sustainability

Article

Attention-Based Multiple Graph Convolutional Recurrent Network for Traffic Forecasting

Lu Liu *, Yibo Cao and Yuhan Dong *

Shenzhen International Graduate School, Tsinghua University, Shenzhen 518000, China
* Correspondence: l-liu20@mails.tsinghua.edu.cn (L.L.); dongyuhan@sz.tsinghua.edu.cn (Y.D.)

Abstract: Traffic forecasting is essential in the development of intelligent transportation systems, as it enables the formulation of effective traffic dispatching strategies and contributes to the reduction of traffic congestion. The abundance of research focused on modeling complex spatiotemporal correlations for accurate traffic prediction, however many of these prior works perform feature extraction based solely on prior graph structures, thereby overlooking the latent graph connectivity inherent in the data and degrading a decline in prediction accuracy. In this study, we present a novel Attention-based Multiple Graph Convolutional Recurrent Network (AMGCRN) to capture dynamic and latent spatiotemporal correlations in traffic data. The proposed model comprises two spatial feature extraction modules. Firstly, a dot product attention mechanism is utilized to construct an adaptive graph to extract the similarity of road structure. Secondly, the graph attention mechanism is leveraged to enhance the extraction of local traffic flow features. The outputs of these two spatial feature extraction modules are integrated through a gating mechanism and fed into a Gated Recurrent Unit (GRU) to make spatiotemporal interaction predictions. Experimental results on two real-world traffic datasets demonstrate the superiority of the proposed AMGCRN over state-of-the-art baselines. The results suggest that the proposed model is effective in capturing complex spatiotemporal correlations and achieving about 1% improvements in traffic forecasting.

Keywords: traffic forecasting; attention mechanism; multiple graphs

check for updates

Citation: Liu, L.; Cao, Y.; Dong, Y. Attention-Based Multiple Graph Convolutional Recurrent Network for Traffic Forecasting. *Sustainability* **2023**, *15*, 4697. https://doi.org/10.3390/su15064697

Academic Editor: Elżbieta Macioszek

Received: 29 January 2023
Revised: 13 February 2023
Accepted: 20 February 2023
Published: 7 March 2023

1. Introduction

The rapid pace of urbanization has resulted in increasing levels of traffic congestion, highlighting the need for innovative solutions in transportation systems. Intelligent transportation systems (ITS) have emerged as a promising solution to address various issues including traffic congestion in the transportation industry. A critical component of ITS deployment is traffic forecasting, which provides a crucial foundation for the formulation of dispatching strategies and control measures. Accurate traffic forecasting not only affects the safe and efficient travel of passengers but also plays a critical role in urban traffic planning. Despite its importance, traffic forecasting remains a challenge due to the dynamic nature of traffic patterns and the complexity of spatiotemporal correlations. These complex factors make it difficult to accurately forecast traffic flow, which is still an ongoing challenge for ITS [1].

Time correlation refers to the observation that similar flow patterns are likely to occur at a given location in consecutive time steps [2]. However, relying solely on time correlation for traffic forecasting is problematic, as unexpected events such as traffic accidents or temporary road closures can disrupt traffic patterns and add noise to the prediction process [3]. Moreover, the ability of time series forecasting approaches is limited to model long-term sequences, making it difficult to extract meaningful long-term time correlations. In contrast, spatial correlation is more prevalent in road networks due to the circulation of vehicles. At a macroscopic level, areas with similar functions are likely to exhibit similar traffic patterns, regardless of whether the nodes are connected, indicating the similarity of road structures.

From a microscopic perspective, there is a higher likelihood of traffic flow interactions between connected or adjacent nodes. Spatial correlation is more complex and implicit than temporal correlation [4,5]. Figure 1 provides a visual representation of a real road network and the traffic flow at three road sections, highlighting the complex spatiotemporal relationship in traffic flow. In the figure, for three randomly selected nodes, the horizontal axis displays the traffic flow data collected by the sensor, aggregated every five seconds, while the vertical axis represents the corresponding traffic flow.

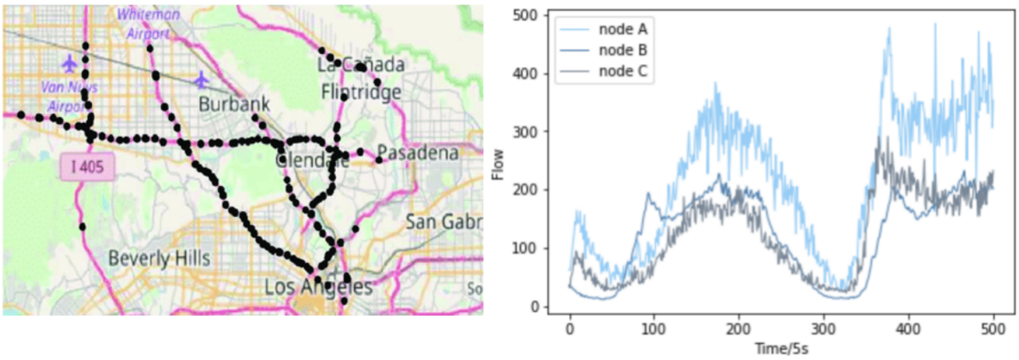

Figure 1. Part of road nodes visualization of the PeMSD4 dataset collected in the San Francisco Bay Area and traffic flow information for three of the nodes.

