Learning Urban Dynamics via Joint Forecasting of Traffic Speed and Ride-Hailing Demand

Hassan Idoudi*[1], S. M. Hassan Mahdavi†[1], Negin Alisoltani‡[2], Jaafar Berrada§[1], and Mostafa Ameli¶[2]

[1]Institut VEDECOM, Human Factors and Economics for Sustainable Mobility, Versailles, France
[2]Université Gustave Eiffel, COSYS-GRETTIA, Paris, France

Short summary

Traffic speed and ride-hailing demand are closely intertwined in dense urban areas, yet they are almost always predicted by separate, unconnected models. This paper presents STT-CTA (Spatial-Temporal Transformer with Cross-Task Attention), a single architecture that jointly predicts hourly traffic speed and ride-hailing demand on a regular urban grid using factorized spatial-temporal self-attention and a learnable-gated cross-task attention module that enables selective information exchange between tasks. We evaluate the model on one full year of data from Chicago's Loop district, combining traffic sensor readings, trip records, weather observations, and public holidays into a shared input representation. The key result is an *asymmetric directional coupling*: letting the demand branch attend to speed signals reduces demand MAPE by 2.2 percentage points and MAE by up to 8.4% at a 6-hour horizon, whereas feeding demand information back into the speed branch degrades speed accuracy. This asymmetry supports a hybrid deployment strategy and carries practical implications for the design of multi-task urban transportation forecasting systems, where different tasks may benefit unequally from shared information.
Keywords: cross-task attention, multi-task learning, ride-hailing demand, spatio-temporal forecasting, traffic speed prediction, Transformer

1 Introduction

Traffic speed forecasting and ride-hailing demand prediction in dense commercial districts are hard problems in their own right, and each becomes harder when treated in isolation. Congestion slows traffic, but it also pushes travelers toward on-demand rides; conversely, ride-hailing vehicles contribute to downtown congestion Erhardt et al. (2019). Despite this documented coupling, operational forecasting systems almost never model the two together—each task has its own model, its own training pipeline, and its own data.

On the research side, deep learning has pushed the state of the art in both areas. Traffic speed prediction evolved from recurrent networks Ma et al. (2015); Cui et al. (2018) to graph convolutional methods Yu et al. (2018); Li et al. (2018); Wu et al. (2019) and more recently to Transformer-based approaches Jiang et al. (2023); Yan et al. (2021); Liu et al. (2023) that capture long-range dependencies without needing a fixed road graph. Demand prediction followed a similar path, from convolutional grid models Zhang et al. (2017); Yao et al. (2018) to graph networks at the zone level Geng et al. (2019). Some multi-task work has coupled demand prediction across transport modes Ye et al. (2021) or across ride-hailing service types Ke et al. (2021); Guo et al. (2024), but none has jointly modeled traffic speed and ride-hailing demand within a shared architecture. Straightforward hard parameter sharing Ma et al. (2019) is also a poor fit when the two tasks pull the loss in different directions Kendall et al. (2018), motivating a selective, gated cross-task communication mechanism.

*Corresponding author. hassan.idoudi@vedecom.fr. ORCID: 0000-0003-2833-9797
†hassan.mahdavi@vedecom.fr. ORCID: 0000-0001-5087-1531
‡negin.alisoltani@univ-eiffel.fr. ORCID: 0000-0003-0458-3074
§jaafar.berrada@vedecom.fr. ORCID: 0000-0001-8016-8847
¶mostafa.ameli@univ-eiffel.fr. ORCID: 0000-0002-2470-6812

We propose STT-CTA, a **Spatial-Temporal Transformer with Cross-Task Attention**, that jointly predicts hourly traffic speed and ride-hailing demand on an urban grid. The work makes four contributions: (1) a data fusion pipeline that combines traffic sensor readings, TNP trip records, weather, and public holidays into a single spatio-temporal tensor; (2) a factorized Transformer encoder with separate spatial and temporal attention that avoids the quadratic cost of joint space-time attention; (3) a learnable-gated cross-task attention module that enables selective, directional information exchange between the speed and demand branches; and (4) an empirical demonstration, to the best of our knowledge the first in the literature, that the inter-task coupling is asymmetric: speed information substantially improves demand forecasting, while demand information degrades speed prediction.

