NATIONAL CENTER FOR UNDERSTANDING FUTURE
TRAVEL BEHAVIOR AND DEMAND

Final Project Report

Investigation of Emerging Sensing and AI/ML Technologies to Enhance the Safety of Vulnerable Roadway Users at Signalized Intersection

BY

Yiqiao Li
Email: yli4@ccny.cuny.edu

Bo Shang
Email: bshang@ccny.cuny.edu

Weicong Feng
Email: wfeng@ccny.cuny.edu

Jie Wei
Email: jwei@ccny.cuny.edu

Camille Kamga
Email: ckamga@utrc2.org

Department of Civil Engineering
The City College of New York
The City University of New York
160 Convent Avenue, New York, NY 10031

May 2026

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. N/A	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Investigation of Emerging Sensing and AI/ML Technologies to Enhance the Safety of Vulnerable Roadway Users at Signalized Intersection		**5. Report Date** May 2026 **6. Performing Organization Code** N/A
7. Author(s) Yiqiao Li, Ph.D., https://orcid.org/0000-0002-2656-9217 Bo Shang, Ph.D., https://orcid.org/0000-0002-5568-1566 Weicong Feng, https://orcid.org/0009-0007-9246-8792 Jie Wei, Ph.D., https://orcid.org/0000-0003-1781-6412 Camille Kamga, Ph.D., https://orcid.org/0000-0002-9223-700X		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Department of Civil Engineering The City College of New York The City University of New York 160 Convent Avenue, New York, NY 10031		**10. Work Unit No. (TRAIS)** N/A **11. Contract or Grant No.** 69A3552344815 and 69A3552348320
12. Sponsoring Agency Name and Address U.S. Department of Transportation University Transportation Centers Program 1200 New Jersey Ave, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final Report, 2024–2026 **14. Sponsoring Agency Code** USDOT OST-R
15. Supplementary Notes N/A		

16. Abstract

Vulnerable roadway users (VRUs)—including pedestrians, cyclists, and micromobility users—face increasing safety risks at signalized intersections. However, traditional crash-based monitoring approaches are limited in their ability to capture near-miss events and behavioral conflicts that precede crashes. This report investigates how emerging LiDAR-based sensing and AI/ML technologies can improve VRU safety observability at signalized intersections through three interconnected research thrusts.

First, a comprehensive survey of infrastructure-based sensing modalities and AI/ML methods is presented to examine the capabilities, limitations, and challenges of existing approaches. Future research directions are identified to improve the robustness, scalability, and effectiveness of AI-powered intersection safety systems. Among the reviewed sensing technologies, LiDAR emerges as a particularly promising primary modality for trajectory-level safety analysis due to its geometric precision and robustness under varying lighting conditions. Second, to evaluate the reliability of LiDAR-based VRU detection under sensor degradation, a systematic study of vertical LiDAR beam loss is conducted across six 3D detection architectures using the KITTI (64-beam) and nuScenes (32-beam) benchmark datasets. The results show that VRU detection performance remains relatively stable up to approximately 20% beam loss but deteriorates rapidly beyond this threshold. Furthermore, contiguous beam loss—commonly caused by sensor occlusion or lens contamination—is found to be substantially more detrimental than dispersed beam loss of equivalent magnitude. These findings provide actionable insights for sensor maintenance, model selection, and operational risk assessment in infrastructure-based sensing systems. Third, a roadside LiDAR-camera data collection site was established at a signalized intersection in New York City, including object-level annotations for pedestrians and cyclists. Building upon this dataset, an end-to-end auditable safety-analysis framework is developed and demonstrated using an 8,000-frame manually annotated roadside LiDAR dataset. The framework integrates 3D detection, multi-object tracking, trajectory refinement, dynamics-aware stabilization, and structured human-in-the-loop quality assurance to transform raw sensor observations into defensible near-miss safety evidence. A case study involving a heavy vehicle-bicycle interaction demonstrates how combined direction-agnostic and longitudinal time-to-collision analyses can reveal lateral-intrusion-dominated conflict mechanisms that single-metric approaches fail to distinguish. Together, these three contributions establish roadside LiDAR as a viable foundation for scalable, interpretable, and auditable VRU safety monitoring at signalized intersections.

17. Key Words	**18. Distribution Statement**
vulnerable roadway users; signalized intersections; traffic safety; AI/ML; roadside sensing; LiDAR; computer vision	No restrictions.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**
Unclassified	Unclassified	58
		22. Price
		N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under Grant No. 69A3552344815 and 69A3552348320 from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

This research was partially supported by the National Center for Understanding Future Travel Behavior and Demand (TBD), a National University Transportation Center sponsored by the U.S. Department of Transportation (USDOT) under grant numbers 69A3552344815 and 69A3552348320. The authors thank the TBD National Center and USDOT for their support of university-based research in transportation, particularly for the funding provided for this project. The authors also acknowledge all project team members and collaborators—including co-authors on the published papers—who contributed to data collection, field deployment, annotation, model development, and manuscript preparation.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Pedestrians, cyclists, and other VRUs remain among the most exposed participants in urban transportation systems, particularly at signalized intersections where turning vehicles, complex signal phases, and dense multimodal interactions create elevated safety risks. Traditional safety monitoring approaches—primarily crash reports and short-duration manual observations—are inherently reactive and unable to capture the near-miss events and behavioral conflicts that often precede serious collisions. This project investigated whether roadside LiDAR sensing, combined with artificial intelligence (AI) and machine learning (ML), can provide continuous, high-resolution, and interpretable safety monitoring capable of supporting proactive intersection safety management.

The project began with a comprehensive assessment of infrastructure-based sensing technologies, including LiDAR, cameras, radar, and thermal sensors, alongside AI/ML methods for detection, tracking, and behavior analysis. The review showed that while each sensing modality offers unique strengths, none independently satisfies all operational requirements for scalable intersection safety monitoring. Cameras provide rich semantic information but are sensitive to poor lighting and adverse weather. Radar performs reliably under challenging environmental conditions but lacks sufficient spatial resolution for detailed VRU trajectory analysis. Thermal sensors support nighttime monitoring but cannot reconstruct accurate motion trajectories. LiDAR emerged as the most promising primary sensing modality because of its ability to capture precise three-dimensional geometry independent of ambient lighting conditions. However, the survey also identified important challenges, including sensor cost, gradual degradation, and the sparse point cloud representation of pedestrians and cyclists. The findings highlighted the importance of future multimodal sensing systems while also demonstrating that LiDAR-first deployments combined with human-in-the-loop validation provide a practical near-term pathway for infrastructure-based safety analytics.

To evaluate the reliability of LiDAR-based VRU detection under real-world operational degradation, the project conducted a systematic investigation of vertical LiDAR beam loss across six state-of-the-art 3D detection architectures using the KITTI and nuScenes benchmark datasets. Simulated degradation experiments demonstrated that VRU detection is substantially more sensitive to sensor degradation than vehicle detection. While vehicle detection remained relatively stable even under severe beam loss, pedestrian and cyclist detection deteriorated rapidly once beam loss exceeded approximately 20%. The study further revealed that contiguous beam loss—commonly caused by dirt, mud, ice, or water occlusion on the sensor aperture—is significantly more damaging than dispersed beam loss of the same magnitude because it creates concentrated geometric blind zones. In addition, beams located in the mid-to-high elevation range were found to be particularly important for accurately capturing pedestrian and cyclist geometry. These findings provide practical guidance for roadway agencies and operators, including the use of approximately 20% beam loss as a maintenance threshold, prioritization of sensor cleaning in critical aperture regions, and consideration of contiguous degradation scenarios during operational risk assessments.

Building upon these findings, the project established a real-world roadside LiDAR-camera data collection site at the signalized intersection of Convent Avenue and West 141st Street in New York City. The deployment integrated a 128-beam Ouster OS-1 LiDAR sensor, a pan-tilt reference camera, and GPU-enabled edge computing infrastructure. Approximately 8,000 LiDAR frames

were manually annotated across four object categories: cars, trucks, bicycles, and pedestrians. Using this dataset, the project developed an end-to-end auditable safety-analysis framework that integrates 3D object detection, multi-object tracking, trajectory refinement, dynamics-aware stabilization, and structured human-in-the-loop quality assurance. Rather than treating AI outputs as unquestionable ground truth, the framework emphasized iterative human review to verify detections, validate trajectories, and refine near-miss evidence through multiple quality assurance rounds.

A detailed heavy vehicle-bicycle near-miss case study demonstrated the value of this interpretable approach. Direction-agnostic time-to-collision metrics identified a critical conflict event, while longitudinal time-to-collision analysis revealed that the event was dominated by lateral intrusion rather than direct rear-end closing behavior. This distinction provided deeper insight into the underlying conflict mechanism and illustrated how relying on a single safety metric can obscure important contextual information needed for effective countermeasure selection.

Overall, the project establishes roadside LiDAR as a viable foundation for scalable, interpretable, and auditable VRU safety monitoring at signalized intersections. The work contributes both methodological and practical insights, including quantified degradation thresholds, sensor maintenance guidance, structured human-in-the-loop validation procedures, and interpretable conflict analysis techniques. Together, these contributions advance the development of infrastructure-based AI safety systems capable of supporting proactive transportation safety management and future connected intersection applications.

1 INTRODUCTION

Signalized intersections are among the most dangerous locations in the urban roadway network for pedestrians and cyclists. Turning conflicts, overlapping signal phases, limited sight lines, and the proximity of fast-moving vehicles to people on foot or on a bicycle create conditions where a momentary failure of attention—by a driver, a pedestrian, or a cyclist—can have irreversible consequences. In the United States, 16% of all pedestrian fatalities and 29% of all cyclist fatalities occur at intersections (National Center for Statistics and Analysis, 2024c, 2024a). In New York State alone, VRUs accounted for 24% of all fatal and serious injury crashes at signalized intersections between 2017 and 2021 (New York State Department of Transportation, 2023). Despite sustained investment in safety programs such as NYC's Vision Zero initiative and the federal Highway Safety Improvement Program, progress has been slow—in part because the data tools available to agencies do not match the scale or resolution of the problem.

The fundamental limitation is that traditional safety monitoring is reactive. Police crash reports document what happened after a collision, not the near-misses, evasive maneuvers, and yielding failures that accumulate at a site long before a crash is recorded. Manual field observation can capture some of this pre-crash behavior, but it is expensive, short in duration, observer-dependent, and impossible to scale across the thousands of intersections that warrant attention (Ismail, Sayed, & Saunier, 2009; Watson, Watson, & Vallmuur, 2015). Agencies trying to prioritize safety investments are therefore working with evidence that may be incomplete: they can see where crashes have already happened, but have limited visibility into where they may be about to happen.

Emerging roadside sensing technology and AI-driven perception methods offer a fundamentally different approach. A fixed sensor deployed at an intersection can observe every interaction that occurs—every pedestrian crossing, every cyclist maneuver, every vehicle turning movement—continuously and without observer fatigue. Machine learning models can detect, classify, and track road users in real time from the resulting data streams, and surrogate safety metrics such as time-to-collision can be computed from the resulting trajectories to flag conflict events before they escalate into crashes. The promise is a shift from reactive crash monitoring to proactive safety auditing: identifying hazardous patterns early enough to act on them.

Among available sensing modalities, LiDAR is a strong candidate for this role at the infrastructure scale. Unlike cameras, it operates independently of ambient lighting and produces precise three-dimensional geometry that supports accurate trajectory reconstruction. Unlike radar, it provides sufficient spatial resolution to detect and classify individual pedestrians and cyclists. Its remaining limitations—cost, sensitivity to weather, and the sparse point clouds it produces for small VRUs—are real constraints that need to be understood and managed carefully. This project takes those constraints seriously: rather than assuming that a deployed LiDAR system will work as well as a laboratory benchmark, it investigates how sensor degradation affects detection performance, what the failure thresholds are, and how an analysis pipeline must be designed to remain trustworthy even when the underlying data is imperfect.

Equally important is the question of *trust*. A safety evidence record that cannot be inspected, challenged, or traced to its source is not useful for agency decision-making. Engineers need to explain their findings to non-specialist reviewers; project managers need to defend prioritization decisions; regulators need to audit the evidence behind safety improvements. This requires not

just automated detection and tracking, but an auditable pipeline in which every step—from raw sensor data to near-miss classification—is transparent, reviewable, and defensible. Building such a pipeline for roadside LiDAR, and demonstrating it on real field data, is the central practical goal of this project.

1.1 Research Contributions

This report advances the application of LiDAR-based sensing and AI/ML methods for VRU safety at signalized intersections through three interconnected research contributions.

The first contribution is a systematic survey of infrastructure-based sensing technologies and AI/ML methods for VRU detection, tracking, and behavioral analysis at signalized intersections (Section 2). The survey evaluates LiDAR, cameras, radar, and thermal sensing against a common set of deployment criteria, examines multi-modal fusion strategies, and maps the behavioral analysis and surrogate safety measure literature. It identifies the key gaps that motivate the subsequent empirical work: no existing infrastructure-side sensing system combines the detection quality, sensor robustness characterization, and auditable pipeline design that practical safety applications require.

The second contribution is a controlled empirical study of how vertical LiDAR beam loss degrades 3D object detection performance, with particular attention to the asymmetry between VRU and vehicle classes (Section 3). Six state-of-the-art detection architectures are evaluated on the KITTI and nuScenes benchmark datasets under progressive beam removal, with up to 100 randomized trials per condition to produce statistically robust estimates. The study establishes a 20% beam-loss maintenance threshold, identifies mid-to-high elevation beams as the most damage-sensitive for VRU detection, and shows that contiguous beam loss is significantly more harmful than dispersed loss of the same magnitude—findings that translate directly into sensor specification and maintenance practice.

The third contribution is an end-to-end auditable roadside LiDAR safety framework, demonstrated on an 8,000-frame manually annotated dataset collected at a signalized intersection in New York City (Section 4). The framework integrates 3D detection, multi-object tracking, trajectory refinement, dynamics-aware stabilization, and a structured iterative human-in-the-loop quality assurance protocol to transform raw LiDAR data into reviewed, defensible near-miss safety evidence. A dual time-to-collision analysis of a heavy vehicle-bicycle conflict reveals a lateral-intrusion mechanism that single-metric approaches cannot distinguish—demonstrating that interpretable conflict diagnostics, not just event flagging, are achievable from roadside LiDAR at the intersection scale.

1.2 Report Organization

Section 2 surveys the technology landscape and analytical foundation, establishing the research gaps this project addresses. Section 3 presents the beam-loss robustness study in full, from methodology and experimental design through results and practitioner guidance. Section 4 presents the NYC intersection dataset, the auditable analysis pipeline, quantitative results across all pipeline components, and a case-study near-miss analysis. Section 5 synthesizes the findings, draws out policy and practice implications, and outlines the most important directions for future work.

2 LITERATURE REVIEW

This section synthesizes the research landscape of AI-enhanced sensing technologies for VRU safety at signalized intersections and establishes the conceptual and technical foundation for the studies presented in Sections 3 and 4. The discussion is primarily based on the project team's published survey, *AI-Enhanced Sensing for Vulnerable Road User Safety at Signalized Intersections: A Survey* (Shang, Li, Amin, Kamga, & Wei, 2025). The survey comprehensively reviewed infrastructure-based sensing modalities, including LiDAR, RGB cameras, radar, thermal sensing, IoT/V2X systems, and emerging multi-modal sensing frameworks, together with AI/ML methods for object detection, trajectory analysis, sensor fusion, behavioral analysis, and crossing-intention prediction.

