Modification and Responses to the Reviewers' Comments and Suggestions
Manuscript ESWA-D-25-28448R2

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:

- Strengthened the problem motivation and aligned it with our core solution in the Abstract. We rewrote the opening of the Abstract to explicitly highlight why existing approaches relying on Euclidean/instantaneous similarity are vulnerable to stop-and-go wave noise and velocity separation, and to directly motivate the Manifold Similarity module as the key mechanism for suppressing false positives.
- Improved the discussion and treatment of severe class imbalance. Beyond our LMF objective (LDAM + Focal), we added a concise discussion of negative sampling as a primary mechanism in graph representation learning, incorporated the reviewer-recommended SOTA references, and explicitly stated the complementary nature of NS and loss re-weighting. We also included NS integration as a concrete future-work direction.
- Enhanced empirical rigor and practicality. We (1) reported training-phase results in an appendix to complement testing-phase performance, (2) added latency and computational-cost reporting (training and inference) to support deployment feasibility, and (3) explicitly listed strengths/limitations and strengthened qualitative misclassification/visualization-based discussion to improve transparency.
- In response to Reviewer #1, we consolidated and clearly summarized our previous revisions (first revision and second revision) addressing the same concerns to present a more coherent logic chain; we then provided a structured, point-by-point reply to the reviewer's questions, and added targeted, up-to-date supporting references to substantiate our arguments and strengthen the manuscript's scholarly grounding.

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. The previous revisions can be found in https://github.com/liuzongshi123/Previous-Revisions.

Comments from Reviewer #1:

Comment 1:

Despite the revisions made, fundamental concerns regarding the methodological novelty and empirical validation of the proposed MS-STGNet framework remain unresolved. The core contribution, the integration of a manifold-similarity prior into a spatiotemporal graph network, is argued to be a conceptual advancement over adaptive adjacency mechanisms like STGAT. However, as noted by Reviewer #1 and evident in the authors' response, this approach essentially initializes the graph from offline-computed manifold distances rather than learning it dynamically from features. The authors' defense positions this as a "problem-driven and physically informed evolution," but the technical distinction appears incremental. The reported performance gains, while consistent, are modest (e.g., ~2 percentage point increases in recall, up to 24% reduction in false alarm rate under specific conditions). In a field densely populated with sophisticated GNN variants, such improvements do not convincingly demonstrate a transformative step or major methodological leap. The framework largely repurposes mature components (ResNet, TCN, fusion gates), and the addition of the manifold module, though beneficial, does not constitute a significant theoretical or architectural breakthrough sufficient for publication in this journal.

Response to Comment 1:

Thank you for the continued and detailed feedback. We acknowledge the reviewer's concern regarding (i) the perceived incremental nature of introducing a manifold-similarity prior and (ii) the interpretation of the empirical gains. To address these points more clearly, we respond in four parts:

(1) Summary of revisions from our first response (clarifying novelty and positioning).

In the previous revision, we strengthened the manuscript's positioning at both the problem and method levels. We emphasized that our target problem is real-time conflict prediction in mixed CAV–HDV freeway traffic, where robustness and false-alarm control are critical and where existing work remains limited. We also explicitly distinguished MS-STGNet from (a) manifold-learning studies that typically use manifold embeddings for clustering/visualization or as auxiliary features, and (b) STGAT-type models that learn adjacency purely from instantaneous node features. Concretely, we expanded Sections 2.2/2.3, revised Section 5.3.1 to explain "manifold-informed prior + lightweight adaptive refinement," and added an explicit discussion in Section 6.5 highlighting how the manifold prior reduces spurious activations and false alarms under speed separation.

(2) Summary of revisions from our second response (making the technical distinctions and evidence more explicit).

In the second revision, we further clarified that MS-STGNet is a problem-driven extension along an established research line, and we articulated three concrete innovations brought by the manifold-similarity module: (i) constructing the initial graph from traffic-state manifolds derived from long-term multi-dimensional flow–speed–occupancy trajectories as an explicit traffic-state prior, (ii) learnable refinement with multi-order message passing to enable information exchange among physically similar yet geographically distant regions, and (iii) suppressing spurious conflict activations, which is also supported by the ablation study showing

the largest degradation when removing this module. To further clarify novelty in context, we also added Table 1 summarizing representative prior studies across six aspects, showing that existing work typically covers only a subset, while MS-STGNet integrates all within one framework.

