

MG-STNET: A Novel Multi-Graph Spatio-Temporal Network for Predicting Traffic Accident Risk[☆]

Guojian Zou^{a,b,c}, Zhiyong Zhou^c, Robert Weibel^c, Ye Li^{a,b}, Ting Wang^{a,b}, Zongshi Liu^{a,b}, Weiping Ding^d and Cheng Fu^{c,*}

^aThe Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, PR China

^bCollege of Transportation Engineering, Tongji University, Shanghai 201804, PR China

^cDepartment of Geography, University of Zurich, Zurich 8057, Switzerland

^dSchool of Artificial Intelligence and Computer Science, Nantong 226019, China

ABSTRACT

All codes and model parameters are available at <https://github.com/zouguojian/Traffic-accident-prediction/tree/main/MG-STNET>.

1. Related work

The related works can be highlighted in two respects: spatio-temporal traffic prediction with deep learning and traffic accident risk forecasting.

1.1. Spatial-temporal traffic prediction with deep learning

Traffic prediction can be described as a spatio-temporal forecasting problem, meaning both spatial and temporal dependencies are critical properties in the modeling process (Liang, Tang, Gao, Wang and Huang, 2023; Zheng, Chai, Duanmu and Katos, 2023). Deep learning has been proven as an effective technique for extracting spatio-temporal correlations compared to statistical and traditional machine learning approaches (Wang, Ren and Li, 2023). Recurrent neural networks (RNNs) are employed to capture temporal correlations. For example, Qu, Lyu, Li, Ma and Fan (2021) introduced Features Injected Recurrent Neural Networks, utilizing stacked RNNs to model series features. Ma, Dai and Zhou (2021) proposed the LSTM_BILSTM model, which combines the advantages of sequential data and the long-term dependencies of both forward and reverse long short-term memory networks (LSTMs), integrated into a traffic flow prediction model. To efficiently capture long-term dependencies, temporal convolutional networks (TCNs) and temporal attention mechanisms are utilized (Lea, Flynn, Vidal, Reiter and Hager, 2017; Shao, Zhang, Wang and Xu, 2022). For example, Ren, Li and Liu (2023) designed parallel TCN components based on dilated causal convolution to obtain improved global features.

Spatial dependency is another critical aspect that garners researchers' attention, with a substantial body of work focusing on effective modeling of spatial correlations and their integration with temporal features (Zou, Lai, Wang, Liu, Bao, Ma, Li and Fan, 2024a; Zou, Lai, Wang, Liu and Li, 2024b), a consideration contributing to our traffic risk forecasting task (Huang, Zhang, Dai and Bo, 2019; Trirat and Lee, 2021). For instance, in the prediction of short-term traffic speed, Lu, Rui and Ran (2020) introduced a Mixed Deep Learning model, combining Convolutional Long Short-Term Memory layers, convolutional layers, and a dense layer within a sequence-to-sequence framework to extract spatio-temporal features integrally. Zou, Lai, Ma, Li and Wang (2023) proposed the Spatio-Temporal Generative Inference Network, leveraging the Spatio-Temporal Correlation Extraction Block to capture complex spatio-temporal dependencies in traffic data. In early research on DL models for traffic accident prediction, some scholars explored the use of position information and time period to describe similar traffic patterns across locations (Ren, Song, Wang, Hu and Lei, 2018). Others employed CNNs to extract the static spatial features of each district (Chen, Fan, Zheng, Xiao, Cheng and Wang, 2018).

[☆] 2010768@tongji.edu.cn (G. Zou); zhiyong.zhou@geo.uzh.ch (Z. Zhou); robert.weibel@geo.uzh.ch (R. Weibel); JamesLI@tongji.edu.cn (Y. Li); 2110763@tongji.edu.cn (T. Wang); chuchuoliu@tongji.edu.cn (Z. Liu); dwp9988@163.com (W. Ding); cheng.fu@geo.uzh.ch (C. Fu)

ORCID(s): 0000-0002-6490-1583 (G. Zou)

Specifically, many works on traffic prediction employ various deep learning frameworks to extract spatio-temporal features and achieve satisfactory performance. However, these intensive and continuous sequence forecasting methods are not directly applicable to traffic accident risk prediction due to the sporadic spatio-temporal distribution and influence of exogenous factors (Trirat, Yoon and Lee, 2023). Accident risk forecasting is a subset of traffic prediction, allowing us to draw insights from the latter. The successful applications of deep learning in traffic prediction underscore its potential in addressing regression problems, offering valuable inspiration and experience to improve the performance of traffic accident risk forecasting.

2. Experiments

2.1. Data description

We experimented on two publicly accessible real-world datasets collected from NYC and Chicago (Wang, Lin, Guo and Wan, 2021). Detailed statistics of both datasets are shown in Table 1. (1) **NYC** This dataset includes five sources of data: traffic accidents, Points of Interest (POI), taxi trips, meteorological features, and road segments. (2) **Chicago** Distinct from NYC, it comprises only four data sources due to the absence of POI information. Specifically, the traffic accident data comprise location, time, and casualty counts. The POI data encompass seven categories: residential areas, schools, social services, transportation, cultural facilities, entertainment, and commerce. Taxi trip records consist of pick-up and drop-off times, longitude, and latitude, facilitating the calculation of inflow to and outflow from each region, serving as human mobility data. Meteorological data comprise temperature and air condition records. Road segment data encompass types, length, width, and snow removal priority of road segments. Additionally, the temporal granularity is hourly. In the experiment, 60% of the data are used as the training set, 20% of the data are used as the validation set, and the remaining 20% are considered as the test set.

Table 1

Statistics of the two datasets.