A central finding of the survey is that no single sensing modality can independently satisfy all operational requirements for reliable VRU safety monitoring at signalized intersections. Cameras provide rich semantic and visual information at relatively low cost but are sensitive to adverse lighting and weather conditions. Radar offers strong robustness under rain and fog while directly measuring object velocity, yet suffers from limited spatial resolution for detailed VRU classification. Thermal cameras support nighttime monitoring but lack sufficient resolution for accurate trajectory reconstruction. LiDAR emerged as a particularly promising primary sensing modality because of its precise 3D geometric perception, centimeter-level positioning accuracy, and independence from ambient lighting conditions. However, the survey also identified several important limitations of LiDAR systems, including higher deployment cost, degradation under heavy rain, sparse VRU point representations, and sensitivity to occlusion and sensor contamination.

The survey further emphasized that future intersection safety systems will likely rely on multi-modal sensing architectures that exploit the complementary strengths of different sensors through sensor fusion and AI-driven analytics. At the same time, the review highlighted the practical trade-offs among sensing accuracy, robustness, interpretability, and deployment cost. In the current study, LiDAR is utilized as the primary sensing modality for trajectory-level safety analysis, while the co-located camera system is primarily used as a human verification and annotation reference rather than for direct sensor fusion. Due to infrastructure deployment constraints, the camera viewing angle was not geometrically aligned with the LiDAR coordinate system, limiting accurate point-level or feature-level fusion between the two modalities. Consequently, the study adopts a LiDAR-first, human-in-the-loop framework in which camera imagery supports manual object verification and trajectory validation to improve the auditability and reliability of the safety analysis pipeline. Future work will investigate more tightly synchronized and geometrically aligned LiDAR-camera configurations to enable robust and cost-efficient multi-modal fusion for infrastructure-based VRU safety monitoring systems.

2.1 Vulnerable Roadway User Safety at Intersections

VRUs—including pedestrians, bicyclists, motorcyclists, and micromobility users—face disproportionate risk in roadway systems. Signalized intersections are especially critical because they concentrate crossing movements, turning conflicts, queue discharge, transit access, and mixed interactions among vehicles and non-motorized users. In the United States, VRUs have represented a growing share of roadway fatalities in recent years. In 2022, 16% of all pedestrian fatalities oc-

curred at intersections (National Center for Statistics and Analysis, 2024c), with even higher shares among cyclists (29%) (National Center for Statistics and Analysis, 2024a) and motorcyclists (36%) (National Center for Statistics and Analysis, 2024b). In New York State, VRUs comprised 24% of all fatal and serious injury crashes at signalized intersections from 2017 to 2021 (New York State Department of Transportation, 2023). Local and federal agencies have responded with comprehensive strategies such as NYC's Vision Zero initiative and FHWA's Highway Safety Improvement Program, yet persistent data gaps in VRU behavior understanding and detection capability continue to hinder effective intervention.

Traditional safety monitoring is poorly matched to this problem. Police crash reports capture post-crash events but miss near-misses, evasive maneuvers, yielding failures, and other precursor events that reveal risk before crashes occur (Ismail et al., 2009; Watson et al., 2015). Manual field observation is labor-intensive, short in duration, and difficult to scale. These limitations, combined with growing urban populations and increased active transportation demand, motivate a shift toward proactive safety analysis using high-resolution sensor data and AI/ML models.

2.2 Surrogate Safety Measures and Conflict Analysis

Because crashes are rare relative to everyday interactions, surrogate safety measures are widely used to infer risk from observed road-user behavior. Common measures include time-to-collision (TTC), post-encroachment time (PET), speed profiles, gap acceptance, exposure, and trajectory-based conflict indicators. Video-based approaches have demonstrated the feasibility of automated pedestrian safety analysis (Ismail et al., 2009), while LiDAR-derived trajectory studies show that infrastructure sensing can support traffic signal performance measures and pedestrian/vehicle trajectory monitoring (Saldivar-Carranza, Zlatkovic, & Stevanovic, 2024).

The safety auditing framework in Section 4 extends this line of work by applying TTC as an interpretable diagnostic within an auditable roadside trajectory pipeline (Shang & Li, 2026). It contrasts direction-agnostic TTC with longitudinal TTC to distinguish lateral-intrusion-dominated interactions from braking-limited interactions—a distinction important for characterizing conflict mechanisms beyond simple proximity alerts.

2.3 Infrastructure-Based Sensing Technologies

The survey compares sensing modalities for VRU detection and monitoring: RGB cameras, thermal cameras, LiDAR, radar, and connected infrastructure (Shang, Li, Amin, et al., 2025). Table 2 summarizes key cost and performance characteristics of each modality.

RGB cameras are low-cost, high-resolution, and semantically rich, supporting mature detection pipelines such as YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016) and Faster R-CNN (Ren, He, Girshick, & Sun, 2015), but are sensitive to lighting, glare, occlusion, and adverse weather. Thermal cameras address nighttime monitoring (Ghose et al., 2019; Kristo, Ivasic-Kos, & Pobar, 2020) but offer lower spatial resolution and struggle with temperature variations. Radar is weather-robust due to its electromagnetic operating principle and directly measures velocity via the Doppler effect; empirical studies confirm that radar maintains $>$95% detection accuracy in heavy rain (40+ mm/h) and dense fog (visibility $<$50 m) (Kang et al., 2025; Sezgin, Vriesman, Steinhauser, Lugner, & Brandmeier, 2023; ?, ?). Under the same conditions, LiDAR performance can drop below 60%,

Table 2: Comparison of Sensing Technologies for VRU Detection at Signalized Intersections

Sensor	Strengths	Limitations	Cost / Frame Rate
RGB Camera	High resolution, rich semantic information	Poor low-light performance, weather-sensitive	\$200–\$1,000 / 30–60 FPS
LiDAR	Precise 3D geometry (1–2 cm accuracy), lighting-invariant	High cost, <60% accuracy in heavy rain	\$4,000–\$20,000 / 5–20 Hz
Thermal	Night vision capability, weather-resistant	Lower spatial resolution, temperature-sensitive	\$1,000–\$5,000 / 30 FPS
Radar	>95% detection accuracy in rain/fog, direct velocity measurement	Low spatial resolution (0.5–1 m at 100 m range)	\$500–\$2,000 / 10–20 Hz

with range reductions exceeding 30% at 40–45 mm/h rainfall and point cloud density decreasing by up to 45% (Kang et al., 2025; Kim, Park, & Kim, 2023). However, radar's spatial resolution of 0.5–1 m at 100 m range limits its ability to distinguish and classify individual VRUs. LiDAR provides precise 3D geometry and operates independent of ambient lighting, making it especially valuable for trajectory reconstruction and geometric conflict analysis (Shang, Li, Amin, et al., 2025). Its constraints—cost, beam resolution, weather sensitivity, and point-cloud sparsity for small VRUs—directly motivate the beam loss study in Section 3.

2.4 LiDAR-Based Object Detection and Point-Cloud Learning

LiDAR-based VRU detection depends on models that transform sparse, irregular 3D point clouds into object classes and bounding boxes. Table 3 summarizes key research contributions in this area.

Table 3: Key Literature on LiDAR-Based VRU Detection

Reference	Contribution
Lu et al. (2023)	Point augmentation strategies to improve 3D VRU detection
Kong et al. (2024)	Improved PointPillars for small 3D object detection
R. Zhang et al. (2025)	CornerPoint3D: corner-based detection for better small-object localization
Nie et al. (2023)	PARTNER: polar representation for LiDAR 3D object detection
Fu et al. (2021)	Improved point pillar architecture tailored to VRU detection
Li et al. (2023)	Point cloud reconstruction for precise VRU classification
L. Zhang et al. (2024)	3D VRU detection from low-resolution LiDAR using PCC and SAM
Saldivar-Carranza et al. (2024)	Multi-LiDAR trajectory monitoring at signalized intersections

Foundational architectures include VoxelNet (Zhou & Tuzel, 2018), which pioneered end-to-end 3D point cloud learning and achieved 89.3% mAP on the KITTI benchmark but required 0.23 s per frame; PointPillars (Lang et al., 2019), which reduced inference to 16 ms through pillar-based encoding, enabling real-time processing; and CenterPoint (Yin, Zhou, & Krahenbuhl, 2020), which introduced center-based detection with 65.5% mAP on nuScenes and superior trajectory consistency for multi-object tracking. Cost trends support broader deployment: solid-state LiDAR units now cost \$4,000–\$8,000, compared to \$15,000–\$20,000 for mechanical systems. Low-resolution sensors, when paired with point cloud completion and segment-based augmentation, can achieve comparable VRU detection at 60–75% cost savings (L. Zhang et al., 2024). Training strategies that

mix synthetic and real data can further improve performance: Jabłoński, Iwaniec, and Zabierowski (2022) demonstrated an F1-score of 0.84 using combined CARLA-generated and real-world training data, outperforming a model trained on real data alone (F1-score of 0.82). Small VRUs can be represented by very few points, making detection sensitive to sensor resolution, viewing angle, occlusion, and degradation—observations central to the beam loss study in Section 3.

2.5 Sensor Fusion and Cooperative Perception

Multimodal sensing combines complementary sensor strengths through three main fusion strategies: data-level fusion (point cloud colorization with camera imagery), feature-level fusion (joint neural network processing of multi-modal representations), and decision-level fusion (independent per-modality detection followed by cross-modal association). Table 4 summarizes representative multi-modal approaches for VRU detection.

Table 4: Representative Sensor Fusion Methods for VRU Detection

Reference	**Sensors**	**Key Characteristics**
Bijelic et al. (2020)	LiDAR, Camera, Thermal, Radar	Multimodal fusion for adverse-weather robustness
Palladin et al. (2024)	LiDAR, Camera, Radar	Sensor-adaptive weighting for 3D VRU detection
Wu et al. (2023)	LiDAR, Camera	Decision-level fusion; LiDAR excels at night, camera by day
Z. Wang et al. (2025)	LiDAR, Camera	Feature-level fusion for joint representation learning

The complementary failure modes of different sensors provide the core motivation for fusion: while radar maintains >95% accuracy in adverse weather conditions where LiDAR performance can fall below 60%, combining both modalities creates robust all-weather detection capabilities. Public datasets such as nuScenes (Caesar et al., 2020) support vehicle-side multi-modal perception research, but a critical gap remains: most existing multimodal datasets are vehicle-side rather than fixed roadside, motivating the infrastructure-side data collection in Section 4.

2.6 Behavioral Analysis and Advanced Applications

AI-enhanced sensing supports behavioral analysis beyond object detection. Mobile device use significantly impairs VRU safety: distracted pedestrians exhibit approximately 20% lower walking speed, increased path deviation, delayed signal response, and higher likelihood of crossing against signals (Nasar, Hecht, & Wener, 2008; De Waard, Schepers, Ormel, & Brookhuis, 2010; Haque, Kidwai, Thapa, Ghani, & Mtapure, 2025). Computer vision systems can detect distracted behavior patterns from RGB imagery (Saenz, Sun, Wu, Zhou, & Yu, 2021), and multi-modal systems can further reinforce this detection through LiDAR-derived speed profiles and radar velocity measurements.

The VRUCrossSafe framework (Abdelrahman, Islam, & Abdel-Aty, 2025) exemplifies the proactive safety potential of AI-enhanced sensing. The system combines YOLOv8 object detection with pose estimation to analyze body posture, head orientation, and trajectory characteristics at 33 FPS, predicting crossing intention before VRUs actually step into the crossing zone. VRUCrossSafe achieves 94.67% crossing intention accuracy with detection typically 2–3 seconds ahead

of actual crossing—sufficient lead time for automated signal activation, connected-vehicle warnings, and safety audit event logging. It also demonstrates a 42.8% improvement in signal cycle efficiency by enabling adaptive timing that responds to actual crossing demand. Smart infrastructure systems extend these capabilities toward active safety support, including pedestrian guidance and adaptive signal control (Kulhandjian, 2024).

2.7 Challenges and Research Gaps

Environmental robustness, occlusion, privacy, cost, and scalability remain persistent challenges across sensing modalities (Bowyer, 2004). Future directions include robust multimodal fusion, edge AI for real-time applications, unified cross-site detection frameworks (Y. Wang et al., 2024), agentic multimodal large language models for explainable safety analysis (Google DeepMind, 2024; Liu, Li, & Wu, 2023), and the integration of autonomous vehicle fleet data to study near-miss scenarios at scale. Table 5 summarizes potential research and deployment trajectories for LiDAR- and Vision-AI-based VRU safety systems across near-term, mid-term, and long-term time horizons. In the near term, research is expected to focus on improving the practicality and deployability of infrastructure-based sensing through low-cost solid-state LiDAR, edge AI processing for low-latency VRU detection, basic V2X-enabled safety alerts, and the use of agentic multimodal large language models (MLLMs) to support more explainable traffic analysis. Over the mid term, advancements are likely to emphasize scalable multi-modal sensing networks, predictive behavioral modeling of VRUs, cooperative infrastructure-vehicle perception systems, and unified cross-domain detection frameworks capable of improving robustness across diverse environments. Looking further ahead, long-term research directions envision fully integrated city-scale digital twins, automated incident prevention systems, self-adapting intelligent intersection infrastructure, and broader zero-fatality transportation ecosystems enabled by tightly coupled sensing, AI reasoning, and connected mobility technologies.

Table 5: Future Research Directions for LiDAR and Vision-AI in VRU Safety

Near-Term (1–2 years)	**Mid-Term (3–5 years)**	**Long-Term (5+ years)**
Low-cost solid-state LiDAR deployment	Multi-modal sensor networks at scale	City-wide digital twin integration
Edge AI for low-latency VRU detection	Predictive VRU behavioral models	Automated incident prevention systems
Basic V2X safety alerts at intersections	Cooperative infrastructure-vehicle systems	Self-adapting intersection infrastructure
Agentic MLLM for explainable traffic analysis	Unified cross-domain detection frameworks	Zero-fatality intersection ecosystems

Three research gaps are particularly relevant to this project. First, there is a growing need for sensing workflows capable of capturing VRU near-miss events and behavioral conflicts at a higher spatial and temporal resolution than traditional crash reports can provide. Second, additional research is needed to better understand the trade-offs between cost, robustness, and detection performance of LiDAR-based VRU safety systems under realistic sensor degradation conditions. Third, future safety applications could benefit from more auditable and interpretable analysis pipelines that integrate object detection, multi-object tracking, structured human review, and surrogate safety measures into a unified framework. Sections 3 and 4 address these gaps directly.

3 LIDAR RELIABILITY AND VRU DETECTION UNDER SENSOR DEGRADATION

3.1 Introduction

The literature review in Section 2 identifies LiDAR beam resolution and sensing reliability as critical deployment considerations, particularly for VRU classes that generate sparse point clouds. Real-world roadside LiDAR systems are subject to progressive sensor degradation: individual beams can fail due to hardware aging, dirt occlusion, water pooling, or contamination on the sensor aperture. When beams are lost, the resulting point cloud is sparser and less geometrically complete, potentially compromising the object detection models that depend on it. Understanding how beam loss affects detection performance—and how that effect differs between vehicles and VRUs—is essential for specifying sensors, setting maintenance thresholds, and selecting detection models for safety-critical deployments.

This section presents a systematic investigation of how LiDAR beam reduction and sensor degradation affect VRU detection performance in infrastructure-based safety applications. The findings consistently show that VRU detection is substantially more sensitive to beam loss than vehicle detection, with detection accuracy remaining relatively stable at moderate degradation levels before deteriorating rapidly beyond critical thresholds. The analysis further demonstrates that degradation patterns matter: contiguous beam loss, which may occur due to sensor occlusion or contamination, produces significantly greater performance degradation than dispersed beam loss of the same magnitude. These findings provide important insights into the robustness limits, maintenance considerations, and operational reliability of LiDAR-based VRU safety systems. The findings presented in this section are based on two peer-reviewed publications produced by the research team (Shang, Li, Wei, & Kamga, 2025; Feng et al., 2026). The earlier study established the sensitivity of VRU detection performance to LiDAR beam density using the KITTI dataset and the PointPillars architecture, while the later journal study expanded the analysis to multiple state-of-the-art 3D detection architectures across both KITTI and nuScenes datasets with more comprehensive statistical evaluation. The present section primarily focuses on the broader journal-scale analysis while referencing the earlier findings where appropriate.

