Lane Change Prediction for Autonomous Driving With Transferred Trajectory Interaction

Yuhuan Lu, Pengpeng Xu, Xinyu Jiang, Ali Kashif Bashir, *Senior Member, IEEE*, Thippa Reddy Gadekallu, *Senior Member, IEEE*, Wei Wang, *Member, IEEE*, and Xiping Hu

***Abstract*— In mixed-autonomy traffic environments, accurately predicting the lane change behavior of human-driven vehicles is critical for ensuring the safety and reliability of autonomous vehicle decision-making. However, existing approaches face two major challenges: 1) they tend to represent the relationships between the target vehicle and surrounding vehicles using parameters like relative position and speed. This approach either requires a fixed number of surrounding vehicles or introduces significant noise by relying on virtual vehicles; and 2) they often fail to fully exploit the vast amount of available vehicle trajectory data, leaving the complexities of vehicular interactions underexplored. To address these issues, this paper presents a novel lane change prediction framework using Transformer-based transfer learning. Our design aims to leverage inter-vehicle interactions learned from trajectory data to improve lane-change prediction accuracy. Specifically, pre-trained trajectory prediction models are used to adapt dynamically to the varying number of surrounding vehicles and to capture interaction context from large sets of trajectory data. We then refine the Transformer model to integrate this context and predict the target vehicle's lane change intentions. The Transformer encoder transforms trajectory interaction context into a lane-change-oriented context using aggregated multi-head attention. The Transformer decoder, in turn, utilizes this context alongside the target vehicle's states through relation-aware multi-head attention to forecast lane change behavior. Extensive experiments on two real-world datasets demonstrate that our proposed framework outperforms state-of-the-art baselines in both accuracy and robustness.**

***Index Terms*— Lane change prediction, transformer, transfer learning, attention mechanism, autonomous driving.**

Received 11 September 2023; revised 27 September 2024; accepted 9 February 2025. This work was supported in part by the Sustainable Development Science and Technology Project of Shenzhen Science and Technology Innovation Commission under Grant KCXFZ2020122-1173411032, in part by Shenzhen Science and Technology Innovation Commission (Stabilization Support Program), and in part by the Natural Science Foundation of Guangdong Province, China, under Project 2023A1515012404. The Associate Editor for this article was Q. Kang. *(Corresponding author: Wei Wang.)*

Yuhuan Lu is with Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, China, and also with the State Key Laboratory of Internet of Things for Smart City and the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: lu.yuhuan@connect.umac.mo).

Pengpeng Xu is with the School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510640, China (e-mail: pengpengxu@yeah.net).

Xinyu Jiang is with Natural Selection Intelligent Company Ltd., Shenzhen 518000, China (e-mail: jiangxinyu1@natureselect.ai).

Ali Kashif Bashir is with the Department of Computing and Mathematics, Manchester Metropolitan University, M15 6BX Manchester, U.K., and also with the Centre for Research Impact and Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab 140401, India (e-mail: dr.alikashif.b@ieee.org).

Thippa Reddy Gadekallu is with the College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China, also with the Division of Research and Development, Lovely Professional University, Phagwara 144001, India, and also with the Center of Research Impact and Outcome, Chitkara University, Rajpura, Punjab 140401, India (e-mail: Thippareddy@ieee.org).

Wei Wang and Xiping Hu are with Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, China, and also with the School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: ehomewang@ieee.org; huxp@bit.edu.cn).

I. INTRODUCTION

AUTONOMOUS driving is expected to be a revolutionary paradigm for intelligent transportation systems. Compared to human-driven vehicles, autonomous vehicles can greatly improve driving safety and operational efficiency in our daily traffic [1], [2], [3]. Nevertheless, it is inevitable that human-driven and autonomous vehicles will coexist on the roads in the near future. This poses a huge burden on the rational decision-making of autonomous vehicles in such mixed-traffic environments [4]. Therefore, to enable autonomous vehicles to make sensible decision-making beforehand, it is crucial to accurately predict the driving behaviors of surrounding vehicles [5], [6]. As one of the most essential driving behaviors, lane change behavior exacerbates traffic oscillations and increases the risk of traffic collisions. According to [7], around 10% of all traffic crashes are caused by lane change behavior. Thus, the accurate prediction of lane change intention of surrounding vehicles can provide vital information for autonomous vehicles to proactively circumvent possible collisions. By integrating lane change prediction into advanced driver-assistance systems (ADASs) [8], [9], autonomous driving can be safer, more efficient, and more comfortable.

As the significance of lane change prediction for autonomous driving is progressively recognized by both industry and academia, accurate lane change prediction has become a topic of great interest in research. Approaches for lane change intention prediction can be classified into four categories: model-based methods, generative methods, discriminative methods, and deep learning methods [10], [11]. Model-based methods rely on different kinds of driver models that describe multiple driving behaviors. A set of similarity metrics is then developed to select the most powerful model to perform the lane change prediction [12], [13], [14]. Although model-based methods have great interpretability and show the capability of stable prediction, they fail to account for the heterogeneity of drivers which results in poor prediction performance. With the rise of advanced machine learning and communication techniques, various generative and discriminative methods have emerged rapidly. Generative methods aim to model the interdependence between driving behavior features [15], [16], however, it is difficult to comprehensively represent the complex relationships between features by Bayesian networks. On the other hand, discriminative methods do not require explicitly modeling interdependence between features. They regard lane change prediction as a classification problem but the forecast performance is prone to noise and outliers inherent in data [17], [18]. In contrast with the above traditional methods, deep learning methods have become increasingly prevalent in lane change prediction due to their strong representation ability and favorable generalization performance. Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) are two typical deep learning models utilized to model the temporal dependence of vehicle trajectories. The deep-seated dependence features are subsequently harnessed to predict when the lane change maneuver is conducted [19], [20], [21]. In light of the tremendous achievement of attention mechanism in the area of pattern recognition [22], [23], some researchers have involved Transformer in lane change prediction [24], [25] to extract long-range time-series dependence and accomplish superior performance to RNN and LSTM.

Although deep learning methods have shown promising results in predicting lane change, there are still research gaps in current lane change prediction models [26], [27], [28]. First, most recent deep learning-based methods tend to model the relative relationships between vehicles, such as relative position and relative speed, among others. However, this representation of relative relationships requires a fixed number of surrounding vehicles [29], which impairs the capture of dynamic interactions between vehicles. To accommodate a more flexible interactional scenario, some research introduces virtual vehicles to complement the interaction information [25], yet the calculation of interaction is sensitive to the setting of virtual vehicles like position and speed. In particular, unreasonable settings will incur severe lane change prediction errors. Second, trajectory data is scarcely considered in present lane change prediction models. Vehicle trajectories contain rich context information, which is beneficial for the representation of interactions between vehicles [30], [31]. Nevertheless, present lane change prediction models have difficulty integrating the trajectory data. Since it is necessary to insert the trajectory data learning module into the lane change prediction model, training such a combined model from scratch is time-consuming and computationally exhaustive. Furthermore, due to the limited number of lane change scenarios, only a small portion of the trajectory data can be utilized for model training, which hinders the in-depth modeling of interactions between vehicles.

To address the above issues, we propose a novel lane change prediction framework based on Transformer-based transfer learning. This approach leverages transfer learning to overcome the limited availability of lane change trajectory data for modeling vehicular interactions. Rather than directly using raw trajectory data, we use transfer learning to incorporate only the learned interaction context into the lane change prediction model, capturing rich inter-vehicle interaction information while mitigating the effects of varying numbers of surrounding vehicles. Furthermore, the Transformer architecture, known for its ability to capture semantic relationships in long sequences [32], [33], is well-suited for handling the temporal dynamics in lane change scenarios. Specifically, we first pre-train trajectory models using a great number of vehicle trajectories. Inspired by the recent progress in vehicle trajectory prediction [34], [35], [36], high-order interactions between vehicles can be extracted from the large volume of trajectory data by Graph Neural Networks (GNNs). Therefore, cutting-edge GNN-based trajectory models are directly used to dynamically adapt to the varying number of surrounding vehicles and capture the vectorized representation of the interaction context. Afterward, a refined Transformer is devised to predict lane change behavior by absorbing the interaction context captured by pre-trained trajectory models. The derived context is fed into the Transformer encoder to obtain a fine-grained lane-change-oriented interaction context. To mitigate the under expression issue of the original multi-head attention in the encoder, an aggregated attention mechanism is developed to improve the representation ability. A relation-aware Transformer decoder is then invented to fuse the lane-change-oriented interaction context stemming from the encoder with the current states of the target vehicle to make the lane change prediction.

Contributions of this article can be summarized as follows:

- We propose a Transformer-based transfer learning framework to achieve accurate lane change prediction. This framework skillfully incorporates the interaction context captured by pre-trained GNN-based trajectory models rather than training the lane change prediction model from the raw trajectory data, which improves both the effectiveness and efficiency of lane change prediction.
- To better convert the trajectory interaction context into the lane-change-oriented interaction context, an aggregated attention mechanism is applied to the multi-head attention in the Transformer encoder. Additionally, we develop relation-aware biases to enable the multi-head attention in the Transformer decoder to account for heterogeneous input features.
- By combining the transfer learning framework with an enhanced Transformer architecture, our approach achieves not only higher lane-change prediction accuracy but also greater stability and robustness across various lane-change scenarios. Comprehensive evaluations on real-world datasets confirm the effectiveness of the proposed framework, demonstrating its superior performance in real-world applications.

The rest of this article is organized as follows. Section II reviews some relevant works. Section III establishes the problem definition. Section IV elaborates on the proposed framework. Section V presents detailed experimental results.

