

Roadside LiDAR for Cooperative Safety Auditing at Urban Intersections: Toward Auditable V2X Infrastructure Intelligence

Bo Shang, Yiqiao Li
The City College of New York
{bshang, yli4}@ccny.cuny.edu

Abstract

Urban intersections expose the limitations of single-vehicle perception under occlusion and partial observability. In this study, we present an auditable roadside LiDAR framework for infrastructure-assisted safety analysis at a signalized urban intersection in New York City, developed and evaluated using real-world data. The proposed framework integrates trajectory construction, iterative human-in-the-loop quality assurance (QA), and interpretable near-miss analytics to produce defensible safety evidence from infrastructure sensing. Using a human-labeled heavy vehicle–bicycle interaction as an anchor case, we show that direction-agnostic time-to-collision (TTC) drops below 1 s, while longitudinal TTC remains above conservative braking thresholds, revealing a lateral-intrusion-dominated conflict mechanism. Beyond individual cases, continuous-window evaluation and multi-round QA analysis demonstrate that the framework systematically reduces failure modes such as track fragmentation, spurious TTC triggers, unstable geometry, and cross-lane false conflicts. These results position roadside LiDAR as a practical post-hoc auditing mechanism for cooperative perception systems, with broader statistical validation discussed. This work provides a pathway toward scalable, data-driven safety auditing of urban intersections, enabling transportation agencies to identify and mitigate high-risk interactions beyond crash-based analyses.

1. Introduction

Autonomous driving has advanced rapidly in 3D detection, tracking, and scene understanding, yet most deployed pipelines remain fundamentally single-vehicle-centric and therefore 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. In this paper, we focus on a key subproblem within full-stack cooperative autonomy: how can roadside sensing provide *auditable* and *interpretable* safety evidence for reviewing safety-critical interactions at intersections? We study auditability and interpretability in the context of infrastructure-based evaluation, with a scope that focuses on evidentiary transparency and reviewability, rather than encompassing broader trustworthiness dimensions such as communication robustness, cybersecurity or regulatory compliance.

We investigate this question through a real-world pilot deployment of a roadside LiDAR–first safety-auditing pipeline at a site-specific urban intersection in New York City. The pipeline integrates 3D object detection, multi-object tracking, conservative trajectory refinement, a dynamics-aware post-tracking stabilization stage, and structured human-in-the-loop Quality Assurance (QA) to construct auditable trajectories for downstream surrogate safety analysis. The primary contribution lies in system design and validation methodology: how standard perception components are constrained, stabilized, reviewed, and interpreted to produce defensible safety evidence.

We use a human-labeled heavy vehicle–bicycle near-miss as an anchor case. Standard direction-agnostic Time-To-Collision (TTC) rapidly collapses below 1 s, while longitudinal TTC relative to the heavy vehicle heading stays above conservative braking thresholds. This divergence suggests a lateral-intrusion-dominated interaction rather than insufficient longitudinal stopping capacity. Beyond this anchor case, we evaluate the tracking pipeline on continuous frame windows with stitched ground-truth trajectories and summarize repeated QA rounds that expose the dominant failure modes of infrastructure-side near-miss mining.

The paper makes four major contributions. First, we present an end-to-end roadside LiDAR pipeline for auditable safety analysis at a real-world urban intersection. Second, we establish an 8,000-frame manually annotated

roadside LiDAR resource with frame-level cuboids plus later trajectory-level QA artifacts. Third, we demonstrate how structured review rounds transform raw detection and tracking outputs into a defensible and auditable near-miss analysis workflow. Fourth, we use longitudinal TTC, contrasted against direction-agnostic TTC, as an explanatory diagnostic for distinguishing braking-limited and lateral-intrusion-dominated interactions in roadside trajectory auditing.

2. Related Work

Cooperative perception and vehicle–infrastructure systems aim to reduce observability gaps through roadside sensing and shared scene context [3, 6, 12]. Public datasets such as DAIR-V2X, V2X-Seq, TUMTraf-V2X, and UrbanIng-V2X have accelerated work on feature fusion, communication efficiency, and cooperative perception benchmarks [10, 14, 15, 17]. In contrast, our work is single-site and narrower in scale, but it contributes a different artifact: a reviewed roadside LiDAR record built for safety auditing rather than only benchmark training and fusion evaluation.

Roadside LiDAR has also been used for intersection monitoring, traffic participant tracking, and occlusion-aware assistance [8, 9, 16]. Meanwhile, surrogate safety measures such as TTC and Post-Encroachment Time (PET) remain standard tools for conflict analysis when crashes are too rare for direct estimation [2, 4, 7]. TTC itself has long been used in essentially longitudinal car-following settings, and later work generalized TTC to unconstrained two-dimensional motion [5, 11]. Our contribution is the use of longitudinal TTC, contrasted against direction-agnostic TTC, as an interpretive diagnostic within an auditable roadside trajectory pipeline. Despite recent advances, 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. This work seeks to fill that gap.