Dataset	NYC	Chicago
Time span	01/01/2013 - 12/31/2013	02/01/2016 - 09/30/2016
Time granularity	1h	1h
Traffic accidents	147k	44k
Taxi orders	173,179k	1,744k
POI	15,625	None
Road segments	103k	56k
Meteorological features	8,760	5,832

2.2. Experimental setup

In the experiments, the entire city is partitioned into 20×20 grid cells, with each cell spanning $2km \times 2km$. Owing to the absence of road networks in certain areas, the NYC dataset contains 243 valid regions, whereas the Chicago dataset comprises 197 valid regions.

In the training stage, we set the maximum epochs to 200, with a batch size of 32 and a learning rate of 0.0005. The Adam optimizer is utilized for weight updates. To align with previous research, we configure the target time slots (Q) as 1, short-term time slots (O) as 4, and long-term time slots (P) as 4. Especially in evaluating the model performance on the validation set after each epoch, the model's weights are saved if the loss is decreased. An early stopping mechanism, named patience, with a set value of 10, is employed to prevent overfitting during training. The training process halts early if the loss remains unchanged for 10 consecutive epochs. Comprehensive model hyperparameters are outlined in Table 2. Following the previous works, in this paper, we maintain the baselines' hyperparameters for the NYC and Chicago datasets. The MG-STNET and baselines are implemented using PyTorch. Model training, validation, and testing utilized one NVIDIA Tesla V100S-PCIE-32GB GPU and 24 CPU cores on the server. The implementation codes for the proposed MG-STNET model and baseline methods are openly accessible on the personal GitHub homepage¹.

¹<https://github.com/zouguojian/Traffic-accident-prediction/tree/main/MG-STNET>

Table 2
Model hyperparameters.

Component	Layer	Hyperparameter	Values (NYC Chicago)	Output dimension
Embedding/Feed/Residual	2D CNNs	Channel Fitter size Number of layers	64/64 $1 \times 1 / 1 \times 1$ 2	[32, 8, 20, 20, 64]
MGraph-STModule	MGSNet	Feed M_{\star}^{left} M_{\star}^{right}	1 $(243 197) \times 100$ $100 \times (243 197)$	[32, 20, 20, 64]
		Bidirectional walks	Orders 2	
	Temporal Blocks	Channel Fitter size Number of blocks Dilation size Padding	64/64/64 $2 \times 2 / 2 \times 2 / 2 \times 2$ 3 0/2/4 1/1/1	
		Channel Fitter size Number of layers Padding Residual	64/64 $3 \times 3 / 3 \times 3$ 2 1×1 1	
		Channel Fitter size Number of blocks Dilation size Padding	64/64/64 $2 \times 2 / 2 \times 2 / 2 \times 2$ 3 0/2/4 1/1/1	
Adaptive channel fusion gate	2D CNNs	Channel Fitter size Number of layers	64 1×1 4	[32, 20, 20, 64]
FC	2D CNNs	Channel Fitter size Number of layers	64/1 $1 \times 1 / 1 \times 1$ 2	[32, 20, 20, 1]
-	-	Batch size	32	-
-	-	\mathcal{L}	10	-
-	-	Dropout	0.0	-
-	-	Decay rate	0.9	-
-	-	Epoch	200	-
-	-	Learning rate	0.00001	-
-	-	λ	0.0001	-
-	-	Patience	10	-
-	-	Optimize method	Adam	-
-	-	Activation function	(ReLU, Sigmoid)	-

2.3. Evaluation metrics

Six metrics are utilized to evaluate the traffic accident risk prediction model's performance (Wang et al., 2021). Accident risk prediction is a regression task, and the root mean square error (RMSE), defined in Equation (1), was used. Inspired by the previous studies, the risk forecasting task can also be deemed as a classification issue. Thus, Recall and mean average precision (MAP) were used to measure the prediction accuracy with high-traffic accident regions, which are defined in Equations (2) and (3). Note that a lower RMSE score indicates more accurate predicted risks in the highest traffic accident regions, while higher Recall and MAP scores mean the model performs better.

$$\text{RMSE} = \sqrt{\frac{1}{D} \sum_{i=1}^D (Y_i - \hat{Y}_i)^2} \quad (1)$$

$$\text{Recall} = \frac{1}{D} \sum_{i=1}^D \frac{|S_i \cap R_i|}{|R_i|} \quad (2)$$

$$\text{MAP} = \frac{1}{D} \sum_{i=1}^D \frac{\sum_{j=1}^{|R_i|} \text{pre}(j) \times \text{rel}(j)}{|R_i|} \quad (3)$$

In Equations (1)-(3), D is the number of test samples, Y_i represents the observed values, and \hat{Y}_i denotes the predicted values for the highest traffic accident regions in sample i . R_i signifies the ground truth, indicating where the traffic accident risks occur within the top L regions, while S_i denotes the set of regions where the predicted risks fall within the top L regions. The term $\text{pre}(j)$ denotes the precision of a cut-off rank list ranging from 1 to j , and $\text{rel}(j)$ signifies the recall value of region j . If traffic accidents occur in this region, $\text{rel}(j)$ equals 1; otherwise, it is 0. Additionally, L is set to 50, indicating that the top-50 regions are considered as the highest traffic accident regions.

GPU memory usage during training (GPU-MUT) and GPU memory usage during inference (GPU-MUI) are utilized as two critical metrics to evaluate the model's computational cost for deep learning approaches. Model parameters are used to assess the scale of the prediction model.

