Dual-View Interaction-Aware Lane Change Prediction for Autonomous Driving

Yuhuan Lu[1,2], Zhen Zhang[1], Rufan Bai[3], Han Liu[4], Wei Wang[1,5*]

[1]Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University
[2]Department of Computer and Information Science, University of Macau
[3]Department of Computer Science and Engineering, Southeast University
[4]School of Software, Dalian University of Technology
[5]School of Medical Technology, Beijing Institute of Technology
yc17462@um.edu.mo, zhangzhen19@lzu.edu.cn, rfbai@seu.edu.cn, liu.han.dut@gmail.com, ehomewang@ieee.org

Abstract

As artificial intelligence techniques evolve, we are approaching a critical moment for the widespread deployment of autonomous vehicles. Subsequently, the emergence of mixed-autonomy traffic environments presents formidable challenges to autonomous vehicles, especially for the accurate prediction of lane change intentions of their surrounding human-driven vehicles, which is crucial for ensuring the safety of autonomous vehicles. Existing lane change prediction models mainly focus on capturing the temporal variations in the movement dynamics of individual vehicles. However, the neglect to consider inter-vehicle interactions hinders their capability in complex lane change scenarios, resulting in suboptimal prediction performance. Moreover, current interaction-aware approaches for autonomous driving fail to explicitly model future interactions between vehicles, leading to unreasonable prediction results that can cause collisions between vehicles. To address the above issues, we propose to incorporate the concept of perceived safety into future interaction modeling and design a dual-view interaction-aware lane change prediction model. We evaluate the proposed model on two real-world datasets and experimental results show that the proposed model achieves average improvements of 11.7-12.4% in classification ability and 75.6-95.7% in forecast ability over the best-performing baselines across the two datasets. The ablation study and investigation into future interaction modeling demonstrate that our model has advantages in interpreting lane change scenarios from a driving safety perspective.

Introduction

With the burgeoning development of artificial intelligence, we are prepared for the wide-scale deployment of autonomous vehicles (AVs). However, the new traffic landscape brought about by the mixed-autonomy traffic scenarios—where AVs and human-driven vehicles (HDVs) coexist on roads—poses a formidable challenge to driving safety and decision-making for AVs. As an extremely frequent driving behavior of HDVs, the lane change maneuver significantly impacts road traffic safety, resulting in over 50,000 accidents in the U.S. each year (Ding et al. 2023). Accordingly, accurate lane change intention prediction of HDVs is paramount for the successful establishment of AVs.

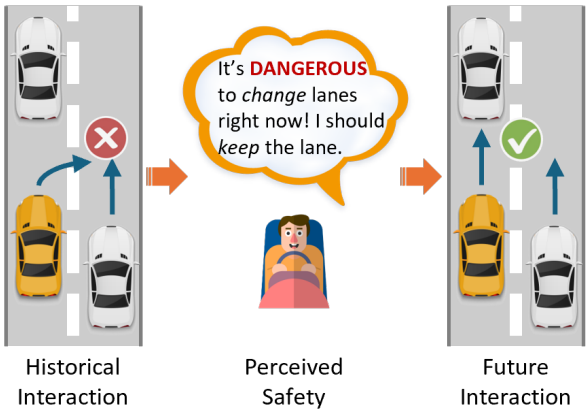


Figure 1: The advantage of modeling the future interaction with perceived safety. Blue arrows denote the predicted intention of vehicles under different interaction settings.

Despite the flourishing of deep learning, existing studies are apt to capture the temporal interdependence in the moving dynamics of individual vehicles to predict their lane change intentions (Shi and Zhang 2021; Xing et al. 2020; Liao et al. 2022). Such a philosophy neglects the interactions between vehicles, which have a direct influence on the motion of contiguous vehicles, thereby yielding suboptimal prediction performance. As illustrated in studies such as (Liang et al. 2020) and (Jia et al. 2023), inter-vehicle interactive activities serve as strong clues for extracting moving patterns of vehicles, and the acquired patterns are conducive to behavior prediction. This spurs us to model inter-vehicle interactions (the mutual influence on the movements of adjacent vehicles) for lane change prediction.

Recent interaction-aware approaches primarily concentrate on comprehensively and appropriately embedding complicated interactions using advanced models based on observed trajectories and driving environments. They either employ Graph Neural Networks (GNNs) to represent the geometric interactions between vehicles (Liang et al. 2020; Gu, Sun, and Zhao 2021; Jia et al. 2023) or utilize the self-attention mechanism to accommodate holistic inter-vehicle

*Corresponding author.

interactive behaviors (Zhou et al. 2022; Hu et al. 2023; Liao et al. 2024b). Although the interaction features learned by these two paradigms achieve commendable performance in motion predictions, we argue that these methods fail to consider possible future interactions between vehicles, which are crucial in lane change scenarios. As shown in Figure 1, the predicted lane change intention of the yellow vehicle can induce a potential collision with its adjacent vehicle if only extracting inter-vehicle interactions from historical trajectories. This confines the prediction to lagged-behind information, which is referred to as modeling historical interactions in this study. Consequently, the crux of the current interaction-aware methods lies in the excessive focus on ***historical interaction*** modeling while disregarding the explicit reflection of plausible ***future interaction***.

Against this backdrop, we resort to the concept of perceived safety for representing future interactions. As elaborated in (Rubagotti et al. 2022; Geisslinger, Poszler, and Lienkamp 2023; Liao et al. 2024b), perceived safety is the embodiment of both instinctive response and cognitive thinking, which can sculpt rational driving behaviors and the judicious decision-making process of human drivers. As portrayed in Figure 1, incorporating the perceived safety of drivers to forge future interactions allows for more sensible intention predictions, thus averting possible conflicts and preferably aligning with real-world situations.