3. Methodology

Figure 1 summarizes the methodology as a perception-to-audit workflow. The perception stack covers Stage 1 detections, Stage 2 identity-consistent tracks, and Stage 3 refined trajectories; human-in-the-loop review then acts as a supervisory validation layer before and during Stage 4 near-miss analytics, which produces the final auditable safety evidence.

3.1. Stage 1: 3D detection.

Each frame is processed using a fine-tuned CenterPoint detector [13] trained on our 8,000 manually annotated roadside LiDAR dataset. For each detected object, the model outputs a class label c , confidence score s , and 3D bounding

box parameters (x, y, z) (center coordinates), (d_x, d_y, d_z) (box dimensions), and ψ (heading angle).

3.2. Stage 2: Multi-object tracking for Data Association.

To establish temporal correspondence, detections are associated across frames using a SORT-style tracking framework [1] that integrates Kalman filtering with bird’s-eye-view (BEV) gating. This step produces stable and persistent track identities, forming the basis for reliable trajectory-level analysis.

3.3. Stage 3: Refinement and stabilization.

While multi-object tracking provides temporally consistent identities, the resulting trajectories often exhibit frame-level noise, orientation jitter, and occasional misalignment due to detection uncertainty and association errors. This noise can propagate into downstream safety metrics (e.g., TTC) and compromise interpretability. Therefore, a refinement stage is necessary to improve temporal stability while preserving audibility.

Branch-level refinement policy. To keep the audited comparisons controlled, we maintain three post-tracking refinement branches under identical upstream detections and audited frame windows. **B0** retains the raw tracked trajectories as a baseline. **B1** introduces a selective correction mechanism that targets potentially unreliable track segments. Suspicious tracklets are identified based on three criteria: (i) *yaw-step discontinuities*, defined as abrupt frame-to-frame changes in the estimated heading angle ψ_t that exceed a predefined threshold and are inconsistent with physically plausible motion; (ii) *registration disagreement*, measured as the mismatch between the predicted state from the tracker (e.g., Kalman filter extrapolation) and the observed detection, typically quantified by BEV spatial residuals or IoU-based inconsistencies; and (iii) a *thresholded suspicion score*, which aggregates normalized indicators (e.g., yaw variation, position error, and detection confidence drops) into a composite reliability score used to flag unstable track segments. For tracklets flagged as suspicious, **B1** applies *capped frame-level corrections*, where adjustments to position and orientation are bounded to prevent over-smoothing or distortion of genuine motion patterns. These corrections are only applied when *registration quality is deemed sufficient*, meaning that the detection-to-track alignment satisfies minimum criteria (e.g., IoU above a threshold or residual below a tolerance), ensuring that corrections are anchored to reliable observations rather than propagating tracking errors. **B2** applies stronger frame-wise temporal smoothing across the entire trajectory, enforcing higher temporal consistency but at the potential cost of attenuating sharp but valid motion changes.

We use **B1** as the practical default because it preserves

Table 1. Positioning of this work relative to representative infrastructure-side datasets.

Resource	Scale / coverage	Sensing setup	Annotation emphasis	Primary tasks	Audit/QA	Distinguishing scope
DAIR-V2X [14]	large real-world V2X benchmark	vehicle + infrastructure multi-view sensing	3D object labels for cooperative perception	cooperative 3D detection	no	broader V2X benchmark scale; not organized around reviewed safety events
V2X-Seq [15]	sequential real-world V2X scenes	vehicle + infrastructure sequential sensing	sequence labels supporting perception and prediction	cooperative perception, forecasting	no	temporal V2X benchmark; audit-oriented near-miss interpretation is not primary
TUMTraf-V2X [17]	urban intersection traffic scenes	multi-sensor vehicle-infrastructure setup	scene-level labels for cooperative scene understanding	cooperative perception at intersections	no	intersection-focused benchmark with broader sensor stack, but not structured around safety auditing
UrbanIng-V2X [10]	broader multi-intersection benchmark	multi-site vehicle-infrastructure sensing	benchmark labels for cooperative perception	multi-intersection cooperative perception	no	larger geographic coverage and benchmark scope than this study
This work	single-site pilot roadside deployment	fixed roadside LiDAR-first sensing	3D boxes, trajectories, reviewed near-miss cases, QA artifacts	detection, tracking, audited near-miss analysis	yes	narrower scale but unique emphasis on auditable trajectories, reviewed conflict interpretation, and hotspot-oriented safety analysis

