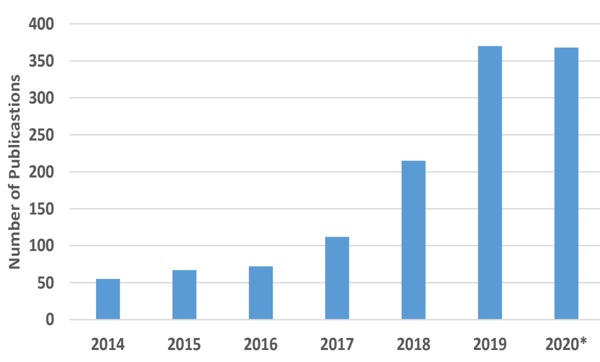
**Dissertation summary**: **Deep Learning for Short-term Network-wide Road Traffic Forecasting**

By Zhiyong Cui

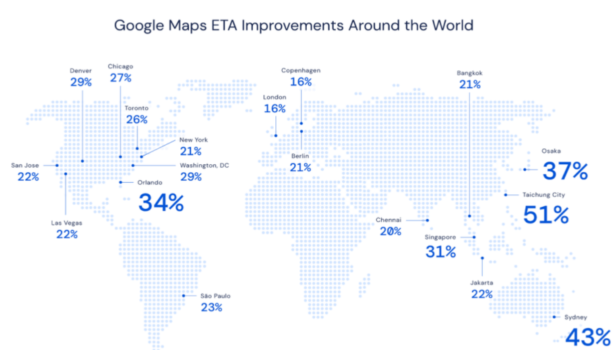
**How will artificial intelligence enhance the understanding of urban traffic?**

The advancement of new smart traffic sensing and communication technologies has stimulated significant growth in the volume and variety of transportation data, which provide us new perspectives to understand the evolving urban traffic patterns and predict those patterns in the foreseeable future. The prediction of traffic is a critical component of modern intelligent transportation systems (ITS) for urban management and control. With the rise of artificial intelligence (AI), short-term traffic prediction-related research attracted more and more attention. Figure 1 (a) shows the number of published papers on the topic of deep learning-based short-term traffic forecasting [1]. It is obvious that there is a surge of studies in this research area starting from 2017. However, before 2016, traffic forecasting research mostly focuses on the prediction of traffic states on a single roadway or in small areas using classical statistical and machine learning models.

However, learning and forecasting network-scale traffic states based on spatial-temporal traffic data is particularly challenging for classical statistical and machine learning models due to the time-varying traffic patterns and the complicated spatial dependencies on road networks. The existence of missing values in traffic data makes this task even harder. Thus, after 2017 a new type of studies, i.e. deep learning-based traffic prediction, has been greatly studied. This research topic is popular not only for the reason that new traffic prediction research challenges appear, including dealing with large-scale traffic network, but also because the traffic prediction research employing novel deep learning technologies has the great potential to be applied into real practice. Figure 1 (6) shows a study on graph neural network-based traffic prediction and estimation times of arrival (ETA) conducted by DeepMind collaborated with Google Maps [2]. It can be observed that the accuracy of real-time ETAs based on novel deep learning methods can be improved by up to 50% compared to the performance of the classical methods.




(a) A history of number of publications with the topics of deep learning-based traffic prediction (retrieved from Scopus in November 2020, Source: [1]).




(b) Deep learning-based traffic prediction application implemented by DeepMind on Google Map (source: [2])

Figure 1. Traffic prediction attracts much attention in recent years partially because of the superior prediction accuracy and the deployable potentials of deep learning models in traffic prediction applications.

**Significance and Objectives**: This dissertation attempts to understand how AI can enhance the understanding of evolving urban traffic patterns and solve the aforementioned traffic prediction challenges by proposing state-of-the-art deep learning methods. Specifically, this research attempts to answer: (1) how to design proper deep learning models to deal with complicated network-wide traffic data and extract comprehensive features to enhance prediction performance, and (2) how to evaluate and apply existing deep learning-based traffic prediction models to further facilitate future research?