In this context, we conceive a dual-view interaction-aware lane change prediction model that combines and embeds both future and historical interactions. By doing this, we aim to not only fortify the capability of accurately predicting lane change intentions but also empower the model to interpret lane change scenarios within the cognitive framework of driving safety. Our contributions can be summarized as follows:

1. **Dual-View Interaction Philosophy.** We revisit the problem of lane change prediction and discover the key limitation in current approaches: the failure to model inter-vehicle interactions, particularly the ignorance of future interactions. Therefore, we propose to model dual-view interactions in this study, encompassing both future and historical interactions. By incorporating the concept of perceived safety to elicit future interactions, we aim to achieve more accurate and socially informed prediction results.
2. **Tailored Model Components.** We design three modules to capture future and historical interactions and to blend the embedded dual-view interaction information. In the Future Interaction Module, four risk metrics are developed to characterize the spatio-temporal perception of driving safety. In the Historical Interaction Module, a novel Graph Transformer is devised to capture the geometric correlations between vehicles. In the Interaction Fusion Module, an edge-biased self-attention layer is contrived to prompt the meld of interaction features while improving the discrimination of connections between different interaction characteristics.
3. **Complete Effectiveness Verification.** We conduct a thorough evaluation of the proposed model to demonstrate its effectiveness on two large-scale real-world datasets. The results show that our model achieves average improvements of 11.7-12.4% in classification ability and 75.6-95.7% in forecast ability over the best-performing baselines across the two datasets. Furthermore, we investigate the Future Interaction Module and corroborate its accomplishment in safety-aware prediction and noise-resistant robustness.

Related Work

Lane Change Prediction. Recurrent Neural Networks (RNN) and their variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have been employed to accommodate the temporal variations in driving features, yielding favorable lane change intention prediction performance (Shi and Zhang 2021; Xing et al. 2020; Do, Han, and Choi 2023; Yuan et al. 2023). To further enhance the ability of RNNs to capture lane change patterns, hierarchical learning is introduced to complement prediction information presented in different lane change scenarios (Wang et al. 2024; Liao et al. 2022; Li et al. 2024). Transformer, a groundbreaking architecture for natural language processing, has extended its applications to computer vision and autonomous driving (Vaswani et al. 2017). It exhibits strong competence in capturing the correlations between input elements with the attention mechanism. Inspired by this, Transformer has also been applied in lane change prediction, resulting in better performance than the RNN family (Gao et al. 2023). However, the above methods hardly consider modeling interactions between vehicles.

Interaction Representation. Representing interactions as a graph and employing GNNs to facilitate information propagation and capture node-level correlations is prevalent in the literature (Liang et al. 2020; Gu, Sun, and Zhao 2021; Jia et al. 2023). Recently, some studies have managed to introduce the self-attention mechanism in Transformers to capture global relationships among vehicles, thus obtaining interaction embeddings with rich context. Moreover, the emergence of generative models provides the potential for robust interaction representation (Liao et al. 2024a). However, the above approaches fail to consider the future interaction, thus precipitating immoderate predictions.

Problem Formulation

In this study, we aim to build a lane change prediction model for an autonomous vehicle (referred to as the *ego vehicle*) that can forecast the short-term lane change intentions of *target vehicles* in the vicinity of the ego vehicle. Assuming that the current time is t, we denote the historical states of target vehicles and their surrounding vehicles by $\mathcal{X}_{tv}^{t-T_h+1:t} = \left\{ \mathbf{X}_{tv,i}^{t-T_h+1:t} | i \in [1,n] \right\}$ and $\mathcal{X}_{sv}^{t-T_h+1:t} = \left\{ \mathbf{X}_{sv,i}^{t-T_h+1:t} | i \in [1,m] \right\}$, respectively. Here, T_h indicates the traceback time window; n and m refer to the total number of target and surrounding vehicles, respectively. More specifically, the historical states can be directly represented

by:

$$\mathbf{X}_{*,i}^{t-T_h+1:t} = \left\{ \mathbf{p}_{*,i}^{t-T_h+1:t}, \mathbf{v}_{*,i}^{t-T_h+1:t}, \mathbf{a}_{*,i}^{t-T_h+1:t} \right\} \quad (1)$$

where $* = tv$ or sv. $\mathbf{p}_{*,i}^{t-T_h+1:t}$, $\mathbf{v}_{*,i}^{t-T_h+1:t}$, and $\mathbf{a}_{*,i}^{t-T_h+1:t}$ denote the time series of 2D coordinates, velocities, and accelerations of vehicle i, respectively. Through capturing intricate relationships among historical states, the proposed lane change prediction model outputs the predicted lane change intentions of target vehicles over a future duration T_f:

$$\begin{aligned} \mathbf{Y}_{tv}^{t+1:t+T_f} &= \left\{ \mathbf{y}_i^{t+1:t+T_f} | i \in [1, n] \right\} \\ \mathbf{y}_i^{t+1:t+T_f} &= \left\{ y_{i,1}^{t+1:t+T_f}, y_{i,2}^{t+1:t+T_f}, y_{i,3}^{t+1:t+T_f} \right\} \end{aligned} \quad (2)$$

where $y_{i,1}^{t+1:t+T_f}$, $y_{i,2}^{t+1:t+T_f}$, and $y_{i,3}^{t+1:t+T_f}$ correspond to the likelihood of three intention categories: lane keeping, left lane change, and right lane change, respectively. Notably, $\sum_{c=1}^{3} y_{i,c}^{t+1:t+T_f} = 1$.