Finally, Section VI concludes the research and provides some future works.

II. Related Work

In this section, we review some critical studies about lane change intention prediction and vehicle trajectory prediction. Furthermore, the development of transfer learning and its applications in intelligent transportation systems are briefly introduced.

A. Lane Change Prediction

In recent decades, lane change prediction has been widely investigated for building efficient, safe, and trustworthy autonomous driving. Some recent works have shown deep learning-based methods have more powerful representation capability than traditional methods [10], [37], [38], especially for extracting the interaction context between vehicles. Multi-Layer Perceptron (MLP) combined with Bayesian regression was exerted to model the lateral movement of vehicles and produce the probability of lane change [39]. To better accommodate the vehicle dynamics, [40] integrated the kinematics equation into MLP, which provides a more accurate prediction. As an efficient feature extractor, the Convolutional Neural Network (CNN) was applied to capture the instantaneous relationships between vehicles [41], [42], which facilitates the intention prediction. Some research [43], [44] also constructed multi-view CNNs to extract the context from both vehicular interaction and ambient environments. Recurrent Neural Network (RNN) is widely used to represent the temporal dependence attributed to its feedback connections [45]. Therefore, RNN is highly appropriate for modeling the temporal motion of vehicles [46]. Reference [47] injected the sequential movement information into RNNs to compute the congestion degree of adjacent lanes of the target vehicle. [48] compared Echo State Network (ESN) with RNN and found that RNN has higher accuracy when predicting the right lane change. The shortage of RNN is easily falling into gradient vanishing or gradient explosion [49]. As a variant of RNN, Long Short-Term Memory (LSTM) uses a gating mechanism to alleviate the problem [50]. Accordingly, LSTM is popular with driver intention prediction. Reference [19] designed an ensemble bi-directional LSTM to account for the varying intention. References [20] and [21] assembled driving environments and vehicle dynamics as united feature vectors, which are then fed into LSTM to determine the feasibility of lane change. Transformer is a new type of deep learning model, which directly captures the interdependence between every two elements in an input sequence. Since its extremely strong long-range information memory, it is extensively used in Natural Language Processing (NLP) and results in the flourishing of pre-trained large language models, such as BERT [51] and GPT [52]. Some research [24], [25] attempted to explore the applications of Transformer in lane change prediction and discovered the superiority of Transformer over classical deep learning models.

In view of the advancement of Transformer, this study also resorts to Transformer to accomplish precise lane change prediction. Distinct from [24] and [25] directly using the primitive Transformer, we ameliorate the multi-head attention in both the encoder and decoder for better interaction context conversion and heterogeneous information adaption, respectively.

B. Vehicle Trajectory Prediction

Physics-based and maneuver-based methods prevail in modeling the individual dynamics of the target vehicle [54], however, they overlook the inter-vehicle interactions and yield suboptimal prediction performance [55], [56], [57]. Recently, researchers have aroused increasing interest in interaction-aware methods due to their representation ability of mutual influence between vehicles. References [58] and [59] construct spatial grid to depict the inter-vehicle interactions. However, this kind of representation hinders the comprehensive interplay between vehicle motions. Representing the interdependence between vehicles as a graph is more helpful for extracting interaction context than using the spatial grid or state cube [60], [61]. The burst of Graph Neural Networks (GNNs) also propels the modeling of non-Euclidean interaction patterns [62], [63], [64]. [65] applied a homogeneous graph to modeling interaction, which is unable to acclimatize to the entering and exiting of surrounding vehicles. To overcome the issue, [66] proposed to utilize the heterogeneous graph to render the interdependence between vehicles, where each node corresponds to a specific vehicle. Afterward, various types of GNNs are put forward to obtain acute interaction context. Reference [67] substituted the aggregation function by spatiotemporal graph architecture, where a kernel function captures the dynamic interactions. Reference [68] incorporated the uncertainty of future motion into GNN to accommodate the evolving nature of interdependence between vehicles. Reference [69] refined [67] through Gated Recurrent Unit (GRU) to further capture the temporal attributes of vehicle behaviors. References [35] and [36] considered the heterogeneous interaction context and developed multi-view GNNs to extract the heterogeneity. Reference [70] proposed to model the hierarchical interaction context from the micro level to the macro level, with logical-physical GNNs.

Against the background of vehicle trajectory prediction, GNN-based models have been shown to be significantly mighty. Thus, this study proposes to pre-train GNN-based vehicle trajectory prediction models to capture the interaction context between vehicles, which is then fed into the refined Transformer by transfer learning to promote lane change prediction.

C. Transfer Learning

In real-world applications, the inadequacy of training data is very common especially labeled data that requires a large amount of human factor. Transfer learning is a remarkable solution to the problem by transferring the knowledge learned from one task to a model on another different task [71]. Through transfer learning, the model capability of the target domain is strengthened with the informative knowledge from multiple related source domains [72]. In this manner, the data scarcity issue is lessened. Nowadays, transfer learning has pervaded diverse tasks in intelligent transportation systems. Reference [73] designed a map-matching model with the pre-trained Transformer, which only uses a limited number of human trajectories for downstream applications. Reference [74] exploited spatiotemporal dependence in urban traffic by cross-city transfer learning to augment the traffic prediction in small cities. Furthermore, researchers have also employed transfer learning to facilitate the advancement of autonomous driving. Reference [75] built a knowledge-enhanced Gaussian Mixture Model (GMM) to accurately identify the braking intensity of drivers. The braking knowledge is transferred among drivers with the probabilistic density ratio. Reference [76] proposed a transfer learning-based online reinforcement learning paradigm to improve the sequential control for intelligent vehicles.

In this study, we resort to transfer learning to enhance lane change prediction with the interaction context captured by pre-trained vehicle trajectory prediction models. For one thing, the potential of trajectory data can be fully exploited. For another thing, the proposed lane change prediction model can focus on learning from lane change behavior without the need to devote time to learning from trajectory data, which elevates lane change pattern recognition.

III. Problem Formulation

In this study, we aim to achieve real-time lane change prediction for autonomous vehicles. As shown in Fig. 1, the vehicle whose driving behaviors have an impact on the ego vehicle (autonomous vehicle) is identified as the target vehicle. To ensure the safety and efficiency of the ego vehicle, it is vital to predict the lane change intention of the target vehicle in a short time. Assume that the ego vehicle is able to detect states (e.g. position and speed) of the target vehicle and surrounding vehicles in its vicinity [24], [25]. The observed states of the vehicle v_i at time slot t are defined as:

$$\mathbf{X}_{v_i}^t = \left[\mathbf{s}_{v_i}^{t-T+1}, \mathbf{s}_{v_i}^{t-T+2}, \ldots, \mathbf{s}_{v_i}^{t}\right] \tag{1}$$

where $\mathbf{s}_{v_i}^t \in \mathbb{R}^F$ denotes the specific states (e.g. position, speed, and acceleration) of the vehicle v_i at time slot t and F denotes the number of state variables. Here, v_0 refers to the target vehicle and v_i $(i > 0)$ refers to the i-th surrounding vehicle. The traceback time window is denoted by T, so $\mathbf{X}_{v_i}^t$ is a matrix of the size $F \times T$. With the aforementioned definitions, this study aims to find an approach $f(\cdot)$ to predict the lane change intention of the target vehicle over the next time slot $t+1$:

$$y_{v_0}^{t+1} = f\left(\mathcal{H}_t\right) \tag{2}$$

where $\mathcal{H}_t \in \mathbb{R}^{N_t \times F \times T}$ is a three-order tensor, such that $\mathcal{H}_t(i,:,:) = \mathbf{X}_{v_i}^t$ and N_t denotes the total number of vehicles between time slots $t-T+1$ and t. Notably, lane change prediction is formulated as a multi-class classification problem in this study. Therefore, $y_{v_0}^{t+1}$ represents the most likely of the three classes: left lane change (left LC), right lane change (right LC), or lane keeping (LK), as presented in Fig. 1.

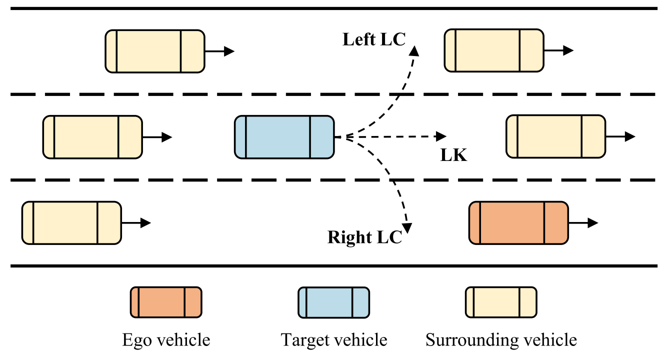


Fig. 1. Lane change prediction schematic.

IV. Methodology

This section introduces a novel Transformer-based transfer learning framework to address the problem outlined in Section III. As depicted in Fig. 2, the proposed lane change prediction framework consists of two components. The first component (A-Trajectory Model Pre-training), involves pre-training a trajectory model on historical data to capture inter-vehicular interactions. The second component (B-Lane Change Prediction), utilizes real-time vehicle information, which is fed into the pre-trained trajectory model to generate real-time interaction context. A refined Transformer is then employed to predict lane change intentions. Specifically, the Transformer encoder transforms the trajectory interaction context into a lane-change-oriented context. The Transformer decoder then merges this context with real-time target vehicle states to estimate the vehicle's lane change intention for the next time step.

