Long-term High-resolution Traffic Flow Forecasting

Michael Shell
School of Electrical and
Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: http://www.michaelshell.org/contact.html

Homer Simpson
Twentieth Century Fox
Springfield, USA
Email: homer@thesimpsons.com

James Kirk
and Montgomery Scott
Starfleet Academy
San Francisco, California 96678–2391
Telephone: (800) 555–1212
Fax: (888) 555–1212

Abstract—**Long-term urban mobility predictions are pivotal for the effective management of urban facilities and services. Traditionally, urban mobility data has been structured into spatiotemporal videos, treating longitude and latitude grids as the fundamental pixels. Consequently, video prediction methods, which rely on Convolutional Neural Networks (CNNs), have played an instrumental role in this field for an extended period. In our research, we propose a novel perspective regarding urban mobility forecasting. Instead of simplifying urban mobility as mere video data, we posit it as a complex super-multivariate time series. Addressing the challenges of prediction requires a comprehensive exploration of global spatiotemporal and frequency-domain correlations. To address this challenge, we introduce an architecture tailored specifically for urban mobility forecasting. Our model comprises three transformer modules operating in the temporal, spatial, and frequency domains, respectively. As a result, it excels in regional mobility pattern modeling and long-term forecasting, surpassing current state-of-the-art methods by 10%. Our approach provides a fresh and promising outlook on the field of urban mobility forecasting.**

I. Introduction

In the realm of urban mobility computing, a diverse array of spatiotemporal data exists, encompassing different organizational structures, scales, and modes. These data are characterized by their dynamic nature, evolving continuously across both time and space. Among the prominent forms of urban dynamic spatiotemporal data are point-based, line-based, and grid-based data [1]. Grid-based data, in particular, involve the division of urban areas into grids based on latitude and longitude coordinates. Each grid contains a wealth of attributes for the current spatiotemporal slot, including latitude and longitude coordinate ranges, Points of Interest (POI), cumulative in/out vehicle counts (in/out traffic flow), and various other relevant information. Predicting grid-based data is crucial as it serves as a foundational framework for spatial analysis and modeling, enabling the assessment, prediction, and management of various urban phenomena, spanning from traffic patterns and congestion hotspots to land use dynamics.

Traditionally, the practice of structuring grid-based mobility data in a video format (T, C, H, W) has naturally emerged due to its alignment with the inherent characteristics of the data. Here, T corresponds to the number of time points, C represents the attributes, and H and W indicate the latitude and longitude dimensions of the urban area. Importantly, recent years have witnessed significant advancements in deep video prediction techniques, with Convolutional Neural Networks (CNNs) serving as their backbone. Consequently, this has led to widespread adoption of CNN-based methods for predicting grid-based mobility data.

A paragraph introducing SimVP, tau and so on. They mistakenly mix up natural video prediction with traffic prediction, which contradicts the 'no free lunch' theorem. Moreover, both of them emphasize short-term forecasting, neglecting the significance of long-term prediction, whereas the traffic domain primarily concerns itself with long-term forecasting

A figure describing that why CNN is not the best choice for it. CNNs excel in image recognition tasks due to their inductive bias for capturing local spatial hierarchies and patterns within grid-like image data. However, when it comes to urban mobility data, which typically involves dynamic spatiotemporal patterns and attributes, CNNs may not be as well-suited for several reasons:

In this paper, inspired by the enduring success of Transformer models in time series modeling, we aim to harness the formidable learning capabilities of Transformers to address the challenge of long-term super-multivariate time series forecasting. For temporal modeling, we divide the time series into patches, which serve as input tokens for the Transformer as well as keeping semantic information [2]. Additionally, we employ a hierarchical feature structure, starting with small-size patches. As the network goes deeper, we use non-overlapping fixed window lengths to merge features from adjacent patches. This enables the model to simultaneously focus on the fine-grained details and the overall trends within the time series. In spatial modeling, we've employed an efficient router module to capture the correlation patterns of super-multivariate time series in each hierarchy with linear complexity. Furthermore, we introduce a frequency-domain denoising module that enhances the model's generalization performance for long-term predictions by sampling the lower frequencies.

We collected traffic flow datasets from three world famous metropolises: Beijing, New York, and Chengdu. Our approach achieved remarkably high predictive performance on these datasets, surpassing CNN-based models by xx%. This success underscores the feasibility of modeling traffic flow as time series data. Furthermore, our model takes into account the interconnections between different grids, outperforming the Channel-Independence Time Series Forecasting model, which neglects the correlations between grids. This demonstrates the necessity of capturing the relationships between super many grids.

