MSinT: a Multi-view Spatio-Informed Transformer for Overweighted Vehicle Forecasting

Ruixuan Zhang, Tu Lan, Zilin Bian*, Kaan Ozbay
New York University

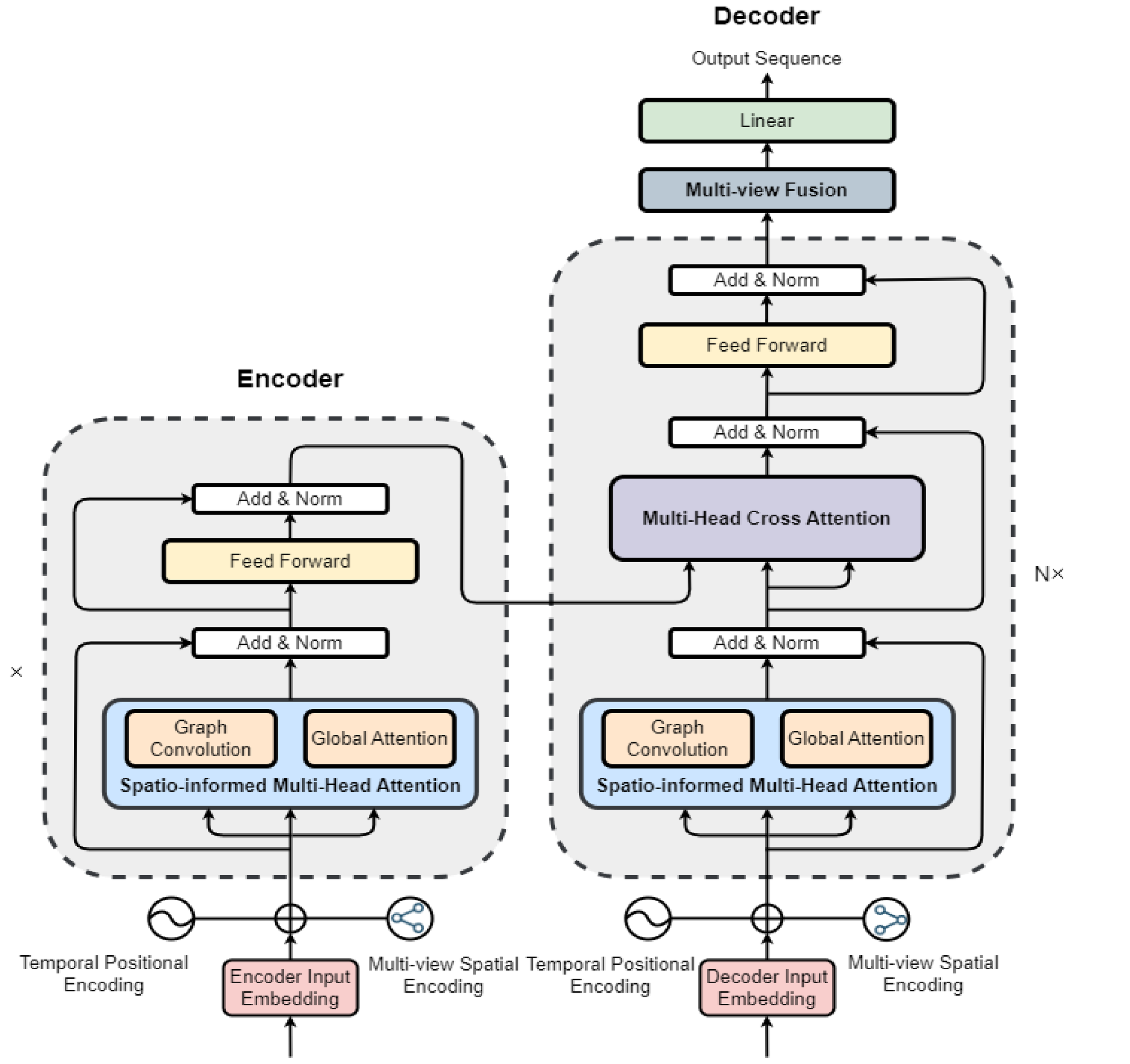
CONNECTED CITIES WITH SMART TRANSPORTATION