2.4. Experimental results

2.4.1. Influence of MG-STNET components

To assess the impact of individual components within the MG-STNET framework on traffic accident risk forecasting, seven variants were compared against the MG-STNET (Table 3),

- **W/O Z-Score**, which substitutes the Z-Score normalization with the min-max normalization method in data preprocessing.
- **W/O Bidirectional similarity matrices**, similar to GSNet, utilizing single-direction similarity graphs for semantic spatial dependencies modeling.
- **W/O Adaptive graphs**, eliminating the adaptive correlation graphs while retaining bidirectional similarity graphs.
- **W/O Temporal Blocks**, akin to GSNet, deploying GRU and attention for extracting temporal correlations.
- **W/O Adaptive channel fusion gate**, employing an addition operation for combining semantic and geographical spatio-temporal features.
- **W/O MGraph-STModule**, solely utilizing the Geographical Spatio-Temporal Module to extract geographical spatio-temporal features.
- **W/O Geo-STModule**, exclusively utilizing the Multi-Graph Spatio-Temporal Module to extract semantic spatio-temporal features.

W/O Z-Score In previous traffic accident risk prediction methods, the min-max method normalizes input variables to ensure each feature contributes uniformly to the learning process, preventing any specific feature from dominating due to its scale and expediting model convergence. However, the performance of *w/o Z-Score* is inferior to MG-STNET on the NYC and Chicago datasets, especially for the regression task. For instance, *w/o Z-Score* increased RMSE by 3.541%, decreased Recall by 2.186%, and improved MAP 0.103%, compared to MG-STNET on the NYC dataset; increased RMSE by 20.348%, decreased Recall and MAP by 5.651% and 20.203% on the Chicago dataset. Experimental results are influenced by the min and max values originating from the training dataset, making it impossible to characterize data distributions in the validation and test datasets, especially if the three datasets have strong biases. These findings suggest that the Z-Score normalization may be more suitable than the min-max method for accident risk prediction.

W/O Bidirectional Similarity Matrices This variant excludes the transpose similarity matrices from the MG-STNET model, resulting in deteriorated performance. For instance, *w/o bidirectional similarity matrices* exhibits

Table 3

The contribution of each component for the MG-STNET.

Dataset	Model	All day			High-frequency accident periods		
		RMSE	Recall	MAP	RMSE	Recall	MAP
NYC	W/o Z-Score	7.3812	34.00%	0.1935	6.5207	34.46%	0.1807
	W/o Bidirectional similarity matrices	7.1304	34.05%	0.1895	6.3965	34.46%	0.1794
	W/o Adaptive graphs	7.1778	34.38%	0.1887	6.4402	34.18%	0.1776
	W/o Temporal Blocks	6.9931	33.70%	0.1886	6.2828	33.69%	0.1795
	W/o Adaptive channel fusion gate	7.0976	34.31%	0.1907	6.3971	34.32%	0.1824
	W/o Geo-STModule	7.0019	33.72%	0.1863	6.2679	34.46%	0.1801
	W/o MGraph-STModule	7.0548	34.31%	0.1952	6.3236	34.11%	0.1860
	MG-STNET	7.1288	34.76%	0.1933	6.4459	35.58%	0.1869
	Gains	-1.940%	1.105%	-0.973%	-2.840%	3.250%	0.484%
Chicago	W/o Z-Score	10.9751	20.87%	0.0786	8.3253	22.09%	0.1013
	W/o Bidirectional similarity matrices	9.3907	21.65%	0.0904	6.7975	23.32%	0.1209
	W/o Adaptive graphs	9.1330	21.71%	0.0930	6.5649	22.50%	0.1132
	W/o Temporal Blocks	9.2540	19.62%	0.0883	6.8220	19.62%	0.0904
	W/o Adaptive channel fusion gate	9.3008	21.05%	0.0877	6.6804	22.50%	0.1137
	W/o Geo-STModule	9.3724	21.53%	0.0933	6.6925	23.32%	0.1169
	W/o MGraph-STModule	9.5245	18.96%	0.0890	6.9606	18.93%	0.0958
	MG-STNET	9.1195	22.12%	0.0985	6.4482	23.18%	0.1206
	Gains	0.148%	1.889%	5.573%	1.778%	-0.600%	-0.248%

higher RMSE values than MG-STNET by 0.022%, and is inferior to MG-STNET by 2.043% and 1.966% in terms of Recall and MAP, respectively, on the NYC dataset; it also has higher values by 2.974%, 2.125%, and 8.223% on the Chicago dataset. The experimental results emphasize the significance of bidirectional relationships in an incomplete symmetry problem. This can complement the lost critical information during the conversion of the symmetric subject into an incomplete symmetry problem, enhancing traffic accident risk forecasting performance.

W/O Adaptive Graphs In the experiment, removing our proposed adaptive graphs from MG-STNET resulted in decreased performance across all evaluation metrics. For example, *w/o adaptive graphs* is worse than MG-STNET by 0.687%, 1.093%, and 2.380% in terms of RMSE, Recall, and MAP, respectively, on the NYC dataset; and by 0.148%, 1.854%, and 5.584% on the Chicago dataset. These results suggest that adaptive graphs address the limitations of predefined similarity graphs from prior knowledge, thus improving the semantic spatial representations and providing insights that may be advantageous for other traffic prediction tasks. Additionally, the variant without adaptive graphs still outperforms GSNet on the NYC and Chicago datasets across all metrics, as illustrated in Fig. 1, underscoring the significance of other improved components within MG-STNET.

W/O Temporal Blocks Notably, the replacement of Temporal Blocks with GRU and attention mechanism did not yield superior performance compared to MG-STNET. For instance, *w/o Temporal Blocks*, led to an improvement in RMSE by 1.904% and decreases in Recall and MAP by 3.049% and 2.431%, respectively, on the NYC dataset; and by 1.474%, 11.302%, and 10.355% on the Chicago dataset when compared to MG-STNET. These results highlight two key points: (1) the characteristics of traffic accident risk underscore the importance of discerning the heterogeneity between daily similarity and short-term traffic patterns, both of which hold significance for risk prediction; (2) RNNs demonstrate adaptability to continuous homogeneous sequences, their effectiveness in handling heterogeneous series, however, is inferior to that of Temporal Blocks.

