MC-Risk: Multi-Component Risk Fields for Risk Identification and Motion Planning

Maximilian Link *
Technical University of Munich
Munich, Germany
link.maximilian@tum.de

Yingjie Xu*
Technical University of Munich
Munich, Germany
yingjie.xu@tum.de

Yingbai Hu
The Chinese University of Hong Kong
Hong Kong, China
yingbai.hu@tum.de

Yinlong Liu †
City University of Macau
Macau, China
ylliu@cityu.edu.mo

Abstract

We present **MC-Risk**, a planner-aligned, multi-component risk field on a bird's-eye-view grid that yields early, calibrated, and class-aware risk localization. MC-Risk linearly composes three interpretable modules: (i) a motorized-agent field that fuses a black-box multimodal trajectory predictor with an analytic Gaussian-torus construction whose lateral width grows with speed/curvature and whose height attenuates with look-ahead; (ii) a VRU risk field that replaces isotropic pedestrian blobs with a forward-biased anisotropic kernel aligned to heading and speed; and (iii) a road penalty field that exploits full HD-map topology, imposing an off-road penalty and lane-aware risk exposure for same/opposite directions. We conduct, to our knowledge, the first standardized quantitative evaluation of a risk-field formulation on RiskBench's collision subset. MC-Risk attains the best overall risk localization and the earliest hazard indication. Finally, we demonstrate a plug-and-play planning interface by using the field as an MPC cost density, enabling risk-aware trajectory generation without additional training.

1 Introduction

Building autonomous vehicles that can localize risk early and reliably in cluttered traffic is a core safety challenge. Most planners need explainable, stable risk/cost evaluation with high spatial resolution that reflects different traffic agents (cars, trucks, motorcycles, pedestrians) and rich map semantics. A wide range of approaches has emerged: classical potential-risk-field formulations provide spatially planner-compatible abstractions of scenario danger [1–4]; data-driven risk assessment methods estimate collision likelihoods or required decelerations from perception and forecasting pipelines [5–8], often leveraging efficient transformer backbones to improve robustness and latency [9]; and recent works couple multimodal forecasting with risk fields to propagate behavioral uncertainty over space and time [10]. Equally important, standardized benchmarks are needed to measure both the quality and timeliness of risk perception and identification. RiskBench addresses this by providing a scaled dataset and a suite of risk-centric metrics [11].

*Equal contribution.
†Corresponding author.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Embodied and Safe-Assured Robotic Systems.

Despite the progress, three points limit the current approaches. First, the balance between interpretability and scenario coverage: learning-based risk predictors expose comprehensive but black-box results, while classical field theories often underfit real road diversity by using class-agnostic envelopes (e.g., fixed vehicle tubes or isotropic pedestrian blobs) that neglect velocity, possible future trajectory, and forward intent [3], so finding the balance between these two aspects is still under exploring. Second, map semantics: penalties for off-road regions and opposing lanes are frequently implemented as a fixed buffer, which either over-penalize plausible maneuvers or weaken the ability to identify true risks [4]. Third, evaluation rigor: to the best of our knowledge, no prior risk-field formulation has been quantitatively evaluated on a standardized and public benchmark. Most reports are qualitative, lacking actor-set alignment [10]. Consequently, it remains difficult to obtain risk maps that are early, comprehensive, calibrated, and to demonstrate these properties under a standardized benchmark.

We introduce **MC-Risk**, a multi-component risk-field representation on a BEV grid that linearly composes three interpretable modules: (i) a motorized-agent field (MAF) that combines a black-box multimodal trajectory predictor with an explainable Gaussian-torus construction. Its lateral width varies with speed and curvature, while height attenuates with look-ahead distance, yielding earlier and more precise risk identification. (ii) a Vulnerable Road User risk field (VRF) that replaces symmetric blobs with a forward-biased, anisotropic kernel aligned to heading and speed, improving risk identification recall for approaching pedestrians or other VRUs. (iii) a road penalty field (RPF) that leverages full HD-map topology, including off-road, same-direction adjacent lanes, and opposite-direction lanes, to encode a structure-aware risk field rather than manually-defined buffers.

We integrate MC-Risk into RiskBench [11] and conduct standardized evaluations on its collision subset, where risk is most prominent. We compute actor-level risk by taking the maximum field value over each actor and thresholding. Across baselines from four families (rule-based, forecast+check, collision anticipation, behavior prediction), MC-Risk achieves the strongest overall risk localization and the earliest hazard indication. Additionally, we integrate the risk map into a standard MPC objective function as a plug-and-play cost, enabling risk-aware motion planning without any additional training.

Contributions of this work.

- We propose a modular, planner-aligned risk-field formulation that couples black-box prediction with explainable analytical fields: velocity-/curvature-conditioned MAF, forward-biased VRF, and topology-aware RPF.
- We conduct a RiskBench-compliant evaluation with a visibility adapter and a unified metric suite (OT-F1, OT-F1-T at 1/2/3 s, wMOTA, PIC), reported on the collision subset, achieving the best overall performance.
- We propose a planning interface that treats the aggregated field as a cost density for MPC, enabling risk-aware trajectory selection without additional training.

2 Related Works

Time/Distance Proximity Methods. Classical indicators—including time headway (TH) [12], time-to-collision (TTC) [13], and their temporally aggregated variants TET/TIT [14], PET [15], etc. offer low-cost signals of closing gaps and exposure [16, 17]. Although intuitive and efficient, they assume simple kinematics, lack uncertainty modeling, and require threshold tuning, limiting the application in multi-agent, multimodal settings.

Formal Safety–Guarantee Methods. Probabilistic survival processes fuse predicted position uncertainty into an integrated collision risk rate (Risk-Spot Detector, RSD) [16]. Responsibility-Sensitive Safety (RSS) derives longitudinal/lateral safe distances and scores risk by violation margin [18]. Hierarchical analytic-network processes quantify scenario risk from layered static/dynamic factors with fuzzy uncertainty handling [19]. These methods provide principled safety guarantees but can be conservative or computationally heavy.

