Modification and Responses to the Reviewers' Comments and Suggestions
Manuscript ESWA-D-25-28448R1

Learning and Predicting Traffic Conflicts in Mixed Traffic: A Spatiotemporal Graph Neural Network with Manifold Similarity Learning

Zongshi Liu, Guojian Zou, Ting Wang, Meiting Tu, Hongwei Wang, Ye Li

Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.

Dear Professor Eklund,

We are very grateful to the Editor and the Reviewers for spending precious time reviewing our work and providing useful comments. We have carefully considered and responded to each of the comments from the reviewers. Below we briefly summarize the major revisions that have been made to improve the quality and clarity of the manuscript:

- Clarified the methodological novelty and positioning of MS-STGNet. We expanded the discussions in Section 2 (including a new summary Table 1) to more clearly distinguish our framework from existing manifold-based traffic models and ST-GNN / STGAT-type graph networks.
- Strengthened validation and data justification. We added new experiments on real-world freeway traffic from the FHWA NGSIM program (Appendix D) and briefly summarized these tests in Section 6.5 to demonstrate transferability from the simulated mixed-traffic corridor to purely human-driven real traffic.

Explanations of what we have changed in response to the reviewers' concerns are given point by point in the following pages. The changes in the revised manuscript have highlighted in blue. We hope these changes will strengthen our manuscript.

Comments from Reviewer #1:

Comment 1:

The core issue lies in the limited novelty of the proposed MS-STGNet framework. The integration of a manifold-similarity prior into a spatiotemporal graph network, while positioned as a key contribution, is conceptually close to existing adaptive adjacency mechanisms in models like STGAT. The authors' response clarifies that their approach initializes the graph from offline-computed manifold distances rather than learning it solely from instantaneous features. However, this distinction appears to be more of a technical variation than a significant conceptual leap. The incremental benefits, as evidenced by the modest performance improvements over strong baselines like STGCN and STGAT, do not sufficiently justify the claim of a major methodological advancement. The framework largely builds upon mature components, and the addition of the manifold-similarity module, while beneficial, does not constitute a transformative step for the field.

Response to Comment 1:

We thank the reviewer for the thoughtful comments on the novelty of MS-STGNet. MS-STGNet is not an isolated design, but a further innovation built on our previously validated multi-graph spatiotemporal framework for accident risk forecasting, which has already been tested on real-world freeway crash data (Zou et al., 2025). In this work, we extend that research line to a more microscopic and complex setting—traffic conflict prediction in mixed CAV–HDV flows—and redesign the spatial modeling component by introducing a manifold-similarity module. Specifically, this manifold-similarity module brings three main innovations:

- It constructs the initial graph from traffic-state manifolds derived from long-term multi-dimensional flow–speed–occupancy trajectories, providing an explicit traffic-state prior for adjacency instead of learning connectivity solely from instantaneous node embeddings as in STGAT-type models.
- It refines this manifold-based graph through learnable weighting and multi-order message passing, so that physically similar but geographically distant regions can effectively exchange information in the mixed CAV–HDV setting.
- By embedding manifold-based state similarity into the spatiotemporal graph, it explicitly encodes the dynamic evolution of traffic states and suppresses spurious conflict activations, which is also confirmed by the strongest performance degradation observed when this module is removed in our ablation study (See Page 21 Table 7).

In this sense, the proposed framework is a problem-driven and physically informed evolution of our prior work, specifically tailored to mixed CAV–HDV conflict prediction.

[Zou, Guojian., Zhou, Zhiyong., Weibel, Robert., Li, Ye., Wang, Ting., Liu, Zongshi., Ding, Weiping., & Fu, Cheng. (2025). Multi-Graph Spatio-Temporal Network for Traffic Accident Risk Forecasting. *Pattern Recognition*, 112784.]

To clarify the position of our work, we have added Table 1 in the revised manuscript to summarize representative prior studies along six key aspects (mixed CAV–HDV traffic, traffic-conflict–based surrogate safety indicators, deep spatiotemporal neural modelling, graph-based representation, manifold-based traffic-state similarity, and explicit treatment of class imbalance). As shown in Table 1, existing studies usually cover only a subset of these aspects, whereas the proposed MS-STGNet jointly incorporates all of them within a single framework, providing a more comprehensive solution for real-time safety prediction in mixed CAV–HDV traffic.