2 Methodology

Problem formulation and data

We focus on Chicago's Loop district (Community Area 32), a compact commercial core that offers a dense set of traffic sensors and a large volume of ride-hailing activity, making it well suited for studying the interaction between the two. Following the grid-based approach common in urban traffic prediction Zhang et al. (2017); Ma et al. (2017), we partition the Loop into a 6×6 grid of 500 m cells. At each hour t, we build a feature tensor $\mathbf{X}_t \in \mathbb{R}^{6\times6\times15}$ from four data sources (Table 1). The prediction task is: given the past 24 hours as input ($T_{\text{in}} = 24$), jointly forecast average traffic speed $\hat{\mathbf{Y}}^{(s)} \in \mathbb{R}^{T_{\text{out}}\times6\times6}$ and ride-hailing demand $\hat{\mathbf{Y}}^{(d)} \in \mathbb{R}^{T_{\text{out}}\times6\times6\times2}$ (pickups and drop-offs separately) for the next $T_{\text{out}} = 6$ hours—a window that is useful for both traffic operations and fleet management.

Table 1: Feature groups ($F = 15$ per cell per hour).

Group	**Features**	**Source**
Speed	avg_speed, min_speed, num_segments, bus_count	Chicago Traffic Tracker
Demand	pickup_count, dropoff_count, avg_trip_duration, avg_trip_distance	TNP Trips
Temporal	hour_of_day, day_of_week, is_weekend, month, is_holiday	Derived / US–IL calendar
Weather	temperature_2m, precipitation	Open-Meteo API

We use two open datasets from the City of Chicago for the full year 2025. The **Traffic Tracker** reports estimated segment speeds roughly every 10 minutes; readings flagged as invalid (speed $= -1$, meaning too few probe vehicles were detected) are dropped, leaving valid records across 33 segments. **TNP trip records** cover ride-hailing trips originating or ending within the Loop bounding box, each entry including a timestamp, anonymized pickup and drop-off centroids, trip duration, and distance. Hourly temperature and precipitation come from the Open-Meteo Archive API, and public holidays are taken from the US federal and Illinois state calendars. For cells with missing speed values we apply forward-fill then backward-fill, which is reasonable given that speeds change gradually over short intervals. All features are normalized using only training-set statistics to avoid leakage. The data is split into training (January 1 – September 10), validation (September 11 – October 16), and a held-out test set (October 17 – December 31).

Model architecture

The STT-CTA architecture (Figure 1) consists of four stages.

Patch embedding and positional encoding. Each (time step, grid cell) pair is projected from the 15 raw features into a $d = 64$-dimensional latent space via a shared linear layer. Learnable spatial embeddings ($\mathbf{P}^{(s)} \in \mathbb{R}^{36\times64}$) and temporal embeddings ($\mathbf{P}^{(t)} \in \mathbb{R}^{24\times64}$) are then added to encode grid location and time-of-day identity, respectively.

Spatial and temporal self-attention. We apply two layers of spatial multi-head self-attention Vaswani et al. (2017) ($N_s = 2$, $h = 4$ heads) so each grid cell can attend to all others at a given