Contact: zb536@nyu.edu

Abstract

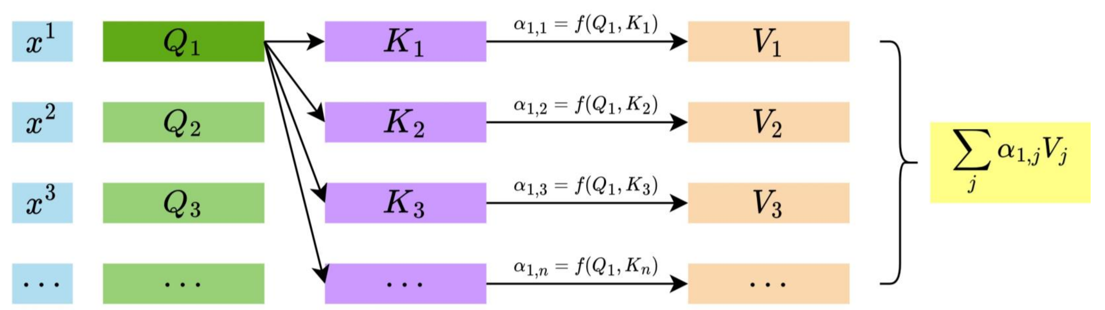
Recognizing the crucial role of Weigh-in-Motion (WIM) systems in infrastructure maintenance and traffic safety, this study addresses the essential task of accurately forecasting overweight vehicle traffic. Leveraging the unique WIM data, which monitors traffic load impacts on aging surface infrastructure, we employ multi-graph learning to enhance predictive performance. We consider multiple graph views to depict spatio-temporal contextual relationships and heterogeneous environmental features surrounding each WIM station, including demographics, land use, road geometry, and infrastructure characteristics. We propose an innovative, end-to-end multi-view transformer framework termed MSinT, which effectively learns the associations of multiple graphs for overweight vehicle prediction. This is achieved by integrating graph convolution networks with a multi-view fusion module. Through extensive experimentation with vehicle records from eight WIM stations, we demonstrate improvement in prediction accuracy compared to state-of-the-art methods. Additionally, our experiment validates the efficiency of our multi-view fusion module and quantifies the contribution of each contextual graph in overweight vehicle prediction.