(3) On the "offline-computed prior vs. dynamic learning" concern (and practical deployability).

We agree that our manifold similarity matrix is pre-computed from traffic-state data (thus already traffic-characteristic and physically meaningful) and is kept fixed during inference for computational stability and real-time constraints; however, this does not imply the prior is immutable in real deployments. In practice, the prior can be periodically recalibrated using newly observed traffic streams (e.g., a sliding time window from detectors), thereby continuously refreshing the traffic-state geometry and updating the prior knowledge base. Furthermore, as we explicitly mentioned in Section 7, periodic/online prior updates are a feasible extension and a direction for our further research.

(4) On the magnitude and significance of improvements (especially false alarms).

We think that the observed gains are convincing for this safety-critical, highly imbalanced task. In related traffic safety/accident modeling literature, improvements are often incremental because baselines are already strong; for example, recent work reports **AUC gains on the order of ~0.3%–4.8%** as meaningful improvements in traffic accident analysis settings (Jiang et.al, 2024). More importantly, in conflict/incident detection, **false alarms are a primary operational bottleneck** (too many false alarms leads to alarm fatigue and undermines trust), and reductions in false-alarm rate are commonly highlighted as practically significant; prior studies report improvements such as **~25% decrease in false-alarm rate** as an important advancement (Nathanail et.al, 2017; Formosa et.al, 2020; Gao et.al, 2020). In our experiments, MS-STGNet provides consistent gains across penetration rates and metrics, and achieves up to a 24% reduction in false-alarm rate under high penetration scenarios. We therefore view the manifold-similarity prior not as a cosmetic tweak, but as a practically meaningful mechanism for suppressing noise-induced spurious activations in mixed traffic.

[Nathanail, E., Kouros, P., & Kopelias, P. (2017). Traffic volume responsive incident detection. Transportation research procedia, 25, 1755-1768.

Formosa, N., Quddus, M., Ison, S., Abdel-Aty, M., & Yuan, J. (2020). Predicting real-time traffic conflicts using deep learning. Accident Analysis & Prevention, 136, 105429.

Gao, Q., Yin, H., & Zhang, W. (2020). Lane departure warning mechanism of limited false Alarm rate using extreme learning residual network and ϵ-greedy LSTM. Sensors, 20(3), 644.

Jiang, X., Chen, X., Wang, H., & Razi, A. (2024, December). Geographical Information Alignment Boosts Traffic Analysis via Transpose Cross-attention. In 2024 IEEE International Conference on Big Data (BigData) (pp. 6031-6036). IEEE.]

We hope the above clarifications adequately address the reviewer's concerns and resolve any remaining confusion.

Comment 2:

The validation strategy, though supplemented, remains a critical weakness undermining the

paper's impact. The authors have added experiments on the NGSIM dataset (pure human-driven traffic) to demonstrate transferability. However, this does not address the primary objective of the study: predicting conflicts in mixed CAV-HDV traffic. The entire model development and core evaluation are still conducted on a single, simulated 14 km freeway corridor under specific demand patterns. The newly added real-world data is not only non-mixed but also represents a fundamentally different and less complex scenario. Consequently, the central claims regarding the model's robustness, stability, and generalizability across diverse mixed-traffic conditions are not substantiated. The promise of the model for "real-time mixed traffic control on intelligent highways" is therefore speculative and not supported by empirical evidence from realistic or varied mixed-traffic environments, simulated or otherwise. This lack of convincing validation on the core problem domain severely limits the paper's practical contribution and readiness.

Response to Comment 2:

Thank you for this important and detailed comment. We agree that validation is central for establishing practical relevance, and we respond in three parts.

(1) On the study scope and the "single corridor" concern (why a freeway corridor with ramps is an appropriate and widely used safety research setting).

Our work targets freeway safety/conflict prediction (including mainline and ramp/merging influence zones), which is a common and practically relevant setting in proactive safety research. Recent crash/conflict risk studies frequently focus on freeway segments and operational treatments (e.g., variable speed limits, weaving/merging areas, tunnels), because these facilities exhibit pronounced stop-and-go waves and concentrated safety risk (Aty et.al, 2024; Chen et.al, 2024; Ma et.al, 2025). Within this scope, we designed a 14 km freeway corridor with ramps to ensure sufficient length for realistic acceleration–cruising–interaction dynamics and to avoid boundary-dominated artifacts, consistent with our simulation setup description in Section 4.1. Similar simulation structures have also been used in previous mixed traffic safety predictions (Lu et.al, 2021). Importantly, our evaluation is not restricted to a single operating point: we test multiple CAV penetration rates (10%–90%) and a wide range of demand levels, and we report cross-run stability (mean±std over multiple seeds) with statistical significance.

