Dual Transformer Based Prediction for Lane Change Intentions and Trajectories in Mixed Traffic Environment

Kai Gao, Xunhao Li, Bin Chen, Lin Hu, *Member, IEEE*, Jian Liu, Ronghua Du, and Yongfu Li, *Senior Member, IEEE*

***Abstract*— In a mixed traffic environment of human and autonomous driving, it is crucial for an autonomous vehicle to predict the lane change intentions and trajectories of vehicles that pose a risk to it. However, due to the uncertainty of human intentions, accurately predicting lane change intentions and trajectories is a great challenge. Therefore, this paper aims to establish the connection between intentions and trajectories and propose a dual Transformer model for the target vehicle. The dual Transformer model contains a lane change intention prediction model and a trajectory prediction model. The lane change intention prediction model is able to extract social correlations in terms of vehicle states and outputs an intention probability vector. The trajectory prediction model fuses the intention probability vector, which enables it to obtain prior knowledge. For the intention prediction model, the accuracy can be improved by designing the multi-head attention. For the trajectory prediction model, the performance can be optimized by incorporating intention probability vectors and adding the LSTM. Verified on NGSIM and highD datasets, the experimental results show that this model has encouraging accuracy. Compared with the model without intention probability vectors, the impact of the model on NGSIM dataset and highD dataset in RMSE is improved by 57.27% and 58.70% respectively. Compared with two existed models, evaluation metrics of the intention prediction can be improved by 7.40-10.09% on NGSIM dataset and 2.17-2.69% on highD dataset within advanced prediction time 1s. This method provides the insights for designing advanced perceptual systems for autonomous vehicles.**

***Index Terms*— Mixed traffic environment, transformer, intention prediction, trajectory prediction, NGSIM, highD.**

I. Introduction

AUTONOMOUS vehicles will be deployed in a complex mixed-traffic environment in the future. It is expected that the driving safety of autonomous vehicles and the safety of the passengers must be ensured, which puts forward higher requirements for the ability to predict the intentions and future driving trajectories of those vehicles that generate the possible lane change actions [1]. In recent years, autonomous vehicles have achieved unprecedented development with the progress of intelligent networking technology and artificial intelligence technology. The lane change behavior of autonomous vehicles in the hybrid human-machine traffic environment is usually conservative [2], i.e., the driving intention of surrounding vehicles may not be correctly judged. The most advanced driving assistance systems (ADAS) can not make accurate predictions. In the future, fully autonomous vehicles should be able to actively perceive changes in the surrounding environment and understand the driving habits and modes of human driving. Considering the current driving safety, vehicles often scratch and collide with other vehicles when moving toward the target lane [3]. It is necessary to predict in advance that the trajectory of the surrounding vehicles will conflict with the movement of other vehicles in the short term, which ensures the optimal lane change decision and direction control. Therefore, designing an advanced real-time prediction algorithm for lane change intentions and lateral trajectories significantly reduces the accident rate and promotes further development.

At present, there are four main models used to predict lane change intentions, including support vector machine (SVM), convolutional neural network (CNN), long short-term memory (LSTM) [4], [45] and Bayesian network (BN) [46]. To reduce the misjudgment rate of model prediction, Lyu et al. [5] proposed an SVM model based on recursive feature elimination (SVM-RFE) based on SVM, and set a lane change rate to reduce the misjudgment rate of lane change behavior prediction. In order to establish an intention prediction model based on LSTM, [6], [7], [8], [9] made corresponding contributions. In [6], the extracted local and global contextual features are put into CNN-LSTM with visual signal features extracted by CNN. While [7] deeply analyzed the lane change task again, redefined the lane change task as a regression problem, and used LSTM to predict the lane change time. In [8], a prediction model, including a primary and adaptive model, is proposed. The basic model is an LSTM prediction model, which reflects the driver's decision-making mode. The adaptive prediction model embeds the adaptive decision threshold into the basic model and updates the threshold in time through the Bayesian inference method. In [9], a data-driven framework based on Inverse Reinforcement Learning and Bidirectional

Manuscript received 15 June 2022; revised 22 October 2022 and 19 January 2023; accepted 20 February 2023. Date of publication 6 March 2023; date of current version 31 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 52172399, in part by the Natural Science Foundation of Hunan under Grant 2021JJ40575, in part by the Open Fund of Hunan Key Laboratory of Smart Roadway and Cooperative Vehicle-Infrastructure Systems (Changsha University of Science and Technology) under Grant kfj190701, and in part by the Postgraduate Scientific Research Innovation Project of Hunan Province under Grant QL20210194. The Associate Editor for this article was X. Di. *(Corresponding author: Bin Chen.)*

Kai Gao, Xunhao Li, Bin Chen, Lin Hu, Jian Liu, and Ronghua Du are with the College of Automotive and Mechanical Engineering, Changsha University of Science and Technology, Changsha 410205, China (e-mail: kai_g@csust.edu.cn; lxhms07@163.com; cbzr520@csust.edu.cn; hulin@csust.edu.cn; liujian_csust@163.com; csdrh@csust.edu.cn).

Yongfu Li is with the Key Laboratory of Intelligent Air-Ground Cooperative Control for Universities in Chongqing, College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: liyongfu@cqupt.edu.cn).

Digital Object Identifier 10.1109/TITS.2023.3248842

Recurrent Neural Network architecture is proposed to predict vulnerable road users' intentions. Moreover, approaches of [10], [11], [12], and [13] extract the interactive features, which are put into different neural network structures, such as graph neural network and depth convolution network.

At the same time, multi-layer perceptron is used in the intention prediction [14], [15]. In the above literature, although interactive features are considered, they fuse all the features and then input them into the network. This disadvantage is that it increases the input dimension and makes it more difficult for the model to extract features. Meanwhile, the machine learning method has achieved good results in lane-change intention prediction, but further research on improving the prediction accuracy and considering the application of other machine learning methods is still of great necessary to the driving safety of autonomous vehicles.

In recent years, scholars have carried out much research on trajectory prediction, such as human trajectory prediction [16], [17], [18] and vehicle trajectory prediction. There are many similarities between human trajectory and vehicle trajectory prediction. Many scholars have improved the prediction method of human trajectory prediction and applied it to vehicle trajectory prediction. Since [16], a new social pooling layer has been introduced to share information between LSTMs, and a human trajectory prediction based on social LSTM is proposed. The social pooling layer is also introduced in the research on trajectory prediction. Deo and Trivedi [19] exploited an LSTM encoder-decoder model, which used the convolution social pooling layer to improve the social pooling layer. The model outputs the multi-modal prediction distribution of future trajectory according to the maneuver category. Zhang et al. [20] presented an extended convolution social pool LSTM model to improve the spatial interaction modelling of vehicles.

Trajectory prediction is a sequence-to-sequence process, so an encoder-decoder structure is widely used in trajectory prediction. In [21], the encoder based on LSTM is used to encode historical tracks into feature vectors, and the decoder generated future track sequences. On this basis, Chen et al. [22] considered multidimensional input information and used three layers of different LSTM to capture the knowledge of spatial, temporal, and trajectory data. Similarly, in [23] and [31], LSTM encoder-decoder structure is used, and the MDN layer is introduced to output the trajectory probability distribution. In [5], according to the results of lane change intention recognition, a trajectory prediction model based on LSTM is established. However, these improvements ignore that intention recognition results can be used as an input feature and impact prediction accuracy.

Attention mechanism has been widely used in intention prediction [24] and trajectory prediction, which can effectively capture the interaction with surrounding vehicles and further enhance the performance of original states input [25], [26]. It is frequently used to process surrounding environment information and combined with convolution network and LSTM network [27], [28]. Messaoud et al. [29] proposed an encoder-decoder structure based on relational recurrent neural networks, which combined the advantages of LSTM block in portraying the time evolution of trajectory and attention mechanism to model the relative interaction among vehicles. This approach outperforms the LSTM encoder-decoder in predicting the RMSE values of the trajectories. Messaoud et al. [30] utilized a non-local multi-head attention mechanism to capture the relative importance of each vehicle. Fu et al. [31] proposed an LSTM-based network with an attention mechanism, which combined temporal and spatial dimensions. The problem of dynamic interference in the autopilot system of the highway was solved. Cai et al. [32] proposed an environmental-attention network model (EA-Net) to obtain complete interactive information, which was more comprehensive and effective in refining the feature information. Although attention mechanisms have been improved and applied to different networks, the features extracted by a single attention mechanism are frequently insufficient. Therefore, fully exploiting the attention mechanism for feature extraction remains a problem worth considering.