Framework

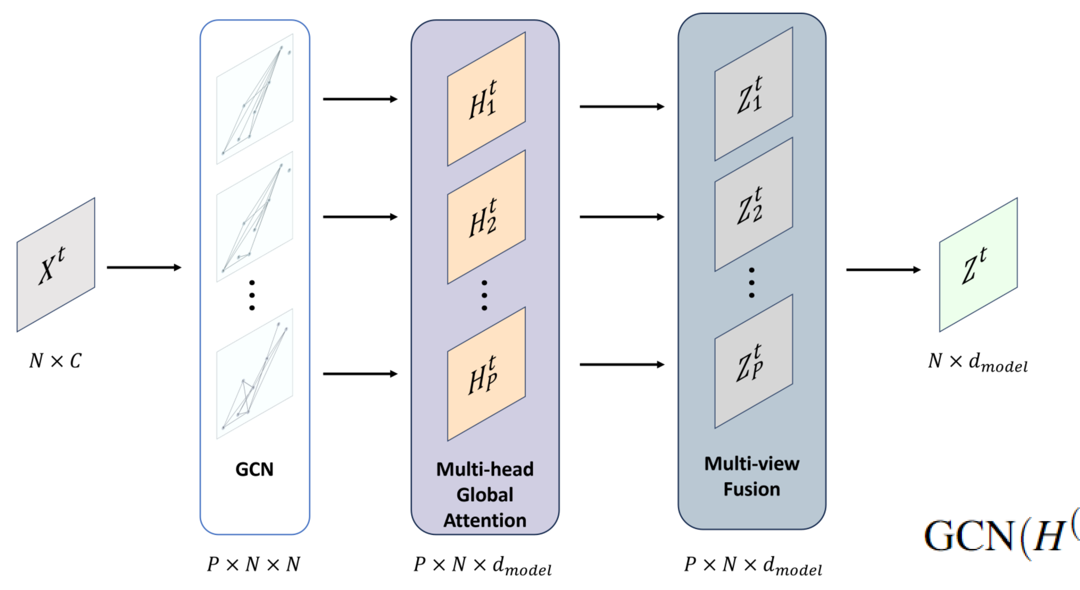


Methodology and Contribution

1. *Transformer as backbone*



2. *Multi-view GCN and fusion blocks*



Multi-view Fusion

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)})$$

$$\mathbf{S} = \mathrm{softmax}\left(\frac{H^{(l)}H^{(l)\top}}{\sqrt{d_{\mathrm{model}}}}\right) \in \mathbb{R}^{N\times N}$$

$$Z^t = \odot(Z_1^t, \cdots, Z_P^t)W^f$$

$$Z_i^t = \mathrm{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}; \mathrm{GCN}(X^t, \tilde{A}_i))$$

$$W^f = \mathrm{softmax}(\Theta)$$

$$\mathrm{GCN}(H^{(l)}, \tilde{A}) = \sigma\left(\left((\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}) \odot \mathbf{S}\right)H^{(l)}W^{(l)}\right)$$

Data collection: We collected data from 8 WIM stations located in New Jersey, covering the period from January 2017 to April 2017.

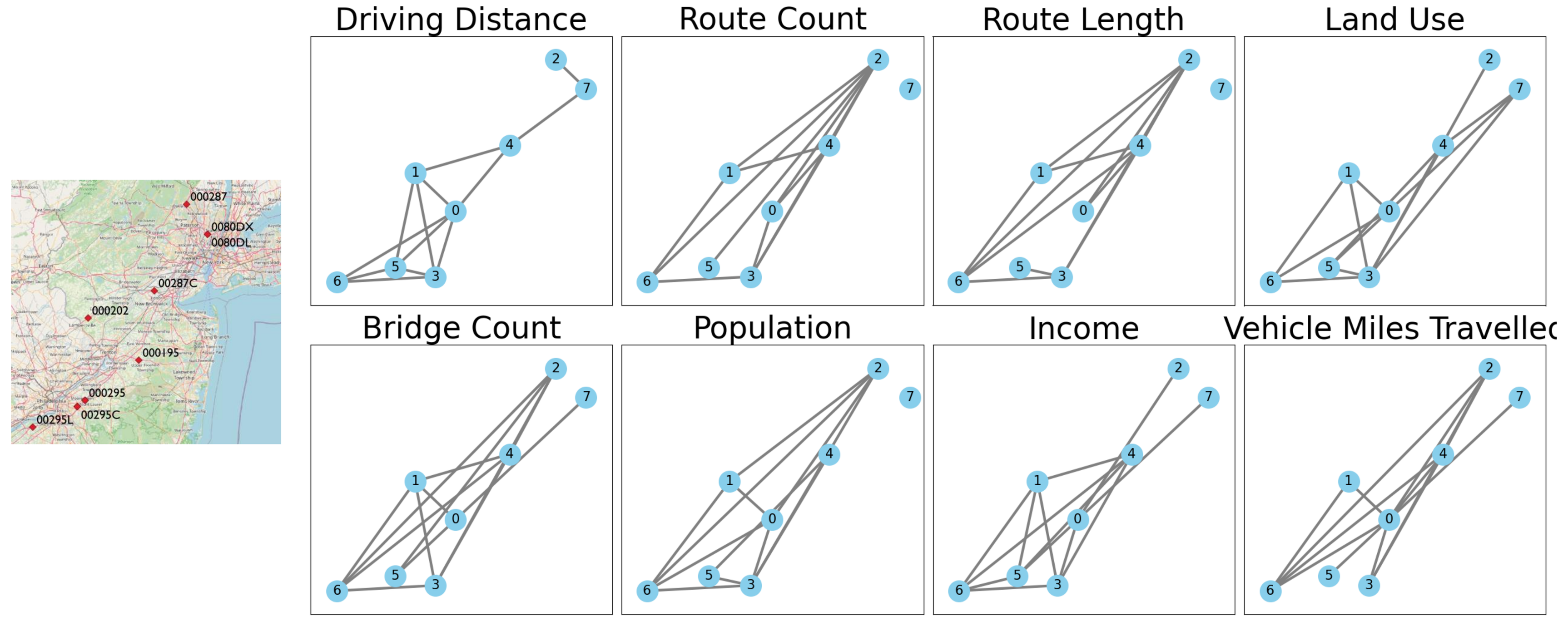
Contribution 1: Our model tackles the spatio-temporal dependencies in WIM data and incorporates the modularized spatial kernel of GCNs with the temporal representation.

