AI Open 6 (2025) 118–129

Contents lists available at ScienceDirect

AI Open

journal homepage: www.elsevier.com/locate/aiopen

SafeCast: Risk-responsive motion forecasting for autonomous vehicles

Haicheng Liao [a,1], Hanlin Kong [b,1], Zhenning Li [c,*], Chengzhong Xu [a]

[a] State Key Laboratory of Internet of Things for Smart City and Department of Computer and Information Science, University of Macau, Macao Special Administrative Region of China

[b] Department of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China

[c] State Key Laboratory of Internet of Things for Smart City and Departments of Civil and Environmental Engineering and Computer and Information Science, University of Macau, Macao Special Administrative Region of China

ARTICLE INFO

Keywords:
Autonomous driving
Motion forecasting
Responsibility-Sensitive Safety
Interaction understanding

ABSTRACT

Accurate motion forecasting is essential for the safety and reliability of autonomous driving (AD) systems. While existing methods have made significant progress, they often overlook explicit safety constraints and struggle to capture the complex interactions among traffic agents, environmental factors, and motion dynamics. To address these challenges, we present SafeCast, a risk-responsive motion forecasting model that integrates safety-aware decision-making with uncertainty-aware adaptability. SafeCast is the first to incorporate the Responsibility-Sensitive Safety (RSS) framework into motion forecasting, encoding interpretable safety rules — such as safe distances and collision avoidance — based on traffic norms and physical principles. To further enhance robustness, we introduce the Graph Uncertainty Feature (GUF), a graph-based module that injects learnable noise into Graph Attention Networks, capturing real-world uncertainties and enhancing generalization across diverse scenarios. We evaluate SafeCast on four real-world benchmark datasets — Next Generation Simulation (NGSIM), Highway Drone (HighD), ApolloScape, and the Macao Connected Autonomous Driving (MoCAD) — covering highway, urban, and mixed-autonomy traffic environments. Our model achieves state-of-the-art (SOTA) accuracy while maintaining a lightweight architecture and low inference latency, underscoring its potential for real-time deployment in safety-critical AD systems.

1. Introduction

The field of autonomous driving (AD) is undergoing rapid advancements, confronting the complex challenge of accurately forecasting vehicle motion. This task, extending beyond mere path plotting, involves intricate decision-making influenced by environmental conditions and a steadfast commitment to safety (Geng et al., 2023). Understanding these variables is essential for developing models that can navigate the dynamic nature of real-world driving.

Existing approaches (Chen et al., 2023) to vehicle motion forecasting have predominantly hinged on the analysis of historical states, leveraging observed patterns to extrapolate future positions. However, these methodologies (Chen et al., 2022b) often prove inadequate for incorporating safety considerations, which are of paramount importance in real-world traffic environments. A common limitation is their inability to explicitly model safety metrics, such as maintaining appropriate longitudinal and lateral distances, which are essential for collision avoidance. For instance, many models lack mechanisms to enforce collision-avoidance principles like safe longitudinal and lateral distances or fail to reconcile safety constraints with the complex interplay of agent intentions, kinematics, and environmental variables. Consequently, these methods risk generating predictions that are statistically plausible yet physically unsafe, particularly in edge cases such as abrupt lane changes, intersections with limited visibility, or high-density traffic scenarios. Moreover, these models frequently fail to account for the intricate interrelationships between safety constraints, motion dynamics, and environmental variables, thereby compromising their resilience and applicability in uncertain or high-risk scenarios.

Recognizing the critical need for enhanced safety and adaptability in forecasting, we introduce **SafeCast**, a risk-responsive motion forecasting framework that jointly considers safety-aware decision-making and environment-driven uncertainty. As shown in Fig. 1, our key insight is to model vehicle motion not only as a physical prediction problem but also as a structured decision-making process, where each motion trajectory reflects rational and lawful driving behavior influenced by traffic rules and situational risks.

Central to the SafeCast model is the integration of the Responsibility-Sensitive Safety (RSS) framework (Shalev-Shwartz et al., 2017), an

* Corresponding author.
E-mail address: zhenningli@um.edu.mo (Z. Li).
[1] Equally Contributed.

https://doi.org/10.1016/j.aiopen.2025.08.001
Received 7 February 2024; Received in revised form 7 August 2025; Accepted 18 August 2025
Available online 27 August 2025
2666-6510/© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

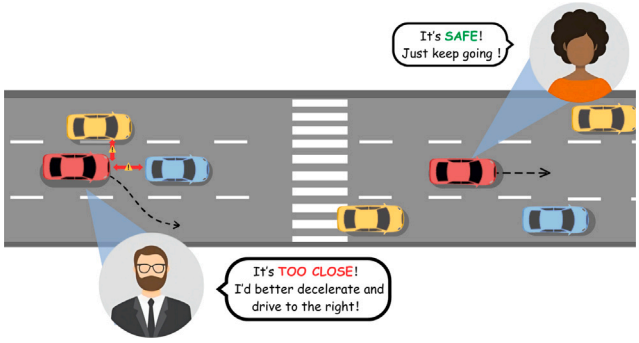


Fig. 1. Illustrative representation of safety considerations impacting the decision-making process and subsequent motion of traffic agents.

innovative approach to autonomous driving safety proposed by Mobileye. The RSS framework offers a structured methodology to ensure safety through a series of formalized rules and logical constructs. These rules are derived from established driving principles and physics, covering crucial elements like maintaining safe distances, adhering to the right of way, managing limited visibility conditions, and responding effectively in potential collision scenarios. Integrating RSS enables our model to interpret vehicle motion as a sequence of safe, deliberate choices made by the driver, deeply embedding a safety-first philosophy into the core of motion forecasting. By embedding RSS as prior domain knowledge, SafeCast significantly enhances the safety robustness of motion forecasting.

To further enhance adaptability and robustness, SafeCast incorporates the Graph Uncertainty Feature (GUF), inspired by the diffusion model (Ho et al., 2020) and Graph Prompt Feature (Fang et al., 2023). The GUF infuses variable-learnable noise into the Graph Attention Network, markedly boosting the model's ability to adjust to the diverse and unpredictable nature of real-world traffic. This feature is vital in addressing common challenges in motion forecasting, such as overfitting and representing complex driving environments accurately. Overall, the primary contributions of this study are summarized as follows:

- We propose SafeCast, the first motion forecasting framework to integrate the RSS safety model, ensuring that predictions are in line with realistic driving behaviors and stringent safety standards.
- We introduce GUF, a graph-based uncertainty modeling technique that enhances the model's robustness by simulating agent-level variability through learnable noise injection, leading to better generalization and more precise predictions in intricate driving scenes.
- We evaluate SafeCast on four datasets — NGSIM, HighD, ApolloScape, and MoCAD — covering a wide range of highway and urban driving scenarios. SafeCast achieves state-of-the-art performance in accuracy and robustness while maintaining low computational overhead, demonstrating its robustness, efficiency, and practical applicability in AD systems.

This paper is organized as follows: Section 2 reviews related work. Section 3 presents the problem formulation and SafeCast architecture. Section 4 reports experimental results and analysis. Section 5 concludes with a summary and implications for autonomous vehicle (AV) technology.

2. Related work

2.1. Motion forecasting for autonomous vehicles

A notable benchmark in motion forecasting is the CS-LSTM model (Deo and Trivedi, 2018a), which introduces a social grid framework to capture spatiotemporal interactions among traffic agents. This approach exemplifies the effective use of CNNs (Wang et al., 2023) and RNNs in handling multimodal trajectory prediction. Building upon this foundation, subsequent research has explored advanced architectures to improve prediction accuracy and contextual reasoning. These include Generative Adversarial Networks (GANs) (Zhao et al., 2019; Zhou et al., 2023), attention-based models (Wang et al., 2024; Chen et al., 2024; Zhang and Li, 2022; Mo et al., 2024), and transformer-based frameworks (Zhou et al., 2022; Li et al., 2023; Huang et al., 2022). For instance, models such as STDAN (Chen et al., 2022b) and GAVA (Liao et al., 2024c) leverage spatio-temporal attention to model interactions effectively. Similarly, FHIF (Zuo et al., 2023) and MHA-LSTM (Messaoud et al., 2020) employ multi-head attention to improve temporal reasoning. Recent innovations include DACR-AMTP (Cong et al., 2024), which utilizes a drivable area class graph to enhance spatial constraints, and iNATran (Cheng et al., 2023), a transformer-based model tailored for interpretable forecasting. BAT (Liao et al., 2024b) incorporates human psychological factors and decision-making uncertainty, while HLTP (Liao et al., 2024a) leverages human visual perception for dynamic scene modeling. Graph-based models such as AI-TP (Zhang et al., 2022) balance computational efficiency with accuracy by explicitly modeling agent interactions. Furthermore, recent works have begun to explore the integration of Large Language Models (LLMs) (Liao et al., 2025) to improve scene understanding and reasoning capabilities in trajectory prediction. Despite these advancements, many deep learning-based methods remain limited by their lack of interpretability and high computational cost. Incorporating domain knowledge and physics-informed constraints offers a promising direction for enhancing transparency, safety awareness, and computational efficiency in motion forecasting for AD systems.

2.2. Safety metrics in motion forecasting

Safety is a foundational requirement in motion forecasting for autonomous driving, necessitating the integration of robust and interpretable safety metrics. Among the most widely recognized are the Responsibility-Sensitive Safety metric (Shalev-Shwartz et al., 2017) and the Safe Force Field (SFF) model (Nistér et al., 2019). RSS provides a mathematically grounded, rule-based framework rooted in kinematic principles, offering explicit guidelines for maintaining safe following distances, lateral spacing, and right-of-way. This level of interpretability ensures a transparent evaluation of safety in diverse traffic scenarios and supports formal compliance with safety standards. In contrast, the SFF model employs a dynamic force field to assess safety, focusing primarily on maintaining safe distances between traffic agents. While both metrics aim to enhance safety, the RSS metric's rigorously defined, physics-based principles make it particularly attractive for ensuring compliance with established safety standards. Its interpretability and clarity offer a distinct advantage, particularly when integrating safety into motion forecasting models for autonomous systems. Recent studies, such as DACR-AMTP (Cong et al., 2024), and CF-LSTM (Xie et al., 2021) also incorporate safety considerations by modeling drivable areas, congestion patterns, or interaction intensity (Huang et al., 2024). However, these approaches often lack the explicitness and generalizability of RSS. Its ability to quantify safety through interpretable kinematic constraints continues to make it a compelling choice for designing safe, collision-free behavior for AVs.