Lane Change Prediction

The overall framework of the proposed lane change prediction model is shown in Figure 2. It comprises three sophisticated and interconnected modules, subtly capturing and blending future and historical interactions. Afterward, a novel intention classifier is curated to simultaneously forecast the likelihood of lane keeping, left lane change, and right lane change.

Future Interaction Module

In this module, we harness the perceived safety of human drivers to characterize future interactions between vehicles. In the traffic safety domain, perceived safety can be quantified by three risk metrics (Minderhoud and Bovy 2001): Time-to-Collision (TTC), Time Exposed Time-to-Collision (TET), and Time Integrated Time-to-Collision (TIT). These three metrics delineate the one-to-one correlation between any pair of vehicles but fail to accommodate the varying number of surrounding vehicles for a target vehicle. To this end, we ameliorate the three metrics:

1) Time-to-Collision (TTC): TTC quintessentially denotes the time remaining until two vehicles collide if they perpetuate their current trajectories. It furnishes a gauge of impending risk and undertakes the functionality of early warning. TTC for vehicle i with respect to vehicle j at time t is computed by:

$$\psi_{TTC_{i,j}^t} = -\frac{d_{i,j}^t}{\hat{d}_{i,j}^t} \quad (3)$$

$d_{i,j}^t$ denotes the distance between vehicles i and j and $\hat{d}_{i,j}^t$ is the change rate of $d_{i,j}^t$:

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

where (p_i^t, v_i^t) and $\left(p_j^t, v_j^t\right)$ are 2D coordinates and velocities of vehicles i and j, respectively. To account for the erratic number of surrounding vehicles, aggregated TTC for vehicle i is calculated with the learnable operation:

$$\psi_{TTC_i^t} = -\sum_{j \in \mathcal{N}_i^t} \alpha_{i,j}^t \psi_{TTC_{i,j}^t} \quad (5)$$

where $\mathcal{N}_i^t$ refers to the set of surrounding vehicles for vehicle i and $\alpha_{i,j}^t$ is the attention weight of vehicle j to vehicle i learned by Graph Attention Networks (GAT) (Veličković et al. 2017).

2) Time Exposed Time-to-Collision (TET): TET weighs the time of exposure to safety-critical TTC values over a specified duration t_H:

$$\psi_{TET_i^t} = \sum_{t_k=t-t_H+1}^{t} \delta_i\left(t_k\right) \cdot \tau_{\text{sc}} \quad (6)$$

where τ_{sc} is set to 0.1s, signifying the minimum time step in which the measured TTC values do not change. $\delta_i\left(t_k\right)$ is a switching variable:

$$\delta_i(t_k) = \begin{cases} 1, & \forall 0 \leq \psi_{TTC_i^{t_k}} \leq \psi_{TTC^*} \\ 0, & \text{else} \end{cases} \quad (7)$$

where ψ_{TTC^*} denotes the threshold safety-critical TTC value, which is set to 2.5s in this study.

3) Time Integrated Time-to-Collision (TIT): TET can not distinguish between diverse risk levels, as it treats all TTC values within the threshold equally. Thus, TIT amalgamates the TTC profiles to preferably represent the degree of safety:

$$\begin{aligned} \psi_{TIT_i^t} = \sum_{t_k=t-t_H+1}^{t} & \left[\psi_{TTC^*} - \psi_{TTC_i^{t_k}}\right] \cdot \tau_{\text{sc}} \\ & \forall 0 \leq \psi_{TTC_i^{t_k}} \leq \psi_{TTC^*} \end{aligned} \quad (8)$$

Increased values of TTC, TET, and TIT suggest a higher exposure to underlying collision risks, resulting in an attenuation in perceived safety.

4) Spatial Risk Perception Index (SRP): The above three metrics evaluate the perceived safety from a temporal perspective. To further capture the spatial distribution of risk levels, we propose SRP by modifying the *Moran's I* (Li, Calder, and Cressie 2007), a prominent spatial autocorrelation measurement:

$$\psi_{SRP_i^t} = \frac{|\mathcal{N}_i^t|}{D_i^t} \frac{\sum_{j \in \mathcal{N}_i^t} \sum_{q \in \mathcal{N}_i^t} \left(d_{j,q}^t \psi_{TTC_{i,j}^t} \psi_{TTC_{i,q}^t}\right)}{\sum_{j \in \mathcal{N}_i^t} \psi_{TTC_{i,j}^t}^2} \quad (9)$$

In this context, $\psi_{SRP_i^t}$ with larger values indicates a less even spatial distribution of collision risks, thereby elevating the perceived safety for performing lane change maneuvers. In addition, D_i^t is the sum of the distances between surrounding vehicles of vehicle i:

$$D_i^t = \sum_{j \in \mathcal{N}_i^t} \sum_{q \in \mathcal{N}_i^t} d_{j,q}^t \quad (10)$$

To mitigate the repercussions of numerical differences on lane change prediction performance, we separately apply min-max normalization to each of the four perceived

