A Cognitive-Driven Trajectory Prediction Model for Autonomous Driving in Mixed Autonomy Environments

Haicheng Liao[1] , Zhenning Li[1] , Chengyue Wang[1] , Bonan Wang[1] , Hanlin Kong[2] , Yanchen Guan[1] , Guofa Li[3] , Zhiyong Cui[4]
[1]University of Macau
[2]University of Electronic Science and Technology of China
[3]Chongqing University
[4]Beihang University
{yc27979, zhenningli, chengyuewang, mc35002, yc37976}@um.edu.com, hanlinkong@foxmail.com, hanshan198@gmail.com, zhiyongc@uw.edu

Abstract

As autonomous driving technology progresses, the need for precise trajectory prediction models becomes paramount. This paper introduces an innovative model that infuses cognitive insights into trajectory prediction, focusing on perceived safety and dynamic decision-making. Distinct from traditional approaches, our model excels in analyzing interactions and behavior patterns in mixed autonomy traffic scenarios. It represents a significant leap forward, achieving marked performance improvements on several key datasets. Specifically, it surpasses existing benchmarks with gains of 16.2% on the Next Generation Simulation (NGSIM), 27.4% on the Highway Drone (HighD), and 19.8% on the Macao Connected Autonomous Driving (MoCAD) dataset. Our proposed model shows exceptional proficiency in handling corner cases, essential for real-world applications. Moreover, its robustness is evident in scenarios with missing or limited data, outperforming most of the state-of-the-art baselines. This adaptability and resilience position our model as a viable tool for real-world autonomous driving systems, heralding a new standard in vehicle trajectory prediction for enhanced safety and efficiency.

1 Introduction

In the evolving landscape of autonomous driving (AD) systems, the challenge of accurately predicting vehicle trajectories is paramount, especially in mixed autonomy environments where autonomous vehicles (AVs) and human-driven vehicles (HVs) dynamically interact. This complex interplay, soon to become an everyday reality on our roads, is marked by its unpredictability, as detailed in studies like [Schwarting *et al.*, 2019] and [Liao *et al.*, 2024b]. Despite the advancements in deep learning for AD, typified in research such as [Sun *et al.*, 2023] and [Geng *et al.*, 2023], a critical gap remains: these models often lack a nuanced understanding of human driving behavior and decision-making processes [Chen *et al.*, 2022a; Liao *et al.*, 2024d; Tian *et al.*, 2024; Guan *et al.*, 2024]. This limitation becomes particularly evident in real-world scenarios where the adaptability of these models to diverse and unforeseen conditions is crucial.

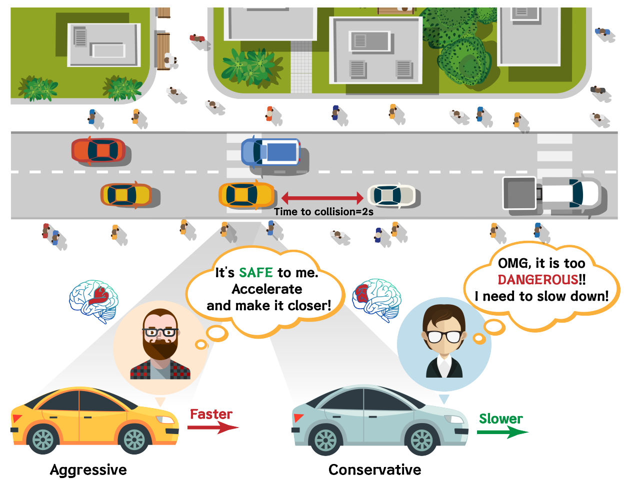

Figure 1: Perceived safety and its influence on decisions of drivers with different driving behaviors.

This backdrop prompts us to ask critical questions about the future trajectory of AD: Is the key to advancing AD not just in accumulating more data or refining algorithms, but in gaining a deeper understanding of the driving environment itself? How can we reshape our models to interpret and respond to the intricate human dynamics that underpin driving? Motivated by these questions, our research embarks on an innovative path. We propose a paradigm shift, extending beyond conventional data-driven approaches to embrace a critical yet often-neglected aspect of driving – the concept of **perceived safety**. This concept, pivotal in shaping driving behaviors and decisions, is deeply rooted in psychological constructs, as detailed in [Rubagotti *et al.*, 2022]. According to the Theory of Planned Behavior, individual actions in driving are influenced by attitudes (driving behaviors towards others), subjective norms (personal evaluation of safety), and perceived behavioral control (confidence in driving ability)

[Ajzen, 1991]. Further depth is added by neuroscientific research, such as studies by [Saadatnejad *et al.*, 2022] and [Kronemer *et al.*, 2022], which unveil that perceived safety is an intricate blend of both conscious and instinctive responses, involving the amygdala's emotional processing and the prefrontal cortex's rational decision-making. Notably, this understanding of perceived safety is exemplified in diverse driving scenarios. For instance, when encountering a close car ahead, different drivers exhibit markedly varied responses, as depicted in Figure 1. An aggressive driver, possibly influenced by sensation-seeking tendencies [Zuckerman, 1990], might quickly swerve, perceiving lower risk. Conversely, a cautious driver, perhaps more risk-averse [Rabin, 2013], might opt for a complete stop. These behaviors, far from being random, are intricately linked to each driver's psychological profile and past experiences, revealing a significant limitation in current AD systems: their inability to account for these complex, cognitive behavioral patterns. In response, our research integrates the concept of perceived safety into trajectory prediction models for AVs. This integration does more than add a new variable; it injects a human-centric perspective into the heart of these systems. By doing so, we aim to enhance the models' ability to interpret driving behaviors, leading to predictions that are not only technically accurate but also richly informed by contextual and psychological insights. Overall, the contributions of this study include:

1. Perceived Safety-Aware Module Development: 1) Quantitative Safety Assessment (QSA): A foundational component that objectively evaluates a driving scenario's safety, leveraging metrics like Time-To-Collision (TTC) and Risk Tendency Index (RTI), thereby offering a perception-neutral safety assessment. 2) Driver Behavior Profiling (DBP): Layering upon QSA's foundation, DBP infuses the human element, differentiating between, for instance, an aggressive driver's risk tolerance and a cautious driver's reservations. This profile captures driving nuances in real-time without the need for manual labeling and selection of time windows.

2. Social Interaction-Aware Module Development: Acknowledging the collective ballet on roads, we have architected a transformer-based framework, named Leanformer, which encapsulates the intricate inter-vehicular interactions, mirroring contemporary shifts in AD research trajectories.

3. Robustness Against Data Inconsistencies: Our model significantly outperforms the SOTA baseline models when tested on the NGSIM, HighD, and MoCAD datasets, respectively. In a significant stride towards practical applicability, it demonstrates unparalleled resilience in scenarios with incomplete or inconsistent data. This adaptability ensures that our trajectory predictions remain accurate and reliable even when faced with real-world data challenges, a feature often overlooked in conventional models.