In real-world scenarios, users are more concerned about potential risk events caused by future traffic peaks. Therefore, we employed a training strategy that, while causing a slight decline in the RMSE and MAE metrics across the entire test dataset, results in the model making more accurate predictions for traffic flow density peaks in both temporal and spatial views. (!!!!!!two figure, temporal more accurate in peak,spatial more accurate in peak region)

We firmly believe that our work not only serves to draw the attention of the urban computing research community towards long-term urban traffic flow prediction but also contributes to the research on cross-transportation-modal perception in smart cities. This, in turn, enhances the convenience for citizens and elevates traffic management of transportation authorities in supercities. Furthermore, we acknowledge that the field of multivariate time series research has primarily been confined to a few datasets with relatively fewer time variables. Traffic grid flow data, due to its flexible resolution and a vast number of time variables, can introduce novel research challenges and opportunities to the multivariate time series research.

II. Related Work

A. Urban Mobility Prediction as Video Prediction

In recent years, urban mobility prediction has become a focal point for machine learning researchers. This interest stems from the natural ability to divide a city into distinct regions using latitude and longitude ranges, akin to treating the entire city as a hyperspectral image. In this analogy, various urban attributes can be likened to unique spectral bands within this "urban hyperspectral image." Additionally, considering the significant breakthroughs achieved by early CNNs in image processing, much of the initial research in this domain was founded on CNN-based methodologies since the pioneering study by [3]. Several following models combine CNN with other structures to improve their ability to capture spatiotemporal dependencies. In some studies [4], [5], Recurrent Neural Networks (RNNs) are employed to model periodic temporal dependencies. ST-GAN [6] employs a generative adversarial training strategy to train a CNN, enabling the learned model to generate realistic simulations.

The authors of DeepSTN+ have noted that CNN struggles to capture long-range spatial dependencies, leading them to incorporate the ConvPlus structure for this purpose [7].

It's worth mentioning that urban mobility datasets like taxbj [3] have also emerged as one of the standard test datasets for generic video prediction algorithms. Numerous video prediction frameworks evaluate their model on urban mobility datasets, including...

While current mainstream frameworks like simvp continue to rely on CNNs, there is a growing trend questioning whether truly universal video prediction genuinely requires CNNs []. Another drawback of existing works is their limited emphasis on long-term forecasting. The majority of them concentrate on short-term predictions, such as the standard experiment prototype with only 4 steps ahead. However, as mentioned in the introduction, it is crucial to forecast mobility patterns several days in advance for effective urban management.

B. Multivariate time series forecasting Framework

Deep neural networks (DNNs), especially Transformer models, have made significant advancements in the field of time series forecasting, with a primary focus on long-term prediction since the inception of early works like Informer [8]. The key to the success of multivariate time series forecasting lies in their ability to model cross-variable correlations. From this perspective, these methods can be broadly categorized into variable-dependent strategies [8]–[11] and variable-independent strategies [2], [12], [13] (To avoid any confusion with the concept of "channels" in the video, we use "variable" instead of the original term 'channels' as mentioned in the original paper [2].). The variable-dependent strategy treats multivariate time series as a whole, aiming to make joint predictions using information from all variables. However, the variable-dependent strategy has been scrutinized for its lack of robustness when faced with distribution shifts among different variables [14]. In contrast, rather than treating multiple variables as a whole, some models adopt a variable-independent strategy [2], [13], applying a univariate forecasting model to multiple variables sharing the same parameters. These approaches overlook the correlations and heterogeneity among different variables but have demonstrated improved performance [14]. However, the variable-independent strategy may lead to an unaffordable computational cost when dealing with a large number of series variables. In the context of urban mobility data, a high-resolution urban grid can result in a substantial number of time series. Moreover, it is well-known that areas with similar semantic and geographical characteristics can exhibit strong correlations. In order to overcome this challenge, we propose spatial module.

C. Deep learning model leveraging frequency-domain information

The Fast Fourier Transform (FFT) algorithm efficiently converts data from the time domain to the frequency domain and serves as a frequency-domain feature extraction module in constructing neural network architectures [10], [15]. Initially proposed as a data-driven method for solving partial differential equations (PDEs), the Fourier Neural Operator (FNO) [16] has subsequently proven effective in image classification [17] and time series forcasting [9], [10], [15].

Given the pronounced temporal periodicity in traffic flow, we opt to utilize FNO1d [16] and FNO3d [16] to model traffic flow as videos and super-multivariate time series respectively in the subsequent experimental section, assessing their feature extraction and prediction performance. Our proposed model draws inspiration from FNO's app roach of sampling the lower frequencies in the spectrum modes. We incorporate FNO-based block into the Transformer architecture.