Potential-Field Methods. Risk is often encoded as spatial fields that repel the ego from hazards while attracting it to goals. Liu et al. define distinct fields for vehicles, pedestrians, and road

boundaries, combined into a unified driving-risk field with optional attractive terms for navigation [4]. NVIDIA's Safety Force Field (SFF) and its differentiable reproductions compute a safety potential over claimed trajectory sets and apply its negative gradient as an intervention force [1, 20]. These fields yield a unified representation but can be sensitive to parameterization and not understanding the scenario comprehensively. EDRF [10] improves realism over purely kinematic fields based on multimodal predicted trajectories (e.g., through QCNet [21], yet its published formulation focuses only on motor vehicles and shows limited map semantics.

Learning-Based Methods. Collision-anticipation networks (e.g., DSA [22], RRL [23]) score collision likelihood from spatiotemporal features without explicit overlap checks. Trajectory predictors (Social-GAN [24], MANTRA [25], QCNet [21]) enable collision checking by future path overlap. Behavior-influence models (BP/BCP) infer risk from predicted deviations to the ego intention via attention or causal masking [26, 27]. Recent systems also embed risk into transformer decoders [5] or use LLMs for evaluation and reasoning [28, 29]. Complementary advances in perception, such as event/frame fusion for detection, offers robust inputs to downstream risk estimation [30, 31]. These methods achieve strong anticipation but are often black-box and may lack planner-ready structure.

We bridge these gaps by coupling a black-box multimodal predictor with an explainable field to produce a class-aware, topology-informed risk map that comprehensively covers all traffic participants and road geometry to remain planner-aligned.

3 Methods

Our goal is to construct a calibrated BEV risk density map $R_{scene}(x, y)$ that motion planners can utilize directly. This section formalizes MC-Risk by composing probability (from trajectory prediction) and consequence (class- and velocity-conditioned severity) into motorized agents, VRU, and road fields. We begin with an overview of the overall definition in Sec. 3.1, followed by detailed descriptions of the three key components: the motorized-agent field based on trajectory prediction in 3.2; the forward-biased anisotropic VRU risk field in Sec. 3.3; the road penalty field via off-road costs in Sec. 3.4.

3.1 Overview

We discretize a local BEV grid $\Omega \subset \mathbb{R}^2$ around the ego vehicle. As suggested by Puphal et al. [32], for visualization or a conceptual combined use outside of the ego-vehicle risk assessment, the field components can be superimposed, so our MC-Risk comprises component fields:

$$R_{scene}(x, y) = \sum_{i=1}^{N_{mot}} \mathrm{MAF}_i(x, y) \; + \; \sum_{k=1}^{N_{VRU}} \mathrm{VRF}_k(x, y) \; + \; \mathrm{RPF}(x, y)\,, \qquad (x, y) \in \Omega, \quad (1)$$

where N_{mot} is the number of motorized agents and N_{VRU} is the number of vulnerable road users in the scene. We denote the field for vehicle i by $\mathrm{MAF}_i(x, y)$ and the field for pedestrian k by $\mathrm{VRF}_k(x, y)$. Each component is a sum over instances (dynamic agents) or static primitives. An overview of the full pipeline is shown in Fig. 1.

3.2 Motorized-agent fields (MAF)

For each motorized agent i (cars, motorcycles, trucks), we couple a black-box multimodal trajectory predictor with an explainable analytical field. Concretely, a learned predictor (e.g., QCNet [33] or any equivalent prediction network) outputs a set of M candidate trajectory centerlines $\{\gamma_i^{(m)}\}_{m=1}^{M}$ with probabilities $\{p_i^{(m)}\}_{m=1}^{M}$. We then transform these hypotheses into a continuous, planner-aligned risk density via a Gaussian-torus construction in Frenet coordinates, following the Enhanced Driving Risk Field (EDRF) principle [10] that risk equals **accident probability** times **accident consequence** [16]. Gaussian parameterizations are widely used in perception-related tasks for compact geometry encoding, which we adopt here for risk field design [34].

Let (s, d) denote longitudinal/lateral Frenet coordinates of a BEV location $(x, y) \in \Omega$ relative to $\gamma_i^{(m)}$. The MAF for hypothesis m, followed by driving-risk probability (DRP) [35], is a line integral of Gaussian cross-sections along the path:

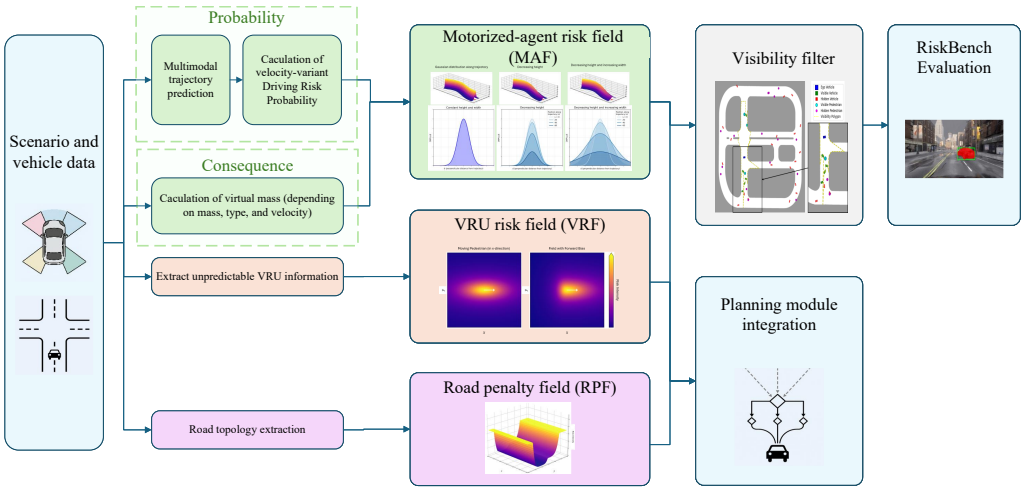

Figure 1: Overall pipeline of **MC-Risk**. Scene and vehicle data feed a multimodal trajectory predictor to compute a velocity-variant driving risk probability, which is combined with a virtual-mass consequence term to form the motorized-agent risk field (MAF). In parallel, road topology is extracted to build a topology-aware road penalty field (RPF), and VRU state/heading are used to construct a forward-biased anisotropic VRU field (VRF) since they are relatively unpredictable. The three fields are linearly composed into a BEV risk map, filtered by a visibility adapter, evaluated with RiskBench metrics, and optionally injected as a cost into an MPC planner.