3. Methodology

3.1. Problem formulation

The primary goal of our research is to forecast the motion of the *target vehicle*, encompassing all traffic agents within the sensing range of the AV in a mixed autonomy environment. Given the observed historical state $\boldsymbol{X}_{0:n}^{t-t_h:t}$ of the target vehicle (denoted as superscript 0)

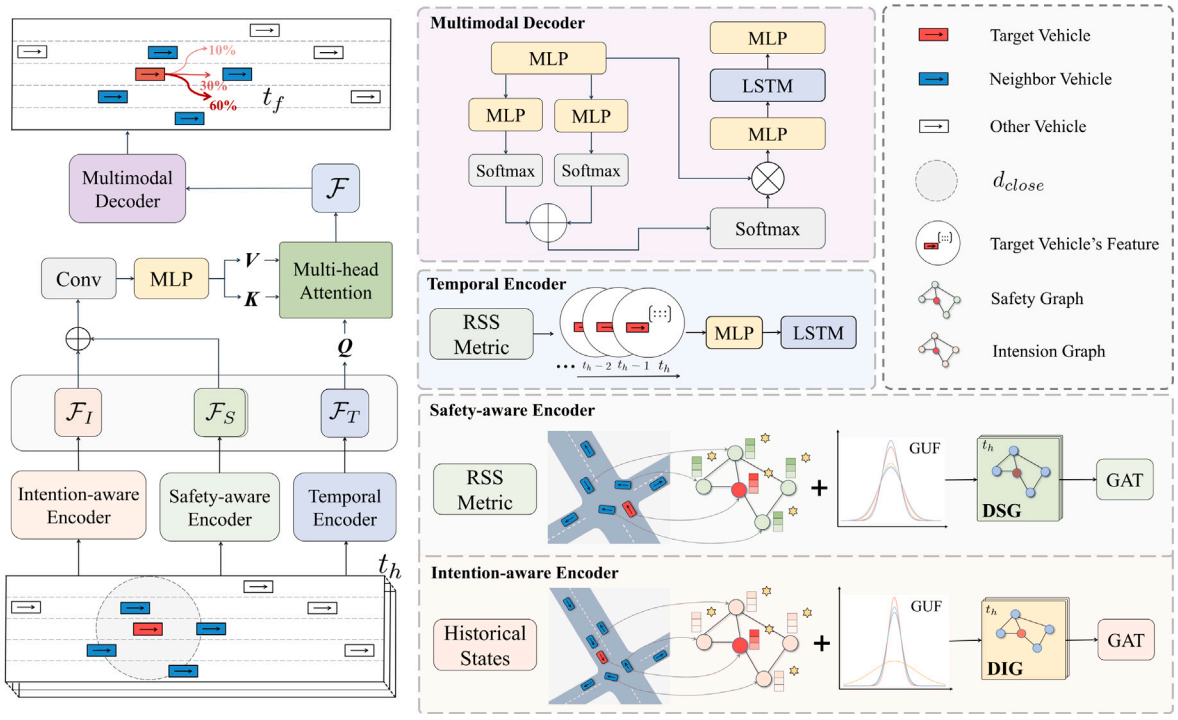


Fig. 2. An overview of the architecture of our proposed model, which includes the intention-aware module, the safety-aware module, the temporal encoder, and the multimodal decoder. The overall process of SafeCast is depicted from bottom to top.

and its surrounding traffic agents (superscripts from 1 to n) over the time interval from $t-t_h$ to t, the task is to predict the future motion of the target vehicle $\boldsymbol{Y}$ over a specified prediction horizon t_h. Moreover, at time step t, the observed historical state $\boldsymbol{X}_{0:n}^{t-t_h:t}$ comprises a series of vectors spanning the past t_h time steps. These vectors encompass position coordinates $\{\boldsymbol{p}_{0:n}^{t-t_h:t}\}$, instant velocity $\{\boldsymbol{v}_{0:n}^{t-t_h:t}\}$, acceleration $\{\boldsymbol{a}_{0:n}^{t-t_h:t}\}$ and vehicle type $\{\boldsymbol{u}_{0:n}^{t-t_h:t}\}$. All these vectors form the input for our proposed model, providing a comprehensive data foundation for motion forecasting.

3.2. Overview

The overall architecture of SafeCast is depicted in Fig. 2. The model adopts an encoder–decoder framework comprising four main components: the **intention-aware module**, **safety-aware module**, **feature fusion module**, and **multimodal decoder**. Each component is designed to capture distinct yet complementary aspects of traffic dynamics, contributing to accurate and efficient motion forecasting for autonomous vehicles. Unlike deterministic forecasting approaches, SafeCast formulates motion prediction as a probabilistic estimation task. Given the historical trajectories of the ego vehicle and surrounding agents, denoted as $\boldsymbol{X}_{0:n}^{t-t_h:t}$, the model estimates a conditional probability distribution over future trajectories, expressed as $P(\boldsymbol{Y}|\boldsymbol{X}_{0:n}^{t-t_h:t})$. This formulation allows SafeCast to account for inherent uncertainty and behavioral variability, producing a multimodal output that captures diverse future maneuvers. A detailed overview of each module is presented in the following section.

3.3. Intention-aware module

This module is designed to tackle the challenges of motion forecasting in highly interactive and dynamic traffic environments. Its core objective is to effectively capture spatiotemporal dependencies and model the uncertainty inherent in agent interactions while maintaining robustness and generalizability to previously unseen scenarios. To this end, we propose two key components: the Dynamic Intention Graph (DIG) and the Uncertainty-aware Graph Attention Network (Uncertainty-aware GAT). Together, these modules empower the SafeCast model to reason over complex agent behaviors and adapt to diverse traffic conditions, significantly improving its performance in real-world applications.

3.3.1. Dynamic intention graph

The DIG is constructed to represent a traffic scene at a specific time step, encompassing the real-time states of the target vehicle and its surrounding agents to represent a traffic scene at a specific time step. At a given time t_k, the DIG, denoted as $\mathcal{G}_{DIG}^{t_k}$ can be formulated:

$$\mathcal{G}_{DIG}^{t_k} = \{\mathcal{V}^{t_k}, \mathcal{E}^{t_k}\} \tag{1}$$

where $\mathcal{V}^{t_k} = \{v_i^{t_k}, i \in [0,n]\}$ is the node set. Each node $v_i^{t_k}$ corresponds to a traffic agent and is characterized by its historical states $\boldsymbol{X}_i^{t_k} = \{\boldsymbol{p}_i^{t_k}, \boldsymbol{v}_i^{t_k}, \boldsymbol{a}_i^{t_k}, \boldsymbol{u}_i^{t_k}\}$. Formally, the edge set $\mathcal{E}^{t_k} = \{e_0^{t_k}, e_1^{t_k} \ldots, e_n^{t_k}\}$ represents the potential interactions between agents, where each edge $e_i^{t_k} = \{v_i^{t_k} v_j^{t_k} \mid (i,j) \in \mathcal{D}^{t_k}, i \neq j\}$ connects the node $v_i^{t_k}$ with other agents that have potential influences with it. Moreover, $\mathcal{D}^{t_k}$ is the set of agents in proximity at time t_k, defined by a shortest distance $d\left(v_i^t, v_j^t\right)$ less than or equal to a pre-established threshold d_{close}. Additionally, the adjacency matrix $\mathcal{A}^{t_k} \in \mathbb{R}^{n\times n}$ for the DIG $\mathcal{G}_{DIG}^{t_k}$ at time t_k can be defined as follows:

$$\mathcal{A}^{t_k}(i,j) = \begin{cases} 1 & \text{if } d\left(v_i^{t_k}, v_j^{t_k}\right) \leq d_{close}, \\ 0 & \text{otherwise} \end{cases} \quad i \neq j,\ i,j \in [0,n] \tag{2}$$

Next, the DIG $\mathcal{G}_{DIG}^{t-t_h:t}$ is fed into our proposed uncertainty-aware GAT to capture the dynamic interactions and influences among different agents in a traffic scene, resulting in the intention-aware features $\mathcal{F}_I$.

