Research Article

Beyond imitation: A generate-and-evaluate framework with spatiotemporal risk assessment for end-to-end motion planning

Ying Liu†, Yi Liu†, Xuewu Lin, Chaoyi Chen, Fang Zhang*, and Jianqiang Wang

Abstract End-to-end motion planning for autonomous vehicles is a critical research frontier, yet prevailing imitation learning approaches are fundamentally constrained by their reliance on regressing to a single, often sub-optimal, ground-truth trajectory. This paradigm fails to address the inherent multi-modality of driving scenarios and the interactive nature of traffic, limiting the system's ability to discover safer or more efficient behaviors. To address these limitations, we propose a novel "generate-and-evaluate" framework for end-to-end motion planning that moves beyond direct imitation. The framework uses the sparse perception and multi-modal candidate-generation foundation of SparseDrive. Its ego-centric prediction module incorporates observed ego-vehicle, surrounding-agent, map, and temporal context. Subsequently, a principled trajectory selection methodology evaluates these candidates by constructing a predictive spatiotemporal risk field and utilizing a holistic cost function that synthesizes safety, efficiency, and comfort. A core innovation is a time-weighted safety metric that enables a more nuanced and proactive risk assessment. Experimental results show lower collision rate and longitudinal acceleration RMS on mini_val than SparseDrive, with nearly unchanged average speed. On val, higher average speed and lower acceleration RMS come with a higher collision rate, indicating a trade-off among the evaluated criteria.

Keywords End-to-end motion planning, Autonomous driving, Multimodal predicted trajectory generation, Trajectory evaluation, Spatiotemporal risk assessment

• Ying Liu is with the College of Engineering, Nanjing Agricultural University, Nanjing 211800, China.
• Yi Liu is with the School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China.
• Xuewu Lin is with Horizon Information Technology Company Ltd., Beijing 100094, China.
• Yi Liu, Chaoyi Chen, and Jianqiang Wang are with the School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China.
• Fang Zhang is with the State Key Laboratory of Intelligent Green Vehicle and Mobility, Tsinghua University, Beijing 100084, China. E-mail: zhang_fang@tsinghua.edu.cn.
† These authors contributed equally to this work.
* To whom correspondence should be addressed.

1 Introduction

Autonomous driving technology has witnessed exponential growth in recent years, holding the promise to revolutionize mobility by significantly improving road safety, enhancing traffic efficiency, and liberating human drivers from the cognitive load of navigation[1,2]. At the heart of this technological frontier is the motion planning and decision-making system, which is tasked with charting a vehicle's course through highly dynamic and unpredictable environments[3,4]. The development of a robust and reliable motion planning stack is paramount to the widespread adoption of fully autonomous vehicles[5].

Historically, the autonomous system architecture for motion planning has been dominated by two competing paradigms: traditional rule-based methods[6–8] and modern end-to-end (E2E) learning-based approaches[9–11]. Rule-based systems employ a modular, cascaded pipeline, typically comprising distinct modules for perception, prediction, planning, and control. At the trajectory planning level, Huang et al. proposed an optimal trajectory planning architecture for autonomous vehicles based on the principle of least action, leveraging the seeking benefits and avoiding harm characteristics of human drivers[12]. Yan et al. proposed a game theory-based trajectory planning framework that can comprehensively improve safety, comfort, and efficiency in complex mixed-traffic environments[13]. While this modularity offers interpretability and allows for the explicit encoding of traffic laws and safety heuristics, it is also a source of significant fragility. Errors can accumulate and propagate through the pipeline, and the hand-crafted rules often struggle to cover the long tail of edge cases encountered in real-world driving, leading to brittle and sometimes unnatural driving behaviors[11]. A comprehensive review of these classical motion planning techniques, including search-based and sampling-based methods, can be found in[14], which highlights their strengths in structured environments but also their limitations in highly dynamic traffic. In contrast, end-to-end approaches seek to overcome these limitations by learning a direct mapping from raw sensor inputs to driving commands. Models like UniAD[10] and VAD[11] have unified perception, prediction, and planning into a sin-

Just Accepted

gle, goal-oriented framework, demonstrating state-of-the-art performance. By training on vast datasets, E2E models can learn complex, nuanced behaviors that are difficult to codify by hand, making them a highly active and promising area of research [15].

Within the end-to-end paradigm, two primary learning strategies have emerged: Reinforcement Learning (RL) [9,16] and Imitation Learning (IL) [10,11]. RL aims to learn an optimal driving policy by interacting with an environment (typically a simulator) and maximizing a cumulative reward signal. For example, Tian et al. proposed a path tracking framework that mimics human learning by initializing a Deep Deterministic Policy Gradient (DDPG) model via behavior cloning [17]. Mao et al. established a Transformer-based unsupervised hierarchical reinforcement learning framework for transforming consecutive multi-frame states into semantically meaningful driving skills [18]. However, its practical application in autonomous driving is hindered by several challenges, including the difficulty of designing a reward function that comprehensively captures all aspects of safe and effective driving, and the significant sim-to-real gap that makes transferring learned policies to the physical world risky [19]. IL, on the other hand, offers a more direct and data-efficient path to deployment. By leveraging large-scale datasets of expert human driving, IL learns to mimic demonstrated behaviors. Pan et al. presented an end-to-end imitation learning system for agile off-road driving that uses low-cost sensors to imitate a model predictive controller, empirically demonstrating that online learning strategies overcome covariate shift to achieve state-of-the-art performance [20]. Wang et al. introduced Social-MAIL, a multiagent imitation learning model that explicitly captures driving sociality by incorporating social value orientation into the reward function and employing a temporal-spatial attention policy network to generate realistic interaction scenarios for autonomous vehicle evaluation [21]. This grounding in real-world data makes IL a more pragmatic and widely adopted choice for developing deployable autonomous driving systems [22].

Nevertheless, conventional IL frameworks are not without their own fundamental limitations. The predominant approach, often termed behavioral cloning, trains a model to regress towards a single ground-truth trajectory recorded from a human driver [10]. This formulation introduces two critical problems. First, it often overlooks the closed-loop, interactive nature of traffic. The model learns to predict the future paths of other agents and the ego-vehicle in an open-loop fashion, failing to account for how the ego-vehicle's own actions will influence the behavior of others. As noted in recent literature [23], this disregard for reciprocal influence can lead to planning failures in dense, interactive scenarios. Second, this approach is constrained by the inherent nature of the ground-truth data. Human driving is fundamentally multi-modal—multiple valid trajectories exist for any given situation—and is not always optimal [24]. By forcing the model to replicate a single, potentially sub-optimal human trajectory, we limit its ability to discover safer, more efficient, or more comfortable solutions. This problem of handling diverse, plausible futures and moving beyond the limitations of a single expert demonstration has been a critical challenge that urgently requires resolution in future research.

To address these deficiencies, this paper proposes a novel planning framework that departs from direct trajectory regression. Instead of learning to produce a single output, our approach begins by thoroughly understanding the interactions between ego-vehicle and others, thereby generating a diverse set of candidate trajectories that embrace the multimodal nature of driving. The learned components retain supervised task losses; a subsequent risk-based evaluation scores and selects an ego trajectory from the candidate set. The core of our contribution is a comprehensive cost function that synthesizes the three pillars of superior driving: safety, efficiency, and comfort. In particular, we introduce a sophisticated, time-weighted safety metric based on potential fields, which enables a more nuanced risk assessment by prioritizing imminent threats over distant ones. This "generate-and-evaluate" paradigm allows our system to explicitly reason about multiple futures and select a trajectory with the lowest chosen composite cost among the generated candidates. The main contributions of this research are as follows:

(1) A "generate-and-evaluate" framework built on the sparse perception and multimodal candidate-generation foundation of SparseDrive, with a subsequent risk-based trajectory-evaluation stage.
(2) Use of the inherited ego-centric scene-context integration for multimodal prediction, exchanging information among observed ego-vehicle, surrounding-agent, map, and temporal representations.
(3) A principled trajectory selection methodology that employs a predictive spatiotemporal risk field and a novel time-weighted safety cost, enabling a balanced and proactive assessment of safety, efficiency, and comfort.

The remainder of this paper is organized as follows. Section 2 presents the detailed formulation of our proposed methodology. Section 3 showcases experimental results and provides a comparative analysis. Finally, Section 4 concludes the paper and outlines directions for future work.

Just Accepted

2 Methodology

This chapter elaborates on our proposed generate-and-evaluate framework for end-to-end autonomous driving, which is designed to generate safe and comfortable ego-vehicle trajectories directly from multi-view camera inputs through learned perception and trajectory generation followed by discrete selection. As illustrated in Fig. 1, our framework is composed of a cascade of three core components: 1) Sparse Scene Perception, 2) Multi-modal Trajectory Prediction, and 3) Spatiotemporal Joint Trajectory Selection. The structure of this chapter follows this pipeline, first introducing the rationale behind the overall architecture and the perception module, followed by an in-depth exploration of the multimodal prediction foundation and our focus on spatiotemporal risk-based trajectory evaluation.

