SAFECODRIVER: POST-INTERVENTION SAFETY WITH COOPERATIVE PERCEPTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Compared with single-vehicle perception, cooperative perception provides additional information about distant or occluded traffic objects, supporting earlier identification of potential conflicts. To respond to these potential conflicts, safety interventions modify a motion planner's nominal trajectory to avoid hazards. Existing methods guide these modifications through safety rules, safety constraints, and trajectory predictions to address potential hazards. However, the modifications themselves alter traffic interactions, which can introduce new conflicts or cause unnecessary slowing or stagnation. This motivates assessing post-intervention safety, i.e., whether intervention addresses hazards without introducing new collisions or stagnation. To address this need, this paper proposes SafeCoDriver, which combines the ego vehicle's onboard observations and cooperative observations to evaluate proposed interventions before execution. Trajectory Conflict Resolution (TCR) refines the complete trajectory, Multidirectional Threat Assessment (MTA) assesses surrounding threats, and Long-Horizon Feasibility Assessment (LHFA) checks predicted collisions and future lane occupancy. By integrating these components and decoupling geometric correction from the warning output, SafeCoDriver accounts for post-intervention safety. Closed-loop experiments in SUMO and open-loop evaluations on DeepAccident show that SafeCoDriver improves safety and reduces stagnation. It achieves the lowest collision rate in SUMO and the lowest waypoint exposure on DeepAccident among the evaluated safety baselines.

1 INTRODUCTION

Cooperative perception extends a vehicle's observations beyond its onboard sensing capability (Xu et al., 2022b;a; Yu et al., 2022; Wang et al., 2024). Observations shared by nearby vehicles can reveal distant or occluded traffic objects before they enter the vehicle's onboard field of view, as illustrated in the left column of Figure 1. These shared observations support timely conflict identification and provide surrounding-traffic context for evaluating potential conflicts. Safety interventions can use this information to modify a motion planner's nominal driving plan to address potential hazards.

Existing safety interventions use safety-distance rules (Shalev-Shwartz et al., 2018), constraints that keep the vehicle within a prescribed safe set (Ames et al., 2017), and predictions of future constraint satisfaction (Mayne et al., 2005; Wabersich & Zeilinger, 2021; Pek & Althoff, 2018). Building on these approaches, recent work addresses road-boundary geometry (Xu et al., 2025), uncertainty across sensing modalities (Chang et al., 2026) and moving obstacles (Park et al., 2026). Cloud-supported planning also tracks changing collision targets under occlusion (Wan et al., 2026). These advances enable safety interventions to address a wider range of hazards in complex traffic scenarios. However, safety interventions may alter traffic interactions, introducing new conflicts or reducing driving efficiency through unnecessary slowing or stagnation. This motivates assessing post-intervention safety, i.e., whether intervention addresses hazards without introducing new collisions or stagnation. Experimental results (Tables 1) show that interventions can cause collisions that would not occur without intervention, or induce unnecessary slowing or stagnation, underscoring the practical importance of addressing these adverse consequences.

The middle column of Figure 1 illustrates these adverse consequences of intervention through two representative scenarios. In the first, braking in response to a front threat can increase rear-end risk

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

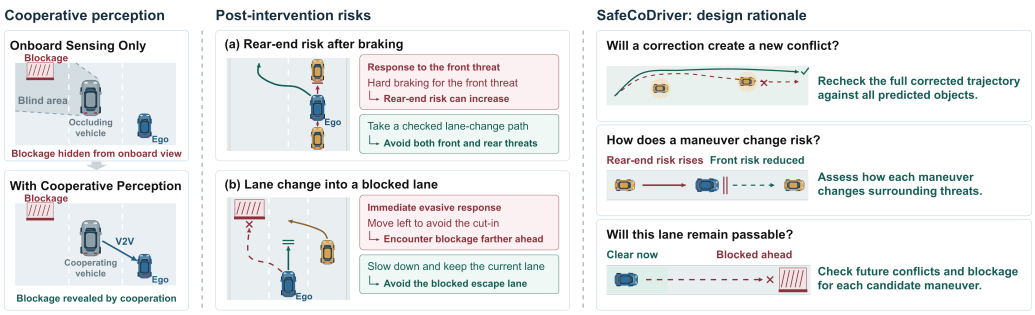

Figure 1: **Cooperative perception for post-intervention safety.** Shared observations reveal a blockage hidden from onboard sensing (left). Braking to avoid a front threat can increase rear-end risk, while a lane change to avoid a cut-in can lead into a blocked lane (middle). These scenarios motivate checks for new conflicts, surrounding threats, and continued passage (right).

when a following vehicle cannot respond in time, illustrating how an intervention can introduce a new conflict. In the second, moving left to avoid a right-side cut-in can lead the vehicle into a blocked lane, illustrating how a correction can reduce driving efficiency. These scenarios highlight the need to assess how interventions change traffic interactions and whether the resulting corrections avoid new conflicts while maintaining continued movement.

To assess and address these adverse consequences of intervention, SafeCoDriver uses onboard and cooperative observations to guide safety interventions. The method is proposed in this paper, which integrates three complementary assessments to resolve the original conflict while preserving safe passage after correction: checking and refining the nominal trajectory, assessing threats from multiple directions, and checking candidate interventions for predicted collisions and future lane occupancy. These assessments guide corrections that account for both newly introduced risks and potential blockages.