3.3.2. Uncertainty-aware graph attention network

We propose a novel framework, the Uncertainty-aware GAT, as shown in Fig. 3, to account for the variables and uncertainties in real-world driving scenes. Traditional graph attention mechanisms assume

deterministic relationships, which limits their ability to generalize in dynamic and unpredictable environments. To overcome this, we integrate a GUF mechanism into the GAT framework. GUF introduces variance-learnable Gaussian noise and a shared linear transformation to realistically simulate unpredictable elements in driving, such as human behavior and potential vehicle maneuvers. This can be represented as follows:

$$\begin{cases} \mathcal{P} = \{p_1, p_2, \dots, p_F\}, \quad p_i \sim \mathcal{N}(\mu, \sigma_i^2) \\ h_i^{t_k} = \mathbf{W}(v_i^{t_k} + \mathcal{P}) \end{cases} \tag{3}$$

Here, $\mathcal{P}$ represents a set of Gaussian noise components, each following a distribution $\mathcal{N}(\mu, \sigma_i^2)$ with mean μ and learnable standard deviation σ_i, where the weight is shared across all nodes at all time steps, and F is the feature dimension of $v_i^{t_k}$. Furthermore, the transformed node features $h_i^{t_k}$ is obtained by applying the shared weight matrix $\mathbf{W}$ to the sum of the node features $v_i^{t_k}$ and the noise components. The attention mechanism quantifies the influence between agents while integrating the uncertainty-enhanced features. The attention coefficient γ_i^j is computed as follows:

$$\gamma_i^j = \phi_{LR}\left(\mathbf{a}^\top (h_i^{t_k} \parallel h_j^{t_k})\right) \tag{4}$$

where ϕ_{LR} refers to the Leaky Rectified Linear Unit activation feature, and $\parallel$ represents the concatenation operation. We implement a masked attention mechanism for preserving vital structural information and concentrate solely on relevant neighboring nodes. The masked attention coefficients α_i^j are computed through the softmax activation feature:

$$\alpha_i^j = \phi_{Softmax}(\gamma_i^j) = \frac{e^{\gamma_i^j}}{\sum_{n \in \mathcal{N}_i} e^{\gamma_i^n}} \tag{5}$$

where $\phi_{Softmax}$ denotes the softmax activation feature, while $\mathcal{N}_i$ represents the neighbors of traffic agent v_i. To improve learning stability and model capacity, multi-head attention aggregates information from multiple perspectives, and residual connections facilitate gradient flow. Specifically, K_{GAT} GAT layers perform attention mechanisms, and their outputs are averaged to obtain the output features:

$$\hat{v}_i^{t_k} = \phi_{ELU}\left(\sum_{head=1}^{K_{GAT}} \frac{\sum_{j \in \mathcal{N}_i} \alpha_i^j \cdot h_j^{t_k}}{K_{GAT}}\right) \tag{6}$$

where ϕ_{ELU} represents the Exponential Linear Unit (ELU) activation function. The output features are processed through a Dropout layer to mitigate overfitting. Next, an additional GAT layer captures high-level interactions among traffic agents. Residual connections are integrated into the architecture to enhance gradient flow and improve training stability. Finally, the refined features are passed through a linear layer for reshaping into the desired format.

3.4. Safety-aware module

Safety is a fundamental aspect of AV motion prediction. Maintaining safe distances between traffic agents ensures reliable and collision-free motion planning, even in complex or dynamic scenarios. To achieve this, the safety-aware module integrates safety distance metrics derived from the RSS framework (as shown in Fig. 4) into the prediction task. These metrics serve as input features, providing the model with prior knowledge of safe distances that guide the prediction process. The module enhances the model's ability to generalize to diverse traffic conditions, including extreme or unseen scenarios, by embedding these physics-based principles.

3.4.1. Safety distance metric

Drawing from foundational physics and practical driving experience, novice drivers are typically advised to maintain a safe following distance to allow sufficient reaction time in case the lead vehicle decelerates unexpectedly. In autonomous driving systems, this principle is operationalized through the definition of safety-aware longitudinal and lateral distances. These metrics formalize the minimum safe spacing required to prevent collisions, enabling the AV to react appropriately to dynamic changes in the environment. This subsection defines these distances precisely (see Fig. 4).

Safe Longitudinal Distance. To ensure safety between two vehicles traveling in the same direction, the minimum safe longitudinal distance is defined as follows:

$$d_{\min}^{\text{lon}} = \left[v_r \rho + \frac{1}{2} a_{\max} \rho^2 + \frac{(v_r + \rho a_{\max})^2}{2 b_{\min}} - \frac{v_f^2}{2 b_{\max}} \right]_+ \tag{7}$$

where $d_{\min}^{\text{lon}}$ denotes the minimum required longitudinal distance to avoid a rear-end collision. The operator $[\cdot]_+$ ensures non-negativity by returning the value itself if positive or zero otherwise. v_f and v_r are the longitudinal velocities of the front and rear vehicles, respectively. ρ is the driver (or system) reaction time, $a_{\max}$ is the maximum acceleration, while $b_{\max}$ and $b_{\min}$ represent the maximum and minimum braking decelerations of the front and rear vehicles, respectively.

Safe Lateral Distance. To maintain a safe buffer between adjacent vehicles, the minimum lateral distance can be formally defined as follows:

$$\begin{cases} v_{1,\rho} = v_1 + \alpha_{\max} \rho \\ v_{2,\rho} = v_2 + \alpha_{\max} \rho \\ d_{\min}^{\text{lat}} = \mu + \left[\frac{(v_1 + v_{1,\rho})}{2} \rho + \frac{v_{1,\rho}^2}{2 \beta_{\min}} - \left(\frac{v_2 + v_{2,\rho}}{2} \rho - \frac{v_{2,\rho}^2}{2 \beta_{\min}} \right) \right]_+ \end{cases} \tag{8}$$

Here, v_1 and v_2 are the lateral velocities of the right and left vehicles, respectively, and ρ is the reaction time. $\alpha_{\max}$ denotes the maximum lateral acceleration, while $\beta_{\min}$ is the minimum lateral deceleration. μ accounts for the necessary lateral safety margin due to natural fluctuations. The operator $[\cdot]_+$ ensures non-negativity. By incorporating these safety-aware metrics directly into the model architecture, SafeCast avoids reliance on post hoc rule-based constraints. Unlike traditional motion planners that apply safety checks after prediction, SafeCast embeds interpretable physical and legal constraints into its learning process. This allows the model to make predictions that are inherently safety-compliant, improving generalization across dense traffic, unstructured road layouts, and edge-case scenarios.

The RSS framework provides a structured, interpretable foundation for motion safety. Its parameters — based on vehicular dynamics and empirical traffic standards — offer consistent safety margins across diverse driving conditions. These fixed settings reduce overfitting, simplify the learning process, and enhance transferability. Furthermore, the mathematical flexibility of RSS allows for future integration with real-time perception and adaptive control modules, enabling dynamic tuning of parameters. This positions RSS not only as a conservative safety baseline but also as a scalable safety prior for real-world autonomous driving systems.

3.4.2. Spatial encoder

To capture safety-critical spatial interactions between the target vehicle and its surrounding agents, we propose the Dynamic Safety Graph (DSG)—a novel graph structure that integrates RSS distance metrics into its representation. At each time step t_k, the DSG is defined as an undirected graph $\tilde{G}^{t_k} = \{\tilde{V}^{t_k}, \tilde{E}^{t_k}\}$, where the node set $\tilde{V}^{t_k} = \{\tilde{v}_i^{t_k} \mid i \in [0, n]\}$ represents all traffic agents. Each node $\tilde{v}_i^{t_k}$ encodes the 2D position $\boldsymbol{p}_i^{t_k}$, along with the agent's dynamically computed maximum safe longitudinal and lateral distances $d_{\max,i}^{\text{lon},t_k}$ and $d_{\max,i}^{\text{lat},t_k}$, respectively:

$$\tilde{v}_i^{t_k} = \left[\boldsymbol{p}_i^{t_k}, d_{\max,i}^{\text{lon},t_k}, d_{\max,i}^{\text{lat},t_k} \right]^\top \tag{9}$$

Correspondingly, edges $\tilde{e}_{ij}^{t_k} \in \tilde{E}^{t_k}$ are created based on spatial proximity: an edge is established between nodes i and j if the Euclidean distance between them, $d(\tilde{v}_i^{t_k}, \tilde{v}_j^{t_k})$, is below a predefined threshold $d_{\text{close}}^{\text{lon}}$. Safety distances are computed dynamically based on lane context—longitudinal safety for agents in the same lane and lateral safety for

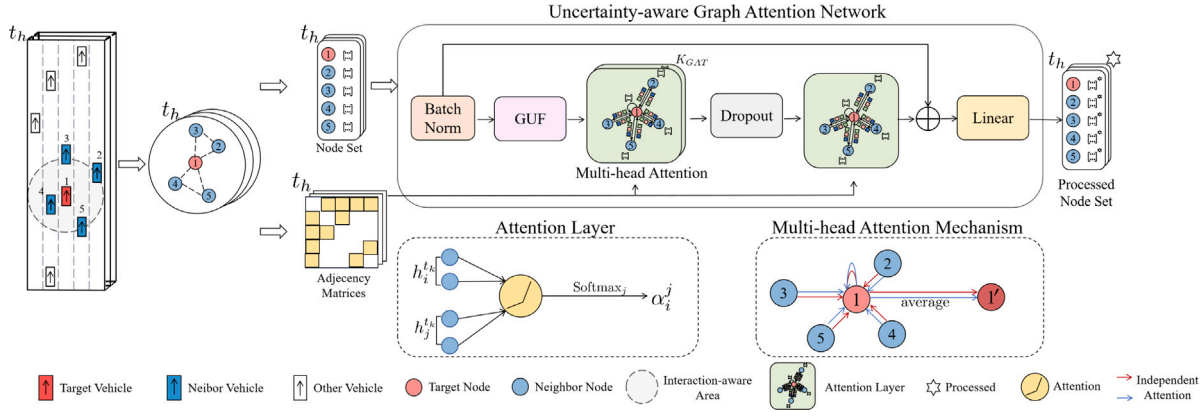


Fig. 3. Illustration of the architecture of the uncertainty-aware GAT. It consists of a batch-normalization layer, a GUF layer, K_{GAT} attention layers for multi-head attention, a dropout layer, another attention layer, and a linear layer. The GUF layer enables the network to estimate and consider the uncertainty of the node features, enhancing its robustness to noisy or incomplete data.

