

# Deep Learning Approaches for Traffic Forecasting: Spatiotemporal Heterogeneity and Model Advances

Shuo Zhu\*

Sichuan University, Chengdu, Sichuan, China

**Abstract.** With the increasing complexity of urban traffic systems, accurate traffic flow forecasting has become a critical component of intelligent transportation systems. Traditional statistical and machine learning models, such as Autoregressive Integrated Moving Average (ARIMA) and random forests, struggle to address the non-stationarity and heterogeneity of traffic data. In recent years, deep learning has significantly advanced spatiotemporal traffic forecasting. This paper provides a comprehensive review of deep learning-based models, including Diffusion Convolutional Recurrent Neural Network (DCRNN), Graph WaveNet, Spatio-Temporal Decoupling Masked Autoencoder (STD-MAE), and Spatio-Temporal Pivot Graph Neural Network (STPGNN), analyzing their structural designs and performance improvements. By examining the theoretical foundations of temporal and spatial heterogeneity, this study highlights how deep models effectively capture multi-scale dependencies, dynamic topologies, and pivotal node behaviors. The paper also discusses challenges such as data noise, computational cost, and real-world deployment. Future research directions include transferable pre-training, frequency-domain modeling, and lightweight architectures to achieve efficient and interpretable traffic forecasting.

## 1 Introduction

Traffic forecasting is an essential component of modern urban transportation systems. It supports a wide range of applications such as route planning, congestion control, and resource scheduling, thereby contributing to the development of intelligent transportation systems. With the continuous expansion of urban areas, traffic flow patterns have become increasingly complex, exhibiting strong nonlinear and uncertain characteristics. Consequently, conventional statistical and shallow learning models often fail to provide reliable predictions. For example, autoregressive integrated moving average (ARIMA) models can represent short-term temporal variations but struggle to accommodate non-stationary fluctuations caused by special events, weather changes, or traffic incidents. Likewise, classical machine learning techniques such as Support Vector Machines (SVM)

\* Corresponding author: 2022141461135@stu.scu.edu.cn

and Random Forests (RF) show some improvement in prediction accuracy but are still limited in capturing intricate spatial correlations between traffic sensors.

The rapid progress of deep learning has greatly reshaped this research field. Recurrent neural networks (RNNs) are effective for sequential modeling, convolutional neural networks (CNNs) extract meaningful spatial representations, and graph neural networks (GNNs) provide a natural way to handle non-Euclidean spatial structures inherent in road networks. Many representative approaches have achieved remarkable results. For instance, Li et al. introduced DCRNN, which formulates traffic flow as a diffusion process on directed graphs and achieves around 17% MAE reduction in 1-hour predictions on the METR-LA dataset, alleviating error propagation in long-term forecasts [1]. Similarly, Wu et al. proposed Graph WaveNet, incorporating dilated causal convolution and an adaptive adjacency matrix, yielding over 20% improvement in 1-hour forecasting accuracy on the PEMS-BAY dataset compared to STGCN, and providing more stable long-term temporal modelling [2]. These advances confirm that deep learning methods possess clear advantages over traditional models for traffic prediction.

Despite these achievements, practical traffic systems are highly dynamic—spatial structures, temporal patterns, and human mobility behaviors evolve continuously. This dynamic nature makes spatiotemporal modeling particularly challenging. Moreover, differences among sensor locations introduce substantial heterogeneity, where neighboring sensors may display distinct traffic dynamics. Additional complications, such as noisy readings, missing data, and distribution shifts between cities, further restrict model generalization. To mitigate these issues, recent works have explored dynamic graph construction, multi-scale spatiotemporal dependencies, and context-aware modeling that integrates external knowledge such as meteorological data, points of interest (POIs), and human activity indicators. In parallel, pre-training and self-supervised learning have emerged as promising strategies to utilize unlabeled data, improving both data efficiency and transferability across regions. A persistent and fundamental challenge lies in spatiotemporal heterogeneity. Temporally, traffic exhibits non-stationary and multi-scale trends, while spatially, dependencies are shaped by node-level functionality and regional context. To address these aspects, researchers have proposed diverse strategies, including enhanced spatiotemporal representation learning, the incorporation of auxiliary contextual signals, and the use of robust training paradigms such as data augmentation and pre-training.