3.2 Methodology

Datasets and Detection Models

The study is conducted on two established benchmark datasets for 3D LiDAR perception. The *KITTI* dataset (Geiger, 2012) provides dense point clouds collected by a Velodyne HDL-64E LiDAR scanner with 64 vertical beams. The training split contains 7,481 labeled frames spanning cars, pedestrians, and cyclists at easy, moderate, and hard difficulty levels. The *nuScenes* dataset (Caesar et al., 2020) uses a Velodyne HDL-32E sensor with 32 channels and a 360° horizontal field of view; the training and validation splits contain 28,130 and 6,019 frames, respectively, across 10 semantic categories including pedestrian, cyclist, motorcycle, and car.

Six state-of-the-art 3D detection architectures are evaluated: **PartA2** employs a hybrid two-stage framework combining point-based and voxel-based processing for precise small-object local-

ization; **PointPillars** transforms point clouds into 2D pseudo-images via vertical pillar encoding for efficient real-time detection; **SECOND** leverages voxelization with sparse 3D convolutions; **PV-RCNN** fuses voxel-level context with point-level precision through voxel-to-point feature aggregation; **SECOND-MultiHead** extends SECOND with parallel class-specific detection heads; and **VoxelNeXt** predicts directly from sparse voxel features in an end-to-end framework. These six models span the major architectural paradigms in modern LiDAR perception and are pre-trained on each dataset without any fine-tuning for degraded inputs, reflecting realistic deployment conditions.

Beam Removal Protocol

Beam loss is simulated by progressively removing complete vertical beams from each point cloud frame at increments of 4 beams. Two beam-identification methods are compared to validate robustness of the simulation to the choice of identification strategy.

Uniform-height-based identification partitions the point cloud into equal-height bins and assigns a beam index to each bin. This method is dataset-agnostic and requires no knowledge of the sensor's angular configuration; it is the primary method used for all comparative experiments reported in this section.

Elevation-angle-based identification uses the sensor's known laser elevation angles to assign each return to its manufactured beam channel, faithfully reproducing the physical beam layout. Comparing the two strategies across all six detectors and all removal counts, AOS and 3D AP degradation curves are visually indistinguishable: the simpler uniform-height proxy reliably reproduces sensor-physics beam assignment for the purpose of degradation analysis (Feng et al., 2026). This correspondence justifies using the computationally cheaper uniform-height method for the large-scale 100-run experiments.

For KITTI (64 beams), removal steps of 4, 8, 12, 16, 20, 24, 28, and 32 beams correspond to removal rates of 6% to 50%. For nuScenes (32 beams), steps of 2, 4, 6, 8, 10, 12, 14, and 16 beams span the same 6%–50% range.

Because the choice of *which* beams to remove at a given count significantly affects performance, the study does not rely on single trials. Instead, at each removal level, beams are randomly sampled and the process is repeated 100 times on KITTI and 10 times on nuScenes (reduced due to the dataset's $\sim$1 TB size). Mean performance and standard deviation are reported for each configuration, providing statistically meaningful estimates rather than anecdotal single-run results.

Evaluation Metrics

$\text{Recall}_{\text{Obj}}$ and $\text{Recall}_{\text{ROI}}$. **$\text{Recall}_{\text{Obj}}$** (referred to as *Recall/RCNN* in the OpenPCDet framework, though the name can be misleading) measures the fraction of annotated ground-truth objects successfully detected in the final model predictions. For single-stage models this is the recall of the entire network; for two-stage models it specifically captures the recall of the final detection head. **$\text{Recall}_{\text{ROI}}$** measures how well the Region Proposal Network covers ground-truth objects—i.e., proposal-level recall before refinement. Both are reported at IoU thresholds of 0.3 and 0.5 (Recall_Obj_0.3 and Recall_Obj_0.5).

Recall-based evaluation is preferred over mean Average Precision in this study for two reasons. First, in transportation safety applications missing an object (false negative) is typically more

consequential than generating a spurious detection (false positive), so maximising coverage—not precision–recall balance—is the primary objective. Second, recall is computationally efficient to compute in real-time monitoring contexts (Feng et al., 2026).

3D Average Precision (3D_R40). The 3D series metrics on KITTI report Average Precision for 3D object detection using the R40 protocol, which evaluates the Precision–Recall curve at 40 uniformly spaced recall positions. Three difficulty levels are defined by truncation, occlusion, and minimum bounding-box height (Table 6).

Table 6: KITTI object difficulty criteria for the Car, Pedestrian, and Cyclist categories (Feng et al., 2026).

Criterion	Easy	Moderate	Hard
Min. bounding-box height (px)	≥ 40	≥ 25	≥ 25
Max. truncation	$\leq 15\%$	$\leq 30\%$	$\leq 50\%$
Max. occlusion	Fully visible	Partly occluded	Heavily occluded

Average Orientation Similarity (AOS_R40). AOS measures the orientation similarity and accuracy between predicted and ground-truth bounding-box angles. The metric captures whether a detector not only localises objects but also predicts their heading direction correctly—a requirement for trajectory prediction and collision-avoidance systems. For a detection set $D(r)$ at recall $r = \frac{TP}{TP+FN}$, the orientation similarity score is:

$$s(r) = \frac{1}{|D(r)|} \sum_{i \in D(r)} \frac{1 + \cos(\Delta\theta_i)}{2} \tag{1}$$

where $\Delta\theta_i$ is the discrepancy between the predicted and ground-truth orientation of detection i. The final AOS score integrates over 40 recall positions:

$$\text{AOS} = \frac{1}{40} \sum_{r \in \{0,\, 0.025,\, \ldots,\, 1\}} \max_{\tilde{r} \geq r} s(\tilde{r}) \tag{2}$$

Difficulty levels and R40 evaluation protocol are the same as for 3D AP above.

AP@0.5m (nuScenes). For nuScenes, performance is reported as **Average Precision at a 0.5 m centre-distance threshold** (AP@0.5m). A detection is counted as a true positive if its predicted centre lies within 0.5 m of the ground-truth centre in the bird's-eye-view plane, irrespective of bounding-box orientation or size; matches with center distance exceeding this threshold are treated as false positives. Because AP metrics at different centre-distance thresholds (0.5 m, 1.0 m, 2.0 m, 4.0 m) are highly correlated across beam-removal conditions, AP@0.5m is used throughout as a representative proxy for observing degradation trends as removed-beam count varies (Feng et al., 2026).

3.3 Results

Variability of Beam-Loss Effects

A key methodological motivation for the randomized multi-run design is that performance varies substantially depending on *which* specific beams are removed, even when the total count is held fixed. Figures 1 and 2 illustrate this variability for PV-RCNN on KITTI. Across ten independent runs each removing four beams, Car AOS/Easy (panel a) shows moderate run-to-run variation; when 32 beams are removed (panel b), the spread between best and worst runs reaches approximately 20 percentage points. This variability motivates the randomized sampling strategy and makes single-run evaluations unreliable for sensor specification decisions.

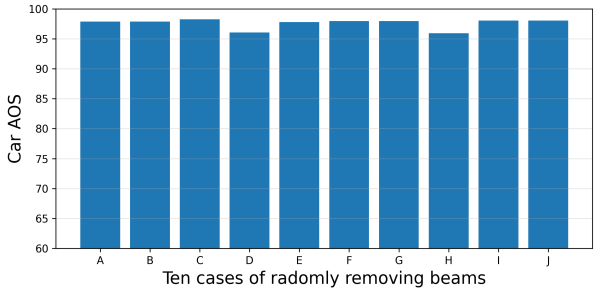


Figure 1: PV-RCNN Car AOS/Easy across ten independent 4-beam removal runs on KITTI. Letters A–J denote individual runs.

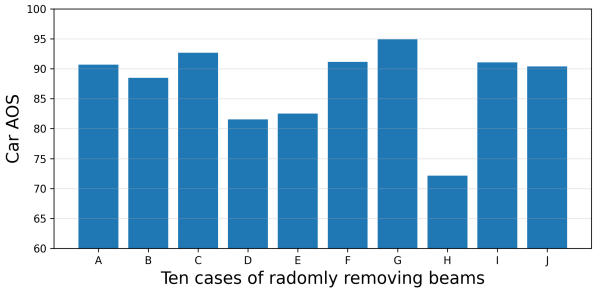


Figure 2: PV-RCNN Car AOS/Easy across ten independent 32-beam removal runs on KITTI, showing up to 20-point spread between best and worst runs.

Beam-Loss Location Sensitivity

Performance degradation is strongly dependent on the vertical location of the missing beams, not only their count. To isolate this effect, the study removes blocks of 4 or 8 beams starting at every possible position in the vertical scan and records the resulting detection performance. Figure 3 shows the progressive loss of pedestrian point-cloud geometry as 0, 8, 16, and 32 beams are removed—demonstrating how rapidly structural cues for human body geometry disappear. Figures 4 and 5 show the SECOND detector's location-sensitivity map for 4-beam and 8-beam block removal, respectively. Each bin on the horizontal axis represents 4 beams; bin 1 is at the bottom (beams 0–3) and bin 16 is at the top (beams 60–63).

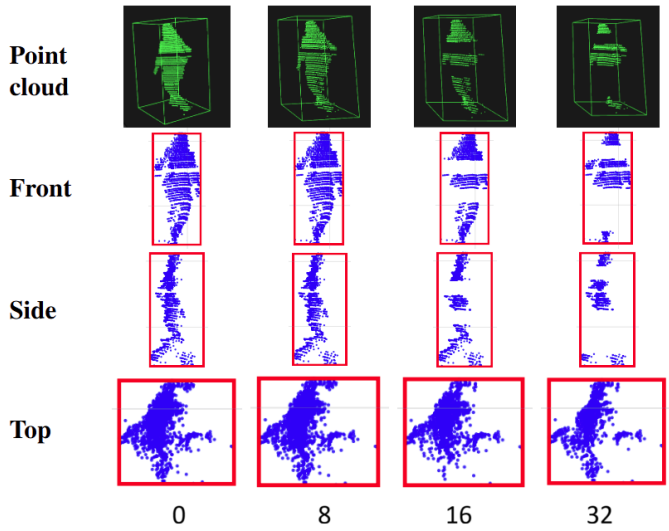


Figure 3: Progressive degradation of a pedestrian point cloud with 0, 8, 16, and 32 beams removed (left to right): raw 3D view, front view, side view, and top view.

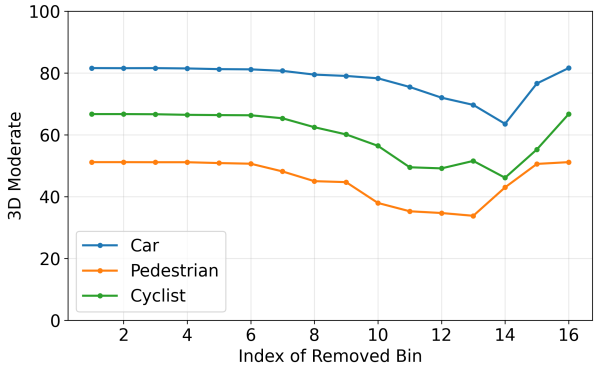


Figure 4: Location sensitivity of SECOND 3D/Moderate with 4-beam block removal. Strongest degradation occurs in high-elevation bins (9–13) corresponding to pedestrian and cyclist torso geometry.

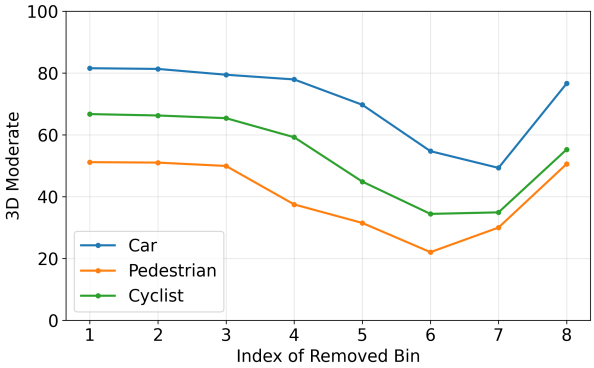


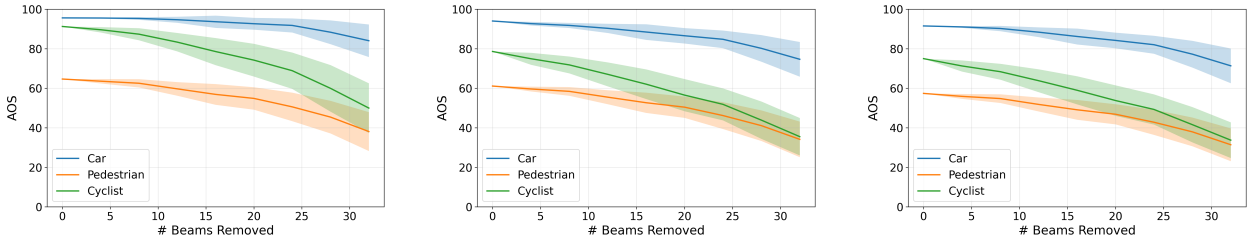
Figure 5: Location sensitivity of SECOND 3D/Moderate with 8-beam block removal. The same high-elevation pattern holds, with steeper absolute drops than the 4-beam case.

For cars, the most severe degradation occurs when bins 11–13 (beams 40–55) are removed; for pedestrians, bins 9–13 (beams 36–55); for cyclists, bins 9–14 (beams 36–59). These mid-to-high elevation beams capture the upper body, head, and frame geometry of VRUs. The 8-beam block experiment confirms the same spatial pattern with steeper absolute performance drops. This finding has a direct implication for maintenance practice: lens cleaning strategies and protection should prioritize high-elevation regions of the LiDAR aperture, and risk assessment should explicitly flag contiguous-loss scenarios—such as mud or water occlusion—because they cluster in the damage-sensitive angular bands.

General Degradation Patterns

Figure 6 presents the complete AOS and 3D AP degradation curves for the SECOND detector on the KITTI dataset as vertical beams are progressively removed from 0 to 32. Lines indicate mean

performance across 100 randomized runs; shaded bands show one standard deviation.