Closed-loop experiments in SUMO (Lopez et al., 2018) and open-loop evaluations on the public DeepAccident dataset (Wang et al., 2024) assess the safety benefits of SafeCoDriver. In SUMO, SafeCoDriver reduces collisions without causing stagnation. It also avoids the new collisions introduced by the evaluated baseline interventions. On DeepAccident, it achieves lower waypoint exposure than nominal plans and all evaluated safety baselines. Additional experiments examine the benefits of cooperative perception and individual components, as well as robustness to observation errors and communication delays.

The main contributions are:

- **Establishing the need to assess intervention-induced interactions.** Safety interventions can introduce collisions or stagnation in scenarios completed safely without intervention. This finding highlights the importance of assessing how interventions change traffic interactions.

- **Assessment of post-intervention safety.** To address newly introduced risks while preserving a passable route, SafeCoDriver integrates trajectory conflict resolution, multidirectional threat assessment, and checks for predicted collisions and future lane occupancy using onboard and cooperative observations. Geometric correction is decoupled from the warning output, separating risk identification from the decision to intervene.

- **Evaluation of safety and continued movement.** Closed-loop SUMO evaluations demonstrate collision-free completion, while open-loop DeepAccident evaluations show lower waypoint exposure than nominal trajectories and the evaluated safety baselines. Ablation studies clarify how the design reduces trajectory conflicts, avoids unnecessary slowing, and preserves passage beyond the immediate conflict.

2 Related Work

Responsibility-Sensitive Safety (RSS) defines braking-distance rules (Shalev-Shwartz et al., 2018), control barrier functions (CBFs) constrain admissible controls (Ames et al., 2017), and model predictive control (MPC) predicts future constraint satisfaction (Mayne et al., 2005; Wabersich & Zeilinger, 2021). These rules, constraints and predictions guide safety interventions. Recent studies address road boundaries (Xu et al., 2025) and sensing uncertainty (Chang et al., 2026), while the Dynamic Parabolic Control Barrier Function (DPCBF) uses relative-motion barriers to handle dynamic obstacles (Park et al., 2026). Under occlusion, cloud-supported planning tracks changing collision targets (Wan et al., 2026); UniE2EV2X predicts accidents from cooperative observations (Li et al., 2024). Map-Assisted Planning (MAP) incorporates map information and collision losses (Yin et al., 2025), while Risk Map as Middleware (RiskMM) couples learned risk representations with MPC (Lei et al., 2026). These advances provide solutions to safety challenges arising in complex traffic scenarios, extending the applicability of safety interventions across diverse conditions. However, the resulting changes to driving plans can alter interactions with surrounding traffic, introduce new conflicts, or cause traffic stagnation or blockage. Assessing and addressing these consequences of intervention remains an open challenge.

3 Method

To address new conflicts and stagnation that intervention may introduce, SafeCoDriver uses observations from onboard and cooperative perception to assess how candidate interventions change traffic interactions and to guide trajectory correction and action selection. Figure 2 presents the overall architecture of SafeCoDriver. Specifically, Trajectory Conflict Resolution (TCR) refines the nominal trajectory and supplies a lane preference, Multidirectional Threat Assessment (MTA) assesses surrounding threats, and Long-Horizon Feasibility Assessment (LHFA) evaluates candidate speed and lane actions for predicted conflicts and future lane occupancy. Geometric correction is decoupled from the warning output, allowing warnings to support risk identification without directly triggering unnecessary interventions. SafeCoDriver outputs a corrected trajectory and a selected driving action to guide vehicle execution toward avoiding new conflicts and maintaining continued movement.

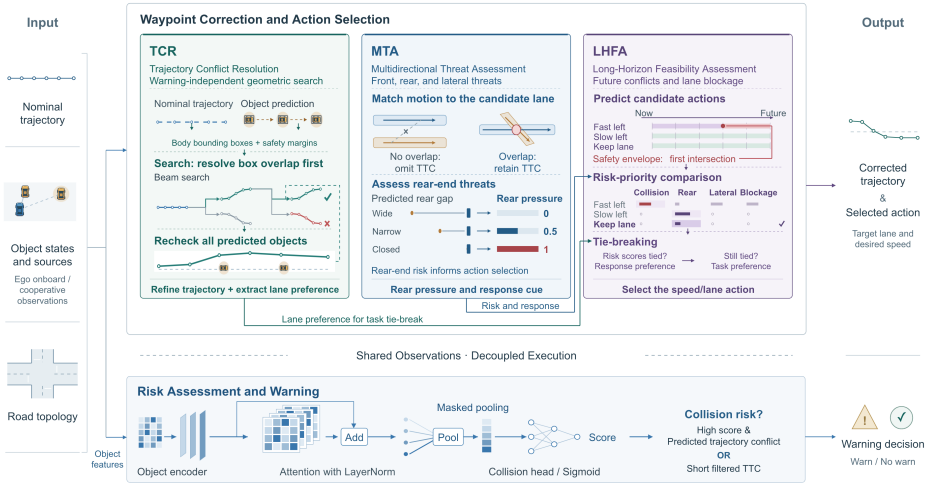

Figure 2: **SafeCoDriver architecture.** The upper branch performs trajectory correction and action selection (Sections 3.2–3.4). TCR refines the nominal trajectory, rechecks all predicted objects and extracts a lane preference. MTA assesses surrounding threats using filtered TTC and rear pressure. LHFA predicts candidate actions and selects the speed/lane action through risk-priority comparison and tie-breaking. The outputs are a corrected trajectory and a selected action. The lower branch performs risk assessment and warning using object features and geometric checks. Both branches share observations, while correction assessment runs independently of the warning decision.