2 Related Work

Motion Prediction For Autonomous Driving. A wealth of recent research has harnessed deep learning architectures to predict future motions of target agents for AVs. Employing a variety of frameworks, these models encompass sequential networks [Chen *et al.*, 2022a; Messaoud *et al.*, 2021], graph neural networks [Rowe *et al.*, 2023], transformer constructs [Chen *et al.*, 2022a; Yin *et al.*, 2021], and generative models [Li *et al.*, 2023]. The crux of these approaches is discerning temporal and spatial dynamics between traffic agents from historical data, aiming to optimize prediction accuracy. Moreover, prior research [Deo and Trivedi, 2018; Messaoud *et al.*, 2019; Chen *et al.*, 2022b] has tapped into the intricate social interplays among traffic agents, unveiling latent insights and enhancing predictive performance.

Perceived Safety Concept. The concept of perceived safety has received extensive attention in psychology and physical human-robot interaction (pHRI) [Guiochet *et al.*, 2017]. In pHRI, perceived safety finds applications in various autonomous physical systems, including mobile robots [Scheunemann *et al.*, 2020], industrial manipulators [Davis, 2023], humanoid robots [Busch *et al.*, 2019], and autonomous driving [Aledhari *et al.*, 2023], to assess and represent people's perception of the level of danger in interactions with robots and their level of comfort during such interactions [Bartneck *et al.*, 2009]. However, the evaluation of perceived safety in trajectory prediction is often subjective and lacks standard metrics. This work introduces a new quantitative criterion for perceived safety, inspired by human decision-making processes, enhancing the model's contextual understanding of behaviors and traffic conditions.

Driving Behavior Understanding. Numerous previous studies [Schwarting *et al.*, 2019; Chandra *et al.*, 2020] have proposed various criteria and metrics to explicitly represent and detect driving behavior, often utilizing classic scales such as Social Value Orientation (SVO) [Murphy *et al.*, 2011], Driving Anger Scale (DAS) [Deffenbacher *et al.*, 1994], Driving Anger Expression (DAE), Driving Style Questionnaire (DSQ) [French *et al.*, 1993], among others. In contrast to these traditional methods, our approach models driving behavior in real-time with adaptive, behavior-aware criteria, eliminating the need for manual labeling and offering enhanced learning flexibility. This method addresses challenges related to shifting behaviors and time window selection, reducing data variability associated with discrete classification.

3 Problem Formulation

Our objective is to predict the trajectories of *target vehicle* (denoted with subscript 0) within the perceptual boundaries of an autonomous vehicle (AV), referred to as the *ego vehicle*, in settings with mixed autonomy. Given the present moment as t, the task is to estimate the probable future trajectory, $\boldsymbol{Y}_0^{t:t+t_f}$, of the target vehicle over the ensuing t_f time intervals. This estimation is based on the historical states of the target vehicle, $\boldsymbol{X}_0^{t-t_h:t}$, and the states of surrounding vehicles (indicated by subscript 1:n), $\boldsymbol{X}_{1:n}^{t-t_h:t}$, within a defined duration t_h. Given the inherent uncertainties in trajectory predictions, our methodology adopts a multimodal prediction framework. This approach discerns various potential maneuvers for the target vehicle and calculates the corresponding probabilities from prior states. As a result, we obtain a spectrum of predictions, each paired with its confidence level.

Central to our model is the use of historical states as inputs. These comprise the 2D position coordinates $p_{0:n}^{t-t_h:t}$, velocity

$v_{0:n}^{t-t_h:t}$, and acceleration $a_i^{t-t_h:t}$ spanning the duration from $t-t_h$ to t. Formally defined, the input is:

$$\boldsymbol{X}_i^{t-t_h:t} = \left\{p_i^{t-t_h:t}, v_i^{t-t_h:t}, a_i^{t-t_h:t}\right\}, \forall i \in [0,n] \tag{1}$$

The resulting output is:

$$\boldsymbol{Y}_0^{t:t+t_f} = \left\{\boldsymbol{p}_0^{t+1}, \boldsymbol{p}_0^{t+2}, \ldots, \boldsymbol{p}_0^{t+t_f-1}, \boldsymbol{p}_0^{t+t_f}\right\} \tag{2}$$

where $\boldsymbol{p}_0^t = \{(p_{0,1}^t; c_{0,1}^t), (p_{0,2}^t; c_{0,2}^t), \cdots, (p_{0,M}^t; c_{0,M}^t)\}$ encompasses both the potential trajectory and its associated likelihood ($\Sigma_1^M c_{0,i}^t = 1$), with M enoting the total number of potential trajectories predicted.

4 Trajectory Prediction Model

Figure 2 illustrates the architecture of our model. Rooted in the encoder-decoder paradigm, the model seamlessly incorporates three novel modules: the Perceived Safety-Aware, the Priority-Aware, and the Interaction-Aware modules. Collectively, these modules are designed to emulate the human decision-making process during driving.

4.1 Perceived Safety Aware Module

Quantitative Safety Assessment. In traffic safety domain, three metrics have gained prominence for their comprehensive portrayal of on-road risks: TTC, Time Exposed Time-to-Collision (TET), and Time Integrated Time-to-Collision (TIT) [Minderhoud and Bovy, 2001]. To align these with our study's objectives, we made slight modifications. Using the 2D position coordinates p_i^t, p_j^t and velocity v_i^t, v_j^t for vehicles i and j at time t, these metrics are formulated as:

1) Time-to-Collision (TTC): TTC is a widely accepted measure used to evaluate the time available before two vehicles collide if they continue on their current trajectories. It offers insights into imminent collision risks and serves as an early warning indicator. TTC for the ith vehicle is computed as:

$$TTC_i^t = -\frac{d_{i,j}^t}{\dot{d}_{i,j}^t} \tag{3}$$

where $d_{i,j}$ represents the distance between vehicles i and j, and $\dot{d}_{i,j}$ is its rate of change:

$$\begin{cases} d_{i,j}^t = \sqrt{\left(p_i^t - p_j^t\right)^\top \left(p_i^t - p_j^t\right)} \\ \dot{d}_{i,j}^t = \frac{1}{d_{i,j}^t}\left(p_i^t - p_j^t\right)^\top \left(v_i^t - v_j^t\right) \end{cases} \tag{4}$$

2) Time Exposed Time-to-Collision (TET): It measures the exposure duration to critical TTC values within t_h. It is the sum product of a switching variable and a time threshold τ_{sc} (set at 0.1s):

$$TET_i^{t_k} = \sum_{t_k=t-t_h}^{t} \delta_i(t_k) \cdot \tau_{\text{sc}} \tag{5}$$

with the switching variable given by:

$$\delta_i(t_k) = \begin{cases} 1 & \forall \quad 0 \le \mathrm{TTC}_i^{t_k} \le \mathrm{TTC}^* \\ 0 & \quad\text{otherwise} \end{cases} \tag{6}$$

In our study, TTC^*, delineating safe and critical thresholds, is set at 3.0s.

3) Time Integrated Time-to-Collision (TIT): An adaptation of TET, TIT integrates TTC profile to evaluate safety levels, factoring in the evolution of each vehicle's TET temporally:

$$TIT_i^{t_k} = \sum_{\tilde{t}=t-t_f}^{t_h} \left[\mathrm{TTC}^* - \mathrm{TTC}_i(t_k)\right] \cdot \tau_{\text{sc}} \tag{7}$$

Elevated values of TTC, TET, and TIT imply sustained exposure to potential collision risks, underscoring a deterioration in perceived safety.