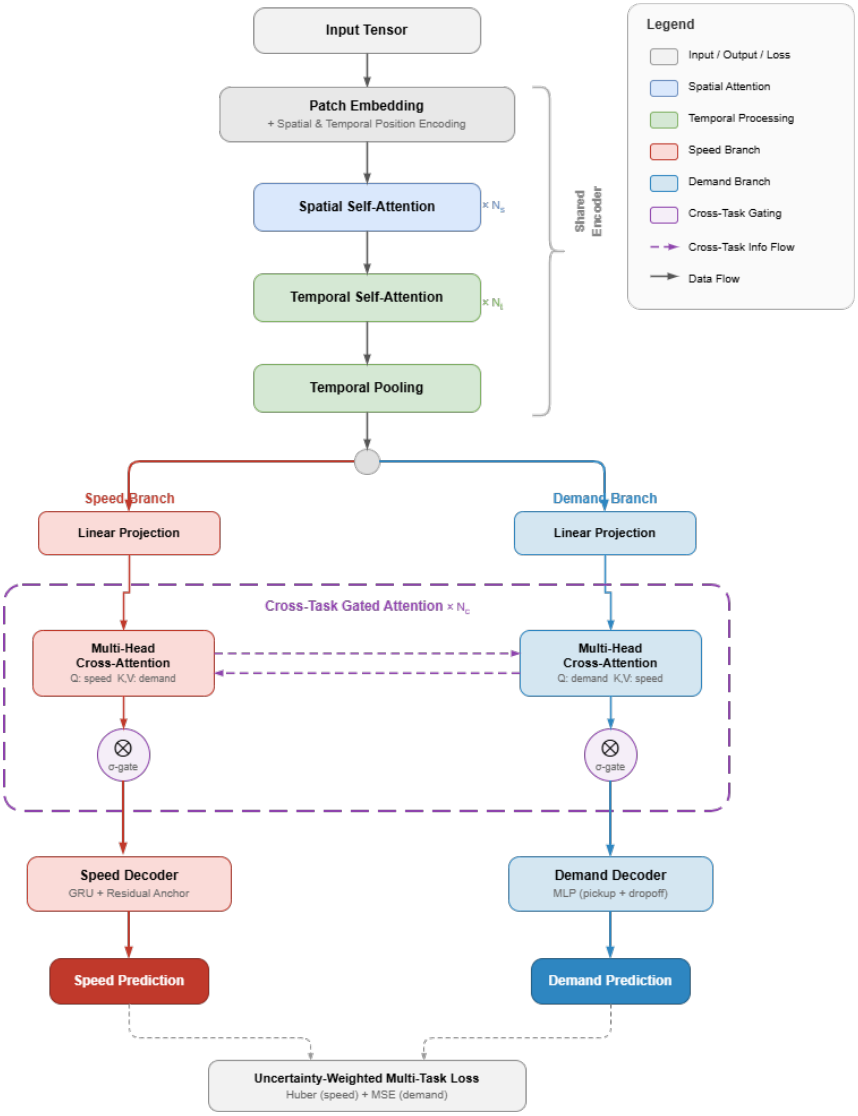


Figure 1: Spatial-Temporal Transformer with gated cross-task attention for joint speed and demand prediction.

time step. This captures spatial dependencies without needing a hand-crafted adjacency matrix, unlike STGCN Yu et al. (2018) or GMAN Zheng et al. (2020). Two temporal attention layers ($N_t = 2$) are then applied per cell to capture recurring daily and weekly patterns Yan et al. (2021). Keeping spatial and temporal attention separate is both computationally cheaper and, in practice, just as accurate as computing them jointly Liu et al. (2023).

Gated cross-task attention. After the shared encoder, task-specific projections produce a speed representation $\mathbf{H}^{(s)}$ and a demand representation $\mathbf{H}^{(d)}$. A bidirectional cross-attention step then lets the two branches exchange information:

$$\tilde{\mathbf{H}}^{(s)} = \mathbf{H}^{(s)} + \sigma(\alpha_s) \cdot \text{CrossAttn}(\mathbf{H}^{(s)}, \mathbf{H}^{(d)}) \tag{1}$$

$$\tilde{\mathbf{H}}^{(d)} = \mathbf{H}^{(d)} + \sigma(\alpha_d) \cdot \text{CrossAttn}(\mathbf{H}^{(d)}, \mathbf{H}^{(s)}) \tag{2}$$

where $\sigma(\cdot)$ is the sigmoid function and α_s, α_d are learnable gate scalars initialized at -5.0 and -2.0. The very small initial value for α_s effectively mutes the demand-to-speed channel at the start of training, while still giving the model the freedom to open it up if the data supports that. The gate for the demand branch starts more open, reflecting our prior expectation that speed is a useful predictor of demand. This is more adaptive than hard parameter sharing Ma et al. (2019) and more general than the same-modality gating used between taxi and TNC demand forecasts Guo et al. (2024). We stack two such cross-task attention layers ($N_c = 2$), each followed by a residual connection and layer normalization.

Prediction heads. Separate MLP decoders take the gated representations and produce the final forecasts: the speed head outputs $\hat{\mathbf{Y}}^{(s)} \in \mathbb{R}^{T_{\text{out}} \times 6 \times 6}$ and the demand head outputs $\hat{\mathbf{Y}}^{(d)} \in \mathbb{R}^{T_{\text{out}} \times 6 \times 6 \times 2}$, with pickups and drop-offs treated as separate channels since their spatial patterns differ.

