GDCRN: Global Diffusion Convolutional Residual Network for Traffic Flow Prediction

Liujuan Chen[1], Kai Han[1(✉)], Qiao Yin[2], and Zongmai Cao[1]

[1] School of Computer Science and Technology/Suzhou Institute for Advanced Study, University of Science and Technology of China, Hefei, China
hankai@ustc.edu.cn

[2] School of Software Engineering, University of Science and Technology of China, Hefei, China

Abstract. Traffic flow prediction is a crucial issue for intelligent transportation system. Because of complicated topological structures of road networks and dynamic spatial-temporal patterns of traffic conditions, predicting flows on the road networks is still a challenging task. Most existing approaches focus on the local spatial-temporal correlations, ignoring the global spatial dependences and the global dynamic spatial-temporal correlations. In this paper, we propose a novel deep learning model for traffic flow prediction, called Global Diffusion Convolution Residual Network (GDCRN), which consists of multiple periodic branches with the same structure. Each branch applies global graph convolution layer to capture both local and global spatial dependencies, and further apply GRes to describe global spatial-temporal correlations simultaneously. Extensive experiments on two real-world datasets demonstrate that our model can capture both the global and local spatial-temporal dependencies dynamically. The experimental results show the effectiveness of our method.

Keywords: Traffic prediction · Spatial-temporal network · Graph convolution network

1 Introduction

Intelligent transportation system (ITS) plays an important role in improving efficiency of traffic management and ensuring traffic safety. Predicting traffic conditions is one of the most important tasks in ITS. It can guide traffic management and help drivers avoid congested roads, such that traffic jams can be avoided or alleviated. The traffic prediction on road networks is a typical spatial-temporal data prediction problem, which aims at predicting future traffic flows by making use of historical traffic data and road networks. The complexity of the traffic prediction problem is mainly affected by the following three factors:

K. Han is the corresponding author.

G. Li et al. (Eds.): KSEM 2020, LNAI 12275, pp. 438–449, 2020.
https://doi.org/10.1007/978-3-030-55393-7_39

1. Traffic conditions are different at various times (i.e. morning peak, noon) and also vary greatly from place to place, which shows strong dynamic characteristics in both spatial and temporal dimensions.
2. In the spatial dimension, complicated correlations are observed among different regions. Congestion can affect the traffic conditions of reachable neighbors with different impact. On the other hand, the traffic conditions of places with far distance may also affect each other. In summary, both long-range and short-range spatial relations between places are important in traffic flow prediction.
3. Temporal dependencies follow complicated periodic patterns. Due to people's regular daily life patterns traffic conditions may show some repeated patterns, such as peak flow in mornings. Moreover, just past traffic conditions inevitably have an influence on future traffic flows. Therefore, the temporal patterns of traffic data are not purely periodic.

In recent years, thanks to the development of sensor networks, ITS systems can obtain massive amounts of real traffic network data, which facilitates traffic prediction. Although deep learning methods have brought breakthroughs in traffic prediction, they still have some limitations. 1) The existing studies [11,14,17] assume that spatial dependency relationships only exist among directly connected or very close nodes. 2) RNN-based methods [15,17] are ineffective to learn long-term periodic dependencies, since they have problems such as gradient explosion/disappearances when capturing long sequences. And it is time consuming to train typical chain structured RNN. 3) Current studies [4] do not capture the global spatial-temporal dependencies in the same time. They also reduce the bidirectional traffic network to undirected graphs, which makes these type of methods less effective in practice.