A. Trajectory Model Pre-Training

To incorporate the informative vehicle trajectories into the lane change prediction model for enlarging the perception field, we pre-train the vehicle trajectory model with historical data and employ it to capture the real-time interaction context. The pre-trained transfer learning philosophy has the advantages of providing abundant vehicle trajectory information for lane change prediction and preventing the unnecessary training overhead that could have been caused by joint training of the trajectory model and lane change prediction model. Let t' be the given current time slot in the historical time horizon. The output of the vehicle trajectory model is the predicted trajectory of the target vehicle v_0 at the next time slot $t'+1$:

$$\mathbf{z}_{v_0}^{t'+1} = g_\phi\left(\mathcal{H}_{t'}\right) \tag{3}$$

where $\mathcal{H}_{t'}$ is the observed multi-vehicle state tensor at time slot t' with the same structure as $\mathcal{H}_t$ in Section III. $g_\phi(\cdot)$ denotes the vehicle trajectory model with the learnable parameter set ϕ. $\mathbf{z}_{v_0}^{t'+1} = \left[\left(z_{v_0,x}^{t'+1}, z_{v_0,y}^{t'+1}\right), \left(z_{v_0,x}^{t'+2}, z_{v_0,y}^{t'+2}\right), \ldots\right]$ stands for the predicted position sequence of the target vehicle, where $z_{v_0,x}^{t'+1}$ and $z_{v_0,y}^{t'+1}$ are the coordinates in the map. At the pre-training stage, the optimization objective is to minimize the trajectory prediction error. Notably, the message passing form of $g_\phi(\cdot)$ is critical to precisely capturing the interaction context. Recently, GNNs have become the model of choice to characterize

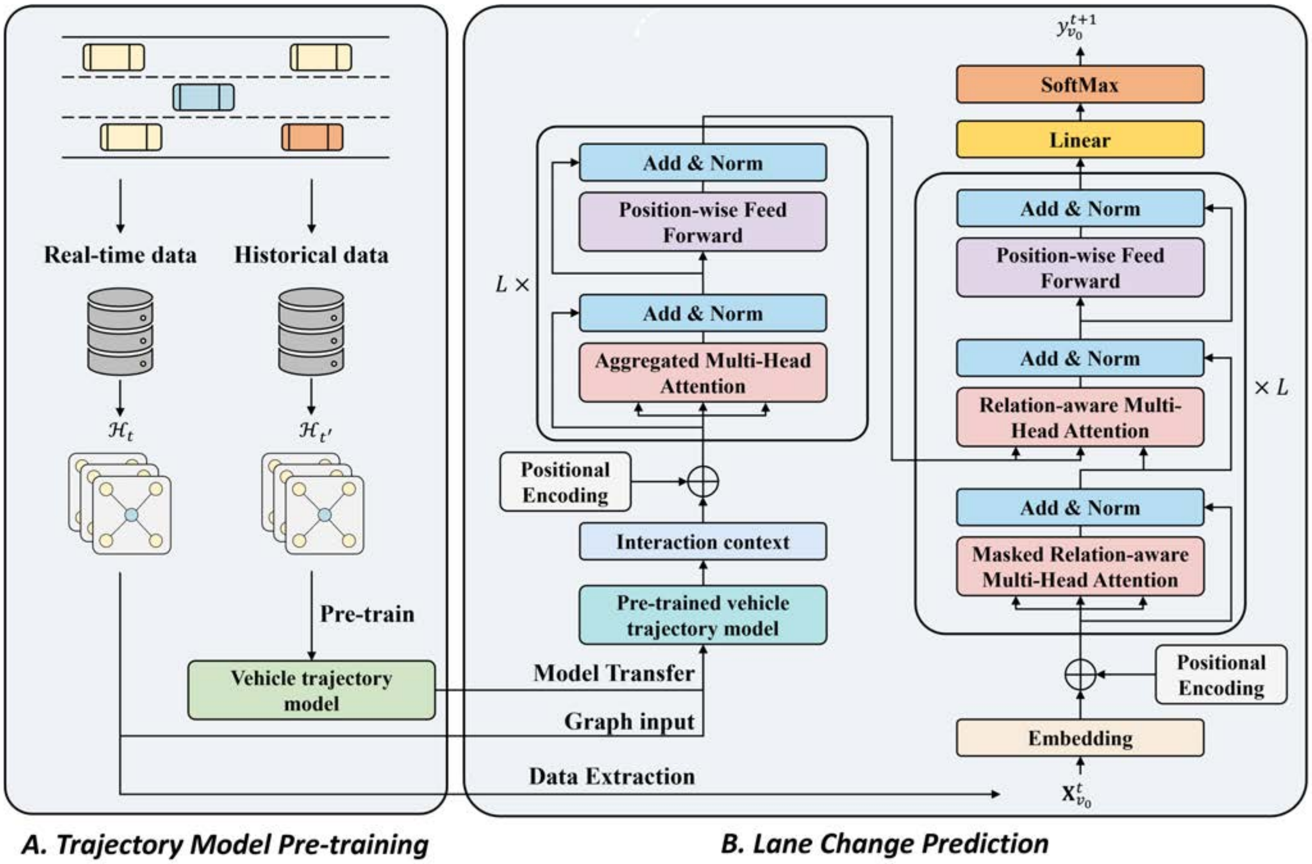


Fig. 2. Overview of the proposed Transformer-based transfer learning framework for lane change prediction. Historical data is used to pre-train the trajectory model while real-time data is fed into both the pre-trained trajectory model and the Transformer decoder to predict the future lane change intention.

inter-vehicle relationships as the non-Euclidean interactions between vehicles can be naturally represented as a graph. Meanwhile, GNNs suit the demand for adaptive modeling of the varying number of surrounding vehicles resulting from vehicle entry and exit. In this study, we adopt five representative GNN-based trajectory models as alternatives of $g_\phi(\cdot)$ to fulfill the self-consistent message passing:

- **VectorNet** [34]. VectorNet is the first to exploit the high-order interactions between vehicles by hierarchical graph neural network. The experimental results demonstrate the superiority of GNN over the traditional RNN and LSTM in representing inter-vehicle interactions.
- **EvolveGraph** [68]. EvolveGraph extends the VectorNet by accounting for the evolving feature of underlying interactions between vehicles. Through the dynamic GNN, the uncertainty and multi-modalities of interactional behaviors can be fully captured.
- **HEAT-I-R** [36]. HEAT-I-R modifies Graph Attention Networks (GAT) to handle the heterogeneity inherent in vehicles and interactions and designs a three-channel framework to jointly model individual dynamics, inter-vehicle interactions and driving environments.
- **HCAGCN** [35]. HCAGCN extends EvolveGraph by further representing the evolving inter-vehicle interactions from both the spatial and temporal domains. In particular, a node attention mechanism is developed to integrate the driving environment feature into the dynamic interaction context.
- **MVHGN** [70]. MVHGN considers the multi-view correlations between vehicles and invents an adaptive logical-physical combined Graph Convolutional Network (GCN) to accommodate heterogeneous vehicles in mixed-traffic environments.

In summary, the pre-trained trajectory model acts as a prior by generating an inter-vehicle interaction context for real-time lane change prediction and can be flexibly replaced with any model based on the GNN paradigm. In this study, we utilize the aforementioned five models as pre-trained models in our experiments.

B. Lane Change Prediction

To achieve accurate lane change prediction, we resort to Transformer to intensively combine the extracted trajectory interaction context with real-time target vehicle states. With the ingenious attention mechanism, Transformer is capable of differentially capturing the information of each part of the input data, which enhances the lane-change prediction accuracy. In the following, we elaborate on each basic module of the proposed Transformer.

1) Model Input: As presented in Fig. 2, the input comprises two components: the trajectory interaction context and target vehicle states fed into the encoder and decoder, respectively. The real-time trajectory interaction context is obtained from the pre-trained trajectory model by:

$$\mathbf{X}_{v_0,g}^t = g_{\phi'}(\mathcal{H}_t) \tag{4}$$

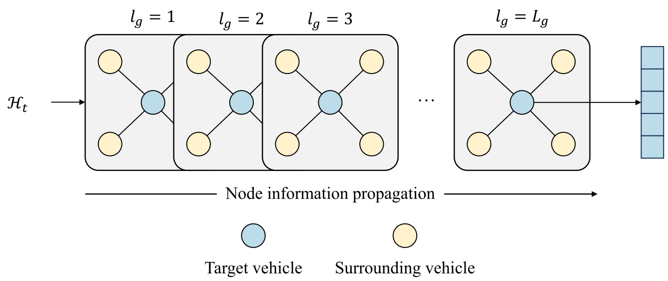


Fig. 3. The process of obtaining trajectory interaction context. Through multi-layer GNN, the trajectory-aware inter-vehicle interactions are captured and extracted from the activation of the target vehicle node in the last (L_g-th) layer.

where $g_{\phi'}$ denotes the GNN-specific component in the pre-trained trajectory model with the transferred GNN parameter set, ϕ'. The detailed process is illustrated in Fig. 3. The derived $\mathbf{X}^t_{v_0,g} \in \mathbb{R}^{F' \times T}$ represents the trajectory interaction context as the input to the refined Transformer encoder, where F' refers to the feature dimension. Correspondingly, the target vehicle states $\mathbf{X}^t_{v_0}$ is fed into the decoder.

2) Embedding Layer: The input layer in the decoder is followed by an embedding layer, which converts the primitive input data into a deep-seated feature vector. The embedding process is formulated as a linear transformation with learnable parameter $\mathbf{W}_e \in \mathbb{R}^{F' \times F}$: $\mathbf{X}^t_{v_0,e} = \mathbf{W}_e \mathbf{X}^t_{v_0}$.