Transformer network is a machine-learning network, which is first proposed by Google [33]. It adopts an encoder-decoder structure and is exclusively based on the attention mechanism. It has achieved positive results in machine translation. A large number of scholars began to study the application of Transformer in image classification [34], trajectory prediction [35], [36], [37], [50] and others afterwards, and achieved favorable results in time series prediction [38], [39]. Inspired by these studies, this paper proposes to use the Transformer network to predict lane change intentions and lateral trajectories of vehicles.

From the above discussion, there are still two research gaps in the existing research. First, the lane change intention and trajectory prediction problems were independently studied in existing studies, i.e., two separate prediction models were built for the intention and trajectory prediction problems. However, these attempts neglect the interaction between intention prediction and trajectory prediction, which sacrifices the prediction accuracy. Second, in existing studies, the target vehicle states and their surrounding vehicles' states are treated as one feature vector to the prediction model, where the couplings among these states are ignored.

To address these issues mentioned above, we propose an integrated approach with a dual Transformer to implement the prediction of lane change intentions and trajectories for target vehicles. First, the target vehicle's lateral position information and interaction information are put into the encoder module in parallel so that the proposed model can fully extract the correlation between the lateral position information and the interactive information when changing lanes. The proposed model first outputs the intention prediction probability vector through the connected layer. After estimating lane change intentions, an intention probability vector is introduced to predict the lateral trajectory. The Concat function is used to splice the probability vector of intention prediction and the input of the encoder module of the trajectory prediction from end to end, which is put into the encoder module of the trajectory prediction as a whole unit. Furthermore,

the trajectory prediction of lane change vehicles through the decoder module is finally output. Moreover, the RMSE can be significantly reduced by fusing the intention probability vector.

Contributions of this article are two folds:

- We propose an integrated approach with dual Transformer to achieve the co-prediction of lane change intentions and trajectories for vehicles in the mixed traffic environment with human-driving vehicles and autonomous vehicles. Experiment results verify that the proposed method has a higher prediction accuracy when compared with existing methods.
- Both target vehicle information and surrounding vehicles' information are considered in the proposed model, where a multi-head attention mechanism is utilized to characterize the social interactions among this information. Then, the proposed prediction can well respond to the change of surrounding environments.

The rest of this article is organized as follows: Section II mainly introduces problem definition and data description. Section III mainly presents the proposed method used in this work. Section IV analyzes the experimental results of intention prediction model and trajectory prediction model. Section V mainly summarizes the content as well as the future research work of this article.

II. Description of Problem and Data

In this section, we give a description of the scenario. In addition, we describe the publicly available the datasets used to evaluate the proposed model.

A. Scenario Description

In this paper, we aim to build a lane change intention and trajectory prediction model for an autonomous vehicle that can almost simultaneous output intentions and trajectories. First, we define a general lane change scenario as shown in Fig.1. The yellow vehicle is the ego vehicle (the autonomous vehicle), and the blue vehicle is the target vehicle (the human vehicle) driving in the adjacent lane. The range is the space in the distance between the ego vehicle and vehicle 1, which can be used by the target vehicle to cut in when it intends to make the left line change and cause potential risks. To ensure the safe driving of the ego vehicle, we establish a lane change prediction model and a trajectory prediction model. For the situation in Fig. 1, the ego vehicle will predict the lane change intentions and trajectories of the blue vehicle. There are three lane change intentions for the blue vehicle (right lane change, lane keeping, left lane change).

This research assumes that the ego vehicle's sensors can obtain the speed and position information of the target vehicle and its surrounding vehicles [8]. Next, the prediction model predicts the lane change intentions and trajectories of the target vehicle. Finally, the ego vehicle will make a rational decision according to the model's prediction results.

B. Dataset

To validate the performance of the algorithm proposed in this paper, the NGSIM [40] dataset I-80, US-101 freeway and highD dataset are selected for training and testing the model. The NGSIM dataset originates from the NGSIM program initiated by the U.S. Federal Highway Administration with a sampling frequency of 10 Hz, and records information including vehicle coordinates, speeds, accelerations, vehicle types, and lane numbers. The highD dataset [41] is a large-scale natural vehicle trajectory dataset of German highways. The sampling frequency of this dataset is 25Hz, and it includes 110,000 vehicles driving data in 6 locations, with a total duration of 16.5 hours. The study section of the two datasets is shown in Fig. 2.

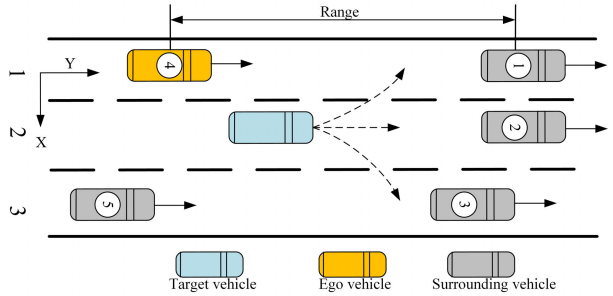
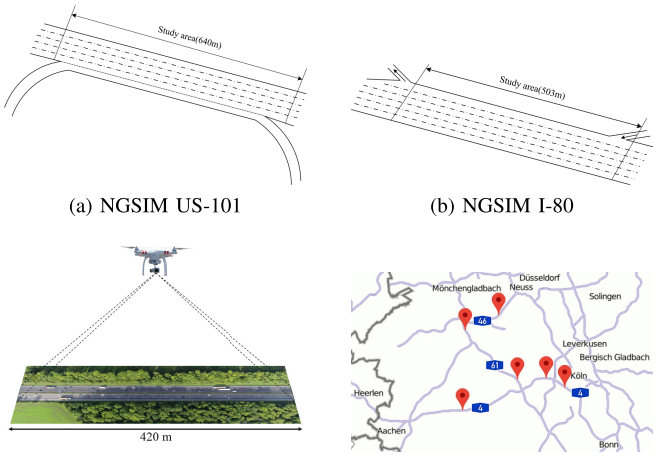


Fig. 1. Diagram of research case.



(a) NGSIM US-101 (b) NGSIM I-80

(c) Collection scenario of highD (d) Collection locations of highD

Fig. 2. Study area schematic and road structure of NGSIM [40] and highD [41].

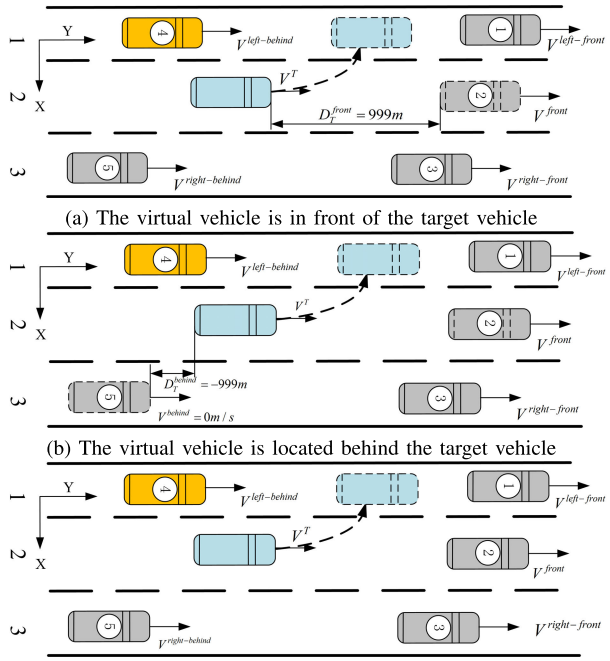
In order to reduce the influence of numerical difference on network training, the lateral position is processed accordingly. For the NGSIM dataset, since the local coordinate $local\_x_t$ at timestamp t given by NGSIM dataset is the distance from the leftmost road boundary, it is known that the road width of US-101 and I-80 is 3.5m, so $local\_x_t$ needs to be converted, as shown in formula (1):

$$\tilde{x}_t^{ng} = local\_x_t - (lane\_ID - 1) \times 3.5/0.3048 \quad (1)$$

where $lane\_ID$ is the lane number of the vehicle, $local\_x_t$ is the local coordinate of the target vehicle, and $\tilde{x}_t^{ng}$ is the lateral position of the target vehicle in the current lane on the NGISM dataset.