In this paper, we propose a global diffusion convolution residual network (GDCRN) to predict traffic flows, which addresses the three shortcomings we have mentioned above. It contains multiple branches with the same structure for capturing information of different time periods, such as hourly period, daily period and weekly period. We propose global graph convolution (GGC) layer, which integrates a novel graph diffusion convolution unit based on three auxiliary matrices. It contains two local adjacency matrices to capture local spatial correlations of the bidirectional traffic network and a global matrix to capture global spatial dependencies. We further apply the attention mechanism to exploit the most important spatial and temporal dependencies. A global residual (GRes) unit is designed to capture global spatial-temporal information. In this way, our model is able to capture more complicated spatial-temporal correlations with better performance. Our contributions are summarized as follows:

– We propose a novel graph convolution layer which considers dynamicity, local and global spatial dependencies simultaneously. A novel GRes module proposed, which consists of a gated convolution to capture the temporal dependencies and a global residual unit to capture the global spatial-temporal correlations.

– We propose an effective and efficient deep learning framework GDCRN that contains multiple branches for capturing informative features of multiple different periods. Each branch is specially designed to capture spatial-temporal information of this time period.
– We evaluate our model on two real datasets and compare it with six baseline methods by three evalution metrics. Extensive experiments verify the effectiveness of our model.

2 Related Work

Accurate prediction of traffic conditions is essential to data-driven urban management. Researchers have made tremendous efforts in traffic prediction [7,8,13]. Statistical regression methods such as ARIMA and its variants [1,13] are representative models in the early studies on traffic prediction. However, they only study traffic time series for each individual location and fail to consider the spatial correlations. Later, some researchers feed spatial features and other external feature information into the traditional machine learning models [7,12]. But it is still difficult to consider the spatial-temporal correlations of high-dimensional traffic data. The prediction performance of traditional machine learning methods heavily depends on feature engineering.

Recently, deep learning methods have brought tremendous advances in traffic prediction, which outperform many traditional methods. Many models integrate convolution neural network (CNN) and recurrent neural network (RNN) to jointly model complex non-linear spatial and temporal dependences in traffic network, and have achieved inspiring success [8,14]. RNN and its variants [3,5] can effectively use the self-circulation mechanism to learn temporal dependence well. CNN treat city traffic network as images by dividing the traffic network into small grids and use CNN to model the non-linear spatial dependencies. However, the grid structure does not hold the real-word conditions, which makes it unsuitable to capture the spatial dependencies of traffic network effectively. The works in [10,17] propose to capture the structural correlations of traffic network by combing RNN and graph convolution network such as GCN [9] and DCNN [2]. GCN and DCNN models capture the dependence of graphs via operating convolution to pass message between the nodes of graphs. However, RNN-based models are difficult to train, computationally heavy and less effective when capturing long-distance contextual temporal information. To solve these challenges, STGCN [16] apply CNN in the time dimension and GCN in spatial dimension, which enable stable gradient and much faster training speed with fewer parameters. ASTGCN [4] further apply the attention mechanism to adjust spatial-temporal dependence dynamically. Although the schemes mentioned above have improved the accuracy of traffic prediction, they still fail to capture the global and local spatial-temporal dependencies simultaneously in the traffic network.

3 Preliminaries

Definition 1 (Traffic Network). *In this study, the traffic topological network can be defined as a weighted bidirectional graph $G = (V, E, A)$, where V is a set of nodes with limited number ($|V| = N$), E is a set of edges that describe the accessible routes between nodes, and $A \in \mathbb{R}^{N \times N}$ indicates the weighted adjacency matrix of G. Specifically, $a_{ij} \in A$ represents the weight from node v_i to v_j.*

Definition 2 (Traffic Data). *Assuming that the network G has N nodes and traffic data $\mathcal{X}$ contains C features (such as flow, occupy, speed), the traffic data of c-th ($c \in (1, \dots, C)$) feature on nodes v_i ($i \in (1, \dots, N)$) at time t can be described as $x_t^{i,c} \in \mathbb{R}$. Then, $X_t^i = (x_t^{i,1}, \dots, x_t^{i,C}) \in \mathbb{R}^C$ denotes the traffic data with all features on node v_i at time t, and $X_t = (X_t^1, \dots, X_t^N) \in \mathbb{R}^{N \times C}$ denotes the traffic data with all features and all nodes at time t. The whole historical traffic data can be denoted by $\mathcal{X} = (X_1, \dots, X_T) \in \mathbb{R}^{N \times C \times T}$.*