Contribution 2: We design a multi-view fusion kernel to capture the distinctive spatial contextual variations surrounding each WIM station to provide better understanding of the spatial dependencies.

Contribution 3: Our model leverages and extends the capabilities of the self-attention mechanism to handle rare events and long temporal dependencies in heterogeneous WIM data.

Experiments, Results and Takeaways

1. Data Processing



2. Performance Comparison of Different Approaches for Total Excess Weight Forecasting.

Transformer-based methods outperform other baselines, which indicates the ability of the self-attention mechanism for better temporal correlation understanding. MSinT achieves the best performance with the most metrics for all forecasting horizons.

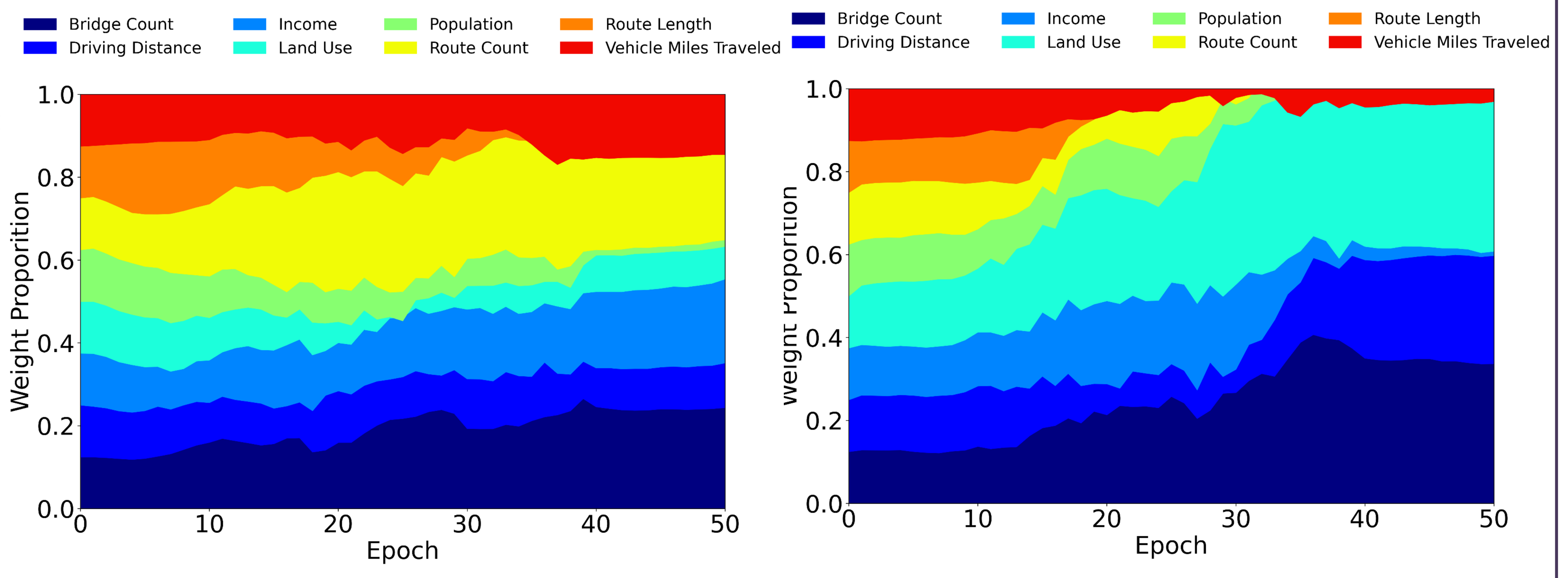
Performance Comparison

T	Metric[a]	GraphWaveNet	MSTGCN	ASTGCN	ASTGNN	*MSinT*
30 min	MAE	24.98	18.39	16.31	16.02	**15.93**
	MAPE(%)	>1,000	>1,000	>1,000	**244.72**	245.82
	RMSE	95.25	63.00	64.43	63.84	**62.94**
1 hour	MAE	25.00	18.91	16.62	**16.39**	16.48
	MAPE(%)	>1,000	603.75	344.73	308.23	**285.02**
	RMSE	95.30	69.38	65.97	65.38	**64.75**
2 hour	MAE	25.02	19.90	17.41	17.17	**17.13**
	MAPE(%)	>1,000	652.32	>1,000	382.59	**377.70**
	RMSE	95.36	**67.12**	68.75	68.65	68.12

[a]MAE and RMSE values are scaled down by a factor of 10^3.

3. Multi-view Interpretation

Evolution of fusion weights in the last encoder and decoder layers during the training process. The input sequence length and prediction horizon are both set to 30. The change of the colored areas represents the weight update of each view after every epoch. At one epoch, a bigger area means a larger weight



(a) Encoder layers (b) Decoder layers

4. Performance comparison for MSinT and two view-specific variants MSinT(I) and MSinT(S).