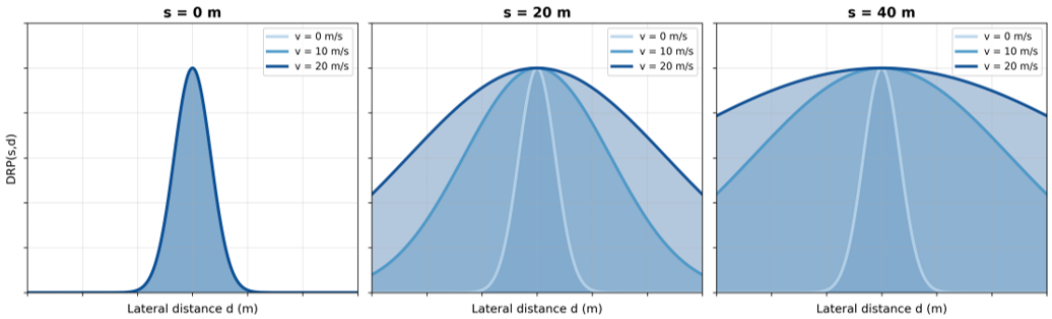

Figure 2: Velocity-variant width in MAF cross-sections. Lateral slices at three path distances (s=0, 20, 40 m) for speeds v=0, 10, 20 m/s. Higher speeds produce larger $\sigma(s)$ via the $k_v|v(s)|$ term in Eq. (4).

$$\mathrm{MAF}_i^{(m)}(x,y) = \mathrm{MAF}_i^{(m)}(s,d) = a(s) \cdot \exp\left(-\frac{d^2}{2\sigma(s)^2}\right) \tag{2}$$

The cross-section are shaped by letting the kernel height decay parabolically and the width grow linearly with path length and average curvature:

$$a(s) = q\,(s - s_{pt})^2, \qquad \sigma(s) = (b + k\,\overline{\kappa_{pt}})\,s + c, \tag{3}$$

where s_{pt} is the total predicted path length and $\overline{\kappa_{pt}}$ its average curvature. The parameter q controls how rapidly the height attenuates to zero at $s = s_{pt}$, while b, k, c control the growth and initial width of the cross-section, encoding greater dispersion for longer and curvier predictions.

Velocity-variant lateral dispersion. To anticipate wider reachable sets for fast or aggressive motion (visualized in Fig. 2), we augment the width function with a speed-sensitive term:

$$\sigma(s) = (b + k\,\overline{\kappa_{pt}})\,s \;+\; k_v\,|v(s)| \;+\; c, \tag{4}$$

where k_v scales the velocity-driven widening. This improves early spatial coverage for high-speed segments (longer braking distances) while keeping interpretability; in practice we bound $\sigma(s) \in [\sigma_{\min}, \sigma_{\max}]$ to avoid over-diffusion.

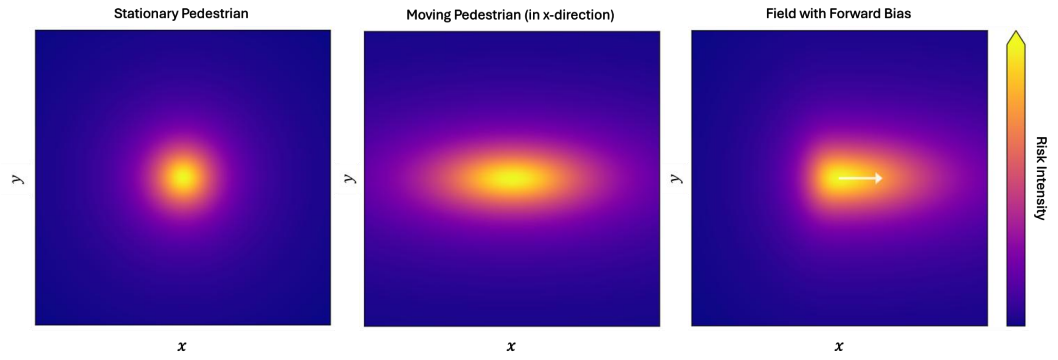

Figure 3: VRF visualization. Left: stationary pedestrian—near-isotropic footprint. Middle: moving pedestrian (baseline)—anisotropic ellipse aligned with heading. Right: forward-biased VRF—kernel shifted by μ_f along heading. Warmer colors indicate higher risk.

Consequence via virtual mass and MAF fusion. Similar to EDRF, we model accident consequence by a virtual-mass term

$$M_i \;=\; m_v\, T\left(\alpha\, v(s=0)^{\beta} + \gamma\right), \tag{5}$$

with tunable parameters α, β, γ, vehicle mass m_v, vehicle type T, and speed $v(s=0)$. In their formulation, virtual mass is calculated using the current velocity, which will result in a lack of foresight, failing to anticipate the severity of collisions caused by vehicles' potential acceleration maneuvers in the future. Instead we adopt a path-aggregated severity $\bar{M}_i^{(m)}$:

$$\bar{M}_i^{(m)} \;=\; \frac{1}{s_{pt}} \int_0^{s_{pt}} M_i(s)\, ds. \tag{6}$$

Finally, we sum over multimodal hypotheses to obtain the motorized-agent field:

$$\mathrm{MAF}_i(x,y) \;=\; \sum_{m=1}^{M} p_i^{(m)} \cdot \bar{M}_i^{(m)} \cdot \mathrm{MAF}_i^{(m)}(x,y), \tag{7}$$

which plugs directly into the scene-level composition of Eq. (1).