2.1 Overall Architecture and Sparse Perception Foundation

The end-to-end paradigm we propose aims to achieve holistic optimization from raw sensor data to final driving decisions. The learned perception and trajectory-generation components are differentiable, whereas the final argmax or argmin selection index is discrete and does not propagate gradients. Our system takes a sequence of multi-view camera images as input and, through a unified network, outputs the ego-vehicle's trajectory for the upcoming seconds.

The cornerstone of our framework is the sparse scene perception module, whose primary responsibility is to transform high-dimensional, raw pixel information into a structured, sparse, and temporally consistent understanding of the driving scene. To provide a high-quality, efficient, and low-latency scene representation for the downstream decision-making and planning stages, we made a deliberate decision to adopt and integrate the advanced sparse perception architecture presented in SparseDrive[23]. As shown in Fig. 1, the ego-vehicle uses 2D images from six views as input. After processing through the backbone network ResNet101, the data enters the sparse scene generator. The entire driving scene is represented as two instance sets: one comprising current and historical instances of dynamic objects, and the other comprising current and historical instances of the static map. Each instance comprises two components: an instance feature, which is a high-dimensional vector encoding semantic information, appearance, category, etc.; and a geometric anchor box/polyline describing the instance's position, size, shape, and pose in 3D space. The model iteratively optimizes this instance set through a multi-layer decoder. At each layer, both instance features and geometric anchors are updated to progressively refine their representation of real-world objects. This technical choice is predicated on several key advantages:

(1) Exceptional Efficiency and Performance: The "sparse-centric" paradigm advocated by SparseDrive abandons the reliance on computationally expensive dense BEV features common in traditional methods. By directly operating on sparse instance queries, it achieves a unified modeling of dynamic agents and static map elements. This approach not only significantly enhances training and inference efficiency but also achieves state-of-the-art perception accuracy (e.g., on mAP and AMOTA metrics), ensuring our downstream modules operate on a high-performance foundation.

(2) Unified and Comprehensive Representation: The perceptual network architecture of SparseDrive elegantly unifies the tasks of detection, tracking, and mapping within a single framework. Its output—a set of dynamic agent instances (vehicles, pedestrians, etc.) and static map instances (lane lines, road boundaries, etc.), each enriched with feature and geometric information—is inherently temporally consistent. This unified and comprehensive sparse representation serves as an ideal input for modeling complex scene interactions and long-term behavior prediction.

It is crucial to emphasize that while a robust perception system is a prerequisite for safe driving, the innovation and exploration of perception algorithms do not constitute the core contribution of this work. Our research focuses on how to more effectively leverage this high-quality sparse scene representation for higher-level reasoning and decision-making. Therefore, we treat the SparseDrive perception module as a highly effective "scene encoder" that provides our system with an accurate description of "what is in the scene and where". The core innovations of this paper are built upon this foundation.

Sparse perception and multimodal candidate generation follow the SparseDrive foundation; our contribution focuses on the subsequent predictive spatiotemporal risk evaluation and ego-trajectory selection.

2.2 Ego-Centric Interactive Multi-modal Trajectory Prediction

Building upon the high-quality sparse scene representation, an autonomous driving system must possess the capability to reason about the future, particularly in predicting the future behaviors of traffic participants. The evolution of a driving scene is characterized by inherent uncertainty and multi-modality. For instance, a vehicle at an intersection might proceed straight, turn left, or turn right, according to the current

Just Accepted

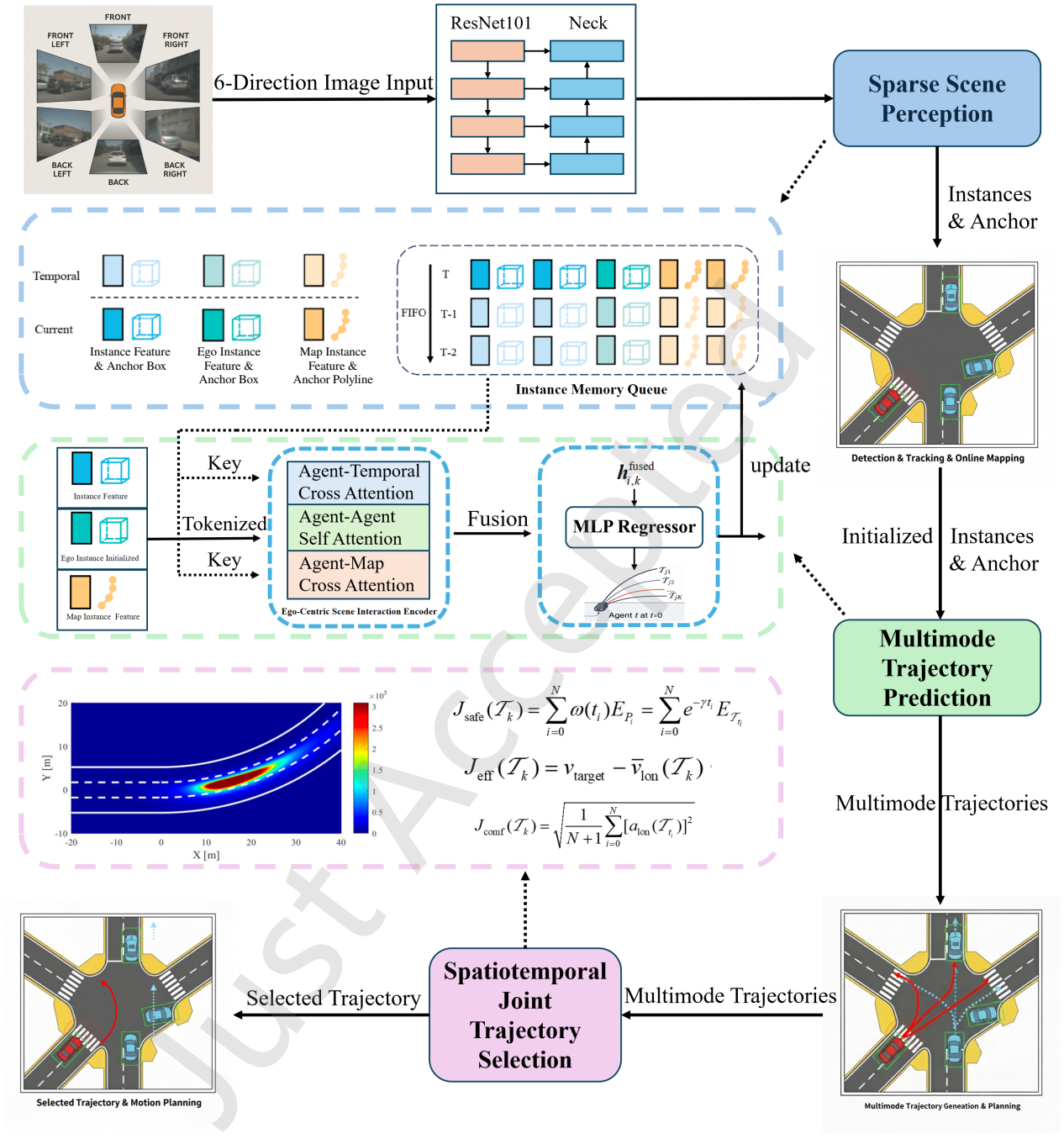


Fig. 1 The framework of our proposed generate-and-evaluate framework for end-to-end autonomous driving.

motion of surrounding interactive vehicles. A reliable prediction module must therefore be able to capture these diverse possibilities and produce a confidence score for each mode.

The Ego-Centric Interactive Prediction module exchanges information among the observed ego-vehicle, surrounding-agent, map, and temporal representations. Surrounding-agent predictions are not recomputed for each candidate ego trajectory, and should not be interpreted as counterfactual responses to individual proposed ego actions. Their predictions are used to construct the environment representation against which the ego candidates are evaluated.

For each relevant vehicle agent $N_{i,\, i\in\mathbb{N}}$ in the scene, our module generates a set of K distinct future trajectories, denoted as $\{\mathcal{T}_{i1}, \mathcal{T}_{i2}, \cdots, \mathcal{T}_{iK}\}$, $i \in \mathbb{N}$, $K \in \mathbb{N}^+$. Each trajectory $\mathcal{T}_{i,k}$ represents a plausible sequence of future coordinates about vehicle i over a time horizon T. A parallel

Just Accepted

classification branch produces a sigmoid confidence for each mode. Each candidate has a mode confidence and is not yet evaluated by the final planning cost. The sigmoid confidences are not normalized trajectory probabilities. The architecture is composed of two primary stages: an Ego-Centric Scene Interaction Encoder and an Interaction-Conditioned Multi-modal Trajectory Decoder.