To process time correlation, various methods have been developed, such as Auto-Regressive Integrated Moving Average (ARIMA [6]), Vector Auto-Regression (VAR) [7], and Kalman Filters [8]. These methods primarily rely on manual feature engineering and have relatively simple model structures, making it challenging to accurately extract the implicit temporal characteristics of traffic flow. With the advancements of deep learning in various engineering fields [9], researchers have started to explore its application in transportation. Deep learning enables end-to-end learning of feature representations from raw traffic data, overcoming the limitations of manual feature engineering. Some models based on Recurrent Neural Networks (RNN), such as Long-Short Term Memory (LSTM) [10] and Gated Recurrent Units (GRU) [2], have been used to extract long-term time correlation from time series data. However, these methods are not designed to learn spatial correlations. To capture spatial features, Ma et al. [11] represented traffic flow as a grid structure based on real road network information, which can be easily processed by convolutional neural networks. They applied convolution kernels to slide across the grid for feature extraction.

Due to the irregular structure of traffic which are non-Euclidean, it is crucial to consider comprehensive spatiotemporal correlations for effective traffic forecasting. In recent years, there has been an increasing trend in the application of Graph Neural Networks (GNNs) for this purpose. Previous works such as [12,13] utilized pre-defined graphs to learn spatial features. However, these graphs are static and primarily based on prior information regarding road network distance, thus lacking the ability to reflect dynamic spatial correlations. On the other hand, [14–17] constructed multi-graph convolutional networks such as similarity graphs, correlation graphs, and topological graphs, thereby allowing for the extraction of spatial features at different scales. Furthermore, [18,19] applied attention mechanisms to temporal and spatial dimensions to adaptively capture the spatiotemporal correlations from the traffic data.

Despite these advances, existing approaches fail to take into account the simultaneous extraction of both global and local features. To address this issue, we propose a novel Attention-based Multiple Graph Convolutional Recurrent Network (AMGCRN) for traffic forecasting. The architecture of AMGCRN consists of two main components: a dynamic multi-graph convolution module based on graph self-attention mechanisms and a time series prediction module based on Gated Recurrent Units (GRU). The former module is capable of learning both global and local spatial features, while the latter one increases the receptive field in the time dimension. These two modules are integrated through an embedded approach, effectively fusing spatial and temporal features to improve the accuracy of traffic flow forecasting.

The contributions of this work can be succinctly stated as follows:

- A novel AMGCRN model is proposed to capture the dynamic spatiotemporal correlations of the traffic flow by simultaneously learning global and local spatial features.
- The AMGCRN models both global and local spatial features through the use of attention mechanisms on node embeddings and input features, respectively.
- The proposed method can adaptively assign different learning weights to the neighbors of nodes, without requiring prior knowledge of the road structure, thereby improving the accuracy of traffic flow forecasting.

The remainder of this paper is organized as follows: Section 2 presents an overview of the related works in this field. In Section 3, the preliminaries are discussed in detail. The proposed method is described and explained in Section 4. The experimental setup and results are presented and analyzed in Section 5. Section 6 presents the results of the ablation studies performed. Finally, the conclusions and future directions for research are discussed in Section 7.

2. Related Works

2.1. Traditional Methods

Traffic forecasting plays a crucial role in mitigating the challenges posed by urban traffic congestion. In previous studies, researchers have predominantly relied on statistical methods such as VAR [7] and ARIMA [6] for traffic flow prediction. However, these methods are limited by certain basic assumptions and may not be capable of capturing the complex spatiotemporal relationships inherent in traffic flow data. With the advent of deep learning, there has been a growing trend in the application of deep learning models in traffic prediction tasks. Researchers have leveraged LSTM, GRU, and other RNN based models to capture long-term temporal dependencies in traffic flow data [2,10,20,21]. In addition, Temporal Convolutional Network (TCN) has been proposed to model the time dependence of longer sequences while reducing memory consumption during training [22].

Despite these advancements, existing methods have limitations in extracting spatial features. In an effort to address this issue, Yao et al. [23] integrated Convolutional Neural Network (CNN) and LSTM to extract both temporal and spatial features. However, the process of transforming the data into a grid structure for network input can result in information loss and errors in prediction.