3.3 Vulnerable-road-user risk fields (VRF)

Pedestrians and other VRUs exhibit higher short-horizon unpredictability than motorized agents, so we adopt an anisotropic, speed-adaptive kernel (illustrations in Fig. 3) inspired by [4]. Let $\mathbf{p} = (x, y)$, $\mathbf{p_k}$ be the VRU position, $\hat{\mathbf{t}}_k$ its heading unit vector, and decompose the BEV offset as longitudinal offset $d_{\parallel} = (\mathbf{p} - \mathbf{p_k}) \cdot \hat{\mathbf{t}}_k$ and lateral offset $d_{\perp} = \left\| (\mathbf{p} - \mathbf{p_k}) - d_{\parallel} \hat{\mathbf{t}}_k \right\|$. The baseline elliptical field is

$$\mathrm{VRF}_k(x,y) = \frac{H}{\left(\dfrac{d_{\parallel}}{\gamma + k_{pl}\,|v_{\parallel}|}\right)^2 + \left(\dfrac{d_{\perp}}{\delta + k_{pw}\,|v_{\perp}|}\right)^2 + 1}, \tag{8}$$

where $v_{\parallel}, v_{\perp}$ are the longitudinal/lateral speed components, H scales overall strength, and $\gamma > \delta$ captures the forward/backward vs. sideways compactness.

Forward-bias To reflect higher risk ahead of a moving VRU, we shift the kernel's center by a speed-proportional parameter $\mu_f = \lambda_f\,|v_{\parallel}|$ along $\hat{\mathbf{t}}_k$:

$$\tilde{d}_{\parallel} = \left(\mathbf{p} - \mathbf{p_k} - \mu_f \hat{\mathbf{t}}_k\right) \cdot \hat{\mathbf{t}}_k, \quad \tilde{d}_{\perp} = \left\| \left(\mathbf{p} - \mathbf{p_k} - \mu_f \hat{\mathbf{t}}_k\right) - \tilde{d}_{\parallel} \hat{\mathbf{t}}_k \right\|, \tag{9}$$

This forward-biased formulation increases the ability to detect risk from approaching VRUs by assuming that most VRUs will maintain their current state of motion, and it integrates into Eq. (1) alongside MAF_i and RPF. A scene overlay including both MAF and VRF is shown in Fig. 4.

3.4 Road penalty field (RPF)

Following Liu et al., we encode a topology-aware road penalty independent of dynamic agents; the map partition and resulting field are shown in Fig. 5. Unlike their formulation, which only poses risk based on road boundaries, we use full HD-map topology to include (i) off-road penalties, (ii) same-direction adjacent lanes, (iii) opposite-direction lanes. Supposing the road is in x direction, we let $d_{\pm}(y)$ be the signed distance to road boundaries (negative inside). $\mathcal{L}$ is the set of lane centerlines with direction attribute, and we partition $\mathcal{L}$ around the ego's current lane into $\mathcal{L}_{\text{same}}$ and $\mathcal{L}_{\text{opp}}$. Using a single compact expression, the RPF is

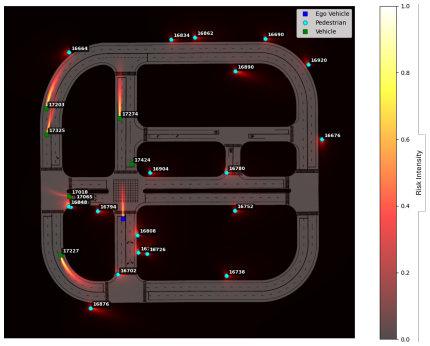

Figure 4: Scene overlay of MAF (vehicles) and VRF (VRUs) with respective ID labeled.

$$RPF(y) \;=\; \lambda_{\text{off}}\,\mathbf{1}[d_{\pm}(y) > 0] \;+\!\sum_{\ell\in\mathcal{L}_{\text{same}}}\!\lambda_{\text{same}}\exp\!\Big(-\tfrac{\text{dist}(y,\ell)^2}{2\sigma_{\text{same}}^2}\Big) \;+\!\sum_{\ell\in\mathcal{L}_{\text{opp}}}\!\lambda_{\text{opp}}\exp\!\Big(-\tfrac{\text{dist}(y,\ell)^2}{2\sigma_{\text{opp}}^2}\Big) \quad (10)$$

where $\mathbf{1}[\cdot]$ is the indicator function and $\text{dist}(y,\ell)$ is the lateral distance to lane centerline ℓ. Hyperparameters $(\lambda_{\text{off}}, \lambda_{\text{same}}, \lambda_{\text{opp}}, \sigma_{\text{same}}, \sigma_{\text{opp}})$ control baseline risk value and spread, typically $\lambda_{\text{off}} >> \lambda_{\text{opp}} > \lambda_{\text{same}}$. The off-road term assigns a large, fixed penalty outside $\mathcal{M}_{\text{road}}$, discouraging driving beyond the drivable boundary. The implementation is also compatible in intersections as shown in Fig. 5, and we finally add RPF to the scene risk field in Eq. (1).

4 Experiments

4.1 Experimental settings

Dataset We evaluate on RiskBench, a CARLA-based benchmark by Kung et al. [11] that couples rich static infrastructure with dynamic agents across diverse scenarios: cut-in, braking, merging, yielding 6916 scenarios with spatiotemporal risk annotations. Despite active research on risk fields and threat assessment, few prior works report standardized, visibility-aware evaluations. Results are often qualitative and omit actor-set alignment. A key contribution of this paper is to adopt RiskBench for per-actor risk quantification, and evaluate the final results with an extended metric suite from RiskBench (**OT-F1**, **OT-F1-T** at 1/2/3 s, **wMOTA**, **PIC**). Because risk is most prominent when hazardous events culminate in contact, we restrict evaluation to the collision subset of RiskBench.

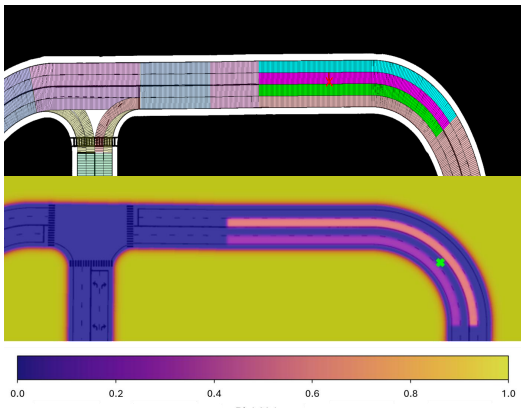

Figure 5: Topology-aware road partition and RPF. Top: CARLA HD-map topology around ego, including the intersection, partitioned into same-/opposite-direction lanes. Bottom: resulting Road Penalty Field (RPF) centered at the ego.

