Diffusion-Based Neural Networks for Traffic Forecasting: A Comprehensive Analysis

Bahne J. Thiel-Peters, Jacob Abb

August 30, 2024

1 Motivation

1.1 Why are traffic forecasts needed?

Traffic forecasting using neural networks is essential for optimizing traffic management, improving urban planning and supporting autonomous vehicles. Traffic management refers to the real-time coordination of traffic flow, including adjusting traffic lights, managing congestion and providing routing advice to drivers (Vlahogianni et al., 2014). By accurately predicting traffic patterns, authorities can proactively address potential issues, reduce the likelihood of traffic jams and improve the overall efficiency of the road network. Forecasting also supports long-term infrastructure planning by identifying where future traffic growth will occur, enabling targeted expansions and upgrades. It enhances public transportation by predicting demand and minimizing delays, while contributing to environmental sustainability by lowering emissions through more efficient traffic flow. Additionally, traffic forecasting improves safety by preventing accidents and supports emergency services with efficient routing. Overall, it plays a crucial role in creating efficient, safe and sustainable transportation systems that can adapt to the needs of growing urban populations (Laña et al., 2018).

1.2 Challenges

The goal of traffic forecasting is to predict future traffic speed across a sensor network based on historical traffic data and the structure of the underlying road networks.

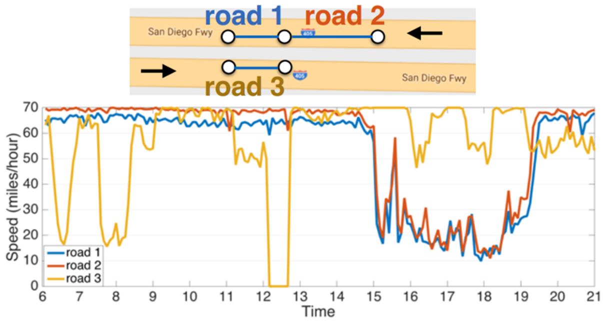

Figure 1: Sensor network example (**“textit –DCRNN”**)

In the following will be stated what challenges are associated to traffic forecasting.
Complex Spatial Dependency: Traffic data is influenced by a network of road sensors that may have complicated interrelationships. For example, two sensors located close to each other in terms of Euclidean distance, similar to *road 1* and *road 3* in Figure 1, may still capture very different traffic patterns due to the actual road network structure. This means that simple proximity does not necessarily correlate with similar traffic conditions, making it difficult to accurately model and predict traffic.
Non-linear Temporal Dependency: Traffic patterns during a normal weekday rush hour are highly predictable, but an unexpected accident or road closure can drastically alter the flow, which creates non-linear changes that are difficult to predict.
Difficulty in Long-Term Forecasting: Predicting traffic conditions over a longer horizon is inherently difficult due to the compounding effects of the aforementioned spatial and temporal dependencies. The fluctuations in traffic patterns make it challenging to maintain accuracy over longer periods (Laña et al., 2018).

2 Related work

2.1 Diffusion Convolutional Recurrent Neural Network (*DCRNN*)

The paper ”Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting” (Li et al., 2018) introduces a novel methodology and architecture that are specifically designed to tackle the challenges of traffic forecasting on road networks.

Methodology and Architecture
The core methodology of the paper revolves around the *Diffusion Convolutional Recurrent Neural Network (DCRNN)*. The *DCRNN* is designed to handle the spatiotemporal dependencies inherent in traffic data by splitting up the problem into spatial- and temporal dependencies and handle them separately. Traditional *convolutional neural networks (CNNs)* assume that data is structured in a (grid-like) euclidean space (Selvi et al., 2021). In general road networks are better represented as graphs where nodes represent traffic sensors and edges represent roads. The paper introduces *Diffusion Convolution* to operate directly on these graphs. This approach models the spatial dependencies by simulating the diffusion process on the graph, allowing the network to consider the direction and distance of traffic flow along the road network by updating the graphs adjacency matrix. To capture temporal dependencies the *DCRNN* integrates *Recurrent Neural Network (RNN)*, specifically a *sequence-to-sequence (Seq2Seq)* model with a *Gated Recurrent Unit (GRU)* (Chung et al., 2014) architecture. This enables the model to learn from past traffic conditions and make predictions about future traffic states. The mentioned challenges from chapter 1.2 are solved by the specific architecture of the *DCRNN*:

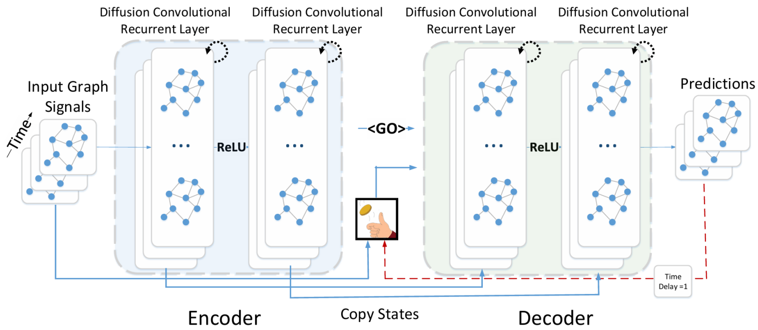

Figure 2: Architecture of *DCRNN* (Li et al., 2018)

The *DCRNN* uses an *Encoder-Decoder Structure* where the encoder processes the input sequence (historical traffic data) and converts it into a compact representation that captures both spatial and temporal patterns. The decoder then takes this encoded information and generates the output sequence which is a prediction of future traffic conditions over multiple time steps. The paper also makes use of a techique called *Scheduled Sampling* (Bengio et al., 2015) in training that gradually shifts from using true outputs to the model's own predictions during training, thereby reducing exposure bias and improving performance during inference which results into greater robustness of the model.

Evaluation
The *DCRNN* was evaluated on two large-scale, real-world traffic datasets: The *METR-LA* dataset contains traffic data gathered from 207 sensors installed on loop detectors across Los Angeles highways. *PEMS-BAY* covers traffic data from the Bay Area with a set of 325 sensors across the road network. Both datasets provide a rich source of data for modeling the spatiotemporal dynamics of traffic flow since they include speed measurements of every 5 minutes. The paper compares various methods for 15-minute, 30-minute, and 1-hour traffic forecasting on the two mentioned datasets, using metrics like *Mean Absolute Error (MAE)*, *Mean Absolute Percentage Error (MAPE)* and *Root Mean Squared Error (RMSE)*. The results show that RNN-based methods generally outperform other approaches, especially in long-term forecasting because the temporal dependencies become more non-linear over time, which also highlights the importance of modeling temporal dependencies. In addition to that *DCRNN* achieves the best performance across all metrics and forecasting horizons which demonstrates its effectiveness in capturing spatiotemporal dependencies.

2.2 Graph-Partitioning-Based *DCRNN*

"Graph-Partitioning-Based Diffusion Convolutional Recurrent Neural Network for Large-Scale Traffic Forecasting" (Mallick et al., 2020) introduces a new method, specifically designed for large highway networks. This approach leverages graph partitioning within the framework of *DCRNN* to improve forecasting accuracy.

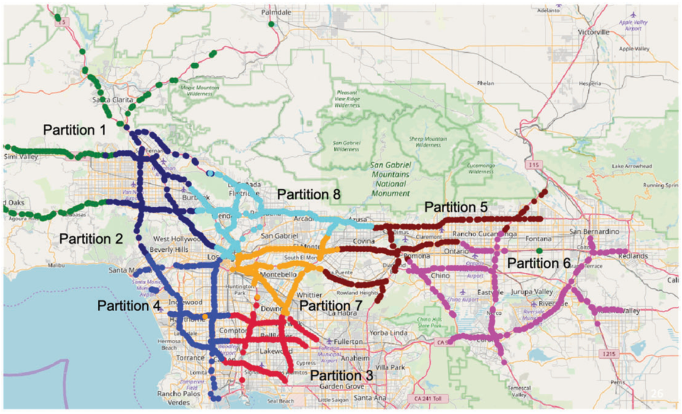

Figure 3: Example result of 8-way-partitioning (Mallick et al., 2020)