2.2.1 *Ego-Centric Scene Interaction Encoder*

To enable interaction-aware predictions, we must first derive feature representations for each agent that are imbued with scene-wide context, especially the influence of the ego-vehicle. We employ a Transformer-based architecture for this purpose [25], leveraging its powerful self-attention and cross-attention mechanism to compute a dense graph of relationships between all scene elements.

The inputs to the encoder are tokenized representations of the ego-vehicle N_0, all other dynamic surrounding vehicles $N_{j,\, j\in\mathbb{N}^+}$, and relevant static map elements $M_{j,\, j\in\mathbb{N}^+}$. Each element's raw attributes (e.g., historical kinematics, lane geometry) are first embedded into a high-dimensional feature space. These feature vectors form an input sequence for the Transformer encoder.

The core of the encoder is the attention mechanism, which computes the updated representation of each element as a weighted sum of all other elements in the scene. The attention operation is formally defined as:

$$\text{Attention}\ (Q,K,V) = \text{softmax}\ (\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

where the query Q, key K, and value V matrices are linear projections of the input feature sequence; d_k is the per-head query/key dimension; n in Equations (2) and (3) is the attention-key count, distinct from the trajectory-mode count K. By applying this mechanism across the entire scene, the encoder explicitly models the critical dependencies:

(1) Ego-Vehicle's Contextual Influence on Other Agents: Through attention, surrounding-agent representations incorporate the observed ego-vehicle state and temporal context. This models statistical dependencies in the observed scene rather than a direct causal response to each proposed maneuver.

(2) Environmental Influence on the Ego-Vehicle's Action Space: Interaction here refers to contextual feature exchange. To enable intelligent planning, the system must understand how the actions of other agents constrain or create opportunities for the ego-vehicle. The model captures this reciprocal effect by allowing the features of surrounding agents to attend to the ego-vehicle's state. This mechanism synthesizes the collective behavior of the traffic scene—such as a group of cars slowing ahead or a vehicle aggressively encroaching from an adjacent lane—into a holistic situational awareness. This enriched understanding is critical, as it informs the ego-vehicle's own predictive model about the evolving dynamics. It effectively encodes how the actions of others define the safe and viable action space for the ego-vehicle, ensuring that its subsequent planning is not only based on its own intent but is also deeply responsive to the unfolding behavior of the entire traffic environment.

After several layers of such encoding, we obtain an interaction-aware context vector h_i^{ctx} for each agent $N_{i,\, i\in\mathbb{N}}$:

$$h_i^{\text{ctx}} = \sum_{j=1}^{n} \alpha_{ij}(W_V X_j) \tag{2}$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n}\exp(e_{ik})} \tag{3}$$

$$e_{ij} = \frac{(W_Q X_i)^T (W_K X_j)}{\sqrt{d_k}} \tag{4}$$

where W_V is a weight matrix used to convert each input X_j into a 'value' vector; α_{ij} is the weight coefficient, whose calculation process clearly shows how the context is formed; W_Q and W_K are also weight matrices, used to generate the 'query' vector and 'key' vector respectively. This interaction-aware context vector h_i^{ctx} is no longer a simple descriptor of the agent itself but a rich representation of its state within the dynamic context of reciprocal interactions, providing a robust foundation for prediction.

2.2.2 *Interaction-Conditioned Multi-modal Trajectory Decoder*

The decoder's function is to leverage the interaction-aware context vector h_i^{ctx} to generate a diverse set of K plausible future trajectories. Here K is the number of modes and k indexes a single mode. To achieve multi-modality, we adopt an approach based on a set of learnable intention anchors or mode queries. We define a set of K learnable query vectors, $\mathcal{Q}_{\text{mode}} = \{q_1, q_2, \cdots, q_K\}$, where $\forall k \in \{1, 2, \cdots, K\}$, $q_k \in \mathbb{R}^d$. Each query vector implicitly learns to represent a distinct behavioral intention, such as lane-keeping, yielding, or lane-changing.

For each agent $N_{i,\, i\in\mathbb{N}}$, the decoder performs K parallel prediction tasks. In each task k, the agent's context vector h_i^{ctx} is combined with a corresponding mode query q_k to form a mode-specific feature vector. This fusion can be achieved through addition or concatenation:

Just Accepted

$$h_{i,k}^{\text{fused}} = \text{Fusion}\ (h_i^{\text{ctx}}, q_k) \tag{5}$$

This fused feature vector $h_{i,k}^{\text{fused}}$ is then processed by a regression head, typically a Multi-Layer Perceptron (MLP), to regress the full trajectory $\mathcal{T}_{i,k}$:

$$\mathcal{T}_{i,k} = \text{MLP}_{\text{reg}}(h_{i,k}^{\text{fused}}) \tag{6}$$

Since the input h_i^{ctx} contains the observed scene information, each generated trajectory represents a plausible future under that context. The final output for agent $N_{i,\, i\in\mathbb{N}}$ is the candidate set, together with a confidence for each candidate, $\mathcal{T}_i = \{\mathcal{T}_{i1}, \mathcal{T}_{i2}, \cdots, \mathcal{T}_{iK}\}, \forall k \in \{1, 2, \cdots, K\}, \mathcal{T}_{ik} \in \mathbb{R}^d$.

This interactive multimodal trajectory generation scheme implicitly forces the different mode queries and their associated MLP weights to specialize in distinct motion patterns. Over time, the model learns to produce a diverse set of trajectories that effectively covers the space of plausible, interaction-conditioned outcomes [26]. This rich set of futures provides a comprehensive basis for the downstream risk assessment and decision-making modules.

2.3 Optimal Trajectory Selection via Spatiotemporal Joint Assessment

Following the generation of multimodal trajectories for both the ego-vehicle and surrounding agents as detailed in Section 2.2, the next critical task is to select a single, optimal trajectory for the ego-vehicle. This selection must be both safe and efficient, navigating a complex and dynamic environment. To achieve this, we propose a two-stage methodology. First, we approximate the environmental futures by selecting one highest-confidence representative trajectory for each surrounding agent and use these to construct a predictive, spatiotemporal risk field. Second, we assess each of the ego-vehicle's candidate trajectories against this dynamic risk field to identify the one that minimizes exposure to potential hazards while maintaining operational efficiency.

2.3.1 Confidence-Based Trajectory Selection and Predictive Risk Field Construction

The foundation of our risk assessment is a predictive risk field that quantifies potential danger across space and time [27]. To construct this field, we first select a representative future from the multimodal predictions for each surrounding agent. From the set of candidate trajectories $\{\mathcal{T}_{j1}, \mathcal{T}_{j2}, \cdots, \mathcal{T}_{jK}\}$ generated for each surrounding agent $N_{j,\, j\in\mathbb{N}^+}$, we select one highest-scoring representative trajectory by employing a comprehensive scoring mechanism.

The general scoring form combines the model confidence, kinematic feasibility, and compliance with traffic rules and social norms as follows:

$$\begin{aligned} Score(\mathcal{T}_{jk}) = \omega_1 \text{P}_{\text{model}}(\mathcal{T}_{jk}) - \omega_2 \text{C}_{\text{kin}}(\mathcal{T}_{jk}) \\ - \omega_3 \text{C}_{\text{soc}}(\mathcal{T}_{jk}) \end{aligned} \tag{7}$$

where $\text{P}_{\text{model}}(\mathcal{T}_{jk})$ is the sigmoid mode confidence, rather than a normalized trajectory likelihood; $\text{C}_{\text{kin}}(\mathcal{T}_{jk})$ is a cost function penalizing kinematic infeasibility, such as excessive acceleration, jerk, or curvature; $\text{C}_{\text{soc}}(\mathcal{T}_{jk})$ is a cost function penalizing violations of social norms and traffic rules, such as departing from the lane center, illegal maneuvers, or unsafe proximity to other agents; ω_1, ω_2, ω_3 are weighting factors that balance the importance of model confidence, physical realism, and rule adherence.

In the present surrounding-agent selection, $\omega_1 = 1$ and $\omega_2 = \omega_3 = 0$, so the highest-confidence mode is selected.

The trajectory with the highest score is chosen as the representative future path for surrounding agent $N_{j,\, j\in\mathbb{N}^+}$: $\mathcal{T}_{j_{\text{selected}}} = \arg\max(Score(\mathcal{T}_{jk}))$.

Using one mode per surrounding agent avoids enumerating joint combinations of modes. The approximation can omit a low-confidence but hazardous future. The selected trajectory is a representative future, and the field around it does not cover every alternative behavior.

Once a representative trajectory is selected for every surrounding agent, these selected paths form the basis for modeling the future evolution of the traffic environment. The risk field of each ego-vehicle's candidate trajectory is modeled as a sequence of discrete matrices for m future time steps, typically at intervals of 0.5 seconds. The total risk at any point (x, y) is a superposition of contributions from both dynamic and static objects.