4) Risk Tendency Index (RTI): To further capture congestion patterns in complex traffic environments, we propose an index between the ith and jth vehicles at time t, denoted as *subjective risk perception indicator* (SPR), i.e. R_i^t, and *dynamic risk volatility indicator* (DRV), i.e. $R_{i,j}^t$, respectively:

$$R_i^t = R_j^t = \left[R_{i,j}^t, \dot{R}_{i,j}^t\right]^T, \forall i,j \in [0,n], i \neq j \tag{8}$$

In this context, the vector $R_{i,j}^t$ with larger values indicates an increased risk of collision, while the vector $\dot{R}^t{}_{i,j}$ characterizes the dynamic congestion conditions in complex traffic scenarios. In addition, the definitions of SPR and DRV are defined as follows:

$$R_{i,j}^t = R_{j,i}^t = \begin{cases} 1/e^{q_{i,j}^t}, q_{i,j}^t > 0 \\ 0, q_{i,j}^t = 0 \end{cases} \tag{9}$$

where the DRV $\dot{R}_{i,j}^t$ represents the gradient to evaluate fluctuations in SPR $R_{i,j}^t$ and can be expressed as follows:

$$\dot{R}^t{}_{i,j} = \dot{R}_{j,i}^t = \begin{cases} 1/e^{\dot{q}_{i,j}^t}, \dot{q}_{i,j}^t > 0 \\ 0, \dot{q}_{i,j}^t = 0 \end{cases} \tag{10}$$

The quantities $q_{i,j}^t$ and $\dot{q}_{i,j}^t$ are calculated based on several critical parameters related to the dynamics of two traffic agents. These parameters include the lateral velocity v_x^t, longitudinal velocity v_y^t, 2D position coordinate p_x^t and p_y^t, as well as the lateral $a_x^t$and longitudinal a_y^t. Mathematically, it can be represented as follows:

$$q_{i,j}^t = \max\left(-\frac{\Delta_{i,j}v_x^t \times \Delta_{i,j}p_x^t + \Delta_{i,j}v_y^t \times \Delta_{i,j}p_y^t}{\Delta_{i,j}v_x^2 + \Delta_{i,j}v_y^2}, 0\right) \tag{11}$$

$$\dot{q}_{i,j}^t = -\frac{\Delta_{i,j}a_x^t \times \Delta_{i,j}p_x^t + \Delta_{i,j}a_y^t \times \Delta_{i,j}p_y^t}{\Delta_{i,j}a_x^2 + \Delta_{i,j}a_y^2} \tag{12}$$

where the $\Delta_{i,j}(\cdot)$ denotes the difference between quantities of the i-th and j-th vehicles. A larger vector $R_{i,j}^t$ indicates a higher risk of collision, while the vector $\dot{R}^t{}_{i,j}$ describes the dynamic congestion conditions in complex traffic scene.

The encoder within the Perceived Safety-Aware module combines the GCNs and the scaled dot-product multi-head self-attention mechanism. To capture the dynamic geometric relationships among traffic agents, we employ a convolutional neural network on a fully connected interaction multigraph. This multigraph operation layer incorporates sequential perceived-safety criteria $\boldsymbol{Q}$ as nodes and adjacency matrix A as graph edges. The GCN layer is given by:

$$\boldsymbol{Z}_i^{k+1} = \phi_{\text{ReLU}}\left(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{Z}_i^k\boldsymbol{W}_i^k\right) \tag{13}$$

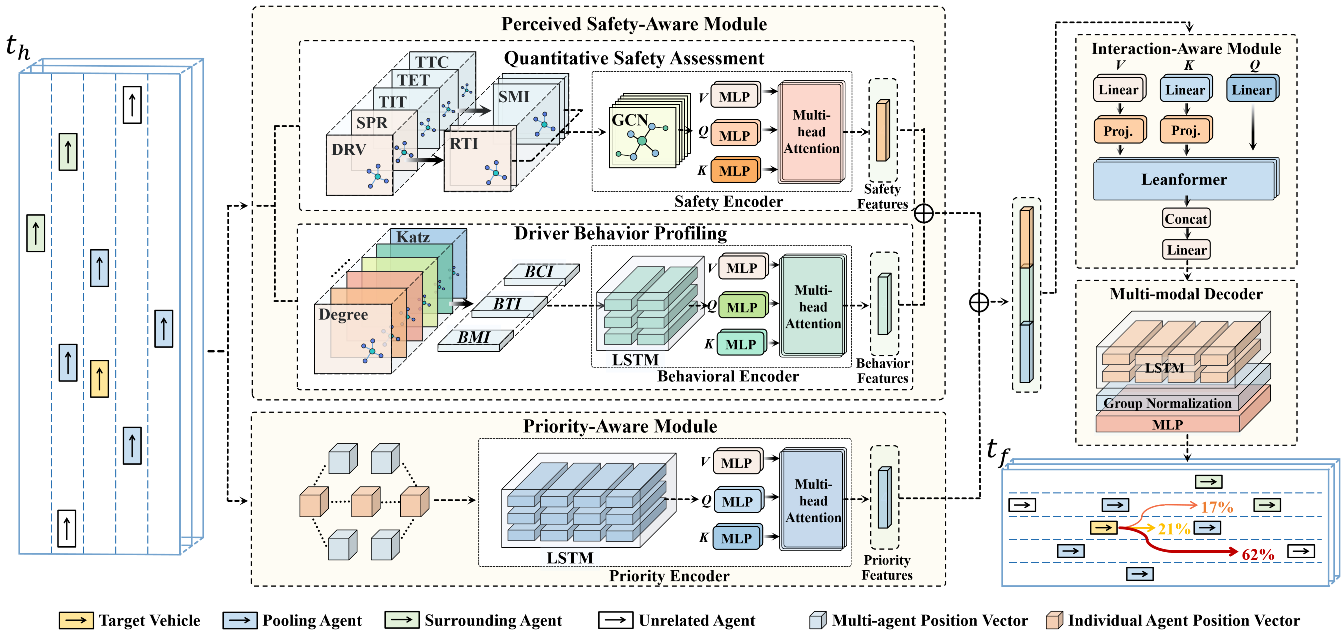

Figure 2: Architecture of proposed trajectory prediction model.

where $\tilde{\boldsymbol{D}}$ is the degree matrix. The multi-head self-attention mechanism then transform the feature matrix $\boldsymbol{Z}_i^{k+1}$ output from GCNs into query, key, and value vectors. The output for the ith agent from this mechanism is denoted as α.

For improved training stability, inspired by ResNet [He *et al.*, 2016], we integrate Gate Linear Units (GLUs) and Layer Normalization to efficiently manage features:

$$\bar{O}^{t-t_h:t} = \phi_{LN}\left(\phi_{MLP}(\phi_{GLUs}(\alpha))\right) \tag{14}$$

In particular, the GLUs layer can be defined as:

$$\phi_{GLUs}(\alpha) = (\alpha W_1 + b_1) \odot \phi_{sigmoid}(\alpha W_2 + b_2) \tag{15}$$

where α represents the safe attention coefficient from the multi-head attention mechanism, W_1 and W_2 are the learnable weight parameters associated with the GLUs layer, b_1 and b_2 are the corresponding biases, $\odot$ denotes element-wise multiplication, $\phi_{sigmoid}$ is the sigmoid activation function, and $\phi_{LN}(\cdot)$ stands for Layer Normalization. Correspondingly, the output of the encoder within the QSA is denoted as $\bar{O}_{safe}^{t-t_h:t}$.