The total of eight views can be grouped into two categories by their attributes: infrastructure-based views and society-based views. Incorporating a better understanding of the social context in addition to infrastructure-based views may further enhance the prediction of overweighted vehicles.

Horizon	30 min			1 hour			2 hour		
Metric[a]	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
MSinT	**15.93**	245.82	**62.94**	16.77	**261.09**	65.59	**17.13**	397.70	**68.12**
MSinT(I)	16.08	239.35	63.83	**16.52**	311.27	**64.68**	17.21	**346.58**	68.14
MSinT(S)	16.02	**231.59**	63.44	16.62	280.61	68.43	17.30	399.24	68.56

[a]MAE and RMSE values are scaled down by a factor of 10^3, and MAPE is expressed as percentages.

5. Takeaways for Overweighted Truck Forecasting

- MSinT exhibit improvements in prediction accuracy, as compared to other state-of-art methods. Further, our ablation study authenticated the efficiency of the multi-view graph learning module and quantified the individual contribution of each graph view.
- This work sets a robust foundation for future efforts in infrastructure maintenance, traffic management, and road safety. There are some limitations such as the selection of additional views relies on subjective human knowledge and the computational complexity scales roughly linearly with the number of views
- One potential direction of future work is to perform large-scale experiments with a more extensive WIM dataset to validate the view selection process. Additionally, enhancing computational efficiency can be achieved by developing a unified early fusion scheme for the multi-view fusion module.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com

- $H = \sigma(AGG(\psi_{s_1}\Gamma_{s_1}\psi_{s_1}^{-1}, \psi_{s_2}\Gamma_{s_2}\psi_{s_2}^{-1}, \ldots, \psi_{s_e}\Gamma_{s_e}\psi_{s_e}^{-1}))$