For the highD dataset, it is a two-way lane, so the lateral position is obtained by formula (2):

$$\tilde{x}_t^{hg} = x_t - x_1 \quad (2)$$



(a) The virtual vehicle is in front of the target vehicle

(b) The virtual vehicle is located behind the target vehicle

(c) No virtual vehicle

Fig. 3. Input features of the model with and without dummy vehicles. The blue represents the predicted target vehicle, the gray represents the surrounding vehicles around it, the blue with a dotted line represents the lane change target position of the target vehicle, and the gray with a dotted line represents the introduced virtual vehicle. For blue vehicles, the dashed black line is the future lateral trajectory. Using the method, we can easily calculate while maintaining the lane change features.

where x_1 is the lateral position of the initial moment in the extracted data, x_t is the local coordinate of the target vehicle, and $\tilde{x}_t^{hg}$ is the lateral position of the target vehicle relative to the initial moment on the highD dataset at timestamp t. Here, $1 \leq t \leq T_h$, T_h is the overall length of the observed historical trajectory.

The positions and speeds of nearest 5 surrounding vehicles (i.e., $m = 5$) are utilized in the calculation of the interaction information of the target vehicle. When the number of surrounding vehicles is less than 5, to ensure the consistency of the interaction features, we consider $5 - m$ virtual vehicles in the interaction information calculation. This is a common practice in the lane change characteristic calculation [8], [23], [47].

In the virtual vehicle setting, we set the velocity and position of the virtual vehicle to a reasonable value, which ensures the virtual vehicles will not influence the prediction model. If the virtual vehicle is located in front of the target vehicle, the vehicle speed is set as $V^f = 999m/s$ and the distance from the target vehicle is set as $D_{tg}^f = 999m$. If the virtual vehicle is located behind the target vehicle, we set $V^b = 0m/s$ and $D_{tg}^b = -999m$ as shown in Fig.3. These settings ensure the virtual vehicle is sufficiently far away from the target vehicle. The features of the interactive information of the target vehicle used in this paper are shown in Table I.

Let $X^{ng}(t)$ and $X^{hg}(t)$ be the historical lateral trajectory at timestamp t for the NGSIM dataset and highD dataset respectively, as shown in formulas (3) and (4):

$$X^{ng}(t) = (\tilde{x}_1^{ng}, \tilde{x}_2^{ng}, \cdots, \tilde{x}_{T_h}^{ng}) \tag{3}$$

TABLE I

DESCRIPTION OF INPUT CHARACTERISTICS AND FEATURE CODE

Feature code	Feature
D_{tg}^f	The longitudinal distance between the target vehicle and the preceding vehicle
V_{tg}^f	The speed difference between the target vehicle and the preceding vehicle
$D_{tg}^{l\text{-}f}$	The longitudinal distance between the target vehicle and the left front vehicle
$V_{tg}^{l\text{-}f}$	The speed difference between the target vehicle and the left front vehicle
$D_{tg}^{r\text{-}f}$	The longitudinal distance between the target vehicle and the right front vehicle
$V_{tg}^{r\text{-}f}$	The speed difference between the target vehicle and the right front vehicle
$D_{tg}^{l\text{-}b}$	The longitudinal distance between the target vehicle and the left rear vehicle
$V_{tg}^{l\text{-}b}$	The speed difference between the target vehicle and the left rear vehicle
$D_{tg}^{r\text{-}b}$	The longitudinal distance between the target vehicle and the right rear vehicle
$V_{tg}^{r\text{-}b}$	The speed difference between the target vehicle and the right rear vehicle

$$X^{hg}(t) = (\tilde{x}_1^{hg}, \tilde{x}_2^{hg}, \cdots, \tilde{x}_{T_h}^{hg}). \tag{4}$$

Let $\hat{X}$ be the lateral trajectory of the target vehicle in the future T_{pre}, as shown in formula (5):

$$\hat{X} = (\hat{x}_{T_h+1}, \hat{x}_{T_h+2}, \cdots, \hat{x}_{T_h+T_{pre}}). \tag{5}$$

Since a long historical trajectory may contain useless information, the length of historical trajectory is extracted as $T_h = 3s$, and the prediction horizon is $T_{pre} = 4s$.

In this study, suppose the lateral trajectory feature vector at timestamp t is $\boldsymbol{X}_n(t)$, where $n = 1, 2, \ldots, N$. N is the training sample size. Equally, suppose the interactive feature vector at timestamp t is:

$$\boldsymbol{S}_n(t) = (D_{tg}^f, V_{tg}^f, D_{tg}^{l-f}, V_{tg}^{l-f}, D_{tg}^{l-b}, V_{tg}^{l-b}, D_{tg}^{r-f}, V_{tg}^{r-f}, D_{tg}^{r-b}, V_{tg}^{r-b}). \tag{6}$$

III. DUAL TRANSFORMER ARCHITECTURE

The input to the model is $\boldsymbol{X}_n(t)$, $\boldsymbol{S}_n(t)$, and $\boldsymbol{\omega}$ (trajectory prediction). The output is $\boldsymbol{\omega}$ (intention prediction) and $\hat{X}$ (trajectory prediction), where $\boldsymbol{\omega} = [a_1, a_2, a_3]^T$ are composed of the probabilities of each intention category, and a_1, a_2, a_3 represent the probability values of lane keeping (LK), right lane change (RLC) and left lane change (LLC) respectively. The maximum value in $\boldsymbol{\omega}$ is the output intention.

To achieve accurate prediction of intentions and trajectories simultaneously, it is required that the connection between the two cannot be broken. Therefore, we propose an ensemble dual Transformer consisting of two components as shown in Fig.4. The intention prediction model is the lane change intention prediction model that reflects basic lane change behaviors, which is trained off-line. The second model is a trajectory prediction model that predicts lane change trajectories when the target vehicle generates lane change intentions, which is combined with the intention results.