Problem 1 (Traffic Flow Prediction). Given a traffic Network G, and its historical signals over past T time slices, i.e. $\mathcal{X} = (X_1, \dots, X_T) \in \mathbb{R}^{N \times C \times T}$. Our problem is to predict the next T_p horizons traffic flow sequences Y on the whole traffic network. The prediction result Y can be defined as $Y = (Y_1, \dots, Y_i, \dots, Y_{T_p}) = (X_{T+1}, \dots, X_{T+j}, \dots, X_{T+T_p}) \in \mathbb{R}^{N \times C \times T_p}$, where $0 < j \leq T_p$, $Y_{T+j} = (Y_{T+j}^1, \dots, Y_{T+j}^N) \in \mathbb{R}^{N \times C}$.

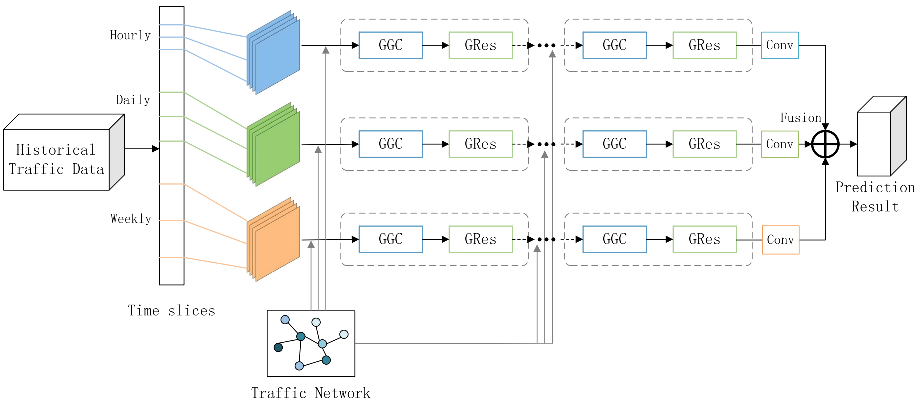


Fig. 1. The architecture of GDCRN.

4 Global Diffusion Convolutional Residual Network

Figure 1 shows the architecture of GDCRN. The inputs of GDCRN are historical traffic data and traffic network, and the outputs are the predictions of the future traffic states. We extract local and global spatial information from topology of traffic network. We set three branches to model the hourly-periodic, daily-periodic and weekly-periodic dependencies. Every branch is able to learn the dynamic spatial-temporal information in every time period by GGC and GRes submodules. A convolution layer is designed in the end to generate prediction results of each branch and keep output shapes consistent. Finally, the outputs of each periodic branch are fused to obtain final prediction results. The detailed mechanism of each module is described in the following subsections.

4.1 Global Graph Convolution Layer

For spatial dimension, the directly connected nodes inevitably affect each other, and those roads which are geographically distant but conveniently reachable are also correlated with each other. So, it is necessary to capture both local and global dependencies. In this paper, we propose a global graph convolution (GGC) unit based on diffusion convolution that simultaneously extract both the local and global spatial dependencies on traffic network. Diffusion convolution [2] is a compositional layers , which smoothes a node's signal by a diffusion process, so that it can directly describe features of bidirectional through multi-dimensional input.

Firstly, three auxiliary matrices are developed to encode spatial information about the topology of the traffic network . For details, we apply forward adjacency matrix A^F and the backward adjacency matrix $A^B = (A^F)^\top$ as local matrices to encode the local spatial proximity. And then we construct a global auxiliary matrix A^G to encode the topological correlations of long distance by a global transform.