For each dynamic object $N_{j,\, j\in\mathbb{N}^+}$, its risk contribution $E_d^j(x, y)$ at a location (x, y) is a composite of a potential energy field $E_{d,p}^j(x, y)$ and a kinetic energy field $E_{d,k}^j(x, y)$, as formulated below:

$$E_d^j(x, y) = E_{d,p}^j(x, y) + E_{d,k}^j(x, y),\ j = 1, 2, \cdots, n \tag{8}$$

where the potential energy field $E_{d,p}^j(x, y)$ quantifies the inherent threat posed by an object's presence, influenced by its mass m and velocity v. It is modeled using a virtual mass M_v^j and its proximity $D(x, y)$ to a given point:

$$\begin{aligned} E_{d,p}^j(x, y) &= \frac{\tau r_a M_v^j}{\vartheta D(x, y)} + \frac{\tau r_a M_v^j}{E_{\max}} \\ M_v^j &= mT(\alpha v^\beta + \gamma) \end{aligned} \tag{9}$$

Just Accepted

where τ, r_a, ϑ, $E_{\max}$, T, α, β, γ are the parameters of the potential energy field $E^j_{d,p}(x,y)$; $D(x,y)$ is the minimum distance from the point (x,y) to the object's boundary.

The aforementioned parameters are primarily set based on vehicle kinematics priors and empirical rules. For instance, τ is set to the average reaction time, and r_a is determined by the physical dimensions of the ego vehicle. Meanwhile, ϑ and $E_{\max}$ are fixed hyperparameters set prior to training to ensure the spatial smoothness of the potential field, rather than learnable network weights. The T and γ in this virtual-mass model are distinct from the trajectory horizon and temporal decay parameter below.

The kinetic energy field $E_{d,ki}(x,y)$ captures the risk associated with the object's motion. To model the anisotropic nature of this risk, we define a weighted distance $D_w(x,y)$:

$$D_w(x,y) = \sqrt{\frac{D_{lo}(x,y)^2}{1-(\frac{D_{la}(x,y)}{w})^2}} \tag{10}$$

where w is the width of the object $N_{j,\,j\in\mathbb{N}^+}$, $D_{lo}(x,y)$ and $D_{la}(x,y)$ are the longitudinal and lateral distance of position (x,y) to the object $N_{j,\,j\in\mathbb{N}^+}$ in its Frenet coordinate system, respectively. The kinetic energy field is then modeled as:

$$E^j_{d,k}(x,y) = \frac{{v_r}^2}{2M^j_v D_w(x,y)(1-\frac{D_w(x,y)}{v_r t})} \tag{11}$$

where v_r is the relative velocity between the object $N_{j,\,j\in\mathbb{N}^+}$ and the ego-vehicle.

Equation (10) is real-valued within the lateral band $|D_{la}| < w$. For positive virtual mass, $v_r > 0$, and $t > 0$, Equation (11) is positive in the interior domain $0 < D_w < v_r t$. Zero relative speed and the singular distance boundaries cannot be treated as ordinary substitution points. These expressions are local analytical risk proxies for the stated domain.

For a static object $N_{n,\,n\in\mathbb{N}^+}$ (e.g., a lane line), its risk contribution $E^j_s(x,y)$ consists only of a potential energy component, representing a repulsive force. It is defined as:

$$E^n_s(x,y) = k_s(\frac{\xi D}{2} - \min\left\{\frac{\xi D}{2}, \mathrm{dis}(x,y)\right\})^2 \tag{12}$$

where the term D corresponds to the physical width of the static object $N_{n,\,n\in\mathbb{N}^+}$, which in many scenarios is the width of the lane line. The parameter ξ is a scaling factor that is constrained to be greater than or equal to one. The function $\mathrm{dis}(x,y)$ calculates the distance from a given position (x,y) to the static object $N_{n,\,n\in\mathbb{N}^+}$. Furthermore, the parameter k_s is configured based on the object's classification. To illustrate, lane lines are considered soft constraints and are thus assigned a relatively low k_s value. In contrast, road barriers are assigned a significantly high k_s value, which supplies a strong but finite deterrent rather than a collision-avoidance guarantee.

Through this risk modeling approach, the relative risk values $E(x,y) = \sum_{j=1}^{N} E^j_d(x,y) + \sum_{n=1}^{M} E^n_s(x,y)$ at both current and future time points along the candidate trajectories of the ego-vehicle can be calculated. Notably, this method comprehensively considers driving environment information, providing a solid theoretical basis for selecting the safest trajectories. Fig. 2 illustrates the fusion risk field between dynamic objects and static lane markings in a single-frame scene.

Here N and M count dynamic and static objects, respectively; the N in the trajectory costs below denotes temporal discretization.

2.3.2 *Spatiotemporal Joint Evaluation and Selection of the Ego-Vehicle Trajectory*

With the predictive risk fields established and the generation of a set of candidate trajectories of $\{\mathcal{T}_1, \mathcal{T}_2, \cdots, \mathcal{T}_K\}$ ego-vehicle, a robust evaluation mechanism is required to select the optimal one for execution. The selection process is formulated as a multi-objective optimization problem, where each trajectory is assessed based on a weighted combination of three critical criteria: safety, efficiency, and comfort. The trajectory that achieves the most desirable balance among these competing objectives is deemed optimal. To this end, we define a comprehensive cost function, J_{total}, which is a linear combination of the cost components for each criterion.

Safety Cost (J_{safe}): The safety of a trajectory is the paramount consideration. We quantify the risk associated with each candidate trajectory by leveraging the safety field method detailed in Section 2.3.1. A candidate trajectory, denoted as $\mathcal{T}_k$, is defined over a future prediction horizon T. To analyze its evolving risk profile, we discretize the trajectory into $N+1$ temporal slices at intervals of Δt, such that $T = N\Delta t$, each path slice considers only the risk field information of the corresponding spatiotemporal slice, as illustrated in Fig. 3 with the candidate curve as an example. The state of the ego-vehicle at each slice i (where $i = 0, 1, \cdots, N$) is represented by the point $P_i = \mathcal{T}_{t_i}$, with $t_i = i\Delta t$.

At each point P_i, we can calculate a risk value, E_{P_i}, using the potential field function from the previous Section 2.3.1. However, risks are not uniform in their importance over the time horizon. A potential collision in the immediate future is

Just Accepted

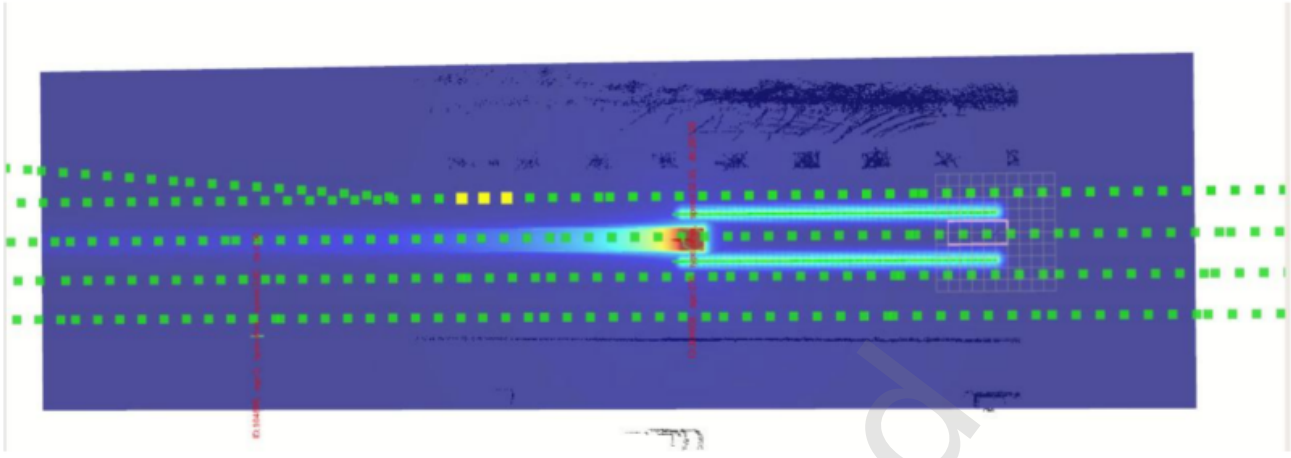


Fig. 2 The fusion risk field between dynamic objects and static lane markings in a single-frame scene.

far more critical than one predicted to occur several seconds later, as the latter allows more time for corrective action. To reflect this temporal urgency, we introduce a time-dependent weighting factor. An exponential decay function is employed to assign higher weights to points closer to the present time $t = 0$.