Building upon this foundation, this paper provides a systematic overview of traffic forecasting research, covering theoretical underpinnings, representative models, ongoing challenges, and prospective directions. The goal is to outline the evolution of this field and offer insights that may guide future studies toward more adaptive and generalizable traffic prediction frameworks.

## 2 Theoretical basis

The core challenge in traffic forecasting lies in spatio-temporal heterogeneity. On the temporal dimension, traffic flow exhibits both regular periodic patterns—such as weekdays versus weekends and morning/evening rush hours—and non-stationary behavior influenced by unexpected events and external conditions. On the spatial dimension, nodes demonstrate entirely distinct flow patterns due to functional and locational differences. Highway corridors, residential streets, and transportation hubs exhibit distinct behaviors, making it difficult for unified modeling approaches to account for all scenarios. To address this, researchers have proposed various improvement strategies focusing on both temporal consistency and spatial heterogeneity.

## 2.1 Temporal heterogeneity

Temporal heterogeneity reflects the dynamic differences of transportation systems across different timescales and cycles, such as changes in traffic patterns during morning and evening peaks, holidays, or special events. This difference makes the traffic time series show significant non-stationary and multi-scale characteristics, so the model needs to be able to adaptively identify and characterize the dependency structure and change law in different time periods. Existing research on temporal heterogeneity focuses on two main areas: multi-scale modeling and temporal perception mechanisms. The former, by introducing multi-scale convolutions, pyramid attention, or decomposition modules, divides the original time series signal into different frequency or periodic components to model short-term fluctuations and long-term trends separately, thereby achieving dynamic adaptation across time scales. For example, Yang divides the time series into multiple patches for separate modelling [3]. The latter focuses on the expression of temporal semantics and transfer learning capabilities. Temporal embedding, periodic encoding, or temporal gating mechanisms are often used to enable the model to adaptively adjust its internal state and parameter distribution when faced with inputs from different time periods, thereby improving the perception and prediction accuracy of non-stationary temporal patterns.

Overall, the theoretical basis for temporal heterogeneity modeling lies in achieving dynamic adaptability and generalization across different time domains through multi-scale decomposition and temporal semantic modulation mechanisms. Future research can further combine frequency-domain modeling with time-varying structure learning to achieve a unified representation of cross-periodic features and efficient modeling of pattern transfer between time periods.

## 2.2 Spatial heterogeneity

Spatial heterogeneity reflects the differences among various nodes in a transportation network in terms of geographical location, functional attributes, and traffic patterns. Such differences can lead to significant inconsistencies in traffic dynamics between adjacent regions. Therefore, the model needs to have the ability to adaptively perceive and adjust to node differences to achieve precise modeling of complex spatial relationships.

Regarding this issue, existing research mainly focuses on two aspects: adaptive structure learning and node importance modeling. The former dynamically builds adjacency relationships that can change over time by learning node embeddings or latent space representations, enabling the model to break through the limitations of fixed graph structures and more flexibly capture the latent correlations among heterogeneous nodes. For instance, Yang construct different latent spaces to capture different spatial dependencies [3]. The core of such methods lies in establishing data-driven graph generation functions through parameterization or attention mechanisms, allowing spatial dependency relationships to self-adjust based on node features.

The latter approach focuses on functional differences among nodes within transportation networks, proposing weighted modeling strategies based on node importance or critical node identification. By assigning differentiated weights to nodes during graph convolution or aggregation processes, the model highlights the role of hub nodes that are more sensitive to global traffic changes, thereby enhancing the overall stability and interpretability of predictions.

Overall, the theoretical foundation of spatial heterogeneity modeling lies in shifting the expression of spatial dependencies from static topology to data-driven dynamic relationship learning through adaptive graph structures and node-differentiated weighting mechanisms.

This approach not only enhances the model's adaptability to complex spatial topologies but also provides crucial insights for subsequent multi-level traffic pattern modeling and regional distribution pattern analysis.