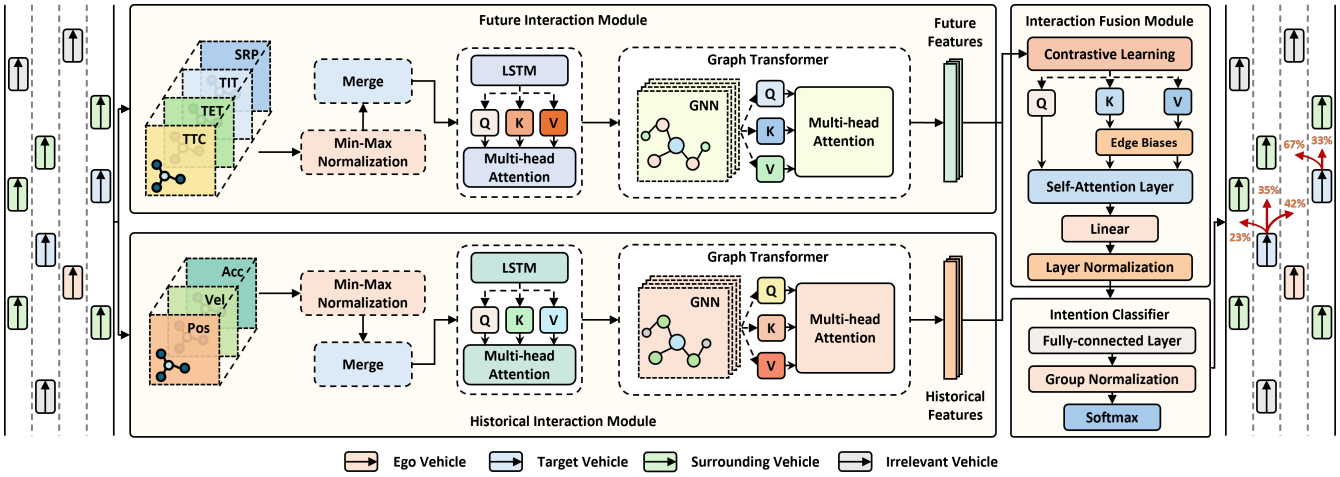


Figure 2: Overall framework of the proposed lane change prediction model.

safety metrics and then merge them into input feature matrices for target vehicles, $\left\{\bar{\mathbf{X}}_{tv,i}^{t-T_h+1:t} | i \in [1,n]\right\}$. Afterward, an LSTM with multi-head attention (Vaswani et al. 2017) is developed to distill the temporal information. We first employ LSTM to obtain the hidden state of each target vehicle, presenting the essence of perceived safety toward future interactions:

$$\mathbf{H}_{F,i}^{t-T_h+1:t} = \varphi_{LSTM}\left(\varphi_{MLP}\left(\bar{\mathbf{X}}_{tv,i}^{t-T_h+1:t}\right)\right) \tag{11}$$

Subsequently, a multi-head attention mechanism is imposed on the derived hidden state to acquire the refined feature vector:

$$\dot{\mathbf{x}}_{tv,i}^{t} = \varphi_{LN}\left(\varphi_{MHA}\left(\mathbf{H}_{F,i}^{t-T_h+1:t}\right)\right) \tag{12}$$

where $\varphi_{LN}(\cdot)$ denotes Layer Normalization (Ba, Kiros, and Hinton 2016) exerted to bolster the training stability. The multi-head attention mechanism can be specified by:

$$\varphi_{MHA}\left(\mathbf{H}_{F,i}^{t-T_h+1:t}\right) = \varphi_{Concat}\left(\mathbf{H}_{F,i}^{t-T_h+1:t}\beta_j|_{j=1}^{h}\right)\mathbf{W}_F \tag{13}$$

where $\beta_j|_{j=1}^{h}$ represents a group of attention weights corresponding to different attention heads, transformed from query, key, and value vectors. $\mathbf{W}_F$ is the learnable parameter.

After attaining the refined feature vectors for target vehicles $\{\dot{\mathbf{x}}_{tv,i}^{t} | i \in [1,n]\}$, we propose to further capture spatial correlations between target vehicles, encoding the deep-seated geometric interdependence. A driving graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is first constructed, where $\mathcal{V}$ denotes the node set corresponding to target vehicles and $\mathcal{E}$ denotes the edge set connecting the nodes. To propagate information among nodes and edges and update their corresponding features, we devise a Graph Transformer that incorporates the multi-head attention mechanism into the message-passing process of a GNN. Specifically, we renovate the *aggregation function* and *update function* in the GNN:

1) Aggregation function: For a node $v \in \mathcal{V}$ in the l-th layer, the input node feature is denoted by $\mathbf{v}^{l-1}$ from the previous layer. Especially, the input to the first layer is the refined feature vectors obtained in Equation (12). We denote the features of neighboring nodes of v that belong to vehicle type j by $\mathbf{V}_j^{l-1}$. The attention score of this specific type for v is calculated by:

$$\mathbf{a}_j = \varphi_{Softmax}\left(\frac{\mathbf{q}_j \mathbf{K}_j}{\sqrt{d_h}}\right) \tag{14}$$

where $\mathbf{q}_j = \mathbf{v}^{l-1}\mathbf{W}_{Q_j}$, $\mathbf{K}_j = \mathbf{V}_j^{l-1}\mathbf{W}_{K_j}$. $\mathbf{W}_{Q_j}$ and $\mathbf{W}_{K_j}$ are the query and key linear transformation matrices for vehicle type j, respectively. In doing so, the attention score becomes type-aware, embracing the prior knowledge of vehicle types. Further, we conduct multi-head attention to collect features of neighboring nodes with various vehicle types for v, resulting in a set of aggregated features $\left\{\mathbf{v}_j^{l-1,\prime} | j \in [1,n_v]\right\}$, where n_v is the number of possible types for v. These aggregated features can be further integrated by:

$$\mathbf{v}^{l-1,\prime} = \varphi_{MLP}\left(\varphi_{Concat}\left(\left\{\mathbf{v}_j^{l-1,\prime} | j \in [1,n_v]\right\}\right)\right) \tag{15}$$

2) Update function: We leverage the Feed-Forward Network (FFN) to update the integrated feature:

$$\mathbf{v}^{l} = \varphi_{LN}\left(\varphi_{FFN}\left(\mathbf{v}^{l-1,\prime}\right)\right) \tag{16}$$