Regarding the concern that the benefits might be incremental, Sections 6.5 and 6.7 show that the manifold-similarity module leads to substantial and consistent gains over strong GNN baselines. In Section 6.5, MS-STGNet surpasses STGCN and STGAT by at least about 2 percentage points on most evaluation metrics across the majority of CAV penetration rates, and achieves up to a 24% reduction in false-alarm rate at high penetration levels. In a highly imbalanced, safety-critical task, such reductions in false positives are practically significant (See Page 18 Table 5). The ablation study in Section 6.7 further demonstrates that removing the manifold-similarity module causes the largest degradation among all tested variants, indicating that this component is the main source of improvement rather than a minor technical tweak (See Page 21 Table 7). Given that deep spatiotemporal models already fit the data strongly, these results show that incorporating the manifold-similarity prior meaningfully advances predictive performance and robustness beyond what existing architectures can achieve on this problem.

Taken together, we view MS-STGNet as a further step along an established research trajectory: it integrates a validated multi-graph spatiotemporal framework with a new manifold-similarity prior specifically designed for mixed CAV–HDV traffic conflicts, thereby improving the fit between problem formulation and model structure and leading to measurable gains in both accuracy and false-alarm control. Below is the revised version:

Revised (Page 5 Line 30—31, Page 6 Line 1—2):

As shown in Table 1, existing studies usually cover only a subset of these aspects, leaving important gaps such as the lack of conflict-based indicators, manifold-based state similarity, or explicit treatment of class imbalance. In contrast, the proposed MS-STGNet jointly incorporates all these aspects within a single framework, providing a more comprehensive solution for real-time safety prediction in mixed CAV–HDV traffic.

Table 1
Summary of key aspects considered in related studies on traffic safety prediction.

Study category	Mixed CAV-HDVtraffic	Traffic-conflict-based surrogate-safety indicator	Deep spa-tiotemporal neural model for safety	Graph-based representa-tion	Manifold-based traffic state similarity prior	Explicit treatment of class imbalance/ zero inflation
Mixed-traffic microsimulation safety studies (e.g., Liu et al. (2018a); Zhou and Zhu (2020); Yao et al. (2023); Chen et al. (2024b))	✓	✓	×	×	×	×
Grid-based deep learning crash-risk prediction (e.g., Ren et al. (2018); Bao et al. (2019); Chen et al. (2018); Hu et al. (2020))	×	×	✓	×	×	✓
ST-GNN-based accident risk models (e.g., Zhou et al. (2020); Yu et al. (2021); Gao et al. (2024); Trirat et al. (2023))	×	×	✓	✓	×	✓
Manifold learning for traffic state/safety modelling (e.g., Wang et al. (2009); Su et al. (2020); Seoa (2023); Liu et al. (2022))	×	×	✓	×	✓	✓
Our framework (MS-STGNet)	✓	✓	✓	✓	✓	✓

Comment 2:

Furthermore, the validation of the proposed model remains a critical weakness. The authors rightly acknowledge the current lack of real-world mixed CAV-HDV datasets and have adopted a simulation-based evaluation strategy calibrated with real trajectory data. However, this does not fully mitigate the concern regarding practical applicability and robustness. The entire evaluation is conducted on a single, simulated 14 km freeway corridor under specific demand patterns. The performance claims, particularly the model's stability and generalizability across diverse conditions, are therefore not substantiated with evidence from real-world or even more varied simulated scenarios. The absence of validation on any real traffic data, even from purely human-driven vehicle scenarios to first establish transferability, significantly undermines the practical impact and readiness of the model for real-world deployment. The promise of future validation, while noted, does not compensate for the present lack of empirical evidence.

Response to Comment 2:

We appreciate the reviewer's concern regarding the validation and practical applicability of the proposed model. In the revised manuscript, we have complemented the simulation-based evaluation by adding experiments on real-world freeway traffic from the FHWA NGSIM program (I-80 and US-101), where all vehicles are human-driven (see Appendix D). Using the same surrogate safety measures and a consistent input–output configuration as in the main experiments, we evaluate MS-STGNet and all baseline models on these independent datasets. The results show that MS-STGNet consistently achieves the highest AUC and overall accuracy and the lowest false-alarm rate, indicating that the proposed framework generalizes beyond the original simulated 14 km mixed CAV–HDV corridor. Since NGSIM contains only purely human-driven traffic and does not directly match our primary focus on mixed traffic flows, we report these experiments in the appendix as supplementary evidence of the model's robustness and extensibility. We have also updated Section 6.5 to briefly summarize these real-world tests and explicitly refer the reader to Appendix D for details. Below is the revised version:

Revised:

In Section 6.5 (Page 19 Line 6—10):

In addition to the simulation-based evaluation on the 14 km mixed CAV–HDV corridor, we further conduct supplementary tests on real-world freeway traffic from the FHWA NGSIM program, to examine the transferability of the proposed framework (see Appendix D). The results show that MS-STGNet still outperforms the baseline models and effectively controls false alarms, providing supporting evidence for its robustness and extensibility beyond the original simulated setting.