2.2. Graph Based Methods

In recent years, graph structure-based deep learning has garnered significant attention in various fields [24]. The graph structure is often utilized to process data with irregular structures, such as molecular data, social networks, point cloud data, etc. [25]. In particular, Graph Convolutional Networks (GCNs) has emerged as a leading method for spatial feature extraction in data with non-Euclidean structures, such as road networks. GCNs can be broadly categorized into two methods: spatial domain GCNs and spectral domain GCNs. Spatial GCNs refer to convolution operations in the spatial domain and involve aggregation of local spatial information from neighboring nodes. On the other hand, spectral GCNs define convolution operations in the spectral domain, based on the properties of the graph Laplacian matrix.

In the field of traffic prediction, dynamic graph convolution has been employed to model spatiotemporal correlations [26]. The Attention-based Spatio-Temporal Graph Convolutional Network (ASTGCN) [13] is proposed to apply temporal and spatial attention mechanisms to capture dynamic spatiotemporal correlations in traffic flow. The Adaptive Graph Convolutional Recurrent Network (AGCRN) [27] utilized a learning graph structure with node embedding to extract spatial features adaptively. The Regularized Spatial-Temporal Graph Learning (RSGL) [28] integrated the road topology graph and learned an adaptive graph to extract both explicit and implicit representations of spatial features. The

Gating Mechanism Attention Network (GMAN) [19] utilized temporal and spatial attention to predict in a spatiotemporal parallel manner and then fused the prediction results using a gating mechanism.

However, these existing methods fall short in their ability to capture both global spatiotemporal correlations and dynamic local spatiotemporal correlations simultaneously. To overcome these limitations, we aim to propose GCN based method, which is capable of modeling both global and local spatiotemporal relations to achieve more accurate traffic flow forecasting.

3. Preliminaries

3.1. Traffic Network

In this study, we define the road network's non-Euclidean structure as an undirected graph $G = (V, E, A)$, where $V = |N|$ represents the combination of nodes on the road, with each node representing an observation point for collecting traffic flow information. The set of edges E represents the connections between nodes, indicating the relationships between road sections. The adjacency matrix $A \in R^{N \times N}$ quantifies the proximity between nodes as follows,

$$A(i,j) = \begin{cases} 1 & \text{if } (v_i, v_j) \in E \\ 0 & \text{else} \end{cases}, \text{for } i,j = 1, \ldots, N \tag{1}$$

3.2. Traffic Flow Forecasting

We define traffic flow forecasting as a sequence-to-sequence mapping problem. For the traffic network G with N nodes, the characteristics of the traffic flow sequence at time T are denoted as $X_T \in \{X_1^T, X_2^T, \ldots, X_N^T\}$. Specifically, $X_i^T \in R^F$ indicates that each node has F-dimensional characteristics at time T. The historical observation sequence χ of τ time slices as input is defined as $\chi = \{X_N^{T-\tau+1}, X_N^{T-\tau+2}, \ldots, X_N^T\} \in R^{\tau \times N \times F}$. The purpose of traffic flow forecasting is to find a mapping function $\mathcal{F}$ to fit the spatiotemporal correlations between historical observation input and future traffic flow data. The traffic flow data of the next P time slices are $\mathcal{Y} = \{Y_N^{T+1}, Y_N^{T+2}, \ldots, Y_N^{T+P}\}$. As a result, the traffic flow forecasting problem is defined as:

$$\{X_N^{T-\tau+1}, X_N^{T-\tau+2}, \ldots, X_N^T\} \quad \xrightarrow{\mathcal{F}} \quad \{Y_N^{T+1}, Y_N^{T+2}, \ldots, Y_N^{T+P}\} \tag{2}$$

4. Proposed Method

In this paper, we propose a new traffic flow forecasting model named AMGCRN, which has an encoder-decoder structure. The overall structures of the AMGCRN and AMGCN are shown in Figure 2. The AMGCN module in the AMGCRN model consists of multiple graph convolution layers that capture both global and local spatial information. It first extracts the global spatial features of the entire road network using the global graph convolution operation and then captures the local spatial features of each node in the road network by using the local graph convolution operation. The global graph convolution operation uses the adjacency matrix of the entire graph as the convolution kernel to perform convolution on all nodes in the graph. In contrast, the local graph convolution operation uses the sub-adjacency matrix centered on each node as the convolution kernel to perform convolution on the neighborhood of each node. By integrating the dynamic global and local spatial information, the AMGCN module can effectively capture the complex spatiotemporal correlations in traffic flow data. Finally, the decoder part of the AMGCRN model uses the extracted traffic flow features to make predictions. It uses a series of 1D convolution layers with increasing kernel size to increase the receptive field of the predicted data and further capture the correlations between the prediction and historical observation data.