[Abdel-Aty, M., Hasan, T., & Anik, B. T. H. (2024). An advanced real-time crash prediction framework for combined hard shoulder running and variable speed limits system using transformer. Scientific Reports, 14(1), 26403.

Chen, K., Xu, C., Liu, P., Li, Z., & Wang, Y. (2024). Evaluating the performance of traffic conflict measures in real-time crash risk prediction using pre-crash vehicle trajectories. Accident Analysis & Prevention, 203, 107640.

Ma, F., Wang, X., & Yang, W. (2025). Real-time accident risk identification for freeway weaving segments based on video analytics. Measurement, 242, 115783.

Lu, Q. L., Yang, K., & Antoniou, C. (2021, September). Crash risk analysis for the mixed traffic flow with human-driven and connected and autonomous vehicles. In 2021 ieee international intelligent transportation systems conference (itsc) (pp. 1233-1238). IEEE.]

(2) Why we added NGSIM experiments (and how we position them).

We added the NGSIM (I-80 and US-101) tests specifically in response to the reviewer's prior request that

"the absence of validation on any real traffic data, even purely human-driven vehicle scenarios to first establish transferability, undermines practical impact."

Accordingly, we complemented the main simulation results with supplementary real-world freeway tests using the same surrogate safety measures (TTC/DRAC/DDR) and aligned input–output configurations, and we reported them in Appendix E (and summarized them briefly in Section 6.5) to demonstrate transferability beyond the original simulated corridor. We fully acknowledge (and explicitly state) that NGSIM is pure HDV and therefore does not replace mixed-traffic field validation; rather, it provides supporting evidence of robustness and extensibility under real-world freeway dynamics.

(3) Why real-world mixed CAV–HDV validation is currently not feasible for our specific prediction target (and how we revised claims accordingly).

At the early stage, our intention was to validate on real-world mixed CAV–HDV data. However, after reviewing available datasets, we found that current data sources typically cannot simultaneously satisfy the two core requirements of our task: (i) truly mixed CAV–HDV traffic and (ii) long, spatially continuous freeway facilities with continuous macroscopic measurements (flow/speed/occupancy) suitable for segment-level conflict prediction over extended time series. This limitation is also consistent with broader reviews noting that empirical safety studies for mixed CAV–HDV environments remain scarce and simulation-based evaluation is often necessary due to data constraints (Ali et.al, 2024). Moreover, even recent CAV-related safety evaluation work explicitly highlights the growing reliance on simulation-based studies in the absence of comprehensive real-world CAV safety data (Zhang et.al, 2024; Do et.al, 2025). Given these constraints, we adopted a hybrid strategy: (a) calibrating HDV behavior using real freeway trajectories (highD) and (b) generating mixed CAV–HDV trajectories with multiple penetration rates, labeling conflicts via physically interpretable SSMs, and then performing macroscopic segment-level prediction along a long corridor.

[Ali, Y., Hussain, F., & Haque, M. M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. Accident Analysis & Prevention, 194, 107378.

Zhang, M., Yang, J., Yang, X., & Duan, X. (2024). Measuring Collision Risk in Mixed Traffic Flow Under the Car-Following and Lane-Changing Behavior. Applied Sciences, 14(23), 11400.

Do, W., Saunier, N., & Miranda-Moreno, L. (2025). Evaluation of conventional surrogate indicators of safety for connected and automated vehicles in car following at signalized intersections. Transportation Research Record, 2679(2), 1118-1133.]

In the conclusion section, we also look forward to testing our method on real-world mixed traffic data in the future. We hope the above clarifications adequately address the reviewer's concerns and resolve any remaining confusion.

Comments from Reviewer #7:

Comment 1:

Add the corresponding graphical abstract.

Response to Comment 1:

Thank you for the suggestion. After checking the journal's author guidelines, a graphical abstract is not required for this submission. Therefore, we did not prepare a separate graphical abstract.

Comment 2:

Improve the abstract's writing by including the following elements: research problem, research gap, study objective, methodology, results, and conclusion.