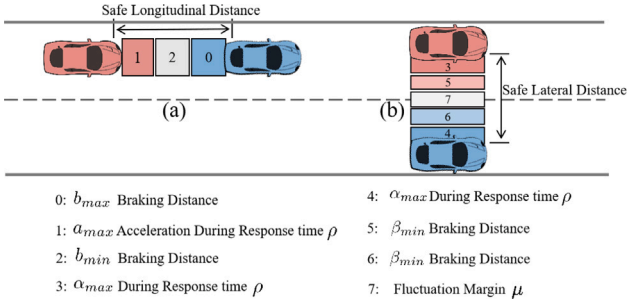
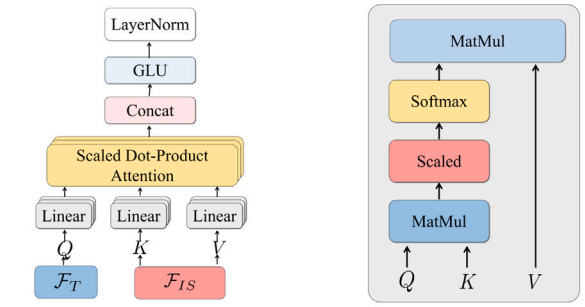


Fig. 4. Illustration of safe longitudinal and lateral distance metrics in the RSS framework: (a) Safe longitudinal distance, accounting for interactions within the same lane. (b) Safe lateral distance, addressing interactions in adjacent lanes.



(a) Multi-head Attention Mechanism (b) Scaled Dot-Product Attention

Fig. 5. Structure of the multi-head attention mechanism: (a) Integration of features $\mathcal{F}_T$ and $\mathcal{F}_{IS}$ through scaled dot-product attention. (b) Illustrates of weight computation.

adjacent-lane interactions. To effectively capture spatial dependencies under uncertainty, we apply the Uncertainty-aware GAT, as introduced in the intention-aware module. This enables the DSG to model both the physical configuration and the safety margin around each agent,

3.4.3. Temporal encoder

To capture the temporal dependencies of the target vehicle, this encoder individually encodes pertinent features of the target vehicle, encompassing both the historical state and the safety distance metric. It applies a Multi-Layer Perceptron-ELU-LSTM structure. Specifically, the historical state of the target vehicle $\mathcal{T}_0 = \{d^{lon,t_k}_{max,i}, d^{lat,t_k}_{max,i}, \boldsymbol{X}_0^{t-t_h:t}\} \in \mathbb{R}^{t_h \times m}$, where m is the feature dimension. We utilize the MLP framework to extract and amplify features with an ELU activation function, followed by the LSTM network to capture feature variations over the time horizon t_h.

$$\mathcal{F}_T = \phi_{LSTM}\left(\phi_{ELU}\left(\phi_{MLP}(\mathcal{T}_0^*, W_m), W_l\right)\right) \tag{10}$$

Here, W_m and W_l represent the learnable weight of the MLP and LSTM, respectively. Moreover, the temporal encoder outputs the temporal feature of the target vehicle, denoted as the $\mathcal{F}_T \in \mathbb{R}^{\mathcal{B} \times t_h \times C}$, where $\mathcal{B}$ is the batch size and C is the hyperparametric dimension for the LSTM network.

3.5. Feature fusion module

The feature fusion module combines the outputs of the intention-aware $\mathcal{F}_I$ and the safety-aware $\mathcal{F}_S$ features using a multi-head scaled dot-product attention mechanism (Fig. 5). The fused high-level feature $\mathcal{F}$ captures contextual and safety-critical information for downstream motion forecasting. Formally, the features $\mathcal{F}_I$ and $\mathcal{F}_S$ are concatenated and passed through a convolutional layer, ReLU activation, and an MLP for initial processing:

$$\mathcal{F}_{IS} = \phi_{MLP}\left(\phi_{ReLU}(\phi_{Conv2D}((\mathcal{F}_I \parallel \mathcal{F}_S)))\right) \tag{11}$$

To enhance interaction modeling, the multi-head attention mechanism projects $\mathcal{F}_T$ and $\mathcal{F}_{IS}$ into query Q, key K, and value V vectors. Moreover, the scaled dot-product operation aggregates these projections across multiple heads to compute the attention output H:

$$\begin{cases} H = \sum_{i=1}^{h} \alpha^i V^i \\ \alpha^i = \phi_{softmax}\left(\dfrac{Q^i K^i}{\sqrt{d_K}}\right) \end{cases} \tag{12}$$

Finally, the fused feature $\mathcal{F}$ undergoes further refinement via a Gated Linear Unit (GLU) and Layer Normalization (LN), ensuring efficient and stable feature extraction for multimodal decoding. Formally,

$$\mathcal{F} = \phi_{LN}\left(\phi_{GLU}(H, W_g)\right) \tag{13}$$

3.6. Multimodal decoder

The multimodal decoder predicts the future motion of the target vehicle, accounting for uncertainty and multimodal behavior. The decoder employs a Gaussian Mixture Model (GMM) to generate lateral

and longitudinal maneuver-based motion in a hierarchical framework:

$$\begin{cases} M = \{M_{lat}, M_{lon}\} \\ M_{lat} = \{m_S, m_R, m_L\} \\ M_{lon} = \{m_A, m_D, m_C\} \end{cases} \tag{14}$$

where M_{lat} includes lateral maneuvers (straight, right turn, left turn) and M_{lon} includes longitudinal maneuvers (acceleration, deceleration, constant speed). The probability distribution of the predicted motion $P(Y|X_{0:n}^{t-t_h:t})$ is computed:

$$P(Y|X_{0:n}^{t-t_h:t}) = P_\epsilon(Y|X_{0:n}^{t-t_h:t}, M)P(M|X_{0:n}^{t-t_h:t}) \tag{15}$$

Here, ϵ denotes the parameters of the bivariate Gaussian distribution associated with the future motion of the target vehicle. It is obtained from the function f_{dec} as follows:

$$\epsilon = f_{dec}(\mathcal{F} \| P(M_{lat}|X_{0:n}^{t-t_h:t}) \| P(M_{lon}|X_{0:n}^{t-t_h:t})) \tag{16}$$

Notably, the function f_{dec} is composed of two MLPs and one LSTM, which can be represented as follows:

$$f_{dec}(\cdot) = \phi_{MLP}(\phi_{LSTM}[\phi_{MLP}(\cdot)]) \tag{17}$$

4. Experiments

4.1. Experiment setup

Datasets. We evaluate the effectiveness of our model on four real-world benchmarks that span diverse traffic complexities and environments, including highways, campuses, and urban streets: the Next Generation Simulation (NGSIM) dataset, the Highway Drone Dataset (HighD), ApolloScape, and the Macao Connected Autonomous Driving (MoCAD) dataset. These datasets provide time-stamped trajectories of traffic participants — including vehicles, pedestrians, and cyclists — represented in a local coordinate frame. Each dataset is partitioned into training, validation, and test sets to ensure experimental consistency and alignment with prior studies.

Temporal Configuration. For NGSIM, HighD, and MoCAD, trajectories are uniformly sampled over 8-s sequences. The initial 3 s ($t_h = 3$) serve as input history, while the following 5 s ($t_f = 5$) form the prediction horizon. This configuration enables robust evaluation in both highway and urban settings. Correspondingly, the ApolloScape benchmark follows the original challenge protocol: Predictions are made over a 3-s future window ($t_f = 3$) given a 3-s observation period ($t_h = 3$). This dataset introduces additional complexity with its inclusion of multiple agent types and diverse scene contexts (e.g., intersections, campus roads, and mixed-use environments).

Evaluation Metrics. For NGSIM, HighD, and MoCAD, we report Root Mean Square Error (RMSE) to quantify trajectory accuracy over time. In ApolloScape, following official standard protocol, we adopt Average Displacement Error (ADE) and Final Displacement Error (FDE) as evaluation metrics. To account for heterogeneous agent types, we compute a weighted sum of ADE and FDE across vehicles, pedestrians, and bicycles. Formally,

$$\begin{aligned} \text{WSADE} &= D_v \cdot \text{ADE}_v + D_p \cdot \text{ADE}_p + D_b \cdot \text{ADE}_b, \\ \text{WSFDE} &= D_v \cdot \text{FDE}_v + D_p \cdot \text{FDE}_p + D_b \cdot \text{FDE}_b, \end{aligned} \tag{18}$$

where $D_v = 0.20$, $D_p = 0.58$, and $D_b = 0.22$ represent the relative weights assigned to vehicles, pedestrians, and bicycles, respectively. These metrics allow for a comprehensive and fair comparison across heterogeneous traffic agents and ensure that model performance reflects safety-critical forecasting across all modalities.

4.2. Implementation and training details

4.2.1. Training details

We conduct the coding and experiments on a device equipped with an NVIDIA GeForce 3090 GPU. We train our model using the Adam optimizer and adopt the CosineAnnealingWarmRestarts strategy for dynamic learning rate scheduling. The initial learning rate is set to $lr = 0.001$, with a batch size of $B = 128$. During training, each input segment is split with a factor of 2 to increase temporal resolution. To reduce the adverse impact of misclassifying maneuver types on motion forecasting accuracy and robustness, we augment the MSE criterion with the NLL criterion in the loss function. Specifically, the proposed loss function $\mathcal{L}$ combines two components: Mean Squared Error (MSE) loss $\mathcal{L}_{MSE}$ and Negative Log-Likelihood (NLL) loss $\mathcal{L}_{NLL}$. These loss functions collectively guide the training process, ensuring that the model accurately forecasts motion and aligns with real-world driving maneuvers.

4.2.2. Data-driven RSS parameter examination

To enhance safety awareness, we embed environment-specific parameters from the RSS framework directly into the training pipeline, distinguishing between highway and urban scenarios. For highways, we use a lateral fluctuation margin of 1.0 m and a reaction time of $\rho = 0.8$ s, while urban settings adopt tighter constraints, including a reduced safety buffer of 2.5 m. The safe longitudinal proximity threshold is set to $d_{close}^{lon} = 2.0$ m. Importantly, acceleration and deceleration capabilities — both longitudinal and lateral — vary across different traffic contexts due to road surface conditions (e.g., dry, wet), weather (e.g., clear, rainy), vehicle properties (e.g., mass, drivetrain), driving styles (e.g., aggressive, conservative), time of day, and scene type (e.g., urban, highway). Rather than using static thresholds (Naumann et al., 2021), we employ a data-driven parameter estimation strategy. Specifically, RSS-relevant parameters such as maximum acceleration, braking limits, and lateral maneuverability are dynamically inferred from historical trajectories of traffic agents. This approach enables the model to adaptively capture scene-specific safety constraints grounded in real-world driving behavior, improving realism and reducing overconservatism. By learning from empirical patterns, the model can better reflect varying traffic densities, agent types, and risk tolerances. Then, these safety-aware parameters are incorporated via the DSG, ensuring physics-consistent safety modeling across diverse and complex driving scenarios.