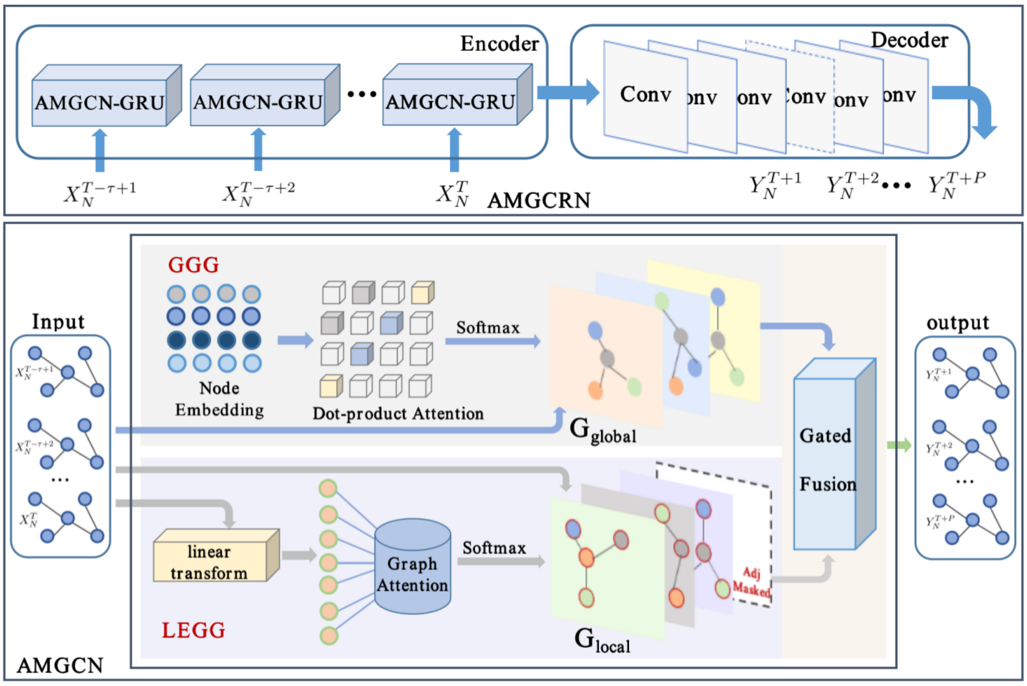

Figure 2. The upper part is the overall framework of the AMGCRN and the lower part is the architecture of AMGCN. AMGCN contains two spatial feature extraction modules, they are GGG and LEGG.

4.1. Global Graph Generation

Previous GCN-based traffic forecasting approaches have utilized pre-defined graphs to establish the correlation or similarity between road nodes. Some studies, such as Li et al. [5], have employed geographical topology to calculate the distance between road nodes in order to construct the graph matrix. However, this approach may introduce noise and worsen the prediction accuracy. Other works, such as [27,29], have measured the similarity or correlation of node series characteristics to determine node proximity. Despite these efforts, these methods may not fully capture the complex implicit spatial correlations and use static pre-defined graphs which are unable to represent dynamic spatial correlations.

To address these challenges, we propose a novel approach that utilizes learnable node embedding parameters, $E \in \mathbb{R}^{N \times d}$, to represent the network structure of each node, instead of solely considering its traffic characteristics. The dot product attention mechanism is then applied to the node embedding parameter vectors to capture the similarity of node network structures in the embedding space, described as:

$$A_{global} = \text{softmax}(\text{ReLU}(EE^T)) \tag{3}$$

where ReLU is an activation function that enhances the model's nonlinearity, to better extract features from sparse data. The softmax function is used to normalize the A_{global}. The global graph convolution operation can be formulated as:

$$X_{global} = (I_N + A_{global})XW + W_b \tag{4}$$

where X is the input feature matrix, W and W_b are learnable parameters.

4.2. Local Enhancement Graph Generation

Due to temporary traffic controls and weather conditions, traffic flow exhibits local implicit spatial correlations, which are complex and difficult to capture. To address this challenge, we introduce the Local Enhancement Graph Generation (LEGG) module, to enhance local spatial feature extraction. The LEGG module utilizes Graph Attention

Networks (GAT) [30] to model the similarity of road node flow characteristics. The input to the LEGG module is the traffic flow, $X_T \in X_1^T, X_2^T, \ldots, X_N^T$, at time T, where N is the number of road nodes and $X_i^T \in \mathbb{R}^F$ for $i \in N$. A learnable linear transformation, W, is applied to the input feature to obtain a higher-level implicit representation, with a dimension of F'. The attention score between two road nodes, i and j, is then defined as:

$$\alpha_{ij} = \frac{\exp(\mathrm{LeakyReLU}(\vec{a}[W\vec{X_i^T} \parallel W\vec{X_j^T}]))}{\sum_{k \in N_i} \exp(\mathrm{LeakyReLU}(\vec{a}[W\vec{X_i^T} \parallel W\vec{X_k^T}]))} \tag{5}$$

where a is a single-layer feedforward neural network with parameter dimension of $2F'$. LeakyReLU activation function is applied to increase nonlinearity and $\parallel$ represents the concatenation operation. Each node can aggregate all normalized first-order neighbors' information including its own. To make the learning process more stable, we use the softmax function to normalize the attention scores. In addition, we propose a masking mechanism to introduce road topology prior information for computing attention scores. The adjacency matrix A obtained from the topological distance of different nodes (sensors) is used as the mask judgment condition and the weight of the two connected nodes is assigned as α_{ij}:

$$\alpha'_{ij} = \begin{cases} \alpha_{ij} & \text{if } A > 0 \\ 0 & \text{else} \end{cases}, \text{for } i, j = 1, \ldots, N \tag{6}$$

after obtaining the attention scores that aggregate the flow features, the output feature of each node is expressed as:

$$X_{local}^i = \sigma\left(\sum_{j \in N_i} \alpha'_{ij} W\vec{X_j^T}\right) \tag{7}$$

where σ is the elu activation.