3.1 Problem Formulation and Shared Observations

At control step k, SafeCoDriver receives $\mathcal{S}_k = (e_k, \mathcal{A}_k, \mathcal{V}_k)$: the ego state e_k, surrounding-object states $\mathcal{A}_k$ from ego onboard sensing and cooperative observations, and their source indicators $\mathcal{V}_k$. A driving planner supplies the nominal ego trajectory $\mathbf{w}^0_{1:T}$, comprising T future positions at times $t_j = j\Delta t$, where Δt is the waypoint interval. TCR produces corrected waypoints $\tilde{\mathbf{w}}_{1:T}$ whose lane preference guides action selection. MTA outputs filtered front/side time-to-collision (TTC) estimates and candidate-specific rear-pressure scores. LHFA selects $a^\star = (\delta, v)$, where δ is the lane offset relative to the current lane and v the target speed. A candidate-action rollout predicts the motion under these commands and the execution limits. An auxiliary branch supplies warning z_k. SafeCoDriver corrects the nominal trajectory and selects driving actions based on surrounding threats, predicted conflicts, and future lane occupancy.

Shared observations and prediction. Safety assessment uses surrounding-object observations from two sources: objects directly observed by the ego vehicle's onboard sensors and objects outside its view whose states are available through cooperative perception. Short-term tracking retains surrounding objects during gaps in these observations (Appendix E.1). Geometric checks use each tracked object's body center p_i relative to the ego body center. Surrounding-object positions are predicted in a frame fixed at the ego vehicle's current pose. Let $\hat{p}_i(t)$ be object i's predicted body center at future time t. The object's velocity in this frame is the sum of its ego-relative velocity u_i and the ego velocity $v_{\text{ego}}e_x$, where e_x is the ego forward unit vector:

$$\hat{p}_i(t) = p_i + (u_i + v_{\text{ego}}e_x)t. \tag{1}$$

An object footprint is a rectangle oriented with the object. Its axis-aligned bounding box encloses the footprint, with edges parallel to the axes of the fixed coordinate frame. A safety envelope conservatively encloses the predicted footprint with an additional safety buffer. TCR checks axis-aligned bounding boxes for overlap and separately requires the predicted center-to-center distance to be at least a safety margin $m_i(t)$. This distance threshold $m_i(t)$ is defined in Eq. (3) and differs from the spatial buffer used to construct safety envelopes. Each component of b^0_i is the sum of the bounding-box half-sizes of the ego vehicle and surrounding object i along the corresponding coordinate axis. Objects predicted to conflict with the ego vehicle's nominal trajectory form the following set:

$$\mathcal{T}_k = \{i : \exists j \in \{1, \dots, T\},\ |w^0_j - \hat{p}_i(t_j)| < b^0_i \ \vee\ \|w^0_j - \hat{p}_i(t_j)\|_2 < m_i(t_j)\}. \tag{2}$$

The coordinate-wise first condition detects body-bounding-box overlap; the second detects a safety-margin violation. These two conditions define a trajectory conflict.

3.2 Trajectory Conflict Resolution

Resolving one trajectory conflict can introduce another at a different waypoint or with another surrounding object. TCR therefore rechecks the complete trajectory after every edit. It first identifies body-bounding-box overlaps and generates edits beyond their boundaries; once overlap is absent, safety-margin violations guide further edits. Each revised trajectory is checked against predictions of all available tracked objects. Candidates are ranked by overlap measures, margin violations, and edit size and smoothness. A fixed number of top-ranked candidates are retained at each search depth. The search returns the best-ranked proposal within its computation budget, supplying corrected trajectory and the lane preference defined below.

Source-Conditioned Safety Margins. TCR assigns a safety margin according to surrounding object i's observation source $\sigma(i) \in \{\text{vis}, \text{coop}\}$, denoting ego-visible and cooperative-only observations:

$$m_i(t) = (m_0 + \Delta_{\sigma(i)})\kappa(v^{\text{app}}_i) + \beta d_i. \tag{3}$$

Here, m_0 is the base margin and $\Delta_{\sigma(i)}$ the source-dependent allowance. The bounded multiplier κ depends on object i's radial approach speed v^{app}_i relative to the ego vehicle. The coefficient β scales the size reserve $d_i = \max(l_i, w_i)$, where l_i and w_i are object i's length and width. Body-bounding-box overlap uses the combined half-extents b^0_i defined in Section 3.1. LHFA's safety envelopes also account for each object's observation source.