Note that we utilize the residual connection (He et al. 2016) between layers to enhance network performance. Through the multi-layer Graph Transformer, high-order geometric relationships with type-aware semantics are extracted, and the final future interaction features for target vehicles are read out from the last layer of the Graph Transformer:

$$\mathbf{X}_{tv,F}^{t} = \left\{\mathbf{x}_{F,i}^{t} | i \in [1,n]\right\} \tag{17}$$

Historical Interaction Module

The historical driving states of surrounding vehicles are momentous for the decision-making of target vehicles in the near future. For instance, vehicles located in adjacent lanes

near the target vehicle may impede the target vehicle's discretionary lane change behaviors. To adapt to such intricate interplay, we introduce the Historical Interaction Module. This sophisticated module converts the driving states of vehicles into high-dimensional interaction features with rich context, capturing the concealed spatio-temporal correlations.

To incorporate the historical states of surrounding vehicles while retaining the distinctiveness of the target vehicle states, we design a local-aware min-max normalization to epitomize the relative states of the target vehicles. For clarity, the relative position of target vehicle i at time t is calculated by:

$$\widetilde{p}_i^t = \frac{p_i^t - \min\left(\left\{p_j^t | j \in \mathcal{N}_i^t\right\}\right)}{\max\left(\left\{p_j^t | j \in \mathcal{N}_i^t\right\}\right) - \min\left(\left\{p_j^t | j \in \mathcal{N}_i^t\right\}\right)} \tag{18}$$

where $\mathcal{N}_j^t$ refers to the surrounding vehicles of i. The operator is implemented on position, velocity, and acceleration across all historical time steps, fulfilling the refined local-aware features for target vehicles. Then the obtained features are fed into a similar branch to that in the Future Interaction Module: an LSTM with multi-head attention and a Graph Transformer to capture the temporal and geometric context, respectively. Finally, the compiled historical interaction features for target vehicles are represented by:

$$\mathbf{X}_{tv,H}^t = \left\{\mathbf{x}_{H,i}^t | i \in [1, n]\right\} \tag{19}$$

Interaction Fusion Module

This module combines the obtained future and historical interaction features, resulting in populated and enriched interaction fusion features for lane change intention prediction. However, the fusion process lacks bound constraints, indicating the gradient engendered by the final prediction results can only lead to minimal parameter learning. Accordingly, we resort to contrastive learning between different target vehicles as an additional learning objective for regularization. We randomly sample n_{CL} target vehicles as a minibatch and define the contrastive loss function as:

$$\begin{aligned}\mathcal{L}_{CL} = &\frac{1}{n_{CL}^2} \sum_{i=1}^{n_{CL}} \sum_{j=1}^{n_{CL}} \left(\varphi_{Sim}\left(\mathbf{x}_{F,i}^t, \mathbf{x}_{H,i}^t\right)\right. \\ &\left. -\varphi_{Sim}\left(\mathbf{x}_{F,i}^t, \mathbf{x}_{H,j}^t\right) + 2\right)\end{aligned} \tag{20}$$

The contrastive loss aims to confine the similarity of positive samples to be greater than that of negative samples, thereby entrenching the interaction fusion. The similarity function is defined as:

$$\varphi_{Sim}\left(\mathbf{u}, \mathbf{z}\right) = -\frac{\mathbf{u}^{\mathrm{T}}\mathbf{z}}{\|\mathbf{u}\|\|\mathbf{z}\|} \tag{21}$$

For each target vehicle i, the comprehensive interaction feature is computed by:

$$\mathbf{x}_{I,i}^t = \varphi_{Concat}\left(\left\{\mathbf{x}_{F,i}^t, \mathbf{x}_{H,i}^t\right\}\right) \tag{22}$$

To further capture the global correlations between interaction features, we introduce an edge-biased self-attention layer. The interaction feature for target vehicle i is updated by:

$$\gamma_{j,o} = \frac{\left(\mathbf{W}_{Q_I}\mathbf{x}_{I,j}^t + \mathbf{b}_{j,o}^Q\right)^\top \left(\mathbf{W}_{K_I}\mathbf{x}_{I,o}^t + \mathbf{b}_{j,o}^K\right)}{\sqrt{d_e}} \tag{23}$$

$$\breve{\mathbf{x}}_{I,i}^t = \sum_{j=1}^{n} \frac{\exp\left(\gamma_{i,j}\right)}{\sum_{k=1}^{n}\exp\left(\gamma_{i,k}\right)} \left(\mathbf{W}_{V_I}\mathbf{x}_{I,j}^t + \mathbf{b}_{i,j}^V\right) \tag{24}$$

where $\mathbf{W}_{Q_I}$, $\mathbf{W}_{K_I}$, and $\mathbf{W}_{V_I}$ are linear transformation matrices of query, key, and value, respectively. $\mathbf{b}_{i,j}^Q$, $\mathbf{b}_{i,j}^K$, and $\mathbf{b}_{i,j}^V$ are edge biases used to discriminate connections between different interaction characteristics. Thereafter, the interaction features culminate in the channel-wide synthesis:

$$\acute{\mathbf{x}}_{I,i}^t = \varphi_{LN}\left(\varphi_{MLP}\left(\breve{\mathbf{x}}_{I,i}^t\right)\right) \tag{25}$$

Intention Classifier

With the obtained final interaction features for target vehicles $\acute{\mathbf{X}}_{tv,I}^t = \left\{\acute{\mathbf{x}}_{I,i}^t | i \in [1, n]\right\}$, the lane change prediction is framed as a classification task. The intention likelihood of target vehicle i is computed by:

$$\omega_i^t = \varphi_{Softmax}\left(\varphi_{GN}\left(\varphi_{MLP}\left(\acute{\mathbf{x}}_{I,i}^t\right)\right)\right) \tag{26}$$

where $\varphi_{GN}(\cdot)$ denotes the Group Normalization (Wu and He 2018) capitalized on for strengthening the training stability. With the parameter-sharing scheme, the intention likelihood of multiple target vehicles can be simultaneously generated. Finally, the cross-entropy loss function is employed to optimize model parameters:

$$\mathcal{L}_{IC} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{3} \bar{y}_{i,c}^{t+1:t+T_f} \log \omega_{i,c}^t \tag{27}$$

where $\bar{y}_{i,c}^{t+1:t+T_f}$ is the ground-truth label for the c-th intention category over the prediction horizon T_f, and $\omega_{i,c}^t$ is the predicted likelihood of the c-th intention category.