Training Scheme

We combine the two task losses using the uncertainty-weighted formulation of Kendall et al. (2018):

$$\mathcal{L} = \frac{1}{2\sigma_s^2}\,\mathcal{L}_s + \frac{1}{2\sigma_d^2}\,\mathcal{L}_d + \log\sigma_s + \log\sigma_d \tag{3}$$

Here $\mathcal{L}_s$ is the Huber loss for speed, chosen for its robustness to large incident-driven residuals, and $\mathcal{L}_d$ is mean-squared error for demand. The noise parameters σ_s and σ_d are learned jointly and act as automatic task weights. We optimize with AdamW Loshchilov & Hutter (2019) (initial learning rate 3×10^{-4}, weight decay 10^{-5}, cosine annealing to 10^{-6}), batch size 64, gradient norm clipping at 1.0, and early stopping with patience 10 on validation loss.

3 Results and discussion

Tables 2 and 3 report results on the held-out test set for four models. The **Speed-Only Transformer** is trained on speed targets only, using speed features plus temporal and weather inputs; it has no demand output. The **Demand-Only Transformer** mirrors this setup for demand. Both single-task models use the same spatial and temporal attention backbone as STT-CTA but see only their own task's features and have no cross-task module. The **Independent Dual Transformer** runs one Transformer per task, each trained separately on the full 15-feature input—same capacity as STT-CTA but no information sharing. Finally, **STT-CTA** (ours) is the joint model with gated bidirectional cross-task attention. We report MAE, RMSE, and MAPE throughout.

Table 2: Test set performance, traffic speed prediction.

Model	MAE (mph)	RMSE (mph)	MAPE (%)
Speed-Only	0.985	1.906	9.1
Independent Dual	**0.967**	**1.888**	**9.0**
STT-CTA (Ours)	0.981	1.925	9.2

Table 3: Test set performance, ride-hailing demand prediction.

Model	MAE (trips)	RMSE (trips)	MAPE (%)
Demand-Only	2.912	13.239	41.9
Independent Dual	2.834	13.469	42.6
STT-CTA (Ours)	**2.662**	**12.934**	**39.7**

The results reveal a clear **asymmetric directional coupling**. On the demand side, STT-CTA is the best model across all metrics. Compared to the Demand-Only baseline it reduces MAPE from 41.9% to 39.7% and RMSE from 13.239 to 12.934 trips. Crucially, it also outperforms the Independent Dual Transformer, which has comparable capacity but no cross-task attention, confirming that the gain comes from leveraging speed information rather than from model size.
The benefit grows with the forecast horizon (Table 4): STT-CTA reduces demand MAE by 5.8% over the Independent Dual at 1 hour ahead, and by 8.4% at 6 hours ahead. Speed dynamics thus act as increasingly valuable leading indicators over longer windows, early congestion signals during the morning peak anticipate demand growth that unfolds over the following hours.
For speed prediction the picture reverses. Both task-isolated models outperform STT-CTA across all four temporal segments (Table 5). Even with the speed gate initialized near zero ($\sigma(\alpha_s) \approx 0.007$), residual demand-to-speed cross-attention introduces gradient interference that degrades

Table 4: Demand MAE by forecast horizon.

Model	1 h	3 h	6 h
Demand-Only	2.680	2.811	3.245
Independent Dual	2.568	2.774	3.236
STT-CTA	**2.419**	**2.644**	**2.962**

Table 5: Speed MAE (mph) by temporal segment.

Model	Weekday	Weekend	Peak	Off-Peak
Speed-Only	0.903	0.994	0.912	0.937
Ind. Dual	**0.880**	**0.975**	**0.882**	**0.917**
STT-CTA	1.025	1.091	1.006	1.060

Table 6: Demand MAE (trips) by temporal segment.

Model	Weekday	Weekend	Peak	Off-Peak
Demand-Only	3.530	3.406	4.218	3.195
Ind. Dual	2.770	3.179	3.619	2.588
STT-CTA	**2.655**	**3.126**	**3.515**	**2.493**

the speed branch throughout training. The absolute MAE gap between the Independent Dual and STT-CTA is widest on weekdays (0.145 mph) and off-peak hours (0.143 mph), where traffic volumes are more variable and harder to predict. This result is physically interpretable: arterial speeds are governed by road capacity, signal timing, and aggregate vehicle flow, none of which ride-hailing demand explains to any meaningful degree, as ride-hailing vehicles account for a small fraction of total traffic in the Loop. The converse does not hold: passengers respond elastically to congestion, so speed serves as a causal upstream driver of demand, but not vice versa.