3) Positional Encoding: To accommodate the positional information of each element in an input sequence, both the encoder and decoder are equipped with a positional encoding layer [32]. Without loss of generality, we describe below the positional encoding process in the encoder. The positional embedding $pe_{pos,i}$ is calculated by:

$$pe_{pos,i} = \begin{cases} \sin\left(\dfrac{pos}{10000^{i/F'}}\right) & \text{when } i \equiv_2 0 \\ \cos\left(\dfrac{pos}{10000^{i/F'}}\right) & \text{otherwise} \end{cases} \tag{5}$$

where pos denotes the position and i stands for the dimension. $i \equiv_2 0$ indicates the dimension i is even. $pe_{pos,i}$ is element of positional embedding matrix $\mathbf{PE}$. The encoding frames a wavelength, which shapes a geometric progression from 2π to $10000 \cdot 2\pi$. The diminishing frequency with position enables each element in a sequence to be assigned a unique embedding. Furthermore, the encoding also possesses a great attribute that $pe_{pos+k,i}$ can be expressed by the linear transformation of $pe_{pos,i}$ for any given k. The output of the positional encoding layer is the sum of trajectory interaction context $\mathbf{X}^t_{v_0,g}$ and the positional embedding matrix $\mathbf{PE}$, denoted by $\mathbf{X}^t_{v_0,gp}$.

4) Refined Multi-Head Attention: The attention mechanism in Transformer plays a critical role in capturing the profound information rooted in a long sequence, provoking an impressive performance in behavior modeling of autonomous driving [24], [25], [77]. The building block of multi-head attention is scaled dot-product attention, as presented in Fig. 4a. The output of it is computed as the weighted sum of input *value* $\mathbf{V}$. The weight for each element in $\mathbf{V}$ is determined by the compatibility degree, attained from the *query* $\mathbf{Q}$ along with the corresponding *key* $\mathbf{K}$. The calculation process can be formulated as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\mathbf{V} \tag{6}$$

where d_k is the dimension of $\mathbf{K}$, serving as the scaling factor. To bolster the representation ability of self-attention, multiple groups of projection from original $\mathbf{Q}$, $\mathbf{K}$ and $\mathbf{V}$ are developed to generate multi-faceted attention weights, as shown in Fig. 4b. Each group of the projection is realized by the linear transformation, which is called an attention head:

$$\text{head}_i = \text{Attention}\left(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V\right) \tag{7}$$

where $\left(\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V\right)$ denotes a specific group of projection parameters. Concatenating multiple attention heads, the output of multi-head attention is:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}\left(\text{head}_1, \ldots, \text{head}_h\right)\mathbf{W}^O \tag{8}$$

where $\mathbf{W}^O$ refers to the projection parameter in the last layer and h denotes the number of attention heads. Unlike RNN, multi-head attention can be parallelized for simultaneously computing activation of any position in an input sequence and thus immensely improves the training efficiency.

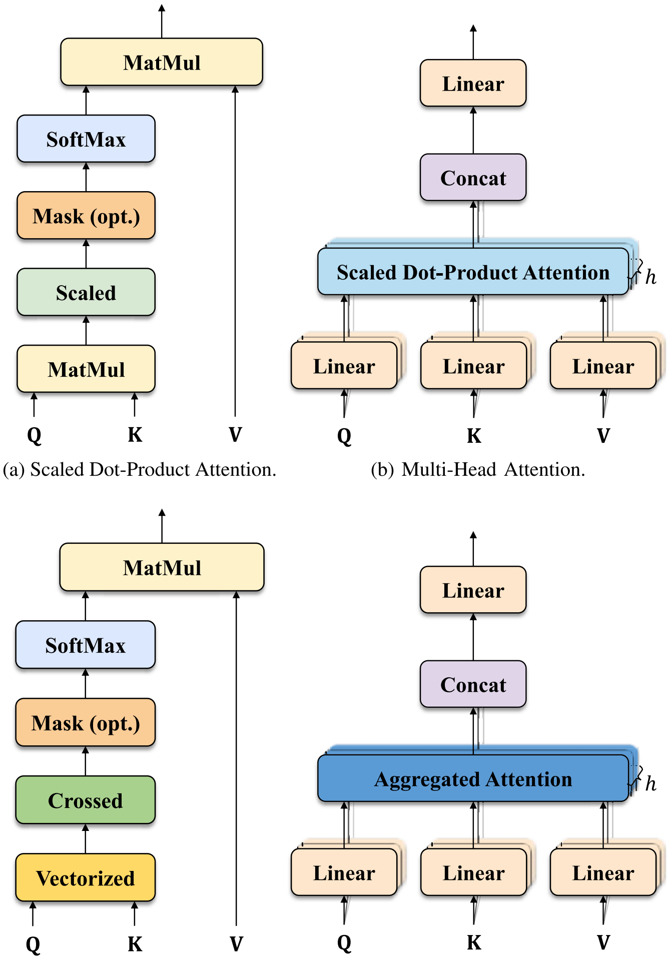
Despite exhibiting the powerful capability of sequential modeling, multi-head attention suffers from separate representation with low-dimensional $\mathbf{Q}$ and $\mathbf{K}$ that inhibits the global information propagation [78]. This issue is more pronounced in the Transformer encoder which aims to convert trajectory interaction context into lane-change-oriented context that requires global information extraction. To address this issue and facilitate context conversion, we propose aggregated multi-head attention. The scaled process in original attention is replaced with the aggregated operation (see Fig. 4c-4d). Specifically, each attention head is regarded as a low-rank distribution mounted on the input $\mathbf{V}$ and we aggregate these distributions with crossed connections to enhance the representation ability of multi-head attention:

$$\begin{pmatrix} \boldsymbol{U}^{(1)} \\ \boldsymbol{U}^{(2)} \\ \vdots \\ \boldsymbol{U}^{(h)} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1h} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{h1} & \lambda_{h2} & \cdots & \lambda_{hh} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{U}}^{(1)} \\ \hat{\boldsymbol{U}}^{(2)} \\ \vdots \\ \hat{\boldsymbol{U}}^{(h)} \end{pmatrix} \tag{9}$$

where $\hat{\boldsymbol{U}}^{(h)} = \mathbf{Q}^{(h)}\mathbf{K}^{(h)\mathrm{T}}$. The updated $\boldsymbol{U}$ has the advantage of enlarging the perception field of multi-head attention, better capturing the holistic information of an input sequence.

In the Transformer decoder, the heterogeneous attributes of the input $\mathbf{V}$ are not accounted for in the vanilla multi-head attention. To adaptively capture the correlations among different attributes of $\mathbf{V}$, a relation-aware multi-head attention is devised with dimension-specific biases. Given a group of projection $\left(\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V\right)$, the scaled dot-product operation in Fig. 4a is upgraded to:

$$\alpha_{jm} = \frac{\left(\mathbf{v}_j \mathbf{W}_i^Q\right)\left(\left(\mathbf{v}_m + \mathbf{e}^K\right)\mathbf{W}_i^K\right)^{\mathrm{T}}}{\sqrt{d_k}} \tag{10}$$



(a) Scaled Dot-Product Attention. (b) Multi-Head Attention.

(c) Aggregated Attention. (d) Aggregated Multi-Head Attention.

Fig. 4. Attention mechanism for vanilla transformer ((a) & (b)) and for refined transformer encoder ((c) & (d)).

where $\mathbf{v}_j$ denotes the j-th input vector (namely j-th row) of $\mathbf{V}$ and $\mathbf{e}^K$ is learnable relation biases, which enables the *key* $\mathbf{K}$ attends to heterogeneous features of $\mathbf{V}$.

5) Feed Forward Network: The output of the refined multi-head attention layer is fed into a position-wise feed-forward network, which comprises two linear transformation layers with a ReLU activation function amid them. The feed-forward network is imposed on each position separately so as to provide the informative mapping for the next encoder or decoder block.

6) Add & Norm Layer: To avert network degradation resulting from the block stacking, add & norm layer is applied by means of residual connection. By eliminating large variations of values transmitting through network layers, the add & norm layer is able to improve the generalization ability of the refined Transformer.

7) Intention Prediction: The hidden state for the next time slot is acquired after the honing of L stacked decoder blocks. Through a linear transformation with a softmax activation function, the lane change intention probability is obtained:

$$\omega_{v_0}^{t+1} = \text{softmax}\left(\mathbf{h}_{v_0}^{t+1}\right) \tag{11}$$

TABLE I

STATISTICS OF THE DATASETS

Dataset	HighD	NGSIM
#Duration (hours)	16.5	1.5
#Lanes (per direction)	2-3	5-6
#Vehicles	110,000	9,206
#Sampling rate (Hz)	25	10
#Lane change samples	8,150	3,310
#Lane keeping samples	12,225	4,965
#Training samples	16,300	6,620
LC+LK (%)	40.0%+60.0%	40.0%+60.0%
#Validation samples	2,037	827
LC+LK (%)	40.0%+60.0%	40.0%+60.0%
#Test samples	2,038	828
LC+LK (%)	40.0%+60.0%	40.0%+60.0%

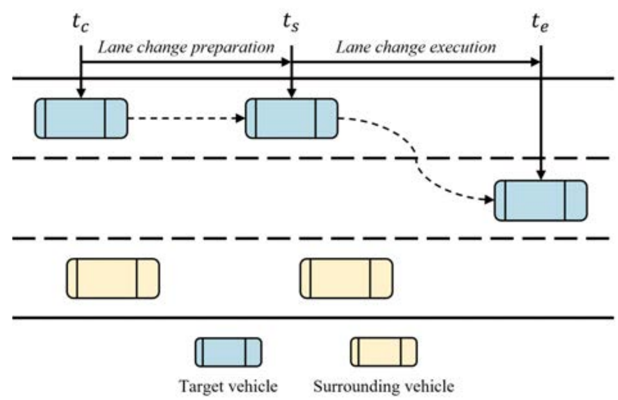


Fig. 5. The example of lane change process.

where $\mathbf{h}_{v_0}^{t+1}$ denotes the linearly transformed hidden state. The calculation process of softmax is:

$$\text{softmax}\left(\mathbf{h}_{v_0,i}^{t+1}\right) = \frac{\exp\left(\mathbf{h}_{v_0,i}^{t+1}\right)}{\sum_{k=1}^{C} \exp\left(\mathbf{h}_{v_0,k}^{t+1}\right)} \tag{12}$$

where $\mathbf{h}_{v_0,i}^{t+1}$ denotes the i-th component of $\mathbf{h}_{v_0}^{t+1}$ and C is the category number referring to three types of lane change intention (as illustrated in Section III). The type corresponding to the largest probability value in $\omega_{v_0}^{t+1}$ is the predicted lane change intention $y_{v_0}^{t+1}$.