TCR bounds the beam search and lateral waypoint displacements. A lane preference is derived from the first K_{pref} corrected waypoints, where K_{pref} is the preference window length. When the maximum absolute lateral edit reaches threshold d_{pref}, the median signed edit determines the preferred direction; otherwise, no lane preference is assigned.

3.3 MULTIDIRECTIONAL THREAT ASSESSMENT

Surrounding objects with short radial time to collision (TTC), estimated from ego-relative distance and approach speed, may follow a separate, nonconflicting path. An intervention can also reduce one threat while increasing another. This motivates avoiding unnecessary interventions that may introduce additional risks. MTA uses relative motion and predicted safety envelopes to distinguish two cases: laterally separated parallel motor vehicles with disjoint safety envelopes, and objects that require conservative TTC checks, including crossing objects, pedestrians, and bicycles. The former are excluded from the front/side TTC assessment, while conservative checks are retained for the latter. For each candidate action, rear pressure is scored using the predicted longitudinal center separation between the ego vehicle and approaching followers at the end of the prediction horizon. The score varies continuously from 0 to 1, with smaller separation yielding higher scores; the values 0, 0.5, and 1 in Figure 2 are illustrative examples.

Let $\tau_{\mathrm{fs}}^{\mathrm{MTA}}$ be the minimum retained TTC for objects in front of or beside the ego vehicle, and let $v_r = \max(v_{\mathrm{ego}}, v_{\min})$ be the reference speed with lower bound $v_{\min}$. The immediate-response indicator $I_{\mathrm{imm}}(a)$ identifies candidate actions that provide insufficient slowing when the retained TTC falls below the response threshold. Given the response threshold τ_{imm} and speed ratio ρ_{imm}, the indicator is defined as

$$I_{\mathrm{imm}}(a) = \mathbf{1}\{\tau_{\mathrm{fs}}^{\mathrm{MTA}} < \tau_{\mathrm{imm}} \ \wedge \ v(a) > \rho_{\mathrm{imm}} v_r\}. \tag{4}$$

Here, $v(a)$ is the candidate's target ego speed; the indicator equals 1 when both conditions hold and 0 otherwise. MTA supplies rear-pressure scores and the immediate-response indicator as inputs to LHFA's candidate-action evaluation.

3.4 LONG-HORIZON FEASIBILITY ASSESSMENT

A lane change that avoids an immediate conflict may encounter blockage beyond the initial avoidance point. LHFA therefore evaluates candidate speed/lane interventions over lookahead horizon H, using MTA's threat scores and TCR's lane preference under the prescribed priority order.

Candidate speeds are paired with lane keeping and road-valid adjacent lane changes to form the candidate action set $\mathcal{C}_{\mathrm{exec}}$. Target speeds are capped at the initial cruise speed, and the vehicle adjusts its speed within the acceleration and braking limits. LHFA checks each candidate continuously over H, including a current-lane reaction step and the complete lateral sweep of a lane change. The safety-envelope checks account for changes in the ego vehicle's orientation during a lane change and include safety buffers based on each surrounding object's observation source. Let $t_i(a)$ denote the predicted time until the ego vehicle's safety envelope first intersects that of surrounding object i under candidate action a. If no intersection is predicted within H, set $t_i(a) = +\infty$ to indicate that no intersection is detected within the prediction horizon. For each candidate action a, the predicted time to the first safety-envelope intersection with surrounding object i is converted into a conflict score $r_i(a)$. The maximum score across surrounding objects, $V_{\mathrm{coll}}(a)$, is the primary risk criterion used by LHFA to compare candidate actions:

$$r_i(a) = \begin{cases} [1 - t_i(a)/H]_+, & t_i(a) < \infty, \\ 0, & \text{otherwise}, \end{cases} \qquad V_{\mathrm{coll}}(a) = \max_i r_i(a). \tag{5}$$

Here, $[x]_+ = \max(x, 0)$. Intersections closer to the start of the prediction horizon receive higher scores, while those closer to its end receive lower scores. The vector $\mathbf{V} = (V_{\mathrm{coll}}, V_{\mathrm{rear}}, V_{\mathrm{lane}}, V_{\mathrm{block}})$ combines this conflict score with scores for rear pressure, lateral safety-envelope intersection, and target-lane blockage. Rear pressure measures followers' predicted terminal separation from ego. Lateral-intersection and blockage scores use the corresponding object subsets from the same safety-envelope checks. These terms distinguish candidates with equal maximum conflict scores.

SafeCoDriver uses a prescribed priority order motivated by minimum-violation planning (Censi et al., 2019; Penlington et al., 2024). Overall conflict risk precedes role-specific threats, the immediate-response preference, and task cost. Role-specific terms distinguish candidates with equal overall risk scores. The first unequal score decides the choice:

$$a^\star = \arg \min_{\mathrm{lex},\, a \in \mathcal{C}_{\mathrm{exec}}} \big(V_{\mathrm{coll}}(a), V_{\mathrm{rear}}(a), V_{\mathrm{lane}}(a), V_{\mathrm{block}}(a), I_{\mathrm{imm}}(a), R_{\mathrm{task}}(a)\big). \tag{6}$$

Task cost $R_{\text{task}}(a)$ encodes task preferences and breaks ties only after the preceding risk scores and immediate-response indicator: it favors ego cruise speed, lane keeping over a lane change, and agreement with the TCR lane preference. At low ego speeds, the lane-change penalty is reduced relative to its cruise-speed value, favoring a moving escape over remaining stopped when risk scores tie. Appendix E.1 defines the scores and tie-breaking rules.