The spatial error maps (Figure 2) support this interpretation. Speed residuals are distributed uniformly across the grid, consistent with the even sensor coverage of major arterials. Demand residuals peak in the central cells co-located with transit interchanges and hotel clusters, where pickup and drop-off volumes are simultaneously highest and most variable. The cross-task attention maps (Figure 3) reinforce these findings: the demand branch assigns strong, spatially structured attention to speed tokens from cells near major arterial intersections, whereas the speed branch's attention to demand tokens remains diffuse and near-uniform, consistent with a gate that never meaningfully opened during training.

No single model wins on both tasks. The most practical path is a **hybrid deployment strategy**: use the Independent Dual Transformer for speed forecasting and STT-CTA for demand forecasting. When a single unified model is required, the speed gate can be clamped to zero after training, restricting information flow to the speed-to-demand direction only; the speed accuracy cost is 0.014 mph in MAE, a negligible trade-off in practice. This extends the multi-task framework of Kendall et al. (2018) and prior joint demand work Ke et al. (2021) to a qualitatively different task pairing that spans traffic state and mobility demand.

4 Conclusions

We presented STT-CTA, a Transformer that jointly forecasts traffic speed and ride-hailing demand on an urban grid, combining sensor data, trip records, weather, and calendar information into a single shared input. The central finding is an **asymmetric directional coupling**: speed information substantially improves demand forecasting, with the gain growing from 5.8% to 8.4% in MAE as the horizon extends from 1 to 6 hours; demand information, by contrast, degrades speed prediction. The learned gate values ($\sigma(\alpha_s) \approx 0.007$, $\sigma(\alpha_d) \approx 0.12$) and the cross-task attention

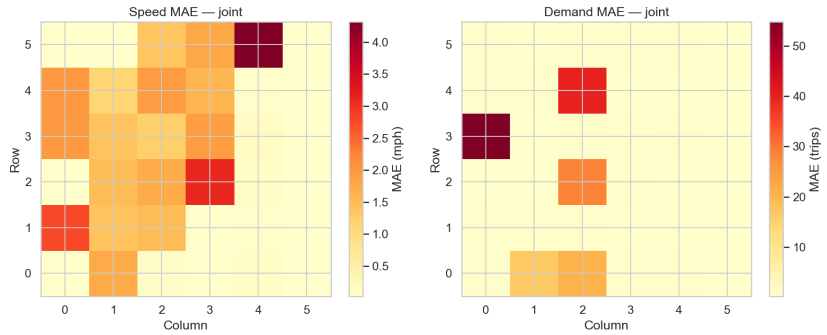


Figure 2: Spatial distribution of speed MAE (left) and demand MAE (right) across the 6×6 Loop grid for STT-CTA.

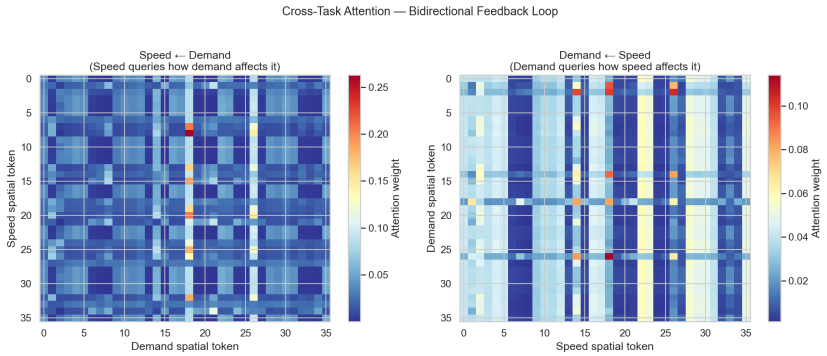


Figure 3: Cross-task attention weights averaged over test samples. The demand branch (right) assigns strong, spatially structured attention to speed tokens from arterial cells; the speed branch (left) shows diffuse, near-uniform attention to demand tokens, consistent with a gate that remained near zero throughout training.

maps independently confirm this asymmetry, indicating that speed acts as a causal upstream predictor of ride-hailing demand, but not vice versa. To our knowledge, this is the first work to jointly model these two tasks and to empirically quantify the directional nature of their coupling. There are clear limits to how far these conclusions can be generalized. The study covers a single small commercial district; whether the same asymmetry holds in larger cities, polycentric areas, or markets with higher ride-hailing penetration is an open question. The grid representation also ignores road topology, which may matter more in other settings. And a single year of data may not capture behavioral shifts driven by pricing changes or longer-term modal trends.
Next steps include scaling to a city-wide grid, comparing grid and graph spatial representations on the same data, and testing whether the coupling direction changes in cities where ride-hailing makes up a larger share of the traffic stream, which would have direct implications for integrated MaaS system design.