Methodology and Architecture

The proposed approach in the paper combines two main techniques: *Diffusion Convolutional Recurrent Neural Network (DCRNN)* and *Graph Partitioning*. The paper uses the same architecture as mentioned in chapter 2.1 since complex spatiotemporal data is given. The authors propose using a *Graph-Partitioning* method based on a spectral clustering-based approach to divide a large highway network into smaller subnetworks. This technique ensures that sensors that are spatially close and share similar traffic patterns are grouped together. Nodes that are included in multiple subgraphs are called *Overlapping Nodes*. These *Overlapping Nodes* help in maintaining the continuity and dependencies of the traffic patterns across different partitions. For example, if two subgraphs represent different sections of a highway, an *Overlapping Node* might be a sensor at an interchange connecting these sections. This sensor helps maintain the flow of information between the subgraphs and ensures that the traffic model accurately reflects how conditions in one section affect the other sections. Furthermore splitting up the large traffic network graph into smaller subgraphs and training them independently, makes the process more computationally feasible and significantly reduces computational overhead.

Evaluation

The approach was tested on a large California highway network with 11,160 sensor locations, which covers a large portion of the California highway network. The dataset contains information on traffic speed, flow and occupancy at each sensor location. The focus of this study was on predicting speed and flow. The traffic data was normalized to ensure that the model training was efficient and that the different features were on a similar scale. In Comparison to the other paper we observe differences as well as similarities. Unlike other methods (2.1) that typically predict either speed or flow, this model is designed to predict both features simultaneously. By handling *Multi-Output Forecasting* we achieve an improvement in the robustness of the traffic predictions.

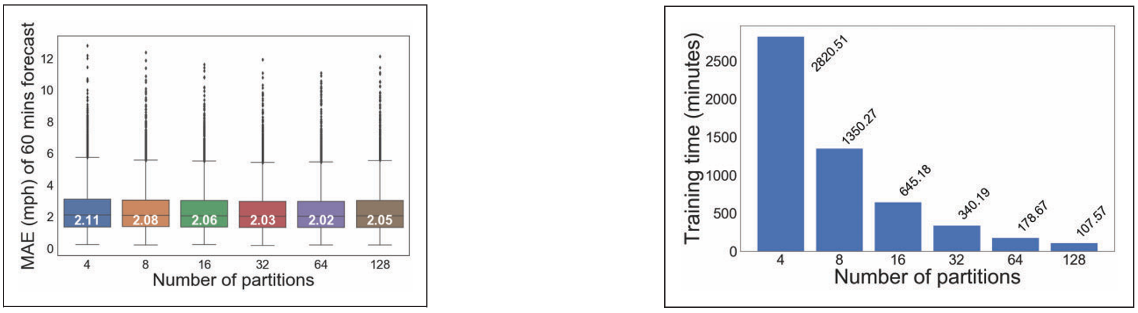

Figure 4: MAE and training time with different numbers of partitions (Mallick et al., 2020)