Driver Behavior Profiling. Unlike traditional methods that classify driver behavior into fixed, predefined categories, we propose a novel dynamic geometric graph (DGG) based real-time profiling approach. This continuum solution addresses the problem of fluctuating behavioral categorization in previous studies. At a given time t, the graph G^t is constructed as $G^t = \{V^t, E^t\}$, where $V^t = \{v_0^t, v_1^t, ..., v_n^t\}$ signifies the node set, and each node v_i^t represents an individual agent. $E^t = \{e_0^t, e_1^t, ..., e_n^t\}$ is the set of edges illustrating potential interactions between agents. Specifically, an edge e_i^t connects node v_i^t to other nodes (agents) potentially influencing it.

Interactions are considered when agents, such as v_i and v_j, are sufficiently close. Specifically, if their distance $d(v_i^t, v_j^t)$ is within a set threshold r, an interaction is inferred. This relationship is represented as $e_i^t = \{v_i^t, v_j^t \mid (j \in N_i^t)\}$ where the neighborhood set N_i^t is defined by:

$$N_i^t = \{v_j^t \in V^t \setminus \{v_i^t\} \mid d(v_i^t, v_j^t) \leq r, \text{ and } i \neq j\} \tag{16}$$

The adjacency matrix A^t of G^t is symmetric. Formally,

$$A^t(i,j) = \begin{cases} d(v_i^t, v_j^t) & \text{if } d(v_i^t, v_j^t) \leq r \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

To further evaluate the behavior of individual agents and reveal important agents and overall connectivity in the traffic graph, centrality measures are used, including degree $J_i^t(D)$, closeness $J_i^t(C)$, eigenvector $J_i^t(E)$, betweenness $J_i^t(B)$, power $J_i^t(P)$, and Katz $J_i^t(K)$ centrality. Formally,

(1) Degree Centrality: Reflects an agent's number of connections, showing its influence on and vulnerability to others.

$$J_i^t(D) = \left|N_i^t\right| + J_i^{t-1}(D) \tag{18}$$

where $|\mathcal{N}i^t|$ denotes the total elements in N_i^t.

(2) Closeness Centrality: Indicates an agent's reachability, suggesting its potential influence over others:

$$J_i^t(C) = \frac{|N_i^t| - 1}{\sum_{\forall v_j^t \in N_i^t} d\left(v_i^t, v_j^t\right)} \tag{19}$$

(3) Eigenvector Centrality: Measures an agent's importance by considering both quantity and quality of connections:

$$J_i^t(E) = \frac{\sum_{\forall v_j^t \in N_i^t} d\left(v_i^t, v_j^t\right)}{\lambda} \tag{20}$$

where λ is the eigenvalue, indicating the collective influence exerted by an agent and its network.

(4) Betweenness Centrality: Highlights agents that act as bridges or bottlenecks in traffic, crucial in congested scenes:

$$J_i^t(B) = \sum_{\forall v_s^t, v_k^t \in V^t} \frac{\sigma_{j,k}(v_i^t)}{\sigma_{j,k}} \tag{21}$$

(5) Power Centrality: Identifies agents in recurrent interactions, hinting at traffic patterns:

$$J_i^t(P) = \sum_k \frac{A_{ii}^k}{k!} \tag{22}$$

where A_{ii}^k denotes the i-th diagonal element of the adjacency matrix raised to the k-th power, while $k!$ signifies the factorial of k, shedding light on its contribution to the network's structural integrity and dynamism.
(6) Katz Centrality: Emphasizes both direct and distant interactions of an agent, capturing intricate driving patterns:

$$J_i^t(K) = \sum_k \sum_j \alpha^k A_{ij}^k + \beta^k, \forall i, j \in [0, n], \text{ where } \alpha^k < \frac{1}{\lambda_{\max}} \tag{23}$$

where n represents the number of agents in the real-time traffic scenario, α^k is the decay factor, β^k denotes the weight assigned to the immediate neighboring agents, and A_{ij}^k is the i,j-th element of the k-th power of the adjacency matrix.

To provide better performance in profiling driving behavior, we draw inspiration from metrics such as speed, acceleration, and jerk. Based on these, we introduce three continuous criteria: Behavior Magnitude Index (BMI) $\mathcal{C}_i^t$, which measures the influence of driving behaviors by evaluating their centrality; Behavior Tendency Index (BTI) $\mathcal{L}_i^t$, which quantifies behavior propensity by calculating time series derivatives, suggesting higher probabilities of specific behaviors with larger derivatives; and Behavior Curvature Index (BCI) $\mathcal{I}_i^t$, which uses the jerk concept to measure the intensity of driving behaviors by calculating the second-order derivatives of continuous centrality measures. Formally,

$$\mathcal{C}_i^t = \left| \mathcal{J}_i^t(\tilde{X}_i^t) \right|^T, \mathcal{L}_i^t = \left| \frac{\partial \mathcal{C}_i^t}{\partial t} \right|^T, \mathcal{I}_i^t = \left| \frac{\partial \mathcal{C}_i^t}{\partial^2 t} \right|^T \tag{24}$$

where $\tilde{X}_i^t = [J_i^t(D), J_i^t(C), J_i^t(E), J_i^t(B), J_i^t(P), J_i^t(K)]$.

In combination, these criteria offer a holistic view of driving behaviors without needing manual labeling during training, addressing issues of changing behavior labels and time window selection.

The encoder of Driver Behavior Profiling comprises two main components: the LSTM and multi-head self-attention mechanism. The behavior-aware criteria are first processed by the LSTM, yielding temporal vectors:

$$\tilde{J}_i^{t-t_h:t} = \phi_{LSTM}\left(h_i^{t-t_h:t}, \phi_{MLP}(J_i^{t-t_h:t}), \phi_{MLP}(\bar{O}_{safety}^{t-th:t})\right) \tag{25}$$

Here, the LSTM encoder updates the hidden state of agent v_i frame-by-frame with shared weights. Using the multi-head self-attention mechanism and GLUs, we compute attention weights for different agent behaviors, producing precise sequential behavioral features in a similar way in QSA:

$$\bar{O}_{behavior}^{t-t_h:t} = \phi_{LN}\left(\phi_{MLP}(\phi_{GLUs}(\alpha^{behavior}))\right) \tag{26}$$

Lastly, the behavioral features are integrated with priority features and contextual features, then forwarded to the Interaction-Aware module for further processing.