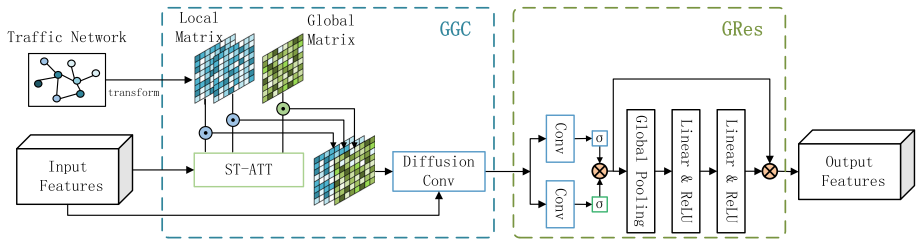


Fig. 2. The substructure of GDCRN module. ST-ATT: Spatial Temporal Attention [4]. Global Pool: Global average pooling layer. Linear & ReLU: Linear Transform and ReLU function

Secondly, traffic conditions of different locations have influence among each other, but not all of these correlations are equally important. Furthermore, correlations between different time horizons are also varying. Therefore, we adopt an attention mechanism (ST-ATT) [4] to generate a spatial-temporal attention matrix β, which can focus on more important spatial-temporal information. Take the temporal attention as an example: $\alpha = \Phi_t \cdot \sigma(((X_{l-1})^T U_1)U_2(U_3 X_{l-1}) + b_t)$, where $X_{l-1} \in \mathbb{R}^{C_{l-1} \times N \times T_{l-1}}$ is the input of the l-th GGC module, σ represents activation function sigmod, $\Phi_t, b_t \in R^{T_{l-1} \times T_{L-1}}, U_1 \in \mathbb{R}^N, U_2 \in \mathbb{R}^{C_{l-1}}, U_3 \in \mathbb{R}^{C_{l-1}}$ are learnable parameters. Each element $\alpha_{i,j}$ represents the strength of the correlation between time i and time j. We apply the normalized temporal attention matrix $\alpha^{'}$ generate dynamic importance-oriented temporal representations $H_t = X_{l-1}\alpha^{'}$. Then, spatial-temporal attention matrix β is generated by the similar attention mechanism based on temporal representations H_t.

Thirdly, we feed the spatial-temporal attention matrix β into following graph convolution layer to adjust correlations between nodes dynamically.

$$\hat{A^F} = (A^F \bigodot \beta); \hat{A^B} = (A^B \bigodot \beta); \hat{A^G} = (A^G \bigodot \beta); \tag{1}$$

where $\bigodot$ is a Hadamard product. By combining importance-oriented diffusion matrices, our innovative graph convolution layer can be formulated as:

$$H_s = \sigma(\sum_{k=0}^{K}(\hat{A}_k^F X_{l-1}\Theta_{k1} + \hat{A}_k^B X_{l-1}\Theta_{k2} + \hat{A}_k^G X_{l-1}\Theta_{k3})) \tag{2}$$

Where K is the diffusion step, $\Theta_{k1}, \Theta_{k2}, \Theta_{k2}$ is a diagonal matrix of learnable parameters, $\hat{A}_k^F$, $\hat{A}_k^B$ and $\hat{A}_k^G$ are the k-th step diffusion matrices, σ is the activation function of graph convolution layer, X_{l-1} is the input of the l-th GGC unit. By applying diffusion convolution operations with attention mechanisms to the input X_{l-1}, H_s can model dynamic local and global spatial dependencies.

4.2 Global Residual Network

The future traffic conditions have a complex non-linear relationship with the previous traffic conditions. To learn informative temporal correlations, we apply a temporal convolution with a gated mechanism. To capture the interdependencies between spatial and temporal dimensions, we propose a global residual unit which uses global information to selectively emphasise spatial-temporal correlations. As shown in the right part of Fig. 2, Global Residual Network (GRes) combines a gated temporal convolution unit with a global residual unit.