Response to Comment 2:

Thank you for the constructive suggestion. We have revised the abstract to more explicitly and systematically cover the research problem, identified gap, study objective, methodology, key quantitative findings, and conclusions/implications. In particular, we clarified the core technical components of MS-STGNet (including manifold-similarity neighbor selection and adaptive fusion) and added representative performance improvements (e.g., reduced false alarm rate and improved AUC/accuracy) to make the contribution and outcomes easier to assess. Below is the revised version:

Revised (Page 1 Line 13—31):

The coexistence of connected and automated vehicles (CAVs) and human-driven vehicles (HDVs) introduces complex non-linear dynamics, characterized by stop-and-go wave noise and velocity separation, making real-time safety risk assessment difficult. Current research on crash/conflict prediction in mixed CAV–HDV traffic remains limited, existing risk assessment models, which predominantly rely on linear Euclidean distances or instantaneous feature similarity, often misinterpret non-conflict fluctuations as crash precursors, resulting in unstable performance and high false alarm rates. To address this, we propose a Manifold Similarity Spatiotemporal Graph Network (MS-STGNet) tailored for robust real-time conflict prediction in mixed freeway traffic. Unlike distinguishing traffic states in a linear space, this model constructs a manifold-based traffic-state similarity graph to capture the intrinsic geometric structure of traffic evolution. It integrates physical adjacency with semantic neighbors and combines residual feature extraction, temporal convolution, and an adaptive fusion gate to learn spatiotemporal risk patterns. We evaluated the framework's performance under mixed traffic scenarios with varying penetration rates of CAVs and HDVs. The experimental results demonstrate that MS-STGNet achieves consistently exceptional and stable performance across varying market penetration levels and traffic scenarios. Compared to state‐of‐the‐art baseline models, it delivers higher predictive accuracy and substantially lower false alarm rates. The methodologies and outcomes presented in this study have the potential to be used for real-time mixed traffic control on intelligent highways and crash prevention in real-time crash

risk warnings at high-risk locations.

Comment 3:

The datasets used should be described, including the input variables and the output (target) variable.

Response to Comment 3:

Thank you for the suggestion. The dataset description, input variables and the target/output definition is already provided in the manuscript, and we have now made this clearer by explicitly pointing to the relevant sections. Specifically, the dataset description was summarized in Section 6.1 (See Page 14 Line 31—37, Page 15 Line 1—9). The input variables are loop-detector–based traffic states—volume, mean speed, and mean occupancy—summarized in Section 6.1 and Table 3 (See Page 14 Line 31—32). The target/output variable is the (binary) occurrence of a traffic conflict in the prediction horizon—summarized in Section 6.3 (See Page 16 Line 1—4).

Comment 4:

The architectures of the models used should be described, both the proposed model and the baseline models.

Response to Comment 4:

Thank you for this helpful comment. In the manuscript, the architecture of the proposed MS-STGNet is described in Section 5.1 (Model architecture overview) and illustrated in Fig. 3, where we summarize the four main components (residual convolutional module, manifold-similarity graph module, TCN layer, and fusion gate mechanism) and explain the overall information flow (See Page 9 Line 16—25). Detailed descriptions of each component are provided in the subsequent subsections of Section 5 (See Page 9 Line 26—36, Page 10—Page 13, Page 14 Line 1—28). For the baseline models, we describe their architectures in Section 6.4, where we list each baseline (SVM, XGBoost, CNN, LSTM-CNN, STGCN, and STGAT) and briefly summarize its modeling structure/mechanism with supporting references to the original studies (See Page 16 Line 5—12, Page 17 Line 1—19).

Comment 5:

The results should be presented for both the training phase and the testing phase.

Response to Comment 5:

Thank you for the suggestion. We agree that reporting results for both phases can help readers better understand model fitting behavior and generalization. In the revised manuscript, we clarify that Table 5 reports the testing-phase performance, and we additionally provide the training-phase results in a new Appendix D (Table D1) for completeness. Below is the revised version:

Revise:

In Section 6.5 (Page 17 Line 21—23):

Table 5 summarizes the testing-phase performance metrics of the proposed MS-STGNet and all baseline models for conflict prediction under five different CAV penetration rates (10%, 30%, 50%, 70%, and 90%), the training-phase results are provided in Appendix D.

In Appendix D (Page 30 Line 1—5, Page 31 Table D1):

For completeness, Table D1 reports the performance of all compared methods on the training set under different CAV penetration rates. Overall, MS-STGNet consistently achieves the best training-phase performance across all metrics than the baselines. These training-phase results, together with the testing-phase results reported in Table 6, provide a clearer view of the model's fitting behavior and generalization performance.