W/O Adaptive Channel Fusion Gate To validate that the contributions of semantic and geographical spatio-temporal features to traffic accident risk prediction are distinct, the adaptive channel fusion gate was replaced by an addition operation, leading to decreased performance. As presented in Table 3, *w/o adaptive channel fusion gate*, there was a decrease in RMSE by 0.438% and reductions in Recall and MAP by 1.295% and 1.345%, respectively, in comparison to MG-STNET on the NYC dataset; and likewise, by 1.988%, 4.837%, and 10.964% on the Chicago dataset. The experimental results highlight the importance of effectively combining semantic and geographical spatio-temporal features. Evidently, employing an adaptive channel fusion gate to assign adaptive weights to these features based on diverse inputs is a significant approach.

W/O Geo-STModule & MGraph-STModule The assessment of the significance of semantic and geographical

Pattern Recognition

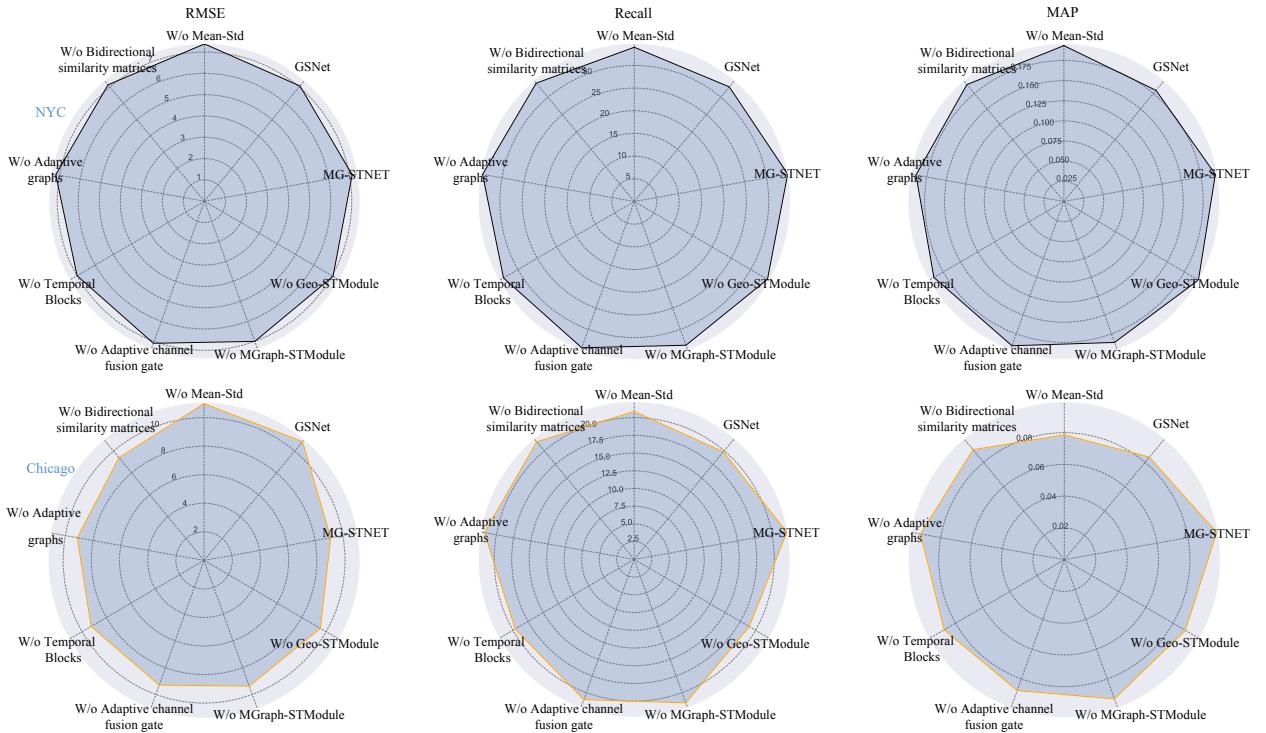


Figure 1: Spider charts of RMSE, Recall, and MAP for GSNet, MG-STNET, and the seven variants, on the two datasets: the first row presents the prediction performance on the NYC dataset; the second row reflects the prediction performance on the Chicago dataset.

spatio-temporal features in traffic accident risk prediction involved the individual removal of the Multi-Graph Spatio-Temporal Module (MGraph-STModule) and Geographical Spatio-Temporal Module (Geo-STModule). Initially, compared to MG-STNET, both variants exhibited decreased performance across all metrics (RMSE, Recall, and MAP). These outcomes affirm the conclusion that both semantic and geographical spatio-temporal features hold critical importance for risk forecasting in GSNet. Subsequently, the *w/o Geo-STModule* and *w/o MGraph-STModule* displayed superior performance compared to GSNet on both the NYC and Chicago datasets, as illustrated in Fig. 1 and Table 3. This signifies the enhanced efficiency of the MGraph-STModule and Geo-STModule over the ST-Sem Module and ST-Geo Module in GSNet. Furthermore, an intriguing observation emerged where MGraph-STModule outperformed the Geo-STModule on the Chicago dataset but was worse on the NYC dataset. These experimental results suggest that the difference in the traffic environment of the two cities, such as the traffic accidents and POIs, may have caused differences in the significance that semantic and geographical spatio-temporal characteristics play in the accident risk prediction task.