Gated Temporal Convolution Unit. Gated mechanisms have powerful ability to control information. We apply two standard convolution operations with different kernel sizes to learn different hidden representations in the time dimension. Then two different activation functions are applied as output gates to learn complex time features. Given the spatial representations H_s, we can formulate the gated temporal convolution unit as:

$$H_{st} = \sigma_1(Conv_1(H_s)) \odot (\sigma_2(Conv_2(H_s))) \tag{3}$$

where σ_1 and σ_2 are the different non-linear activations, σ_1 is RELU function and σ_2 is tangent hyperbolic function, $\odot$ is element-wise product, $Conv_1$ and

$Conv_2$ are the standard convolution functions. H_{st} can model both spatial and temporal dependencies.

Global Residual Unit. To improve the sensitivity of global spatial-temporal correlations in our model, we design a global residual unit to exploit informative features and suppress less useful ones. Firstly, a global average pooling layer is used to capture global contextual spatial-temporal dependencies directly among all nodes and all time horizons. To limit the model complexity and improve the generalization ability of the model, we use a linear transformation for decreasing dimension and a ReLU function, which can be defined as:

$$f(x) = ReLU(Wx) \tag{4}$$

where $W \in \mathbb{R}^{C_r \times C}$ is learning parameters, C is the input dimension, C_r is the output dimension and $C_r < C$. Different from SElayer [6], we use two same transformations in Fig. 2 instead of different ones, which has been proved by experiments the former performs better. Given local spatial-temporal representations H_{st}, the core of the global residual unit can be defined as:

$$H_o = f(f(GlobalPooling(H_{st}))) \otimes H_{st}, \tag{5}$$

where $\otimes$ is the element-wise product. H_o can further model global dynamic spatial-temporal dependencies. Then, a residual mechanism and LayerNorm are applied to improve generalization performance.

4.3 Fusion Mechanism

To ensure that multiple branches can be effectively merged, we apply a convolution layer at the end of each branch. The output prediction results of the three branches $(\hat{Y_h}, \hat{Y_d}, \hat{Y_w})$ have same shape. Finally, we fuse prediction results of multiple periods to capture global temporal correlations by learning weights and generate the final prediction result $\hat{Y}$, which can be formulated as:

$$\hat{Y} = W_h \odot \hat{Y_h} + W_d \odot \hat{Y_d} + W_w \odot \hat{Y_w}, \tag{6}$$

where $W_h, W_d, W_w \in \mathbb{R}^{N \times T_p}$ are the learning parameters.

5 Experiment Evaluation

5.1 Experiment Settings

DataSet. We verify GDCRN on two large real world highway traffic datasets, PeMSD4 and PeMSD8, released by [4]. PEMSD4 and PeMSD8 records two months of statistics on traffic data on the highways of California. Table 1 presents the details of the two datasets. The traffic data are aggregated every 5 min.

Table 1. The detailed information of the dataset.

Dataset	PeMSD8	PeMSD4
Locations	San Francisco Bay area, California	San Bernardino, California
Detectors	170	307
Time interval	12	12
Time span	01/01/2018-28/02/2018	07/01/2016-31/08/2016

Network Structure and Hyperparameter Settings. We implemented our model in Python with MXNet 1.6.0. We set up three different periodic branches for the model by week, day, and hour. We set the input period length of three branches as: $T_w = 2, T_d = 2, T_h = 1$. Each branch contains two GDCRN blocks. For graph convolution, we construct the global spatial matrix A_G by a random walk with path length $q = 3$ and set graph convolution layers with diffusion step $k = 3$. For gated temporal convolution unit, we set one with 64 filters and the kernel size 3×3, and another with 64 filters and the kernel size 1×1. In the first GDCRN block of branches, we set the strides of temporal convolution as the length of input period (i.e., 2, 2, 1). For the output convolution layer of each branch, we use 12 (prediction horizons) filters with kernel size 1×64. For training phase, the batch size is 16, learning rate is 0.001 and epochs are 50. We split dataset in chronological order with 70% for training, and 20% for testing, and the remaining data for validating.