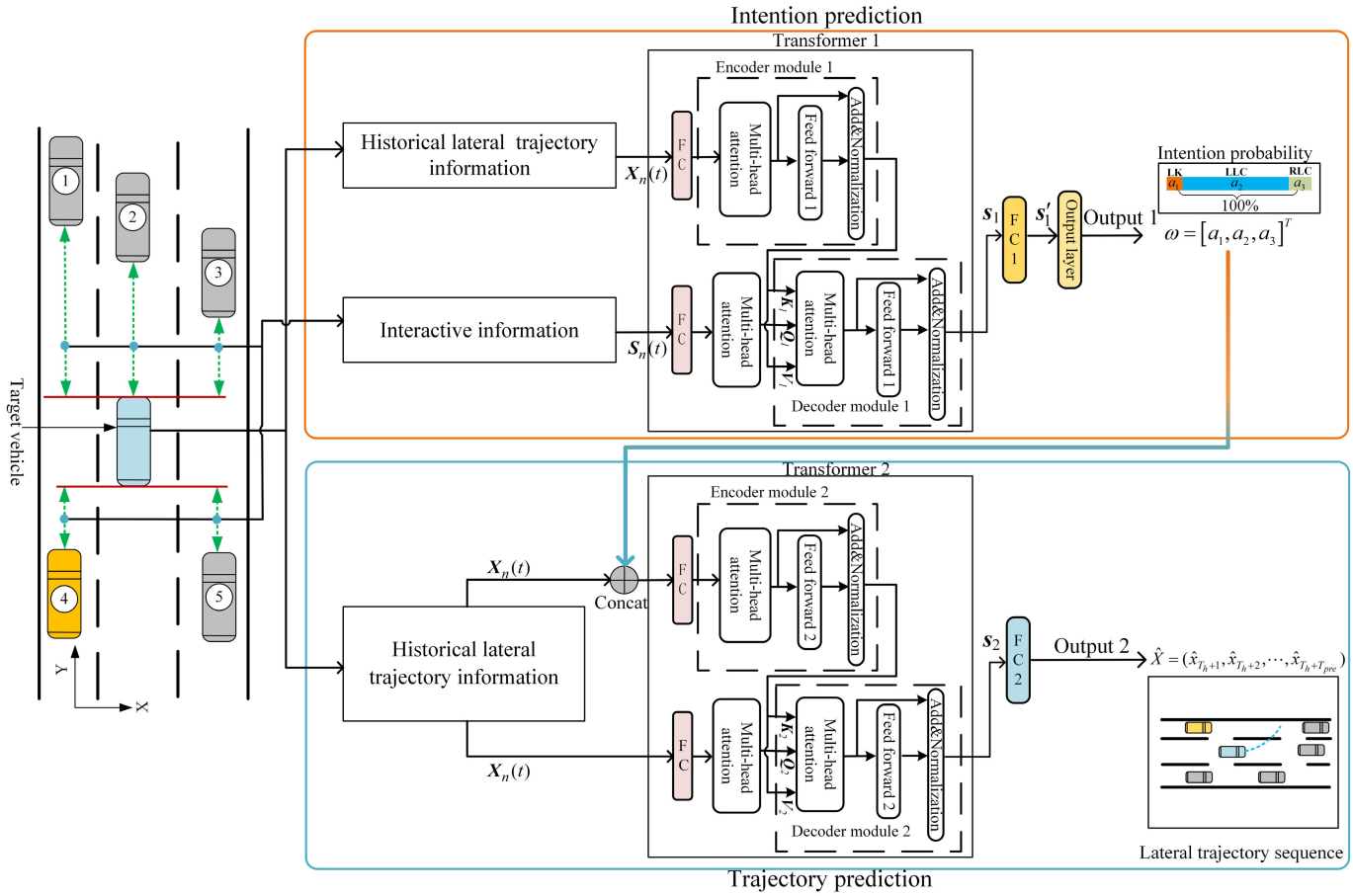


Fig. 4. Overall process of intention prediction and trajectory prediction of the proposed model. The proposed model consists of two parts: Transformer 1 and Transformer 2, and the input consists of three parts: lateral position trajectory, interactive information and intention probability vector. The output1 is the intention probability vector. Feed forward 1 is implemented using two 1D convolutional layers with a convolution kernel size of 8. Feed forward 2 uses two LSTM layers with 128 units to implement. The output dimensions of FC1 and FC2 are 32 and 40, respectively. The output 2 is the lateral trajectory. The green two-way arrow represents the interaction between the target vehicle and 5 surrounding vehicles. Concat is a splicing function. The encoder extracts features from the input, the decoder decodes vectors and extracting features, and finally outputs intention probability vectors and lateral trajectory sequences. Our method can be used as a multi-channel model that combines different input information, using intention probability vectors to improve the quality of trajectory prediction.

A. Transformer Network

Transformer is an encoder-decoder structure, in which the encoder and decoder are composed of six layers. It contains three modules: a multi-head attention module, a feed-forward full connection module, and two residual connections after each previous block.

Each multi-head attention module is concatenated by h Scaled Dot-Product Attentions (SDPA). SDPA is the transposed dot product of the query matrix $\boldsymbol{Q}$ and the key matrix $\boldsymbol{K}$ through the scaled dot-product. The correlation of all vectors in the two matrices is calculated. And the dimensions of $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\boldsymbol{V}$ are also calculated. d_k scales $\boldsymbol{Q}$ and $\boldsymbol{K}$ to reduce the impact of dimensional changes on the dot product value. Then using the softmax function maps the value to [0,1]. At the same time, in order to prevent over-fitting, set the Dropout layer, and its value is set to 0.1. Finally, $\boldsymbol{V}$ is weighted using its output as shown in formula (7):

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Dropout}\left[\text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}}\right)\right] \bullet \boldsymbol{V} \tag{7}$$

where query, key and value matrices are denoted by $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\boldsymbol{V}$, respectively.

The multi-head attention module can calculate the correlation at any position in the sequence. Because it has multiple heads, it can fuse the features focused on different heads. Compared with SDPA, the multi-head attention module focuses on the global information of the sequence more efficiently, and better handles long sequence data. Specifically, it is a linear projection of the input matrix. Then, the created $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\boldsymbol{V}$ are fed into the SDPA. Next, this process is repeated h times. Finally, they are stitched by the Concatenate layer to obtain the global self-attentive feature maps, as shown in formulas (8) and (9):

$$\text{Multihead}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Concat}(head_1, \cdots, head_h)\boldsymbol{W}^o \tag{8}$$

$$head_i = \text{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^{\boldsymbol{Q}}, \boldsymbol{K}\boldsymbol{W}_i^{\boldsymbol{K}}, \boldsymbol{V}\boldsymbol{W}_i^{\boldsymbol{V}}), \quad i = 1, 2, \ldots, h \tag{9}$$

and $\text{Attention}_i(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})$ denotes the self-attention of the ith head. Parameter matrices $\boldsymbol{W}_i^{\boldsymbol{Q}} \in \mathbb{R}^{d_{model} \times d_k}$, $\boldsymbol{W}_i^{\boldsymbol{K}} \in \mathbb{R}^{d_{model} \times d_k}$, $\boldsymbol{W}_i^{\boldsymbol{V}} \in \mathbb{R}^{d_{model} \times d_v}$, and $\boldsymbol{W}^o \in \mathbb{R}^{hd_v \times d_{model}}$.

For input information, the proposed model extracts different feature matrices through three different fully connected layers. Through the above formulas, the proposed model can calculate the self-attention of information between different moments, so as to obtain the time-dimension movement pattern of the vehicle, and complete the time series modeling.

The encoder module mainly creates the representation of the input, which enables the model to memorize and acquire the features of the context vector. It includes a multi-head attention module. The decoder module, which is responsible for outputting the results, contains two multi-head attention modules.

Add&Normalization in the encoder-decoder module is the residual concatenation and layer normalization, which takes charge of concatenating the input and output of the previous module and then performing layer normalization.

At the same time, we improve the Feed Forward Network layer of the traditional Transformer. Considering that LSTM achieves excellent results in trajectory prediction, the convolutional layers are linearly transformed using LSTM layers instead. The one in the intention prediction, however, remains unchanged.

B. Intention Prediction

Previous studies have demonstrated the effectiveness of using trajectory features, and interaction features [8], [23]. Despite the progress, it is still tricky for current feature aggregation methods to focus on the correlation between different feature information. Hence, we apply the Transformer to learn the correlation between the lateral and interactive trajectory features and achieve more precise intention prediction. The yellow frame of Fig.4 illustrates the overview of trajectory prediction.

The intention prediction is based on the observed trajectory information of the target vehicle and its interactive information with the surrounding vehicles to predict the next driving intentions of the target vehicle, as shown in the intention prediction model of Fig. 4. Thus, our goal is to learn a function $\boldsymbol{\omega}_{\boldsymbol{X_n(t)},\boldsymbol{S_n(t)}} = \mathcal{F}(\boldsymbol{X_n(t)}, \boldsymbol{S_n(t)})$, where $\mathcal{F}$ is the intention prediction model, and $\boldsymbol{\omega}_{\boldsymbol{X_n(t)},\boldsymbol{S_n(t)}} \in \mathbb{N}^{1\times 3}$ is the estimated intention probability, which is called the intention probability vector.

The whole prediction architecture is built based on Transformer. In the encoder module 1, three inputs of the multi-head attention $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\boldsymbol{V}$ are derived from the same input features $\boldsymbol{X_n}(t)$. For embeddings $\boldsymbol{X_n}(t)$, we use full connected layers. They are represented as formula (10):

$$\begin{cases} \bar{\boldsymbol{Q}} = \boldsymbol{X}_n(t)\boldsymbol{W}^{\boldsymbol{Q}} \\ \bar{\boldsymbol{K}} = \boldsymbol{X}_n(t)\boldsymbol{W}^{\boldsymbol{K}} \\ \bar{\boldsymbol{V}} = \boldsymbol{X}_n(t)\boldsymbol{W}^{\boldsymbol{V}} \end{cases} \tag{10}$$

where $\boldsymbol{W}^{\boldsymbol{Q}}$, $\boldsymbol{W}^{\boldsymbol{K}}$ and $\boldsymbol{W}^{\boldsymbol{V}}$ represent the learnable parameter matrices of different full connected layers.