The California highway network is split into partitions ranging from 2 to 128. The results in figure 4 showed that increasing the number of partitions generally decreased the *MAE* due to more effective separation of sensor locations and easier model training for smaller partitions. Remarkably is that beyond 64 partitions the accuracy declined as spatially correlated nodes were split across partitions which shows a tipping point in partitioning efficiency. Training time significantly decreased as the number of partitions increased and achieved the best balance of accuracy and speed using 64

partitions. Error analysis revealed that factors such as sensor type, location and traffic dynamics influenced large errors, which highlights the complexity of accurately forecasting traffic patterns across diverse conditions. As a consequence 64 partitions were adopted as the optimal configuration for subsequent experiments, balancing computational efficiency with forecasting accuracy.
The paper compares the proposed method against several baseline models, including *Historical Average (HA)*, *Vector Autoregression (VAR)*, *Graph Convolutional Network (GCN)* and *Seq2Seq* models. Experiments conducted on a high-performance computing platform showed that the proposed method outperformed these baselines in most cases. The graph-partitioning-based *DCRNN* proved effective in capturing spatial and temporal dependencies which improves both forecast accuracy and computational efficiency. The architecture in combination with the overlapping nodes scaled well to large networks like California's, demonstrating significant improvements in traffic prediction and model manageability.

2.3 Adaptive Graph Neural Diffusion (AGND)

In the paper titled "Adaptive Graph Neural Diffusion for Traffic Demand Forecasting" (Wu et al., 2023), the authors propose a method called *Adaptive Graph Neural Diffusion (AGND)*, which integrates a graph neural diffusion process with attention mechanisms to better model spatial-temporal dynamics in traffic systems.

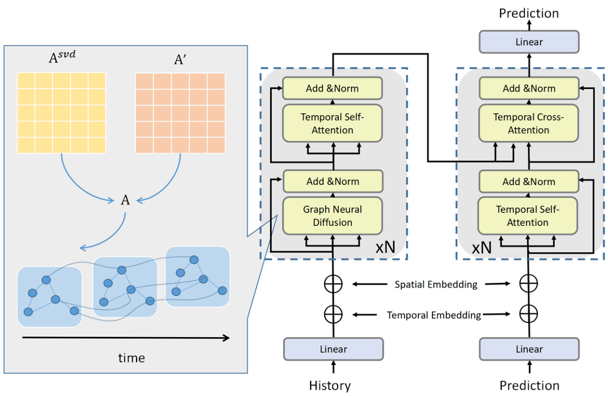

Figure 5: Architecture of *AGND* (Wu et al., 2023)

Methodology and Architecture
The proposed *AGND* framework in Figure 5 is designed around an encoder-decoder architecture to better handle complex long term dependencies. The core of the *AGND* lies in its application of graph neural diffusion to model complex spatial relationships as a diffusion process, similar to the method used in 2.1. Unlike most other approaches that rely on a static, heuristically defined adjacency matrix, *AGND* generates this matrix in a data-driven manner using *Singular Value Decomposition (SVD)*. This approach captures global spatial relations by transforming historical traffic data into a low-dimensional space, which is then used to construct the adjacency matrix. The *Attention Mechanisms* in *AGND* are crucial for enhancing the model's ability to focus on relevant spatial and temporal features. The *Spatial Attention Mechanism* operates by computing attention weights across different locations in the traffic network at each time step. Instead of applying a uniform focus across all nodes, the *Spatial Attention* allows the model to emphasize certain nodes based on their relevance to the current traffic conditions. This helps in dynamically adjusting the importance of different spatial relations, leading to more accurate modeling of traffic flow. The *Temporal Attention Mechanism* addresses the challenge of capturing long-term dependencies in time-series data. The models weights of different time steps differs and depends on their relevance to the prediction task. This mechanism allows *AGND* to effectively consider the entire history of traffic data which results in reducing information loss over long sequences and improving the accuracy of the forecasts. By combining both *Spatial* and *Temporal Attention* the model gains a nuanced understanding of how traffic patterns evolve over time and space which is critical for accurate forecasting.

Evaluation
The *AGND* model was evaluated on two real-world datasets, *NYC Citi Bike* and *NYC Taxi*.