5.2 Measurement and Baseline Methods

In our experiment, we use three most-widely adopted evaluation metrics, Mean Absolute Error (MAE), Root MeanSquare Error (RMSE), and Mean Absolute Percentage Error (MAPE) to measure our scheme and others. We compare GDCRN with following 6 baseline methods:

- HA: Historical Average model uses the average value of the last 12 time slices as the next prediction value.
- ARIMA [13]: Auto-Regressive Integrated Moving Average method is a widely used time series regression model.
- LSTM [5]: Long Short Term Memory network, which is a spacial RNN model.
- STGCN [16]: Spatio-Temporal Graph Convolutional Network applies purely convolutional structures with a gating mechanism to extract spatial-temporal features simultaneously.
- T-GCN [17]: Temporal Graph Convolutional Network combines with GCN and GRU.
- ASTGCN [4]: Attention based Spatial-Temporal Graph Convolution Network use spatial-temporal mechanism in graph convolution with Chebyshev polynomials approximation.

5.3 Experimental Results

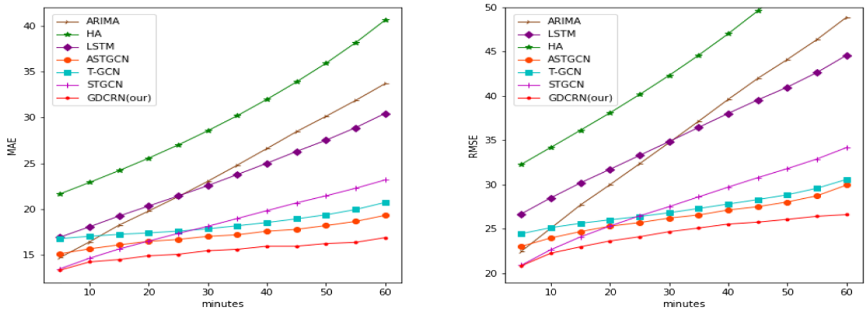
Performance Comparison. We compare the performance of our GDCRN and 6 baseline methods for 15-min, 30-min, 60-min predictions on PEMSD4 and PEMSD8 datasets. Table 2 shows the average results of traffic flow prediction performance on the three prediction intervals. Our GDCRN model obtains superior results on two datasets. It can be seen that GDCRN significantly outperforms the approaches that only take temporal features into account (HA, ARIMA, LSTM).

Table 2. The performance of our model and baselines on different predicting intervals

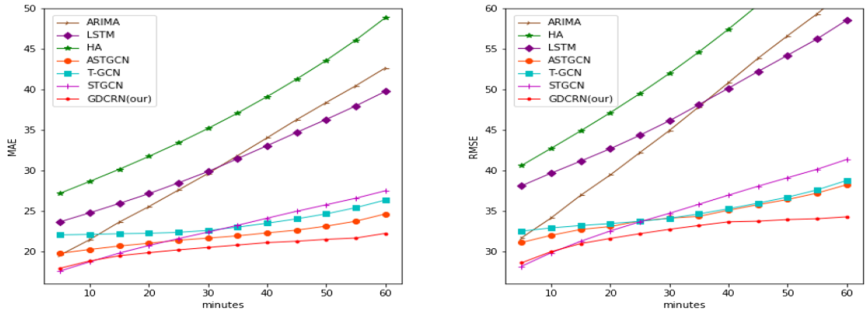
Dataset	Models	15 min			30 min			60 min		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
PEMSD8	HA	22.92	34.22	14.31	24.97	37.33	15.59	30.03	44.98	18.84
	ARIMA	16.46	25.15	10.06	18.93	29.00	11.54	24.08	35.85	14.83
	LSTM	18.09	28.48	11.86	19.76	30.9	13.85	23.37	36.03	15.5
	T-GCN	17.03	25.06	13.59	17.32	25.74	13.50	18.30	27.29	14.37
	STGCN	14.51	22.58	9.49	15.87	24.58	10.90	18.16	27.08	13.84
	ASTGCN	15.66	23.95	10.22	16.14	24.82	10.51	17.03	26.27	11.17
	GDCRN	**14.03**	**22.03**	**9.31**	**14.59**	**23.11**	**9.74**	**15.39**	**24.59**	**10.40**
PEMSD4	HA	28.64	42.78	19.45	31.02	46.29	21.17	26.83	55.02	25.66
	ARIMA	21.53	34.34	13.78	24.54	38.50	15.82	30.89	47.66	20.41
	LSTM	24.77	39.67	15.94	26.61	42.10	17.06	30.06	46.05	19.54
	T-GCN	22.09	32.90	18.20	22.23	33.34	17.95	23.34	34.97	18.94
	STGCN	18.85	30.00	13.09	20.49	32.19	13.90	23.07	36.92	16.75
	ASTGCN	19.82	31.98	14.33	20.78	32.92	14.8	21.91	34.75	15.81
	GDCRN	**18.65**	**29.91**	**12.98**	**19.37**	**31.13**	**13.46**	**20.40**	**32.81**	**14.25**