Evaluation metrics Following RiskBench [11] and its adaptations [36], we use actor-aligned metrics for spatial accuracy and temporal consistency. At each frame t, we compute an actor-level risk score $r_{i,t} = \max_{(x,y)\in\Omega} R_{scene}(x,y)$ as the maximum risk over the BEV grid Ω, and an actor is labeled risky if $r_{i,t} > \tau$. Sweeping τ yields precision $= TP/(TP{+}FP)$, recall $= TP/(TP{+}FN)$, and $F1 = \frac{2\cdot\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}}$. We report **OT-F1** $= \max_\tau F1(\tau)$ and **OT-F1-T** is computed over the last T seconds before the ego's closest approach to maximum risk to evaluate temporal consistency. The proactivity of risk anticipation is summarized by $PIC = -\sum_{t=1}^{T}\exp\!\big(-\frac{T-t}{T}\big)\cdot\log\big(F_{1,t}\big)$, penalizing a false identification more when ego vehicle is closer to the risk than far from the risk. Robust tracking quality is assessed

via **wMOTA** $= 1 - \frac{\sum_t \left[w_p(FN_t + IDsw_t^p) + w_n(FP_t + IDsw_t^n)\right]}{w_p\, GT_t^p + w_n\, GT_t^n}$, which addresses the imbalance between risky and non-risky samples.

Visibility adapter To enable the occlusion-aware and visibility-scoped risk identification, we apply a lightweight line-of-sight adapter used only at evaluation time (Fig. 6). We create a road mask from HD map data and take non-drivable areas as opaque. Motorized agents are added as dynamic occluders (pedestrians are non-occluding in our setup). From the ego position, we cast a certain number of rays uniformly over 2π angle and stop each ray at the first opaque pixel or a fixed max range. The endpoints define a convex visibility polygon. At each time step, we filter both the scene risk map and the actor set by this polygon: actors outside are ignored for metric computation, thus preventing access to off-screen actors, decreasing the false positive rate, and increasing computational efficiency. This adapter aligns our evaluation with a risk model's visibility property while leaving the underlying risk model unchanged.

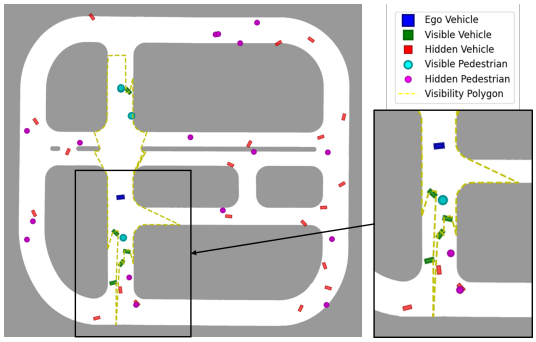

Figure 6: Visibility filtering utilizing the visibility polygon.

4.2 Quantitative results

Table 1: Comparison of Risk Object Identification (ROI) metrics between **MC-Risk** and RiskBench baseline models. Baselines are ordered by **OT-F1** score. The best performer of each metric is marked in bold, and the second best performer is marked in italic.

Model	**OT-F1** (↑)	**wMOTA** (↑)	**PIC** (↓)	**OT-F1-1s** (↑)	**OT-F1-2s** (↑)	**OT-F1-3s** (↑)
Range	*65.67%*	**77.09%**	30.15	*81.97%*	71.79%	*67.07%*
Kalman filter	58.98%	72.33%	*28.49*	80.33%	*69.39%*	64.25%
DSA	58.19%	66.85%	39.10	79.24%	62.78%	55.31%
Social-GAN	51.97%	68.25%	46.36	64.29%	53.98%	51.55%
QCNet	51.82%	68.35%	48.05	65.44%	56.09%	52.26%
MANTRA	51.07%	67.75%	46.26	63.50%	53.36%	50.39%
BCP	32.56%	57.80%	64.54	38.70%	29.93%	28.34%
Random	27.75%	1.80%	48.40	33.55%	31.53%	30.65%
BP	17.33%	53.47%	73.10	15.69%	13.66%	14.88%
MC-Risk	**68.62%**	*74.03%*	**14.78**	**90.80%**	**79.23%**	**73.16%**

We adopt nine reference methods integrated in RiskBench [11] as baselines, covering four families: (i) **Rule-based**: *Random* and *Range* (fixed-distance filter) risk object identification provides the simplest baselines. (ii) **Trajectory prediction and collision checking**: *Kalman* [37], *Social-GAN* [24], *MANTRA* [25], and *QCNet* [33] first predict agent motion and then simply declare risk via trajectory overlap with ego. (iii) **Collision anticipation**: *DSA* [22] learns to score collision likelihood from spatiotemporal features without explicit overlap tests. (iv) **Behavior prediction-based**: *BP* [26] and *BCP* [27] infer risk from predicted influence on the ego's intended behavior.

We use the RiskBench implementations and default settings, apply our visibility adapter, restrict to the collision subset, and evaluate with the unified metric suite described above.

Table 1 shows that **MC-Risk** delivers the strongest overall risk localization and the earliest hazard indication on the collision subset. It achieves the best **OT-F1** (68.62%), surpassing the strongest baseline (Range, 65.67%) by +2.95 points, while obtaining the lowest **PIC** (14.78), a ∼48% reduction versus Kalman filter (28.49), indicating early and reliable risk identification for the ego vehicle to react. Consistency metrics are also the best: **OT-F1-1s/2s/3s** reach 90.80/79.23/73.16%, improving upon Range by +8.83/+7.44/+6.09 points, respectively, showing the stability of the model. The only metric where MC-Risk is not the top performer is **wMOTA**: it ranks second (74.03%) and is within 3.06 points of Range (77.09%), suggesting slightly less temporal persistence in tracking