2.4.2. Case visualization

The performance of various models on the NYC and Chicago datasets has been presented, showcasing metrics such as RMSE, Recall, and MAP in Table ???. Analyses and scientific issues have been elucidated in Sections ??, ??, and 2.4.1. Practically, a more comprehensive discussion of detailed prediction results is required for managers to formulate effective management strategies. This information also serves to alert travelers, encouraging cautious driving behavior and aiding in route planning guidance. For instance, weekly traffic risk forecasting reflects overall performance, effectively evaluating its application value from a macro perspective. Daily traffic prediction assesses real-time accuracy through regression and classification from a micro perspective. This approach fosters an understanding of traffic accident distribution across various regions in the transportation network, enabling a comprehensive analysis of temporal and spatial dynamics in traffic accident forecasting. Hence, samples from one week were randomly chosen from the test dataset and visualized to validate their application value, offering insights for managerial decision-making. Two

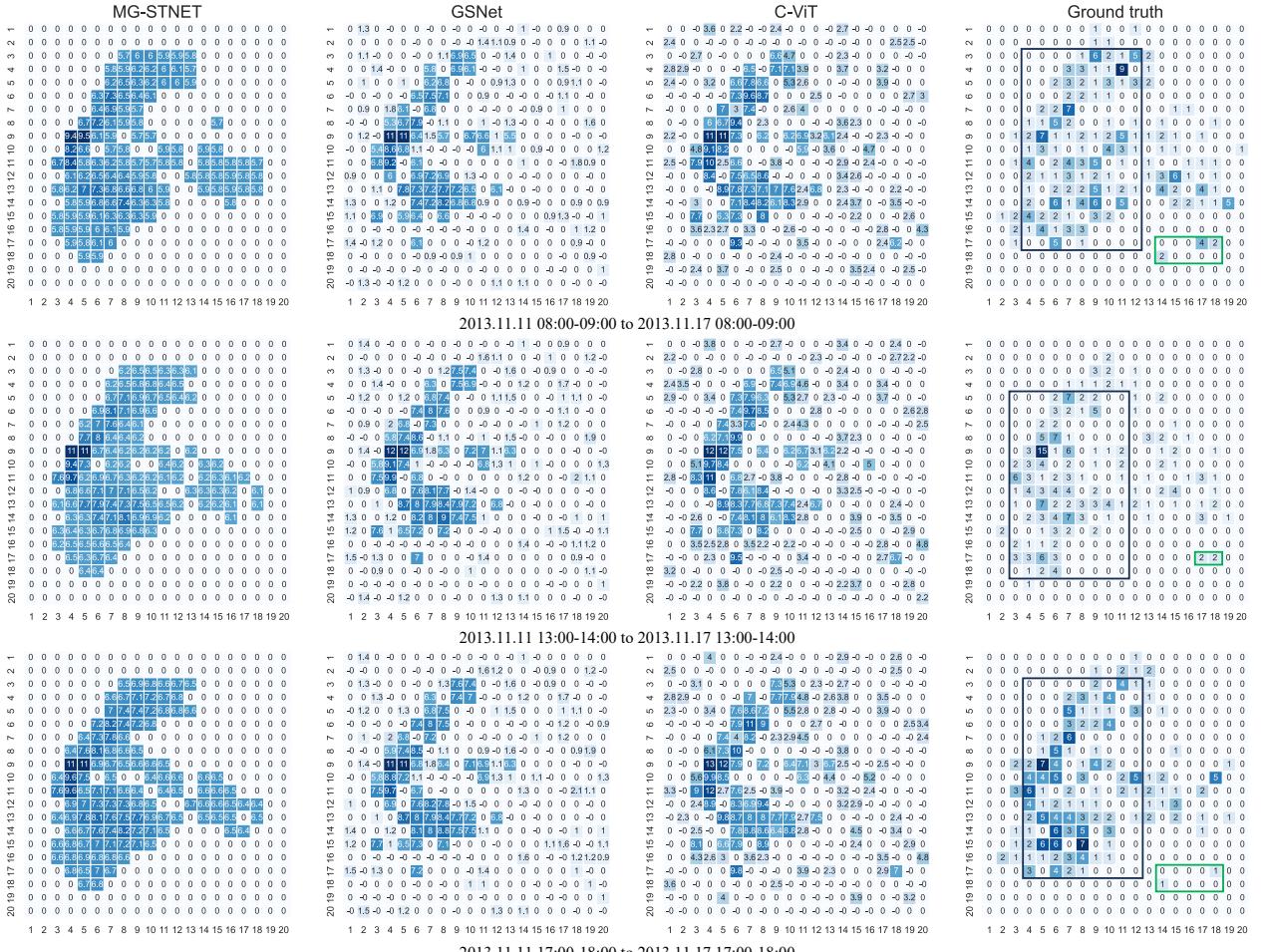


Figure 2: The visualization of traffic accident risk prediction for the same time in a week on the NYC dataset includes three selected periods: high-risk (i.e., 8:00-9:00 and 17:00-18:00) and general (i.e., 13:00-14:00) periods.

particular implications for traffic accident risk prediction and management are identified below.

Firstly, the blue blocks in MG-STNET align well with the ground truth in contrast to baselines, depicted in Fig. 2. Nevertheless, the scores assigned to high-risk regions within the black box are lower than those in GSNet and C-ViT on the NYC dataset, and vice versa in Chicago (as illustrated in Fig. 4). The visualizations align with the experimental results in Table ??, illustrating that MG-STNET holds more advantages in traffic accident classification, that is, classifying whether accidents occur or not. Despite a slightly inferior performance in terms of RMSE compared to C-ViT, it does not compromise its position in the traffic accident risk prediction task. It is noteworthy that regions with low scores are concealed, resulting in a few low real-risk regions in the green box that are overlooked by MG-STNET, as depicted in Fig. 2, but this phenomenon is not observed in the Chicago dataset. Looking back to the yearly traffic accident risk distributions shown in Fig. ?? (a), it becomes evident that this outcome is primarily attributed to the unbalanced distribution of traffic accidents. This imbalance poses a challenge in identifying accident patterns in low-risk regions due to insufficient accident samples during training. However, given that high-risk regions are more crucial for transportation management, the proposed MG-STNET is still suitable for predicting areas that transportation governors should prioritize.