4.2.3. Implementation details

We provide further details on the implementation and important parameter settings as follows:

Intention-aware Module: This module constructs the Dynamic Interaction Graph $\mathcal{G}_{DIG}^{t-t_h:t}$, following the edge construction strategy of GRIP (Li et al., 2019) with an interaction radius $d_{close} = 25$. It is then processed by an uncertainty-aware GAT, where the number of attention heads K_{GAT} is set to 2, and the mean variance of the learnable Gaussian noise, introduced via the GUF, is initialized to 0.

Safety-aware Module: This module consists of a spatial encoder and a temporal encoder. The spatial encoder builds the DSG $\mathcal{G}_{DSG}^{t-t_h:t}$ based on RSS metrics and processes it using the same uncertainty-aware GAT. The temporal encoder employs an LSTM with a hidden size of $C = 64$, capturing temporal dependencies in the safety-aware features.

Feature Fusion Module: This module fuses three feature types: $\mathcal{F}_I$ (intention-aware), $\mathcal{F}_S$ (safety-aware), and $\mathcal{F}_T$ (temporal). Feature integration employs a Conv2D operation with a kernel size of 3 and padding of 1, as described in Eq. (11), along with a multi-head dot-product attention mechanism comprising 4 heads.

Multi-modal Decoder: This component introduces maneuver-aware multimodality to support diverse motion outcomes. It uses an LSTM

Table 1
Evaluation results in the ApolloScape dataset. $ADE_{v/p/b}$ and $FDE_{v/p/b}$ are the ADE and FDE metrics for the vehicles, pedestrians, and bicycles, respectively. **Bold** and underlined values represent the best and second-best performance in each category.

Method	Publication	Backbone	WSADE	ADEv	ADEp	ADEb	WSFDE	FDEv	FDEp	FDEb
TrafficPredict (Ma et al., 2019)	AAAI'19	LSTM	8.5881	7.9467	7.1811	12.8805	24.2262	12.7757	11.1210	22.7912
TPNet (Fang et al., 2020)	CVPR'20	CNN	1.2800	2.2100	0.7400	1.8500	2.3400	3.8600	1.4100	3.4000
S2TNet (Chen et al., 2021)	ACML'21	Transformer	1.1679	1.9874	0.6834	1.7000	2.1798	3.5783	1.3048	3.2151
MSTG (Tang et al., 2023)	IET-ITS'23	LSTM	1.1546	1.9850	0.6710	1.6745	2.1281	3.5842	1.2652	3.0792
TP-EGT (Yang et al., 2024)	IEEE-TITS'24	Transformer	1.1900	2.0500	0.7000	1.7200	2.1400	3.5300	1.2800	3.1600
SafeCast	–	GAT + GUF	**1.1253**	**1.9372**	**0.6561**	**1.6247**	**2.1024**	**3.5061**	**1.2524**	**3.0657**

Table 2
Evaluation results for our proposed model and other SOTA baselines on the NGSIM, HighD, and MoCAD datasets across various prediction horizons. The evaluation metric is RMSE (m), where lower values indicate better performance. Some results are unspecified ("–"). **Bold** values denote the best performance, and underlined values indicate the second-best performance in each category. "AVG" represents the average RMSE across all horizons.

Dataset	Model	Prediction horizon (s)					
		1	2	3	4	5	AVG
NGSIM	S-LSTM (Alahi et al., 2016)	0.65	1.31	2.16	3.25	4.55	2.38
	M-LSTM (Deo and Trivedi, 2018b)	0.58	1.26	2.12	3.24	4.66	2.37
	CS-LSTM (Deo and Trivedi, 2018a)	0.61	1.27	2.09	3.10	4.37	2.29
	MATF-GAN (Zhao et al., 2019)	0.66	1.34	2.08	2.97	4.13	2.24
	DRBP (Gao et al., 2023)	1.18	2.83	4.22	5.82	–	3.51
	MFP (Tang and Salakhutdinov, 2019)	0.54	1.16	1.89	2.75	3.78	2.02
	NLS-LSTM (Messaoud et al., 2019)	0.56	1.22	2.02	3.03	4.30	2.23
	MHA-LSTM (Messaoud et al., 2020)	0.41	1.01	1.74	2.67	3.83	1.93
	WSiP (Wang et al., 2023)	0.56	1.23	2.05	3.08	4.34	2.25
	CF-LSTM (Xie et al., 2021)	0.55	1.10	1.78	2.73	3.82	2.00
	STDAN (Chen et al., 2022b)	0.42	1.01	1.69	2.56	3.67	1.87
	DACR-AMTP (Cong et al., 2024)	0.57	1.07	1.68	2.53	3.40	1.85
	HTPF (Wang et al., 2024)	0.49	1.09	1.78	2.62	3.65	1.92
	BAT (Liao et al., 2024b)	**0.23**	**0.81**	1.54	2.52	3.62	1.74
	FHIF (Zuo et al., 2023)	0.40	0.98	1.66	2.52	3.63	1.84
	SafeCast	0.40	0.90	**1.42**	**2.05**	**2.89**	**1.54**
HighD	S-GAN (Gupta et al., 2018)	0.30	0.78	1.46	2.34	3.41	1.66
	CS-LSTM (Deo and Trivedi, 2018a)	0.22	0.61	1.24	2.10	3.27	1.49
	DRBP (Gao et al., 2023)	0.41	0.79	1.11	1.40	–	0.93
	NLS-LSTM (Messaoud et al., 2019)	0.20	0.57	1.14	1.90	2.91	1.34
	MHA-LSTM (Messaoud et al., 2020)	0.19	0.55	1.10	1.84	2.78	1.29
	HTPF (Wang et al., 2024)	0.16	0.47	0.94	1.58	2.36	1.10
	STDAN (Chen et al., 2022b)	0.19	0.27	0.48	0.91	1.66	0.70
	iNATran (Chen et al., 2022a)	**0.04**	**0.05**	0.21	0.54	1.10	0.39
	DACR-AMTP (Cong et al., 2024)	0.10	0.17	0.31	0.54	1.01	0.43
	GaVa (Liao et al., 2024c)	0.17	0.24	0.42	0.86	1.31	0.60
	SafeCast (small)	0.24	0.39	0.57	0.70	1.08	0.58
	SafeCast	0.06	0.10	**0.14**	**0.35**	**0.79**	**0.29**
MoCAD	CS-LSTM (Deo and Trivedi, 2018a)	1.45	1.98	2.94	3.56	4.49	2.88
	MHA-LSTM (Messaoud et al., 2020)	1.25	1.48	2.57	3.22	4.20	2.54
	NLS-LSTM (Messaoud et al., 2019)	0.96	1.27	2.08	2.86	3.93	2.22
	CF-LSTM (Xie et al., 2021)	0.72	0.91	1.73	2.59	3.44	1.88
	WSiP (Wang et al., 2023)	0.70	0.87	1.70	2.56	3.47	1.86
	STDAN (Chen et al., 2022b)	0.62	**0.85**	1.62	2.51	3.32	1.78
	SafeCast (small)	0.70	1.20	1.87	2.60	3.42	1.96
	SafeCast	**0.50**	**0.85**	**1.57**	**2.19**	**2.99**	**1.62**

with a hidden size of 64 to produce a multimodal distribution over future trajectories.

SafeCast (small): We propose a lightweight variant of our model, SafeCast (small), which omits the temporal encoder from the safety-aware module and merges the intention-aware and safety-aware components into a unified uncertainty-aware GAT. The input node set is redefined as $\mathcal{V}^{t_k} = \{v_i^{t_k} \mid t_k \in [t-t_h, t], i \in [0, n]\}$, where each node $v_i^{t_k}$ includes $\{\boldsymbol{X}_i^{t_k}, d_{max,i}^{lon,t_k}, d_{max,i}^{lat,t_k}\}$. The resulting representation $F_{IS} \in \mathbb{R}^{\mathcal{B}\times t_h\times n\times C}$ is projected to $F_T \in \mathbb{R}^{\mathcal{B}\times t_h\times C}$ by selecting the 0-th index along the agent dimension, corresponding to the ego vehicle. This compact variant retains core capabilities while significantly improving computational efficiency.

4.3. Evaluation results

4.3.1. Compare to SOTA baselines

Tables 2 and 1 summarizes the performance of SafeCast against state-of-the-art (SOTA) baselines across four benchmark datasets: NGSIM, HighD, ApolloScape, and MoCAD. The results demonstrate that SafeCast consistently achieves superior accuracy, outperforming existing methods in both short-term and long-term motion forecasting tasks.

On the NGSIM dataset, SafeCast delivers substantial gains over earlier baselines (2017–2020), with improvements of at least 10.89% for short-term and 23.14% for long-term predictions. Compared to more recent methods (2021–2024), it maintains a strong lead, reducing 5-s RMSE by a minimum of 0.51 m and achieving at least 15.00% improvement in long-horizon accuracy. These enhancements are particularly significant for reducing accident risk in complex, high-speed driving scenarios. Correspondingly, in the HighD dataset, although short-term predictions are competitive across models, SafeCast shows marked improvements in long-term accuracy, with a minimum RMSE reduction

Table 3
Comparative evaluation of our proposed models with selected baselines. Highlighting the prediction accuracy metric (Average RMSE) and model complexity measured by the number of parameters (#Param.). **Bold** and underlined values represent the best and second-best performance.