Finally, the output of the encoder module 1 $\boldsymbol{K_1}$ and $\boldsymbol{V_1}$ can be processed as formula (11).

$$\boldsymbol{K_1} = \boldsymbol{V_1} = \Phi_1(\bar{\boldsymbol{Q}}, \bar{\boldsymbol{K}}, \bar{\boldsymbol{V}}) \tag{11}$$

where $\Phi_1(\cdot)$ is defined as formulas (12) and (13).

$$\Phi_1(x) = \text{LayerNorm}(Tanh(x))\boldsymbol{W}_1^{CNN})\boldsymbol{W}_2^{CNN}) + x)) \tag{12}$$

$$x = \text{Multihead}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) \tag{13}$$

in which LayerNorm and Tanh denote Layer Normalization and activate function [49], respectively, and $\boldsymbol{W}_1^{CNN}$ and $\boldsymbol{W}_2^{CNN}$ are the weight matrices of CNN networks.

In the decoder module 1, there are two input sources, and one of which is the interactive information $\boldsymbol{S}_n(t)$. We use the same method to embed $\boldsymbol{S}_n(t)$ as formula (14).

$$\begin{cases} \tilde{\boldsymbol{Q}} = \boldsymbol{S}_n(t)\boldsymbol{W}^{\boldsymbol{Q}} \\ \tilde{\boldsymbol{K}} = \boldsymbol{S}_n(t)\boldsymbol{W}^{\boldsymbol{K}} \\ \tilde{\boldsymbol{V}} = \boldsymbol{S}_n(t)\boldsymbol{W}^{\boldsymbol{V}} \end{cases}. \tag{14}$$

Meanwhile, we use formula (15) to derive $\boldsymbol{Q}_1$.

$$\boldsymbol{Q}_1 = \text{Multihead}(\tilde{\boldsymbol{Q}}, \tilde{\boldsymbol{K}}, \tilde{\boldsymbol{V}}) \tag{15}$$

To generate the feature of the entire target vehicle, decoder module 1 is applied to aggregate lateral trajectory features and interactive features. Formally, the decoder module 1 aggregates $\boldsymbol{K}_1$, $\boldsymbol{V}_1$ and $\boldsymbol{Q}_1$ with formula (16).

$$\boldsymbol{s}_1 = \Phi_1(\boldsymbol{Q}_1, \boldsymbol{K}_1, \boldsymbol{V}_1) \tag{16}$$

Next, we determine the intention probability vector $\boldsymbol{\omega}$ by a fully connected layer and an output layer. The output layer consists of a fully connected layer with a softmax activation function, which is utilized to calculate the probability of lane change intentions as shown in formula (17).

$$\begin{cases} \boldsymbol{s}_1' = \text{FC1}\ (\boldsymbol{s}_1) \\ \omega = \text{softmax}(\text{FC}(\boldsymbol{s}_1')) \end{cases} \tag{17}$$

where both FC1 and FC are the full connection layer, and their output space dimensions are 32 and 3, respectively. The softmax activation function is defined as formula (18).

$$\text{softmax}(x) = \frac{e^{x_i}}{\sum_{i=1}^{C} e^{x_i}} \tag{18}$$

where C is the category number of lane change intentions. For multi-lane highways, vehicle lane change intentions have generally three types, namely, right lane change, lane keeping and left lane change. Then, C is set to three here. The softmax activation function would transform the input to a non-negative number for generating probability.

For training loss, the intention prediction can often be regarded as a classification problem. In classification, the commonly used loss functions of two classifications are the logarithmic loss function and the log-like hood loss function, and the commonly used cross entropy loss function of three classifications. This article studies the prediction of LLC, LK, and RLC, thus the cross entropy loss function is used as shown in formula (19).

$$\mathcal{L}_1 = -\sum_{i=1}^{C} y_i \log\ y_i' \tag{19}$$

where y_i is the true label, y_i' is the probability for the ith class, and $C = 3$.

In summary, the target vehicle states and its surrounding vehicles' states are supplied to the prediction model respectively in the proposed method, where a multi-head attention mechanism is utilized to capture the social correlations among these states. Therefore, The dimensionality of the input vector of the encoder module 1 is reduced and the inference time of the model is improved.

C. Trajectory Prediction

Attention mechanism has been widely used to enhance the effect of trajectory prediction [31], [32]. Inspired by them, we propose Transformer-based method. The blue frame of Fig.4 illustrates the overview of trajectory prediction.

The trajectory prediction is to use the historical lateral trajectory information of the target vehicle and the intention probability vector as input. The output is the lateral trajectory of vehicle lane change in the next period T_{pre}. Hence, the problem can be formulated as formula (20).

$$\hat{X}_{\boldsymbol{X}_n(t)),\boldsymbol{\omega}} = \mathcal{F}(\boldsymbol{X}_n(t)), \boldsymbol{\omega}). \tag{20}$$

The encoder module 2 is responsible for encoding $\boldsymbol{X}_n(t)$ into a fixed-length context vector. Next, the decoder module 2 is not only responsible for encoding $\boldsymbol{X}_n(t)$ into the $\boldsymbol{Q}_2$ value of the next multi-head attention module, but also for receiving the output vector of the encoder module 2. The important features are extracted to predict the future trajectory.

Specifically, $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ are also obtained by full connected layer through formula (21). Note that $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ in formulas (10), (14), and (21) do not share weights.

$$\begin{cases} \hat{\boldsymbol{Q}} = \text{Concat}(\boldsymbol{X}_n(t), \boldsymbol{\omega})\boldsymbol{W}^{\boldsymbol{Q}} \\ \hat{\boldsymbol{K}} = \text{Concat}(\boldsymbol{X}_n(t), \boldsymbol{\omega})\boldsymbol{W}^{\boldsymbol{K}} \\ \hat{\boldsymbol{V}} = \text{Concat}(\boldsymbol{X}_n(t), \boldsymbol{\omega})\boldsymbol{W}^{\boldsymbol{V}} \end{cases}. \tag{21}$$

We then use the same steps as intention prediction, but the difference is that the Feed forward 2 is an LSTM network, such as formula (22).

$$\boldsymbol{K}_2 = \boldsymbol{V}_2 = \Phi_2(\hat{\boldsymbol{Q}}, \hat{\boldsymbol{K}}, \hat{\boldsymbol{V}}) \tag{22}$$

where $\Phi_2(\cdot)$ is defined as formulas (23) and (24).

$$\Phi_2(x) = \text{LayerNorm}(Tanh(x))\boldsymbol{W}_1^{LSTM})\boldsymbol{W}_2^{LSTM}) + x)) \tag{23}$$

$$x = \text{Multihead}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) \tag{24}$$

where $\boldsymbol{W}_1^{LSTM}$ and $\boldsymbol{W}_2^{LSTM}$ are the weight matrices of LSTM networks.

Then, we apply formula (8) to gain $\boldsymbol{Q}_2$. The output then of the decoder module 2 can be obtained as formula (25).

$$\boldsymbol{s_2} = \Phi_2(\boldsymbol{Q}_2, \boldsymbol{K}_2, \boldsymbol{V}_2) \tag{25}$$

In the trajectory prediction, RMSE loss function is often used to train the trajectory model [31], [32], because it can well reflect the average distance error between the predicted coordinates and the actual coordinates as shown in formula (26):

$$\mathcal{L}_2 = \sqrt{\frac{1}{T_{pre}} \sum_{t=1}^{T_{pre}} ((x_t - \tilde{x}_t)^2)} \tag{26}$$

where x_t and $\tilde{x_t}$ represent the predicted lateral position value and actual lateral position value in the prediction horizon T_{pre} respectively.

As explained above, we add the intention probability vector $\boldsymbol{\omega}$ to enhance the prior knowledge learning capability. The fusion of the attention and the LSTM network may be beneficial for the model to improve the prediction performance. It is worth mentioning that although the proposed model is trained on historical data, the proposed model achieves positive results on the dataset, and the inference time of the proposed model does not affect the real-time performance.

IV. Experiments and Results

In this section, we compare the proposed model with the existing models and analyze the results quantitatively. Then the proposed model is verified by real data. All models are trained under the framework of Tensorflow 2.4.0 [43]. The hardware environment of the experimental operation is Intel core i5-6300HQ CPU @2.3GHz and NVIDIA GTX960M.