4.3. Graph Fusion Mechanism

The adaptive gating network is designed to learn the relationship between the outputs of the two spatial feature extraction modules and determine the importance of each module for the final prediction. The gating weight W_g is calculated as a function of the output from the two spatial feature extraction modules, represented as X_{local} and X_{global}, respectively. The gating weight is then used to calculate the weighted sum of the two features, producing the final fused output. Formally, the gating weight W_g can be expressed as:

$$W_g = \sigma[W_{g2}(\mathrm{ReLU}(W_{g1}X^o + b_1)) + b_2] \tag{8}$$

where W_{g1} and W_{g2} are learnable weight matrices and σ is the sigmoid activation function. The final fused output can then be represented as:

$$X_{spa} = W_{gG} \odot X_{global} + W_{gL} \odot X_{local} \tag{9}$$

where $\odot$ represents element-wise multiplication. This graph fusion mechanism allows the network to weigh the importance of each spatial feature extraction module for the final prediction and effectively captures both local and global traffic flow patterns to improve traffic prediction.

4.4. Spatio-Temporal Fusion Mechanism

The GRU architecture allows the model to capture long-term dependencies in time, while still preserving the most recent information. Additionally, the use of the AMGCN as the gating mechanism provides a unique fusion of temporal and spatial information, which can lead to better results compared to other models that only rely on traditional GRU structures or use MLP layers for gating. The final output of the spatiotemporal fusion mechanism is the hidden state h_t, which is used to make predictions for future traffic flow. Specifically:

$$\begin{aligned} z_t &= \sigma(FG([X_{0:t}, h_{t-1}])) \\ r_t &= \sigma(FG([X_{0:t}, h_{t-1}])) \\ \hat{h_t} &= \tanh(FG([X_{0:t}, r_t \odot h_{t-1}])) \\ h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \hat{h_t} \end{aligned} \tag{10}$$

where FG refers to the two graph generation and graph fusion operations mentioned above. z_t and r_t are the update gate and reset gate in GRU, respectively, which are used to filter past information. $\hat{h_t}$ is the candidate hidden state calculated according to Equation (9) and the h_0 tensor is initialized to 0.

5. Experiments

5.1. Datasets

To evaluate the performance of the proposed model, we conduct experiments on two real-world datasets, PeMSD4 [13] and PeMSD8 [17] consisting of 307 sensors and 170 sensors, respectively. The detailed information of these datasets is presented in Table 1. To test the performance of the model, we randomly split the nodes of each class into three parts, with a 6:2:2 ratio, for training, validation, and testing, respectively. We focus on short-term traffic forecasting, e.g., one hour prediction, in which we predict the traffic of the next hour given the historical traffic data of the past hour.

Table 1. Details of PeMSD4 and PeMSD8.

Datasets	Nodes	Edges	Time Steps	Time Range
PeMSD4	307	340	16992	01/01/2018–02/28/2018
PeMSD8	170	295	17856	07/01/2016–08/31/2016

5.2. Baselines

To evaluate the performance of the proposed method, we compare the experimental results with several traditional methods and graph-based models, which are as follows:

- HA: Historical Average is based on integrating and averaging historical information to predict future information.
- VAR [7]: Vector Auto-Regression predicts interconnected time series by capturing hidden relationships in time series.
- SVR [31]: Support Vector Regression is a supervised learning model with a related learning algorithm, which is used to analyze data used for classification and regression analysis.
- LSTM [32]: Long Short Term Memory networks are a special recurrent neural network, which solves the problem of long-term dependence in time series prediction by introducing forgetting gates.
- TCN [22]: Temporal Convolutional Neural Network achieves the effect of capturing long-term dependent information through causal convolution.
- DCRNN [5]: Diffusion Convolutional Recurrent Neural Network is a sequence-to-sequence structure that models traffic flow as a diffusion process on a directed graph, capable of capturing both spatial and temporal correlations.
- STGCN [4]: Spatio-Temporal Graph Convolutional Network combines graph convolution and gated time convolution, which can extract the most useful spatial features and continuously capture the most basic time features.
- ASTGCN [13]: Attention-based Spatio-Temporal Graph Convolutional Network designs a spatial attention mechanism and a temporal attention mechanism to simultaneously extract spatiotemporal correlation.