Model	Average RMSE (m)			#Param. (K)
	NGSIM	HighD	MoCAD	
CS-LSTM (Deo and Trivedi, 2018a)	2.29	1.49	2.88	194.92
CF-LSTM (Xie et al., 2021)	1.99	1.17	1.88	387.10
WSiP (Wang et al., 2023)	2.25	1.44	1.86	300.76
STDAN (Chen et al., 2022b)	1.87	0.70	1.78	486.22
SafeCast (small)	1.64	0.58	1.96	**135.83**
SafeCast	**1.54**	**0.29**	**1.62**	234.55

Table 4
Performance comparison of proposed models with the SOTA baselines on NGSIM dataset. The batch size for evaluation is set to 128. A total of 10 batches are predicted, and the average inference time (s) and average RMSE (m) of the predictions are calculated for each baseline.

Model	Average RMSE (m)	Inference time (s)
M-LSTM (Deo and Trivedi, 2018b)	2.37	**0.14**
S-LSTM (Alahi et al., 2016)	2.38	0.30
CS-LSTM (Deo and Trivedi, 2018a)	2.29	0.37
MHA-LSTM (Messaoud et al., 2020)	1.93	0.23
TS-GAN (Wang et al., 2020)	2.06	0.39
DACR-AMTP (Cong et al., 2024)	1.85	0.26
STDAN (Chen et al., 2022b)	1.87	0.21
SafeCast (small)	1.70	0.17
SafeCast	**1.54**	0.20

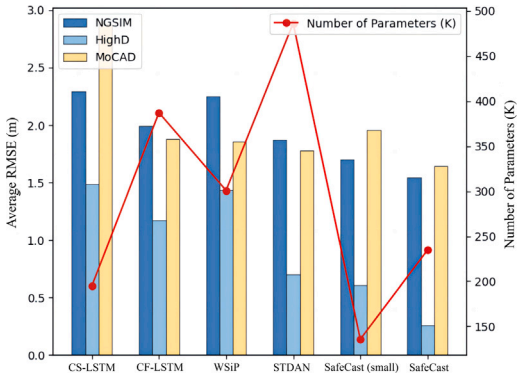


Fig. 6. Comparative analysis of the performance of the models and their parametric counts.

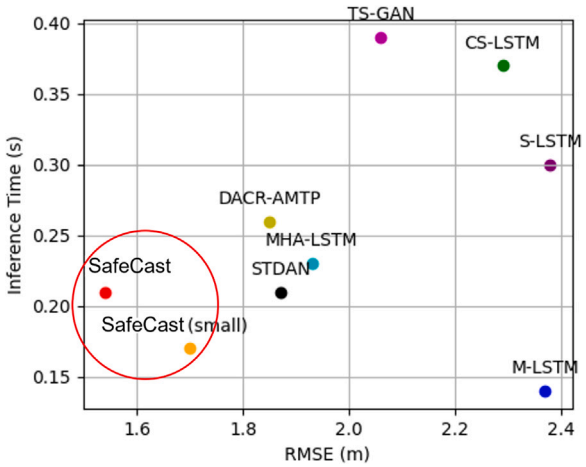


Fig. 7. Inference speed versus prediction accuracy comparison between SafeCast and open-source baselines. Models closer to the origin achieve superior performance. The top two performers are marked in a red circle.

of 21.78%. Notably, even the lightweight SafeCast (small) variant outperforms most full-scale models, demonstrating the efficiency and scalability of our SafeCast model.

Moreover, on the ApolloScape dataset, a challenging urban dataset with diverse agent types (vehicles, pedestrians, and cyclists), SafeCast achieves the best overall performance. As shown in Table 1, it records a WSADE of 1.1253 and WSFDE of 2.1024, outperforming the previous best baseline (MSTG) by 2.5% and 1.2%, respectively. It also establishes new benchmarks across all agent-specific ADE and FDE metrics, underscoring its effectiveness in modeling fine-grained, heterogeneous interactions. The integration of uncertainty-aware GAT and Global Uncertainty Fusion proves especially beneficial for dynamic agents like pedestrians and cyclists. Finally, in the urban, multi-agent environments of the MoCAD dataset, SafeCast maintains its leading performance. It achieves an average RMSE improvement of at least 8.90% over competing methods and reduces long-term RMSE by 0.32–0.33 m. These results further validate SafeCast's capacity to handle dense, interactive traffic scenes with high temporal variability.

Overall, SafeCast demonstrates outstanding generalization and robustness across diverse driving scenarios, consistently advancing the state of the art in both accuracy and reliability for motion forecasting in real-world AD systems.

4.3.2. *Inference speed and model complexity*

To further demonstrate the scalability and efficiency of our SafeCast and SafeCast (small), we conduct a detailed benchmark analysis of their inference speed and complexity against other SOTA models. The results are presented in Table 4 and Fig. 7. Notably, the efficiency metrics are not commonly provided in this field, and our access to the source code of various models is limited. Therefore, our comparison primarily focuses on available open-source models.

When compared to the fastest inference model, M-LSTM, SafeCast and SafeCast (small) exhibit slightly slower speeds. However, they demonstrate superior performance on the NGSIM dataset, with an average RMSE 35.02% higher than M-LSTM. Impressively, SafeCast and SafeCast (small) improve inference speed by at least 4.76% and 19.05%, respectively, compared to other SOTA models, while ensuring better prediction performance. The inference speed is highly dependent on the specific algorithm implementation and the computing hardware

Table 5
Different methods and components of ablation study.

Components	Ablation methods						
	A	B	C	D	E	F	G
Intention-aware encoder	✘	✔	✔	✔	✔	✔	✔
Safety-aware spatial encoder	✔	✘	✔	✔	✔	✔	✔
Safety-aware temporal encoder	✔	✔	✘	✔	✔	✔	✔
Multi-modal decoder	✔	✔	✔	✘	✔	✔	✔
Multi-head attention mechanism	✔	✔	✔	✔	✘	✔	✔
Graph uncertainty feature method	✔	✔	✔	✔	✔	✘	✔

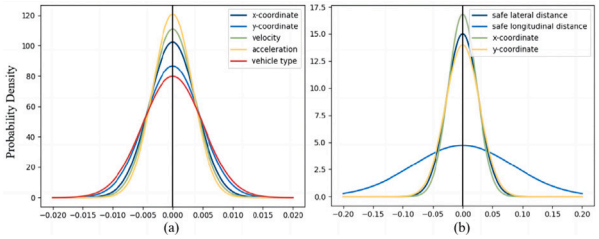


Fig. 8. The visualization shows the variable-learnable noise distribution produced by the GUF layer, illustrating how the noise adapts to different contexts: (a) intention-aware module, (b) safety-aware module.

used, and it can vary significantly even for identical model architectures. Moreover, the parameter analysis depicted in Table 3 and Fig. 6 illustrates that SafeCast achieves high precision in motion forecasting with a modest parameter quantity (just 234.55K). This figure represents a reduction of 39.41% and 22.01%, compared to the leading baselines CF-LSTM and WSiP, respectively. Notably, SafeCast (small) has the fewest parameters while outperforming most of the state-of-the-art baseline models regarding prediction accuracy. These findings underscore the efficient and streamlined design of SafeCast (small), highlighting its prowess in motion forecasting.

4.4. Ablation study

To evaluate the individual contributions of each component in our model, we conduct a series of ablation studies. Table 5 outlines the specific modifications for each method: Method A excludes the Intention-aware module; Method B omits the spatial encoder from the Safety-aware module, while Method C removes the temporal encoder; Method D eliminates the multimodal probabilistic maneuver prediction in the decoder; Method E replaces the multi-head attention mechanism with a convolutional neural network; Method F excludes the GUF method; and Method G includes all components as the complete model.

The ablation results on the NGSIM dataset, shown in Table 6, highlight the importance of each component in improving model performance. All models were fully trained, and the results reflect their best performance on the test set. Excluding the intention-aware module (Method A) leads to a performance drop of at least 10.73% in long-term predictions, while removing safety-aware features (Method B) results in a 4.74% decrease compared to the full model (Method G). Method D exhibits the most substantial performance decline, with a reduction of at least 7.50%, underscoring the significance of multimodal probabilistic maneuver prediction in the decoder. Furthermore, incorporating the temporal encoder within the Safety-aware module proves essential for capturing the temporal dependencies of the target vehicle. The comparison between Method C (without the temporal encoder) and Method G highlights its importance, particularly for short-term prediction accuracy. The benefits of the GUF method in handling uncertainties are evident in Method F, where excluding GUF results in noticeable performance deterioration. Finally, replacing the multi-head attention mechanism with a convolutional neural network in Method E causes a performance decrease of at least 4.2%. These findings emphasize the critical role of each component in motion forecasting, as demonstrated by the superior results of the Method G.

4.5. Qualitative results

4.5.1. Qualitative results of GUF noise distribution

We integrate the GUF into both the intention-aware module and the spatial encoder of the safety-aware module. This design addresses a core challenge in modeling traffic dynamics: the lack of universally accepted standards for noise distribution in historical state elements — such as velocity, acceleration, and 2D positional coordinates — within the same traffic scenario. Instead of relying on fixed noise thresholds, GUF introduces learnable Gaussian noise parameters that dynamically adapt to the characteristics of the input data. This flexibility allows the prediction model to better capture uncertainty and improve generalization.

To assess the effectiveness of GUF, we visualize the learned noise distributions for both modules in Fig. 8. These distributions exhibit normal-like behavior, with the range and variance varying across different input features. In Fig. 8(a), representing the intention-aware module, the learned noise for acceleration differs from that of the y-coordinate, illustrating GUF's capacity to differentiate uncertainty based on the relational significance of each feature. The relatively compact distributions in this module are likely due to the strong numerical and semantic correlations among position, velocity, and acceleration.

In contrast, Fig. 8(b) shows the learned noise distributions in the spatial encoder, which incorporates features such as coordinates and safety distances. These features have weaker interdependencies compared to the intention-aware module, resulting in broader and more diverse noise ranges. This suggests that the spatial encoder captures more latent, abstract safety representations beyond conventional physical dynamics. Overall, the GUF enables the model to flexibly adapt its noise modeling strategy to reflect the underlying complexity of intention and safety features, supporting more robust and context-aware prediction.