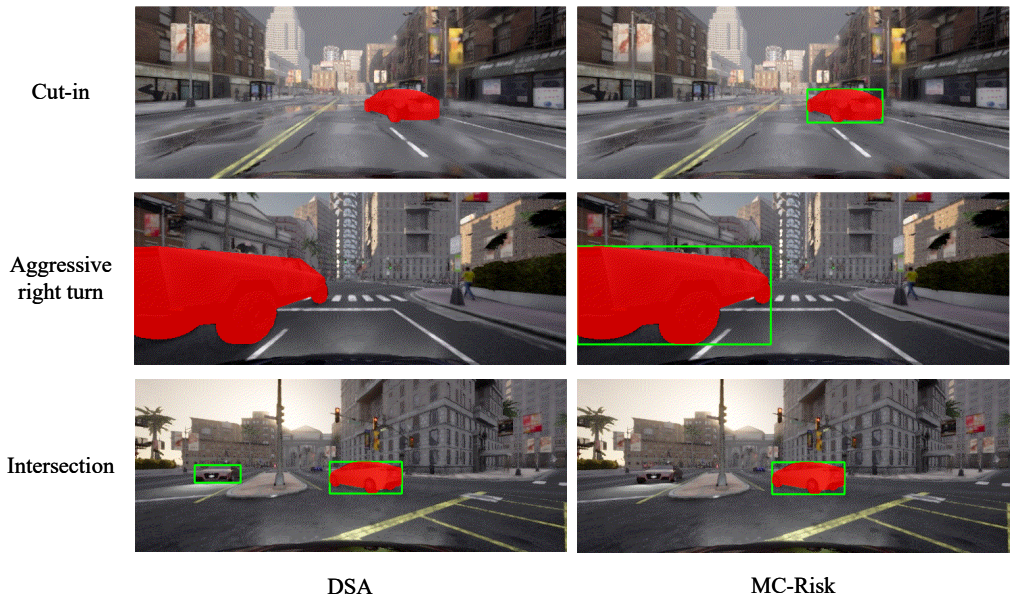

Figure 7: Qualitative comparison of risk identification on three representative scenarios. Red overlays denote the ground-truth risky actor from RiskBench, and green boxes mark the predicted risky object at a fixed threshold.

robustness but a better precision–recall balance overall. In summary, treating risk identification as a simple downstream of the computed risk field (actorwise max + threshold) already achieves strong performance, validating the MC-Risk field formulation for this subtask.

4.3 Qualitative results

For qualitative results, we take the DSA model as the baseline. As shown in Figure 7, the red layers identify the ground-truth risky object from RiskBench, and the green bounding boxes indicate risky objects derived from our MC-Risk with a tuned threshold. We compare DSA and MC-Risk on three challenging situations: close cut-in, intersection, and aggressive right turn.

Close cut-in: A neighboring vehicle executes a late merge into the ego lane with small headway and a tight lateral gap. DSA does not identify risk until the intruder is already merging, whereas MC-Risk places early risk along the intruder's predicted centerline, with velocity-conditioned widening that projects risk exposure into the ego lane ahead of the merge.

Aggressive right turn: A nearby vehicle starts a sharp right turn across the ego lane at high entry speed and high curvature. DSA again lags and still cannot capture the risk in the displayed time step. MC-Risk expands laterally on high-curvature segments, highlighting the future risk overlap region before the turn is executed.

Intersection: An oncoming vehicle approaches the intersection, turns right, and decelerates at the same time. The ego vehicle is turning left and trying to enter the same target lane as the other vehicle. Both models can detect the risk, but DSA produces a false positive on the stopping vehicle on the opposite lane; MC-Risk avoid this via reduced risk spread from a low-speed (stopped) agent, yielding a cleaner focus on truly threatening actors.

These results align with the quantitative gains in early-warning metrics, showing that MC-Risk provides earlier and cleaner localization and avoids false alarms on negative agents.

4.4 Ablation Studies

Table 2 quantifies each component's contribution:

MAF velocity-variant width. Removing the speed-conditioned lateral widening lowers **OT-F1** from 68.62% to 65.09 (-3.53) and **wMOTA** from 74.03% to 72.09 (-1.94), while **PIC** rises from 14.78 to

Table 2: Component ablations on **MC-Risk**. A checkmark indicates the component is enabled. Metrics are still computed on the collision subset with visibility-scoped, actor-aligned evaluation. The best performer of each metric is marked in bold.

Model	MAF vel-var	VRF	RPF	OT-F1 (↑)	wMOTA (↑)	PIC (↓)	OT-F1-1s (↑)	OT-F1-2s (↑)	OT-F1-3s (↑)
MC-Risk (full)	✓	✓	✓	**68.62%**	**74.03%**	**14.78**	**90.80%**	**79.23%**	**73.16%**
w/o vel-variant width (MAF)	×	✓	✓	65.09%	72.09%	20.40	84.73%	76.05%	69.82%
w/o VRU risk field (VRF)	✓	×	✓	61.42%	70.81%	22.87	71.23%	65.96%	62.02%
w/o road penalty field (RPF)	✓	✓	×	68.62%	74.01%	19.66	90.80%	79.23%	73.16%

20.40 (+5.62). Lead-time scores also drop, indicating that velocity-variant width is key for predicting risk early along fast segments and reducing the weight of low-speed objects.

VRU modeling. In the VRF ablation, VRUs are also using an MAF to indicate their risks instead of the forward-biased anisotropic VRF. This causes the largest degradation: **OT-F1** 68.62% → 61.42 (-7.20), **wMOTA** 74.03% → 70.81 (-3.22), and **PIC** 14.78 → 22.87 (+8.09). Early-warning quality also drops sharply, showing that a forward-biased VRU field is crucial for timely, precise pedestrian/bicycle risk.

Road penalty field (RPF). Removing the static/topology-aware road layer leaves **OT-F1** (68.62%) and **wMOTA** (74.01%) essentially unchanged, and the lead-time scores are identical to full. However, **PIC** worsens (14.78 → 19.66; +4.88). A plausible explanation is that the RPF layer adds early risk where edge proximity is inherent, which advances correct identifications in time without changing which actor is named or how stable identities are tracked. The effect targets timing more than classification or tracking, hence the integration of RPF shows a clear improvement in the PIC metric.