Method		SVR	DCRNN	STG2Seq	TGCN	ASTGCN	STID	GWNET	CCRNN	STGODE	AGND
NYC Citi Bike	MAE	2.3218	1.8954	2.1969	1.9817	1.8961	2.1160	1.7221	1.7404	1.7328	**1.6788**
	RMSE	3.9240	3.2094	3.8733	3.4256	3.4009	3.3322	2.8997	2.8382	2.8423	**2.8115**
	MAPE	0.6452	0.6179	0.6395	0.6046	0.6112	**0.5215**	0.5986	0.5836	0.5893	0.5818
NYC Taxi	MAE	10.2658	8.4274	6.9588	6.4831	6.6995	6.4514	5.6106	5.4979	5.5326	**5.4417**
	RMSE	19.4368	14.7926	13.1335	11.8229	13.5320	11.9850	9.9738	9.5631	**9.5611**	9.8878
	MAPE	1.0477	0.4179	0.5020	0.4845	0.4753	0.4758	0.3974	0.4811	0.4196	**0.3882**

Figure 6: Evaluation and comparison of *AGND* (Wu et al., 2023)

Figure 6 shows that *AGND* outperformes other state-of-the-art models. Besides that *AGND* surpasses *DCRNN* across all metrics and both datasets. The use of a data-driven adjacency matrix and attention mechanisms was highlighted as critical to this improvement and allowed the model to adapt to varying traffic conditions more effectively than traditional methods. In conclusion, the *AGND* framework significantly advances traffic demand forecasting by effectively integrating adaptive diffusion processes with sophisticated attention mechanisms. These innovations allow the model to capture the complex spatial-temporal dependencies inherent in traffic systems more accurately, resulting to better forecasting outcomes. The framework also provides a versatile approach that can be adapted to other spatiotemporal prediction tasks. This demonstrates its broader applicability and potential impact across various domains.

2.4 *DCRNN* for speech emotion recognition (*SER*)

The paper "A *DCRNN*-based ensemble classifier for speech emotion recognition in Odia language" (Swain et al., 2022) proposes an ensemble classifier model for *SER* based on a *DCRNN*. The study is conducted on the Odia language[1] and aims to improve the accuracy of *SER* by leveraging a combination of deep learning techniques and ensemble methods.

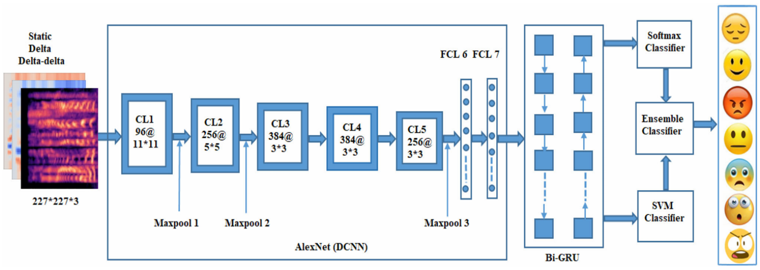

Figure 7: Architecture of *DCRNN* in speech emotion recognition (Swain et al., 2022)

Approach
This study utilized a pre-trained AlexNet model for feature extraction which was then fine-tuned for the emotion recognition task. The study presents an ensemble classifier based on a *DCRNN*. Its architecture combines *Deep Convolutional Neural Network (CNNs)* with *Bi-directional Gated Recurrent Units (Bi-GRUs)*. The CNN is used to extract utterance-level log Mel-spectrograms which are a type of audio representation that shows how the energy of sound frequencies evolves over time and is especially useful for speech and music analysis. By including their first and second derivatives (Static, Delta, and Delta-delta), it creates 3-D log Mel-spectrograms which then serve as the input to the model. The *Bi-GRU* captures long-term temporal dependencies in the extracted features. The ensemble classifier combines *Softmax* and *Support Vector Machine (SVM)* to improve the final recognition accuracy.