To hasten model training, we employ a multi-task training strategy by alternating the training procedures of intention classification and contrastive learning. The corresponding parameters in the two pipelines are updated alternatively until the intention classification pipeline reaches convergence.

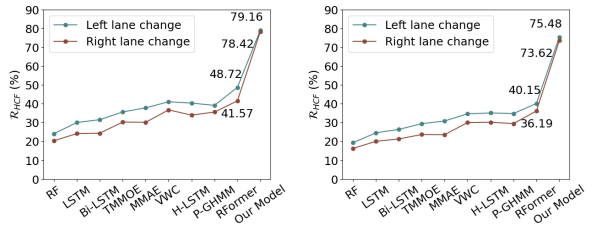
Experiments

Experimental Setup

We evaluate the proposed model on two large-scale real-world datasets, NGSIM (Deo and Trivedi 2018) and HighD (Krajewski et al. 2018). Referring to current studies (Gao et al. 2023), the lane change intention triggers 2s before lane change execution. Accordingly, we label the duration of 2s before and 4s after the start of the lane change maneuver as a lane change sample (Li et al. 2024). In this context, we obtain 10,126 lane change/lane keeping samples for the NGSIM dataset and 15,432 lane change/lane keeping samples for the HighD dataset. Among these samples, 80% are used for model training whereas the remaining 20% are used

Method	NGSIM Dataset				HighD Dataset			
	Left Lane Change		Right Lane Change		Left Lane Change		Right Lane Change	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
RF	69.24%	68.37%	65.33%	63.64%	59.78%	60.42%	56.32%	58.17%
LSTM	77.41%	75.68%	70.36%	72.03%	70.31%	68.25%	67.42%	65.99%
Bi-LSTM	79.87%	80.31%	75.24%	73.54%	76.45%	75.27%	71.12%	71.53%
TMMOE	78.32%	80.34%	75.21%	76.94%	76.75%	76.37%	72.69%	72.56%
MMAE	82.17%	78.65%	76.39%	82.31%	78.42%	79.27%	74.63%	77.65%
VWC	84.31%	86.54%	80.62%	84.11%	80.37%	81.27%	78.46%	79.53%
H-LSTM	80.29%	85.74%	82.64%	86.75%	82.57%	81.30%	78.24%	79.07%
P-GHMM	83.26%	83.69%	85.18%	86.72%	81.45%	83.64%	83.37%	81.09%
RFormer	88.46%	86.38%	87.10%	88.35%	88.37%	86.59%	85.10%	84.99%
Our Model	**97.62%**	**98.33%**	**97.27%**	**97.98%**	**97.17%**	**97.02%**	**96.86%**	**96.72%**

Table 1: Classification ability comparison on NGSIM and HighD datasets. The underline and **bold** values denote the best results in baselines and overall, respectively.



(a) NGSIM dataset. (b) HighD dataset.

Figure 3: Forecast ability comparison on NGSIM and HighD datasets.

for assessment. Note that NGSIM and HighD have 3 and 2 vehicle types, respectively. Precision and Recall are employed to appraise the classification ability. To further measure the forecast ability, specifically how early the model can detect the lane change intention of a target vehicle, we propose $\mathcal{R}_{HCF}$:

$$\mathcal{R}_{HCF} = \frac{1}{N_{LC}} \sum_{i=1}^{N_{LC}} \delta_{HCF}(R_i)$$
$$\delta_{HCF}(R_i) = \begin{cases} 1, & \text{if } R_i \geq 80\% \\ 0, & \text{else} \end{cases} \tag{28}$$

where R_i is the ratio of the period from the first correctly predicted timestamp to the end of the lane change to 6s. N_{LC} is the number of lane change samples in the test set. Moreover, we instantiate the proposed model on benchmark hardware (Intel Xeon5320@2.20GHz, 512GB RAM@4800Hz, NVIDIA GeForce RTX 4090 24GB, Ubuntu 21.04), implemented using Adam optimizer with a batch size of 64 and a learning rate ranging from 10^{-3} to 10^{-5}. Three key hyperparameters in our model are set up: the number of Graph Transformer layers is set to 6; the number of edge-biased self-attention layers is set to 4; and the embedding dimension is set to 256. We compare our model with a range of state-of-the-art baselines: RF (Sun et al. 2021); LSTM (Shi and Zhang 2021); Bi-LSTM (Xing et al. 2020); TMMOE (Yuan et al. 2023); MMAE (Do, Han, and Choi 2023); VWC (Wang et al. 2024); H-LSTM (Liao et al. 2022); P-GHMM (Li et al. 2024); RFormer (Gao et al. 2023). Notably, the results for the lane keeping category are omitted in the experiments due to the page limitation. However, this category is still included in the calculation of overall prediction performance in the following.