To address those key challenges in short-term road traffic forecasting problems, this work develops deep learning models and applications to 1) extract comprehensive features from complex spatial-temporal data to enhance prediction performance, 2) address the missing value issue in traffic forecasting tasks, and 3) deal with multi-source data, evaluate existing deep learning-based traffic forecasting models, share model results as benchmarks, and apply those models into practice.

**Technical Approach**: This work makes original methodological contributions to short-term network-wide traffic forecasting research from several perspectives: The traffic feature learning can categorized as learning traffic data as spatial-temporal matrices and learning the traffic network as a graph. 1) Stacked bidirectional recurrent neural network is proposed to capture bidirectional temporal dependencies in traffic data. 2) To learn localized features from the topological structure of the road network, two deep learning frameworks incorporating graph convolution and graph wavelet operations, respectively, are proposed to learn the interactions between roadway segments and predict their traffic states. 3) To deal with missing values in traffic forecasting tasks, an imputation unit is incorporated into the recurrent neural network to increase prediction performance. 4) Further, to fill in missing values in the graph-based traffic network, a graph Markov network is proposed, which can infer missing traffic states step by step along with the prediction process. In summary, the proposed graph-based neural network models not only achieve superior forecasting performance but also increase the interpretability of the interaction between road segments during the forecasting process.

**Application Context and Potential**: In this dissertation, from the practical perspective, an open-source data and model sharing platform, TraffiX.ai[1], for evaluating existing traffic forecasting models as benchmarks is established to further facilitate future research. It has been pre-selected as the online platform by the TRB AED 50 Standing committee on Artificial Intelligence and Advanced Computing Applications for people to gain hands-on experience of learning and using AI technologies for ITS problem-solving. Additionally, a Traffic Performance Score (TPS) measurement platform [2] is presented which has the capability of taking the proposed network-wide traffic prediction models into practice. Currently, the most recent deep learning-based traffic prediction models, such as recurrent neural network-based methods, encoder-decoder structured methods, and Transformer-based methods [3], are deployed on this TPS platform. This dissertation is the first research that proposes deep learning-based methods and applying them to practical AI platforms.

**Dissertation Scope, Overview, and Contributions**: This dissertation has introduced a series of deep learning-based models for solving the short-term network-wide traffic forecasting problems under two scenarios, i.e. traffic forecasting with and without missing values. It starts with a review of the historical and current status of spatial-temporal traffic data learning and prediction. Then, the introduced traffic forecasting research can be categorized into two types, **learning traffic as a matrix** using RNN-based networks and **learning traffic as a graph** using graph-based neural networks. Furthermore, it has also introduced a set of methodologies and platforms to process, integrate, and formulate transportation data for enhancing prediction performance and sharing experimental results as benchmarks to facilitate future research. The primary contributions of this dissertation and the organization of the remaining chapters are presented in Figure 2.