Secondly, MG-STNET exhibits satisfactory performance in predicting the next time slot, as depicted in Fig. 3 and 5. For instance, MG-STNET demonstrates comparable prediction results to GSNet and C-ViT in the NYC dataset, but as shown in Fig. 5, it outperforms them in the Chicago dataset. Notably, MG-STNET accurately identifies traffic accidents

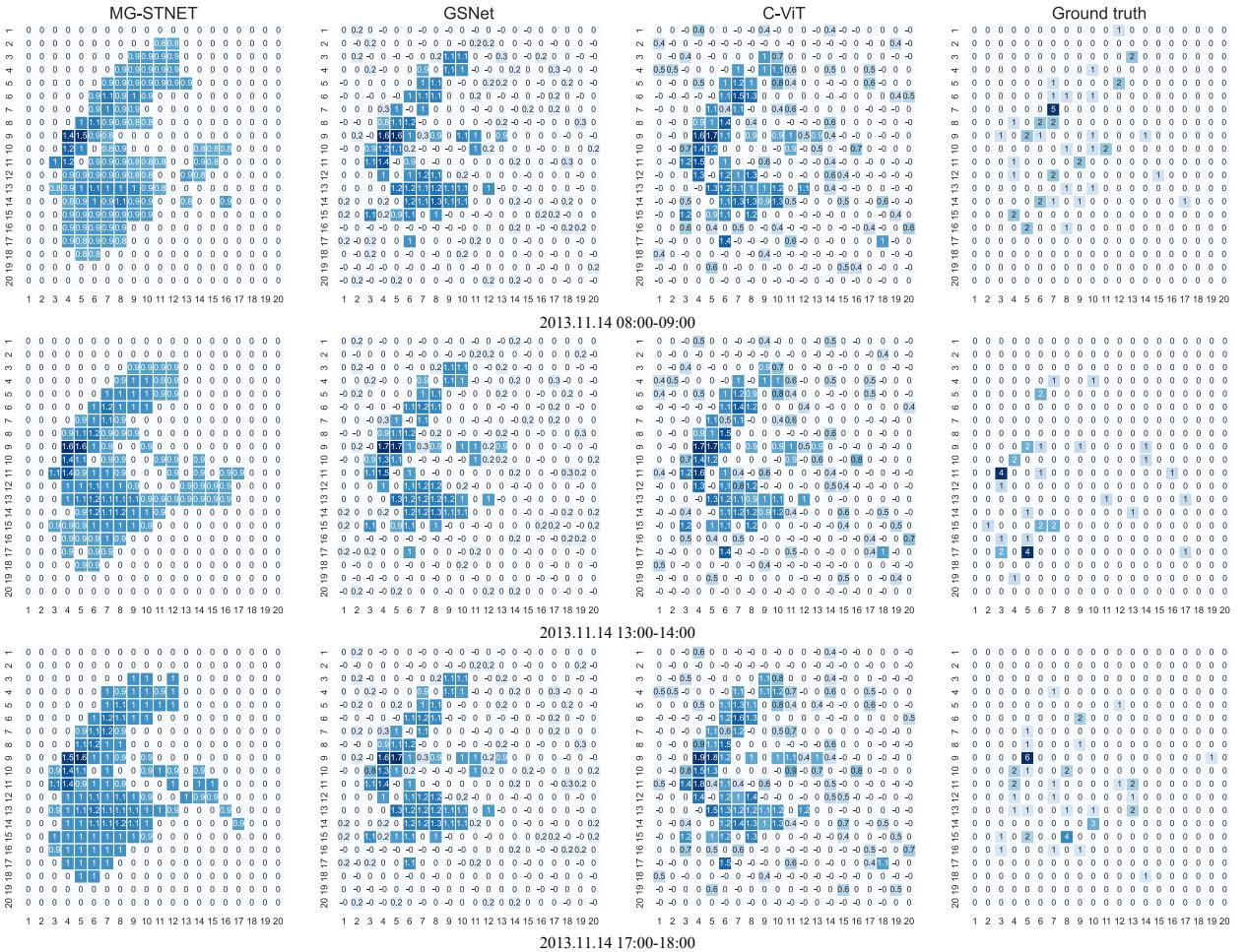


Figure 3: Examples of traffic accident risk prediction visualization on Thursdays in the NYC dataset. Three periods are selected: high-risk (i.e., 8:00-9:00 and 17:00-18:00) and general (i.e., 13:00-14:00) periods.

within the red box (i.e., Fig. 5) and assigns low-risk scores when no accidents occur. These comparisons illustrate that MG-STNET performs competitively in predicting traffic accident risk, offering a more reliable ability to recognize accidents. These advantages are crucial for practical applications. Furthermore, based on real-time prediction results, managers could implement traffic control strategies and prevention plans dynamically before traffic accidents occur.

Acknowledgement

This work was supported in part by the scholarship of the China Scholarship Council under Grant 202306260111, in part by the Project of the National Key Research and Development Program of China under Grant 2018YFB1601301, in part by the Project of the National Key Research and Development Program of China under Grant 2023YFC3305802, in part by the National Natural Science Foundation of China under Grant 71961137006.

References

- Chen, C., Fan, X., Zheng, C., Xiao, L., Cheng, M., Wang, C., 2018. Sdcae: Stack denoising convolutional autoencoder model for accident risk prediction via traffic big data, in: 2018 sixth international conference on advanced cloud and big data (CBD), IEEE. pp. 328–333.
- Huang, C., Zhang, C., Dai, P., Bo, L., 2019. Deep dynamic fusion network for traffic accident forecasting, in: Proceedings of the 28th ACM international conference on information and knowledge management, pp. 2673–2681.