(a) AOS Easy (b) AOS Moderate (c) AOS Hard

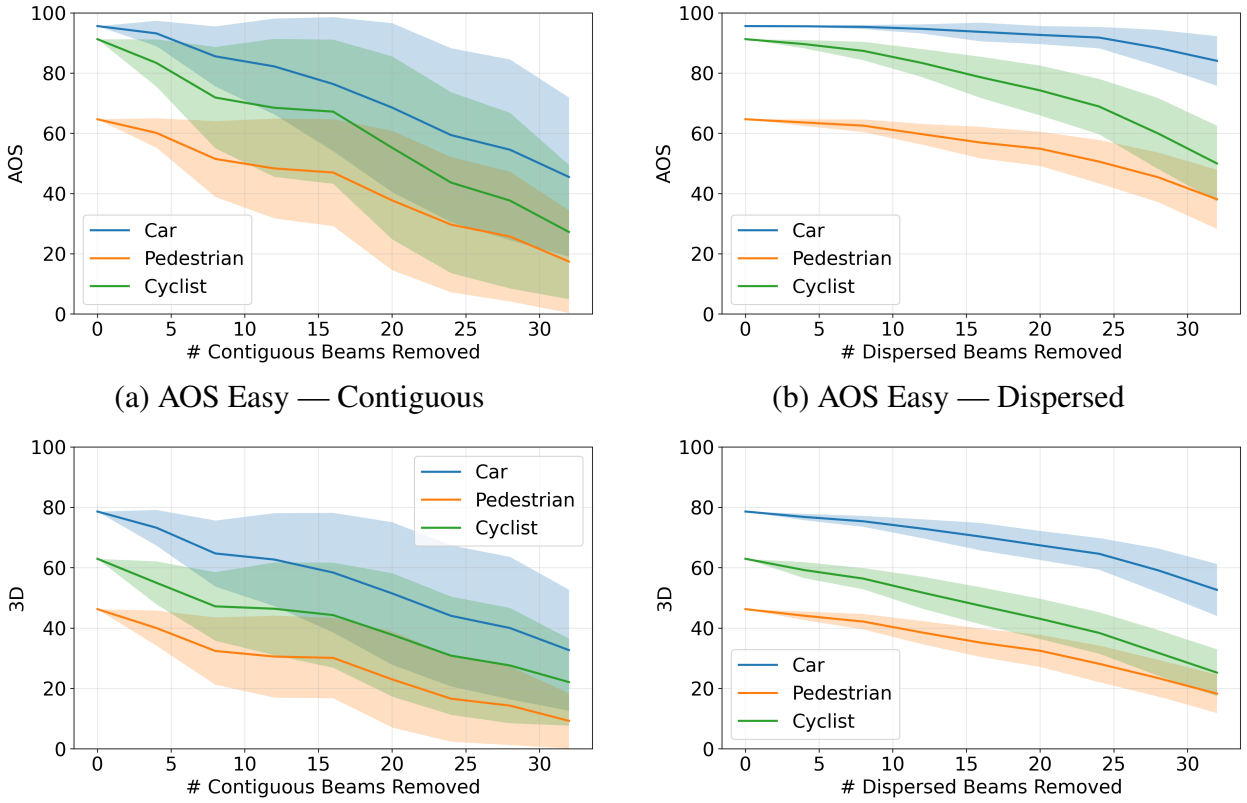
(d) 3D Easy (e) 3D Moderate (f) 3D Hard

Figure 6: AOS (top row) and 3D AP (bottom row) of the SECOND detector on KITTI for easy, moderate, and hard cases. Blue: car; orange: pedestrian; green: cyclist. Lines are mean across 100 runs; bands are one standard deviation.

Three patterns stand out. First, car detection (blue) remains robust throughout: even at 50% beam removal (32 beams), car AOS declines by less than 20% in the easy case, reflecting the large, geometrically distinctive shape of vehicles. Second, VRU detection degrades sharply. Cyclist AOS drops by more than 50% when 32 beams are removed (easy case: 82.94% $\rightarrow$ 40.85%; hard case: 62.92% $\rightarrow$ 25.23%), and the decline is steeper in harder conditions where targets are already partially occluded or distant. Pedestrian AOS also degrades substantially but remains slightly more resilient than cyclist—the latter's bicycle frame provides even fewer point returns than a standing pedestrian, making it more sensitive to any beam reduction. Third, standard deviation widens as more beams are removed: with more possible combinations of missing beams, individual run outcomes diverge, confirming that performance becomes less predictable at high degradation levels.

Contiguous vs. Dispersed Beam Loss

Beyond how many beams are lost and where, the spatial *pattern* of loss also matters. Contiguous loss—where adjacent beams are all disabled, as occurs with mud occlusion or water pooling on the lens—differs from dispersed loss, where missing beams are spread across the vertical scan. Figure 7 compares these two modes for the SECOND detector.



(a) AOS Easy — Contiguous

(b) AOS Easy — Dispersed

(c) 3D Hard — Contiguous

(d) 3D Hard — Dispersed

Figure 7: Comparison of contiguous (left) and dispersed (right) beam loss on SECOND evaluated on KITTI. Top row: AOS/Easy; bottom row: 3D AP/Hard. Contiguous loss consistently produces greater performance degradation and higher variance.

Both loss patterns produce similar qualitative trends—VRUs degrade faster than cars, and degradation accelerates at higher removal counts—but contiguous loss consistently incurs more severe performance drops and higher variance than dispersed loss of the same total beam count. Examining the SECOND detector on KITTI as an illustrative case: at 16 beams removed (25% loss), contiguous removal reduces Cyclist AOS/Easy by approximately 15 percentage points more than dispersed removal; at 32 beams (50% loss), the gap widens further, and the standard deviation under contiguous removal is roughly twice that under dispersed removal. The same ordering holds for Pedestrian AOS and for 3D AP at all difficulty levels.

The physical interpretation is straightforward. Dispersed loss removes one beam at regular angular intervals, preserving the coarse geometric skeleton of an object: the detector still receives point returns spanning the object's full height, just at lower density. Contiguous loss, by contrast, creates a complete vertical gap in the scan—removing all returns from a specific height range and destroying structural cues such as head and shoulder geometry for pedestrians, or wheel and frame geometry for cyclists. This is precisely the failure mode produced by mud occlusion, ice buildup, or standing water on a lens aperture, all of which block a contiguous angular band.

This finding is consequential for maintenance planning and environment-specific sensor specification. Sensors deployed in environments prone to frontal occlusion (mud splatter, ice formation,

low-angle rain, or bird fouling) face a materially more severe risk profile than sensors in open, clean-air deployments, even when the total fraction of lost beams is identical. Maintenance intervals and alert thresholds should therefore be calibrated to the *pattern*, not merely the *count*, of beam failures reported by sensor diagnostics.

Quantitative Thresholds and Cross-Dataset Comparison

Table 7 reports 3D AP at Moderate difficulty (R40) for four detectors on KITTI across four beam-completeness levels. Values are averaged over 100 randomized beam-removal runs.

Table 7: Impact of beam removal on KITTI 3D AP/Moderate (R40). Values are mean AP (%) over 100 randomized runs (Feng et al., 2026).

#	Object Class	PartA2	PointPillars	PV-RCNN	SECOND
100% beams (intact)					
1	Pedestrian	59.76	51.43	55.18	51.18
2	Cyclist	70.16	62.92	69.72	66.72
3	Car	82.95	78.40	84.36	81.59
95% beams (4 beams removed)					
4	Pedestrian	56.73	49.98	52.58	48.92
5	Cyclist	67.44	61.26	67.21	63.00
6	Car	82.31	77.08	82.66	79.90
80% beams (12 beams removed)					
7	Pedestrian	48.99	45.26	46.72	42.87
8	Cyclist	59.42	56.27	59.98	55.17
9	Car	78.93	73.38	80.43	76.39
50% beams (32 beams removed)					
10	Pedestrian	20.39	22.35	21.23	20.48
11	Cyclist	27.92	28.26	28.51	26.97
12	Car	62.49	52.85	62.52	56.82

Across all four detectors, performance remains relatively stable when up to 12 beams (80% intact, approximately 20% removed) are lost, then drops sharply once 32 beams (50%) are removed. The 20% threshold is consistent across models and categories, though VRU categories decline more steeply than cars at every removal level. Notably, PartA2 achieves the highest baseline accuracy for pedestrians (59.76%) and cyclists (70.16%) but also sustains the largest absolute drop; PointPillars starts lower but declines more slowly in relative terms, particularly for cyclists at heavy degradation (28.26% vs. PartA2's 27.92% at 50% beams).

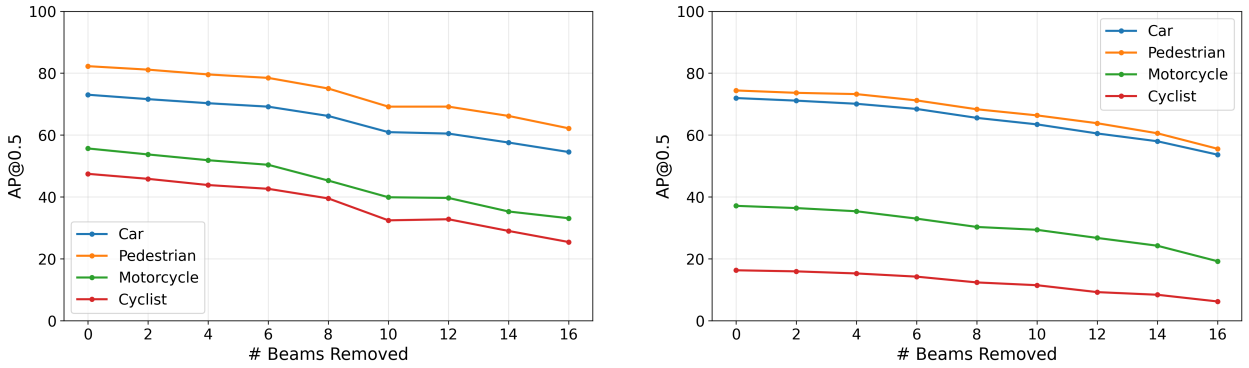
Table 8 reports AP@0.5m on the nuScenes dataset for VoxelNeXt and SECOND evaluated on the 32-beam sensor platform. Values are averaged over 10 randomized runs.

The 20% beam-loss threshold holds on nuScenes: performance is relatively stable from 100% to 80% intact beams, then accelerates its decline. Crucially, cyclist detection on nuScenes is already weaker at baseline—SECOND achieves only 16.31% AP@0.5m under intact conditions—and degrades to 6.23% at 50% beam loss, a relative decline of 62%. This underscores that cyclist detection is fragile on lower-resolution sensors even under ideal conditions and cannot tolerate meaningful degradation.

Table 8: Impact of beam removal on nuScenes AP@0.5m. Values are mean AP (%) over 10 randomized runs (Feng et al., 2026).

#	Object Class	VoxelNeXt	SECOND
100% beams (intact)			
1	Pedestrian	82.26	74.41
2	Cyclist	47.47	16.31
3	Car	73.03	71.96
90% beams (2 beams removed)			
4	Pedestrian	81.14	73.64
5	Cyclist	45.86	15.96
6	Car	71.61	71.12
80% beams (6 beams removed)			
7	Pedestrian	78.48	71.19
8	Cyclist	42.63	14.25
9	Car	69.18	68.44
50% beams (16 beams removed)			
10	Pedestrian	62.19	55.57
11	Cyclist	25.45	6.23
12	Car	54.53	53.67

Figure 8 compares VoxelNeXt and SECOND-MultiHead across all removal levels on nuScenes. While SECOND-MultiHead performs worse than VoxelNeXt at baseline (particularly for cyclists and motorcycles), it shows a slower rate of degradation as beam count decreases—suggesting that multi-head architectures with class-specific detection branches may be more resilient to sensor degradation despite lower peak accuracy.



(a) VoxelNeXt AP@0.5m (b) SECOND-MultiHead AP@0.5m

Figure 8: Performance trends on nuScenes as a function of removed beams for VoxelNeXt (left) and SECOND-MultiHead (right). SECOND-MultiHead exhibits slower degradation despite lower baseline accuracy.

Figure 9 compares all four KITTI detectors side by side at intact, 8-beam removed, 20-beam removed, and 32-beam removed conditions. The upper-left panel confirms that most models achieve similar performance on intact data except PointPillars, which starts lower. The lower-right panel

shows that at 32-beam removal PointPillars converges toward the performance of other models—its relative degradation is smaller than for the higher-baseline models—confirming the resilience advantage of its pillar-based encoding under sparse inputs.

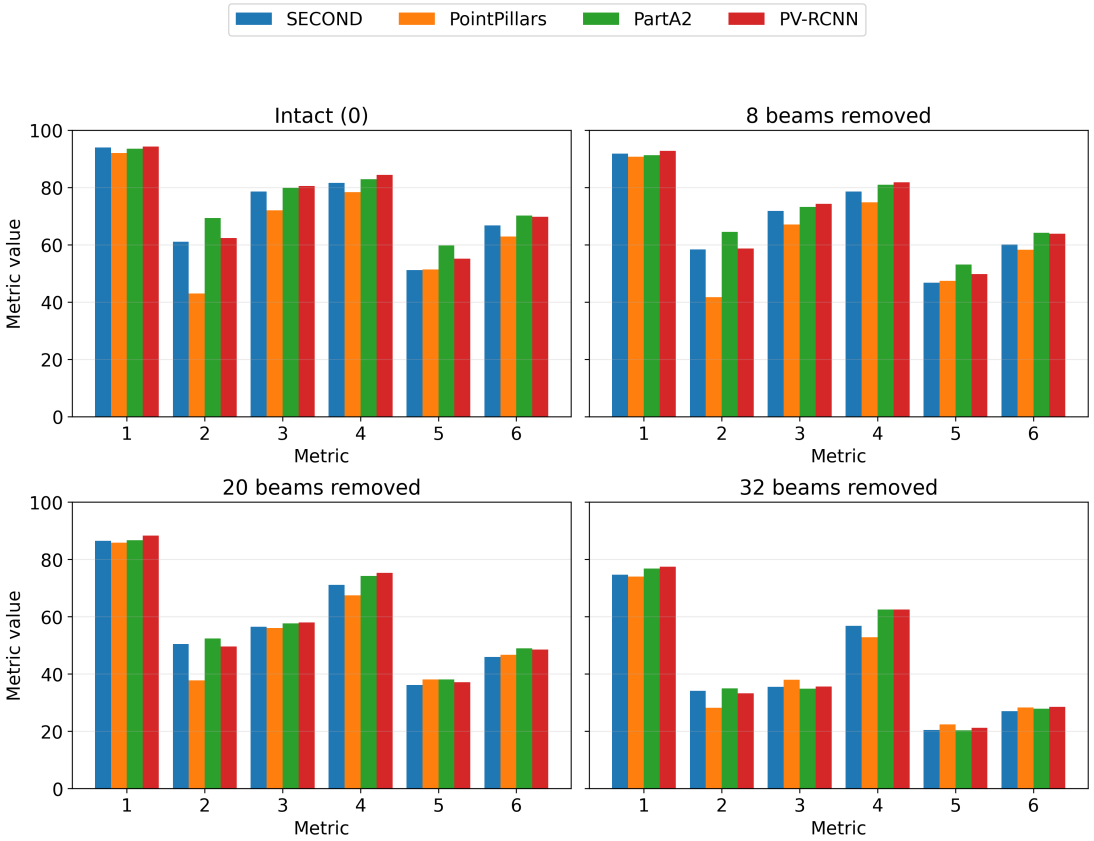


Figure 9: Performance comparison of four detectors (PartA2, PointPillars, PV-RCNN, SECOND) on KITTI at intact and three beam-removal levels. Metrics: Car AOS, Pedestrian AOS, Cyclist AOS, Car 3D, Pedestrian 3D, Cyclist 3D.

3.4 Conclusion

This study yields the following conclusions for agencies planning or maintaining roadside LiDAR systems:

- **VRUs are disproportionately affected.** Car detection remains relatively stable up to 50% beam loss, while cyclist AOS drops by more than 50% at the same removal level. Sensor configurations adequate for vehicle monitoring may fall below acceptable thresholds for VRU safety at the same degradation level.
- **The 20% beam-loss threshold.** Detection performance across all detectors and both datasets remains relatively stable until approximately 20% of beams are lost. Beyond this point, VRU detection degrades rapidly and variance increases substantially. Agencies should treat 20% beam loss as a maintenance trigger for roadside LiDAR systems monitoring pedestrians and cyclists.