A. Experimental Details

The implementation parameters of the proposed model are different from those of the traditional Transformer. In the intention prediction, $d_{model} = 256$ and $h = 4$. In the trajectory prediction, $d_{model} = 320$ and $h = 5$. d_{model} is the output dimension of multi-head attention modules. In this paper, we set $d_k = d_v = d_{mdoel}/h$. The details of the intention prediction training are Adam optimizer [42], learning rate $l_r = 0.0001$, and batch size $batch\_size = 400$. The details of the trajectory prediction are Adam optimizer, learning rate $l_r = 0.005$, and batch size $batch\_size = 15$.

B. Data Processing

The lane change time is defined as the lane in which the vehicle is located changed (*lane_ID* change in the two datasets). The advanced prediction time (T) is the time before lane change time, as shown in Fig. 5. The driver's reaction time is 0.3-1.35s and the braking system's action time is about 0.15s [48]. Hence, in order to ensure that the system has sufficient reaction time and braking time, we choose different T ($T = 0s, 0.5s, 1s, 1.5s, 2s$) to evaluate the performance of the proposed model. According to the definition of lane change time, a lane change trajectory with a total length of 25s is obtained from the NGSIM dataset, from data 24s before the lane change time to data 1s after the lane change time. Accordingly, datasets with different data lengths D ($D = 3s, 6s, 9s, 12s, 15s, 18s, 21s$) are generated, respectively, according to the T before lane change time. However, the road length collected by the highD dataset is

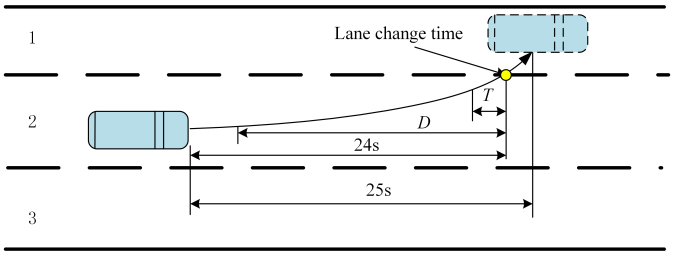


Fig. 5. An example for generating training sets and test sets with different data lengths on the NGSIM dataset.

much smaller than that of the NGSIM dataset, so we select a lane change trajectory of 10s from the highD dataset and choose $D = 6s$. For the NGSIM dataset, there are 4965 samples in all categories, of which 1655 LLC samples,1655 LK samples, and 1655 RLC samples in each data length. For the highD dataset, there are 4075 LLC samples, 4075 LK samples, and 4075 RLC samples, for a total of 12225 samples across all categories. 80% of all screened samples are randomly selected as the training set and 20% as the test set. Specifically, there are 3972 training samples and 993 test samples in the NGSIM dataset, and 9780 training samples and 2445 test samples in the highD dataset. The data labels are one-hot encoded, with (0, 1, 0) for LLC, (0, 0, 1) for RLC, and (1, 0, 0) for LK.

C. Performance Evaluation Metrics of Intention and Trajectory Prediction

Performance evaluation metrics of intention prediction [10] is illustrated as formulas (27)-(30).

$$\text{Average accuracy: } ACC = \frac{TP\_L + TP\_C + TP\_R}{N_{test}} \tag{27}$$

where TP_L is samples whose real labels represent LLC and prediction is also LLC. TP_C is samples whose real labels represent LK and prediction is also LK. TP_R is samples whose true labels represent RLC and prediction is also RLC. N_{test} is the number of testing set samples.

$$\text{precision} = \frac{TP}{TP + FP} \tag{28}$$

$$\text{recall} = \frac{TP}{TP + FN} \tag{29}$$

$$\text{F1} = 2 * \frac{\text{recall} * \text{precision}}{\text{recall} + \text{precision}} \tag{30}$$

where TP represents the number of samples whose true label is positive and predicted label to be positive. FP represents the number of samples whose true label is negative but predicted label to be positive. FN represents the number of samples whose true label is positive but predicted label is negative.

This paper has three lane change intentions, so macro average F1 (Macro_F1) is taken in formula (31).

$$\text{Macro\_F1} = \frac{\text{F1\_L} + \text{F1\_C} + \text{F1\_R}}{3} \tag{31}$$

where F1_L, F1_C, and F1_R are the F1 value of LLC, LK, and RLC, respectively.

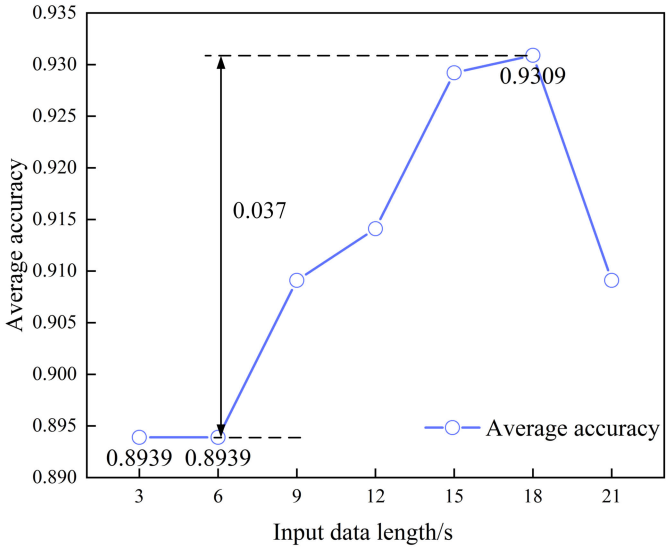


Fig. 6. The influence of different data lengths on average accuracy in intention prediction. We choose 18s as the input length of the model for the NGSIM dataset.

We evaluate the prediction results of the trajectory prediction model using root-mean-squared error (RMSE) [32], [50], [51], which measures the difference between the predicted and actual position at different timestamps, as shown in formula (32).

$$RMSE = \sqrt{\frac{1}{N_{test}} \left(\sum_{j=1}^{N_{test}} \frac{1}{T_{pre}} \sum_{t=1}^{T_{pre}} \left(x_t^{(j)} - \tilde{x}_t^{(j)} \right)^2 \right)} \tag{32}$$

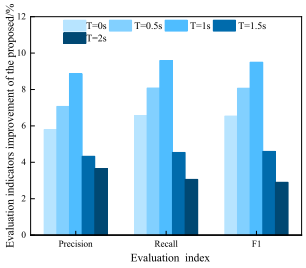
where N_{test} is the number of testing samples, $x_t^{(j)}$ and $\tilde{x}_t^{(j)}$ represent, respectively, the predicted lateral position value and actual lateral position value at t timestamps for the jth testing sample. T_{pre} is the prediction horizon, T_{pre} ranges from 1s to 4s.

D. Performance Analysis of Intention Prediction

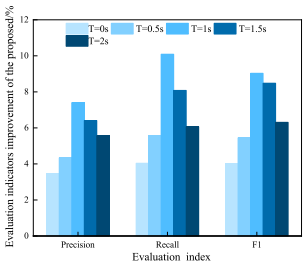
1) Effect of Input Data Length on Average Accuracy of the Proposed Model: In this section, we experimentally pick the optimal data length for the NGSIM dataset. The test set with $T = 1s$ is chosen to verify the effect of inputting different data lengths on the model, and the results are shown in Fig. 6.

As can be seen from Fig. 6, the length of the input data affects the average accuracy of the prediction. After the data length is greater than 6s, the average accuracy of the model gradually increases as the input data length D increases. The average accuracy peaks when the data length $D = 18s$, which is 0.037 higher than that of $D = 6s$. Then the average accuracy starts to decrease when the data length continues to increase. Therefore, the data length $D = 18s$ is chosen for the NGSIM dataset.