In Appendix D (Page 30 Line 1—15, Page 31 Line 1—16):

To further evaluate the scalability of the model and its adaptability to real-world conditions, we tested the MS-STGNet model and the baseline models on a real-world trajectory dataset. Specifically, we employed the trajectory data from the Next Generation Simulation (NGSIM) program of the U.S. Federal Highway Administration (FHWA, 2006), which has been widely used for the analysis of longitudinal car-following, lane-changing maneuvers, and *traffic flow characteristics (Chen et al., 2021; Zong et al., 2024). The highway subset of NGSIM comprises the I-80 and US-101 datasets, both collected using cameras mounted on the rooftops of adjacent high-rise buildings. The I-80 dataset covers an approximately 500 m freeway segment with six unidirectional lanes, with a total observation duration of 45 minutes (Fig. D1(a)). The US-101 trajectory dataset was collected on a roughly 640 m freeway segment with six lanes, with the same 45-minute observation period (Fig. D1(b)). Because the dimensions of the traffic statistics and the road segment lengths differ from those of our own dataset, we adopted a slightly modified data-processing procedure. The spatial aggregation interval was set to 100 m, and traffic flow, speed, and occupancy were aggregated over 3-minute windows. The indicators used to quantify traffic conflicts were kept consistent with the original study (TTC, DRAC, DDR), and the input–output configurations of all models were aligned with the original setting. The experimental results are reported in Table D1.*

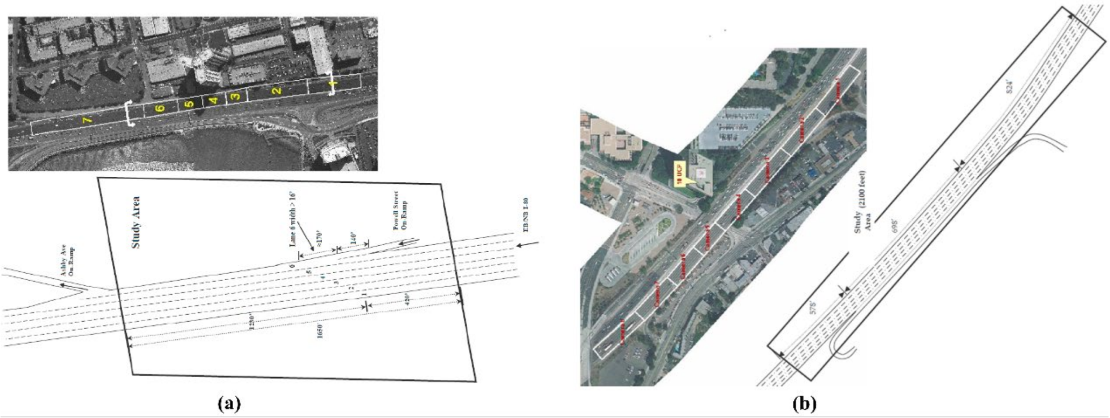

Fig. D1. Collection Scenarios of open-source datasets (a) NGSIM, I-80; (b) NGSIM US-101.

As shown in Table D1, the proposed MS-STGNet still exhibits consistently superior performance on both the I-80 and US-101 sub-datasets. On I-80, MS-STGNet achieves a recall of 76.6% compared with 75.3% for STGAT and 74.0% for STGCN, while reducing the false alarm rate from 18.0% (STGAT) to 17.1% and improving AUC, accuracy, and G-mean from 77.6%, 81.0%, and 77.4% to 79.6%, 84.0%, and 80.4%, respectively. A similar pattern is observed on US-101, where MS-STGNet improves recall from 74.3% to 76.2%, decreases the false alarm rate from 20.4% to 16.9%, and further increases AUC, accuracy, and G-mean from 78.2%, 80.0%, and 76.8% to 80.2%, 83.2%, and 79.5%. These consistent gains over both traditional machine-learning methods (SVM, XGBoost) and strong deep and graph-based baselines (CNN, LSTM-CNN, STGCN, STGAT) confirm that MS-STGNet maintains strong scalability and robustness when transferred to real-world freeway traffic conditions.

Table D1
Performance on NGSIM dataset.