- **Location matters as much as quantity.** High-elevation beams (bins 9–14 in a 64-beam system) corresponding to human torso and head geometry are the most damage-sensitive. Maintenance prioritization and risk assessment should focus on mid-to-high elevation aperture regions.
- **Contiguous loss is more harmful than dispersed loss.** Sensors in occlusion-prone environments (mud, ice, rain) face a higher effective risk than the raw beam-count fraction alone would suggest. System-level risk assessments should include contiguous-loss scenarios.
- **Higher resolution does not mean more robust.** Comparing KITTI (64-beam) and nuScenes (32-beam) results, higher-resolution sensors show no systematic advantage in relative resilience to beam loss. Raw beam count is not a reliable proxy for system reliability.
- **Model selection depends on deployment context.** PartA2 may be preferred when maximum VRU detection accuracy under intact or lightly degraded conditions is the primary objective. PointPillars may be a suitable choice for deployments where sensor reliability is uncertain or maintenance intervals are long, given its comparatively slower relative degradation. For nuScenes-type 32-beam environments, VoxelNeXt achieves higher baseline VRU accuracy; SECOND-MultiHead degrades more slowly and may be preferable in harsh environments.

These findings directly informed the design decisions in Section 4: the 20% threshold set the baseline for sensor health monitoring, the identification of high-elevation beams as the most damage-sensitive zone guided sensor mounting and aperture protection protocols, and the relative resilience ranking between model architectures motivated the selection of CenterPoint as the primary detection backbone for the NYC intersection deployment.

4 AN END-TO-END ROADSIDE LIDAR SAFETY AUDITING FRAMEWORK

4.1 Introduction

Autonomous driving and connected infrastructure systems have advanced rapidly in 3D detection, tracking, and scene understanding, yet most deployed perception pipelines remain fundamentally single-vehicle-centric and are constrained by occlusion, limited viewpoints, and partial observability. These limitations are especially acute at urban intersections, where safety-critical interactions often unfold across crosswalks, turning lanes, and transient blind zones. Vehicle-to-Everything (V2X) and infrastructure-assisted perception offer a complementary paradigm by extending observability beyond the ego vehicle. This study addresses a key subproblem within this broader infrastructure intelligence agenda: how can roadside sensing provide *auditable* and *interpretable* safety evidence for reviewing safety-critical interactions at intersections? Auditability is studied in the context of evidentiary transparency and reviewability—the ability of reviewers to inspect, challenge, and trace the safety evidence record—rather than encompassing broader V2X trustworthiness dimensions such as communication robustness or cybersecurity.

The literature review in Section 2 identifies a critical gap in the current research landscape: most existing infrastructure-side sensing datasets and perception systems are oriented toward cooperative perception and broad-scale training rather than auditable safety analysis. Transportation agencies need not just detection and tracking outputs but defensible evidence records—traceable, interpretable, and reviewable by both engineers and non-specialist decision-makers—that can support project prioritization, safety audit, and regulatory review.

This section presents an end-to-end roadside LiDAR safety-auditing framework for infrastructure-assisted analysis at a signalized urban intersection in New York City. The work was published by Shang and Li as *Roadside LiDAR for Cooperative Safety Auditing at Urban Intersections: Toward Auditable V2X Infrastructure Intelligence* (Shang & Li, 2026) at the CVPR DriveX Workshop (2026). The paper makes four contributions: (1) an end-to-end roadside LiDAR pipeline for auditable safety analysis at a real-world urban intersection; (2) an 8,000-frame manually annotated roadside LiDAR dataset with frame-level cuboid labels plus trajectory-level QA artifacts; (3) a demonstration of how structured human review rounds transform raw detection and tracking outputs into a defensible near-miss analysis workflow; and (4) the use of longitudinal TTC contrasted against direction-agnostic TTC as an explanatory diagnostic for distinguishing braking-limited and lateral-intrusion-dominated conflict mechanisms. The primary contribution lies not in a novel perception backbone but in a system design and validation methodology showing how standard perception components can be constrained, stabilized, reviewed, and interpreted to produce defensible safety evidence.

4.2 Related Work

Cooperative perception and vehicle–infrastructure systems aim to reduce observability gaps at complex intersections through roadside sensing and shared scene context (Ji et al., 2024; Huang et al., 2023; Yazgan et al., 2024). Roadside LiDAR has been applied to intersection monitoring, traffic participant tracking, and occlusion-aware assistance at signalized sites (J. Zhang et

al., 2026; Lin et al., 2023; Mo et al., 2024). Surrogate safety measures such as TTC and Post-Encroachment Time (PET) are established tools for conflict analysis when crashes are too rare for direct statistical estimation (Gettman & Head, 2003; Hydén, 1987; Johnsson et al., 2021); TTC was originally formulated for longitudinal car-following settings and later generalized to unconstrained two-dimensional motion (Jansson, 2005; Ward, Agamennoni, Worrall, Bender, & Nebot, 2015).

Table 9 positions this work relative to representative infrastructure-side datasets and frameworks. Existing V2X and cooperative perception resources such as DAIR-V2X (?, ?), V2X-Seq (Yu et al., 2023), TUMTraf-V2X (Zimmer et al., 2024), and UrbanIng-V2X (Sekaran et al., 2025) provide large-scale multi-sensor data for cooperative perception, forecasting, and multi-site detection benchmarks. However, none of these resources incorporate structured human review, conflict auditing, or defensible near-miss evidence production as primary objectives. There remains a lack of an auditable pipeline that unifies roadside trajectory construction, structured human review, and near-miss interpretation into a transparent and reviewable artifact. The present work is distinguished by its emphasis on auditable trajectories, iterative human-in-the-loop QA, reviewed conflict interpretation, and hotspot-oriented safety analysis at a single well-characterized intersection.

Table 9: Positioning of this framework relative to representative infrastructure-side datasets (Shang & Li, 2026).

Resource	Scale	Sensing	Annotation	Primary tasks	Audit/QA
DAIR-V2X (?, ?)	Large real-world V2X benchmark	Vehicle + infra multi-view	3D object labels	Cooperative 3D detection	No
V2X-Seq (Yu et al., 2023)	Sequential V2X scenes	Vehicle + infra sequential	Sequence labels	Perception, forecasting	No
TUMTraf-V2X (Zimmer et al., 2024)	Urban intersection scenes	Multi-sensor V-I setup	Scene-level labels	Cooperative perception	No
UrbanIng-V2X (Sekaran et al., 2025)	Multi-intersection benchmark	Multi-site V–I sensing	Benchmark labels	Multi-intersection perception	No
This work	Single-site pilot deployment	Fixed roadside LiDAR-first	3D boxes, trajectories, QA artifacts	Detection, tracking, audited near-miss analysis	**Yes**

4.3 Dataset Development

Study Site

The main field site is the signalized intersection at Convent Ave & W 141st St near the City College of New York, selected to support observation of diverse roadway users in a complex urban

environment with street parking, transit stops, and frequent pedestrian-vehicle interactions. The deployment uses a fixed roadside Ouster OS-1-128 LiDAR sensor (128 vertical beams) operating at 10 Hz in a fixed elevated oblique-view configuration that covers the intersection center and all four crosswalk approaches. A Pan-Tilt camera is co-deployed on the same sensor pole to provide synchronized video for human-in-the-loop annotation validation and quality assurance. A GPU-enabled edge computing unit, housed at the base of the sensor installation, handles on-site data logging and pre-processing. All object positions, trajectories, and near-miss measurements are reported in a common bird's-eye-view (BEV) coordinate system aligned to this intersection and held fixed throughout the study. Figure 10 shows the physical deployment. Prior to the main deployment, a temporary on-campus synchronized LiDAR-camera setup produced approximately 45 minutes of data for system training and validation. Table 10 summarises the two collection sites.

Table 10: Study site characteristics

Site	Location	Sensor(s)	Collection period	Notes
Temporary on-campus setup	CCNY campus area	LiDAR and camera	Approximately 45 minutes	Used for system training and validation
Main urban intersection site	Convent Ave & W 141st St, near CCNY	Ouster OS-1-128 LiDAR at 10 Hz (primary); camera (validation)	Field deployment period	Fixed elevated oblique-view; 8,000 annotated LiDAR frames

Figure 10: Roadside sensing deployment at Convent Ave and W 141st St, New York City. The 128-beam Ouster OS-1 LiDAR and Pan-Tilt camera for ground-truth validation are mounted on the sensor pole; a GPU-enabled edge computing unit is installed at the base for on-site data logging and pre-processing.

Data Collection and Sensors

LiDAR point-cloud data served as the primary source for roadside perception and safety auditing. Video was collected in synchronization with LiDAR and used for human-in-the-loop validation and annotation—specifically, to check LiDAR detections, resolve ambiguous object classes, interpret trajectories, and support quality assurance. The field dataset encompasses approximately 45 minutes of roadside LiDAR data from the main intersection site. Table 11 summarises all data sources, collection periods, and key variables used across this study.

Table 11: Data sources and variables

Data source	Unit of observation	Time span	Key variables	Processing status
Public benchmark datasets (Section 3)	LiDAR frames/sequences	Beam-count and beam-loss studies	3D point clouds, object labels, beam-loss variants	Used for robustness evaluation
Temporary on-campus collection	LiDAR-camera frames	Approximately 45 minutes	Point clouds and video	Used for system training and validation
CCNY urban intersection dataset	LiDAR frames	Field deployment	Cars, trucks, bicycles, pedestrians; trajectories; validation video	Approximately 8,000 frames manually annotated

Annotation Pipeline

The annotation workflow combines automated model assistance with structured human review. Pre-trained detection and tracking models accelerated the ground-truthing process. Human annotators then reviewed and corrected model outputs using synchronized video as a validation reference. This human-in-the-loop pipeline supports class labeling, trajectory checking, and resolution of ambiguous roadway users. Approximately 8,000 LiDAR frames were manually annotated across four classes: cars, trucks, bicycles, and pedestrians (Shang & Li, 2026). Figure 11 shows a representative frame with human-labeled 3D cuboid annotations for a truck and bicycle in bird's-eye view. Qualitative detection results on the annotated dataset are shown later in Figure 13 in the Results section. Quality control proceeds in two stages: (1) human-reviewed ground-truth frame-label review used for detector training and held-out evaluation, and (2) trajectory-level re-review applied specifically to audited interaction windows for near-miss validation. This two-stage approach distinguishes the dataset from benchmark-only infrastructure datasets by combining frame-level labels with trajectory-QA artifacts and reviewed near-miss analysis.

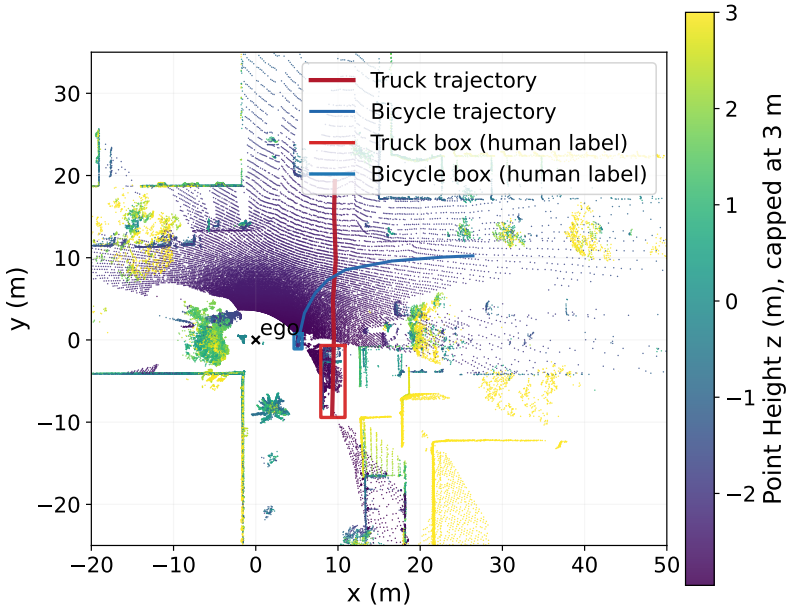


Figure 11: Roadside LiDAR point cloud in bird's-eye view with human-labeled 3D cuboid annotations for a truck (larger box, upper area) and a bicycle (smaller box, lower area). This frame is representative of the annotation style used across all 8,000 labeled frames.

4.4 Methodology

Framework Architecture

Figure 12 summarizes the perception-to-audit workflow. Raw LiDAR point clouds enter a four-stage pipeline. **Stage 1 (Detection):** each frame is processed by a CenterPoint 3D detector to produce per-frame object detections with class labels, confidence scores, 3D bounding boxes, and heading angles. **Stage 2 (Tracking):** detections are associated across frames using a SORT-style Kalman filter tracker to produce identity-consistent tracklets. **Stage 3 (Trajectory Refinement):** raw tracklets are stabilized using registration-guided refinement (B1) and dynamics-aware smoothing to suppress orientation jitter and enforce physically plausible motion. **Stage 4 (Safety Analytics):** refined trajectories are mined for near-miss events using TTC and BEV separation thresholds, and candidate events are interpreted using both direction-agnostic and longitudinal TTC. A cross-cutting human-in-the-loop quality assurance layer acts as a supervisory validation layer throughout all stages, ensuring that the final near-miss evidence records are traceable, inspectable, and defensible.

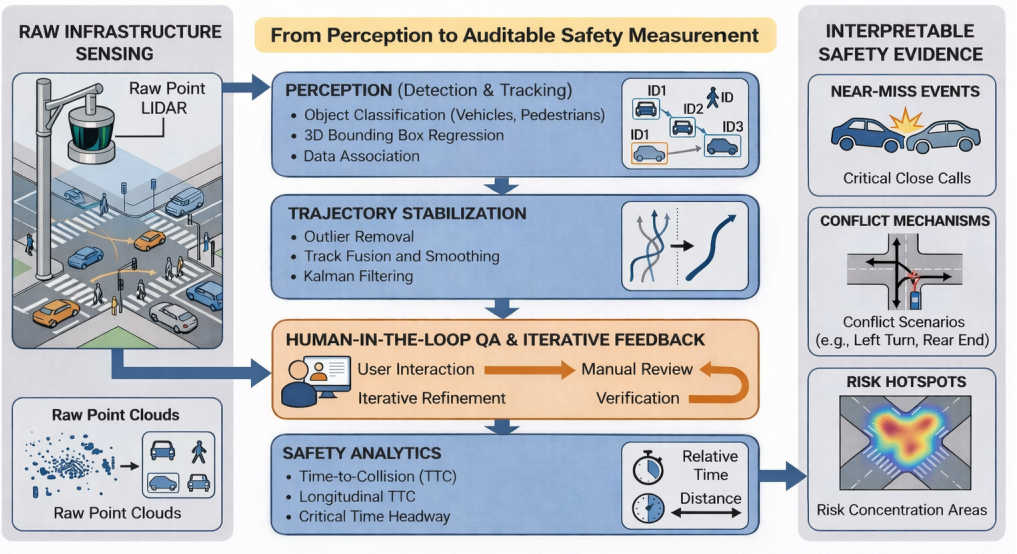


Figure 12: Conceptual overview of the roadside LiDAR safety-auditing pipeline. Raw point clouds are converted into detections, trajectories, and stabilized motion estimates, then reviewed through a human-in-the-loop QA layer and interpreted with surrogate-safety analytics to produce auditable safety evidence.

3D Object Detection

Each LiDAR frame is processed by a fine-tuned CenterPoint detector trained on the 8,000-frame annotated dataset. For each detected object the model outputs a class label c, confidence score s, 3D bounding box center (x, y, z), box dimensions (d_x, d_y, d_z), and heading angle ψ. Car and truck detections are generally strong; pedestrian-scale and low-overlap cases remain more fragile, consistent with the sparsity challenge characterized in Section 3.

Multi-Object Tracking

Detections are associated across frames using a SORT-style tracking framework with Kalman filtering and bird's-eye-view (BEV) gating. This step establishes temporally consistent track identities as the basis for trajectory-level analysis. A predicted box is counted as a match when its BEV center lies within 1.5 m of the corresponding ground-truth center.

Trajectory Refinement and Dynamics-Aware Stabilization

While multi-object tracking provides temporally consistent identities, raw trajectories exhibit frame-level noise, orientation jitter, and occasional misalignment that can propagate into TTC estimates and compromise interpretability. Two levels of refinement address this.