4.5.2. Qualitative results for multimodal forecasting

Fig. 9 illustrates the multimodal probabilistic forecasting performance of the SafeCast model on the NGSIM dataset. The heatmaps depict the GMM's predictions at each time step, with brighter colors indicating higher probabilities for the forecasted motion. These results confirm that the most probable forecasts closely match the ground truth, demonstrating the impressive performance of SafeCast.

4.5.3. Qualitative results for real-world datasets

We present qualitative visualizations of SafeCast's motion forecasting in real-world safety-critical scenarios, with a focus on the impact of incorporating the Responsibility-Sensitive Safety (RSS) metric. Safety-critical scenarios refer to situations where collision avoidance and maintaining safe distances are essential—such as high-density traffic, sudden deceleration of leading vehicles, or abrupt lane changes. To evaluate the effect of integrating RSS, we compare the performance of SafeCast with and without safety constraints, as illustrated in Fig. 10(a1), (a2), and (a3).

In Fig. 10(a1), the target vehicle is closely surrounded by other vehicles in the front and adjacent lanes. SafeCast, when incorporating the RSS metric, calculates both longitudinal and lateral safety distances, prompting the vehicle to decelerate appropriately and maintain a safe buffer, resulting in a predicted trajectory closely aligned with the ground truth. In contrast, SafeCast without the RSS module (SafeCast (-RSS)) disregards these constraints and fails to model safe spacing, leading to significant deviation from the true path.

Fig. 10(a2) presents a scenario with minimal surrounding traffic, allowing the target vehicle to move more freely. In this case, SafeCast correctly identifies the lack of immediate obstacles and generates an extended motion forecast consistent with the ground truth. Meanwhile,

Table 6
Ablation results for different models. **Bold** values represent the best performance. (Evaluation metric: RMSE (m)).

Prediction horizon (s)	Method A (ΔMethod G)	Method B (ΔMethod G)	Method C (ΔMethod G)	Method D (ΔMethod G)	Method E (ΔMethod G)	Method F (ΔMethod G)	Method G (Full model)
1	$0.42_{\downarrow 5.00\%}$	$0.42_{\downarrow 5.00\%}$	$0.45_{\downarrow 12.50\%}$	$0.43_{\downarrow 7.50\%}$	$\mathbf{0.40}_{\downarrow 0.00\%}$	$0.42_{\downarrow 5.00\%}$	**0.40**
2	$0.96_{\downarrow 6.67\%}$	$0.94_{\downarrow 4.44\%}$	$0.95_{\downarrow 5.56\%}$	$1.02_{\downarrow 13.33\%}$	$0.92_{\downarrow 2.22\%}$	$0.94_{\downarrow 4.44\%}$	**0.90**
3	$1.56_{\downarrow 9.86\%}$	$1.49_{\downarrow 4.93\%}$	$1.49_{\downarrow 4.93\%}$	$1.71_{\downarrow 20.42\%}$	$1.45_{\downarrow 2.11\%}$	$1.51_{\downarrow 6.34\%}$	**1.42**
4	$2.27_{\downarrow 10.73\%}$	$2.15_{\downarrow 4.88\%}$	$2.12_{\downarrow 3.41\%}$	$2.59_{\downarrow 26.34\%}$	$2.08_{\downarrow 1.46\%}$	$2.13_{\downarrow 3.90\%}$	**2.05**
5	$3.22_{\downarrow 11.42\%}$	$3.02_{\downarrow 4.50\%}$	$2.96_{\downarrow 2.42\%}$	$3.70_{\downarrow 28.03\%}$	$2.93_{\downarrow 1.38\%}$	$3.01_{\downarrow 4.15\%}$	**2.89**

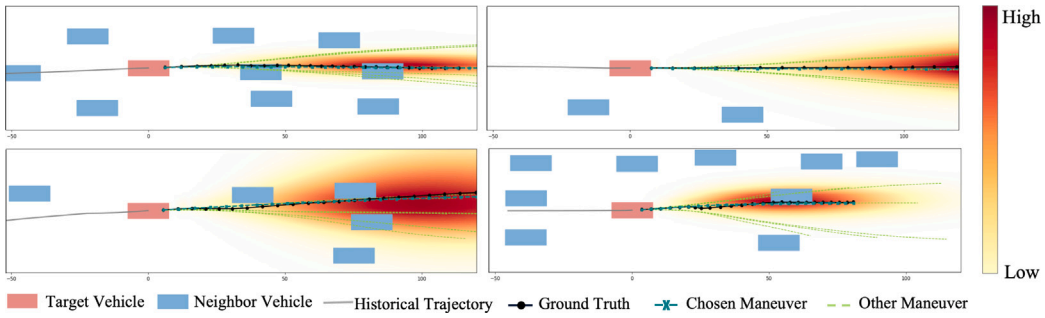


Fig. 9. Multi-modal probabilistic motion forecasting of the target vehicle. Heat maps show the Gaussian Mixture Model (GMM) of the predictions, with brighter regions indicating higher probabilities.

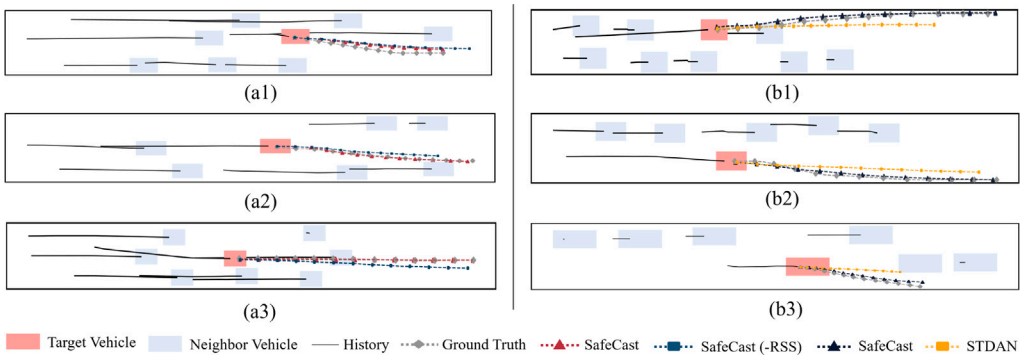


Fig. 10. Qualitative results of our proposed models on the NGSIM dataset. (a1)–(a3): Visualization results of motion forecasting between SafeCast and SafeCast (-RSS) with and without RSS metrics; (b1)–(b3): Comparison of visualization results between SafeCast and STDAN. The target vehicle is depicted in red, while its surrounding vehicles are shown in light blue.

SafeCast (-RSS) underestimates the available motion space, highlighting its inability to adaptively recognize safe conditions.

Fig. 10(a3) further demonstrates SafeCast's handling of lateral safety. When a nearby vehicle is in an adjacent lane, SafeCast correctly infers the importance of lateral separation, predicting a cautious path that preserves lane integrity and avoids potential conflict. Conversely, SafeCast (-RSS) forecasts an unsafe lane-change maneuver, diverging significantly from the ground truth and exposing a failure in lateral risk awareness.

To further highlight SafeCast's robustness, we compare it with a state-of-the-art baseline, STDAN, under similar conditions, as shown in Fig. 10(b1)–(b3). Across various complex scenarios — including lane merging and traffic-dense environments — SafeCast consistently produces forecasts more closely aligned with the ground truth. These visual comparisons underscore the model's ability to incorporate real-time safety considerations, and its superior reliability in dynamic and uncertain traffic contexts. Overall, these qualitative results demonstrate SafeCast's adaptability, scene comprehension, and safety-awareness, reinforcing its effectiveness for real-world motion forecasting.

5. Conclusion

Accurate motion forecasting of surrounding vehicles is a pivotal challenge in achieving fully autonomous driving. To tackle this challenge, we propose SafeCast, a novel Risk-Responsive motion forecasting model, designed to enhance prediction accuracy while prioritizing safety. This model integrates the Responsibility-Sensitive Safety model to characterize the safety attributes of traffic agents while employing a graph-based framework for real-time modeling of their intentions and safety features. In addition, we propose a variable-learnable noise approach to simulate the inherent uncertainties of traffic environments, which enhances the adaptability and robustness of our model. We also implement a feature fusion strategy to navigate the intricate relationships among various features. Furthermore, we utilize a GMM-based decoder for multimodal prediction to improve the model's forecasting precision. Evaluated on the NGSIM, HighD, ApolloScape, and MoCAD real-world datasets, SafeCast achieves SOTA performance while maintaining a relatively lightweight architecture and quick inference speeds,

demonstrating its robustness, efficiency, and practical applicability in autonomous driving systems.

Limitations and Future Work. While SafeCast demonstrates superior performance and practical utility, we recognize several limitations in its current formulation. The design of the RSS-based module prioritizes a trade-off between interpretability, real-world applicability, and computational efficiency. While the data-driven strategy is productive, it may not fully capture the nuanced variability of real-world driving behavior across diverse conditions. In the future, we plan to extend the temporal observation window to better capture long-term behavioral patterns, refine the granularity of RSS parameterization to account for varying driver profiles and vehicle capabilities, and incorporate richer contextual information, such as road surface conditions, weather, and vehicle types. These enhancements are expected to further improve the adaptability, robustness, and fidelity of the safety-aware module in real-world driving scenarios.

CRediT authorship contribution statement

Haicheng Liao: Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Hanlin Kong:** Software, Methodology, Investigation. **Zhenning Li:** Formal analysis, Conceptualization. **Chengzhong Xu:** Project administration, Methodology.

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.