Table E1
Performance of Different Models during the training phase.

Penetration rates	Metric	SVM	XGBoost	CNN	LSTM-CNN	STGCN	STGAT	MS-STGNet
10%	Recall	0.553	0.605	0.745	0.760	0.804	0.826	**0.850**
	False alarm rate	0.388	0.384	0.175	0.161	0.120	0.131	**0.116**
	AUC	0.618	0.672	0.785	0.801	0.851	0.840	**0.878**
	Accuracy	0.628	0.667	0.847	0.861	0.886	0.888	**0.915**
	G-mean	0.566	0.626	0.776	0.799	0.817	0.842	**0.872**
30%	Recall	0.617	0.675	0.786	0.772	0.832	0.841	**0.862**
	False alarm rate	0.364	0.321	0.155	0.142	0.134	0.132	**0.108**
	AUC	0.609	0.690	0.803	0.818	0.841	0.864	**0.876**
	Accuracy	0.626	0.697	0.832	0.821	0.849	0.885	**0.900**
	G-mean	0.631	0.683	0.816	0.824	0.853	0.870	**0.884**
50%	Recall	0.610	0.631	0.798	0.833	0.836	0.858	**0.883**
	False alarm rate	0.404	0.290	0.120	0.120	0.118	0.113	**0.094**
	AUC	0.630	0.732	0.840	0.834	0.850	0.888	**0.912**
	Accuracy	0.628	0.706	0.843	0.855	0.907	0.890	**0.930**
	G-mean	0.625	0.660	0.816	0.835	0.872	0.875	**0.897**
70%	Recall	0.625	0.641	0.811	0.784	0.811	0.848	**0.875**
	False alarm rate	0.394	0.299	0.143	0.148	0.125	0.117	**0.085**
	AUC	0.632	0.705	0.816	0.803	0.840	0.879	**0.907**
	Accuracy	0.629	0.729	0.846	0.822	0.851	0.870	**0.898**
	G-mean	0.632	0.692	0.827	0.832	0.846	0.855	**0.881**
90%	Recall	0.645	0.661	0.823	0.801	0.844	0.862	**0.891**
	False alarm rate	0.345	0.291	0.146	0.125	0.112	0.107	**0.081**
	AUC	0.645	0.717	0.822	0.839	0.846	0.883	**0.910**
	Accuracy	0.626	0.726	0.844	0.866	0.907	0.897	**0.917**
	G-mean	0.624	0.672	0.860	0.868	0.877	0.889	**0.907**

Comment 6:

In addition to computational cost, model latency should also be analyzed, and a corresponding complexity analysis should be conducted.

Response to Comment 6:

Thank you for this valuable comment. We agree that model latency and complexity are critical factors for practical deployment. We clarify that model complexity is reflected by the number of trainable parameters, which has already been reported in Section 6.6 and Table 5 (See Page 19 Line 32—38). In addition, to explicitly analyze model latency, we have extended the computational evaluation by reporting both training cost (per epoch) and inference cost, which directly characterize the time efficiency of each model during training and deployment. These newly added metrics, together with GPU memory usage and parameter size, provide a more comprehensive analysis of computational efficiency, model complexity, and latency. Below is the revised version:

Revised (Page 19 Line 32—50, Page 20 Table 6):

In real-world deployment, predictive accuracy is the primary requirement for traffic safety applications, while the hardware cost of the deployed model constitutes a secondary but still crucial consideration for practical implementation. To highlight the computational overhead of different approaches, Table 6 reports five indicators under five CAV penetration-rate scenarios: GPU-MUT (peak GPU memory usage during training), GPU-MUI (peak GPU memory usage during inference), number of trainable parameters, training cost, and inference cost. For the classical machine-learning baselines (SVM and XGBoost), GPU-based indicators are omitted ("–") because they are trained and executed on CPU and their memory footprint is negligible compared with deep models in our setting.

Overall, MS-STGNet achieves a favorable balance between model expressiveness and computational efficiency. Although its training cost per epoch is slightly higher than that of lightweight CNN-based baselines, MS-STGNet consistently requires fewer parameters than other deep graph-based models such as STGCN and STGAT, indicating a lower structural complexity. This compact design contributes to moderate GPU memory consumption while maintaining strong predictive performance. From a latency perspective, MS-STGNet remains computationally efficient during inference. As shown in Table 6, its inference cost is comparable to or lower than that of other spatiotemporal graph models, and significantly lower than methods with more complex graph attention or message-passing mechanisms. While certain learning-based methods (e.g., CNN) exhibit lower latency due to their simpler architectures, they do so at the expense of predictive accuracy.