C. Model Training

Instead of using the cross-entropy loss that is prevalent in lane change prediction [24], [25], we propose to leverage the focal loss function [79]. Focal loss is designed to handle class imbalance by assigning more weights to minority categories while down-weighting majority categories. The focal loss is defined as:

$$\mathcal{L} = -\frac{1}{M} \sum_{i} \phi_i \left(1 - y_i\right)^{\gamma} \log\left(y_i\right) \tag{13}$$

where y_i stands for the predicted probability of the ground-truth intention type of the i-th training sample. M denotes the total number of training samples. ϕ and γ are the fixed balancing factor and the tunable focusing parameter, respectively. Benefiting from the above two regulatory factors, the focal loss is capable of adaptively adjusting the optimization contributions from different training samples and thus elevates training efficiency and model performance.

V. Experimental Results

In this section, extensive experiments are conducted on real-world data to verify the performance of the proposed framework. All models are trained and evaluated by the PyTorch package.[1]

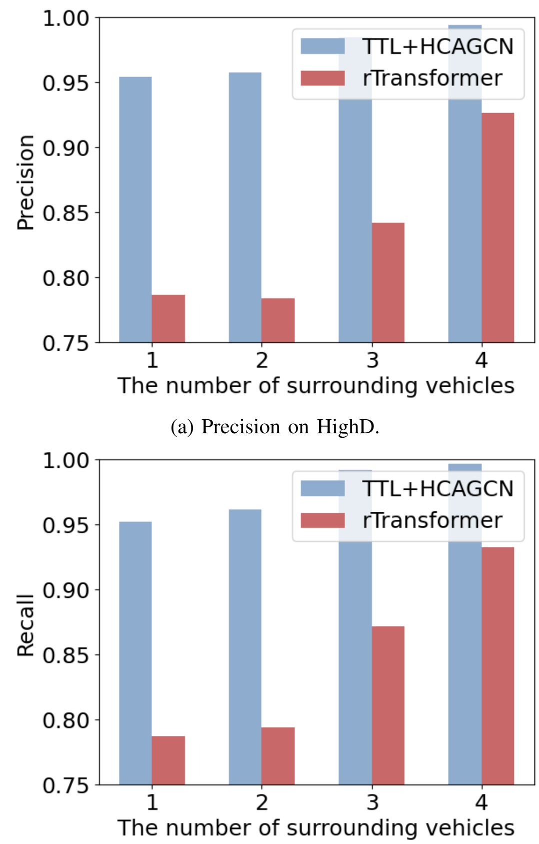
A. Experimental Setup

1) Datasets: To verify the proposed approach, we conduct experiments on two commonly used public datasets, HighD [80] and NGSIM [81]. For each dataset, 80% of the total samples are randomly selected as the training set. The remaining 20% samples are uniformly divided into the validation set and the test set. Table I shows the detailed statistics of the datasets. For model training and evaluation, it is essential to label lane change trajectories. Accordingly, we first examine the real-world lane change process. Fig. 5 illustrates a typical lane change scenario, divided into three key phases: (1) before t_c, the target vehicle maintains its lane; (2) from t_c to t_s, the vehicle observes its surroundings and prepares for the lane change maneuver; and (3) from t_s to t_e, the vehicle executes and completes the lane change by t_e. Since only t_s and t_e are provided in the datasets and t_c is undefined, we must determine t_c to label the lane change process. Based on previous studies on the temporal dynamics of lane changes [82], we assume a 2-second period between t_c and t_s. We label the time frame (t_c, t_e) as the lane change trajectory.

2) Baselines: We compare the proposed Transformer-based Transfer Learning framework (**TTL**), which incorporates different pre-trained vehicle trajectory models, with a sizable collection of state-of-the-art lane change prediction techniques:

- **SWA** [14]. SWA is a model-based method, which includes the steering wheel angle in vehicular kinematic description to improve lane change prediction performance. The steering wheel angle of a vehicle is calculated by its lateral and longitudinal velocities.
- **BN** [15]. BN is a generative method, in a bid to mimic the human reasoning process when executing the lane change maneuver. With the decomposed probability inference of the Bayesian network, BN is able to be aware of the dynamics of surrounding vehicles.
- **HMM** [83]. HMM supposes a latent variable to account for various observed lane change behaviors and integrates Gaussian Mixture Models to further interpret the lane change intention. Hidden Markov Models are then leveraged to accommodate latent states and estimate the lane change maneuver.
- **EBiRNN** [19]. EBiRNN is an ensemble bidirectional LSTM model to precisely forecast the lane change intention. LSTM is utilized to cope with the sequence of driving behaviors, capturing the temporal dependence and behavioral patterns.

[1] https://github.com/pytorch/pytorch



(a) Precision on HighD.

(b) Recall on HighD.

Fig. 6. The variation in lane change prediction performance with the number of surrounding vehicles. Notably, we only present the results on HighD as the NGSIM dataset contains only multi-surrounding-vehicle lane change scenarios.

- **ESN-RNN** [48]. ESN-RNN compares the lane change prediction performance of ESN and RNN and selects a more suitable one for each driving scenario. Under the circumstances, the left lane change maneuver is predicted by ESN while the right lane change behavior is estimated by RNN.
- **AttnLSTM** [21]. AttnLSTM incorporates the attention mechanism into LSTM to achieve accurate lane change prediction. The attention mechanism computes impacting weights of driving features at different time slots on the lane change intention, enhancing the discrimination ability of LSTM.
- **Transformer** [24]. Transformer models the time-series driving behaviors with multi-head attention, resulting in a powerful generalization ability. Notably, this is a decoder-only Transformer with a pooling layer to output the probability of lane change.
- **rTransformer** [25]. rTransformer employs the complete encoder-decoder structure to augment the lane change prediction. The lateral trajectory information captured by the Transformer encoder is regarded as the prior knowledge for the Transformer decoder, supporting the extraction of inter-vehicle interactions.

TABLE II

INTENTION PREDICTION PERFORMANCE ON BOTH HIGHD AND NGSIM

Method	HighD				NGSIM			
	Lane Change		Lane Keeping		Lane Change		Lane Keeping	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
SWA	0.5398	0.6491	0.6815	0.6933	0.6240	0.6124	0.6427	0.6436
BN	0.6019	0.6933	0.7121	0.7058	0.6730	0.6812	0.7024	0.6997
HMM	0.6138	0.7013	0.7044	0.7129	0.6835	0.7012	0.6714	0.6931
EBiRNN	0.7428	0.7512	0.7836	0.7617	0.7983	0.7716	0.8018	0.8034
ESN-RNN	0.7219	0.7285	0.7416	0.7440	0.7712	0.7837	0.8012	0.7996
AttnLSTM	0.7918	0.8017	0.8135	0.8029	0.8316	0.8284	0.8501	0.8496
Transformer	0.8615	0.8754	0.8862	0.8903	0.9014	0.9048	0.8915	0.8987
rTransformer	0.8736	0.8817	0.8905	0.9052	0.9093	0.9136	0.9057	0.9011
TTL+VectorNet	0.9716	0.9783	0.9802	0.9814	0.9730	0.9695	0.9786	0.9719
TTL+EvolveGraph	0.9796	0.9824	0.9815	0.9833	0.9806	0.9865	0.9812	0.9796
TTL+HEAT-I-R	0.9841	0.9878	0.9883	0.9865	0.9901	0.9919	0.9907	0.9925
TTL+MVHGN	0.9879	0.9903	**0.9915**	**0.9932**	0.9916	0.9920	0.9904	0.9918
TTL+HCAGCN	**0.9883**	**0.9914**	0.9901	0.9920	**0.9927**	**0.9926**	**0.9937**	**0.9933**

TABLE III

LANE CHANGE PREDICTION PERFORMANCE BREAKDOWN ON BOTH HIGHD AND NGSIM

Method	HighD				NGSIM			
	Left Lane Change		Right Lane Change		Left Lane Change		Right Lane Change	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
SWA	0.5463	0.6619	0.5281	0.6220	0.6381	0.6311	0.6094	0.6017
BN	0.6218	0.7025	0.5907	0.6854	0.6841	0.6907	0.6654	0.6723
HMM	0.6305	0.7124	0.5988	0.6916	0.6892	0.7202	0.6784	0.6833
EBiRNN	0.7511	0.7724	0.7337	0.7309	0.8081	0.7826	0.7894	0.7631
ESN-RNN	0.7362	0.7388	0.7086	0.7164	0.7852	0.8042	0.7608	0.7665
AttnLSTM	0.8026	0.8201	0.7813	0.7802	0.8431	0.8328	0.8250	0.8179
Transformer	0.8821	0.8875	0.8432	0.8690	0.9104	0.9123	0.8925	0.8966
rTransformer	0.8857	0.8937	0.8665	0.8784	0.9174	0.9199	0.8987	0.9082
TTL+VectorNet	0.9723	0.9805	0.9701	0.9762	0.9725	0.9665	0.9734	0.9721
TTL+EvolveGraph	0.9803	0.9815	0.9787	0.9836	0.9802	0.9880	0.9811	0.9842
TTL+HEAT-I-R	0.9882	0.9864	0.9810	0.9885	0.9911	0.9924	0.9889	0.9913
TTL+MVHGN	0.9892	0.9917	0.9863	0.9894	**0.9924**	0.9925	0.9908	0.9913
TTL+HCAGCN	**0.9896**	**0.9925**	**0.9871**	**0.9905**	0.9923	**0.9931**	**0.9930**	**0.9920**

3) Evaluation Metrics: Two prevailing evaluation metrics, i.e. Precision and Recall, are used in the experiments:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{14}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{15}$$

where TP denotes the number of positive samples predicted correctly, FP denotes the number of negative samples wrongly predicted as positive, and FN stands for the number of positive samples wrongly predicted as negative.