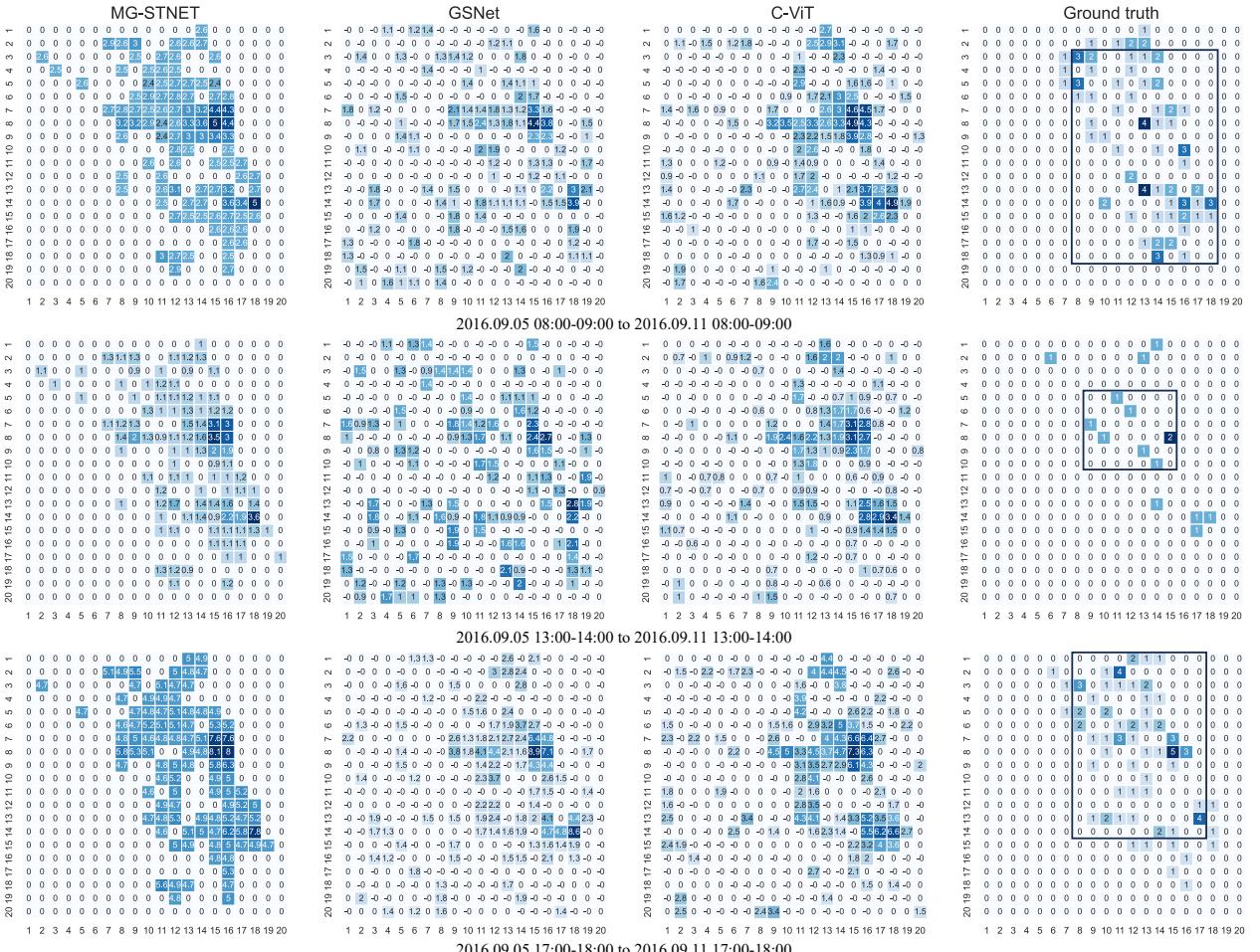


Figure 4: The visualization of traffic accident risk prediction for the same time in a week on the Chicago dataset.

- Lea, C., Flynn, M.D., Vidal, R., Reiter, A., Hager, G.D., 2017. Temporal convolutional networks for action segmentation and detection, in: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 156–165.
- Liang, J., Tang, J., Gao, F., Wang, Z., Huang, H., 2023. On region-level travel demand forecasting using multi-task adaptive graph attention network. *Information Sciences* 622, 161–177.
- Lu, W., Rui, Y., Ran, B., 2020. Lane-level traffic speed forecasting: a novel mixed deep learning model. *IEEE transactions on intelligent transportation systems* 23, 3601–3612.
- Ma, C., Dai, G., Zhou, J., 2021. Short-term traffic flow prediction for urban road sections based on time series analysis and lstm_bilstm method. *IEEE Transactions on Intelligent Transportation Systems* 23, 5615–5624.
- Qu, L., Lyu, J., Li, W., Ma, D., Fan, H., 2021. Features injected recurrent neural networks for short-term traffic speed prediction. *Neurocomputing* 451, 290–304.
- Ren, H., Song, Y., Wang, J., Hu, Y., Lei, J., 2018. A deep learning approach to the citywide traffic accident risk prediction, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 3346–3351.
- Ren, Q., Li, Y., Liu, Y., 2023. Transformer-enhanced periodic temporal convolution network for long short-term traffic flow forecasting. *Expert Systems with Applications* 227, 120203.
- Shao, Z., Zhang, Z., Wang, F., Xu, Y., 2022. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1567–1577.
- Trirat, P., Lee, J.G., 2021. Df-tar: a deep fusion network for citywide traffic accident risk prediction with dangerous driving behavior, in: Proceedings of the Web Conference 2021, pp. 1146–1156.
- Trirat, P., Yoon, S., Lee, J.G., 2023. Mg-tar: Multi-view graph convolutional networks for traffic accident risk prediction. *IEEE Transactions on Intelligent Transportation Systems* 24, 3779–3794.
- Wang, B., Lin, Y., Guo, S., Wan, H., 2021. Gsnet: learning spatial-temporal correlations from geographical and semantic aspects for traffic accident