By jointly considering model complexity (parameters), computational cost, and latency (training and inference time), the results demonstrate that MS-STGNet achieves a well-balanced trade-off, making it suitable for real-time and large-scale traffic conflict prediction applications.

Table 6
The computational performance of different models on dataset.

Penetration rates	Metric	SVM	XGBoost	CNN	LSTM-CNN	STGCN	STGAT	MS-STGNet
10%	GPU-MUT	—	—	4,333MiB	4,443MiB	5,574MiB	5,802MiB	**5,031MiB**
	GPU-MUI	—	—	2,283MiB	2,799MiB	4,446MiB	3,986MiB	**3,359MiB**
	Parameters	—	—	298,742	346,251	594,758	528,759	**490,154**
	Training cost	—	—	16.013s	16.255s	19.773s	20.748s	**18.684s**
	Inference cost	—	—	2.226s	2.384s	4.374s	4.148s	**3.857s**
30%	GPU-MUT	—	—	4,419MiB	4,530MiB	5,684MiB	5,917MiB	**5,130MiB**
	GPU-MUI	—	—	2,328MiB	2,854MiB	4,534MiB	4,065MiB	**3,425MiB**
	Parameters	—	—	304,621	353,064	606,462	539,164	**499,800**
	Training cost	—	—	16.331s	16.573s	20.163s	21.159s	**19.052s**
	Inference cost	—	—	2.270s	2.431s	4.461s	4.230s	**3.933s**
50%	GPU-MUT	—	—	3,496MiB	3,584MiB	4,497MiB	4,681MiB	**4,059MiB**
	GPU-MUI	—	—	1,842MiB	2,258MiB	3,587MiB	3,216MiB	**2,710MiB**
	Parameters	—	—	241,008	279,335	479,816	426,572	**395,428**
	Training cost	—	—	12.920s	13.112s	15.952s	16.739s	**15.074s**
	Inference cost	—	—	1.796s	1.923s	3.529s	3.347s	**3.112s**
70%	GPU-MUT	—	—	3,085MiB	3,162MiB	3,968MiB	4,130MiB	**3,581MiB**
	GPU-MUI	—	—	1,625MiB	1,992MiB	3,165MiB	2,838MiB	**2,391MiB**
	Parameters	—	—	212,656	246,474	423,370	376,390	**348,909**
	Training cost	—	—	11.401s	11.568s	14.076s	14.769s	**13.299s**
	Inference cost	—	—	1.584s	1.697s	3.114s	2.953s	**2.745s**
90%	GPU-MUT	—	—	2,983MiB	3,058MiB	3,837MiB	3,994MiB	**3,463MiB**
	GPU-MUI	—	—	1,572MiB	1,927MiB	3,061MiB	2,744MiB	**2,312MiB**
	Parameters	—	—	205,636	238,338	409,394	363,965	**337,392**
	Training cost	—	—	11.024s	11.188s	13.611s	14.283s	**12.861s**
	Inference cost	—	—	1.533s	1.641s	3.011s	2.856s	**2.655s**

Comment 7:

Perform the appropriate statistical analysis to determine whether there is a significant difference between the proposed method and the baseline models.

Response to Comment 7:

Thank you for the suggestion. We have already conducted statistical analyses to examine whether the performance differences between MS-STGNet and the baseline models are statistically significant. As stated in Section 6.5 (Performance comparison), each result in Table 5 is reported as the mean ± standard deviation over five independent runs with different random seeds, and statistical tests across these five runs confirm that the improvements over baseline models are significant at the 5% level ($p < 0.05$) for all reported metrics and penetration-rate scenarios. Meanwhile, a comparative analysis of the performance of different models on various evaluation indicators was conducted (See Page 17 Line 20—48, Page 19 Line 1—30).

Comment 8:

The strengths of the proposed approach, as well as its limitations, should be clearly listed.