Warning-independent correction. Predicted geometric conflicts may require correction even when no warning is issued, while issuing a warning does not necessarily require a trajectory change. The correction branch therefore runs independently of the warning output. The auxiliary warning combines an attention-based score from surrounding-object features with the ego nominal trajectory's geometric conflicts and MTA-filtered TTC. An ablation study evaluates the benefit of decoupling correction from warning. Appendix E.2 gives the fusion rule, network and training settings.

Applying the intervention. The selected action is designed to be executable. If execution fails, feasible alternatives are reassessed using the same risk criteria.

4 Experiments

Closed-loop experiments in SUMO (Lopez et al., 2018) and open-loop evaluations on the public DeepAccident dataset (Wang et al., 2024) evaluate the safety and effectiveness of SafeCoDriver's interventions. Matched driving episodes demonstrate the need to assess post-intervention safety by showing that safety interventions can introduce collisions or stagnation into runs successfully completed without intervention. Overall comparisons show that SafeCoDriver reduces collision rates in simulated driving and waypoint exposure on public driving data relative to the evaluated safety baselines. Further experiments examine the benefits of cooperative perception, individual component contributions, and robustness to position errors and communication delays.

4.1 Experimental Setup

All compared safety methods use onboard and cooperative observations, except the onboard-only control. The evaluation includes eight SUMO scenario families with 150 matched episodes each and 104 DeepAccident evaluation scenarios comprising 8,193 frames. Within each benchmark, methods share the same observation inputs, nominal driving plans and evaluation settings. Scenarios used for development, warning-model training and checkpoint selection are separate from those used for evaluation. All method parameters remain fixed during evaluation. Appendix E.2 describes warning-network training and checkpoint selection; Appendix E.3 describes baseline adaptations and RiskMM-SC training.

Baselines. The baselines include RSS braking-distance rules (Shalev-Shwartz et al., 2018), CBF-SC's barrier-constrained speed correction (Ames et al., 2017), DPCBF's relative-motion barriers (Park et al., 2026), UniE2EV2X-SC's polygon-distance check (Li et al., 2024), MAP-SC's overlap loss (Yin et al., 2025), and RiskMM-SC's learned risk and MPC costs (Lei et al., 2026). SC means *Safety Component Adaptation*. Joint MPC is the joint lane-and-speed baseline implemented here using predictive constrained optimization (Mayne et al., 2005; Wabersich & Zeilinger, 2021). Appendix E.3 specifies each adaptation and action mapping. Full denotes complete SafeCoDriver; Nominal denotes driving without intervention.

Metrics. Collision rate (Coll.) is the fraction of episodes with an ego-involving simulator collision. Intervention-added collision rate (ACR) counts new collisions among matched episodes collision-free under nominal driving (310 in the main benchmark). Completion (Compl.) requires collision-free arrival or continued driving without terminal stagnation through the evaluation window. Terminal stagnation (Stag.) means no arrival and less than 1 m traveled in the final 5 s. Progress averages traveled distance in meters, assigning zero to failed-completion episodes. Arrival means reaching the route end; targeted tests observe arrival times through 60 s. Waypoint exposure (WPC) is the fraction of proposed ego waypoints within 2 m of a predicted object center. On DeepAccident, a warning, geometric conflict or waypoint intervention activates a response. Detection (Det), false alarm (FA) and activation (Act) measure response rates over dangerous frames (1–30 frames before collision), remaining frames and all frames, respectively. Appendices A and E.4 detail the metrics and evaluation parameters.

4.2 Intervention-Added Collisions and Stagnation

Matched SUMO experiments show that safety interventions can introduce collisions and stagnation in scenarios completed safely without intervention, demonstrating the need to assess intervention-induced changes in traffic interactions. Table 1 groups nominally collision-free episodes by completion or stagnation. Every evaluated safety baseline introduces collisions in the completed group. RSS and CBF-SC also cause collision-free terminal stagnation, whereas SafeCoDriver completes all 310 episodes across both groups without collision or stagnation. These results demonstrate SafeCoDriver's ability to intervene without introducing new collisions or stagnation. Examining the failures, RSS causes rear-end collisions through braking, illustrating the interaction in Figure 1. Joint MPC introduces conflicts with slower vehicles through lane changes where lateral space is limited. With the same observation inputs, Joint MPC also introduces collisions in vehicle-queue scenarios that are collision-free without intervention. These findings reinforce the need to assess intervention-induced conflicts and stagnation. Appendix C details the RSS rear-end and Joint MPC lane-change cases.

Table 1: Intervention-added collisions and stagnation in matched episodes. Overall ACR pools both nominally collision-free groups (310 episodes).