Three post-tracking refinement branches are maintained under identical upstream detections and audited frame windows. **B0** retains raw tracked trajectories as a baseline. **B1** introduces selective correction: suspicious tracklets are flagged based on yaw-step discontinuities, registration disagreement (Kalman-predicted vs. observed state mismatch), and a composite thresholded suspicion score; capped position and orientation corrections are applied only when registration quality is sufficient, anchoring adjustments to reliable observations. **B2** applies stronger frame-wise temporal smoothing enforcing higher consistency but potentially attenuating sharp valid motion changes.

B1 is used as the practical default because it preserves pair-level comparisons while reducing orientation noise.

A separate dynamics-aware stabilization module suppresses high-frequency noise while enforcing physically plausible motion. Position is stabilized using a centered moving-average followed by a constant-velocity prior:

$$\bar{\mathbf{p}}_t = \frac{\sum_{\tau \in \mathscr{W}_t} w_{t,\tau}\, \mathbf{p}_\tau}{\sum_{\tau \in \mathscr{W}_t} w_{t,\tau}}, \qquad \hat{\mathbf{p}}_t = \tilde{\mathbf{p}}_{t-1} + \mathbf{v}_{t-1}\Delta t, \qquad \tilde{\mathbf{p}}_t = (1-\alpha)\,\bar{\mathbf{p}}_t + \alpha\,\hat{\mathbf{p}}_t, \tag{3}$$

Here $\bar{\mathbf{p}}_t$ is the temporally smoothed position, $\hat{\mathbf{p}}_t$ is the constant-velocity prediction, and $\tilde{\mathbf{p}}_t$ is the final blended estimate. $\mathscr{W}_t$ is a centered temporal window of default width 9 frames; $w_{t,\tau}$ are Gaussian weights centered at t that down-weight distant frames; $\mathbf{v}_{t-1}$ is the finite-difference velocity at the previous frame; $\Delta t = 0.1$ s is the inter-frame interval at 10 Hz; and $\alpha \in [0,1]$ is a fixed blend coefficient that trades off between the smoothed observation (low α, more responsive to measurement) and the constant-velocity prediction (high α, more temporally stable). Heading ψ_t is stabilized analogously using circular averaging to avoid angular wrap-around discontinuities, with an adaptive blend coefficient $\beta_t \in [0,1]$ that increases as a function of estimated speed: at low speed the heading prediction is unreliable, so β_t is small and the raw measurement is trusted; at higher speed the constant-velocity heading extrapolation is reliable, so β_t approaches 1. Object dimensions (d_x, d_y, d_z) are stabilized component-wise via robust ℓ_1 estimation across the track's temporal window, rejecting outlier box-size estimates caused by partial occlusion.

Safety Analytics: TTC Formulation

Candidate pairs are first screened by direction-agnostic TTC and minimum BEV separation to identify events of interest, then interpreted with longitudinal TTC to distinguish braking-limited risk from lateral-intrusion-dominated interactions. The two metrics answer complementary questions: direction-agnostic TTC asks "are these two objects closing on each other?"; longitudinal TTC asks "is the lead vehicle being overtaken along its own heading direction?"

For a candidate pair indexed by $i \in \{1,2\}$, let $\mathbf{p}_i$ and $\mathbf{v}_i$ denote the BEV position and velocity of object i; $\Delta\mathbf{p} = \mathbf{p}_2 - \mathbf{p}_1$ is the relative position vector and $\Delta\mathbf{v} = \mathbf{v}_2 - \mathbf{v}_1$ is the relative velocity. The effective radius of each object is $r_i = \frac{1}{2}\sqrt{d_{x,i}^2 + d_{y,i}^2} + \beta$, where $d_{x,i}$ and $d_{y,i}$ are the BEV box dimensions and $\beta = 0.3$ m is a fixed buffer added to account for box-size uncertainty. The signed size-adjusted separation between the two objects is:

$$s_t = \|\Delta\mathbf{p}(t)\| - \big(r_1(t) + r_2(t)\big), \tag{4}$$

so that $s_t < 0$ indicates estimated overlap between the two bounding boxes. Direction-agnostic TTC estimates the time until the objects' separation reaches zero under constant-velocity assumptions:

$$\mathrm{TTC}(t) = \begin{cases} \dfrac{-\Delta\mathbf{p}(t)\cdot\Delta\mathbf{v}(t)}{\|\Delta\mathbf{v}(t)\|^2}, & \Delta\mathbf{p}(t)\cdot\Delta\mathbf{v}(t) < 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{5}$$

The condition $\Delta\mathbf{p}(t)\cdot\Delta\mathbf{v}(t) < 0$ holds when the objects are moving toward each other (the relative

velocity has a component pointing against the relative position vector); when they are diverging, TTC is defined as $+\infty$.

To distinguish conflict mechanisms, longitudinal TTC ($\mathrm{TTC}_{\parallel}$) is computed along the heading direction of the heavy vehicle. Let $\mathbf{p}_h$ and $\mathbf{v}_h$ denote the position and velocity of the heavy vehicle, and $\mathbf{p}_b$ and $\mathbf{v}_b$ the position and velocity of the bicycle. The unit heading vector is $\hat{\mathbf{e}}_{\parallel} = \mathbf{v}_h / \|\mathbf{v}_h\|$. The longitudinal separation and relative closing speed along this axis are:

$$d_{\parallel}(t) = \big(\mathbf{p}_b(t) - \mathbf{p}_h(t)\big) \cdot \hat{\mathbf{e}}_{\parallel}, \qquad v_{\parallel}(t) = \big(\mathbf{v}_b(t) - \mathbf{v}_h(t)\big) \cdot \hat{\mathbf{e}}_{\parallel}, \tag{6}$$

$$\mathrm{TTC}_{\parallel}(t) = \begin{cases} d_{\parallel}(t) \,/\, -v_{\parallel}(t), & v_{\parallel}(t) < 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{7}$$

$\mathrm{TTC}_{\parallel}$ measures only the longitudinal closing component; a conflict where the bicycle crosses laterally into the truck's path will show a low direction-agnostic TTC but a large (or infinite) $\mathrm{TTC}_{\parallel}$, directly exposing the lateral-intrusion mechanism.

Human-in-the-Loop Quality Assurance

Structured QA is a defining feature of the framework. Automated review queues export short tracklet windows with aligned point-cloud context, box overlays, and candidate-event metadata, so reviewers inspect the same evidence used by downstream mining. Each queued item is checked for identity continuity, pose plausibility, split/merge failures, and whether the apparent event reflects a true conflict, a TTC misuse case, or a geometry artifact. Reviewer outcomes are stored as structured correction records—keep/reject decisions, dominant failure tags, and required track corrections—rather than free-form notes, so they can be fed back into the pipeline systematically.

4.5 Experimental Setup

The CenterPoint detector was fine-tuned on the 8,000-frame annotated dataset, with a held-out human-reviewed subset used for evaluation. Detection performance is reported as 3D Average Precision at IoU threshold 0.70 (R40). Tracking performance is evaluated on continuous frame windows against ground-truth trajectories stitched across consecutive frames, with a BEV center-distance matching threshold of 1.5 m. Three trajectory refinement branches (B0, B1, B2) are evaluated to characterize the trade-off between geometric stability and recall. Stabilization parameters include a temporal window of 9 frames and a constant-velocity blend coefficient α; near-miss candidate pairs are gated by a TTC threshold of 2 s and a BEV separation of zero margin.

Three forms of evidence are reported: (1) detector performance against human-reviewed ground truth; (2) branch-level tracking and refinement comparisons on audited and continuous frame windows with stitched ground-truth trajectories; and (3) human cross-checks for reviewed near-miss candidates. This deployment is intentionally single-site: by fixing the sensor, calibration frame, and site geometry, the study establishes a controlled audit reference rather than a general-purpose multi-site benchmark. The results should therefore be read as feasibility and auditability evidence for a single-site pilot rather than a comprehensive cross-site generalization study.

4.6 Results

Detection Performance

Table 13 summarizes quantitative results across pipeline components. Vehicle AP at IoU 0.70 reaches 76.38%. Class-specific native AP is 89.75% for car, 54.55% for truck, and 60.10% for pedestrian, at a throughput of 38.31 fps.

Figure 13 shows qualitative detection examples: a high-overlap truck match, a high-overlap car match, a car case with localization offset, and a single-pedestrian case with moderate localization offset. This illustrates both the deployment-specific strengths and the residual fragility of small-object detection in infrastructure-side geometry.

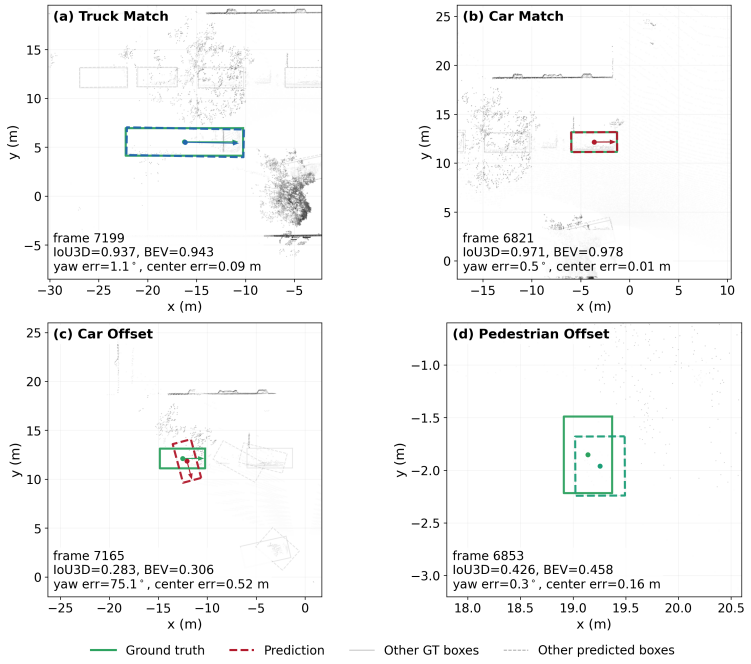


Figure 13: CenterPoint detection examples from the 8,000-frame model: (a) high-overlap truck match, (b) high-overlap car match, (c) car case with localization offset, (d) single-pedestrian case with moderate localization offset.

Tracking Performance

Table 12 reports tracking performance on continuous frame windows. Performance is highest on the shorter first-2,000-frame slice and declines on longer windows, primarily through accumulated false positives and long-tail missed detections rather than catastrophic geometric instability. These results underscore the importance of the trajectory refinement and QA stages for producing reliable safety evidence.

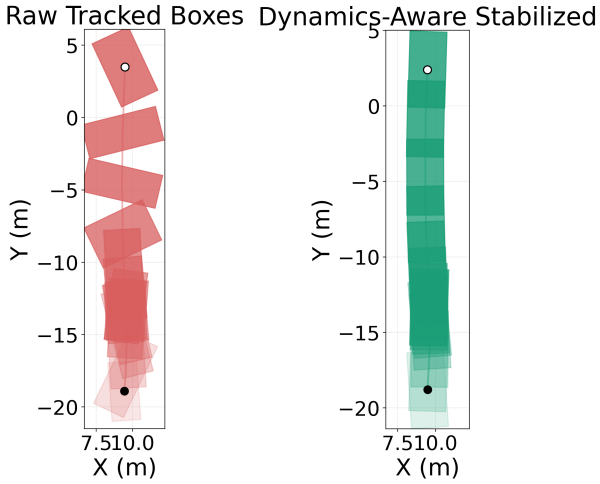
Trajectory Refinement and Stabilization

As reported in Table 13, B1 preserves the audited F1 of B0 (0.9825) while modestly reducing yaw jitter from 15.19° to 14.04°; B2 achieves stronger smoothing (3.92° yaw_{p95}) at a small cost in F1 (0.9780).

Table 12: Tracking evaluation on continuous frame windows with stitched ground-truth trajectories (Shang & Li, 2026).

Window	Precision	Recall	F1	Yaw p_{95}
First 2,000 frames	0.6947	0.9858	0.8150	3.88°
First 4,000 frames	0.5734	0.8677	0.6905	6.50°
Sequential validation slice	0.6036	0.6726	0.6362	3.47°

Figure 14 shows the dynamics-aware stabilization effect on a representative car track: raw tracked geometry (left) exhibits visible jitter in box orientation, whereas the stabilized output (right) enforces consistent heading and dimension estimates. Heading-motion error decreases from 10.45° to 2.36° and the heading-step measure drops from 0.0251 to 0.0040 rad.



(a) Raw tracked trajectory

(b) Dynamics-aware stabilized trajectory

Figure 14: Effect of dynamics-aware stabilization on a representative car track. The stabilized output (b) suppresses orientation jitter and enforces consistent box geometry relative to the raw tracking output (a).

Table 13: Compact quantitative summary of pipeline and audit-layer components (Shang & Li, 2026).

Component	Result
Detector (human-reviewed GT holdout)	Vehicle AP@0.70 (R40): 76.38%; Car: 89.75%, Truck: 54.55%, Pedestrian: 60.10%; 38.31 fps
B0 baseline refinement	F1: 0.9825; yaw$_{p95}$: 15.19°
B1 registration-guided refinement	F1: 0.9825; yaw$_{p95}$: 14.04°
B2 temporal smoothing	F1: 0.9780; yaw$_{p95}$: 3.92°
Dynamics-aware stabilization	Heading-motion error: 10.45° → 2.36°; heading-step: 0.0251 → 0.0040 rad

Quality Assurance Outcomes

Table 14 summarizes the four iterative QA rounds. Failure modes evolved from initial track fragmentation (Round 000), to TTC misuse from near-static or yielding cases (Round 001), to clearance ambiguity and anti-repeat sampling (Round 002), to box geometry instability and cross-lane false conflicts (Round 003). This progression reflects the pipeline becoming more refined at each stage: early rounds eliminate gross failures, while later rounds expose subtle issues in geometry and conflict classification.

Table 14: Iterative trajectory-QA validation rounds (Shang & Li, 2026).

Round	Cases	Issue	Main change or finding
000	10	Tracking breaks	No true near-miss; track fragmentation filtered before metric tuning
001	10	TTC misuse / weak distance	Motion filtering and TTC gating reduced over-triggering in near-static or yielding cases
002	3	TTC / clearance ambiguity	Anti-repeat sampling removed replay pairs; one near-miss remained but clearance was unreliable
003	3	Box geometry / lane conflict	Stationary-aware TTC and BEV clearance reduced false positives; geometry still unstable

Table 15 reports the overall human cross-check outcomes. Across 26 completed dashboard review decisions spanning rounds 000–003, the process confirmed one true near-miss case, one provisional/borderline positive, and one case deferred pending better geometry resolution. The anchor truck–bicycle interaction was confirmed by both model-based and human-reviewed assessment.

Table 15: Human cross-check summary for near-miss validation (Shang & Li, 2026).

Cross-check item	Result	Evidence
Completed dashboard review decisions	26	Rounds 000–003
Confirmed true near-miss decisions	1	Round 002
Provisional / borderline positives	1	Round 001
Deferred decisions (geometry unreliable)	1	Round 002
Anchor-case class-pair agreement	Yes	Truck–bicycle

Near-Miss Analytics

Figure 15 shows the three surrogate-safety curves for the anchor truck–bicycle interaction. Direction-agnostic TTC (middle panel) decreases sharply from above 2 s to a minimum near 0.62 s, confirming a genuine near-miss event. Longitudinal TTC (bottom panel) shows only a shallow minimum around 3.1 s—well above conservative braking thresholds—indicating that the heavy vehicle was not closing longitudinally at a rate requiring emergency braking. This divergence between the two TTC formulations reveals a lateral-intrusion-dominated conflict mechanism: the bicycle's trajectory crossed into the truck's path from the side, compressing the safety margin despite sufficient longitudinal stopping capacity.