Response to Comment 8:

Thank you for this helpful suggestion. In the manuscript, we have clarified and explicitly listed the strengths and limitations of the proposed MS-STGNet framework in the Conclusion (Section 7). In this section, we summarize the main technical advantages (e.g., the manifold-similarity mechanism for more informative neighbor selection and improved robustness, especially in mixed traffic) and highlight the practical implications for proactive freeway safety management (See Page 26 Line 19—25). In addition, we now present the limitations as clearly enumerated items in the same section to improve readability and transparency, and we briefly outline corresponding future research directions (See Page 27 Line 8—18).

Comment 9:

According to Table 7, there is still considerable room for improvement. Therefore, the misclassified records should be analyzed, and future work aimed at improving the results should be described.

Response to Comment 9:

Thank you for this helpful suggestion. We would like to clarify that Table 7 reports the ablation study, rather than the final performance ceiling of the proposed method. In Section 6.7, we explicitly state that the ablation study is conducted by progressively removing key components to quantify their individual contributions, and Table 7 provides the corresponding statistical results. Therefore, the "performance drop" in Table 7 indicates how each removed component degrades performance relative to the complete MS-STGNet, which further supports the necessity and effectiveness of the proposed modules. Meanwhile, the reviewer's constructive request for misclassification analysis, we have strengthened the discussion using our qualitative visualization-based analysis (Section6.8, 6.9, 6.10) the speed heatmaps and detection outcomes show that baseline models (STGCN/STGAT) still produce a small number of false positives in stop-and-go waves, whereas MS-STGNet better suppresses such noise and achieves more reliable recognition across scenarios. We also highlight in the main results discussion that the manifold similarity module helps reduce misjudgments in conflict-prone traffic flow (See Page 20 Line 1—12, Page 21 Line 1—15, Page 22 Line 1—15). Finally, regarding future work for improvement, we have clarified that a key direction is to further enhance the manifold-similarity component by moving beyond a purely prior/predefined similarity structure and enabling more adaptive (or learnable) updates, since relying only on prior knowledge can limit discovery of latent spatial correlations (See Page 27 Line 18—26).

Comment 10:

Include in the conclusion the findings of the statistical analysis in accordance with the revised version (Comment 7).

Response to Comment 10:

Thank you for the suggestion. We have added a concise statement in Section 7 explicitly summarizing the statistical significance findings reported in Section 6.5 / Table 5. In particular, we now state that the improvements of MS-STGNet over the baselines are statistically significant at the 5% level ($p < 0.05$) based on five independent runs with different random seeds.Below is the revised version:

Revised (Page 27 Line 1—2):

Moreover, statistical tests across five independent runs confirm that the performance improvements of MS-STGNet over the baselines are statistically significant at the 5% level ($p < 0.05$) across all penetration-rate scenarios.

Comments from Reviewer #8:

Comment 1:

The current abstract is technically sound but opens broadly with general limitations: "The coexistence... introduces significant uncertainties for real-time safety risk assessment." This can be strengthened by explicitly linking the problem to the model's core solution: the Manifold Similarity module. The authors should incorporate the insight from their own analysis and ablation studies directly into the problem statement in the Abstract.

Suggestion: Rephrase the problem to emphasize that existing models fail because they rely on linear distances or instantaneous feature similarity, making them vulnerable to false-positive predictions due to stop-and-go wave noise or velocity separation (which MS-STGNet is designed to suppress).

Response to Comment 1:

We sincerely thank the reviewer for this insightful suggestion. We agree that the original opening was too general and did not sufficiently highlight the specific technical gap that our Manifold Similarity module addresses. We have rewritten the first half of the Abstract to explicitly state that the non-linear dynamics of mixed traffic (e.g., stop-and-go waves and velocity separation) cause traditional models relying on Euclidean distance or instantaneous similarity to generate high false alarm rates. This provides a much stronger motivation for the introduction of the MS-STGNet. Below is the revised version:

Revised (Page 1 Line 13—31):

