieving the lowest MAE, RMSE and MAPE. Conversely, the ScR-DTG, GCN-LSTM model exhibits superior predictive performance in the 30-60minute interval. This disparity arises because the model proposed here incorporates a time convolution method for extracting temporal features, which offers enhanced performance in local temporal extraction and greater computational efficiency. In contrast, LSTM models excel in capturing longer-term dependencies within the time series. Consequently, the model presented in this study is

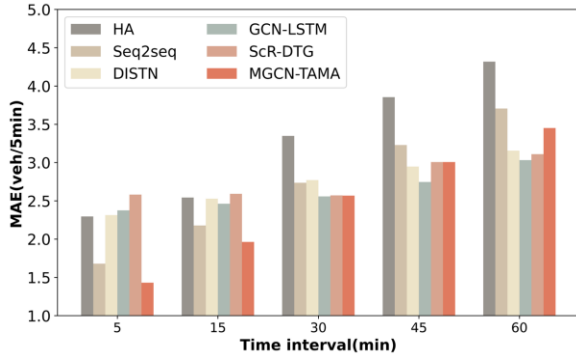


Figure 5. MAE prediction evaluation results in the next 1 hour

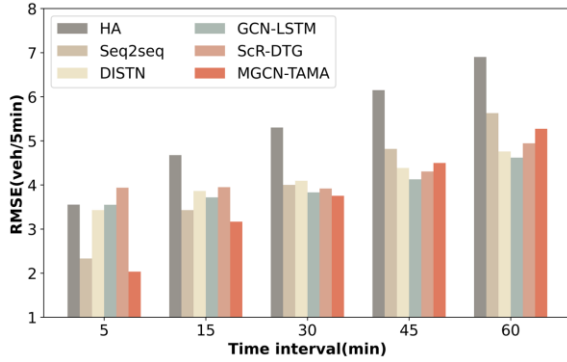


Figure 6. RMSE prediction evaluation results in the next 1 hour

significantly more efficient and accurate for real-time and rapid prediction of road network traffic conditions.

Prediction results of west approach of NO.15 intersection are depicted in Fig.7-9, which features three designated lanes for left, straight, and right turns. These figures compare the actual traffic values of these lanes with the predicted values for the next 5, 30, and 60 minutes. Each of the three turning lanes experiences different levels of traffic: the straight lane is the busiest, while the right-turn lane sees the least traffic. The results indicate that the model proposed in this study makes accurate predictions across lanes with varying traffic volumes, successfully forecasting both peak and off-peak traffic flows up to one hour in advance.

D. Ablation result

To demonstrate the individual contribution of each component in our proposed method, we conducted ablation experiments, the results of which are depicted in Fig.10-11. Each bar in the figure represents the performance of the model after the removal of a specific component. The findings indicate that the temporal-aware improved attention module exerts the most significant impact on the model's performance. Additionally, both the signal matrix and the attention matrix, which represent dynamic spatial information, contribute similarly to the model. Notably, the signal matrix demonstrates superior performance over the attention matrix in the 15-minute prediction evaluations. Our analysis suggests that, within the 15-minute time frame, traffic flow evolution is still largely confined to a smaller geographical area. Here, the signal matrix, which dynamically controls the propagation of upstream and downstream traffic flow and leverages