The total safety cost for a trajectory $\mathcal{T}_k$, denoted as $J_{\text{safe}}(\mathcal{T}_k)$, is calculated as the weighted sum of the risk values at each discrete point:

$$J_{\text{safe}}(\mathcal{T}_k) = \sum_{i=0}^{N} \omega(t_i) E_{P_i} = \sum_{i=0}^{N} e^{-\gamma t_i} E_{\mathcal{T}_{t_i}} \tag{13}$$

where $t_i = i\Delta t$ is the timestamp of the ith slice; $E_{P_i} = E_{\mathcal{T}_{t_i}}$ is the risk value at the vehicle's position at time t_i; γ is a positive decay-rate parameter. A larger γ places a stronger emphasis on immediate safety.

A larger positive γ suppresses the contribution of later time slices more strongly. When immediate risk dominates the candidate scores, the planner may favor harder braking. A smaller positive γ retains more influence from later risks, giving the planner room for smoother adjustments over the horizon. We selected the time-decay parameter γ by grid search to balance near-term collision avoidance with longer-term interactions.

Efficiency Cost (J_{eff}): Efficiency in motion planning is primarily associated with maintaining progress towards a destination. A common and effective proxy for this is the vehicle's longitudinal velocity. We define the efficiency metric as the average longitudinal velocity over the trajectory's horizon. Since our goal is to formulate a cost function to be minimized, we define the efficiency cost as the signed difference between a reference velocity v_{target} and the mean longitudinal velocity. This reference is constant across candidates within a scene. The efficiency cost $J_{\text{eff}}(\mathcal{T}_k)$, is thus defined as:

$$J_{\text{eff}}(\mathcal{T}_k) = v_{\text{target}} - \bar{v}_{\text{lon}}(\mathcal{T}_k) \tag{14}$$

where $\bar{v}_{\text{lon}}(\mathcal{T}_k)$ is the average longitudinal velocity over the trajectory. For the discretized trajectory, this is calculated as:

$$\bar{v}_{\text{lon}}(\mathcal{T}_k) = \frac{1}{N+1} \sum_{i=0}^{N} v_{\text{lon}}(\mathcal{T}_{t_i}) \tag{15}$$

where $v_{\text{lon}}(\mathcal{T}_{t_i})$ is the longitudinal velocity at point P_i. Minimizing this signed cost favors higher mean longitudinal speed; it is not a symmetric target-speed penalty and does not independently enforce a speed limit.

Comfort Cost (J_{comf}): The comfort cost measures longitudinal acceleration magnitude rather than its rate of change. A smooth ride is characterized by minimal longitudinal and lateral accelerations. For this evaluation, we focus on longitudinal comfort, as it is often the most significant factor in on-road driving scenarios like acceleration and braking. The Root Mean Square (RMS) of the longitudinal acceleration is a standard and effective metric for quantifying ride discomfort.

The comfort cost J_{comf}, is defined as the RMS of the longitudinal acceleration a_{lon} along the trajectory:

$$J_{\text{comf}}(\mathcal{T}_k) = \sqrt{\frac{1}{N+1} \sum_{i=0}^{N} [a_{\text{lon}}(\mathcal{T}_{t_i})]^2} \tag{16}$$

where $a_{\text{lon}}(\mathcal{T}_{t_i})$ is the longitudinal acceleration at the discrete point P_i. By minimizing this value, the planner penalizes trajectories with aggressive acceleration or braking ma-

Just Accepted

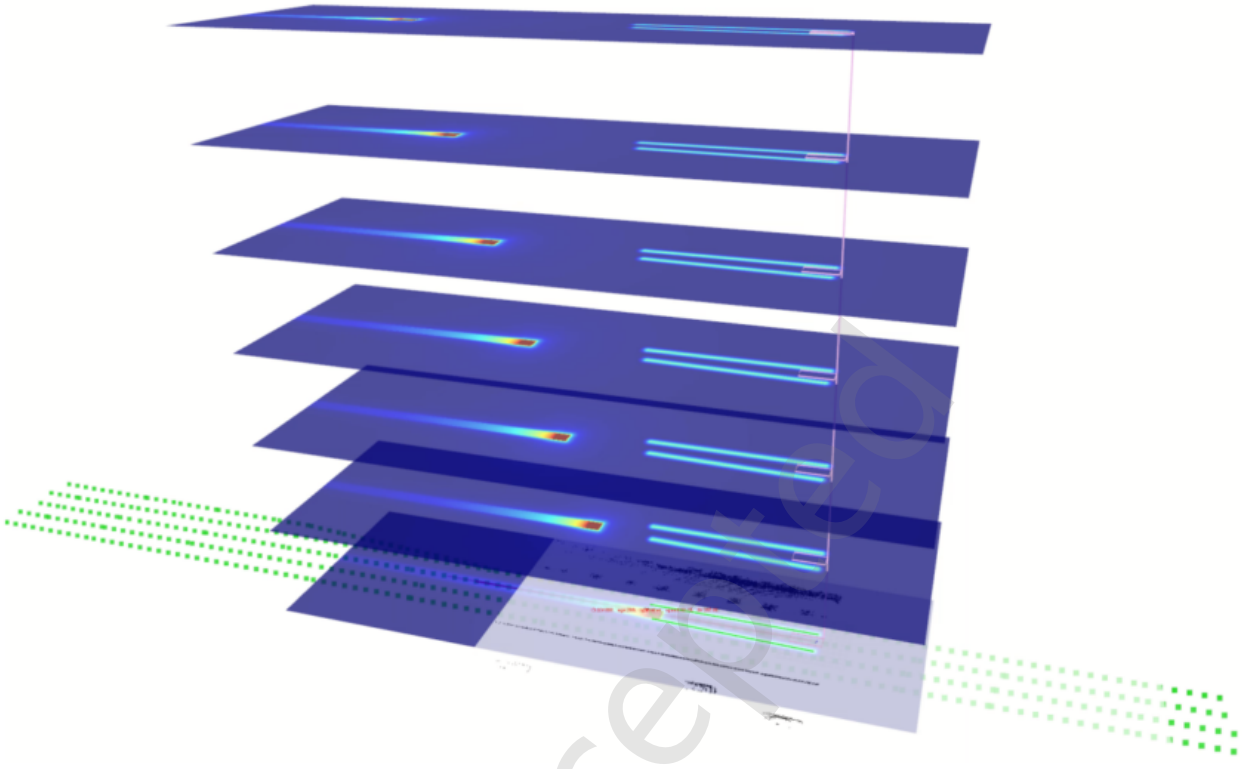


Fig. 3 The comprehensive spatiotemporal-discretised slice of the predictive driving risk field.

neuvers, favoring those that provide a smoother and more comfortable experience for the occupants.

The final step is to synthesize these individual cost components into a single, comprehensive cost function J_{total}, which allows for a holistic evaluation of each candidate trajectory. This is achieved through a weighted sum of the safety, efficiency, and comfort costs:

$$J_{\text{total}}(\mathcal{T}_k) = \omega_{\text{safe}} J_{\text{safe}}(\mathcal{T}_k) + \omega_{\text{eff}} J_{\text{eff}}(\mathcal{T}_k) + \omega_{\text{comf}} J_{\text{comf}}(\mathcal{T}_k) \tag{17}$$

where non-negative weights ω_{safe}, ω_{eff}, and ω_{comf} represent the relative importance of each objective. In our baseline evaluation, we adopted a balanced set of weights to simulate standard driving behavior. However, this framework allows for flexible switching of driving styles by adjusting these weights: for instance, a higher ω_{eff} for a more aggressive style, or a higher ω_{comf} for a conservative, comfort-oriented style, with ω_{safe} typically assigned the highest priority.

The weights must be interpreted together with the scaling of their component costs; coefficients applied after different cost transformations are not directly equivalent to the linear weights in Equation (17).

The optimal trajectory $\mathcal{T}^*$, is the one from the set of all candidate trajectories $\{\mathcal{T}_1, \mathcal{T}_2, \cdots, \mathcal{T}_K\}$ that minimizes this total cost:

$$\mathcal{T}^* = \arg\min J_{\text{total}}(\mathcal{T}_k) \tag{18}$$

This selected trajectory $\mathcal{T}^*$ represents the best-compromise solution, optimally balancing the requirements of safety, efficiency, and comfort according to the predefined weights, and is then forwarded to the vehicle controller for execution.

3 Experiment

3.1 Experimental Settings

3.1.1 Datasets

We evaluate the proposed framework on the nuScenes v1.0 dataset, using the val and mini_val evaluation collections. The six camera inputs are CAM_FRONT, CAM_FRONT_RIGHT, CAM_FRONT_LEFT, CAM_BACK, CAM_BACK_LEFT, and CAM_BACK_RIGHT. The data were collected in Boston and Singapore.

The processed val and mini_val evaluation collections contain 6,019 and 81 scenario cases, respectively. We use the term processed scenario cases to distinguish these counts

Just Accepted

from official continuous-recording scene records and logs.