### **3 Case studies**

#### **3.1 Case Study 1: Diffusion Convolutional Recurrent Neural Network (DCRNN) for traffic flow prediction**

Urban traffic forecasting relies on accurately modeling long-term temporal patterns. Li et al. proposed the DCRNN model, which models traffic flow propagation as a diffusion process on a directed graph [1]. The researchers used the METR-LA and PEMS-BAY datasets, which include flow and speed data from hundreds of traffic sensors, and constructed an adjacency matrix based on road network topology. In their implementation, the DCRNN incorporates diffuse convolutions into a gated recurrent unit (GRU), creating a diffuse convolutional GRU (DCGRU), which can simultaneously capture spatial diffusion and temporal dependencies. To mitigate error accumulation in multi-step forecasting, the model adopts a sequence-to-sequence architecture and introduces scheduled sampling. Experimental results show that in a one-hour forecast task on the METR-LA dataset, the DCRNN reduces the mean average efficiency (MAE) by approximately 17% compared to the LSTM model, effectively improving the stability and temporal consistency of long-term forecasts [4].

#### **3.2 Case Study 2: Graph WaveNet for adaptive spatio-temporal modeling**

Although traditional RNN-based methods can handle time series information, their efficiency is limited and they are difficult to parallelize. The Graph WaveNet proposed by Wu et al. utilizes expanded causal convolution instead of the cyclic structure, enabling the model to cover long historical sequences with a relatively shallow number of network layers and enhancing the nonlinear expression ability through the gating mechanism [2]. The researchers also designed an adaptive adjacency matrix learning module, which utilizes node embeddings to automatically discover potential spatial dependencies rather than relying entirely on static road networks. The model adopts residual join and skip join to ensure stable training and alleviate vanishing gradients. The experimental results show that in the 1-hour prediction of the PIMs-Bay dataset, Graph WaveNet reduces the MAE by more than 20% compared with STGCN, significantly improving the temporal consistency and generalization performance of long prediction tasks [5].

#### **3.3 Case Study 3: Spatial-Temporal-Decoupled Masked Pre-training (STD-MAE) for spatiotemporal forecasting**

In recent years, a key challenge in research has been how to better capture the spatiotemporal heterogeneity in traffic data. Gao et al. proposed the Spatial-Temporal-Decoupled Masked Pre-training (STD-MAE) framework, which systematically introduces self-supervised pretraining into traffic prediction for the first time [6]. The core idea of this method is to learn context-rich spatiotemporal representations through masked reconstruction and to model the temporal and spatial dimensions separately in a decoupled manner. Specifically, STD-MAE consists of a spatial autoencoder (S-MAE) and a temporal autoencoder (T-MAE), which force the model to learn long-range spatial and temporal dependencies by randomly masking sensor sequences or time slices,

respectively. This spatiotemporal-decoupled masking mechanism effectively captures cross-dimensional heterogeneous features in traffic data and prevents the "spatiotemporal illusion" caused by input length limitations in conventional end-to-end models. In downstream tasks, the spatial and temporal representations learned by STD-MAE can be seamlessly integrated into any predictor (e.g., DCRNN, Graph WaveNet, MTGNN) to provide long-range context enhancement. Experimental results show that, on the PEMS04 dataset, DCRNN assisted by STD-MAE reduced the MAE from 19.63 to 18.65, demonstrating its effectiveness in enhancing predictor performance.

### **3.4 Case Study 4: Spatio-Temporal Pivotal Graph Neural Network (STPGNN) for key node modeling**

In the study of spatial heterogeneity modeling, identifying and focusing on key nodes in traffic networks is a key approach to improving global prediction accuracy. To this end, Kong et al. proposed the STPGNN (Stage-Program-Optimized Neural Network) at AAAI [7]. This approach emphasizes that not all nodes in a traffic network contribute equally; rather, some key nodes have stronger aggregation and diversion effects in traffic distribution. The STPGNN first locates these nodes using a key node identification module (PIM). It then employs a key graph convolution module (PGCM) to perform more fine-grained modeling of the key node neighborhood. A parallel framework is then used to model the features of key nodes and ordinary nodes separately and then fuse them. This hierarchical modeling approach ensures the representation of global dependencies while enhancing the modeling accuracy of key locations. Experimental results show that on the PEMS03 dataset, the STPGNN achieves a MAE reduction of approximately 9% compared to the next best method, and an improvement of over 14% on the TaxiBJ dataset, demonstrating its outstanding advantages in spatial heterogeneity modeling.