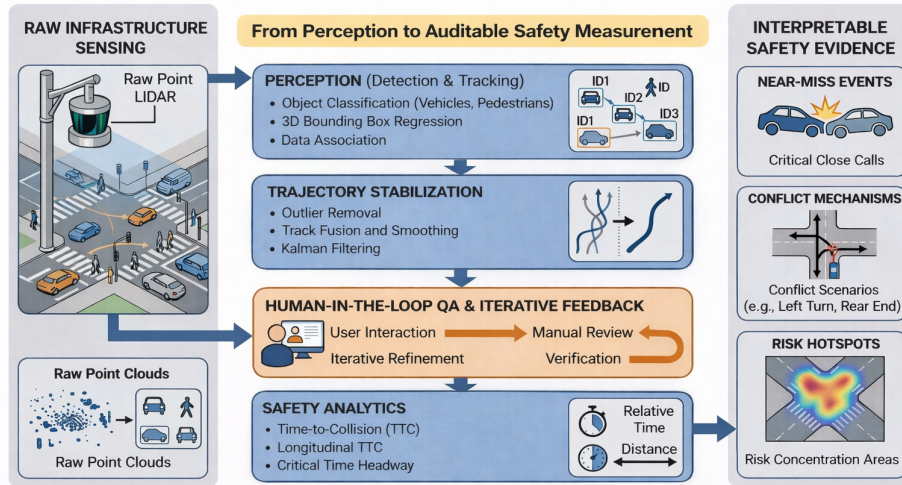


Figure 1. Conceptual overview of the roadside LiDAR safety-auditing architecture. Raw point clouds are converted into detections, trajectories, and stabilized motion estimates, then reviewed through a human-in-the-loop QA layer and interpreted with surrogate-safety analytics to produce auditable safety evidence such as reviewed near-miss cases, conflict mechanisms, and risk hotspots.

audited pair-level summaries while modestly reducing orientation jitter. Longitudinal TTC is then applied to the selected audited pairs as an interpretive diagnostic rather than a standalone binary event threshold.

Dynamics-aware stabilization extension. Beyond the B0/B1/B2 branch comparison, we also evaluate a separate dynamics-aware stabilization module. This extension is intended to suppress high-frequency noise while enforcing physically plausible motion, but because it can materially change the recovered anchor-case geometry, we report it separately rather than fold it into the default audited branch. Let $\mathbf{p}_t = (x_t, y_t)^\top$ denote the observed box center (BEV center), ψ_t the heading, and $\mathbf{d}_t = (d_{x,t}^{\text{aligned}}, d_{y,t}^{\text{aligned}}, d_{z,t})^\top$

the aligned dimensions. Let \mathcal{W}_t denote a centered temporal window.

Position smoothing and stabilization. A local estimator is first constructed as a weighted average:

$$\bar{\mathbf{p}}_t = \frac{\sum_{\tau \in \mathcal{W}_t} w_{t,\tau} \mathbf{p}_\tau}{\sum_{\tau \in \mathcal{W}_t} w_{t,\tau}}, \quad w_{t,\tau} \geq 0, \quad (1)$$

where $w_{t,\tau}$ denotes the temporal smoothing weight. In the current implementation, $w_{t,\tau} = 1$ for frames $\tau \in \mathcal{W}_t$ and $w_{t,\tau} = 0$ otherwise, yielding a centered moving average with a default window size of 9 frames. The smoothed position is then passed through a separate constant-velocity prior, so temporal averaging and motion regularization are

applied in two successive steps.

Let $\tilde{\mathbf{p}}_t$ denote the stabilized position and define the finite-difference velocity

$$\mathbf{v}_{t-1} = \frac{\tilde{\mathbf{p}}_{t-1} - \tilde{\mathbf{p}}_{t-2}}{\Delta t}. \quad (2)$$

A constant-velocity prediction is given by

$$\hat{\mathbf{p}}_t = \tilde{\mathbf{p}}_{t-1} + \mathbf{v}_{t-1}\Delta t. \quad (3)$$

The stabilized estimate is obtained via convex combination:

$$\tilde{\mathbf{p}}_t = (1 - \alpha)\bar{\mathbf{p}}_t + \alpha\hat{\mathbf{p}}_t, \quad \alpha \in [0, 1]. \quad (4)$$

Heading stabilization. To account for angular periodicity, heading is smoothed using circular averaging:

$$\bar{\psi}_t = \text{atan2}\left(\sum_{\tau \in \mathcal{W}_t} w_{t,\tau} \sin \psi_\tau, \sum_{\tau \in \mathcal{W}_t} w_{t,\tau} \cos \psi_\tau\right). \quad (5)$$