- STSGCN [17]: Spatial-Temporal Synchronous Graph Convolutional Network uses multiple local spatiotemporal feature extraction modules to capture the heterogeneity of long-term spatiotemporal network data.
- AGCRN [27]: Adaptive Graph Convolutional Recurrent Network proposes a method for adaptively constructing spatial correlation and adopting spatiotemporal embedding method for traffic prediction.

5.3. Metrics

We adopt multiple performance metrics to jointly evaluate the model performance, including Mean Absolute Errors (MAE), Mean Absolute Percentage Errors (MAPE), and Root Mean Square Errors (RMSE). The calculation of each evaluation metric is as follows:

$$\begin{aligned} MAE &= \frac{1}{N}\sum_{i=1}^{N}|Y_{Ti} - Y_{Pi}| \\ MAPE &= \frac{100\%}{N}\sum_{i=1}^{N}|\frac{Y_{Ti} - Y_{Pi}}{Y_{Ti}}| \\ RMSE &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{Ti} - Y_{Pi})^2} \end{aligned} \tag{11}$$

where Y_P is the predicted value and Y_T is the ground truth and N is the number of nodes. The smaller the error value, the better the model performance. These metrics provide a comprehensive evaluation of the model's performance, taking into account the magnitude and percentage of prediction errors, as well as the square error of prediction values.

5.4. Experiment Settings

- Data Preprocessing: The collected traffic data from the sensors is processed by being windowed and aggregated into 5-min intervals. For the purpose of predicting the traffic flow of the next hour based on the input data of the previous hour, the dimensions of each input and label data instance are $N \times 12 \times 1$, where N represents the number of nodes.
- Loss Function: In order to train the model, we adopt L1 loss function and apply Adam optimizer to optimize its convergence. The batch size is set as 32 for PeMSD4 dataset and 64 for PeMSD8 dataset, respectively. The learning rate is set to 0.003. The model is trained for 200 epochs on PeMSD4 and 300 epochs on PeMSD8.
- Hardware Support: The proposed model is implemented on a combination of one GeForce RTX 2080 Ti GPU and one Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz.

5.5. Results

We evaluate the proposed model against ten baseline models in terms of MAE, RMSE, and MAPE on both PeMSD4 and PeMSD8 datasets. Table 2 summarizes the average results of traffic flow forecasting performance in the next hour separated as 12 time steps. Numerical results show that: (1) The GCN-based methods exhibit better performance than traditional methods (HA, VAR, and SVR) and deep learning methods (LSTM and TCN), demonstrating the importance of spatial feature extraction for traffic flow prediction tasks and the suitability of GCN for this task. (2) The designed multiple graph convolution module in the proposed model can effectively extract and integrate spatial features from multiple attributes, leading to better modeling of spatial correlation and spatiotemporal dependencies.

The proposed model demonstrates significantly improvement performance on the PeMSD4 dataset and a slight improvement in MAE on PeMSD8 dataset, compared to other models. To further illustrate the results, we display the forecasting performance of five GCN-based baselines at each horizon on both PeMSD4 and PeMSD8 datasets in Figure 3.

Table 2. Average performance comparison of different models on PeMSD4 and PeMSD8. Our model significantly improved prediction accuracy on PeMSD4 and we also achieved a slight improvement in MAE on PeMSD8. Ours metrics are highlighted in the table.

Model	Dataset	PeMSD4			PeMSD8		
	Metrics	MAE	RMSE	MAPE	MAE	RMSE	MAPE
HA		38.03	59.24	27.88%	34.86	52.04	24.07%
VAR		24.54	38.61	17.24%	19.19	29.81	13.10%
SVR		28.70	44.56	19.20%	23.25	36.16	14.64%
LSTM		26.77	40.65	18.23%	23.09	35.17	14.99%
TCN		23.22	37.26	15.59%	22.72	35.79	14.03%
DCRNN		21.22	33.44	14.17%	16.82	26.36	10.92%
STGCN		21.16	34.89	13.83%	17.50	27.09	11.29%
ASTGCN		22.93	35.22	16.56%	18.25	28.06	11.64%
STSGCN		21.19	33.65	13.90%	17.13	26.86	10.96%
AGCRN		19.83	32.26	12.97%	15.95	25.22	10.09%
AMGCRN (ours)		**19.52**	**31.76**	**12.90%**	**15.85**	**25.32**	**10.18%**