3.1.2 Implementation Details

Our model is implemented using the PyTorch framework. Consistent with the base setting of previous work SparseDrive [23], we employ ResNet101 initialized with ImageNet weights as the image backbone. The input image resolution is set to 512×1408 to ensure sufficient perception granularity for distant small objects.

All experiments are conducted on a high-performance cluster equipped with 4 NVIDIA A100 (80GB) GPUs. Stage 1 learns perception. Stage 2 initializes from Stage 1 and trains perception, prediction, and planning using supervised task losses. Closest-mode assignment defines classification targets and selects regression outputs; gradients pass through these outputs rather than the discrete assignment index. Final risk-based selection is used during inference, not as an additional training loss. Table 1 lists training hyperparameters.

The model predicts six modes per surrounding agent and six ego candidates per navigation command. Agent prediction covers 12 future steps (6 s), and ego planning covers six steps (3 s), at 0.5 s intervals. Risk construction uses the current state and the first six predicted agent steps. Features have 256 channels, with eight attention heads, three interaction blocks, and an observation queue of four entries including the current observation.

The numerical risk field uses a 640×640 grid spanning -50 to 50 m on each axis, a detection-confidence threshold of 0.3, and an ego-footprint scale factor of 1.1. Its potential calculation uses a scale coefficient of 0.013, an inverse-distance exponent of 1.63, a multiplicative factor of 1, and a fixed virtual mass of 1000. The numerical potential-field calculation clips its raw potential at 1500 before normalization. The inverse-distance exponent is distinct from the time-decay rate in Equation (13). These numerical settings should not be identified directly with all analytical coefficients in Equations (9)–(17); clipping limits scale but does not define the singular boundaries of Equations (10) and (11).

3.1.3 Evaluation Metrics

A core premise of this work is that human driving demonstrations are not always optimal. Consequently, traditional imitation learning metrics—such as L2 displacement error, minADE, minFDE, MR, and EPA—are insufficient for our purpose, as they fundamentally penalize any deviation from the logged human trajectory, even if the model's behavior is safer or smoother.

We use the following driving-quality metrics to assess the selected trajectories in open-loop evaluation; they do not establish superiority to the logged human driver. Instead, we focus on three practical dimensions of autonomous driving quality:

Safety (Collision Rate): The primary metric. We evaluate the Collision Rate (CR) by checking if the ego-vehicle's planned trajectory intersects with any ground-truth objects or map boundaries.

The ego trajectory is evaluated over a 3-second horizon at six future timestamps separated by 0.5 s. Let $S_{T-\text{collision}}$ denote the collision count at time T and $S_{\text{total - samples}}$ the number of evaluated planning instances. The expression for CR is:

$$\text{CR} = \frac{\sum_{T} S_{T-\text{collision}}}{S_{\text{total - samples}}}, \quad T = \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0\} \tag{19}$$

Counts are summed over the six timestamps, so one planning instance can contribute more than once. CR is neither the proportion of trajectories with any collision nor a time average; multiplication by 100 gives the percentage scale in the tables. Timestamps are not excluded solely because the logged ground-truth ego trajectory also collides. Collision checking uses an oriented ego rectangle of 4.084×1.85 m with a 0.5 m forward reference offset, at the same 0.5 s intervals.

Efficiency: Measured by the ability to maintain a reasonable driving speed without unnecessary stops. We report average longitudinal velocity (ALV), calculated as the mean longitudinal velocity along each trajectory and then averaged over valid evaluated planning instances.

Comfort: RMSA is the root mean square of longitudinal acceleration, rather than its rate of change. Dataset-level RMSA averages the per-trajectory RMSA values over valid evaluated planning instances. A lower value indicates smaller longitudinal acceleration magnitude. The per-trajectory expression is:

$$RMSA = \sqrt{\frac{1}{N}\sum_{i=1}^{N} a_i^2} \tag{20}$$

where a_i denotes the instantaneous longitudinal acceleration value of the ego vehicle at the i-th sampling point; N is the total number of sampling points.

Equation (20) uses N values indexed from 1 to N; sums indexed from 0 to N in the trajectory costs contain $N + 1$ values.

Just Accepted

Table 1 Training hyper-parameters.

Training stage	Batch size	Epochs	Lr	Backbone lr scale	Weight decay
Stage 1	4	80	3×10^{-4}	0.1	1×10^{-3}
Stage 2	4	10	3×10^{-4}	0.1	1×10^{-3}

3.2 Comparative Experiment

We compare the proposed framework with SparseDrive[23] and UniAD 2.0[10] on nuScenes val, and with SparseDrive on mini_val. UniAD 2.0 is included in the val comparison only.

3.2.1 Main Results in Complex Corner Cases

The mini_val evaluation collection contains the selected processed scenario cases. Table 2 summarizes the performance comparison between our method and the baseline SparseDrive.

In complex scenarios, safety is the paramount differentiator. As shown in Table 2, our method achieves a 33.5% reduction in Collision Rate (0.242%→0.161%) compared to the baseline. The risk field provides a trajectory-scoring cost rather than a hard collision-rejection constraint. Furthermore, our method improves ride comfort (RMSA: 0.457→0.403 m/s^2) without compromising traffic efficiency, achieving smoother trajectory output.

3.2.2 Main Results in Large-scale General Scenarios

To evaluate general driving quality, we use the nuScenes val collection of 6,019 processed scenario cases and compare with SparseDrive and UniAD 2.0. The results are presented in Table 3. This is an open-loop comparison.

On the val benchmark, the results show a trade-off among the evaluated criteria.

Safety Trade-off: Our method has a collision rate of 0.204%, compared with 0.104% for SparseDrive and 0.530% for UniAD 2.0. Relative to SparseDrive, higher ALV and lower RMSA are obtained at a higher collision rate; these results do not establish Pareto optimality or improvement in every criterion.

Active Efficiency: We achieve an ALV of 5.440 m/s, compared with 5.290 m/s for UniAD 2.0 and 5.242 m/s for SparseDrive. Higher average speed should be interpreted together with collision incidence.

Dominance in Ride Comfort: Our method achieves an exceptionally low RMSA of 0.194 m/s^2, which is 30.7% lower than SparseDrive (0.280 m/s^2) and 61.4% lower than UniAD 2.0 (0.502 m/s^2). This indicates smaller longitudinal acceleration magnitude under the stated metric; it does not establish the causal origin of the differences.

3.3 Qualitative Results

3.3.1 Visualization

To demonstrate how our Generate-and-Evaluate framework transcends simple imitation, Fig. 4 visualizes the decision-making process in a typical "Turn Right" scenario.

Risk-Based Decision Making (Top-Right). The core of our approach is the spatiotemporal risk field shown in the top-right panel. Risk values are normalized and rendered using a Jet colormap, where red indicates high-energy collision risks and blue represents safe navigable space. Unlike behavioral cloning which blindly regresses to human trajectories, our planner objectively evaluates candidate paths against this risk field. As illustrated, the ego-vehicle (red icon) identifies a "low-energy valley", effectively avoiding the high-risk zones generated by static obstacles and dynamic agents. This visualization illustrates candidate evaluation, rather than quantitative superiority to the logged human trajectory.

Perception and Verification (Top-Left & Bottom). The Top-Left panel displays the Ground Truth trajectories. While provided for reference, our model's planned path is not constrained to strictly mimic the human driver but is optimized for safety and comfort within the risk field. The Bottom panels display the synchronized 6-camera inputs. The robust 3D detections (bounding boxes) serve as the precise geometric foundation for constructing the risk field, ensuring that the planner's "energy valleys" correspond to physical reality.

3.3.2 Case Study: Interactive Decision Making in Complex Constraints

To illustrate candidate evaluation in complex traffic dynamics, we analyze this challenging scenario involving multi-agent interaction and strict static constraints, as depicted in Fig. 4.

Scenario Description. The ego-vehicle faces a highly constrained environment: Left: A double yellow line strictly prohibits overtaking or lane deviation. Front: A slow-moving vehicle obstructs the current lane, significantly impeding traffic flow. Right: A pedestrian stands approximately one lane-width away, exhibiting an intention to cross, while a large truck is approaching from the rear-right.

Baseline Comparison. Figure 4 does not include matched baseline rollouts; the case therefore illustrates the scoring

Just Accepted

Table 2 Main results on the nuScenes mini_val evaluation collection.

Models/Configuration	Collision Rate (%) ↓	Efficiency (ALV, m/s) ↑	Comfort (RMSA, m/s^2) ↓
SparseDrive	0.242	5.411	0.457
Ours	0.161	5.409	0.403

Table 3 Main results on the nuScenes val evaluation collection.