### **3.5 Comparative analysis of different spatio-temporal models**

Different models differ significantly in their temporal and spatial modeling strategies. DCRNN ensures temporal consistency through diffuse convolutions and recurrent structures. Graph WaveNet improves the efficiency of long-term time series modeling by leveraging dilated convolutions and adaptive graph structures. STD-MAE leverages a self-supervised mask reconstruction mechanism to provide contextual enhancement in both spatial and temporal dimensions, and STPGNN further deepens spatial heterogeneity modeling at the key node level. These methods share a common approach: they all offer unique solutions to spatiotemporal heterogeneity and demonstrate significant improvements across various tasks and datasets.

### **3.6 Comparative analysis of different spatio-temporal models**

To systematically compare the structural differences and performance of current mainstream spatiotemporal prediction models, Table 1 summarizes the representative models mentioned in this article and their key features. From a methodological perspective, DCRNN and Graph WaveNet represent two typical frameworks: recurrent and convolutional. The former achieves joint spatiotemporal modeling through a diffusion process and sequence-to-sequence structure, while the latter relies on dilated convolutions and adaptive graph structures to enhance the ability to capture long-term dependencies. STD-MAE introduces a multi-scale spatiotemporal self-supervision mechanism to enhance generalization performance, while STPGNN models spatial heterogeneity through key node identification and differentiated convolution. Prediction results show that these models

significantly outperform earlier baseline methods across various datasets, validating the effectiveness of spatiotemporal structure optimization and adaptive learning strategies in traffic forecasting.

**Table 1.** Comparison of Representative Models

Model	Method Highlights	Temporal Dependency	Spatial Dependency	Representative Improvement
DCRNN	Diffusion convolution + Seq2Seq	GRU long-term memory + scheduled sampling	Based on diffusion graph convolution	On METR-LA, 1-hour prediction MAE ↓17%
Graph WaveNet	Graph convolution + adaptive adjacency	Captures temporal dependency via convolution	Dynamically learns inter-node correlations	On PEMS-BAY, MAE ↓20%+
STD-MAE	Spatiotemporal decomposition + masked autoencoder	Temporal encoding module (T-MAE)	Spatial encoding module (S-MAE)	On PEMS04, MAE improved by 5%
STPGNN	Graph structure perception + distribution discrepancy optimization	Captures temporal dependency via convolution	Captures spatial dependency through hierarchical graph convolution	On PEMS03, MAE improved by 9%

## 4 Challenges and limitations

### 4.1 Data issue

Traffic data is generally noisy and missing. During the sensor collection process, it may be affected by environmental interference, equipment aging or network transmission delay, resulting in deviation or even loss of observed values. For instance, abnormal fluctuations in traffic flow at intersections may result from sensor malfunctions rather than actual congestion, which can interfere with the model's learning of effective patterns. In addition, in some areas, the distribution of sensors is sparse, resulting in a lack of data and making it difficult to comprehensively reflect the real traffic dynamics. In this case, even if the model structure is advanced, its performance in scenarios with insufficient data is still greatly reduced. This problem is particularly evident in cross-city or small-sample prediction tasks. As Zhang et al. pointed out in the review, the noise and missing problems of traffic data are one of the core factors affecting the stability and generalization ability of the model [8]. Liu et al. also emphasized that the sparse distribution of sensors and the imbalance of regional data are the main reasons for the decline in cross-city prediction performance [9].