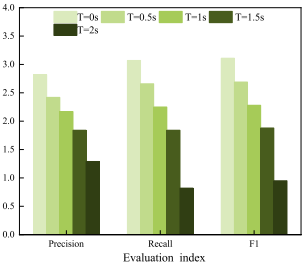
2) Result Analysis of Intention Prediction: The accuracy of the intention prediction has an essential impact on the quality of the trajectory prediction. We compare the precision, recall, and F1 score of CNN [6] and LSTM [23] in the two datasets. Table II shows that T is inversely proportional to its performance evaluation metrics, and the closer the lane change



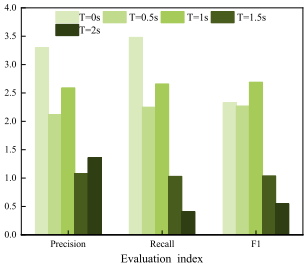
(a) Compared with CNN on NGSIM dataset



(b) Compared with LSTM on NGSIM dataset



(c) Compared with CNN on highD dataset



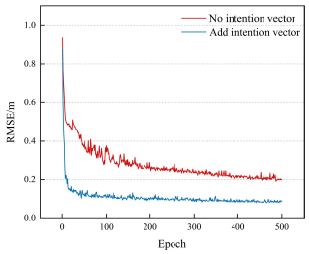
(d) Compared with LSTM on highD dataset

Fig. 7. The improvement of evaluation indicators of the proposed model compared to CNN and LSTM on NGSIM and highD datasets. Our method can drastically increase the precision, the recall and F1.

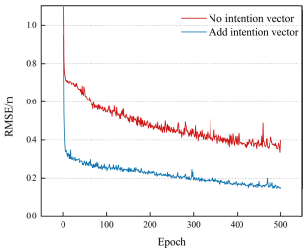
time is, the more accurate the prediction result is. The reason for this is that the closer the lane change time is, the closer the vehicle's lateral position is to the target lane. Moreover, the proposed model outperforms the other two models for LLC, LK, and RLC at all T in the two datasets.

For $T = 1s$, the average precision, the average recall, and Macro_F1 of the proposed model in the two datasets are greater than 90% in the two datasets, which indicates that the proposed model has good intention prediction ability. For $T = 2s$, the proposed model's average precision and average recall can reach about 80% on the NGISM dataset and 90% on the highD dataset. The decrease of each evaluation metric mainly lies in the fact that the lane change in the highway environment finishes a complete lane change process in 5s on average [43]. Assuming that 2.5s before and 2.5s after the lane change, most vehicles do not have a large lateral displacement at $T = 2s$. However, the surrounding environment has lane change conditions, so the model mainly relies on the interaction information between the target vehicle and surrounding vehicles to make predictions.

Besides, to verify the proposed model's stability, it has different degrees of improvement at different T compared with the remaining models. According to formula (33), Fig. 7 gives the improvement rate of the proposed model relative to LSTM and CNN at different T in the two datasets. It can be seen that the proposed model has different degrees of improvement at different T. In the NGSIM dataset, it can be concluded that $\Delta I_T^{M,1}$ and $\Delta I_T^{M,2}$ are the greatest at $T = 1s$, and they are in the range of 8.86-9.59% and 6.41-8.48%, respectively. $\Delta I_T^{M,1}$ is a larger increase at $T = 0 \sim 1s$, but $\Delta I_T^{M,2}$ has greatly improvement at $T = 1 \sim 2s$. Besides, $\Delta I_T^{M,1}$ and $\Delta I_T^{M,2}$ are at 2.9-3.65% and 5.57-6.31% when $T = 2s$, respectively. In the highD dataset, $\Delta I_T^{M,1}$ and $\Delta I_T^{M,2}$ reach 2.17-3.11%



(a) On NGSIM dataset



(b) On highD dataset

Fig. 8. Comparison of loss values of the trajectory prediction training set with intention probability vector. The plot shows that adding intention probability vector can guarantee a lower RMSE value for the trajectory prediction than no intention probability vector.

and 2.12-3.48% at $T = 0 \sim 1s$,respectively, but they only achieve 0.82-1.88% and 0.41-1.08% at $T = 1.5s, 2s$. Therefore, compared with the NGSIM dataset, the improvement of the proposed model is slight on the highD dataset. This outcome may be due to differences in the accuracy of the datasets. These results imply that the accuracy of the intention recognition for RLC, LLC, and LK of target vehicles can be improved.

$$\Delta I_T^{M,k} = I_T^M - I_T^k, T = 0s, 0.5s, 1s, 1.5s, 2s, k = 1, 2 \tag{33}$$

where $\Delta I_T^{M,1}$ and $\Delta I_T^{M,2}$ is the improvement of the evaluation metrics of the proposed model relative to CNN and LSTM, respectively, I_T^M is the average value of the evaluation metrics of the proposed model at different T, I_T^k is the average value of the evaluation metrics of the compared models.

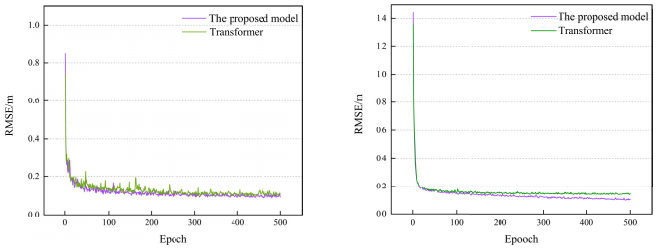
E. Trajectory Prediction Results and Analysis

1) Influence of the Intention Probability Vector and LSTM on the Trajectory Prediction: As stated in Fig. 8, from the decreasing trend of the loss value during the training process, the intention probability vector accelerates the convergence speed and reduces the RMSE value. In the NGSIM and highD datasets, the RMSE value of adding the intention probability vector (take the average of the last 5 epochs) is reduced by 57.27% and 58.70% compared to the RMSE value without the intention probability vector, respectively. Therefore, the inclusion of the intention probability vector is helpful to the results of the trajectory prediction. The reason may be that the intention probability vector is equivalent to a direction vector. When it is fed into the proposed model as a feature, it will have an indicative effect on the trajectory prediction of lane change vehicles, which allows the model to have some prior knowledge. As a result,the model is better able to predict the trajectories of lane change vehicles.

Similarly, Fig. 9 presents the comparison of the proposed model, adding LSTM as the feed-forward layer and Transformer. Compared with Transformer, the RMSE value of the proposed model is reduced by 7.52% and 27.3% on the NGSIM and highD datasets, respectively. Therefore, adding LSTM improves the model's ability to extract sequence features, giving the proposed model better prediction performance.

TABLE II
COMPARISON OF PRECISION, RECALL AND F1 OF DIFFERENT LANE CHANGING TYPES OF EACH MODELS ON THE NGSIM AND THE HIGHD DATASETS