Overall Performance

Table 1 exhibits the classification performance on two datasets and we highlight the best-performing result for each task on each dataset. We observe that our model consistently outperforms all baselines, achieving an 11.7-12.4% improvement on average over the best-performing baselines across the two datasets. Notably, the interaction-aware baseline (RFormer) also surpasses other baselines in most tasks. The above observations further evidence that capturing intricate interactions between vehicles is instrumental in lane change prediction. Moreover, the classification performance of our model on HighD remains comparable to that on NGSIM, further manifesting the robustness of dual-view interaction modeling in lane change prediction.

We also investigate the forecast ability of our model, with the results shown in Figure 3. Our model consistently outperforms all baselines, yielding an average improvement of 75.6-95.7% over the best-performing baselines across the two datasets. The tremendous progress is also attributed to the design ethos of dual-view interactions and exquisite model components. We discover that RFormer is still the best-performing baseline, indicating that interaction representation is also the hinge of raising forecast ability in lane change prediction. These observations confirm that our model effectively serves as an early-warning indicator for possible lane change maneuvers, further enhancing the safety of autonomous driving.

Ablation Study

We conduct a thorough ablation study to examine the effectiveness of individual components of our model in lane change prediction. The ablation results are presented in Table 2. We consider three variants to quantify their impact on classification performance. Specifically, *w/o FI* removes

Method	NGSIM Dataset				HighD Dataset			
	Left Lane Change		Right Lane Change		Left Lane Change		Right Lane Change	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Our Model	**97.62%**	**98.33%**	**97.27%**	**97.98%**	**97.17%**	**97.02%**	**96.86%**	**96.72%**
w/o FI	89.04%	88.21%	87.34%	86.59%	87.46%	87.39%	86.21%	85.90%
w/o GT	92.77%	93.05%	92.86%	92.57%	92.38%	92.50%	91.87%	91.46%
w/o EB	95.03%	95.27%	95.10%	95.58%	94.93%	94.87%	94.33%	94.52%

Table 2: Ablation results for different variants on NGSIM and HighD datasets.

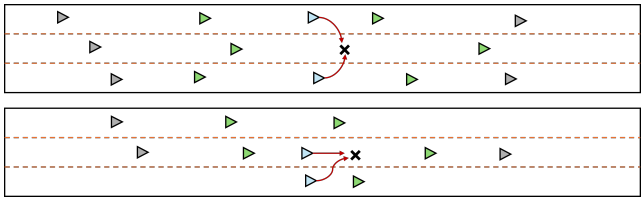


Figure 4: Realistic cases of safety-critical lane change scenarios from the HighD dataset. Blue, green, and grey triangles denote target, surrounding, and irrelevant vehicles, respectively. In such cases, the predicted intention of target vehicles can easily lead to conflict points.

Dataset	Model			
	P-GHMM	RFormer	*w/o FI*	**Our Model**
NGSIM	16%	17%	24%	**78%**
HighD	11%	10%	19%	**74%**

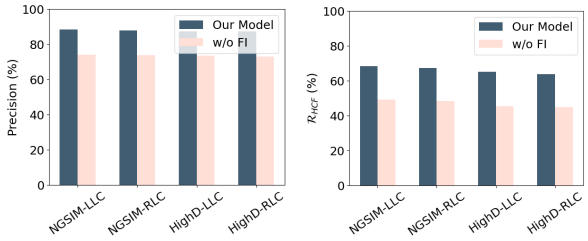
Table 3: Comparison of prediction performance in safety-critical scenarios.

the Future Interaction Module and suffers from a marked degradation in performance, with a decline of 10.5% on both datasets. This underlines the significance of incorporating future interactions for lane change prediction. *w/o GT* and *w/o EB* remove the Graph Transformer and edge biases, respectively. These two variants also parade a pronounced erosion of model performance, which signifies the critical role of both the Graph Transformer and edge biases in capturing and fusing interactions, respectively.

Investigation into Future Interaction

We display the efficacy of the Future Interaction Module in safety-aware prediction and noise-resistant robustness.

Safety-aware Prediction. Safety-critical scenarios are often deemed as corner cases for lane change prediction. As shown in Figure 4, the predicted intention of target vehicles can simply result in hazardous conflicts since current lane change prediction models are incapable of interpreting the psychological activities of human drivers. To more intuitively unveil the limitation, we manually label 100 safety-critical lane change samples in the test set of each dataset and compare the performance of different models on these samples. The prediction results are presented in Table 3. We discover that our model significantly surpasses baselines, with nearly 7x the prediction accuracy of the best-performing baselines on the HighD dataset. Additionally, we observe that the *w/o FI* variant only bears a slight improvement compared to the best-performing baselines. These observations indicate that modeling perceived safety in future interactions facilitates the understanding of safety-critical scenarios, implicitly enhancing the comprehension of social norms in real-world driving environments.



(a) Precision. (b) $\mathcal{R}_{HCF}$.

Figure 5: Robustness analysis of Future Interaction Module against data noise. LLC and RLC mean left and right lane changes, respectively.

Noise-resistant Robustness. Data noise is prevalent in autonomous driving, so we apply Gaussian noise to 10% of training data in each dataset to evaluate the robustness of our model. As shown in Figure 5, the ability of our model to resist data noise primarily stems from capturing future interactions. This implies that modeling perceived safety endows our model with a deliberate consideration of surrounding environments, enabling it to make sensible predictions.

Conclusion

In this study, we design a dual-view interaction-aware model aimed at accurate lane change intention prediction. We are the first to propose modeling future interactions with the concept of perceived safety. This design ethos not only boosts lane change prediction accuracy but also informs the model to socially interpret complex lane change scenarios. Extensive experiments on two real-world datasets demonstrate the advancement of our model. To sum up, our findings suggest promising potential for future research, representing future interaction with perceived safety and integrating future interaction into various behavior prediction models for more assured autonomous driving.