4.5 Planning integration

In this section, we demonstrate an example to show how the MC-Risk can serve as a planner cost, leaving closed-loop evaluations to future work. Let $\mathcal{R}_t(x, y) = R_{\text{scene}}(x, y)$ denote the BEV risk density at time t and let $\xi = \{x_t\}_{t=0}^{H}$ be the ego state trajectory to be planned over horizon H with controls $u = \{u_t\}_{t=0}^{H-1}$. We formulate a standard Model Predictive Control (MPC) [38] problem with an additional risk-guidance term:

$$\begin{aligned} \min_{u,\xi} \quad & J(\xi, u) = \alpha \sum_{t=0}^{H} \hat{\mathcal{R}}_t(x_t) + \beta \sum_{t=0}^{H-1} \|u_t\|_R^2 + \gamma \sum_{t=1}^{H-1} \|u_t - u_{t-1}\|_S^2 \\ \text{s.t.} \quad & x_{t+1} = f(x_t, u_t), \\ & x_t \in \mathcal{X}, \\ & u_t \in \mathcal{U}. \end{aligned} \tag{11}$$

Here f is a kinematic bicycle model with standard bounds $\mathcal{X}, \mathcal{U}$. R, S are positive-definite weights, and $\hat{\mathcal{R}}_t(x_t)$ samples the risk field around the vehicle footprint (e.g., max- or area-average over a swept polygon). We use (α, β, γ) to trade off risk avoidance, control effort, and comfort, so no additional training is required.

5 Conclusion

This work introduced **MC-Risk**, a modular BEV risk field that couples learned multimodal forecasting with explainable analytic components to provide early, calibrated, and planner-ready risk maps. The field comprises a velocity/curvature-aware motorized-agent kernel, a forward-biased anisotropic VRU model, and a topology-informed road penalty term. Together, these yield a representation suitable for direct consumption by downstream planners. We performed a standardized, visibility-consistent evaluation on the RiskBench collision subset and achieved the strongest overall risk localization ability with the highest stability.

Limitations Closed-loop planning results using MC-Risk are not yet reported. Future directions include (i) closed-loop planning evaluations with on-policy setup in dense traffic; (ii) transfer to real-world datasets and hardware-in-the-loop tests.

References

[1] D. Nistér, H.-L. Lee, J. Ng, and Y. Wang, "The safety force field," *NVIDIA White Paper*, vol. 15, 2019.

[2] J. Wang, J. Wu, X. Zheng, D. Ni, and K. Li, "Driving safety field theory modeling and its application in pre-collision warning system," *Transportation research part C: emerging technologies*, vol. 72, pp. 306–324, 2016.

[3] J. Wang, J. Wu, and Y. Li, "The driving safety field based on driver–vehicle–road interactions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2203–2214, 2015.

[4] P. Liu, Y. Chang, J. Gao, G. Du, Z. Su, M. Liu, and W. Liu, "Research on local path planning of unmanned vehicles based on improved driving risk field," *Scientific Reports*, vol. 14, no. 1, p. 29153, 2024.

[5] Q. Wang, D. Xu, G. Kuang, C. Lv, S. E. Li, and B. Nie, "Risk-aware vehicle trajectory prediction under safety-critical scenarios," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[6] J. Li, X. Wang, and T. Zhang, "A spatiotemporal learning approach to safety-oriented individualized driving risk assessment in a vehicle-to-everything (v2x) environment," *IET Intelligent Transport Systems*, vol. 18, no. 12, pp. 2459–2484, 2024.

[7] L. Shi, C. Qian, and F. Guo, "Real-time driving risk assessment using deep learning with xgboost," *Accident Analysis Prevention*, vol. 178, p. 106836, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457522002718

[8] H. Hu, Q. Wang, M. Cheng, and Z. Gao, "Cost-sensitive semi-supervised deep learning to assess driving risk by application of naturalistic vehicle trajectories," *Expert Systems with Applications*, vol. 178, p. 115041, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421004826

[9] H. Cao, Z. Qu, G. Chen, X. Li, L. Thiele, and A. Knoll, "Ghostvit: Expediting vision transformers via cheap operations," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 6, pp. 2517–2525, 2024.

[10] J. Jiang, Z. Han, Y. Wang, M. Cai, Q. Meng, Q. Xu, and J. Wang, "Edrf: Enhanced driving risk field based on multimodal trajectory prediction and its applications," in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2024, pp. 2287–2293.

[11] C.-H. Kung, C.-C. Yang, P.-Y. Pao, S.-W. Lu, P.-L. Chen, H.-C. Lu, and Y.-T. Chen, "Riskbench: A scenario-based benchmark for risk identification," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 14 800–14 807.

[12] J. Jansson, *Collision avoidance theory with application to automotive collision mitigation*. Linkopings Universitet (Sweden), 2005.

[13] J. C. Hayward, "Near miss determination through use of a scale of danger," 1972.

[14] M. M. Minderhoud and P. H. Bovy, "Extended time-to-collision measures for road traffic safety assessment," *Accident Analysis & Prevention*, vol. 33, no. 1, pp. 89–97, 2001.

[15] B. L. Allen, B. T. Shin, and P. J. Cooper, "Analysis of traffic conflicts and collisions," Tech. Rep., 1978.

[16] T. Puphal, M. Probst, and J. Eggert, "Probabilistic uncertainty-aware risk spot detector for naturalistic driving," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 3, pp. 406–415, 2019.

[17] C. Wang, Y. Xie, H. Huang, and P. Liu, "A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling," *Accident Analysis & Prevention*, vol. 157, p. 106157, 2021.

[18] P. Guo, Q. Zhu, and X. Wu, "Responsibility-sensitive collision risk assessment and maneuvering safety evaluation," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 99–104.

[19] Z. Wei, H. Zhou, and R. Zhou, "Risk and complexity assessment of autonomous vehicle testing scenarios," *Applied Sciences*, vol. 14, no. 21, 2024.

[20] H. Suk, T. Kim, H. Park, P. Yadav, J. Lee, and S. Kim, "Rationale-aware autonomous driving policy utilizing safety force field implemented on carla simulator," *arXiv preprint arXiv:2211.10237*, 2022.

[21] Z. Zhou, J. Wang, Y.-H. Li, and Y.-K. Huang, "Query-centric trajectory prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 17 863–17 873.

[22] F.-H. Chan, Y.-T. Chen, Y. Xiang, and M. Sun, "Anticipating accidents in dashcam videos," in *Asian conference on computer vision*. Springer, 2016, pp. 136–153.