4.2 Priority-Aware Module

In light of recent advances in cognitive studies, it has become evident that the spatial positioning of vehicles within a scene can variably influence the behavior and decisions of a target vehicle. For instance, vehicles located directly in the anticipated trajectory path tend to exert greater influence relative to those situated behind. Furthermore, during overtaking maneuvers, vehicles positioned on the left may carry augmented significance. Recognizing these spatial intricacies, we introduce the Priority-Aware Module. This sophisticated module adeptly transforms the spatial coordinates of agents, encoding them into high-dimensional positional vectors, thereby yielding detailed positional features. Our pooling mechanism adeptly amalgamates dynamic positional information from the encompassing traffic scenario, effectively capturing both singular and multi-agent positional vectors. This mechanism emphasizes the dynamic nuances of position data, accommodating historical agent states, denoted as $\boldsymbol{S}_i^{t_k}$, as well as the intricate spatial interplay symbolized by $\boldsymbol{P}_{i,j}^{t_k}$. Mathematically, these relationships are represented as $\boldsymbol{S}_i^{t_k} = \{p_i^{t_k} - p_i^{t_k-1}, v_i^{t_k} - v_i^{t_k-1}, a_i^{t_k} - a_i^{t_k-1}\}$ and $\boldsymbol{P}_{i,j}^{t_k} = \{p_i^{t_k} - p_j^{t_k}, v_i^{t_k} - v_j^{t_k}, a_i^{t_k} - a_j^{t_k}\}$, where $t_k \in [t-t_h, t]$. Within this module, the encoder, which amalgamates both LSTM and multi-head attention mechanisms, meticulously processes the dynamic positional vectors. This processing involves transmuting discrete position vectors into a more continuous spatio-temporal domain, thereby enhancing the representation of temporal and spatial dynamics. At each discrete temporal instance t, the encoder assimilates recent historical position vectors via an LSTM network:

$$O_{priority}^{t-t_h:t} = \phi_{LSTM}\left(\bar{\boldsymbol{h}}_i^{t-t_h:t-1}, \boldsymbol{S}_i^{t-t_h:t-1}, \boldsymbol{P}_{i,j}^{t-t_h:t-1}\right) \tag{27}$$

Subsequently, the LSTM's output is channeled through a multi-head attention mechanism, culminating in the synthesis of refined priority features:

$$\bar{O}_{priority}^{t-t_h:t} = \phi_{LN}\left(\phi_{MLP}(\phi_{GLUs}(\alpha^{priority}))\right) \tag{28}$$

4.3 Interaction-Aware Module

To better understand the synergistic influence of surrounding vehicles' risk levels, positions, and one's own behavior on the target vehicle's future trajectory, we introduce the Leanformer in this module, which is a lightweight version of the Transformer, allowing for a favorable trade-off between accuracy and efficiency. Mathematically,

$$O^{t-t_h:t} = \phi_{Leanformer}\left(\bar{O}_{safety}^{t-t_h:t} \| \bar{O}_{behavior}^{t-t_h:t} \| \bar{O}_{priority}^{t-t_h:t}\right) \tag{29}$$

This equation captures the integrated effect of safety perception and behavioral tendencies on the target agent's trajectory.

4.4 Multimodal Decoder

The decoder, rooted in a Gaussian Mixture Model (GMM) with multimodality, employs a dedicated LSTM and a fully connected layer. It processes the composite interactive vectors $\bar{O}$ to forecast the target vehicle's trajectory. The predicted trajectory, $\boldsymbol{Y}_{\mathbf{0}}^{\boldsymbol{t:t+t_f}}$, is determined by:

$$\boldsymbol{Y}_{\mathbf{0}}^{\boldsymbol{t:t+t_f}} = F_{\theta}\left(F_{\theta}(\bar{O})\right) \tag{30}$$

where $F_{\theta}(\cdot) = \phi_{ReLU}\left(\phi_{MLP}\left[\phi_{\text{GN}}\left(\phi_{LSTM}(\cdot)\right)\right]\right)$ Moreover, ϕ_{ReLU} is the ReLU activation, and ϕ_{GN} is Group Normalization, used for improved training stability. The decoder's output comprises multiple future trajectories for target vehicle.

5 Experiments

5.1 Experimental Setup

We carried out experiments using three prominent datasets: NGSIM [Deo and Trivedi, 2018], HighD [Krajewski *et al.*, 2018], and MoCAD [Liao *et al.*, 2024b], adhering to a common dataset division framework. Trajectories were split into 8s segments, where the initial 3s served as historical data, and the succeeding 5s were used for assessment, forming the *complete* test set. The Root Mean Square Error (RMSE) was employed as the evaluation metric. Recognizing a gap in research concerning data omissions in trajectory prediction, we established the *missing* test set, which was further categorized into three subsets based on the duration of data omissions: *drop 3-frames*, *drop 5-frames*, and *drop 8-frames*. Omissions were purposefully made around the midpoint of the historical trajectory. For instance, in the *drop 5-frames* subset, data ranging from the $(t-8)$th to the $(t-12)$th frame was excluded. To manage these omissions, we employed simple linear interpolation. Furthermore, we train our model on only 25% of the dataset and evaluate it on *complete* test sets to evaluate its ability to handle unfamiliar data and manage data omissions. Additionally, our model is trained on four A40 48G GPUs. We employ the Adam optimizer and utilize CosineAnnealingWarmRestarts for the scheduler.

Model	Prediction Horizon (s)				
	1	2	3	4	5
S-GAN [Gupta *et al.*, 2018]	0.57	1.32	2.22	3.26	4.40
CS-LSTM [Deo and Trivedi, 2018]	0.61	1.27	2.09	3.10	4.37
MATF-GAN [Zhao *et al.*, 2019]	0.66	1.34	2.08	2.97	4.13
DRBP[Gao *et al.*, 2023]	1.18	2.83	4.22	5.82	-
MFP [Tang and Salakhutdinov, 2019]	0.54	1.16	1.89	2.75	3.78
NLS-LSTM [Messaoud *et al.*, 2019]	0.56	1.22	2.02	3.03	4.30
MHA-LSTM [Messaoud *et al.*, 2021]	0.41	1.01	1.74	2.67	3.83
WSiP [Wang *et al.*, 2023]	0.56	1.23	2.05	3.08	4.34
CF-LSTM [Xie *et al.*, 2021]	0.55	1.10	1.78	2.73	3.82
STDAN [Chen *et al.*, 2022b]	0.42	1.01	1.69	2.56	3.67
iNATran [Chen *et al.*, 2022a]	0.39	0.96	1.61	2.42	3.43
BAT (25%) [Liao *et al.*, 2024b]	<u>0.31</u>	<u>0.85</u>	1.65	2.69	3.87
BAT [Liao *et al.*, 2024b]	**0.23**	**0.81**	1.54	2.52	3.62
FHIF [Zuo *et al.*, 2023]	0.40	0.98	1.66	2.52	3.63
DACR-AMTP [Cong *et al.*, 2023]	0.57	1.07	1.68	2.53	3.40
Our Model	<u>0.31</u>	**0.81**	**1.44**	**2.08**	**2.85**
Our Model (25%)	0.43	0.94	1.57	2.55	3.32
Our Model (drop 3-frames)	0.40	0.88	<u>1.53</u>	<u>2.34</u>	<u>2.99</u>
Our Model (drop 5-frames)	0.44	0.92	1.57	2.42	3.34
Our Model (drop 8-frames)	0.48	0.98	1.68	2.47	3.52

Table 1: Evaluation of the proposed model and baselines on the NGSIM dataset's *complete* and *missing* test sets over a 5-second prediction horizon. The accuracy metric is RMSE (m). Cases marked as ('-') indicate unspecified values. Bold and underlined values represent the best and second-best performance in each category.