Compared to the spatial-temporal models, the prediction results of GDCRN excels RNN-based scheme T-GCN and also performs better than CNN-based schemes STGCN and ASTGCN. As for STGCN, GDCRN achieves bigger enhancement on the 60-min horizons than 15-min horizons. Since GDCRN introduces attention mechanism and GRes module to capture global spatial-temporal correlations, so that our model can better hold the long-term traffic pattern. ASTGCN utilizes GCN to describe spatial dependencies. However, GCN regards the traffic network as a unidirectional graph, which is not practical for real-world traffic network. In contrast, GDCRN adopts global diffusion convolution, which is able to handle bidirectional network and can capture global spatial correlation directly. Therefore, combining with the ability to obtain local and global spatial-temporal correlations on the bidirectional network, GDCRN is able to perform better regarding all the metrics for all predicting horizons.

Figure 3 illustrates the changes of prediction performance of our model and other baseline models as the prediction temporal interval increases. We have two valuable observations which further confirm the superiority of our model.



(a) The prediction results on PeMSD8



(b) The prediction results on PeMSD4

Fig. 3. Performance changes of different methods as the predicting horizon increases

Firstly, the growth trends of prediction error of GDCRN are smaller than almost all methods, indicating that our model is insensitive to prediction time interval. Secondly, GDCRN achieves the best forecasting performance in all time dimensions, especially for the long-term prediction. Specifically, the differences between DGCRN and other baseline methods are more significant as the prediction time interval increases, which shows that the scheme of our GDCRN model has advantages not only in short-term predictions, but also in long-term predictions. All the experiment results above demonstrate the advantages of our model in capture spatial-temporal correlation of the highway traffic data.

Ablation Study. In order to verify the effectiveness of every components on our model, we compare the following four variants of our model.

- ChebNet, which replaces diffusion convolution with ChebNet.
- No-GRU, which removes global residual unit in GRes module.
- No-A_G, which removes global spatial matrix in diffusion convolution unit.
- No-Gate, which removes Gate mechanism in temporal convolution unit.

Table 3. Performance of variants of GDCRN on different predicting intervals

Dataset	Models	15 min			30 min			60 min		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
PEMSD8	chebNet	14.83	23.01	10.14	15.65	24.37	10.70	16.83	26.27	11.60
	No-GRU	14.87	23.21	10.28	15.86	25.12	10.68	17.20	27.54	11.68
	No-A_G	14.18	22.21	9.55	14.72	23.26	10.00	15.51	24.71	10.68
	No-Gate	14.22	22.22	9.43	14.79	23.35	9.81	15.65	24.94	10.45
	GDCRN	**14.03**	**22.03**	**9.31**	**14.59**	**23.11**	**9.74**	**15.39**	**24.59**	**10.40**
PEMSD4	chebNet	19.08	30.37	14.01	19.86	31.60	14.82	20.96	33.35	15.88
	No-GRU	19.42	30.73	14.97	20.22	31.98	15.89	21.32	33.73	17.04
	No-A_G	18.73	30.08	13.37	19.52	31.35	14.23	20.53	33.03	15.14
	No-Gate	18.91	30.08	13.82	19.62	31.30	14.27	20.55	32.94	14.82
	GDCRN	**18.65**	**29.91**	**12.98**	**19.37**	**31.13**	**13.46**	**20.40**	**32.81**	**14.25**