Safety policy	Nominal completed		Nominal stagnant	Overall ACR↓
	Added collisions↓	Added stagnation↓	Added collisions↓	
RSS	4/154	2/154 (1.30%)	**0/156**	1.29%
CBF-SC	11/154	9/154 (5.84%)	123/156	43.23%
DPCBF	1/154	**0/154 (0.00%)**	**0/156**	0.32%
UniE2EV2X-SC	10/154	**0/154 (0.00%)**	**0/156**	3.23%
MAP-SC	2/154	**0/154 (0.00%)**	41/156	13.87%
RiskMM-SC	139/154	**0/154 (0.00%)**	156/156	95.16%
Joint MPC	2/154	**0/154 (0.00%)**	28/156	9.68%
SafeCoDriver	**0/154**	**0/154 (0.00%)**	**0/156**	**0.00%**

4.3 Overall Safety and Progress

SafeCoDriver completes all 1,200 SUMO episodes without collisions and achieves the lowest DeepAccident WPC, reducing exposure from 3.767% under nominal plans to 0.380% (Table 2). These results demonstrate safety benefits in both simulated driving and public-data evaluation. The baseline comparisons highlight the balance between safety and continued movement: DPCBF achieves a low collision rate but completes only 29.8% of episodes because of frequent stagnation, whereas Joint MPC achieves comparable Progress but incurs 83 collisions. SafeCoDriver combines collision-free driving with full completion and the highest Progress. On DeepAccident, MAP-SC has the lowest false-alarm rate but also the lowest detection rate. RSS and RiskMM-SC reach 100% detection with 100% false alarms and substantially higher WPC than SafeCoDriver. SafeCoDriver combines 95.5% detection, 84.8% false alarms and the lowest WPC, balancing detection and unnecessary responses. Independent seeds reproduce collision-free completion (Appendix B.1).

Table 2: Main comparison with cooperative observations. 1200 matched SUMO episodes and 104 DeepAccident scenarios. ACR uses 310 nominally collision-free episodes. Act has no preferred direction.

	SUMO					DeepAccident			
Method	Coll.↓	ACR↓	Compl.↑	Progress↑	WPC↓	WPC↓	Det↑	FA↓	Act
No safety intervention	74.17%	–	12.8%	22.02	17.798%	3.767%	–	–	0.0%
RSS (2018)	12.50%	1.29%	12.3%	6.05	0.069%	9.309%	**100.0%**	100.0%	100.0%
CBF-SC (2017)	47.83%	43.23%	26.3%	27.77	1.267%	4.049%	97.4%	93.8%	94.2%
DPCBF (2026)	0.25%	0.32%	29.8%	39.13	0.488%	4.566%	99.2%	97.3%	97.6%
UniE2EV2X-SC (2024)	41.08%	3.23%	24.2%	16.15	0.063%	1.278%	91.8%	89.1%	89.4%
MAP-SC (2025)	41.08%	13.87%	58.0%	98.05	0.053%	0.585%	53.8%	**37.9%**	39.9%
RiskMM-SC (2026)	73.75%	95.16%	26.2%	31.71	1.655%	3.606%	**100.0%**	100.0%	100.0%
Joint MPC	6.92%	9.68%	93.1%	177.59	0.069%	0.915%	96.2%	82.1%	83.9%
SafeCoDriver (ours)	**0.00%**	**0.00%**	**100.0%**	**178.23**	**0.043%**	**0.380%**	95.5%	84.8%	86.2%

4.4 VALUE OF COOPERATIVE PERCEPTION

Table 3 examines the contribution of cooperative observations. Compared with onboard sensing alone, cooperative observations improve collision avoidance, completion and Progress in SUMO, while reducing waypoint exposure on both benchmarks. In blockage scenarios like the one illustrated in Figure 1, cooperative perception provides observations of an adjacent-lane obstruction in the ego vehicle's blind spot. SafeCoDriver uses this information to assess the escape route and select a correction that avoids both predicted conflicts and subsequent stagnation. These results demonstrate the value of cooperative observations in improving intervention safety and preserving continued movement.

Table 3: Cooperative vs. onboard-only SafeCoDriver. 1200 SUMO episodes and 104 DeepAccident scenarios.

	SUMO					DeepAccident
Variant	Coll.↓	ACR↓	Compl.↑	Progress↑	WPC↓	WPC↓
SafeCoDriver (onboard-only)	12.50%	**0.00%**	87.5%	163.10	0.085%	0.522%
SafeCoDriver (cooperative)	**0.00%**	**0.00%**	**100.0%**	**178.23**	**0.043%**	**0.380%**

4.5 COMPONENT BENEFITS

Table 4 tests four component controls on all eight SUMO families and DeepAccident; Tables 5 and 6 add 200 targeted episodes each to examine MTA and LHFA in their intended traffic conditions. Removing TCR increases WPC on both benchmarks, showing its role in reducing trajectory conflicts. MTA avoids repeated braking and improves Progress while maintaining collision-free completion. LHFA and warning-independent correction reduce collisions. The MTA control's slightly lower WPC comes with slower driving, whereas the LHFA control's low Progress partly reflects collision failures counted as zero distance. Full combines collision avoidance with continued movement. Appendices E.3 and B.2 specify the controls and further search, horizon and margin comparisons.

Table 4: Component ablations with cooperative observations. The MTA control replaces MTA with conventional nondirectional TTC rather than removing TTC response. 1200 matched SUMO episodes and 104 DeepAccident scenarios.