[23] K.-H. Zeng, S.-H. Chou, F.-H. Chan, J. Carlos Niebles, and M. Sun, "Agent-centric risk assessment: Accident anticipation and risky region localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2222–2230.

[24] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2255–2264.

[25] F. Marchetti, F. Becattini, L. Seidenari, and A. D. Bimbo, "Mantra: Memory augmented networks for multiple trajectory prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 7143–7152.

[26] C. Li, Y. Meng, S. H. Chan, and Y.-T. Chen, "Learning 3d-aware egocentric spatial-temporal interaction via graph convolutional networks," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 8418–8424.

[27] C. Li, S. H. Chan, and Y.-T. Chen, "Who make drivers stop? towards driver-centric risk assessment: Risk object identification via causal inference," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 711–10 718.

[28] S. You, X. Luo, X. Liang, J. Yu, C. Zheng, and J. Gong, "A comprehensive llm-powered framework for driving intelligence evaluation," *arXiv preprint arXiv:2503.05164*, 2025.

[29] Z. Zhou, H. Huang, B. Li, S. Zhao, Y. Mu, and J. Wang, "Safedrive: Knowledge-and data-driven risk-sensitive decision-making for autonomous vehicles with large language models," *arXiv preprint arXiv:2412.13238*, 2024.

[30] H. Cao, G. Chen, J. Xia, G. Zhuang, and A. Knoll, "Fusion-based feature attention gate component for vehicle detection based on event camera," *IEEE Sensors Journal*, vol. 21, no. 21, pp. 24 540–24 548, 2021.

[31] H. Cao, Z. Zhang, Y. Xia, X. Li, J. Xia, G. Chen, and A. Knoll, "Embracing events and frames with hierarchical feature refinement network for object detection," in *European Conference on Computer Vision*. Springer, 2024, pp. 161–177.

[32] T. Puphal, M. Probst, and J. Eggert, "Probabilistic uncertainty-aware risk spot detector for naturalistic driving," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 3, p. 406–415, Sep. 2019. [Online]. Available: http://dx.doi.org/10.1109/TIV.2019.2919465

[33] Z. Zhou, J. Wang, Y.-H. Li, and Y.-K. Huang, "Query-centric trajectory prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 17 863–17 873.

[34] H. Cao, G. Chen, Z. Li, Q. Feng, J. Lin, and A. Knoll, "Efficient grasp detection network with gaussian-based grasp representation for robotic manipulation," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 3, pp. 1384–1394, 2023.

[35] S. Kolekar, J. De Winter, and D. Abbink, "Human-like driving behaviour emerges from a risk-based driver model," *Nature communications*, vol. 11, no. 1, p. 4850, 2020.

[36] P.-Y. Pao, S.-W. Lu, Z.-Y. Lu, and Y.-T. Chen, "Potential field as scene affordance for behavior change-based visual risk object identification," *arXiv preprint arXiv:2409.15846*, 2024.

[37] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

[38] B. Kouvaritakis and M. Cannon, "Model predictive control," *Switzerland: Springer International Publishing*, vol. 38, no. 13-56, p. 7, 2016.

A Experiments compute resources

The experiments are conducted on a laptop with an Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz with 13GB of system memory, which is used both for testing and debugging. No GPU resource is required for the implementation.

B Societal Impacts

MC-Risk can advance road safety by delivering earlier, cleaner risk cues without sacrificing planner compatibility, especially for pedestrians and cyclists. Its forward-biased VRU modeling and topology-aware road layer help vehicles anticipate conflicts at crosswalks, merges, and opposing lanes, reducing high-severity collisions. Because the method is modular and light-weight, it is practical for embedded systems and accessible to most operating environments. Finally, a benchmarked risk map approach enables reproducible research and further comparison in this area.

C Evaluation metrics in detail

We evaluate along two complementary axes: spatial accuracy and temporal consistency, following RiskBench [11] with adaptations from Pao et al. [36].

Detection Accuracy. Let TP, FP, and FN denote true positives, false positives, and false negatives identifying risk objects, respectively. Precision and recall are

$$\text{precision} = \frac{TP}{TP + FP}, \qquad \text{recall} = \frac{TP}{TP + FN}. \tag{12}$$

The F1-score, the harmonic mean of precision and recall, is

$$F1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{13}$$

Overall Thresholded F1 (OT-F1). We sweep a decision threshold τ over all unique predicted risk scores and compute $F1(\tau)$. The overall thresholded score is

$$\text{OT-F1} = \max_{\tau} \; F1(\tau). \tag{14}$$

PIC. The PIC metric [11, 36] penalizes late errors more strongly. Using per-frame $F1_t$ (at the global threshold that maximizes OT-F1), we define

$$\text{PIC} = -\sum_{t=1}^{T} \exp\left(-\frac{T-t}{T}\right) \cdot \log\big(F1_t\big). \tag{15}$$

Lower values indicate earlier and more consistent correct identifications.

Weighted Multi-Object Tracking Accuracy (wMOTA). To measure temporal stability with class imbalance, we adopt wMOTA from [36]. For each frame t,

$$PM_t = w_p\big(FN_t + IDsw_t^p\big), \qquad NM_t = w_n\big(FP_t + IDsw_t^n\big), \tag{16}$$

where $IDsw_t^p$ and $IDsw_t^n$ count identity switches among ground-truth risky and non-risky actors, respectively, and w_p, w_n are class weights. With GT_t^p and GT_t^n the ground-truth counts,

$$\text{wMOTA} = 1 - \frac{\sum_t \big(PM_t + NM_t\big)}{\sum_t \big(w_p\, GT_t^p + w_n\, GT_t^n\big)}. \tag{17}$$

Higher wMOTA reflects fewer misses and switches, accounting for the prevalence of non-risky actors.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made accurately reflect the paper's contribution and scope.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitation is discussed in the last section of the paper.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: All theoretical results contain full formula derivation and verification.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Formulas are all provided, input and output formats are completely defined.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Code will be published upon acceptance.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: Data splits used and experiment setting in RiskBench is introduced in the experiment section.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: This study does not report error bars.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Computational resources is provided in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The study conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The societal impacts are included in the appendix.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Those assets are all cited or mentioned properly.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: No new assets are released.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: It does not involve human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.