Equation (5) computes a weighted circular mean of yaw angles over a temporal window by averaging their sine and cosine components and recovering the orientation via the *atan2* function. This formulation avoids discontinuities due to angular wrap-around and yields a stable and physically consistent estimate of heading. Let ψ_t^{motion} denote the motion-induced heading (e.g., path tangent). Define the wrapped angular difference

$$\Delta\psi(a, b) := \text{wrap}(a - b) \in (-\pi, \pi].$$

A blended heading is first computed as

$$\psi_t^{\text{blend}} = \bar{\psi}_t + \beta \Delta\psi(\psi_t^{\text{motion}}, \bar{\psi}_t), \quad \beta \in [0, 1]. \quad (6)$$

In the current implementation, β is adaptive rather than fixed. Let s_t denote speed, s_{\min} the minimum speed at which motion heading becomes reliable, and let r_t denote the resulting heading-reliability coefficient:

$$r_t = \min\left(\frac{s_t - s_{\min}}{s_{\min} + 0.75}, 1\right), \quad (7)$$

$$\alpha_t = \alpha_{\text{low}} + (\alpha_{\text{high}} - \alpha_{\text{low}})r_t.$$

Here α_t denotes the implementation-side raw-yaw blend factor, while r_t measures heading reliability from motion evidence. with defaults $\alpha_{\text{low}} = 0.35$, $\alpha_{\text{high}} = 0.08$, and $s_{\min} = 0.75$ m/s. Because the implementation parameterizes the blend from the motion-heading side, the equivalent coefficient in Eq. (6) is

$$\beta_t = \begin{cases} 1 - \alpha_t, & \text{if } |\Delta\psi(\bar{\psi}_t, \psi_t^{\text{motion}})| \leq \epsilon_\psi, \\ 1, & \text{otherwise,} \end{cases} \quad (8)$$

$$\epsilon_\psi = 20^\circ.$$

Thus β_t increases with motion reliability: at low speed the blend remains closer to the raw smoothed heading, while at higher speed it shifts toward the motion-induced heading.

To enforce a bounded turn rate, the update is constrained by projection onto the admissible increment set:

$$\tilde{\psi}_t = \tilde{\psi}_{t-1} + \Pi_{[-\omega_{\max}\Delta t, \omega_{\max}\Delta t]}(\Delta\psi(\psi_t^{\text{blend}}, \tilde{\psi}_{t-1})), \quad (9)$$

where ω_{\max} is the maximum allowed angular velocity, so the proposed heading increment is clipped to the interval $[-\omega_{\max}\Delta t, \omega_{\max}\Delta t]$: the lower bound $-\omega_{\max}\Delta t$ is the largest allowed clockwise step and the upper bound $+\omega_{\max}\Delta t$ is the largest allowed counterclockwise step. Here $\Pi_{[a,b]}(x) = \min(\max(x, a), b)$ denotes scalar clipping onto $[a, b]$. In the current implementation this is realized as a per-frame angular-step limit, with default $\omega_{\max}\Delta t = 2.5^\circ$ for consecutive frames and proportional scaling for larger frame gaps.

Optionally, $\tilde{\psi}_t$ is back-propagated from the first reliable-motion frame to stabilize low-speed initialization.

Dimension stabilization. Object dimensions are stabilized componentwise via robust estimation:

$$\tilde{d}_i = \arg \min_d \sum_t \rho(d_{i,t} - d), \quad i \in \{x, y, z\}, \quad (10)$$

and $\tilde{\mathbf{d}} = (\tilde{d}_x, \tilde{d}_y, \tilde{d}_z)^\top$, where $\rho(\cdot)$ is a convex robust loss (e.g., ℓ_1). $d_{i,t}$ is the observed size at time t . d is candidate "true" dimension.

3.4. Cross-Cutting Audit Layer: Human QA.

Automated review queues are exported as short tracklet windows with aligned point-cloud context, box overlays, and candidate-event metadata so that 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 / lane-assignment artifact. Reviewer outcomes are stored as structured correction records rather than free-form notes, including keep / reject decisions, dominant failure tags, and any required track-level corrections. These records are then fed back into the pipeline in two ways: they correct the audited track outputs for reported examples, and they update the near-miss screening logic by exposing recurring false-positive modes such as fragmented tracks, invalid closing-motion assumptions, and unstable box geometry.

3.5. Stage 4: Near-Miss Analytics.

Candidate pairs are first screened by direction-agnostic TTC and minimum separation, then interpreted with longitudinal TTC to distinguish longitudinal braking-limited risk from lateral-intrusion-dominated interactions. We use



Figure 2. Study intersection and roadside sensing setup used for the pilot deployment.