	SUMO					DeepAccident
Variant	Coll.↓	ACR↓	Compl.↑	Progress↑	WPC↓	WPC↓
w/o TCR	**0.00%**	**0.00%**	**100.0%**	**178.23**	0.242%	3.767%
w/o MTA	**0.00%**	**0.00%**	**100.0%**	164.49	**0.029%**	**0.380%**
w/o LHFA	52.33%	64.52%	47.7%	79.52	0.593%	**0.380%**
w/o warning decoupling	2.25%	5.48%	97.8%	152.37	0.956%	1.235%
SafeCoDriver (Full)	**0.00%**	**0.00%**	**100.0%**	**178.23**	0.043%	**0.380%**

MTA: avoiding unnecessary slowing. Across 200 targeted episodes, MTA recognizes that queued vehicles in an adjacent lane do not obstruct the ego vehicle's path, avoiding the repeated braking triggered by nondirectional TTC. Both configurations remain collision-free, while MTA shortens mean arrival time from 40.55 to 18.61 s and increases Progress by 110.20 m (Table 5). These results show that MTA distinguishes nearby traffic from threats to the ego vehicle's movement, improving driving efficiency without sacrificing collision avoidance (Appendix B.3).

LHFA: reducing blockage and recovery time. Across 200 targeted episodes, LHFA checks whether a lane change that avoids a cut-in leads toward a distant blockage. With TCR and MTA unchanged, extending the checking horizon from 1.5 to 5 s eliminates stagnation at the end of the fixed evaluation window, raises completion from 47.5% to 100%, and shortens mean arrival time from 34.22 to 24.21 s (Table 6). With rear-pressure checking unchanged, longer envelope checking reduces blockage-related waiting and improves recovery. On the main benchmark, extending the checking horizon reduces action PFE, the proportion of predicted states with physical-body overlap, from 6.811% to 0.388%, alongside lower Progress. In the targeted blockage tests, jointly shortening the envelope and rear-pressure checking horizons reduces arrival time. Together, these comparisons illustrate the safety–efficiency trade-off associated with prediction horizons (Appendix B.4).

Table 5: Effect of MTA on unnecessary slowing in 200 paired adjacent-lane queue scenarios.

Metric	w/o MTA	SafeCoDriver (Full)
Coll.↓	**0.0%**	**0.0%**
Stag.↓	**0/200**	**0/200**
Compl.↑	**100.0%**	**100.0%**
Progress↑	109.12	**219.31**
Arrival (s)↓	40.55	**18.61**

Table 6: Effect of LHFA on blockage avoidance in 200 cut-in scenarios with a blocked adjacent lane.

Metric	w/ short-horizon LHFA (1.5 s)	SafeCoDriver (Full)
Coll.↓	**0.0%**	**0.0%**
Stag.↓	105/200	**0/200**
Compl.↑	47.5%	**100.0%**
Progress↑	105.00	**205.98**
Arrival (s)↓	34.22	**24.21**

4.6 Robustness to Position Errors and Communication Delays

Table 7 evaluates joint position errors and communication delay in cooperative observations, while onboard observations remain current. DPCBF and Joint MPC are selected for the lowest baseline collision rate and highest baseline completion and Progress in the main benchmark, respectively. Under joint errors, SafeCoDriver completes all 200 SUMO episodes with a responsive following vehicle without collisions or stagnation. DPCBF maintains a low collision rate but frequently stagnates, limiting completion to 41%. Joint MPC achieves slightly higher Progress and lower WPC, but records seven collisions, including two introduced into nominally collision-free episodes. On DeepAccident, SafeCoDriver retains the lowest WPC among the three methods, although exposure rises from 0.380% with clean observations to 0.496%. These results show that its safety and continued-movement benefits persist under the tested input errors (Appendices B.6 and D.1).

Table 7: Robustness to a 1 m cooperative-position bias and a 0.3 s communication delay. SUMO uses 200 matched episodes across five scenario families; ACR uses 160 nominally collision-free episodes. DeepAccident uses 104 scenarios with 8,193 frames.

	SUMO					DeepAccident
Method	Coll.↓	ACR↓	Compl.↑	Progress↑	WPC↓	WPC↓
DPCBF	0.50%	**0.00%**	41.00%	61.94	0.4759%	4.603%
Joint MPC	3.50%	1.25%	96.50%	**223.72**	**0.0012%**	1.003%
SafeCoDriver	**0.00%**	**0.00%**	**100.00%**	218.34	0.0033%	**0.496%**

5 Conclusion

Matched experiments show that safety interventions may alter traffic interactions and introduce collisions or stagnation in scenarios completed safely without intervention, establishing the need for post-intervention safety assessment. SafeCoDriver addresses this problem using onboard and cooperative observations to integrate trajectory refinement, multidirectional threat assessment, and checks for predicted collisions and lane occupancy. Geometric correction is decoupled from warning output to support risk identification without directly triggering unnecessary interventions.

Closed-loop SUMO experiments demonstrate collision-free completion on the main benchmark, while open-loop DeepAccident evaluations show lower waypoint exposure than nominal plans and the evaluated baselines. Component studies support the benefits for safety and continued movement. Limitations include simplified motion predictions and collisions under combined position errors and communication delays, motivating further evaluation with uncertain observations and responsive traffic.