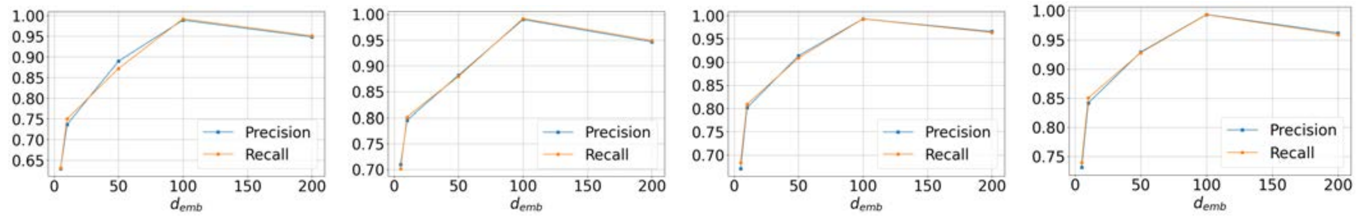
4) Implementation Details: The Adam optimizer is utilized to minimize the training loss and the learning rate is set to 0.0001. The training epochs are set as 500 and 400 for HighD and NGSIM datasets, respectively. The batch size is set to 256. The embedding size d_{emb} for the input is set as 100. The number of attention heads h is set to 4. The dimension of attention-related variables d_{model} is uniformly set as 200. Furthermore, the time interval is set to 0.2s and the traceback time window T is set to 4s. All experiments are implemented on a server with an NVIDIA A100 GPU of 40GB memory.

B. Intention Prediction Performance

In this experiment, we compare various combinations of TTL and pre-trained vehicle trajectory models (introduced in Section IV-A) against all baselines on different intention prediction tasks. Table II shows the intention prediction results on both datasets. The best-performing method in each case is highlighted.

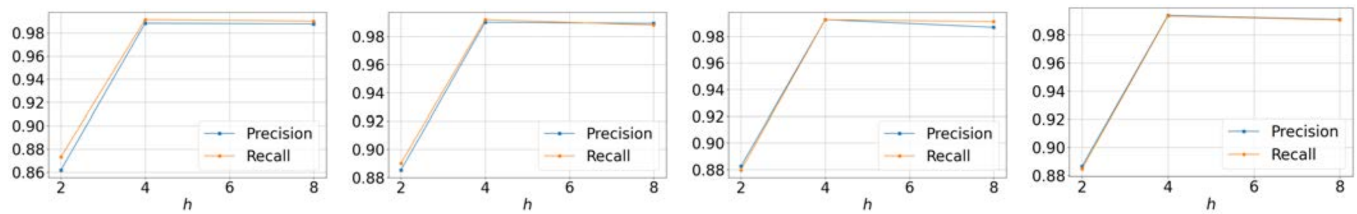
We observe that deep learning-based baselines (EBiRNN, ESN-RNN, AttnLSTM, Transformer, and rTransformer) consistently achieve better results than the model-based method (SWA) and generative methods (BN and HMM) in all cases, which verifies the capability of deep learning models. Among deep learning-based baselines, methods that integrate attention mechanism (AttnLSTM, Transformer, and rTransformer) perform better than bare RNNs (EBiRNN and ESN-RNN) in all cases, indicating that attention mechanism encourages the comprehension of driving intention. Apparently, the TTL family consistently outperforms all baselines on both datasets, demonstrating the powerful prediction ability of TTL. TTL+HCAGCN achieves the best performance in most cases, showing an improvement of 10.6% over the best-performing baseline (rTransformer) on average. Moreover, the TTL family exhibits robust and stable performance in predicting driving intention on different datasets, while baselines produce significantly different results on HighD and NGSIM.

We further examine the impact of introducing virtual vehicles in the best-performing baseline, rTransformer. Specifically, we plot the variation in lane change prediction



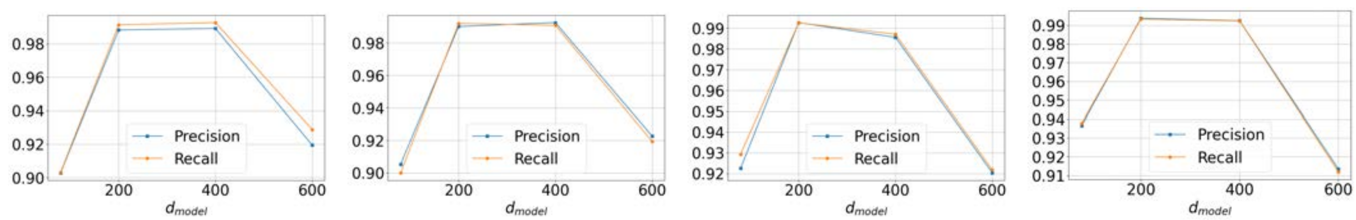
(a) Lane change prediction on HighD.
(b) Lane keeping prediction on HighD.
(c) Lane change prediction on NGSIM.
(d) Lane keeping prediction on NGSIM.

Fig. 7. Impact of the embedding size d_{emb}.



(a) Lane change prediction on HighD.
(b) Lane keeping prediction on HighD.
(c) Lane change prediction on NGSIM.
(d) Lane keeping prediction on NGSIM.

Fig. 8. Impact of the number of attention heads h.



(a) Lane change prediction on HighD.
(b) Lane keeping prediction on HighD.
(c) Lane change prediction on NGSIM.
(d) Lane keeping prediction on NGSIM.

Fig. 9. Impact of the dimension of attention variables d_{model}.

performance based on the number of surrounding vehicles, as shown in Fig. 6. We observe that both our approach and rTransformer perform worse with fewer surrounding vehicles compared to scenarios with more vehicles. This is expected, as lane change prediction becomes more challenging in free-flow environments where drivers tend to make discretionary lane changes. However, rTransformer's performance declines significantly. This is due to the introduction of virtual vehicles to supplement interaction information, which also introduces noise into the lane change modeling process.

C. Performance Across Different Categories of Lane Change

To further explore the effectiveness of TTL, we look into the breakdown of performance on different categories of the lane change maneuver. Table III shows the prediction results in different lane change categories on both datasets. The best-performing method in each case is highlighted.

Similar to the trend of intention prediction, deep learning-based baselines perform better than model-based and generative methods in all cases. Additionally, the attention mechanism has an immense effect on distinguishing between driving maneuvers of left lane change and right lane change, as evidenced by the satisfactory performance of AttnLSTM, Transformer, and rTransformer. The TTL family also outperforms all baselines in predicting whether left lane change or right lane change on both datasets. TTL+HCAGCN still achieves the best performance, showing an improvement of 10.5% over the best-performing baseline (rTransformer) on average. Generally, it is more difficult to accurately predict the right lane change behavior due to the higher randomness degree. However, the TTL family realizes extremely close performance when predicting left and right lane change maneuvers.

D. Parametric Studies

The embedding size d_{emb}, the number of attention heads h, and the dimension of attention variables d_{model} are three vital hyperparameters in TTL. Therefore, we investigate the impact of them on the prediction performance of TTL. Without loss of generality, we present the results of the best-performing method, namely TTL+HCAGCN.

First, specifying the number of attention heads $h = 4$ and the dimension of attention variables $d_{model} = 200$, we vary the embedding size d_{emb} from 5 to 200 on a log scale, and exhibit its impact on intention prediction for both datasets in Fig. 7. It can be observed that the performance ascends rapidly in the beginning and then reaches the peak when $d_{emb} = 100$ in all cases. Second, specifying the embedding size $d_{emb} = 100$ and the dimension of attention variables $d_{model} = 200$, we vary the number of attention heads h from 2 to 8, and plot its impact on prediction performance for both datasets in Fig. 8. Similar to the tendency of d_{emb}, the performance rises sharply in the beginning and remains steady when $h \geq 4$ in all cases. Finally, specifying the embedding size $d_{emb} = 100$ and the

TABLE IV
ABLATION STUDY ON INTENTION PREDICTION PERFORMANCE

Method	HighD				NGSIM			
	Lane Change		Lane Keeping		Lane Change		Lane Keeping	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
TTL_E+HCAGCN	0.9547	0.9608	0.9565	0.9659	0.9561	0.9627	0.9588	0.9634
TTL_D+HCAGCN	0.9405	0.9471	0.9654	0.9692	0.9473	0.9515	0.9626	0.9663

TABLE V
ABLATION STUDY ON LANE CHANGE PREDICTION PERFORMANCE BREAKDOWN

Method	HighD				NGSIM			
	Left Lane Change		Right Lane Change		Left Lane Change		Right Lane Change	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
TTL_E+HCAGCN	0.9726	0.9763	0.9318	0.9455	0.9679	0.9782	0.9436	0.9471
TTL_D+HCAGCN	0.9509	0.9554	0.9417	0.9502	0.9528	0.9553	0.9526	0.9580

TABLE VI
TIME-CONSUMING COMPARISON (S)

Method	HighD	NGSIM
AttnLSTM	0.009	0.004
Transformer	0.021	0.012
rTransformer	0.028	0.019
TTL+HCAGCN	0.029	0.021

number of attention heads $h = 4$, we vary the dimension of attention variables d_{model} from 80 to 600, and show its impact on intention prediction for both datasets in Fig. 9. We discover that the best performance is achieved when $d_{model} = 200$ in most cases. Accordingly, in all experiments, the embedding size d_{emb}, the number of attention heads h, and the dimension of attention variables d_{model} are set to 100, 4, and 200, respectively.