TTC as a standard surrogate safety measure from the traffic-conflict literature [2, 4, 7], while minimum separation is treated here as a practical implementation-specific BEV clearance proxy rather than a canonical published metric. For a candidate pair indexed by $i \in \{1, 2\}$, let $\Delta \mathbf{p} = \mathbf{p}_2 - \mathbf{p}_1$, $\Delta \mathbf{v} = \mathbf{v}_2 - \mathbf{v}_1$, and $r_i = \frac{1}{2} \sqrt{d_{x,i}^2 + d_{y,i}^2} + \beta$, where $d_{x,i}, d_{y,i}$ are the BEV box dimensions of participant i , β is a fixed radius buffer (0.3 m in the current implementation), and t indexes frames within the interaction window. We use

$$s_t = \|\Delta \mathbf{p}(t)\| - (r_1(t) + r_2(t)), \quad (11)$$

$$\text{TTC}(t) = \begin{cases} \frac{-\Delta \mathbf{p}(t) \cdot \Delta \mathbf{v}(t)}{\|\Delta \mathbf{v}(t)\|^2}, & \Delta \mathbf{p}(t) \cdot \Delta \mathbf{v}(t) < 0, \\ +\infty, & \text{otherwise,} \end{cases} \quad (12)$$

and summarize each candidate pair by $\min_t s_t$ and $\min_t \text{TTC}(t)$ over the interaction window. Negative minimum-separation values therefore indicate overlap of the size-adjusted BEV proxies, not literal physical penetration.

4. Experimental Setup

Study site and sensing. Data were collected at the signalized intersection of Convent Ave & W 141st St near the City College of New York using a fixed roadside Ouster OS-128 LiDAR operating at 10 Hz. The deployment is a fixed elevated oblique-view setup covering the intersection center and crosswalk approaches. All object positions, trajectories, and near-miss measurements are reported in one common BEV coordinate system aligned to this intersection and held fixed throughout the study.

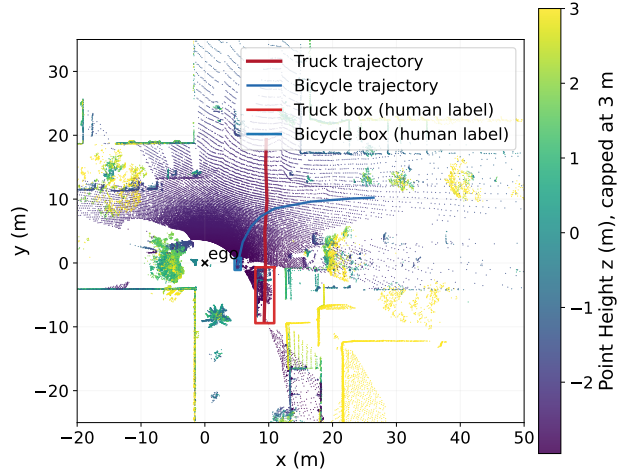


Figure 3. Representative roadside LiDAR point cloud from the audited interaction window, shown in bird's-eye view with human-labeled truck and bicycle cuboids.

Annotation resource. The 8,000 study set consists of frame-level 3D cuboid annotations for four coarse classes: car, truck, pedestrian, and bicycle. Source annotations were created in MATLAB LiDAR Labeler at the native 10 Hz frame rate, then converted into the formats used for detector training and audit evaluation. Quality control has two stages: human-reviewed ground-truth frame-label review for detector evaluation and later trajectory-level re-review for audited windows. This differs from benchmark-only infrastructure datasets by combining frame labels with later trajectory-QA artifacts and reviewed near-miss analysis.

Evaluation protocol. We report three forms of evidence: detector performance against human-reviewed ground truth, branch-level tracking and refinement comparisons on audited and continuous frame windows with stitched ground-truth trajectories, and human cross-checks for reviewed near-miss candidates. The present paper is a single-site pilot, so these results should be interpreted as feasibility and auditability evidence rather than a comprehensive generalization study.

This deployment choice is intentional: 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 trade-off is that synchronization error, calibration maintenance across many intersections, and bandwidth-constrained V2X delivery are left to future systems work.