Acknowledgments

This research is supported by Science and Technology Development Fund of Macau SAR (File no. 0122/2024/RIB2, 0215/2024/AGJ, 001/2024/SKL, 0021/2022/ITP), State Key Lab of Intelligent Transportation System (2024-B001), Jiangsu Provincial Science and Technology Program (BZ2024055), University of Macau (SRG2023-00037-IOTSC, MYRG-GRG2024-00284-IOTSC), and Shenzhen-Hong Kong-Macau Science and Technology Program Category C (SGDX20230821095159012).

References

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S., 2016. Social lstm: Human trajectory prediction in crowded spaces. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 961–971.

Chen, W., Wang, F., Sun, H., 2021. S2tnet: Spatio-temporal transformer networks for trajectory prediction in autonomous driving. In: Proceedings of the 13th Asian Conference on Machine Learning. Vol. 157, pp. 454–469.

Chen, H., Wei, C., Liu, Y., Hu, C., Zhang, X., 2024. Vulnerable traffic participant trajectory prediction based on gate recurrent unit-attention and ameliorative social force model. IEEE Trans. Transp. Electr. 10 (4), 9396–9405, Art no. 2513912.

Chen, X., Zhang, H., Deng, F., Liang, J., Yang, J., 2023. Stochastic non-autoregressive transformer-based multi-modal pedestrian trajectory prediction for intelligent vehicles. IEEE Trans. Intell. Transp. Syst. 25 (5), 3561–3574.

Chen, X., Zhang, H., Zhao, F., Cai, Y., Wang, H., Ye, Q., 2022a. Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for internet of vehicles. IEEE Trans. Instrum. Meas. 71, 1–12, Art no. 2513912.

Chen, X., Zhang, H., Zhao, F., Hu, Y., Tan, C., Yang, J., 2022b. Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for internet of vehicles. IEEE Trans. Intell. Transp. Syst. 23 (10), 19471–19483.

Cheng, H., Liu, M., Chen, L., Broszio, H., Sester, M., Yang, M.Y., 2023. Gatraj: A graph- and attention-based multi-agent trajectory prediction model. ISPRS J. Photogramm. Remote Sens. 205, 163–175.

Cong, P., Xiao, Y., Wan, X., Deng, M., Li, J., Zhang, X., 2024. DACR-AMTP: Adaptive multi-modal vehicle trajectory prediction for Dynamic Drivable Areas based on collision risk. IEEE Trans. Intell. Veh. 9 (9), 5339–5360.

Deo, N., Trivedi, M.M., 2018a. Convolutional social pooling for vehicle trajectory prediction. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. CVPRW, pp. 1549–15498.

Deo, N., Trivedi, M.M., 2018b. Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs. In: 2018 IEEE Intelligent Vehicles Symposium. IV, pp. 1179–1184.

Fang, L., Jiang, Q., Shi, J., Zhou, B., 2020. Tpnet: Trajectory proposal network for motion prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6797–6806.

Fang, T., Zhang, Y., YANG, Y., Wang, C., Chen, L., 2023. Universal prompt tuning for graph neural networks. In: Advances in Neural Information Processing Systems. Vol. 36, Curran Associates, Inc, pp. 52464–52489.

Gao, K., Li, X., Chen, B., Hu, L., Liu, J., Du, R., Li, Y., 2023. Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment. IEEE Trans. Intell. Transp. Syst. 24 (6), 6203–6216.

Geng, M., Li, J., Xia, Y., Chen, X.M., 2023. A physics-informed transformer model for vehicle trajectory prediction on highways. Transp. Res. Part C: Emerg. Technol. 154, 104272.

Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., Alahi, A., 2018. Social GAN: Socially acceptable trajectories with generative adversarial networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2255–2264.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. Vol. 33, Curran Associates, Inc, pp. 6840–6851.

Huang, Z., Karkus, P., Ivanovic, B., Chen, Y., Pavone, M., Lv, C., 2024. Dtpp: Differentiable joint conditional prediction and cost evaluation for tree policy planning in autonomous driving. In: 2024 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 6806–6812.

Huang, Z., Mo, X., Lv, C., 2022. Multi-modal motion prediction with transformer-based neural network for autonomous driving. In: 2022 International Conference on Robotics and Automation. ICRA, pp. 2605–2611.

Li, R., Shi, H., Fu, Z., Wang, Z., Lin, G., 2023. Weakly supervised class-agnostic motion prediction for autonomous driving. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 17599–17608.

Li, X., Ying, X., Chuah, M.C., 2019. Grip: Graph-based interaction-aware trajectory prediction. In: 2019 IEEE Intelligent Transportation Systems Conference. ITSC, IEEE, pp. 3960–3966.

Liao, H., Kong, H., Wang, B., Wang, C., Ye, W., He, Z., Xu, C., Li, Z., 2025. CoT-drive: Efficient motion forecasting for autonomous driving with LLMs and chain-of-thought prompting. IEEE Trans. Artif. Intell. http://dx.doi.org/10.1109/TAI.2025.3564594.

Liao, H., Li, Y., Li, Z., Wang, C., Cui, Z., Li, S.E., Xu, C., 2024a. A cognitive-based trajectory prediction approach for autonomous driving. IEEE Trans. Intell. Veh. 9 (4), 4632–4643.

Liao, H., Li, Z., Shen, H., Zeng, W., Liao, D., Li, G., Xu, C., 2024b. Bat: Behavior-aware human-like trajectory prediction for autonomous driving. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38, pp. 10332–10340.

Liao, H., Liu, S., Li, Y., Li, Z., Wang, C., Li, Y., Li, S.E., Xu, C., 2024c. Human observation-inspired trajectory prediction for autonomous driving in mixed-autonomy traffic environments. In: 2024 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 14212–14219.

Ma, Y., Zhu, X., Zhang, S., Yang, R., Wang, W., Manocha, D., 2019. Trafficpredict: Trajectory prediction for heterogeneous traffic-agents. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33, pp. 6120–6127.

Messaoud, K., Yahiaoui, I., Verroust-Blondet, A., Nashashibi, F., 2019. Non-local social pooling for vehicle trajectory prediction. In: 2019 IEEE Intelligent Vehicles Symposium. IV, IEEE, pp. 975–980.

Messaoud, K., Yahiaoui, I., Verroust-Blondet, A., Nashashibi, F., 2020. Attention based vehicle trajectory prediction. IEEE Trans. Intell. Veh. 6 (1), 175–185.

Mo, X., Liu, H., Huang, Z., Li, X., Lv, C., 2024. Map-adaptive multimodal trajectory prediction via intention-aware unimodal trajectory predictors. IEEE Trans. Intell. Transp. Syst. 25 (6), 5651–5663.

Naumann, M., Wirth, F., Oboril, F., Scholl, K.-U., Elli, M.S., Alvarez, I., Weast, J., Stiller, C., 2021. On responsibility sensitive safety in car-following situations - A parameter analysis on german highways. In: 2021 IEEE Intelligent Vehicles Symposium. IV, pp. 83–90.

Nistér, D., Lee, H.-L., Ng, J., Wang, Y., 2019. The Safety Force Field, Vol. 15, NVIDIA White Paper, pp. 1–27.

Shalev-Shwartz, S., Shammah, S., Shashua, A., 2017. On a formal model of safe and scalable self-driving cars. http://dx.doi.org/10.48550/arXiv.1708.06374, CoRR abs/1708.06374. arXiv:1708.06374.

Tang, C., Salakhutdinov, R.R., 2019. Multiple futures prediction. In: Proceedings of the Advances in Neural Information Processing Systems. Vol. 32, pp. 15424–15434.

Tang, L., et al., 2023. Trajectory prediction for autonomous driving based on multiscale spatial-temporal graph. IET Intell. Transp. Syst. 17 (2), 386–399.

Wang, R., Wang, S., Yan, H., Wang, X., 2023. Wsip: Wave superposition inspired pooling for dynamic interactions-aware trajectory prediction. Proc. AAAI Conf. Artif. Intell. 37 (4), 4685–4692.

Wang, M., Zhang, L., Chen, J., Zhang, Z., Wang, Z., Cao, D., 2024. A hybrid trajectory prediction framework for automated vehicles with attention mechanisms. IEEE Trans. Transp. Electr. 10 (3), 6178–6194.

Wang, Y., Zhao, S., Zhang, R., Cheng, X., Yang, L., 2020. Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion. IEEE Trans. Intell. Transp. Syst. 23 (1), 236–248.

Xie, X., Zhang, C., Zhu, Y., Wu, Y.N., Zhu, S.-C., 2021. Congestion-aware multi-agent trajectory prediction for collision avoidance. In: 2021 IEEE International Conference on Robotics and Automation. ICRA, pp. 13693–13700.

Yang, B., Fan, F., Ni, R., Wang, H., Jafaripournimchahi, A., Hu, H., 2024. A multi-task learning network with a collision-aware graph transformer for traffic-agents trajectory prediction. Trans. Intell. Transp. Syst. 25 (7), 6677–6690.

Zhang, K., Li, L., 2022. Explainable multimodal trajectory prediction using attention models. Transp. Res. Part C: Emerg. Technol. 143, 103829.

Zhang, K., Zhao, L., Dong, C., Wu, L., Zheng, L., 2022. AI-TP: Attention-based interaction-aware trajectory prediction for autonomous driving. IEEE Trans. Intell. Veh. 8 (1), 73–83.

Zhao, T., Xu, Y., Monfort, M., Choi, W., Baker, C., Zhao, Y., Wang, Y., Wu, Y.N., 2019. Multi-agent tensor fusion for contextual trajectory prediction. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 12118–12126.

Zhou, Z., Wang, J., Li, Y.-H., Huang, Y.-K., 2023. Query-centric trajectory prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17863–17873.

Zhou, Z., Ye, L., Wang, J., Wu, K., Lu, K., 2022. HiVT: Hierarchical vector transformer for multi-agent motion prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 8823–8833.

Zuo, Z., Wang, X., Guo, S., Liu, Z., Li, Z., Wang, Y., 2023. Trajectory prediction network of autonomous vehicles with fusion of historical interactive features. IEEE Trans. Intell. Veh. 9 (1), 2171–2183.