Reproducibility Statement

Appendix A specifies the evaluation protocols and metrics. Appendix E describes the implementation, warning-network training, baseline adaptations and parameter settings. Appendices B–D present supplementary experiments and describe experiments' conditions and control configurations.

AI USE STATEMENT

Generative AI tools assisted with data analysis, translation, manuscript editing and formatting, code inspection and debugging, and limited coding assistance. The overall research design and primary implementation were carried out by the authors, who take responsibility for the final analyses, code, and manuscript.

REFERENCES

Aaron D. Ames, Xiangru Xu, Jessy W. Grizzle, and Paulo Tabuada. Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control*, 62(8): 3861–3876, 2017.

Andrea Censi, Konstantin Slutsky, Tichakorn Wongpiromsarn, Dmitry Yershov, Scott Pendleton, James Fu, and Emilio Frazzoli. Liability, ethics, and culture-aware behavior specification using rulebooks. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 8536–8542, 2019.

Pei Yu Chang, Qizhe Xu, Vishnu Renganathan, and Qadeer Ahmed. Risk aware safe control with multi-modal sensing for dynamic obstacle avoidance. *Control Engineering Practice*, 175:107114, 2026. doi: 10.1016/j.conengprac.2026.107114.

Mingyue Lei, Zewei Zhou, Hongchen Li, Jiaqi Ma, and Jia Hu. Risk map as middleware: Toward interpretable cooperative end-to-end autonomous driving for risk-aware planning. *IEEE Robotics and Automation Letters*, 11(1):818–825, 2026. doi: 10.1109/LRA.2025.3636031.

Zhiwei Li, Bozhen Zhang, Lei Yang, Tianyu Shen, Nuo Xu, Ruosen Hao, Weiting Li, Tao Yan, and Huaping Liu. Unified end-to-end V2X cooperative autonomous driving. arXiv preprint arXiv:2405.03971, 2024.

Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. Microscopic traffic simulation using SUMO. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2575–2582, 2018. doi: 10.1109/ITSC.2018.8569938.

D. Q. Mayne, M. M. Seron, and S. V. Raković. Robust model predictive control of constrained linear systems with bounded disturbances. *Automatica*, 41(2):219–224, 2005.

Hun Kuk Park, Taekyung Kim, and Dimitra Panagou. Beyond collision cones: Dynamic obstacle avoidance for nonholonomic robots via dynamic parabolic control barrier functions. In *IEEE International Conference on Robotics and Automation*, 2026.

Christian Pek and Matthias Althoff. Computationally efficient fail-safe trajectory planning for self-driving vehicles using convex optimization. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1447–1454, 2018.

Matteo Penlington, Alessandro Zanardi, and Emilio Frazzoli. Optimization of rulebooks via asymptotically representing lexicographic hierarchies for autonomous vehicles. *arXiv preprint arXiv:2409.11199*, 2024.

Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. On a formal model of safe and scalable self-driving cars. arXiv preprint arXiv:1708.06374v6, 2018.

Kim Peter Wabersich and Melanie N. Zeilinger. A predictive safety filter for learning-based control of constrained nonlinear dynamical systems. *Automatica*, 129:109597, 2021. doi: 10.1016/j.automatica.2021.109597.

Keke Wan, Wei Zhong, Shuyan Li, Fachao Jiang, Yuming Ge, Danyang Chang, Ao Zhang, Chen Lv, and Bolin Gao. A cloud-supported proactive safety decision-making and control system for rear-end collision avoidance in occluded emergency scenarios: Design and joint validation across OEMs. *Accident Analysis & Prevention*, 236:108682, 2026. doi: 10.1016/j.aap.2026.108682.

Tianqi Wang, Sukmin Kim, Wenxuan Ji, Enze Xie, Chongjian Ge, Junsong Chen, Zhenguo Li, and Ping Luo. DeepAccident: A motion and accident prediction benchmark for V2X autonomous driving. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(6):5599–5606, 2024. doi: 10.1609/aaai.v38i6.28370.

Jianye Xu, Chang Che, and Bassam Alrifaee. A real-time control barrier function-based safety filter for motion planning with arbitrary road boundary constraints. In *2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2818–2825, 2025. doi: 10.1109/ITSC60802.2025.11423203.

Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma. V2X-ViT: Vehicle-to-everything cooperative perception with vision transformer. In *Computer Vision – ECCV 2022*, pp. 107–124. Springer, 2022a.

Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. OPV2V: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication. In *2022 International Conference on Robotics and Automation (ICRA)*, pp. 2583–2589, 2022b. doi: 10.1109/ICRA46639.2022.9812038.

Huilin Yin, Yiming Kan, and Daniel Watzenig. MAP: End-to-end autonomous driving with map-assisted planning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pp. 1823–1830, 2025.

Haibao Yu, Yizhen Luo, Mao Shu, Yiyi Huo, Zebang Yang, Yifeng Shi, Zhenglong Guo, Hanyu Li, Xing Hu, Jirui Yuan, and Zaiqing Nie. 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)*, pp. 21361–21370, 2022.