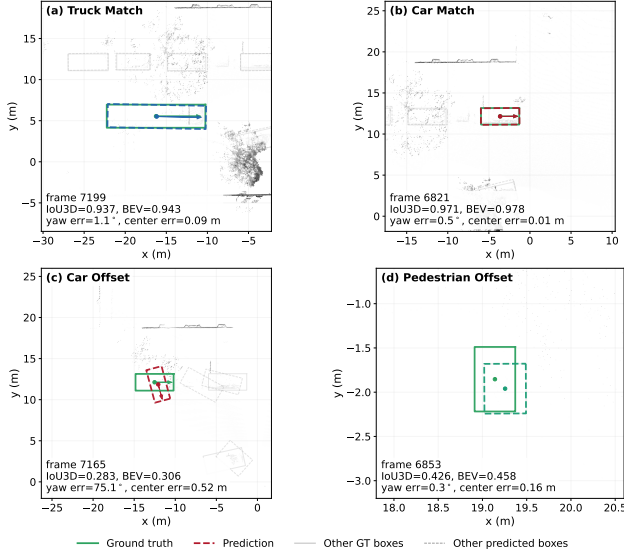


Figure 4. Qualitative detection examples from the CenterPoint model trained on the 8,000-frame dataset, evaluated on the consecutive validation split: (a) a high-overlap truck match, (b) a high-overlap car match, (c) a car case with noticeable localization offset, and (d) a single-pedestrian case with moderate localization offset. Car and truck detections are generally strong, while small-object localization remains more fragile.

Table 2. Compact quantitative summary of the main pipeline and audit-layer components.

Component	Result
Detector (human-reviewed GT holdout)	Vehicle AP@0.70 (R40) 76.38%; native AP: 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

Table 3. Tracking evaluation on continuous frame windows against ground-truth trajectories stitched across consecutive frames. A predicted box is counted as a match when its BEV center lies within 1.5 m of the corresponding ground-truth box center.

Window	Precision	Recall	F1	yaw _{p95}
First 2000 frames	0.6947	0.9858	0.8150	3.88°
First 4000 frames	0.5734	0.8677	0.6905	6.50°
Sequential validation slice	0.6036	0.6726	0.6362	3.47°

Table 4. Human cross-check summary for near-miss validation.

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 pending better geometry	1	round 002
Anchor-case class-pair agreement	yes	truck–bicycle

Table 5. Iterative trajectory-QA validation rounds for refining the near-miss review pipeline.

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.

5. Results

5.1. Detection, Tracking, and QA

Table 2 shows that the paper’s strongest quantitative support is not detector novelty but the combination of reviewed trajectory construction and controlled refinement. B1 preserves the audited F1 of the baseline while modestly reducing yaw jitter, whereas B2 achieves stronger smoothing at a small cost in F1. The dynamics-aware branch further improves geometric consistency, but because it materially changes the anchor near-miss geometry, we report it as a separate engineering extension rather than substitute it into the core audited comparison.

Figure 4 is consistent with a local-domain deployment story rather than cross-site generalization: the detector localizes cars and trucks well in the study geometry, but pedestrian-scale and weak-overlap cases remain more brittle. Table 3 broadens the evidence beyond the anchor interaction. Performance is strongest on the shorter first-2000 slice and declines on longer windows, mainly through accumulated false positives and missed long-tail cases rather than catastrophic geometric instability. The iterative QA process in Table 5 is equally important: dominant failure modes evolved from track fragmentation to TTC misuse, then to geometry instability and different-lane false conflicts. The public QA dashboard records 26 completed re-

Table 6. Anchor near-miss comparison for the same truck–bicycle interaction. Minimum separation is a signed size-adjusted clearance; negative values in Δ indicate that the model predicts a tighter or more overlapping 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 7. Multi-case audited near-miss summary from the same run under the final conservative trajectory pipeline. ‘Min sep.’ denotes signed size-adjusted clearance, so negative values indicate estimated overlap under the box proxy.

Case	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

view decisions across rounds 000–003.¹

These results also clarify what is and is not being claimed. We do not present an apples-to-apples leaderboard comparison against recent cooperative fusion methods because those works use different sensing stacks, synchronization assumptions, communication protocols, and target tasks. Instead, the present evaluation focuses on what is directly valid under this roadside audit setup: detector performance against human-reviewed ground truth, controlled branch-level refinement trade-offs, broader continuous-window tracking behavior with stitched ground-truth trajectories, and reviewed model-versus-human near-miss timing agreement.

5.2. Near-Miss Interpretation

For the anchor truck–bicycle interaction, the model recovers the same critical timing within two frames of the human reference (Table 6) but estimates a tighter minimum gap. This behavior is consistent with residual box-geometry uncertainty even when interaction timing is largely preserved.

Table 7 shows that the audited analysis is not limited to one pair type. Within the same deployment, the reviewed set also contains truck–truck, car–bicycle, car–car, and truck–car interactions. These cases are still too few for a full statistical study, but they help demonstrate that the audit pipeline is being exercised across multiple interaction types rather than only the anchor truck–bicycle example.