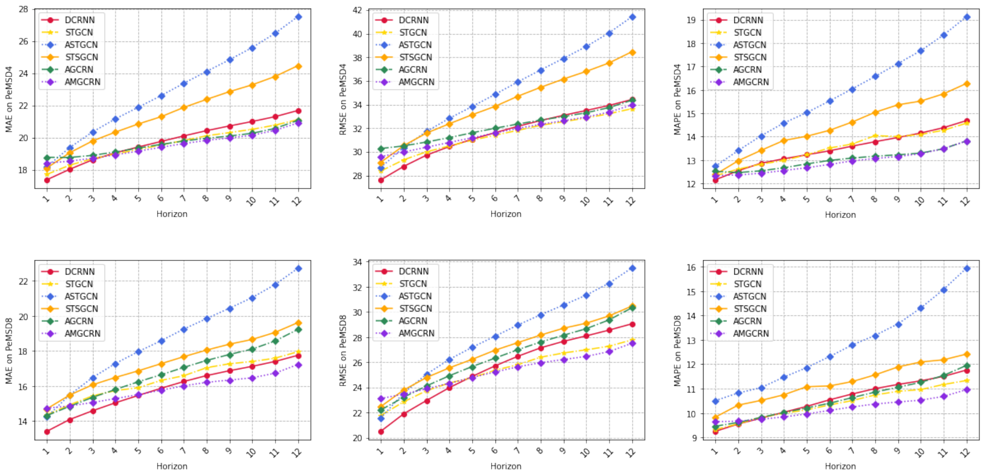

Figure 3. Forecasting performance at each horizon comparing with other baselines. AMGCRN achieved the slowest rise and the most stable prediction performance.

6. Ablation Study

In order to assess the contribution of each component in the proposed model, we conduct an ablation study on both the PeMSD4 and PeMSD8 datasets. The ablation study removes two components of the AMGCRN model: (1) w/Global: The GGG module is removed from AMGCRN. (2) w/Local: The LEGG module is removed from AMGCRN. Table 3 summarizes the average results of the ablation study, and the forecasting performance at each horizon is further displayed in Figure 4.

The results of ablation study further demonstrate the validity of the model design, highlighting the importance of both the GGG and LEGG modules in improving the prediction performance. The proposed AMGCRN model offers a new solution for effectively modeling complex spatiotemporal correlations in traffic flow forecasting.

Table 3. Average performance comparison of ablation study on PeMSD4 and PeMSD8. The proposed GGG and LEGG can jointly improve the accuracy of traffic flow forecasting.

Model	**Dataset**		**PeMSD4**			**PeMSD8**	
	Metrics	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
w/Global		20.20	32.63	13.30%	18.55	29.57	12.60%
w/Local		19.83	32.26	12.97%	15.95	25.22	10.09%
AMGCRN (ours)		**19.52**	**31.76**	**12.90%**	**15.85**	**25.32**	**10.18%**

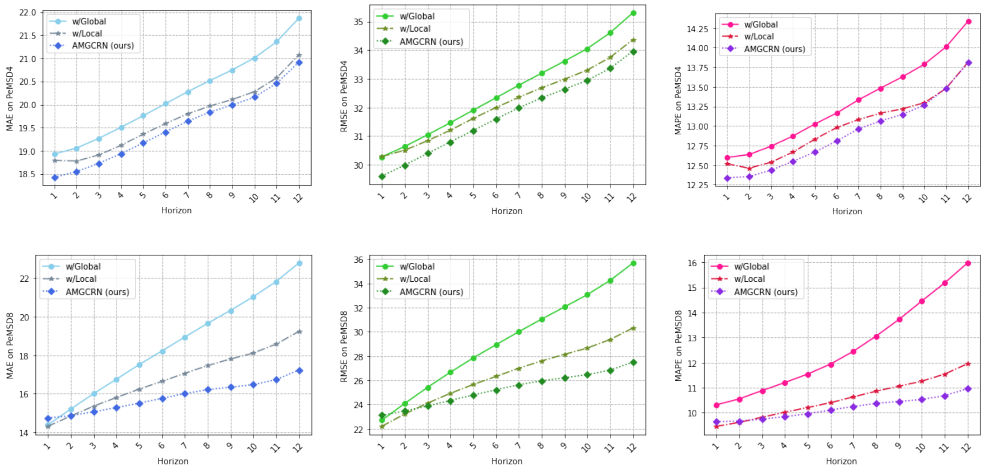

Figure 4. Forecasting performance of the ablation study at each horizon. The proposed GGG and LEGG can significantly reduce the prediction error at each horizon.

7. Conclusions

In this work, we proposed a novel and innovative model of AMGCRN for modeling complex spatiotemporal correlations in traffic flow forecasting. The AMGCRN model effectively captures the global road network structure similarity and the local features of traffic flow through the integration of two distinct spatial feature extraction modules of GGG and LEGG, which are combined through an adaptive gating mechanism and embedded into a GRU-based time series prediction network. Empirical results from extensive experiments on two real-world datasets demonstrate the superiority of the proposed AMGCRN model compared to other state-of-the-art baselines in terms of performance. These results highlight the effectiveness and feasibility of the proposed approach in modeling complex spatiotemporal correlations in traffic flow forecasting.

In future work, we aim to explore incorporating external factors such as weather and POIs into the AMGCRN model to further enhance its accuracy in traffic prediction tasks. Furthermore, the proposed AMGCRN model has the potential to be adapted and applied to other traffic prediction problems, such as Origin-Destination (OD) prediction, thereby demonstrating its versatility and wide-ranging impact.