E. Ablation Studies

In this experiment, we investigate the effectiveness of the proposed aggregated multi-head attention and relation-aware multi-head attention. Two variants TTL_E+HCAGCN and TTL_D+HCAGCN are created. TTL_E+HCAGCN represents using the aggregated multi-head attention in the encoder while keeping the original multi-head attention in the decoder. TTL_D+HCAGCN embodies utilizing the relation-aware multi-head attention in the decoder while maintaining the original multi-head attention in the encoder. Tables IV-V present the experimental results.

In Table IV, we observe that TTL_E+HCAGCN achieves a lower variance of prediction performance between lane change and lane keeping compared to TTL_D+HCAGCN, manifesting that aggregated multi-head attention is critical for improving the lane change prediction performance since it can fluently convert the trajectory interaction context into lane-change-oriented interaction context. In Table V, we discover that TTL_D+HCAGCN outperforms TTL_E+HCAGCN in stability when predicting different categories of the lane change maneuver. This implies that relation-aware multi-head attention enhances the robustness of TTL in predicting lane change behaviors by smoothing attention weights to accommodate heterogeneous input features.

F. Comparison of Inference Time

Inference time is a key factor influencing the wide deployment of the proposed framework. Therefore, we calculate the time consumption of different methods and the results are shown in Table VI. Notably, the computing server used in this study is on par with mainstream autonomous driving platforms. It can be found that the inference time for the best-performing method TTL+HCAGCN is only 0.029s and 0.021s on HighD and NGSIM, respectively, comparable to the best-performing baseline (rTransformer). This means that the proposed framework TTL is efficient enough to predict the lane change intention in real time.

VI. CONCLUSION

Existing lane-change prediction models often fail to fully utilize the large volume of vehicle trajectory data and rely on rigid representations of traffic environments. To address these limitations, we propose a novel Transformer-based transfer learning framework that incorporates inter-vehicle interaction context learned from pre-trained trajectory models, rather than directly from raw trajectory data. This approach enhances both the effectiveness and efficiency of lane change prediction. Our extensive evaluation on two real-world datasets demonstrates the superiority of this framework across multiple intention prediction tasks. Compared to eight state-of-the-art baselines, our model consistently outperforms the best deep learning-based approach, achieving at least 10.6% and 10.5% improvements in predicting driving intention and lane change type, respectively. Additionally, our approach shows strong robustness and stability across different datasets and tasks, benefiting from its adaptability to varying traffic environments through interaction context. Lastly, we assess computational efficiency and confirm that the proposed method is fast enough to support real-time predictions.

In the future, we plan to further investigate the lane change intention prediction problem by accounting for extreme weather conditions and uncertainties in prediction results.

REFERENCES

[1] Y. Xing, C. Lv, H. Wang, D. Cao, E. Velenis, and F.-Y. Wang, "Driver activity recognition for intelligent vehicles: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5379–5390, Jun. 2019.

[2] R. Chai, D. Liu, T. Liu, A. Tsourdos, Y. Xia, and S. Chai, "Deep learning-based trajectory planning and control for autonomous ground vehicle parking maneuver," *IEEE Trans. Autom. Sci. Eng.*, vol. 20, no. 3, pp. 1633–1647, Jul. 2022.

[3] R. Chai, A. Tsourdos, H. Gao, Y. Xia, and S. Chai, "Dual-loop tube-based robust model predictive attitude tracking control for spacecraft with system constraints and additive disturbances," *IEEE Trans. Ind. Electron.*, vol. 69, no. 4, pp. 4022–4033, Apr. 2022.

[4] W. Ding, J. Chen, and S. Shen, "Predicting vehicle behaviors over an extended horizon using behavior interaction network," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 8634–8640.

[5] P. Karle, M. Geisslinger, J. Betz, and M. Lienkamp, "Scenario understanding and motion prediction for autonomous vehicles—Review and comparison," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 16962–16982, Oct. 2022.

[6] R. Chai, A. Tsourdos, H. Gao, S. Chai, and Y. Xia, "Attitude tracking control for reentry vehicles using centralised robust model predictive control," *Automatica*, vol. 145, Nov. 2022, Art. no. 110561.

[7] M. Rahman, M. Chowdhury, Y. Xie, and Y. He, "Review of microscopic lane-changing models and future research opportunities," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1942–1956, Dec. 2013.

[8] D. Tran, J. Du, W. Sheng, D. Osipychev, Y. Sun, and H. Bai, "A human-vehicle collaborative driving framework for driver assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3470–3485, Sep. 2019.

[9] I. Kotseruba and J. K. Tsotsos, "Attention for vision-based assistive and automated driving: A review of algorithms and datasets," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 19907–19928, Nov. 2022.

[10] Y. Xing et al., "Driver lane change intention inference for intelligent vehicles: Framework, survey, and challenges," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4377–4390, May 2019.

[11] R. Song and B. Li, "Surrounding vehicles' lane change maneuver prediction and detection for intelligent vehicles: A comprehensive review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6046–6062, Jul. 2022.

[12] C. Schlenoff, R. Madhavan, and Z. Kootbally, "PRIDE: A hierarchical, integrated prediction framework for autonomous on-road driving," in *Proc. IEEE Int. Conf. Robot. Autom., ICRA.*, Jul. 2006, pp. 2348–2353.

[13] D. D. Salvucci, H. M. Mandalia, N. Kuge, and T. Yamamura, "Lane-change detection using a computational driver model," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 49, no. 3, pp. 532–542, Jun. 2007.

[14] K. Schmidt, M. Beggiato, K. H. Hoffmann, and J. F. Krems, "A mathematical model for predicting lane changes using the steering wheel angle," *J. Saf. Res.*, vol. 49, pp. 85–90, Jun. 2014.

[15] G. Weidl, A. L. Madsen, S. Wang, D. Kasper, and M. Karlsen, "Early and accurate recognition of highway traffic maneuvers considering real world application: A novel framework using Bayesian networks," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 3, pp. 146–158, Fall. 2018.

[16] L. Klitzke, C. Koch, and F. Köster, "Identification of lane-change maneuvers in real-world drivings with hidden Markov model and dynamic time warping," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–7.

[17] P. Kumar, M. Perrollaz, S. Lefèvre, and C. Laugier, "Learning-based approach for online lane change intention prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2013, pp. 797–802.

[18] H. Woo et al., "Lane-change detection based on vehicle-trajectory prediction," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 1109–1116, Apr. 2017.

[19] Y. Xing, C. Lv, H. Wang, D. Cao, and E. Velenis, "An ensemble deep learning approach for driver lane change intention inference," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102615.

[20] W. Wang, T. Qie, C. Yang, W. Liu, C. Xiang, and K. Huang, "An intelligent lane-changing behavior prediction and decision-making strategy for an autonomous vehicle," *IEEE Trans. Ind. Electron.*, vol. 69, no. 3, pp. 2927–2937, Mar. 2022.

[21] Z.-N. Li, X.-H. Huang, T. Mu, and J. Wang, "Attention-based lane change and crash risk prediction model in highways," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 22909–22922, Dec. 2022.

[22] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[23] J. Jumper et al., "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

[24] H. Guo, M. Keyvan-Ekbatani, and K. Xie, "Lane change detection and prediction using real-world connected vehicle data," *Transp. Res. C, Emerg. Technol.*, vol. 142, Sep. 2022, Art. no. 103785.

[25] K. Gao et al., "Dual transformer based prediction for lane change intentions and trajectories in mixed traffic environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 6, pp. 6203–6216, Jun. 2023.

[26] R. Chai, A. Tsourdos, S. Chai, Y. Xia, A. Savvaris, and C. L. P. Chen, "Multiphase overtaking maneuver planning for autonomous ground vehicles via a desensitized trajectory optimization approach," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 74–87, Jan. 2023.

[27] R. Chai, H. Niu, J. Carrasco, F. Arvin, H. Yin, and B. Lennox, "Design and experimental validation of deep reinforcement learning-based fast trajectory planning and control for mobile robot in unknown environment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 4, pp. 5778–5792, Apr. 2022.

[28] R. Chai, A. Tsourdos, A. Savvaris, S. Chai, Y. Xia, and C. L. P. Chen, "Design and implementation of deep neural network-based control for automatic parking maneuver process," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1400–1413, Apr. 2022.

[29] Q. Shi and H. Zhang, "An improved learning-based LSTM approach for lane change intention prediction subject to imbalanced data," *Transp. Res. C, Emerg. Technol.*, vol. 133, Dec. 2021, Art. no. 103414.