### 4.2 Model issues

One of the major challenges of deep learning models in traffic prediction is the excessive number of parameters and high computational cost. For instance, although complex

architectures based on convolution and attention mechanisms can capture more patterns, they often require expensive hardware support and long training times, which is not conducive to practical deployment. Another key issue is that the model has difficulty capturing complex spatiotemporal relationships. The dynamics and diversity of transportation networks make a single mechanism (such as static adjacency or a fixed time window) insufficient to model global and local dependencies. For instance, circular structures tend to overlook spatial interactions at long distances, while convolutional structures may lose fine-grained information in long-term time series modeling. These limitations lead to the model still having bottlenecks when dealing with non-stationary traffic patterns and cross-scenario generalization. Relevant review studies also pointed out that existing deep models still have structural limitations in capturing multi-scale spatiotemporal dependencies and improving generalization [10].

### 4.3 Application and Deployment Issues

At present, most traffic prediction models are still at the laboratory verification stage and there is still a gap from practical application. On the one hand, due to the lack of large-scale tests across cities and multiple scenarios, the generalization ability of the model has not been fully verified. On the other hand, traffic data involves user privacy and travel safety, and there are still obstacles to data sharing at the regulatory and ethical levels. Finally, the reasoning time of some models is relatively long, which is difficult to meet the real-time requirements of intelligent transportation systems and limits their direct application in dispatching and control

## 5 Conclusion

This paper systematically reviews the main research progress in the field of traffic prediction in recent years, with a focus on analyzing the theoretical basis and representative models of deep learning methods in dealing with spatio-temporal heterogeneity. Research shows that temporal heterogeneity stems from the dynamic changes of traffic flow in multi-scale and non-stationary environments, while spatial heterogeneity is manifested as functional differences and complex dependencies among different nodes. In response to these challenges, researchers have proposed a variety of innovative mechanisms, including cyclic structure based on diffusion process (DCRNN), adaptive Graph convolution and dilation convolution (Graph WaveNet), spatial-temporal decoupling self-supervised pre-training (STD-MAE), and critical node modeling mechanism (STPGNN). The robustness and prediction accuracy of the model have been significantly enhanced.

However, traffic prediction still faces problems such as data noise, model complexity and actual deployment. Future research can be further explored in the following directions: First, construct transferable spatio-temporal pre-trained models to enhance cross-city generalization; Secondly, integrate frequency-domain and dynamic graph structure learning to capture higher-level dependencies; Thirdly, develop efficient and interpretable lightweight models to facilitate their practical application. Overall, deep spatio-temporal prediction is moving from structural optimization to a new stage of knowledge fusion and general representation.

## References

1. Y. Li, R. Yu, C. Shahabi, Y. Liu, Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. Proc. Int. Conf. Learn. Represent. (ICLR) (2018)

2. Z. Wu, S. Pan, G. Long, J. Jiang, C. Zhang, Graph WaveNet for deep spatial-temporal graph modeling. arXiv preprint arXiv:1906.00121 (2019)
3. S. Yang, Q. Wu, Y. Wang, T. Lin, SSGCRTN: A space-specific graph convolutional recurrent transformer network for traffic prediction. *Applied Intelligence* **54**(22), 11978–11994 (2024)
4. S. Hochreiter, J. Schmidhuber, Long short-term memory. *Neural Computation* **9**(8), 1735–1780 (1997)
5. B. Yu, H. Yin, Z. Zhu, Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 (2017)
6. X. Rao, S. Shang, R. Jiang, L. Chen, P. Han, Traffic forecasting with patch-based graph convolutional recurrent network. *GeoInformatica*, 1–30 (2025)
7. H. Gao, R. Jiang, Z. Dong, J. Deng, Y. Ma, X. Song, Spatial-temporal-decoupled masked pre-training for spatiotemporal forecasting. arXiv preprint arXiv:2312.00516 (2023)
8. Y. Zhang, X. Kong, W. Zhou, J. Liu, Y. Fu, G. Shen, A comprehensive survey on traffic missing data imputation. *IEEE Trans. Intell. Transp. Syst.* (2024)
9. R. Liu, S. Y. Shin, A review of traffic flow prediction methods in intelligent transportation system construction. *Applied Sciences* **15**(7), 3866 (2025)
10. A. A. Kashyap, S. Raviraj, A. Devarakonda, S. R. Nayak K, S. KV, S. J. Bhat, Traffic flow prediction models – A review of deep learning techniques. *Cogent Engineering* **9**(1), 2010510 (2022)