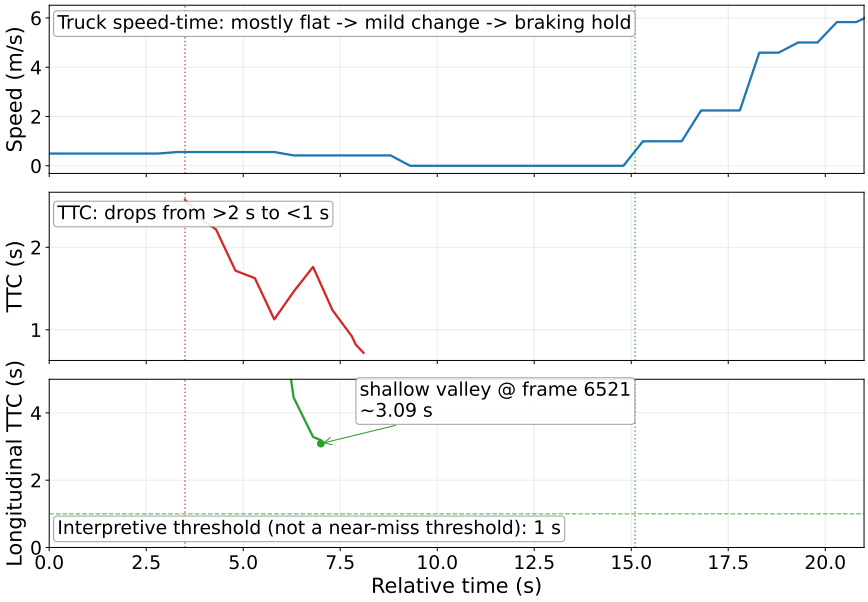


Figure 15: Anchor heavy vehicle–bicycle interaction. Top: heavy-vehicle speed. Middle: direction-agnostic TTC drops sharply below 1 s. Bottom: longitudinal TTC remains above braking thresholds, revealing a lateral-intrusion-dominated conflict mechanism.

Table 16 compares the model-recovered interaction metrics against the human-reviewed reference for the same anchor case. The model recovers the critical timing within two frames (0.2 s) of the human reference, confirming that the pipeline can reproduce human-reviewed conflict timing with high fidelity. The model estimates a tighter minimum gap ($\Delta = -0.61$ m) than the human reference, reflecting residual box-geometry uncertainty in the model's size estimates.

Table 16: Anchor truck–bicycle near-miss: model vs. human comparison (Shang & Li, 2026). Negative Δ indicates the model predicts a tighter interaction than the human reference.

Metric	Human	Model	Δ
Min TTC (s)	0.62	0.55	-0.07
Min separation (m)	0.81	0.20	-0.61
Min center distance (m)	6.47	5.61	-0.86
Min TTC timing	Reference	+2 frames	+0.2 s
Min sep. timing	Reference	+2 frames	+0.2 s

Table 17 shows that the audited analysis extends beyond the anchor case to multiple interaction types within the same deployment run. The truck–truck case (Case 02) exhibits very short minimum TTC (0.131 s) with negative separation (-4.017 m), suggesting extreme proximity that warrants further geometric review. The car–bicycle case (Case 03) shows a low TTC (0.455 s) but a large clearance (5.870 m), indicating that the TTC trigger may reflect converging trajectories rather than a true near-miss. These examples demonstrate that the audit pipeline is being exercised across multiple class-pair interaction types, and that the human QA layer is essential for distinguishing genuine conflicts from geometric or TTC-misuse artifacts.

Table 17: Multi-case audited near-miss summary from the same deployment run (Shang & Li, 2026). Negative minimum separation indicates estimated overlap under the BEV box proxy.

Case	Interaction pair	Min TTC (s)	Min sep. (m)
01	Truck–bicycle	0.552	0.201
02	Truck–truck	0.131	−4.017
03	Car–bicycle	0.455	5.870
04	Car–car	2.563	12.563
05	Truck–car	1.249	5.009

4.7 Discussion

Figure 16 shows the near-miss spatial overlay from post-8,000-frame data at the same intersection. Although the later events are not individually human-validated, the hotspot concentration persists in the same central conflict zone near the crosswalk and downstream of the stop bar, rather than dispersing uniformly across the scene. This spatial recurrence supports the broader claim that roadside sensing can expose site-specific conflict structure over time.

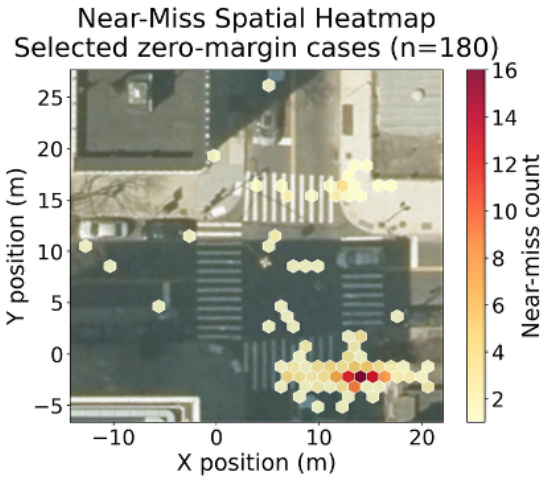


Figure 16: Vehicle–VRU near-miss spatial overlay ($n = 180$ zero-margin cases) from extended roadside LiDAR data.

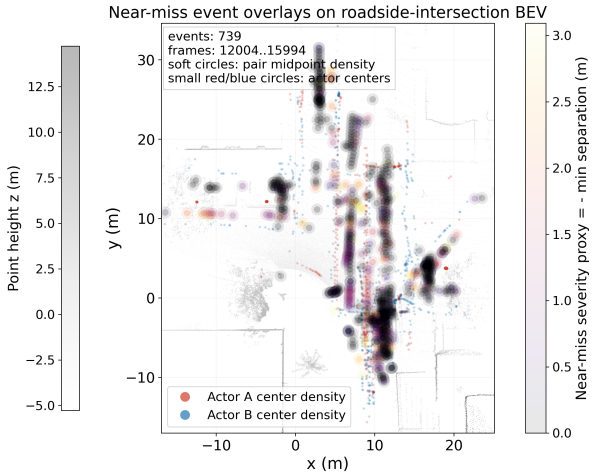


Figure 17: Post-8,000-frame vehicle–VRU near-miss overlay. Automatically detected events remain concentrated in the same central intersection conflict zone.

At larger scale, automatically mined movement summaries over the first 16,000 predicted frames yielded 336 movement-valid tracks, 6,046 quality-filtered candidate pairs, and 604 near-miss events. These aggregate statistics serve as recurrence indicators rather than a human-validated benchmark, confirming that the anchor interaction is not an isolated artifact of one short window but is part of a broader pattern of vehicle–VRU conflict at this intersection.

The iterative QA process reveals an important lesson: the difficulty of safety auditing shifts as the pipeline matures. Early rounds expose fundamental tracking and detection failures; later rounds expose subtle issues in geometry estimation, TTC metric selection, and conflict classification at the boundary of what counts as a genuine event. This progression—from gross failures to nuanced interpretation challenges—is characteristic of auditable safety analysis and motivates the

structured, multi-round review approach adopted here.

The primary contribution of this framework lies in its auditable infrastructure-side workflow, which improves the reviewability and interpretability of roadside trajectories through conservative trajectory refinement, iterative quality assurance, and mechanism-aware near-miss analysis. Direct comparison with recent cooperative perception and sensor-fusion frameworks is challenging because those studies often rely on different sensing configurations, synchronization assumptions, communication protocols, and application objectives. Accordingly, the present evaluation focuses on aspects that can be rigorously validated within the current roadside safety-audit setting, including detector performance against human-reviewed ground truth, branch-level trajectory refinement trade-offs, continuous-window tracking behavior, and reviewed agreement between model-estimated and human-interpreted near-miss timing.

Key limitations include data coverage at a single site, dependence on sensor placement and field of view, annotation effort that limits dataset scale, and residual box-geometry uncertainty that affects separation estimates for small objects such as bicycles. Future work should expand the pipeline to larger multi-intersection studies with more extensive model-versus-human near-miss validation, stronger identity-consistent tracking metrics across longer sessions, and explicit treatment of synchronization, communication, and calibration constraints required for scalable V2X deployment. LiDAR-camera fusion for precise object sizing remains an important complementary direction.

Table 18 summarizes the project milestones and outputs, tracing the progression from survey and benchmark studies through field deployment and safety auditing.

Table 18: Project milestones, datasets, and research outputs

Year	Milestone	Key Outputs and Contributions
2025	Literature survey and technology assessment	Comprehensive review of AI-enhanced sensing technologies for VRU safety at signalized intersections, covering LiDAR, cameras, radar, thermal sensing, sensor fusion, trajectory analysis, crossing-intention prediction, and AI-driven safety systems. Published in *Procedia Computer Science* (Shang, Li, Amin, et al., 2025).
2025	Initial LiDAR robustness evaluation	Conducted an initial investigation of LiDAR robustness for VRU detection under reduced beam-density conditions using the KITTI dataset and PointPillars architecture. Results showed that VRU detection performance degrades substantially faster than vehicle detection under sensor degradation conditions (Shang, Li, Wei, & Kamga, 2025).
2025	Infrastructure deployment and field data collection	Established a roadside LiDAR-camera sensing platform at a signalized intersection in New York City, including GPU-enabled edge computing infrastructure and synchronized multimodal data collection. Approximately 45 minutes of roadside LiDAR-camera data were collected to support trajectory analysis and annotation development.
2025	Annotation and trajectory analysis pipeline development	Developed a semi-assisted annotation workflow integrating pre-trained detection models with structured human review to support VRU trajectory validation and safety-event auditing.
2025	LiDAR vehicle classification under limited data	Developed LiDAR-based vehicle classification methods under limited-data conditions, with related findings accepted for presentation at the TRB Annual Meeting.
2026	Comprehensive LiDAR beam-loss robustness evaluation	Expanded the LiDAR degradation analysis to six state-of-the-art 3D detection architectures across KITTI and nuScenes datasets with up to 100 randomized beam-removal trials per condition. Identified approximately 20% beam-loss stability thresholds and demonstrated the stronger impact of contiguous beam degradation on VRU detection performance. Published in *Pattern Recognition Letters* (Feng et al., 2026).
2026	CCNY urban intersection LiDAR dataset	Developed an approximately 8,000-frame manually annotated roadside LiDAR dataset containing cars, trucks, bicycles, and pedestrians collected at a signalized urban intersection in New York City.
2026	Auditable roadside safety-analysis framework	Developed an end-to-end roadside LiDAR safety-auditing framework integrating 3D detection, multi-object tracking, trajectory refinement, dynamics-aware stabilization, and structured human-in-the-loop quality assurance for near-miss analysis at signalized intersections. Results presented at the CVPR DriveX Workshop (Shang & Li, 2026).

5 CONCLUSIONS AND POLICY IMPLICATIONS

5.1 Conclusions

This project examined whether roadside LiDAR, combined with modern AI/ML methods, can provide the continuous, interpretable, and defensible safety evidence needed to better protect vulnerable roadway users (VRUs) at signalized intersections. Through three interconnected research thrusts, the findings demonstrate that roadside LiDAR represents a promising and practical foundation for infrastructure-based VRU safety monitoring, while also highlighting important considerations related to sensing robustness, deployment conditions, and analysis workflow design.

The literature review showed that no single sensing modality can fully satisfy all operational requirements for VRU safety analysis. Cameras provide rich semantic information but remain sensitive to lighting and weather conditions. Radar offers strong environmental robustness but limited spatial resolution for detailed VRU characterization. Thermal sensing improves nighttime visibility but lacks precise trajectory reconstruction capability. Among the reviewed technologies, LiDAR emerged as a particularly strong candidate for infrastructure-side trajectory analysis because of its precise 3D geometric perception, lighting independence, and growing deployment feasibility. At the same time, the survey emphasized that future systems will likely benefit from multi-modal sensing and AI-driven fusion frameworks that combine the complementary strengths of different sensing technologies.

The LiDAR robustness analysis further demonstrated that VRU detection performance is substantially more sensitive to sensor degradation than vehicle detection performance. Across multiple detection architectures and benchmark datasets, pedestrian and cyclist detection remained relatively stable under moderate degradation conditions but deteriorated rapidly once beam loss exceeded approximately 20%. The results also showed that degradation patterns are important: contiguous beam loss caused by lens occlusion or contamination produced substantially greater performance degradation than dispersed beam loss of equivalent magnitude. In addition, beams located in the mid-to-high elevation region were found to be particularly important for preserving pedestrian and cyclist geometry. These findings provide practical guidance for sensor maintenance, deployment planning, and operational risk assessment for roadside LiDAR systems.

Building upon these findings, the project developed and demonstrated an auditable roadside safety-analysis framework using a real-world roadside LiDAR deployment at a signalized intersection in New York City. The framework integrated 3D object detection, multi-object tracking, trajectory refinement, dynamics-aware stabilization, and structured human-in-the-loop quality assurance to generate reviewable and defensible trajectory-level safety evidence. Rather than treating AI outputs as unquestionable ground truth, the framework emphasized iterative human review as a central component of the safety-analysis process. The near-miss case study involving a heavy vehicle and bicycle further demonstrated the importance of mechanism-aware interpretation: while direction-agnostic time-to-collision identified the interaction as a critical near-miss event, longitudinal time-to-collision analysis revealed the conflict to be dominated by lateral intrusion rather than direct longitudinal closing behavior. This distinction provides more interpretable insight into conflict mechanisms and can better support countermeasure selection and communication with transportation practitioners.

Taken together, the project establishes a coherent pathway from roadside sensing to interpretable safety intelligence for VRU monitoring at signalized intersections. The contributions include a comprehensive assessment of AI-enhanced sensing technologies, a systematic evaluation of LiDAR robustness under degradation conditions, and the development of an auditable infrastructure-based safety-analysis framework grounded in structured human review. While additional work is needed to improve scalability, multi-modal integration, and long-term deployment robustness, the results demonstrate the potential of roadside LiDAR and AI-driven analytics to support more proactive, data-driven, and interpretable approaches to transportation safety management.

5.2 Policy and Practice Implications

The findings from this project carry concrete implications for transportation agencies at multiple stages of the safety monitoring lifecycle.

Sensor procurement and specification. Agencies may benefit from looking beyond raw beam count when evaluating LiDAR sensors for VRU monitoring. The beam-loss study found that higher-resolution sensors showed no systematic robustness advantage in relative terms; the vertical distribution of beam coverage in the elevation range corresponding to human-scale geometry appears more relevant. When specifying LiDAR sensors, agencies could consider requesting documentation of beam distribution in the 15–45° elevation range above the horizontal and could include VRU-class detection performance—not only vehicle performance—in acceptance testing criteria.

Maintenance thresholds and inspection protocols. The study results suggest that approximately 20% beam loss may serve as a useful maintenance trigger for roadside LiDAR systems monitoring pedestrians and cyclists. Maintenance inspection could prioritize the mid-to-high elevation region of the sensor aperture, where beam loss appears to have the greatest impact on VRU detection. Sensors in occlusion-prone environments—intersections exposed to truck splash, winter road salt, or seasonal bird activity—may benefit from more frequent inspection intervals and from evaluation under contiguous-loss risk scenarios, not only average degradation assumptions.