[30] Q. Xue, Y. Xing, and J. Lu, "An integrated lane change prediction model incorporating traffic context based on trajectory data," *Transp. Res. C, Emerg. Technol.*, vol. 141, Aug. 2022, Art. no. 103738.

[31] J. Do, K. Han, and S. B. Choi, "Lane change–intention inference and trajectory prediction of surrounding vehicles on highways," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 7, pp. 3813–3825, Jul. 2023.

[32] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jun. 2017, pp. 5998–6008.

[33] Y. Lu et al., "Hyper-relational interaction modeling in multi-modal trajectory prediction for intelligent connected vehicles in smart cites," *Inf. Fusion*, vol. 114, Feb. 2025, Art. no. 102682.

[34] J. Gao et al., "VectorNet: Encoding HD maps and agent dynamics from vectorized representation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11525–11533.

[35] Y. Lu, W. Wang, X. Hu, P. Xu, S. Zhou, and M. Cai, "Vehicle trajectory prediction in connected environments via heterogeneous context-aware graph convolutional networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8452–8464, Aug. 2022.

[36] X. Mo, Z. Huang, Y. Xing, and C. Lv, "Multi-agent trajectory prediction with heterogeneous edge-enhanced graph attention network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9554–9567, Jul. 2022.

[37] J. Schlechtriemen, F. Wirthmueller, A. Wedel, G. Breuel, and K.-D. Kuhnert, "When will it change the lane? A probabilistic regression approach for rarely occurring events," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 1373–1379.

[38] X. Duan, C. Sun, D. Tian, J. Zhou, and D. Cao, "Cooperative lane-change motion planning for connected and automated vehicle platoons in multi-lane scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7073–7091, Jul. 2023.

[39] S. Yoon and D. Kum, "The multilayer perceptron approach to lateral motion prediction of surrounding vehicles for autonomous vehicles," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 1307–1312.

[40] Y. Hu, W. Zhan, and M. Tomizuka, "Probabilistic prediction of vehicle semantic intention and motion," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 307–313.

[41] D. Lee, Y. P. Kwon, S. McMains, and J. K. Hedrick, "Convolution neural network-based lane change intention prediction of surrounding vehicles for ACC," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.

[42] S. Mozaffari, E. Arnold, M. Dianati, and S. Fallah, "Early lane change prediction for automated driving systems using multi-task attention-based convolutional neural networks," *IEEE Trans. Intell. Vehicles*, vol. 7, no. 3, pp. 758–770, Sep. 2022.

[43] S. Casas, W. Luo, and R. Urtasun, "IntentNet: Learning to predict intention from raw sensor data," in *Proc. Conf. Robot Learn.*, 2018, pp. 947–956.

[44] Q. Deng, J. Wang, K. Hillebrand, C. R. Benjamin, and D. Söffker, "Prediction performance of lane changing behaviors: A study of combining environmental and eye-tracking data in a driving simulator," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3561–3570, Aug. 2020.

[45] A. Jain, A. Singh, H. S. Koppula, S. Soh, and A. Saxena, "Recurrent neural networks for driver activity anticipation via sensory-fusion architecture," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 3118–3125.

[46] A. Zyner, S. Worrall, and E. Nebot, "A recurrent neural network solution for predicting driver intention at unsignalized intersections," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1759–1764, Jul. 2018.

[47] L. Li, W. Zhao, C. Xu, C. Wang, Q. Chen, and S. Dai, "Lane-change intention inference based on RNN for autonomous driving on highways," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5499–5510, Jun. 2021.

[48] K. Griesbach, M. Beggiato, and K. H. Hoffmann, “Lane change prediction with an echo state network and recurrent neural network in the urban area,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6473–6479, Jul. 2022.

[49] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, “Independently recurrent neural network (IndRNN): Building a longer and deeper RNN,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5457–5466.

[50] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social LSTM: Human trajectory prediction in crowded spaces,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 961–971.

[51] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” 2018, *arXiv:1810.04805*.

[52] T. Brown et al., “Language models are few-shot learners,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.

[53] Y. Lu et al., “Knowledge-driven lane change prediction for secure and reliable Internet of Vehicles,” *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 4, 2025, doi: 10.1109/TITS.2025.3526341.

[54] S. Lefèvre, D. Vasquez, and C. Laugier, “A survey on motion prediction and risk assessment for intelligent vehicles,” *ROBOMECH J.*, vol. 1, no. 1, pp. 1–14, Dec. 2014.

[55] M. Althoff, O. Stursberg, and M. Buss, “Model-based probabilistic collision detection in autonomous driving,” *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 299–310, Jun. 2009.

[56] M. Ardelt, C. Coester, and N. Kaempchen, “Highly automated driving on freeways in real traffic using a probabilistic framework,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1576–1585, Dec. 2012.

[57] X. Li, Z. Sun, D. Cao, Z. He, and Q. Zhu, “Real-time trajectory planning for autonomous urban driving: Framework, algorithms, and verifications,” *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 2, pp. 740–753, Apr. 2016.

[58] N. Deo and M. M. Trivedi, “Convolutional social pooling for vehicle trajectory prediction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1468–1476.

[59] X. Mo, Y. Xing, and C. Lv, “Interaction-aware trajectory prediction of connected vehicles using CNN-LSTM networks,” in *Proc. 46th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2020, pp. 5057–5062.

[60] X. Jia, P. Wu, L. Chen, Y. Liu, H. Li, and J. Yan, “HDGT: Heterogeneous driving graph transformer for multi-agent trajectory prediction via scene encoding,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13860–13875, Nov. 2023.

[61] Y. Lu, Q. Lin, H. Chi, and J.-Y. Chen, “Automatic incident detection using edge-cloud collaboration based deep learning scheme for intelligent transportation systems,” *Int. J. Speech Technol.*, vol. 53, no. 21, pp. 24864–24875, Nov. 2023.

[62] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, “A comprehensive survey on graph neural networks,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Mar. 2020.

[63] Y. Lu, H. Ding, S. Ji, N. N. Sze, and Z. He, “Dual attentive graph neural network for metro passenger flow prediction,” *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13417–13431, Aug. 2021.

[64] J. Chen, W. Wang, K. Yu, X. Hu, M. Cai, and M. Guizani, “Node connection strength matrix-based graph convolution network for traffic flow prediction,” *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 12063–12074, Sep. 2023.

[65] F. Diehl, T. Brunner, M. T. Le, and A. Knoll, “Graph neural networks for modelling traffic participant interaction,” in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 695–701.

[66] X. Mo, Y. Xing, and C. Lv, “ReCoG: A deep learning framework with heterogeneous graph for interaction-aware trajectory prediction,” 2020, *arXiv:2012.05032*.

[67] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, “Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 14424–14432.

[68] J. Li, F. Yang, M. Tomizuka, and C. Choi, “Evolvegraph: Multi-agent trajectory prediction with dynamic relational reasoning,” in *Proc. NeurIPS*, 2020, pp. 19783–19794.

[69] Z. Sheng, Y. Xu, S. Xue, and D. Li, “Graph-based spatial–temporal convolutional network for vehicle trajectory prediction in autonomous driving,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 17654–17665, Oct. 2022.

[70] D. Xu, X. Shang, H. Peng, and H. Li, “MVHGN: Multi-view adaptive hierarchical spatial graph convolution network based trajectory prediction for heterogeneous traffic-agents,” *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 6, pp. 6217–6226, Jun. 2023.

[71] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[72] F. Zhuang et al., “A comprehensive survey on transfer learning,” *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jul. 2020.

[73] Z. Jin, J. Kim, H. Yeo, and S. Choi, “Transformer-based map-matching model with limited labeled data using transfer-learning approach,” *Transp. Res. C, Emerg. Technol.*, vol. 140, Jul. 2022, Art. no. 103668.

[74] Y. Huang, X. Song, Y. Zhu, S. Zhang, and J. J. Q. Yu, “Traffic prediction with transfer learning: A mutual information-based approach,” *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 8, pp. 8236–8252, Aug. 2023.

[75] Z. Li, J. Gong, C. Lu, and J. Li, “Personalized driver braking behavior modeling in the car-following scenario: An importance-weight-based transfer learning approach,” *IEEE Trans. Ind. Electron.*, vol. 69, no. 10, pp. 10704–10714, Oct. 2022.

[76] X. Liang, Y. Liu, T. Chen, M. Liu, and Q. Yang, “Federated transfer reinforcement learning for autonomous driving,” in *Federated and Transfer Learning*. Cham, Switzerland: Springer, 2022, pp. 357–371.

[77] Y. Hu et al., “Planning-oriented autonomous driving,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 17853–17862.

[78] N. Shazeer, Z. Lan, Y. Cheng, N. Ding, and L. Hou, “Talking-heads attention,” 2020, *arXiv:2003.02436*.

[79] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[80] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, “The highD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems,” in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2118–2125.

[81] *Ngsim—Next Generation Simulation*, U.S. Dept. Transp., Washington, DC, USA, 2008.

[82] T. Chen, X. Shi, and Y. D. Wong, “Key feature selection and risk prediction for lane-changing behaviors based on vehicles’ trajectory data,” *Accident Anal. Prevention*, vol. 129, pp. 156–169, Aug. 2019.

[83] Y. Zhang, Q. Lin, J. Wang, S. Verwer, and J. M. Dolan, “Lane-change intention estimation for car-following control in autonomous driving,” *IEEE Trans. Intell. Veh.*, vol. 3, no. 3, pp. 276–286, Sep. 2018.