To interpret the interaction dynamics beyond direction-agnostic safety screening, we compute the longitudinal

¹trajectory-qa.aimobilitylab.xyz/

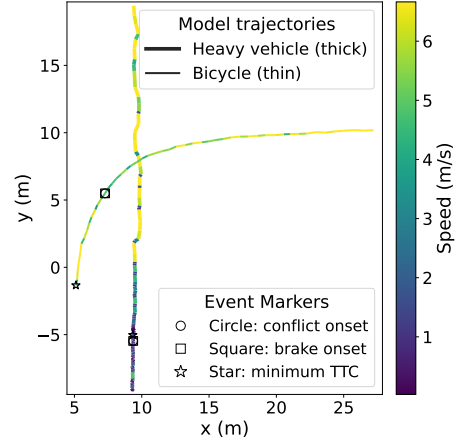


Figure 5. Spatiotemporal interaction window for the anchor heavy vehicle–bicycle encounter. Trajectories are colored by speed, and event markers indicate conflict onset, brake onset, and minimum TTC.

time-to-collision (TTC_{\parallel}) for the heavy vehicle. Let \mathbf{v}_h denote the heavy vehicle velocity and $\hat{\mathbf{e}}_{\parallel} = \mathbf{v}_h / \|\mathbf{v}_h\|$ its unit heading direction. For bicycle position \mathbf{p}_b and heavy-vehicle position \mathbf{p}_h ,

$$d_{\parallel}(t) = (\mathbf{p}_b(t) - \mathbf{p}_h(t)) \cdot \hat{\mathbf{e}}_{\parallel},$$

$$v_{\parallel}(t) = (\mathbf{v}_b(t) - \mathbf{v}_h(t)) \cdot \hat{\mathbf{e}}_{\parallel},$$

$$TTC_{\parallel}(t) = \begin{cases} \frac{d_{\parallel}(t)}{-v_{\parallel}(t)}, & v_{\parallel}(t) < 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Figure 5 shows the interaction geometry in space and time, while Fig. 6 summarizes the corresponding surrogate-safety curves. Direction-agnostic TTC decreases sharply from above 2 s to a minimum near 0.62 s, while longitudinal TTC shows only a shallow minimum around 3.1 s. This discrepancy suggests that the near-miss is not primarily constrained by the heavy vehicle’s longitudinal braking authority. Instead, it is more consistent with rapid lateral intrusion by the bicycle, which compresses the available safety margin despite sufficient stopping capacity along the truck heading. We therefore use longitudinal TTC here as a mechanism-explanation tool, not as a replacement for the primary event-mining criterion.

At larger scale within the same deployment, 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. We use this aggregate only as a recurrence indicator rather than a human-validated benchmark, but it supports the view that the anchor case is not an isolated artifact of one short window.

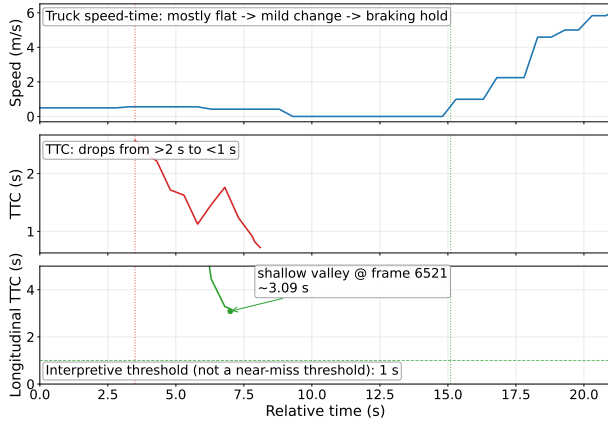


Figure 6. Anchor heavy vehicle–bicycle interaction. The x-axis shows relative time in seconds from the start of the audited interaction window. Top: heavy-vehicle speed. Middle: direction-agnostic TTC rapidly drops below 1 s. Bottom: longitudinal TTC remains above conservative braking thresholds, suggesting a lateral-intrusion-dominated interaction.

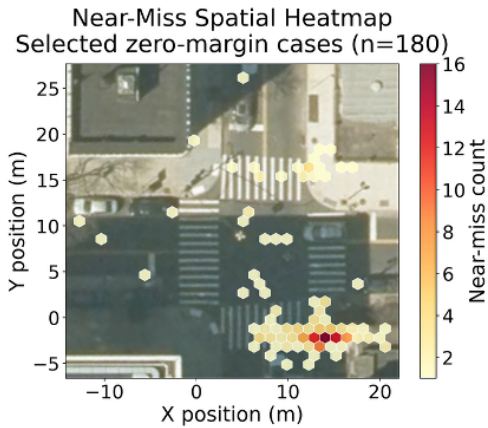


Figure 7. Post-8,000-frame vehicle–VRU near-miss overlay mined from subsequent roadside LiDAR data outside the labeled set. The label $n = 180$ denotes the number of selected zero-margin cases plotted in this heatmap. Automatically detected events remain concentrated within the same central conflict zone.