Sub-datasets	**Metric**	**SVM**	**XGBoost**	**CNN**	**LSTM-CNN**	**STGCN**	**STGAT**	**MS-STGNet**
I-80	Recall	0.502	0.570	0.692	0.709	0.740	0.753	$\mathbf{0.766}_{\uparrow 1.73\%}$
	False alarm rate	0.464	0.436	0.224	0.220	0.195	0.180	$\mathbf{0.171}_{\uparrow 5.00\%}$
	AUC	0.557	0.614	0.740	0.749	0.768	0.776	$\mathbf{0.796}_{\uparrow 2.58\%}$
	Accuracy	0.562	0.628	0.774	0.779	0.819	0.810	$\mathbf{0.840}_{\uparrow 2.56\%}$
	G-mean	0.512	0.565	0.721	0.738	0.778	0.774	$\mathbf{0.804}_{\uparrow 3.34\%}$
US-101	Recall	0.492	0.560	0.682	0.699	0.730	0.743	$\mathbf{0.762}_{\uparrow 2.56\%}$
	False alarm rate	0.469	0.436	0.232	0.219	0.182	0.204	$\mathbf{0.169}_{\uparrow 7.14\%}$
	AUC	0.547	0.604	0.730	0.739	0.758	0.782	$\mathbf{0.802}_{\uparrow 2.56\%}$
	Accuracy	0.552	0.618	0.764	0.769	0.809	0.800	$\mathbf{0.832}_{\uparrow 2.84\%}$
	G-mean	0.502	0.555	0.711	0.728	0.764	0.768	$\mathbf{0.795}_{\uparrow 3.52\%}$

From the results reported in Table D1, we can again observe that MS-STGNet achieves the most pronounced improvement in terms of reducing the false positive rate. Accordingly, we further visualized the spatiotemporal speed heat maps and the corresponding conflict detection outcomes (Fig. D2). The visualizations indicate that, under real-world testing conditions, the three architectures yield almost identical predictions of traffic conflicts. Nevertheless, within the stop-and-go waves, a small number of false positives can still be observed for the STGAT and STGCN models. This finding further corroborates that, by incorporating flow similarity to suppress noise, the MS-STGNet framework attains robust predictive performance with strong adaptability and reliability across diverse traffic scenarios.

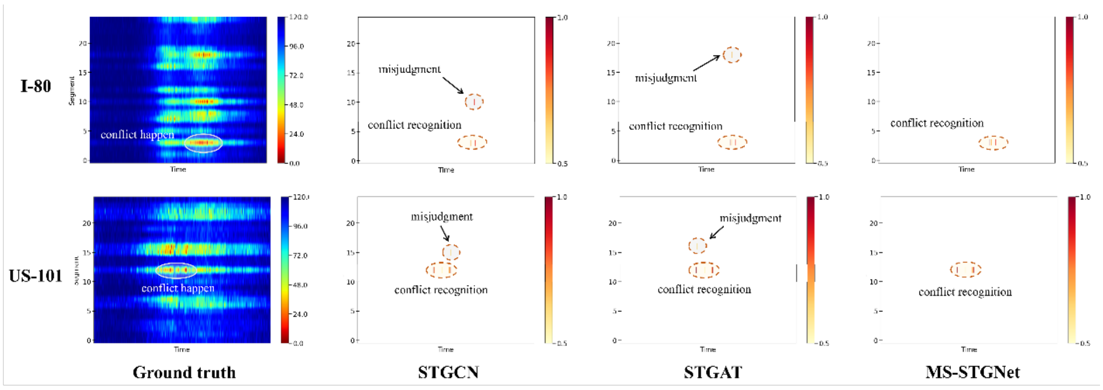

Fig. D2. Speed heatmap and conflict recognition result for STGCN, STGAT, and MS-STGNet under NGSIM dataset.

Comments from Reviewer #3:

Comments:

Dear Authors,

thank you very much for a very thorough review! I like the paper a lot! It is very good written, very nicely explained.

Yours,

AE

Response to Comments:

We sincerely thank the Associate Editor for the positive evaluation and encouraging comments. We are very grateful for your time and consideration, and we greatly appreciate your support for our work. In the revised manuscript, we have further clarified and strengthened the presentation of the methodological innovations and empirical results, and we have carefully addressed all comments from the reviewers through additional analyses, experiments, and clarifications. We hope that these revisions satisfactorily resolve the remaining concerns and that the revised manuscript will be favorably considered for publication.

Comments from Reviewer #5:

Comments:

Comment 1: The authors explain that current datasets have limitations for their specific problem. They strengthened the limitations section and the conclusion, clarifying that their hybrid strategy is currently the best possible approach. This comment is considered addressed.

Comment 2: The authors incorporated Section 6.1 (Data Preparation). They examined the influence of velocity dispersion and volume on the likelihood of confrontation. This comment is considered addressed.

Comment 3: The authors created Section 6.6 (Computation cost). They compared GPU memory usage (training and inference) and the number of parameters between their model and the baselines. They demonstrated that their model is lighter in parameters and memory than the direct competitors (STGCN/STGAT), although heavier than simpler models (CNN). This comment is considered addressed.

Response to Comments:

We sincerely thank the reviewer for carefully checking our revisions and for the positive assessment that Comments 1–3 have been satisfactorily addressed. We are grateful for your constructive feedback, which has helped us clarify the limitations, enrich the data analysis, and provide a more complete discussion of computational cost.