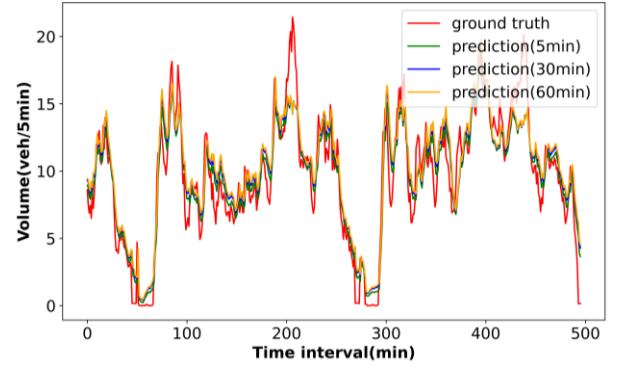


Figure 7. Comparison of predicted results in the left lane

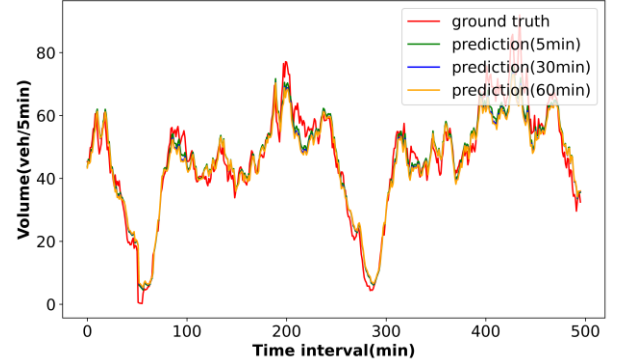


Figure 8. Comparison of predicted results in the straight lane

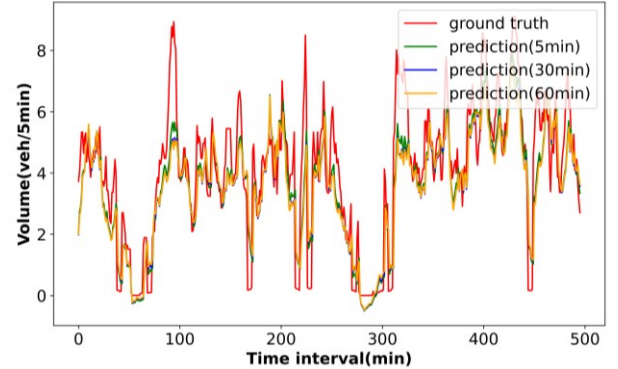


Figure 9. Comparison of predicted results in the right lane

theoretical traffic flow knowledge, provides more reliable support in this scenario.

CONCLUSION

This study proposed a hybrid model that integrates multi-graph convolutional networks with a temporal-aware multi-head attention mechanism, aimed at enhancing lane-level traffic flow prediction in signalized network. The proposed model seeks to effectively capture the complex spatiotemporal dependencies characteristic of urban traffic, potentially improving prediction accuracy and computational efficiency. Notably, experimental results show that our model achieves a significant improvement in forecasting traffic flow within the first 30 minutes, with the lowest MAE recorded at 2.5649 vehicle in each 5 minutes interval. This performance

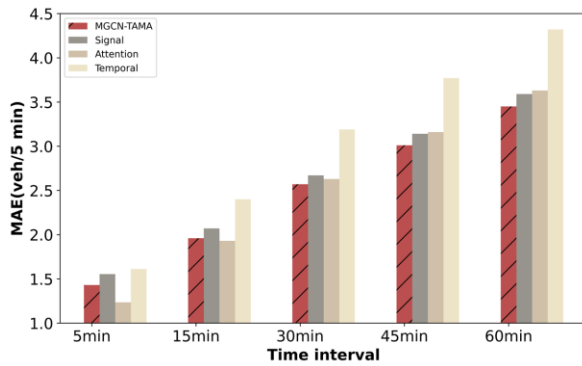


Figure 10. MAE of ablation experiment results

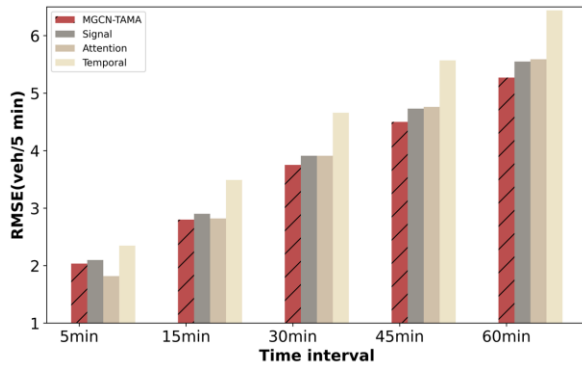


Figure 11. RMSE of ablation experiment results

demonstrates a substantial enhancement compared to traditional methods.

The incorporation of multiple graphs representing different spatial relationships, along with the innovative use of temporal convolution methods to extract temporal features, plays a crucial role in these results. While the findings are promising, ongoing efforts to refine the model's adaptability to real-time traffic changes remain crucial. Future research could include integrating real-time data inputs from a variety of sources such as IoT devices and vehicle sensors, aiming to refine the model's accuracy under dynamic traffic conditions. Such advancements would further enable the model to adjust predictions effectively in real-time, enhancing its utility for intelligent transportation systems.