Figure 7 provides the qualitative recurrence view behind that aggregate summary. Although these later events are not human-validated one by one, the hotspot concentration persists in the same central intersection region rather than dispersing uniformly across the scene. That recurrence supports the broader claim that roadside sensing can expose site-specific conflict structure over time, even when the present paper remains focused on post-hoc auditing rather than real-time intervention.

6. Discussion and Conclusion

This pilot shows how roadside LiDAR can be used as an external safety-audit reference at a complex urban intersection. The main contribution is not a new perception backbone, but a workflow that makes infrastructure-side trajectories more reviewable through conservative refinement, repeated QA, and mechanism-aware near-miss interpretation.

The evaluation remains intentionally limited. Although we include detector results against human-reviewed ground truth, audited branch comparisons, stitched-window tracking evaluation, and human cross-checks, the paper does not yet provide a large-scale event-level benchmark across vehicle classes, traffic conditions, or multiple sites. The present claims should therefore be read as evidence of feasibility, auditability, and initial explanatory value rather than definitive reliability or full-stack trustworthiness guarantees.

Future work should expand the current pipeline to larger multi-intersection studies with more extensive model-versus-human near-miss validation, stronger identity-consistent tracking metrics, and explicit treatment of synchronization, communication, and calibration constraints. A longer journal version is ongoing work and will elaborate the broader dataset comparison, full QA dashboard evidence, post-8,000 hotspot recurrence analysis, and additional engineering extensions such as learned residual stabilization.

References

- [1] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3464–3468, 2016.
- [2] D. Gettman and L. Head. Surrogate safety measures from traffic simulation models. Technical Report FHWA-RD-03-050, Federal Highway Administration (FHWA), 2003. HTML landing page: <https://www.fhwa.dot.gov/publications/research/safety/03050/07.cfm>.
- [3] T. Huang et al. Vehicle-to-everything cooperative perception for autonomous driving: A survey. arXiv preprint, 2023.
- [4] Christer Hydén. The swedish traffic conflicts technique. Technical report / calibration study, 1987.
- [5] Jonas Jansson. *Collision Avoidance Theory: with Application to Automotive Collision Mitigation*. Phd dissertation, Linköping University, Linköping, Sweden, 2005.
- [6] Y. Ji et al. Toward autonomous vehicles: A survey on cooperative vehicle-infrastructure system. *iScience*, 2024.
- [7] C. Johnsson et al. A relative approach to the validation of surrogate measures of safety. *Accident Analysis & Prevention*, 2021.
- [8] C. Lin et al. Vehicle detection and tracking with roadside lidar using low-latency edge computing. *Sensors*, 2023.
- [9] Y. Mo et al. Enhanced perception for autonomous vehicles at intersections using roadside sensing and v2x communication. *Sensors*, 2024.

- [10] Karthikeyan Chandra Sekaran, Markus Geisler, Dominik Rößle, Adithya Mohan, Daniel Cremers, Wolfgang Utschick, Michael Botsch, Werner Huber, and Torsten Schön. Urbaning-v2x: A large-scale multi-vehicle, multi-infrastructure dataset across multiple intersections for cooperative perception. arXiv preprint, 2025. Code/dataset: <https://github.com/thi-ad/UrbanIng-V2X>.
- [11] James R. Ward, Gabriel Agamennoni, Stewart Worrall, Asher Bender, and Eduardo Nebot. Extending time to collision for probabilistic reasoning in general traffic scenarios. *Transportation Research Part C: Emerging Technologies*, 51:66–82, 2015.
- [12] M. Yazgan et al. Collaborative perception datasets in autonomous driving. arXiv preprint, 2024.
- [13] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11784–11793, 2021.
- [14] Haibao Yu et al. Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [15] Haibao Yu et al. 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)*, 2023.
- [16] Jiaxing Zhang, Chengjun Ge, Wen Xiao, Miao Tang, Jon Mills, Benjamin Coifman, and Nengcheng Chen. Roadside lidar-based scene understanding toward intelligent traffic perception: A comprehensive review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2026.
- [17] Walter Zimmer et al. Tumtraf v2x cooperative perception dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. Project page: <https://tum-traffic-dataset.github.io/tumtraf-v2x/>; arXiv: <https://arxiv.org/abs/2403.01316>.