Model	Prediction Horizon (s)				
	1	2	3	4	5
S-GAN [Gupta *et al.*, 2018]	0.30	0.78	1.46	2.34	3.41
WSiP [Wang *et al.*, 2023]	0.20	0.60	1.21	2.07	3.14
CS-LSTM [Deo and Trivedi, 2018]	0.22	0.61	1.24	2.10	3.27
MHA-LSTM [Messaoud *et al.*, 2021]	0.19	0.55	1.10	1.84	2.78
NLS-LSTM [Messaoud *et al.*, 2019]	0.20	0.57	1.14	1.90	2.91
DRBP[Gao *et al.*, 2023]	0.41	0.79	1.11	1.40	-
EA-Net [Cai *et al.*, 2021]	0.15	0.26	0.43	0.78	1.32
CF-LSTM [Xie *et al.*, 2021]	0.18	0.42	1.07	1.72	2.44
STDAN [Chen *et al.*, 2022b]	0.19	0.27	0.48	0.91	1.66
iNATran [Chen *et al.*, 2022a]	0.04	**0.05**	0.21	0.54	1.10
DACR-AMTP [Cong *et al.*, 2023]	0.10	0.17	0.31	0.54	1.01
BAT (25%) [Liao *et al.*, 2024b]	0.14	0.34	0.65	0.89	1.27
BAT [Liao *et al.*, 2024b]	0.08	0.14	<u>0.20</u>	0.44	0.62
GaVa [Liao *et al.*, 2024c]	0.17	0.24	0.42	0.86	1.31
Our Model	**0.04**	<u>0.09</u>	**0.19**	**0.31**	**0.45**
Our Model (25%)	0.09	0.22	0.41	0.64	0.95
Our Model (drop 3-frames)	<u>0.05</u>	0.12	0.21	<u>0.33</u>	<u>0.49</u>
Our Model (drop 5-frames)	0.17	0.31	0.44	0.65	0.98
Our Model (drop 8-frames)	0.18	0.47	0.84	1.27	1.74

Table 2: Evaluation of our model and SOTA baselines on HighD.

5.2 Experimental Results

Performance Comparison on *Complete* Test Set. As illustrated in Table 1, our model surpasses the best of SOTA baselines with noteworthy gains of 10.6% and 16.2% for short-term (1s-3s) and long-term predictions (4s-5s) respectively on the NGSIM dataset. Moreover, Table 2 highlights fewer prediction inaccuracies on the HighD dataset than NGSIM, due to HighD's superior trajectory precision and larger sample size. Notably, our model's long-term forecasting outperforms SOTA baselines, showing RMSE improvements of 27.4% for a 5s prediction horizon. Additionally, as shown in Table 3, our model notably excels on busy urban roads, surpassing SOTA baselines by at least 12.2% for short-term predictions and reducing long-term prediction errors by at least 0.57 meters on MoCAD, potentially decreasing traffic accident risks.

Performance Comparison on *Missing* Test Set. Tables 1, 2, and 3 spotlight our model's robustness, even in the face of incomplete datasets. Consistently, across the *drop 3-frames* and *drop 5-frames* datasets, it outperforms all baselines. Notably, within the *drop 3-frames* context, our model even eclipses several leading SOTA benchmarks that were evaluated using *complete* test sets. However, no model is without its exceptions. In the *drop 5-frames* dataset, while our model largely dominates, it trails behind DACR-AMTP on NGSIM and the short-term forecasts of iNATran on HighD. Still, these are isolated instances, and the overarching trend confirms the model's robustness and versatility. A deeper dive reveals a predictable trend: the model's performance is inversely proportional to the omission of input data. Yet, even under a stringent scenario like the *drop 8-frames*—where half the data is missing—our model stands its ground, frequently matching or surpassing most SOTA benchmarks.

Performance on a Limited 25% Training Set. To challenge our model's adaptability, we trained it using a quarter of the available training set from the NGSIM, HighD, and MoCAD, yet evaluated its performance on the *complete* test set. Impressively, as shown in Tables 1, 2, and 3, even with this limited training data, our model delivered RMSE values that were notably lower than most baseline models. Such results underscore our model's efficiency and robustness in trajectory

Model	Prediction Horizon (s)				
	1	2	3	4	5
S-GAN [Gupta *et al.*, 2018]	1.69	2.25	3.30	3.89	4.69
CS-LSTM [Deo and Trivedi, 2018]	1.45	1.98	2.94	3.56	4.49
MHA-LSTM [Messaoud *et al.*, 2021]	1.25	1.48	2.57	3.22	4.20
NLS-LSTM [Messaoud *et al.*, 2019]	0.96	1.27	2.08	2.86	3.93
WSiP [Wang *et al.*, 2023]	0.70	0.87	1.70	2.56	3.47
CF-LSTM [Xie *et al.*, 2021]	0.72	0.91	1.73	2.59	3.44
STDAN [Chen *et al.*, 2022b]	0.62	0.85	1.62	2.51	3.32
BAT (25%) [Liao *et al.*, 2024b]	0.65	0.99	1.89	2.81	3.58
BAT [Liao *et al.*, 2024b]	0.35	0.74	1.39	2.19	2.88
HLTP [Liao *et al.*, 2024a]	0.55	0.76	1.44	2.39	3.21
Our Model	**0.30**	**0.65**	**1.07**	**1.66**	**2.31**
Our Model (25%)	0.53	0.94	1.38	2.27	3.14
Our Model (drop 3-frames)	0.35	0.78	1.23	1.77	2.42
Our Model (drop 5-frames)	0.46	0.92	1.29	1.86	2.74
Our Model (drop 8-frames)	0.68	1.04	1.73	2.17	3.02

Table 3: Evaluation of our model and SOTA baselines on MoCAD.

prediction. This performance indicates a promising potential: our model might substantially cut down on the data demands typically associated with training autonomous vehicles, particularly in scenarios that are data-scarce or unconventional.

Dataset	Time (s)	Model					
		A	B	C	D	E	F
NGSIM	1	0.47	0.45	0.41	0.43	0.39	**0.31**
	2	1.01	0.97	0.89	0.93	0.88	**0.81**
	3	1.50	1.69	1.77	1.54	1.59	**1.44**
	4	2.58	2.70	2.39	2.47	2.27	**2.08**
	5	3.32	3.41	3.10	3.26	2.98	**2.85**
HighD	1	0.06	0.0	0.04	0.05	0.05	**0.04**
	2	0.13	0.14	0.10	0.12	0.11	**0.09**
	3	0.25	0.25	0.21	0.24	0.22	**0.19**
	4	0.40	0.42	0.36	0.38	0.36	**0.31**
	5	0.58	0.67	0.51	0.55	0.49	**0.45**

Table 4: Ablation results for various models on NGSIM and HighD.

5.3 Qualitative Results

It is an intrinsic understanding that vehicles in closer proximity often pose greater risks. However, an aspect frequently neglected by many models is that even distant vehicles, based on their driving behavior, can significantly influence the ego vehicle's trajectory. In our observations, driving behavior substantially impacts risk evaluations. Specifically, drivers with a more aggressive demeanor tend to escalate the risk levels. Drawing a direct comparison, our model's predictions stand out in their precision and alignment with reality, especially when juxtaposed against its counterparts like Stdan and WSiP. Taking the right merge scenario as a case in point: where Stdan and WSiP show discrepancies, it forecasts with minimal deviation, resonating closely with the ground truth. This differential is not just a testament to our model's accuracy but also its adeptness at capturing the nuanced effects of driving behaviors, a facet that seems muted in other models.