Models/Configuration	Collision Rate (%) ↓	Efficiency (ALV, m/s) ↑	Comfort (RMSA, m/s^2) ↓
UniAD 2.0	0.530	5.290	0.502
SparseDrive	0.104	5.242	0.280
Ours	0.204	5.440	0.194

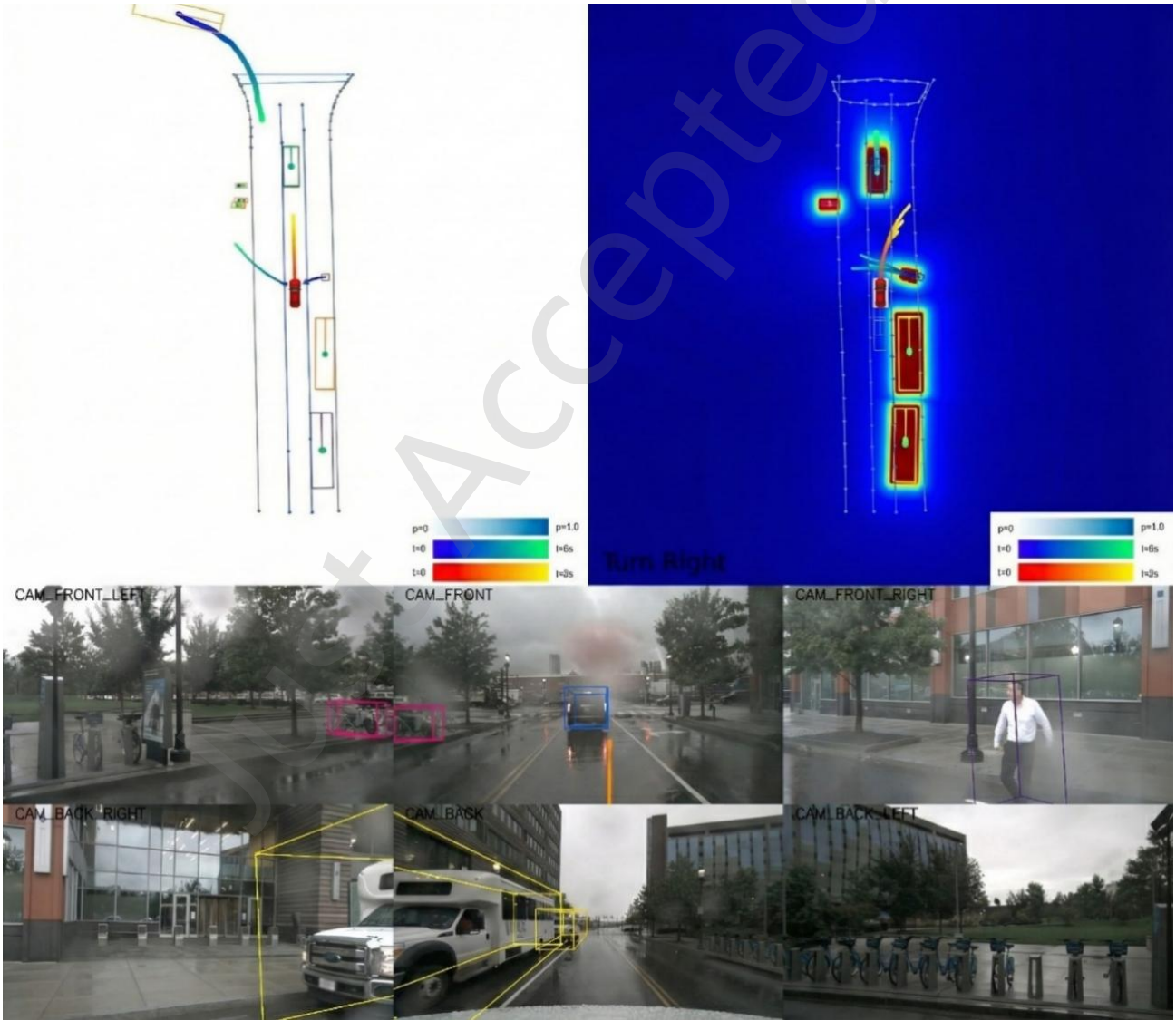


Fig. 4 Visualization of a "Turn Right" scenario. Top-Left: Ground truth trajectories provided for reference, denotes the trajectories of the ego-vehicle, pedestrians and vehicles in the opposite lane. Top-Right: The generated Spatiotemporal Risk Field (Red: High Risk; Blue: Low Risk). The planner identifies a safe corridor based on objective risk minimization rather than pure imitation. Bottom: Synchronized surround-view camera inputs with 3D detections, which form the basis for risk assessment.

Just Accepted

mechanism rather than a measured advantage over a particular baseline.

Ours: Risk-Aware Active Planning. In contrast, our Generate-and-Evaluate framework executes a precise acceleration and lane-change maneuver to the right. This decision, while seemingly aggressive, is theoretically grounded in the Spatiotemporal Risk Field: Unified Constraint Mapping: The algorithm accurately perceives the double yellow line as a finite risk penalty that discourages nearby candidates. Dynamic Risk Integration: Simultaneously, the slow-moving front vehicle generates a high-energy zone directly ahead. Crucially, the risk field calculates that the pedestrian's kinetic energy distribution (based on their speed and distance) does not yet encroach upon the target lane's safe valley, and the gap ahead of the rear-right truck is sufficient. Candidate Selection: By synthesizing these heterogeneous factors—road rules (lines), dynamic obstacles (vehicles), and vulnerable road users (pedestrians)—into a single scalar risk value, the planner identifies that the "accelerate-and-merge" trajectory offers the lowest cost among the generated candidates.

This case illustrates how road constraints and predicted agent motion enter candidate scoring. The selected trajectory minimizes the chosen cost within the generated candidate set; the example does not establish a global optimum, guaranteed traffic-rule compliance, or superiority to human driving.

4 Conclusion

In conclusion, this paper presents a novel "Generate-and-Evaluate" framework for end-to-end motion planning that fundamentally transcends the limitations of traditional imitation learning. By adding risk-based candidate evaluation after supervised trajectory generation on the SparseDrive foundation, our approach effectively addresses the inherent multi-modality and interactivity of complex driving scenarios. The integration of an ego-centric interactive prediction module with a principled spatiotemporal risk field enables the system to generate diverse plausible futures and select a low-cost trajectory from the generated candidates under the selected risk model and weights. Experimental results on nuScenes mini_val show reduced collision rate and longitudinal acceleration RMS with nearly unchanged speed relative to SparseDrive. On val, higher ALV and lower RMSA come with a higher collision rate, indicating a trade-off among the evaluated criteria.

Using one representative future per agent can omit hazardous alternatives, and the risk field is not a complete representation of environmental uncertainty.

It is worth noting that the proposed trajectory generation and risk assessment mechanisms are primarily based on highly parallel, lightweight tensor operations. On existing computing platforms, the parallel evaluation of multiple candidate trajectories adds only a few milliseconds of computational overhead. This fully meets the strict high-frequency real-time requirements of autonomous driving systems, laying a solid foundation for its deployment in real vehicles or closed-loop simulations.

In the future, we aim to expand the validation scope of our framework beyond the current open-loop evaluation on open-source datasets. While the current results demonstrate strong theoretical and empirical performance, the dynamic nature of autonomous driving necessitates rigorous closed-loop testing to fully verify the planner's responsiveness to cumulative errors and continuous environmental feedback. Therefore, we plan to deploy our model within high-fidelity simulators, such as CARLA, and eventually on physical autonomous vehicles. This progression will allow us to assess the framework's robustness in real-time interactive settings and address the challenges of the sim-to-real gap, further bridging the divide between theoretical motion planning and practical deployment.

Replication and data sharing

The code of the proposed method is available upon request by emailing the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Tsinghua University Motor Corporation Joint Research Center for AI Technology Automated Vehicle.

This research was funded by the National Natural Science Foundation of China (No. 52131201) and Zhongguancun High-end Leading Talents Aggregation Project (No. 202504841093).

Just Accepted

Reference

[1] C. Gao, G. Wang, W. Shi, Z. Wang, and Y. Chen, Autonomous driving security: State of the art and challenges, *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7572–7595, 2022.

[2] J. Zhao, W. Zhao, B. Deng, Z. Wang, F. Zhang, W. Zheng, W. Cao, J. Nan, Y. Lian, and A. F. Burke, Autonomous driving system: A comprehensive survey, *Expert Syst. Appl.*, vol. 242, Art. no. 122836, 2024.

[3] Y. Dong, B. van Arem, and H. Farah, Toward developing socially compliant automated vehicles: Advances, expert insights, and a conceptual framework, *Commun. Transp. Res.*, vol. 5, Art. no. 100207, 2025.

[4] J. Hu, Y. Wang, S. Cheng, J. Xu, N. Wang, B. Fu, Z. Ning, J. Li, H. Chen, C. Feng, and Y. Zhang, A survey of decision-making and planning methods for self-driving vehicles, *Front. Neurorobot.*, vol. 19, Art. no. 1451923, 2025.