Evaluation
The proposed method was evaluated on both the *Odia dataset*, which was specifically developed for this study and contains seven emotional states and the *RAVDESS dataset*, which covers eight different emotional states. Previous studies as well as the mentioned papers from 2.1 and 2.2 often focused on a single dataset, potentially limiting their generalizability across different contexts and emotions. The ensemble classifier outperformed individual classifiers, achieving 85.31% accuracy on the Odia dataset and 77.54% accuracy on the RAVDESS dataset. These results demonstrate that the proposed model effectively captures emotional features and performs well compared to traditional and also state-of-the-art methods. Finally the paper demonstrates that a well-optimized

[1]Odia language: The official language spoken in eastern India.

deep learning model with an ensemble of classifiers and pre-trained feature extraction methods can significantly outperform traditional models, especially on diverse datasets like *Odia* and *RAVDESS*.

3 Conclusion

The exploration of various methodologies for traffic forecasting, particularly focusing on the *DCRNN* and its adaptations, highlights innovative developments in modeling the complex spatiotemporal dependencies characteristic of traffic networks. The *DCRNN* effectively addresses the complex spatial correlations inherent in road networks through diffusion convolution while leveraging an *RNN* architecture to model temporal dependencies. This integrated approach has proven highly effective in mitigating challenges such as non-linear traffic patterns and long-term forecasting difficulties and positions the *DCRNN* as a leading model in the domain of traffic forecasting.

The graph-partitioning-based *DCRNN* introduced a significant enhancement by enabling the model to scale efficiently to large traffic networks. This was achieved by partitioning the network into smaller, more manageable subgraphs and using *Overlapping Nodes* to ensure the continuity of traffic patterns across these partitions. The evaluation on the expansive *California highway* network underscored the effectiveness of this approach which shows improvements in forecast accuracy and computational efficiency. The method's ability to maintain performance while significantly reducing training time makes it a practical solution for large-scale traffic forecasting.

A further advancement is seen in the Adaptive Graph Neural Diffusion (*AGND*) model which introduces attention mechanisms to the *DCRNN* framework. The *AGND* enhances the model's ability to capture long-term dependencies by dynamically adjusting the importance of different spatial and temporal features and the usage of data-driven methods to construct the adjacency matrix. This innovation allows the *AGND* to better adapt to varying traffic conditions which leads to improved forecasting performance across multiple datasets. The success of the *AGND* model in outperforming traditional *DCRNN* approaches across various metrics highlights the potential of attention-based mechanisms in spatiotemporal forecasting tasks.

Lastly, the application of *DCRNN* in fields beyond traffic forecasting, such as in speech emotion recognition (*SER*), showcased the model's versatility. The ensemble classifier for *SER*, which combines *DCRNN* with other deep learning techniques, achieved notable improvements in accuracy across diverse datasets. This demonstrates the *DCRNN*'s potential as a general-purpose model capable of handling complex temporal sequences in various domains.

In conclusion, the advancements in *DCRNN* and its adaptations have significantly enhanced the capabilities of traffic forecasting models, making them more scalable, accurate and adaptable to complex real-world scenarios. These innovations not only set the *DCRNN*'s position as a state-of-the-art model for spatiotemporal forecasting but also open new directions for its application in other fields and underscore its broad utility for future research.

References

Bengio, S., Vinyals, O., Jaitly, N., & Shazeer, N. (2015). Scheduled sampling for sequence prediction with recurrent neural networks.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling.

Laña, I., Del Ser, J., Vélez, M., & Vlahogianni, E. I. (2018). Road traffic forecasting: Recent advances and new challenges.

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting.

Mallick, T., Balaprakash, Rask, E., & Macfarlane, J. (2020). Graph-partitioning-based diffusion convolutional recurrent neural network for large-scale traffic forecasting.

Selvi, K. T., Saranya, S. M., & Thamilselvan, R. (2021). Diffusion convolution recurrent neural network – a comprehensive survey.

Swain, M., Maji, B., Kabisatpathy, P., & Routray, A. (2022). A dcrnn-based ensemble classifier for speech emotion recognition in odia language.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going.

Wu, Y., Zhang, X., & Wang, Y. (2023). Adaptive graph neural diffusion for traffic demand forecasting.