Acknowledgments

This study was funded in part by Natural Science Foundation of China under 52102400 and 62402103, in part by the Shenzhen Science and Technology Innovation Commission (Stabilisation Support Programme), and in part by Natural Science Foundation of Jiangsu Province under BK20241273.

References

Ba, J. L.; Kiros, J. R.; and Hinton, G. E. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450*.

Deo, N.; and Trivedi, M. M. 2018. Convolutional social pooling for vehicle trajectory prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 1468–1476.

Ding, H.; Lu, Y.; Sze, N.; Antoniou, C.; and Guo, Y. 2023. A crash feature-based allocation method for boundary crash problem in spatial analysis of bicycle crashes. *Analytic methods in accident research*, 37: 100251.

Do, J.; Han, K.; and Choi, S. B. 2023. Lane change–intention inference and trajectory prediction of surrounding vehicles on highways. *IEEE Transactions on Intelligent Vehicles*.

Gao, K.; Li, X.; Chen, B.; Hu, L.; Liu, J.; Du, R.; and Li, Y. 2023. Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment. *IEEE Transactions on Intelligent Transportation Systems*.

Geisslinger, M.; Poszler, F.; and Lienkamp, M. 2023. An ethical trajectory planning algorithm for autonomous vehicles. *Nature Machine Intelligence*, 5(2): 137–144.

Gu, J.; Sun, C.; and Zhao, H. 2021. Densetnt: End-to-end trajectory prediction from dense goal sets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 15303–15312.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hu, Y.; Yang, J.; Chen, L.; Li, K.; Sima, C.; Zhu, X.; Chai, S.; Du, S.; Lin, T.; Wang, W.; et al. 2023. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17853–17862.

Jia, X.; Wu, P.; Chen, L.; Liu, Y.; Li, H.; and Yan, J. 2023. Hdgt: Heterogeneous driving graph transformer for multi-agent trajectory prediction via scene encoding. *IEEE transactions on pattern analysis and machine intelligence*.

Krajewski, R.; Bock, J.; Kloeker, L.; and Eckstein, L. 2018. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *2018 21st international conference on intelligent transportation systems (ITSC)*, 2118–2125. IEEE.

Li, H.; Calder, C. A.; and Cressie, N. 2007. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geographical analysis*, 39(4): 357–375.

Li, Z.; Wang, Y.; Zuo, Z.; and Hu, C. 2024. A Hierarchical Intention Inference Model for Connected and Automated Vehicles. *IEEE Transactions on Intelligent Vehicles*.

Liang, M.; Yang, B.; Hu, R.; Chen, Y.; Liao, R.; Feng, S.; and Urtasun, R. 2020. Learning lane graph representations for motion forecasting. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, 541–556. Springer.

Liao, H.; Li, X.; Li, Y.; Kong, H.; Wang, C.; Wang, B.; Guan, Y.; Tam, K.; Li, Z.; and Xu, C. 2024a. Characterized Diffusion and Spatial-Temporal Interaction Network for Trajectory Prediction in Autonomous Driving. *arXiv preprint arXiv:2405.02145*.

Liao, H.; Li, Z.; Wang, C.; Wang, B.; Kong, H.; Guan, Y.; Li, G.; Cui, Z.; and Xu, C. 2024b. A Cognitive-Driven Trajectory Prediction Model for Autonomous Driving in Mixed Autonomy Environment. *arXiv preprint arXiv:2404.17520*.

Liao, X.; Wang, Z.; Zhao, X.; Zhao, Z.; Han, K.; Tiwari, P.; Barth, M. J.; and Wu, G. 2022. Online prediction of lane change with a hierarchical learning-based approach. In *2022 International Conference on Robotics and Automation (ICRA)*, 948–954. IEEE.

Minderhoud, M. M.; and Bovy, P. H. 2001. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention*, 33(1): 89–97.

Rubagotti, M.; Tusseyeva, I.; Baltabayeva, S.; Summers, D.; and Sandygulova, A. 2022. Perceived safety in physical human–robot interaction—A survey. *Robotics and Autonomous Systems*, 151: 104047.

Shi, Q.; and Zhang, H. 2021. An improved learning-based LSTM approach for lane change intention prediction subject to imbalanced data. *Transportation research part C: emerging technologies*, 133: 103414.

Sun, Q.; Wang, C.; Fu, R.; Guo, Y.; Yuan, W.; and Li, Z. 2021. Lane change strategy analysis and recognition for intelligent driving systems based on random forest. *Expert Systems with Applications*, 186: 115781.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Wang, X.; Zan, Y.; Zheng, J.; Wenjuan, E.; Tao, Y.; Xiao, Y.; and Li, T. 2024. Variable Weight Combination Model for Lane-changing Prediction of Human-Driven Vehicle. *IEEE Transactions on Intelligent Vehicles*.

Wu, Y.; and He, K. 2018. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, 3–19.

Xing, Y.; Lv, C.; Wang, H.; Cao, D.; and Velenis, E. 2020. An ensemble deep learning approach for driver lane change intention inference. *Transportation Research Part C: Emerging Technologies*, 115: 102615.

Yuan, R.; Abdel-Aty, M.; Xiang, Q.; Wang, Z.; and Gu, X. 2023. A Temporal Multi-Gate Mixture-of-Experts Approach for Vehicle Trajectory and Driving Intention Prediction. *IEEE Transactions on Intelligent Vehicles*.

Zhou, Z.; Ye, L.; Wang, J.; Wu, K.; and 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*, 8823–8833.