Table 3 compares the average performance of every variant over different prediction interval. We can find that GDCRN achieves the best prediction performance. The predicting results of GDCRN excels the ChebNet model, which verifies that capturing bidirectional spatial dependencies is very necessary and useful for prediction tasks on real traffic networks. Compared with the No-GRU model, GDCRN has better prediction precision and is insensitive to prediction interval, which proves that capturing global spatial-temporal features are important for traffic prediction. The GDCRN are superior to No-A_G model, indicating the effectiveness of capture global spatial correlations. In summary, the GDCRN can achieve the best results regardless of the prediction horizon, and each component of our model make sense.

6 Conclusion

In this paper, we propose a novel global diffusion convolution residual network for traffic prediction. Based on the spatial topological structure of the traffic network, we propose a novel graph convolution layer, which leverages global and local information of spatial structure. To exploit informative features, we design a global residual network GRes and combine it with GGC module to capture both global and local spatial-temporal correlations. Experiments on two large-scale real-world datasets verify the effectiveness of our model. Furthermore, GDCRN is a generalized spatial-temporal network prediction framework, and has the potential to be applied to other similar prediction problems such as taxi demand forecasting.

Acknowledgement. This work is partially supported by The National Key R&D Program of China under Grant 2018AAA0101200, National Natural Science Foundation of China (NSFC) under Grant No. 61772491, No. U170921, Anhui Initiative in Quantum Information Technologies AHY150300 and the Fundamental Research Funds for the Central Universities.

References

1. Ahmed, M.S., Cook, A.R.: Analysis of freeway traffic time-series data by usingBox-Jenkins techniques. No. 722 (1979)
2. Atwood, J., Towsley, D.: Diffusion-convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1993–2001 (2016)
3. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
4. Guo, S., Lin, Y., Feng, N., Song, C., Wan, H.: Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proceedings of the AAAI Conference on Artificial Intelligence. **33**, pp. 922–929 (2019)
5. Hochreiter, S.: Schmidhuber: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
6. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
7. Jeong, Y.S., Byon, Y.J., Castro-Neto, M.M., Easa, S.M.: Supervised weighting-online learning algorithm for short-term traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems **14**(4), 1700–1707 (2013)
8. Ke, J., Zheng, H., Yang, H., Chen, X.M.: Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transportation Research Part C: Emerging Technologies **85**, 591–608 (2017)
9. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
10. Li, Y., Yu, R., Shahabi, C., Liu, Y.: Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. arXiv preprint arXiv:1707.01926 (2017)
11. Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., Wang, Y.: Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. Sensors **17**(4), 818 (2017)
12. Van Lint, J., Van Hinsbergen, C.: Short-term traffic and travel time prediction models. Artificial Intelligence Applications to Critical Transportation Issues **22**(1), 22–41 (2012)
13. Williams, B.M., Hoel, L.A.: Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. Journal of transportation engineering **129**(6), 664–672 (2003)
14. Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems. pp. 802–810 (2015)
15. Yao, H., Tang, X., Wei, H., Zheng, G., Li, Z.: Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In: AAAI Conference on Artificial Intelligence (2019)
16. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 (2017)
17. Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., Li, H.:T-gcn: A temporal graph convolutional network for traffic prediction. IEEETransactions on Intelligent Transportation Systems (2019)