[1] https://ai.uwstarlab.org/
[2] http://tps.uwstarlab.org/

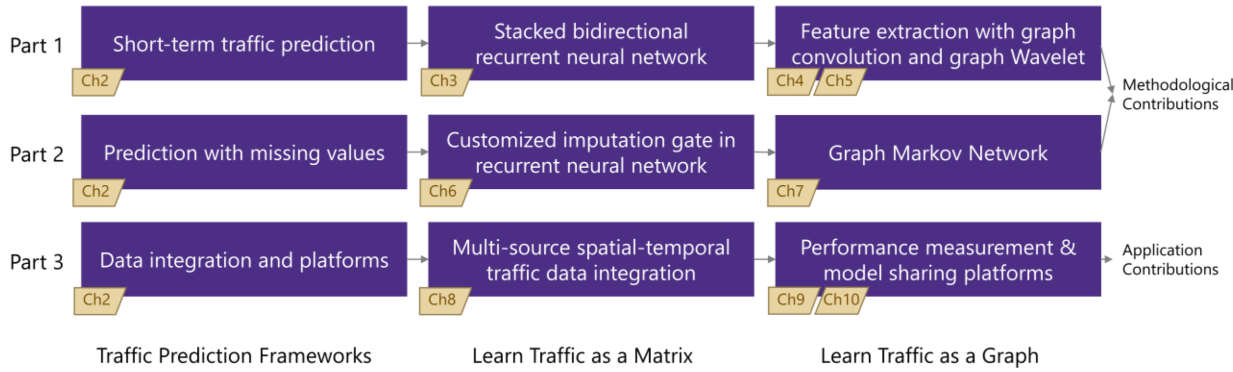



Figure 2. Dissertation organization and summary of contribution

**Part 1: Prediction**. The first part of the dissertation proposes three types of neural network structures based on the network-wide traffic state datasets and their various characteristics to achieve specific traffic prediction goals. Chapter 3 describes a stacked bidirectional recurrent neural network to take the bidirectional dependencies in both time series and road segments into consideration to achieve good prediction performance [4]. To incorporate roadway physical properties and extract localized features from the road network, We learn the traffic network as graph. Chapter 4 presents the traffic graph convolutional recurrent neural network, which incorporates traffic speed and roadway length into the graph convolution operation [5]. Chapter 5 moves one step further to apply graph wavelet operation into the recurrent neural network structure to dynamically capture the localized features in traffic network [6].

**Part 2: Prediction with Missing Values**. The second part mainly explores the challenges of missing values in traffic prediction tasks by designing neural network models. Chapter 6 adds an imputation unit in the LSTM structure to infer the missing values during the prediction process [7]. Learning traffic network as graph, Chapter 7 presents a graph Markov process, which describes the network-wide traffic state transition process [8]. Based on the graph Markov process, a graph Markov network is proposed with the capability of imputing missing values and predicting future traffic states at the same time.

**Part 3: Integration and Platforms**. The third part of the dissertation mainly describes how the proposed methods in the previous two parts can be shared and applied into practice to benefit future research. Chapter 8 describes a multi-source traffic data integration framework whose generated integrated data can be a source of traffic prediction tasks [9]. Chapter 9 introduces an open source platform, TraffiX.ai, which shares standardized traffic prediction datasets and publishes the source code and evaluation results of existing deep learning traffic prediction models as benchmarks. Chapter 10 demonstrates a network-wide traffic data based performance measurement platform which analyzes traffic performance under special conditions and applies the proposed the proposed traffic prediction methods into real applications.

**Discussion and Concluding Remarks**: This dissertation attempts to bridge the gap between classical transportation data modeling and the developing trend in the traffic pattern prediction and estimation field. It not only proposes deep learning-based models to deal with specific traffic prediction tasks, but also applies those models into practice by establishing data sharing and model evaluation platforms. The remaining gaps in the traffic prediction are also discussed in this dissertation.

**Reference**


[1] E. L. Manibardo, I. Laña, and J. Del Ser, "Deep learning for road traffic forecasting: Does it make a difference?," *IEEE Trans. Intell. Transp. Syst.*, 2021.

[2] O. Lange and L. Perez, "No Title." https://deepmind.com/blog/article/traffic-prediction-with-advanced-graph-neural-networks.

[3] A. Vaswani *et al.*, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[4] Z. Cui, R. Ke, and Y. Wang, "Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction," 2018, [Online]. Available: http://arxiv.org/abs/1801.02143.

[5] Z. Cui, K. Henrickson, R. Ke, and Y. Wang, "Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting," *IEEE Trans. Intell. Transp. Syst.*, 2019.

[6] Z. Cui, R. Ke, Z. Pu, X. Ma, and Y. Wang, "Learning traffic as a graph: A gated graph wavelet recurrent neural network for network-scale traffic prediction," *Transp. Res. Part C Emerg. Technol.*, vol. 115, p. 102620, 2020.

[7] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," *Transp. Res. Part C Emerg. Technol.*, vol. 118, p. 102674, 2020.

[8] Z. Cui, L. Lin, Z. Pu, and Y. Wang, "Graph Markov network for traffic forecasting with missing data," *Transp. Res. Part C Emerg. Technol.*, vol. 117, p. 102671, 2020.

[9] Z. Cui, K. Henrickson, S. A. Biancardo, Z. Pu, and Y. Wang, "Establishing multisource data-integration framework for transportation data analytics," *J. Transp. Eng. Part A Syst.*, vol. 146, no. 5, p. 4020024, 2020.