REPRODUCIBILITY STATEMENT

The conference paper is eligible for the Reproducible Research Prize. The code and data associated with this work are available at `https://github.com/hassanidoudi/Joint-Forecasting-of-Traffic-Speed-and-Ride-Hailing-Demand`. The repository is prepared at SRL 5.

Conflict of Interest

The authors declare no potential conflict of interest.

Author Contribution Statement

All authors contributed to all aspects of this study, including the conception and design of the methodology, the analysis and interpretation of results, and manuscript preparation.

Acknowledgements

This research was developed as part of the European Union's Horizon Europe project MetaCCaze – Flexibly Adapted Meta Innovations, Use Cases, Collaborative Business and Governance Models to Accelerate Deployment of Smart and Shared Zero Emission Mobility for Passengers and Freight (Grant Agreement No. 101139678). The work was carried out within the framework of the MetaCCaze project, with Hassan Idoudi, S. M. Hassan Mahdavi, and Jaafar Berrada contributing as project partners. Additional support was provided by co-authors affiliated with Université Gustave Eiffel. Negin Alisoltani has been supported by the French I-SITE FUTURE Fund under the project CAMPUS.

References

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2018). Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. *arXiv preprint arXiv:1801.02143*.

Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion? *Science Advances*, *5*(5). doi: 10.1126/sciadv.aau2670

Geng, X., Li, Y., Wang, L., Zhang, L., Yang, Q., Ye, J., & Liu, Y. (2019). Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the aaai conference on artificial intelligence* (pp. 3656–3663).

Guo, Y., Chen, Y., & Zhang, Y. (2024). Enhancing demand prediction: A multi-task learning approach for taxis and tncs. *Sustainability*, *16*(5), 2065.

Jiang, J., Han, C., Zhao, W. X., & Wang, J. (2023). PDFormer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the aaai conference on artificial intelligence* (pp. 4365–4373).

Ke, J., Feng, S., Zhu, Z., Yang, H., & Ye, J. (2021). Joint predictions of multi-modal ride-hailing demands: A deep multi-task multi-graph learning-based approach. *Transportation Research Part C: Emerging Technologies*, *127*, 103063.

Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7482–7491).

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *Proceedings of the international conference on learning representations.*

Liu, H., Dong, Z., Jiang, R., Deng, J., Deng, J., Chen, Q., & Song, X. (2023). Spatio-temporal adaptive embedding makes vanilla transformer sota for traffic forecasting. In *Proceedings of the 32nd acm international conference on information and knowledge management* (pp. 4125–4129).

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. In *Proceedings of the international conference on learning representations.*

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, *17*(4), 818.

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187–197.

Ma, X., Yu, H., Wang, Y., & Wang, Y. (2019). Multi-task learning for traffic speed and flow prediction. *IEEE Access*, *7*, 80007–80016.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (Vol. 30).

Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for deep spatial-temporal graph modeling. In *Proceedings of the 28th international joint conference on artificial intelligence* (pp. 1907–1913).

Yan, H., Ma, X., & Pu, Z. (2021). Learning dynamic and hierarchical traffic spatiotemporal features with transformer. *IEEE Transactions on Intelligent Transportation Systems*, *23*(11), 22386–22399.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., ... Li, Z. (2018). Deep multi-view spatial-temporal network for taxi demand prediction. In *Proceedings of the aaai conference on artificial intelligence* (pp. 2588–2595).

Ye, J., Sun, L., Du, B., Fu, Y., & Xiong, H. (2021). Coupled layer-wise graph convolution for transportation demand prediction. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 35, pp. 4617–4625).

Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th international joint conference on artificial intelligence* (pp. 3634–3640).

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the aaai conference on artificial intelligence* (pp. 1655–1661).

Zheng, C., Fan, X., Wang, C., & Qi, J. (2020). GMAN: A graph multi-attention network for traffic prediction. In *Proceedings of the aaai conference on artificial intelligence* (pp. 1234–1241).