Author Contributions: Methodology, L.L.; data curation, Y.C.; formal analysis, Y.D.; writing—original draft preparation, L.L.; writing—review and editing, Y.D. and Y.C.; supervision, Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lv, Z.; Xu, J.; Zheng, K.; Yin, H.; Zhao, P.; Zhou, X. Lc-rnn: A deep learning model for traffic speed prediction. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018.
2. Fu, R.; Zhang, Z.; Li, L. Using LSTM and GRU neural network methods for traffic flow prediction. In Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 11–13 November 2016; pp. 324–328.
3. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
4. Yu, B.; Yin, H.; Zhu, Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv* **2017**, arXiv:1709.04875.
5. Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv* **2017**, arXiv:1707.01926.
6. Williams, B.M.; Hoel, L.A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *J. Transp. Eng.* **2003**, *129*, 664–672. [CrossRef]
7. Zivot, E.; Wang, J. Vector autoregressive models for multivariate time series. In *Modeling Financial Time Series with S-PLUS®*; Springer: New York, NY, USA, 2006; pp. 385–429.
8. Lippi, M.; Bertini, M.; Frasconi, P. Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 871–882. [CrossRef]
9. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
10. Chen, Y.y.; Lv, Y.; Li, Z.; Wang, F.Y. Long short-term memory model for traffic congestion prediction with online open data. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; pp. 132–137.
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* **2017**, *17*, 818. [CrossRef]
12. Pan, Z.; Liang, Y.; Wang, W.; Yu, Y.; Zheng, Y.; Zhang, J. Urban traffic prediction from spatiotemporal data using deep meta learning. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 1720–1730.
13. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 922–929.
14. Liu, C.; Xiao, Z.; Wang, D.; Wang, L.; Jiang, H.; Chen, H.; Yu, J. Exploiting Spatiotemporal Correlations of Arrive-Stay-Leave Behaviors for Private Car Flow Prediction. *IEEE Trans. Netw. Sci. Eng.* **2021**, *9*, 834–847. [CrossRef]
15. Li, M.; Zhu, Z. Spatial-temporal fusion graph neural networks for traffic flow forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 4189–4196. [CrossRef]
16. Chen, W.; Chen, L.; Xie, Y.; Cao, W.; Gao, Y.; Feng, X. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 3529–3536. [CrossRef]
17. Song, C.; Lin, Y.; Guo, S.; Wan, H. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 914–921. [CrossRef]
18. Guo, S.; Lin, Y.; Wan, H.; Li, X.; Cong, G. Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting. *IEEE Trans. Knowl. Data Eng.* **2021**, 34, 5415–5428. [CrossRef]
19. Zheng, C.; Fan, X.; Wang, C.; Qi, J. Gman: A graph multi-attention network for traffic prediction. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 1234–1241. [CrossRef]
20. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
21. Cui, Z.; Ke, R.; Pu, Z.; Wang, Y. Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transp. Res. Part C Emerg. Technol.* **2020**, *118*, 102674. [CrossRef]
22. Lea, C.; Flynn, M.D.; Vidal, R.; Reiter, A.; Hager, G.D. Temporal convolutional networks for action segmentation and detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 156–165.
23. Yao, H.; Wu, F.; Ke, J.; Tang, X.; Jia, Y.; Lu, S.; Gong, P.; Ye, J.; Li, Z. Deep multi-view spatial-temporal network for taxi demand prediction. *Proc. AAAI Conf. Artif. Intell.* **2018**, *32*. [CrossRef]
24. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and locally connected networks on graphs. *arXiv* **2013**, arXiv:1312.6203.
25. Wang, X.; Bo, D.; Shi, C.; Fan, S.; Ye, Y.; Philip, S.Y. A survey on heterogeneous graph embedding: Methods, techniques, applications and sources. *IEEE Trans. Big Data* **2022**, 1. [CrossRef]

26. Han, L.; Du, B.; Sun, L.; Fu, Y.; Lv, Y.; Xiong, H. Dynamic and multi-faceted spatiotemporal deep learning for traffic speed forecasting. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Online, 14–18 August 2021; pp. 547–555.
27. Bai, L.; Yao, L.; Li, C.; Wang, X.; Wang, C. Adaptive graph convolutional recurrent network for traffic forecasting. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17804–17815.
28. Yu, H.; Li, T.; Yu, W.; Li, J.; Huang, Y.; Wang, L.; Liu, A. Regularized graph structure learning with semantic knowledge for multi-variates time-series forecasting. *arXiv* **2022**, arXiv:2210.06126.
29. Liu, L.; Chen, J.; Wu, H.; Zhen, J.; Li, G.; Lin, L. Physical-virtual collaboration modeling for intra-and inter-station metro ridership prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, 23, 3377–3391. [CrossRef]
30. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
31. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* **1996**, *9*.
32. Graves, A. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin, Heidelberg. 2012; pp. 37–45.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.