[5] W. Chu, K. Yang, S. Li, and X. Tang, Motion planning for autonomous driving with real traffic data validation, *Chin. J. Mech. Eng.*, vol. 37, no. 1, Art. no. 6, 2024.

[6] Q. Gu, G. Bai, G. Wang, Y. Meng, and L. Liu, 3D search based hybrid optimal trajectory planning for autonomous LHD in turning maneuvers, *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 17704–17716, 2022.

[7] Q. Liu, J. Zhang, J. Ge, C. Chang, Z. Li, S. Li, and L. Li, Integrating spatial-temporal risk maps with candidate trajectory trees for explainable autonomous driving planning, *Commun. Transp. Res.*, vol. 5, Art. no. 100161, 2025.

[8] X. Tang, K. Yang, H. Wang, W. Yu, X. Yang, T. Liu, and J. Li, Driving environment uncertainty-aware motion planning for autonomous vehicles, *Chin. J. Mech. Eng.*, vol. 35, no. 1, Art. no. 120, 2022.

[9] S. Chen, M. Wang, W. Song, Y. Yang, Y. Li, and M. Fu, Stabilization approaches for reinforcement learning-based end-to-end autonomous driving, *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 4740–4750, 2020.

[10] Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang, L. Lu, X. Jia, Q. Liu, J. Dai, Y. Qiao, and H. Li, Planning-oriented autonomous driving, in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, Canada, 2023, pp. 17853–17862.

[11] B. Jiang, S. Chen, Q. Xu, B. Liao, J. Chen, H. Zhou, Q. Zhang, W. Liu, C. Huang, and X. Wang, VAD: Vectorized scene representation for efficient autonomous driving, in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Paris, France, 2023, pp. 8306–8316.

[12] H. Huang, Y. Liu, J. Liu, Q. Yang, J. Wang, D. Abbink, and A. Zgonnikov, General optimal trajectory planning: Enabling autonomous vehicles with the principle of least action, *Engineering*, vol. 33, pp. 63–76, 2024.

[13] Y. Yan, L. Peng, T. Shen, J. Wang, D. Pi, D. Cao, and G. Yin, A multi-vehicle game-theoretic framework for decision making and planning of autonomous vehicles in mixed traffic, *IEEE Trans. Intell. Veh.*, vol. 8, no. 11, pp. 4572–4587, 2023.

[14] D. González, J. Pérez, V. Milanés, and F. Nashashibi, A review of motion planning techniques for automated vehicles, *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 1135–1145, 2016.

[15] L. Chen, P. Wu, K. Chitta, B. Jaeger, A. Geiger, and H. Li, End-to-end autonomous driving: Challenges and frontiers, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 10164–10183, 2024.

[16] Z. Huang, J. Zhang, R. Tian, and Y. Zhang, End-to-end autonomous driving decision based on deep reinforcement learning, in *Proc. 5th Int. Conf. Control, Autom. Robot. (ICCAR)*, Beijing, China, 2019, pp. 658–662.

[17] Y. Tian, X. Cao, K. Huang, C. Fei, Z. Zheng, and X. Ji, Learning to drive like human beings: A method based on deep reinforcement learning, *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6357–6367, 2022.

[18] Z. Mao, Y. Liu, and X. Qu, Integrating big data analytics in autonomous driving: An unsupervised hierarchical reinforcement learning approach, *Transp. Res. Part C Emerg. Technol.*, vol. 162, Art. no. 104606, 2024.

[19] S. Gu, L. Yang, Y. Du, G. Chen, F. Walter, J. Wang, and A. Knoll, A review of safe reinforcement learning: Methods, theories, and applications, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 11216–11235, 2024.

[20] Y. Pan, C. A. Cheng, K. Saigol, K. Lee, X. Yan, E. A. Theodorou, and B. Boots, Imitation learning for agile autonomous driving, *Int. J. Robot. Res.*, vol. 39, nos. 2–3, pp. 286–302, 2020.

[21] S. Wang, Y. Ni, C. Miao, J. Sun, and J. Sun, A multiagent social interaction model for autonomous vehicle testing, *Commun. Transp. Res.*, vol. 5, Art. no. 100183, 2025.

[22] P. Czechowski, B. Kawa, M. Sakhai, and M. Wielgosz, Deep reinforcement and IL for autonomous driving: A review in the CARLA simulation environment, *Appl. Sci.*, vol. 15, no. 16, Art. no. 8972, 2025.

[23] W. Sun, X. Lin, Y. Shi, C. Zhang, H. Wu, and S. Zheng, SparseDrive: End-to-end autonomous driving via sparse scene representation, in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Atlanta, GA, USA, 2025, pp. 8795–8801.

[24] Y. Zheng, R. Liang, K. Zheng, J. Zheng, L. Mao, J. Li, W. Gu, R. Ai, S. E. Li, X. Zhan, and J. Liu, Diffusion-based planning for autonomous driving with flexible guidance, in *Proc. 13th Int. Conf. Learn. Represent. (ICLR)*, Singapore, 2025, [Online]. Available: https://openreview.net/forum?id=wM2sfVgMDH.

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, in *Adv. Neural Inf. Process. Syst.*, vol. 30, Long Beach, CA, USA, 2017, pp. 5998–6008.

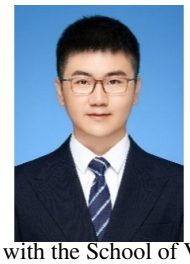
[26] D. L. Rong, Y. F. Wu, W. J. Du, C. C. Yang, S. Jin, M. Xu, and F. J. Wang, Smart prediction-planning algorithm for con-

Just Accepted

nected and autonomous vehicle based on social value orientation, *J. Intell. Connect. Veh.*, vol. 8, no. 1, Art. no. 9210053, 2025.

[27] Z. J. Wang, G. Q. Lu, J. H. Wang, H. T. Tan, and R. J. Tang, Trajectory prediction of human-driven vehicles on the basis of risk field theory and interaction multiple models, *J. Intell. Connect. Veh.*, vol. 8, no. 1, Art. no. 9210052, 2025.

Author biography



Ying Liu received the B.E. degree in vehicle and mechanical engineering from Hefei University of Technology, Hefei, China, in 2016, and the Ph.D. degree in cyber science and engineering from Southeast University, Nanjing, China, in 2023. From Apr. 2024 to Apr. 2026, He was a postdoctoral researcher with the School of Vehicle and Mobility, Tsinghua University, Beijing, China. He is currently an Assistant Professor with the College of Engineering, Nanjing Agricultural University. His current research interests include intelligent and connected vehicles, multi-vehicle cooperative decision making and control, vehicle dynamics and control.



Yi Liu received the B.E. degree in automation engineering from Nanjing University of Aeronautics and Astronautics, in 2019, a M.S. in electrical and computer engineering from the University of Illinois, in 2021, and now pursuing a joint Ph.D. degree with the South China University of Technology and Tsinghua University. He also worked as an assistant engineer with School of Vehicle and Mobility, Tsinghua University, from 2021 to 2023. His study mainly focuses on intelligent vehicle dynamic planning, scenario generation and risk quantification.



Xuewu Lin received the B.E. degree from the Beijing University of Technology, Beijing, China, in 2018, and the M.S. degree from Tsinghua University, Beijing, China, in 2021. He is currently an Engineer with Horizon Information Technology Company Ltd., Beijing. His research interests include object detection, multiobject tracking, and trajectory prediction.



Chaoyi Chen received the B.E. degree from Tsinghua University, Beijing, China, in 2016, the M.S. from Tsinghua University, Beijing, China and RWTH Aachen University, Aachen, Germany in 2019, and the Ph.D. degrees from Tsinghua University, Beijing, China in 2023. He is currently a Assistant Research Fellow with the School of Vehicle and Mobility, Tsinghua University. He was a recipient of the Shuimu Scholarship, the Scholarship of Strategic Partnership RWTH Aachen University and Tsinghua University. His research interests include vehicular network, control theory and cooperative control.



Fang Zhang received the B.E. degree in automotive engineering from Tsinghua University, Beijing, China, in 2012, and the Ph.D. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2018. From 2018 to 2020, he was a postdoctoral researcher with the School of Vehicle and Mobility, Tsinghua University, China. He is currently an Associate Researcher with State Key Laboratory of Intelligent Green Vehicle and Mobility, Tsinghua University. His current research interests include intelligent and connected vehicles, e2e autonomous driving, V2X cooperative driving.



Jianqiang Wang received the B. Tech. and M.S. degrees from Jilin University of Technology, Changchun, China, in 1994 and 1997, respectively, and the Ph.D. degree from Jilin University, Changchun, in 2002. He is currently a Professor of the School of Vehicle and Mobility, Tsinghua University, Beijing, China. His active research interests include intelligent vehicles, driving assistance systems, and driver behavior.

Just Accepted