5.4 Ablation Studies

We conduct a comprehensive ablation study to assess each component's contribution to our model, with results summarized in Table 4. Model F, which includes all components, consistently outperforms other variations across all metrics, demonstrating their collective value in achieving optimal performance. Model A, excluding the DBP within the Perceived

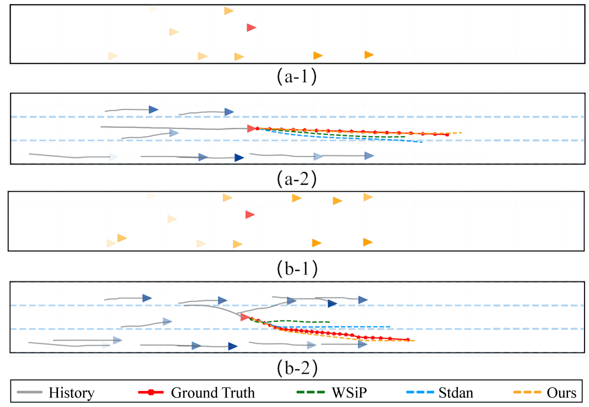

Figure 3: Visual insights from the NGSIM dataset depict two complex driving scenarios: (a) driving straight and (b) merging to the right. The target vehicle is marked by red triangles, with surrounding vehicles represented by other triangles. Subfigures (a-1) and (b-1) present the outcomes of driver behavior profiling, whereas (a-2) and (b-2) illustrate the perceived safety metrics for each vehicle alongside their predicted trajectories. A vehicle shaded in deeper brown signals potential aggressive behavior, while a darker blue shade indicates a heightened risk to the target vehicle and vice versa.

Safety-Aware module, shows significant performance degradation, particularly in short-term predictions, with drops of at least 13.9% and 24.0% on the NGSIM and HighD, respectively. This underscores the critical role of DBP in enhancing trajectory prediction accuracy. Model B, a reduced version of Model F without the Perceived Safety-Aware module, shows a significant reduction in RMSE, particularly for long-term predictions, with improvements of at least 19.6% and 26.2% on NGSIM and HighD datasets, respectively. This emphasizes the importance of perceived safety factors in trajectory prediction, especially for long-term predictions. Model C, which uses absolute coordinates instead of relative positions, demonstrates notable reductions in prediction metrics, underscoring the importance of spatial relationships for accuracy. Model D, lacking the Interaction-Aware module, shows performance drops of at least 6.5% and 20.0% on the NGSIM and HighD datasets for short-term predictions, and at least 12.6% and 18.2% for long-term predictions, respectively. Finally, Model E, with reduced multimodal prediction in the decoder, experiences at least 4.4% and 8.2% performance degradation on the NGSIM and HighD datasets, respectively.

6 Conclusion

This work presents a novel approach rooted in cognitive insights, emphasizing the crucial role of perceived safety in driving behaviors. Our Perceived Safety-Aware Module harmoniously merges Quantitative Safety Assessment and Driver Behavior Profiling, offering a detailed perspective on safety perceptions in driving. Rigorous evaluations on NGSIM, HighD, and MoCAD highlight our model's robustness and adaptability, even under data constraints and data missing. In conclusion, our findings suggest a promising trajectory for future research, bridging computational strength with human cognition to drive safety and efficiency in AVs.

Acknowledgements

This research is supported by the Science and Technology Development Fund of Macau SAR (File no. 0021/2022/ITP, 0081/2022/A2, 001/2024/SKL), Shenzhen-Hong Kong-Macau Science and Technology Program Category C (SGDX20230821095159012), and University of Macau (SRG2023-00037-IOTSC). Haicheng Liao and Zhenning Li contributed equally to this work. Please ask Dr. Zhenning Li (zhenningli@um.edu.mo) for correspondence.

References

[Ajzen, 1991] Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

[Aledhari *et al.*, 2023] Mohammed Aledhari, Mohamed Rahouti, Junaid Qadir, Basheer Qolomany, Mohsen Guizani, and Ala Al-Fuqaha. Motion comfort optimization for autonomous vehicles: Concepts, methods, and techniques. *IEEE Internet of Things Journal*, 2023.

[Bartneck *et al.*, 2009] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81, 2009.

[Busch *et al.*, 2019] Baptiste Busch, Giuseppe Cotugno, Mahdi Khoramshahi, Grigorios Skaltsas, Dario Turchi, Leonardo Urbano, Mirko Wächter, You Zhou, Graham Asfour, et al. Evaluation of an industrial robotic assistant in an ecological environment. In *RO-MAN*, 2019.

[Cai *et al.*, 2021] Yingfeng Cai, Zihao Wang, Hai Wang, Long Chen, Yicheng Li, Miguel Angel Sotelo, and Zhixiong Li. Environment-attention network for vehicle trajectory prediction. *IEEE Transactions on Vehicular Technology*, 2021.

[Chandra *et al.*, 2020] Rohan Chandra, Uttaran Bhattacharya, Trisha Mittal, Aniket Bera, and Dinesh Manocha. Cmetric: A driving behavior measure using centrality functions. In *IEEE/RSJ IROS*. IEEE, 2020.

[Chen *et al.*, 2022a] Xiaobo Chen, Huanjia Zhang, Feng Zhao, Yingfeng Cai, Hai Wang, and Qiaolin Ye. Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for internet of vehicles. *IEEE Transactions on Instrumentation and Measurement*, 2022.

[Chen *et al.*, 2022b] Xiaobo Chen, Huanjia Zhang, Feng Zhao, Yu Hu, Chenkai Tan, and Jian Yang. Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for internet of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):19471–19483, 2022.

[Cong *et al.*, 2023] Peichao Cong, Yixuan Xiao, Xianquan Wan, Murong Deng, Jiaxing Li, and Xin Zhang. Dacr-amtp: Adaptive multi-modal vehicle trajectory prediction for dynamic drivable areas based on collision risk. *IEEE Transactions on Intelligent Vehicles*, 2023.

[Davis, 2023] Jenny L Davis. Role-taking and robotic form: An exploratory study of social connection in human-robot interaction. *International Journal of Human-Computer Studies*, page 103094, 2023.

[Deffenbacher *et al.*, 1994] Jerry L Deffenbacher, Eugene R Oetting, and Rebekah S Lynch. Development of a driving anger scale. *Psychological reports*, 74(1):83–91, 1994.

[Deo and Trivedi, 2018] Nachiket Deo and Mohan M Trivedi. Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1468–1476, 2018.

[French *et al.*, 1993] Davina J French, Robert J West, James Elander, and John Martin Wilding. Decision-making style, driving style, and self-reported involvement in road traffic accidents. *Ergonomics*, 36(6):627–644, 1993.

[Gao *et al.*, 2023] Kai Gao, Xunhao Li, Bin Chen, Lin Hu, Jian Liu, Ronghua Du, and Yongfu Li. Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment. *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[Geng *et al.*, 2023] Maosi Geng, Yong Chen, Yingji Xia, and Xiqun Michael Chen. Dynamic-learning spatial-temporal transformer network for vehicular trajectory prediction at urban intersections. *Transportation research part C: emerging technologies*, 156:104330, 2023.