REFERENCES

- [1] W. Jiang and J. Luo, "Graph Neural Network for Traffic Forecasting: A Survey," *Expert Syst. Appl.*, vol. 207, p. 117921, Nov. 2022, doi: 10.1016/j.eswa.2022.117921.
- [2] L. Bai, L. Yao, S. S. Kanhere, X. Wang, and Q. Z. Sheng, "STG2Seq: Spatial-temporal Graph to Sequence Model for Multi-step Passenger Demand Forecasting," arXiv, May 24, 2019. Accessed: Apr. 15, 2024. [Online]. Available: <http://arxiv.org/abs/1905.10069>
- [3] F. Zhou, Q. Yang, T. Zhong, D. Chen, and N. Zhang, "Variational Graph Neural Networks for Road Traffic Prediction in Intelligent Transportation Systems," *IEEE Trans. Ind. Inform.*, vol. 17, no. 4, pp. 2802–2812, Apr. 2021, doi: 10.1109/TII.2020.3009280.
- [4] C. Salort Sánchez, A. Wieder, P. Sottovia, S. Bortoli, J. Baumbach, and C. Axenie, "GANNSTER: Graph-Augmented Neural Network Spatio-Temporal Reasoner for Traffic Forecasting," in *Advanced Analytics and Learning on Temporal Data*, vol. 12588, V. Lemaire, S. Malinowski, A. Bagnall, T. Guyet, R. Tavenard, and G. Ifrim, Eds., in Lecture Notes in Computer Science, vol. 12588, Cham: Springer International Publishing, 2020, pp. 63–76. doi: 10.1007/978-3-030-65742-0_5.
- [5] B. Zhao, X. Gao, J. Liu, J. Zhao, and C. Xu, "Spatiotemporal Data Fusion in Graph Convolutional Networks for Traffic Prediction," *IEEE Access*, vol. 8, pp. 76632–76641, 2020, doi: 10.1109/ACCESS.2020.2989443.
- [6] W. Alajali, W. Zhou, S. Wen, and Y. Wang, "Intersection Traffic Prediction Using Decision Tree Models," *Symmetry*, vol. 10, no. 9, p. 386, Sep. 2018, doi: 10.3390/sym10090386.
- [7] W. Li, X. "Jeff" Ban, J. Zheng, H. X. Liu, C. Gong, and Y. Li, "Real-Time Movement-Based Traffic Volume Prediction at Signalized Intersections," *J. Transp. Eng. Part Syst.*, vol. 146, no. 8, p. 04020081, Aug. 2020, doi: 10.1061/JTEPBS.0000384.
- [8] M. S. Ghanim and K. Shaaban, "Estimating Turning Movements at Signalized Intersections Using Artificial Neural Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1828–1836, May 2019, doi: 10.1109/TITS.2018.2842147.
- [9] S. Lee, K. Xie, D. Ngoduy, and M. Keyvan-Ekbatani, "An advanced deep learning approach to real-time estimation of lane-based queue lengths at a signalized junction," *Transp. Res. Part C Emerg. Technol.*, vol. 109, pp. 117–136, Dec. 2019, doi: 10.1016/j.trc.2019.10.011.
- [10] R. Rahman and S. Hasan, "Real-time signal queue length prediction using long short-term memory neural network," *Neural Comput. Appl.*, vol. 33, no. 8, pp. 3311–3324, Apr. 2021, doi: 10.1007/s00521-020-05196-9.
- [11] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1644, no. 1, pp. 132–141, Jan. 1998, doi: 10.3141/1644-14.
- [12] S. Sun, C. Zhang, and G. Yu, "A Bayesian Network Approach to Traffic Flow Forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 124–132, Mar. 2006, doi: 10.1109/TITS.2006.869623.
- [13] B. Yu, H. Yin, and Z. Zhu, "Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Jul. 2018, pp. 3634–3640. doi: 10.24963/ijcai.2018/505.
- [14] R. Rahman *et al.*, "A Data-Driven Network Model for Traffic Volume Prediction at Signalized Intersections," *J. Big Data Anal. Transp.*, vol. 4, no. 2–3, pp. 135–152, Dec. 2022, doi: 10.1007/s42421-022-00059-2.
- [15] M. Xu, T. Z. Qiu, J. Fang, H. He, and H. Chen, "Signal-control refined dynamic traffic graph model for movement-based arterial network traffic volume prediction," *Expert Syst. Appl.*, vol. 228, p. 120393, Oct. 2023, doi: 10.1016/j.eswa.2023.120393.
- [16] Y. Jia, J. Wu, and M. Xu, "Traffic Flow Prediction with Rainfall Impact Using a Deep Learning Method," *J. Adv. Transp.*, vol. 2017, pp. 1–10, 2017, doi: 10.1155/2017/6575947.
- [17] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," arXiv, Feb. 22, 2017. Accessed: Apr. 15, 2024. [Online]. Available: <http://arxiv.org/abs/1609.02907>
- [18] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language Modeling with Gated Convolutional Networks," arXiv, Sep. 08, 2017. Accessed: Apr. 15, 2024. [Online]. Available: <http://arxiv.org/abs/1612.08083>
- [19] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," arXiv, Apr. 19, 2018. Accessed: Apr. 15, 2024. [Online]. Available: <http://arxiv.org/abs/1803.01271>
- [20] N. Mahmoud, M. Abdel-Aty, Q. Cai, and J. Yuan, "Predicting cycle-level traffic movements at signalized intersections using machine learning models," *Transp. Res. Part C Emerg. Technol.*, vol. 124, p. 102930, Mar. 2021, doi: 10.1016/j.trc.2020.102930.