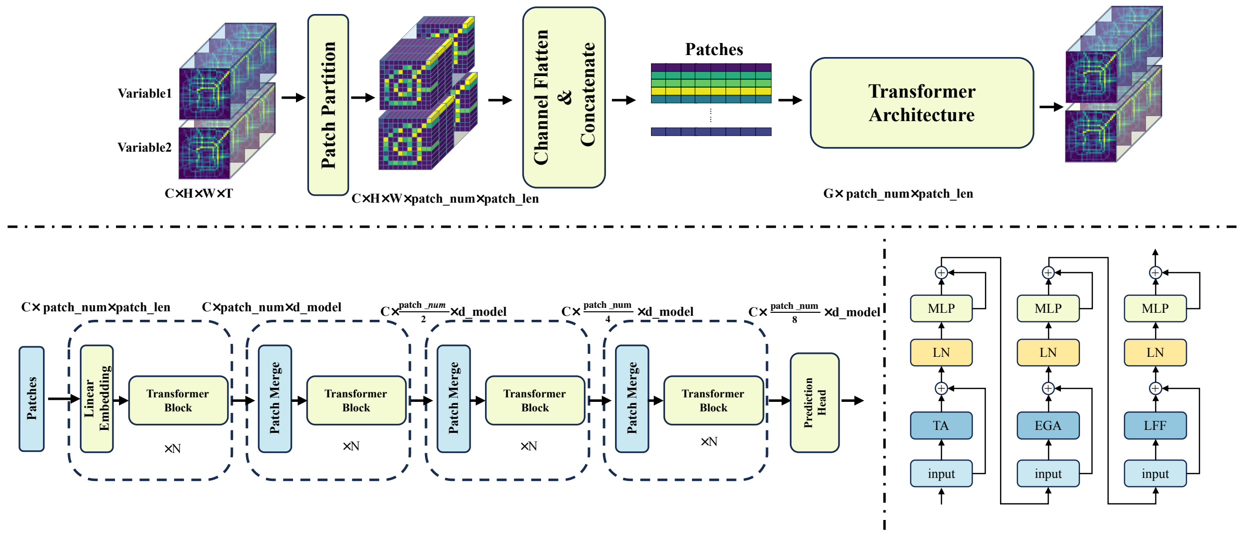

Fig. 1. Architecture.

III. Methodology

IV. Conclusion

The conclusion goes here.

Acknowledgment

The authors would like to thank...

References

[1] J. Wang, J. Jiang, W. Jiang, C. Han, and W. X. Zhao, "Towards efficient and comprehensive urban spatial-temporal prediction: A unified library and performance benchmark," *arXiv preprint arXiv:2304.14343*, 2023.

[2] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A time series is worth 64 words: Long-term forecasting with transformers," *arXiv preprint arXiv:2211.14730*, 2022.

[3] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[4] A. Zonoozi, J.-j. Kim, X.-L. Li, and G. Cong, "Periodic-crn: A convolutional recurrent model for crowd density prediction with recurring periodic patterns." in *IJCAI*, vol. 18, 2018, pp. 3732–3738.

[5] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 5668–5675.

[6] H. Zhang, Y. Wu, H. Tan, H. Dong, F. Ding, and B. Ran, "Understanding and modeling urban mobility dynamics via disentangled representation learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2010–2020, 2020.

[7] Z. Lin, J. Feng, Z. Lu, Y. Li, and D. Jin, "Deepstn+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 1020–1027.

[8] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.

[9] H. Wu, J. Xu, J. Wang, and M. Long, "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting," *Advances in Neural Information Processing Systems*, vol. 34, pp. 22 419–22 430, 2021.

[10] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *International Conference on Machine Learning*. PMLR, 2022, pp. 27 268–27 286.

[11] Z. Zhang, X. Wang, and Y. Gu, "Sageformer: Series-aware graph-enhanced transformers for multivariate time series forecasting," *arXiv preprint arXiv:2307.01616*, 2023.

[12] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[13] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 121–11 128.

[14] L. Han, H.-J. Ye, and D.-C. Zhan, "The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting," *arXiv preprint arXiv:2304.05206*, 2023.

[15] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," *arXiv preprint arXiv:2210.02186*, 2022.

[16] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," *arXiv preprint arXiv:2010.08895*, 2020.

[17] W. Johnny, H. Brigido, M. Ladeira, and J. C. F. Souza, "Fourier neural operator for image classification," in *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE, 2022, pp. 1–6.