Pattern Recognition

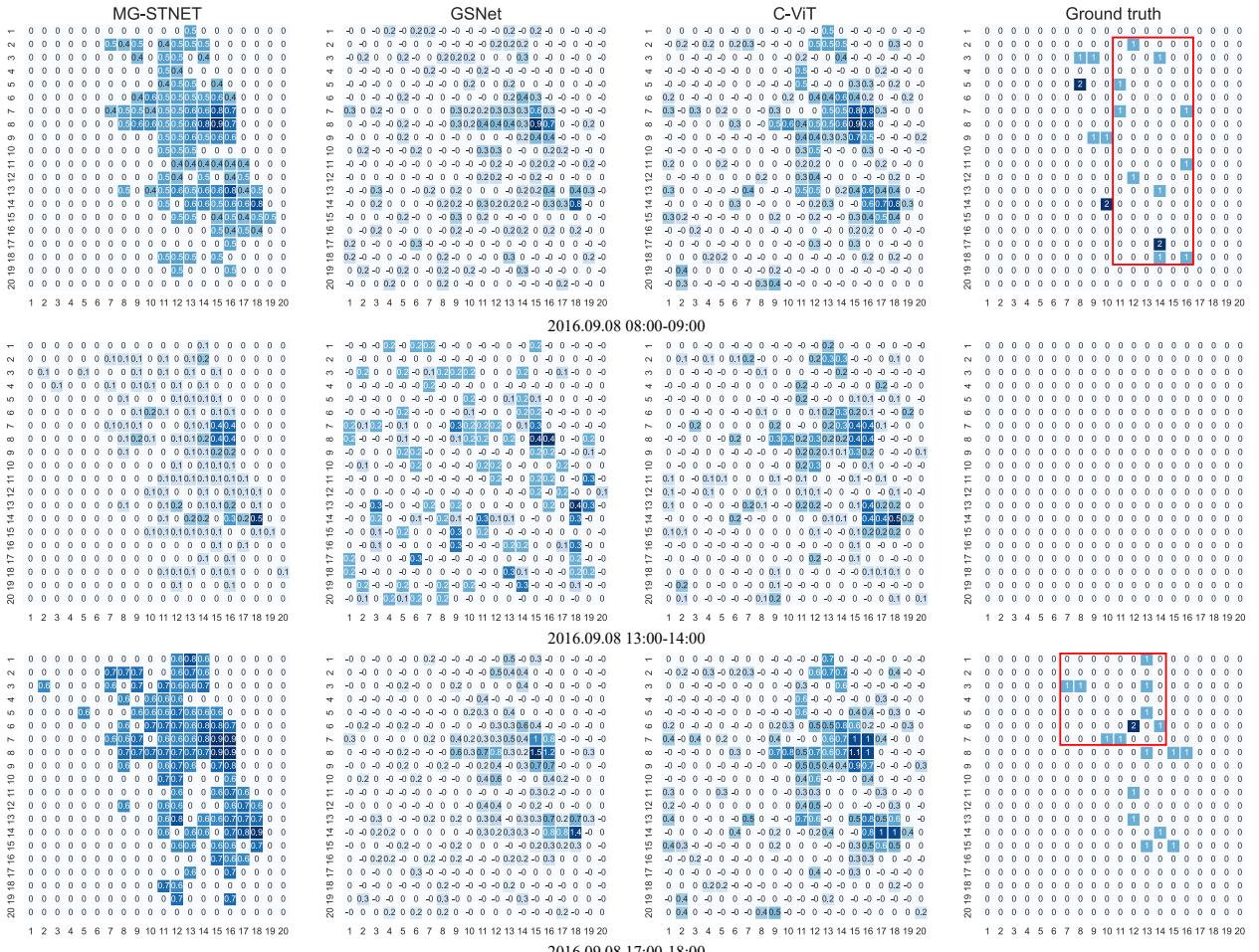


Figure 5: Examples of traffic accident risk prediction visualization on Thursdays in the Chicago dataset.

- risk forecasting, in: Proceedings of the AAAI conference on artificial intelligence, pp. 4402–4409.
- Wang, Y., Ren, Q., Li, J., 2023. Spatial-temporal multi-feature fusion network for long short-term traffic prediction. Expert Systems with Applications 224, 119959.
- Zheng, G., Chai, W.K., Duanmu, J.L., Katos, V., 2023. Hybrid deep learning models for traffic prediction in large-scale road networks. Information Fusion 92, 93–114.
- Zou, G., Lai, Z., Ma, C., Li, Y., Wang, T., 2023. A novel spatio-temporal generative inference network for predicting the long-term highway traffic speed. Transportation research part C: emerging technologies 154, 104263.
- Zou, G., Lai, Z., Wang, T., Liu, Z., Bao, J., Ma, C., Li, Y., Fan, J., 2024a. Multi-task-based spatiotemporal generative inference network: A novel framework for predicting the highway traffic speed. Expert Systems with Applications 237, 121548.
- Zou, G., Lai, Z., Wang, T., Liu, Z., Li, Y., 2024b. Mt-stnet: A novel multi-task spatiotemporal network for highway traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems 25, 8221–8236. doi:10.1109/TITS.2024.3411638.