The coexistence of connected and automated vehicles (CAVs) and human-driven vehicles (HDVs) introduces complex non-linear dynamics, characterized by stop-and-go wave noise and velocity separation, making real-time safety risk assessment difficult. Current research on crash/conflict prediction in mixed CAV–HDV traffic remains limited, existing risk assessment models, which predominantly rely on linear Euclidean distances or instantaneous feature similarity, often misinterpret non-conflict fluctuations as crash precursors, resulting in unstable performance and high false alarm rates. To address this, we propose a Manifold Similarity Spatiotemporal Graph Network (MS-STGNet) tailored for robust real-time conflict prediction in mixed freeway traffic. Unlike distinguishing traffic states in a linear space, this model constructs a manifold-based traffic-state similarity graph to capture the intrinsic geometric structure of traffic evolution. It integrates physical adjacency with semantic neighbors and combines residual feature extraction, temporal convolution, and an adaptive fusion gate to learn spatiotemporal risk patterns. We evaluated the framework's performance under mixed traffic scenarios with varying penetration rates of CAVs and HDVs. The experimental results demonstrate that MS-STGNet achieves consistently exceptional and stable performance across varying market penetration levels and traffic scenarios. Compared to state - of - the - art baseline models, it delivers higher predictive accuracy and substantially lower false alarm rates. The methodologies and outcomes presented in this study have the potential to be used for real-time mixed traffic control on intelligent highways and crash prevention in real-time crash risk warnings at high-risk locations.

Comment 2:

The core task is binary classification (Conflict/Non-Conflict) on a highly imbalanced dataset (zero-inflation, $\approx 1:25$ to $1:38$ ratio). While the authors address this using a custom loss function ($\mathcal{L}_{LDAM} + \mathcal{L}_{Focal}$), GNNs often use specialized Negative Sampling (NS) techniques to handle sparse positive relationships (conflicts) against overwhelming negative relationships (non-conflicts). The suggested SOTA papers are highly relevant to this domain and should be discussed.

The authors should add a brief discussion of Negative Sampling methods in the context of GNNs for classification on unbalanced graphs.

- Acknowledge that NS techniques are a primary mechanism for handling imbalance in graph tasks. Discuss how the suggested methods (Layer-diverse and Diverse Negative Samples) aim to sample "hard negatives" or "diverse negatives" to force the model to learn better separation boundaries.

- The authors should briefly contextualize the relevance of NS to their work. For instance, while MS-STGNet addresses imbalance via a sophisticated loss function, future work could explore if combining this loss with hard negative sampling further enhances performance or stability.

Suggested References:

- Learning from the Dark: Boosting Graph Convolutional Neural Networks with Diverse Negative Samples. AAAI 2022.

- Layer-diverse Negative Sampling for Graph Neural Networks.TMLR 2024.

Response to Comment 2:

Thank you for this insightful suggestion. We agree that negative sampling (NS) is a primary mechanism in graph representation learning for handling highly imbalanced supervision, especially when positive signals are sparse relative to abundant negatives. In the revised manuscript, we have added a brief discussion of NS in the context of GNN-based imbalanced learning and incorporated the recommended state-of-the-art references. We also clarified the positioning of our method: MS-STGNet currently addresses class imbalance mainly at the objective level via the proposed LMF loss (LDAM + Focal), and we expanded the explanation of why combining LDAM (margin adjustment for inter-class imbalance) with Focal loss (down-weighting easy samples) is beneficial. While we did not explicitly adopt NS in the current experiments, we emphasize that NS is orthogonal and complementary to loss re-weighting. Accordingly, we added a future-work direction to explore integrating advanced NS techniques with our loss framework to better identify informative “hard negatives” under highly dynamic traffic conditions. Below is the revised version:

Revised:

In Section 2.2 (Page 4 Line 29—34):

Beyond loss re-weighting, negative sampling (NS) is a common mechanism in graph representation learning to address highly imbalanced supervision, by selecting informative

negative instances rather than treating all negatives equally. Duan et al. (2022) proposed boosting GCNs with diverse negative samples to prevent the model from overfitting to easy negatives, thereby enhancing representation quality. Recent work by (Duan et al., 2024) introduced Layer-diverse Negative Sampling, which adapts sampling strategies across different GNN layers to capture multi-scale structural information.

In Section 5.6 (Page 14 Line 16—21):

To be specific, Focal loss was originally designed to address the class imbalance problem in object detection and small-sample classification by reducing the relative weight of easy-to-classify samples. However, while focal loss adjusts for sample difficulty, it does not explicitly address inter-class imbalances, which can result in biased decision boundaries favoring majority classes. Apart from that, LDAM loss focuses on mitigating inter-class imbalance by dynamically assigning larger decision margins to minority classes, reducing their generalization error relative to majority classes. But it does not account for sample-level difficulty, potentially overlooking hard examples within a class.

In Section 7 (Page 27 Line 24—26):

5) Exploring the integration of advanced Negative Sampling techniques with our loss framework to further improve the identification of "hard negatives" in highly oscillatory traffic flows.