Safety auditing and near-miss evidence production. The auditable pipeline developed in this project offers a replicable template for intersection-level safety analysis. The human-in-the-loop QA process, structured as iterative review rounds with explicit outcome classification, played an important role in producing evidence that was both technically defensible and interpretable to non-specialist reviewers. Agencies considering LiDAR-based safety audit programs may find it useful to plan for reviewer time as a substantive component of program budgets, rather than treating it as optional. Video co-deployment for annotation validation can improve dataset trustworthiness and conflict verification without requiring a fully automated video analytics deployment—a potentially useful consideration where camera coverage, privacy regulation, or infrastructure constraints limit video's role as a primary sensing stream.

Conflict diagnostics and countermeasure design. The dual TTC approach—contrasting direction-agnostic TTC with longitudinal TTC—offers agencies a way to distinguish lateral-intrusion conflicts from braking-limited ones. This distinction may have implications for countermeasure selection: lateral-intrusion conflicts may point toward geometric interventions such as protected bike

lanes, leading pedestrian intervals, and improved sight-line management, while braking-limited conflicts may call for speed management and approach-zone warning systems. Relying on only a single TTC metric risks collapsing this distinction, potentially making it harder to match countermeasure type to conflict mechanism.

5.3 Limitations and Future Work

Several limitations should be acknowledged when interpreting the current findings. First, data collection was conducted at a single signalized intersection, which limits the generalizability of the results across different roadway geometries, traffic compositions, environmental conditions, and regional climates. Second, although the approximately 8,000-frame annotation dataset represents a substantial effort, the labeling and review process was performed by a relatively small team under controlled quality-assurance conditions. Scaling the framework to larger multi-intersection deployments will require more efficient annotation tools, reviewer training protocols, and scalable quality-control workflows. Third, detection performance for smaller VRU classes—particularly bicycles under challenging viewing geometries and sparse point-cloud conditions—remains less reliable than vehicle detection, highlighting an important challenge for safety-critical infrastructure deployment. In addition, while cameras were co-deployed with LiDAR during data collection, direct LiDAR-camera fusion was not implemented in the current framework because the camera viewpoints were not geometrically aligned with the LiDAR coordinate system to support accurate feature-level or point-level fusion. Consequently, the camera system primarily served as a reference source for human verification and annotation rather than as an integrated perception modality. Finally, the safety-analysis framework was demonstrated using a limited number of reviewed near-miss events, including one detailed heavy vehicle-bicycle interaction case study, which establishes proof of concept but does not yet constitute a large-scale statistical validation.

Several future research directions emerge from this work. Expanding the framework to multi-intersection deployments is an immediate priority to evaluate cross-site generalizability, compare conflict patterns under diverse roadway conditions, and accumulate a larger reviewed near-miss dataset for broader safety analysis. Future deployments will also explore more tightly synchronized and geometrically aligned LiDAR-camera configurations to enable robust multi-modal fusion for improved VRU detection accuracy, object geometry estimation, and trajectory consistency, particularly for smaller VRU classes such as bicycles and pedestrians. Beyond traditional sensor fusion, dual LiDAR-derived representations—combining 3D point clouds with 2D range-image processing pipelines—may provide an alternative pathway for improving robustness when full multi-modal fusion is impractical. On the analytical side, extending the time-to-collision framework to more diverse pedestrian-vehicle and multi-party interactions, together with developing semi-automated conflict classification approaches that reduce reviewer burden while preserving interpretability, will be important for scaling infrastructure-based safety auditing. Longer term, integration with V2X communication systems and connected-vehicle infrastructure could enable real-time sharing of roadside safety observations with approaching vehicles, transforming the current auditing-oriented framework into a proactive and adaptive intersection safety intervention system.

References

Abdelrahman, A. S., Islam, Z., & Abdel-Aty, M. (2025). Vrucrosssafe for crossing intention prediction of vulnerable road users for improving safe crossing at intersections. *npj Sustainable Mobility and Transport*, *2*, 1–15.

Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., & Heide, F. (2020). Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 11682–11692).

Bowyer, K. W. (2004). Face recognition technology: Security versus privacy. *IEEE Technology and Society Magazine*, *23*, 9–19.

Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... Beijbom, O. (2020). nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 11621–11631).

De Waard, D., Schepers, P., Ormel, W., & Brookhuis, K. (2010). Mobile phone use while cycling: Incidence and effects on behaviour and safety. *Ergonomics*, *53*, 30–42.

Feng, W., Li, Y., & Wei, J. (2026). Impact of lidar beam loss on 3d object detection: A systematic analysis of vulnerable road user safety. *Pattern Recognition Letters*, *205*, 155–161. doi: 10.1016/j.patrec.2026.04.030

Fu, J., Ren, G., Chen, Y., & Liu, S. (2021). *Improved pillar with fine-grained feature for 3d object detection.* (arXiv:2110.06049)

Geiger, A. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the 2012 ieee conference on computer vision and pattern recognition (cvpr)* (pp. 3354–3361). IEEE Computer Society. doi: https://doi.org/10.1109/CVPR.2012.6248074

Gettman, D., & Head, L. (2003). *Surrogate safety measures from traffic simulation models* (Tech. Rep. No. FHWA-RD-03-050). Federal Highway Administration (FHWA). Retrieved from https://ntlrepository.blob.core.windows.net/lib/38000/38000/38015/FHWA-RD-03-050.pdf (HTML landing page: https://www.fhwa.dot.gov/publications/research/safety/03050/07.cfm)

Ghose, D., Desai, S. M., Bhattacharya, S., Chakraborty, D., Fiterau, M., & Rahman, T. (2019). *Pedestrian detection in thermal images using saliency maps.* (arXiv:1904.06859)

Google DeepMind. (2024). *Gemini: A family of highly capable multimodal models.* (arXiv:2312.11805)

Haque, F., Kidwai, F. A., Thapa, I., Ghani, S., & Mtapure, L. M. (2025). Modeling and evaluating the impact of mobile usage on pedestrian behavior at signalized intersections: A machine learning perspective. *Future Transportation*, *5*.

Huang, T., et al. (2023). *Vehicle-to-everything cooperative perception for autonomous driving: A survey.* arXiv preprint. Retrieved from https://arxiv.org/abs/2310.03525

Hydén, C. (1987). *The swedish traffic conflicts technique.* Technical report / calibration study. Retrieved from https://www.ictct.net/wp-content/uploads/SMoS_Library/LIB_Hyden_1987.pdf

Ismail, K., Sayed, T., & Saunier, N. (2009). Automated pedestrian safety analysis using video data in the context of scramble phase intersections. In *Annual conference of the transportation*

association of canada.

Jabłoński, P., Iwaniec, J., & Zabierowski, W. (2022). Comparison of pedestrian detectors for LiDAR sensor trained on custom synthetic, real and mixed datasets. *Sensors*, *22*(18), 7014.

Jansson, J. (2005). *Collision avoidance theory: with application to automotive collision mitigation* (PhD dissertation, Linköping University, Linköping, Sweden). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-91385

Ji, Y., et al. (2024). Toward autonomous vehicles: A survey on cooperative vehicle-infrastructure system. *iScience*. Retrieved from https://www.sciencedirect.com/science/article/pii/S2589004224009738 doi: 10.1016/j.isci.2024.109751

Johnsson, C., et al. (2021). A relative approach to the validation of surrogate measures of safety. *Accident Analysis & Prevention*. Retrieved from https://www.sciencedirect.com/science/article/pii/S000145752100381X doi: 10.1016/j.aap.2021.106511

Kang, J., Hamidi, O., Vanäs, K., Eidebåg, T., Nilsson, E., & Friel, R. (2025). Effects of Dust and Moisture Surface Contaminants on Automotive Radar Sensor Frequencies. *Sensors*, *25*(7), 2192.

Kim, J., Park, B.-j., & Kim, J. (2023). Empirical Analysis of Autonomous Vehicle's LiDAR Detection Performance Degradation for Actual Road Driving in Rain and Fog. *Sensors*, *23*(6), 2972.

Kong, W., Du, Y., He, L., & Li, Z. (2024). Improved 3d object detection based on pointpillars. *Electronics*, *13*, 2915.

Kristo, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal object detection in difficult weather conditions using yolo. *IEEE Access*, *8*, 125459–125476.

Kulhandjian, H. (2024). *Smart robot design and implementation to assist pedestrian road crossing* (Tech. Rep. No. MTI-2353). Mineta Transportation Institute.

Lang, A., Vora, S., Caesar, H., Zhou, L., Yang, J., & Beijbom, O. (2019). Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 12697–12705).

Li, Y., Tok, A. Y., Sun, Z., Ritchie, S. G., & Allu, K. R. (2023). Lidar vehicle point cloud reconstruction framework for axle-based classification. *IEEE Sensors Journal*, *23*, 11168–11180.

Lin, C., et al. (2023). Vehicle detection and tracking with roadside lidar using low-latency edge computing. *Sensors*. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC10575351/ doi: 10.3390/s23198125

Liu, H., Li, C., & Wu, Q. (2023). *Visual instruction tuning.* (arXiv:2304.08485)

Lu, W., Zhao, D., Premebida, C., Zhang, L., Zhao, W., & Tian, D. (2023). Improving 3d vulnerable road user detection with point augmentation. *IEEE Transactions on Intelligent Vehicles*, *8*, 3489–3505.

Mo, Y., et al. (2024). Enhanced perception for autonomous vehicles at intersections using roadside sensing and v2x communication. *Sensors*. Retrieved from https://pmc.ncbi.nlm.nih.gov/articles/PMC10856888/ doi: 10.3390/s24030931

Nasar, J. L., Hecht, P., & Wener, R. (2008). Mobile telephones, distracted attention, and pedestrian safety. *Accident Analysis & Prevention*, *40*, 69–75.

National Center for Statistics and Analysis. (2024a). *Bicyclists and other cyclists: 2022 data* (Tech. Rep. No. DOT HS 813 591). National Highway Traffic Safety Administration. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/Publication/813591

National Center for Statistics and Analysis. (2024b). *Motorcycles: 2022 data* (Tech. Rep. No. DOT HS 813 589). National Highway Traffic Safety Administration. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/Publication/813589

National Center for Statistics and Analysis. (2024c). *Pedestrians: 2022 data* (Tech. Rep. No. DOT HS 813 590). National Highway Traffic Safety Administration. Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/Publication/813590

New York State Department of Transportation. (2023). *Appendix 2: Vulnerable road user safety assessment* (Tech. Rep.). New York State Department of Transportation.

Nie, M., Xue, Y., Wang, C., Ye, C., Xu, H., Zhu, X., ... Zhang, L. (2023). Partner: Level up the polar representation for lidar 3d object detection. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 3801–3813).

Palladin, E., Dietze, R., Narayanan, P., Bijelic, M., & Heide, F. (2024). Samfusion: Sensor-adaptive multimodal fusion for 3d object detection in adverse weather. In *European conference on computer vision* (pp. 484–503).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 779–788).

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).

Saenz, H., Sun, H., Wu, L., Zhou, X., & Yu, H. (2021). Detecting phone-related pedestrian distracted behaviours via a two-branch convolutional neural network. *IET Intelligent Transport Systems*, *15*, 147–158.

Saldivar-Carranza, E. D., Zlatkovic, M., & Stevanovic, A. (2024). Vehicle and pedestrian traffic signal performance measures using lidar-derived trajectory data. *Sensors*, *24*(19), 6410. doi: 10.3390/s24196410

Sekaran, K. C., Geisler, M., Rößle, D., Mohan, A., Cremers, D., Utschick, W., ... Schön, T. (2025). *Urbaning-v2x: A large-scale multi-vehicle, multi-infrastructure dataset across multiple intersections for cooperative perception.* arXiv preprint. Retrieved from https://arxiv.org/abs/2510.23478 (Code/dataset: https://github.com/thi-ad/UrbanIng-V2X)

Sezgin, F., Vriesman, D., Steinhauser, D., Lugner, R., & Brandmeier, T. (2023). Safe Autonomous Driving in Adverse Weather: Sensor Evaluation and Performance Monitoring. *arXiv preprint arXiv:2305.01336*.

Shang, B., & Li, Y. (2026). Roadside lidar for cooperative safety auditing at urban intersections: Toward auditable v2x infrastructure intelligence. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 757–765).

Shang, B., Li, Y., Amin, A. G., Kamga, C., & Wei, J. (2025). Ai-enhanced sensing for vulnerable road user safety at signalized intersections: A survey. In *The 22nd international conference on mobile systems and pervasive computing (mobispc)* (Vol. 265, pp. 350–357). doi: 10

.1016/j.procs.2025.07.191

Shang, B., Li, Y., Wei, J., & Kamga, C. (2025). Investigating the lidar beam count needed for effective vulnerable road user detection. In *The 22nd international conference on mobile systems and pervasive computing (mobispc)* (Vol. 265, pp. 358–365). doi: 10.1016/j.procs .2025.07.192

Wang, Y., Zou, Z., Ye, X., Tan, X., Ding, E., & Zhao, C. (2024). *Uni2det: Unified and universal framework for prompt-guided multi-dataset 3d detection.* (arXiv:2409.20558)

Wang, Z., Huang, X., & Hu, Z. (2025). Attention-based lidar-camera fusion for 3d object detection in autonomous driving. *World Electric Vehicle Journal*, *16*, 306.

Ward, J. R., Agamennoni, G., Worrall, S., Bender, A., & Nebot, E. (2015). Extending time to collision for probabilistic reasoning in general traffic scenarios. *Transportation Research Part C: Emerging Technologies*, *51*, 66–82. doi: 10.1016/j.trc.2014.11.002

Watson, A., Watson, B., & Vallmuur, K. (2015). Estimating under-reporting of road crash injuries to police using multiple linked data collections. *Accident Analysis & Prevention*, *83*, 18–25. doi: 10.1016/j.aap.2015.06.011

Wu, A., Banerjee, T., Chen, K., Rangarajan, A., & Ranka, S. (2023). A multi-sensor video/lidar system for analyzing intersection safety. In *2023 ieee 26th international conference on intelligent transportation systems* (pp. 1158–1165).

Yazgan, M., et al. (2024). *Collaborative perception datasets in autonomous driving.* arXiv preprint. Retrieved from https://arxiv.org/abs/2404.14022

Yin, T., Zhou, X., & Krahenbuhl, P. (2020). *Center-based 3d object detection and tracking.* (arXiv:2006.11275)

Yu, H., et al. (2023). V2x-seq: A large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (cvpr).* Retrieved from https://openaccess.thecvf.com/content/CVPR2023/html/Yu_V2X-Seq_A_Large-Scale_Sequential_Dataset_for_Vehicle-Infrastructure_Cooperative_Perception_and_CVPR_2023_paper.html

Zhang, J., Ge, C., Xiao, W., Tang, M., Mills, J., Coifman, B., & Chen, N. (2026). Roadside lidar-based scene understanding toward intelligent traffic perception: A comprehensive review. *ISPRS Journal of Photogrammetry and Remote Sensing*. Retrieved from https://www.sciencedirect.com/science/article/pii/S0924271626000122 doi: 10.1016/j.isprsjprs.2026.01.012

Zhang, L., Yu, X., Aboah, A., & Adu-Gyamfi, Y. (2024). *3d object detection and high-resolution traffic parameters extraction using low-resolution lidar data.* (arXiv:2401.06946)

Zhang, R., Guan, R., Chen, X., Prugel-Bennett, A., & Cai, X. (2025). *Cornerpoint3d: Look at the nearest corner instead of the center.* (arXiv:2504.02464)

Zhou, Y., & Tuzel, O. (2018). Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4490–4499).

Zimmer, W., et al. (2024). Tumtraf v2x cooperative perception dataset. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (cvpr).* Retrieved from

https://openaccess.thecvf.com/content/CVPR2024/papers/Zimmer_TUMTraf_V2X_Cooperative_Perception_Dataset_CVPR_2024_paper.pdf (Project page: https://tum-traffic-dataset.github.io/tumtraf-v2x/ ; arXiv: https://arxiv.org/abs/2403.01316)