[Guan *et al.*, 2024] Yanchen Guan, Haicheng Liao, Zhenning Li, Guohui Zhang, and Chengzhong Xu. World models for autonomous driving: An initial survey. *arXiv preprint arXiv:2403.02622*, 2024.

[Guiochet *et al.*, 2017] Jérémie Guiochet, Mathilde Machin, and Hélène Waeselynck. Safety-critical advanced robots: A survey. *Robotics and Autonomous Systems*, 2017.

[Gupta *et al.*, 2018] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE CVPR*, 2018.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Krajewski *et al.*, 2018] Robert Krajewski, Julian Bock, Laurent Kloeker, and Lutz Eckstein. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *ITSC*, 2018.

[Kronemer *et al.*, 2022] Sharif I Kronemer, Mark Aksen, Julia Z Ding, Jun Hwan Ryu, Qilong Xin, Zhaoxiong Ding, Jacob S Prince, Hunki Kwon, Aya Khalaf, Sarit Forman, et al. Human visual consciousness involves large scale cortical and subcortical networks independent of task report and eye movement activity. *Nature Communications*, 13(1):7342, 2022.

[Li *et al.*, 2023] Peizheng Li, Shuxiao Ding, Xieyuanli Chen, Niklas Hanselmann, and Juergen Gall. Powerbev:

A powerful yet lightweight framework for instance prediction in bird's-eye view. In *Proceedings of IJCAI*, 2023.

[Liao *et al.*, 2024a] Haicheng Liao, Yongkang Li, Zhenning Li, Chengyue Wang, Zhiyong Cui, Shengbo Eben Li, and Chengzhong Xu. A cognitive-based trajectory prediction approach for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, pages 1–12, 2024.

[Liao *et al.*, 2024b] Haicheng Liao, Zhenning Li, Huanming Shen, Wenxuan Zeng, Dongping Liao, Guofa Li, and Chengzhong Xu. Bat: Behavior-aware human-like trajectory prediction for autonomous driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 10332–10340, 2024.

[Liao *et al.*, 2024c] Haicheng Liao, Shangqian Liu, Yongkang Li, Zhenning Li, Chengyue Wang, Bonan Wang, Yanchen Guan, and Chengzhong Xu. Human observation-inspired trajectory prediction for autonomous driving in mixed-autonomy traffic environments. *arXiv preprint arXiv:2402.04318*, 2024.

[Liao *et al.*, 2024d] Haicheng Liao, Huanming Shen, Zhenning Li, Chengyue Wang, Guofa Li, Yiming Bie, and Chengzhong Xu. Gpt-4 enhanced multimodal grounding for autonomous driving: Leveraging cross-modal attention with large language models. *Communications in Transportation Research*, 4:100116, 2024.

[Messaoud *et al.*, 2019] Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, and Fawzi Nashashibi. Non-local social pooling for vehicle trajectory prediction. In *IEEE Intelligent Vehicles Symposium (IV)*, 2019.

[Messaoud *et al.*, 2021] Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, and Fawzi Nashashibi. Attention based vehicle trajectory prediction. *IEEE Transactions on Intelligent Vehicles*, 6(1):175–185, 2021.

[Minderhoud and Bovy, 2001] Michiel M Minderhoud and Piet HL Bovy. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention*, 33(1):89–97, 2001.

[Murphy *et al.*, 2011] Ryan O Murphy, Kurt A Ackermann, and Michel JJ Handgraaf. Measuring social value orientation. *Judgment and Decision making*, 6(8):771–781, 2011.

[Rabin, 2013] Matihew Rabin. Risk aversion and expected-utility theory: A calibration theorem. In *Handbook of the fundamentals of financial decision making: Part I*, pages 241–252. World Scientific, 2013.

[Rowe *et al.*, 2023] Luke Rowe, Martin Ethier, Eli-Henry Dykhne, and Krzysztof Czarnecki. Fjmp: Factorized joint multi-agent motion prediction over learned directed acyclic interaction graphs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13745–13755, 2023.

[Rubagotti *et al.*, 2022] Matteo Rubagotti, Inara Tusseyeva, Sara Baltabayeva, Danna Summers, and Anara Sandygulova. Perceived safety in physical human–robot interaction—a survey. *Robotics and Autonomous Systems*, 151:104047, 2022.

[Saadatnejad *et al.*, 2022] Saeed Saadatnejad, Mohammadhossein Bahari, Pedram Khorsandi, Mohammad Saneian, Seyed-Mohsen Moosavi-Dezfooli, and Alexandre Alahi. Are socially-aware trajectory prediction models really socially-aware? *Transportation research part C: emerging technologies*, 141:103705, 2022.

[Scheunemann *et al.*, 2020] Marcus M Scheunemann, Raymond H Cuijpers, and Christoph Salge. Warmth and competence to predict human preference of robot behavior in physical human-robot interaction. In *IEEE RO-MAN*, 2020.

[Schwarting *et al.*, 2019] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, pages 24972–24978, 2019.

[Sun *et al.*, 2023] Jianhua Sun, Yuxuan Li, Liang Chai, and Cewu Lu. Modality exploration, retrieval and adaptation for trajectory prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Tang and Salakhutdinov, 2019] Charlie Tang and Russ R Salakhutdinov. Multiple futures prediction. *Advances in neural information processing systems*, 32, 2019.

[Tian *et al.*, 2024] Chunlin Tian, Zhan Shi, Zhijiang Guo, Li Li, and Chengzhong Xu. Hydralora: An asymmetric lora architecture for efficient fine-tuning, 2024.

[Wang *et al.*, 2023] Renzhi Wang, Senzhang Wang, Hao Yan, and Xiang Wang. Wsip: Wave superposition inspired pooling for dynamic interactions-aware trajectory prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4685–4692, 2023.

[Xie *et al.*, 2021] Xu Xie, Chi Zhang, Yixin Zhu, Ying Nian Wu, and Song-Chun Zhu. Congestion-aware multi-agent trajectory prediction for collision avoidance. In *ICRA*, 2021.

[Yin *et al.*, 2021] Ziyi Yin, Ruijin Liu, Zhiliang Xiong, and Zejian Yuan. Multimodal transformer networks for pedestrian trajectory prediction. In *IJCAI*, 2021.

[Zhao *et al.*, 2019] Tianyang Zhao, Yifei Xu, Mathew Monfort, Wongun Choi, Chris Baker, Yibiao Zhao, Yizhou Wang, and Ying Nian Wu. Multi-agent tensor fusion for contextual trajectory prediction. In *Proceedings of the IEEE/CVF CVPR*, 2019.

[Zuckerman, 1990] Marvin Zuckerman. The psychophysiology of sensation seeking. *Journal of personality*, 58(1):313–345, 1990.

[Zuo *et al.*, 2023] Zhiqiang Zuo, Xinyu Wang, Songlin Guo, Zhengxuan Liu, Zheng Li, and Yijing Wang. Trajectory prediction network of autonomous vehicles with fusion of historical interactive features. *IEEE Transactions on Intelligent Vehicles*, 2023.