Dataset	Model	Intention	T=2s			T=1.5s			T=1s			T=0.5s			T=0s		
			Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
NGSIM	Proposed model	LLC	91.84%	68.18%	78.26%	93.10%	81.82%	87.10%	96.88%	93.94%	95.38%	95.38%	93.94%	94.66%	98.51%	100.00%	99.25%
		RLC	76.92%	90.91%	83.33%	84.51%	90.91%	87.59%	90.28%	98.48%	94.20%	94.20%	98.48%	96.30%	94.29%	100.00%	97.06%
		LK	81.69%	87.88%	84.67%	82.61%	86.36%	84.44%	93.55%	87.88%	90.63%	92.19%	89.39%	90.77%	100.00%	92.42%	96.06%
		Average	**83.48%**	**82.32%**	**82.09%**	**86.74%**	**86.36%**	**86.38%**	**93.57%**	**93.43%**	**93.40%**	**93.93%**	**93.94%**	**93.91%**	**97.60%**	**97.47%**	**97.46%**
	CNN	LLC	83.64%	69.70%	76.03%	84.48%	74.24%	79.03%	91.53%	81.82%	86.40%	93.33%	84.85%	88.89%	98.39%	92.42%	95.31%
		RLC	74.03%	86.36%	79.72%	75.64%	89.39%	81.94%	75.95%	90.91%	82.76%	77.78%	95.45%	85.71%	82.28%	98.48%	89.66%
		LK	81.82%	81.82%	81.82%	87.10%	81.82%	84.38%	86.67%	78.79%	82.54%	89.47%	77.27%	82.93%	94.74%	81.82%	87.80%
		Average	**79.83%**	**79.29%**	**79.19%**	**82.41%**	**81.82%**	**81.78%**	**84.71%**	**83.84%**	**83.90%**	**86.86%**	**85.86%**	**85.84%**	**91.80%**	**90.91%**	**90.92%**
	LSTM	LLC	82.22%	56.06%	66.67%	86.67%	59.09%	70.27%	94.23%	74.24%	83.05%	94.92%	84.85%	89.60%	98.41%	93.94%	96.12%
		RLC	66.67%	87.88%	75.82%	68.18%	90.91%	77.92%	74.12%	95.45%	83.44%	79.01%	96.97%	87.07%	85.71%	100.00%	92.31%
		LK	84.85%	84.85%	84.85%	86.15%	84.85%	85.50%	90.16%	83.33%	86.61%	94.83%	83.33%	88.71%	98.28%	86.36%	91.94%
		Average	**77.91%**	**76.26%**	**75.78%**	**80.33%**	**78.28%**	**77.90%**	**86.17%**	**84.34%**	**84.37%**	**89.59%**	**88.38%**	**88.46%**	**94.13%**	**93.43%**	**93.46%**
HighD	Proposed model	LLC	98.64%	88.96%	93.55%	98.76%	97.55%	98.15%	99.38%	98.77%	99.08%	98.79%	100.00%	99.39%	99.39%	99.39%	99.39%
		RLC	97.33%	89.57%	93.29%	99.37%	96.93%	98.14%	99.39%	99.39%	99.39%	99.39%	99.39%	99.39%	100.00%	99.39%	99.69%
		LK	81.77%	96.32%	88.45%	94.67%	98.16%	96.39%	98.17%	98.77%	98.47%	99.38%	98.16%	98.77%	98.78%	99.39%	99.08%
		Average	**92.58%**	**91.62%**	**91.76%**	**97.60%**	**97.55%**	**97.56%**	**98.98%**	**98.98%**	**98.98%**	**99.18%**	**99.18%**	**99.18%**	**99.39%**	**99.39%**	**99.39%**
	CNN	LLC	92.36%	88.96%	90.63%	93.57%	98.16%	95.81%	94.77%	100.00%	97.31%	92.09%	100.00%	95.88%	92.09%	100.00%	95.88%
		RLC	96.58%	86.50%	91.26%	96.97%	98.16%	97.56%	96.99%	98.77%	97.87%	98.19%	100.00%	99.09%	97.60%	100.00%	98.79%
		LK	84.95%	96.93%	90.54%	96.73%	90.80%	93.67%	98.68%	91.41%	94.90%	100.00%	89.57%	94.50%	100.00%	88.96%	94.16%
		Average	**91.29%**	**90.80%**	**90.81%**	**95.76%**	**95.71%**	**95.68%**	**96.81%**	**96.73%**	**96.70%**	**96.76%**	**96.52%**	**96.49%**	**96.57%**	**96.32%**	**96.28%**
	LSTM	LLC	90.74%	90.18%	90.46%	96.36%	97.55%	96.95%	94.15%	98.77%	96.41%	93.64%	99.39%	96.43%	93.64%	99.39%	93.64%
		RLC	92.55%	91.41%	91.98%	96.95%	97.55%	97.25%	96.99%	98.77%	97.87%	98.19%	100.00%	99.09%	95.32%	100.00%	98.19%
		LK	90.36%	92.02%	91.19%	96.25%	94.48%	95.36%	98.03%	91.41%	94.60%	99.33%	91.41%	95.21%	99.31%	88.34%	99.33%
		Average	**91.22%**	**91.21%**	**91.21%**	**96.52%**	**96.52%**	**96.52%**	**96.39%**	**96.32%**	**96.29%**	**97.06%**	**96.93%**	**96.91%**	**96.09%**	**95.91%**	**97.06%**



(a) On NGSIM dataset (b) On highD dataset

Fig. 9. Comparison of loss values of the proposed model and Transformer. Our use of LSTM as a feed-forward layer of Transformer is beneficial for improving the prediction effect.

2) RMSE Analysis of Trajectory Prediction Results: We choose KF and LSTM as baseline models to evaluate the performance of the proposed model, which allows us to find the best one. Table III compares the results of the proposed model with KF and LSTM [23] in the two datasets. On the NGSIM dataset, the RMSE of the proposed model is reduced by 1.72 and 0.5 compared to KF and LSTM, respectively. As a result, the proposed model shows better prediction performance for both short- and long-horizon predictions. Compared to the NGSIM dataset, the proposed model shows a slight improvement in prediction on the highD dataset. To visualize the prediction results, a segment of the actual trajectory is selected from the test set to be compared with the predicted trajectory, as shown in Fig. 10. Compared to LSTM and KF, the predicted trajectory of the proposed model matches with the actual lateral trajectory, which shows that the proposed model has better performance in the long prediction horizon. The reason is that the proposed model separates the encoded output from the decoded sequence, while LSTM accumulates both to its hidden state, thus controlling what to remember or forget each time [36].

F. Comparison of the Inference Time of Each Model

Intention prediction and trajectory prediction require models that are fast and computationally efficient. The inference time of the model in trajectory prediction and intention prediction are compared in Table IV. Trajectory prediction is the inference time for predicting the future 4s trajectory of the target vehicle. Intention prediction is the inference time for predicting the intention probability of a target vehicle with

TABLE III
COMPARISON OF DIFFERENT PREDICTION HORIZON OF EACH MODEL ON THE PREDICTION PERFORMANCE ON THE NGSIM AND THE HIGHD DATASETS

Dataset	Prediction horizon(s)	RMSE(m)		
		KF	LSTM	Proposed model
NGSIM	1	2.19	1.68	1.18
	2	3.70	3.36	2.83
	3	5.02	4.71	4.22
	4	6.08	6.82	5.82
highD	1	0.56	0.44	0.41
	2	1.01	0.82	0.79
	3	1.4	1.15	1.11
	4	1.75	1.43	1.4

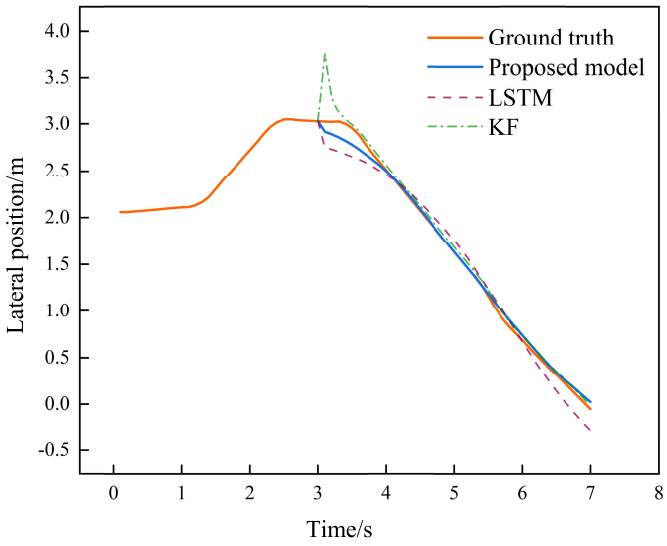


Fig. 10. Comparison between predicted lateral trajectory and actual lateral trajectory. The length of the historical lateral trajectory put into the proposed model is 3s. The trajectory predicted by the proposed model is closer to the ground truth.

input data of length 18s. From Table IV, in the trajectory prediction, KF has the fastest inference speed, followed by LSTM, and the proposed model is slightly higher than LSTM by 0.013s. In the intention prediction, the proposed model takes 0.014s and 0.011s more time than LSTM and CNN, respectively. The inference time consumed by the proposed model is highly close to other methods in both trajectory and intention prediction. It takes 0.13s to complete the full trajectory prediction with the proposed model. The reason for the slightly higher inference time of the proposed model is the increased inference time due to a large number of matrix operations within it. However, the dimensionality of the input is considerably reduced by feeding the lateral position information and the surrounding information into the model separately and in parallel.

G. Model Effect Verification in Different Lane Change Scenarios

In order to verify the application capability of the model in real scenarios, a trajectory sequence of LLC and a trajectory sequence of RLC are selected from the test set for intention and trajectory prediction. The sliding time window method updates the input with an intention prediction frequency of 1s. When RLC and LLC probability reaches 90% or more, the proposed model performs trajectory prediction in the prediction horizon of 1s and 3s. For the convenience of description, we mark the time on the abscissa in Fig. 11 (a) as $T1$, and the time on the abscissa in Fig. 11 (b) as $T2$.

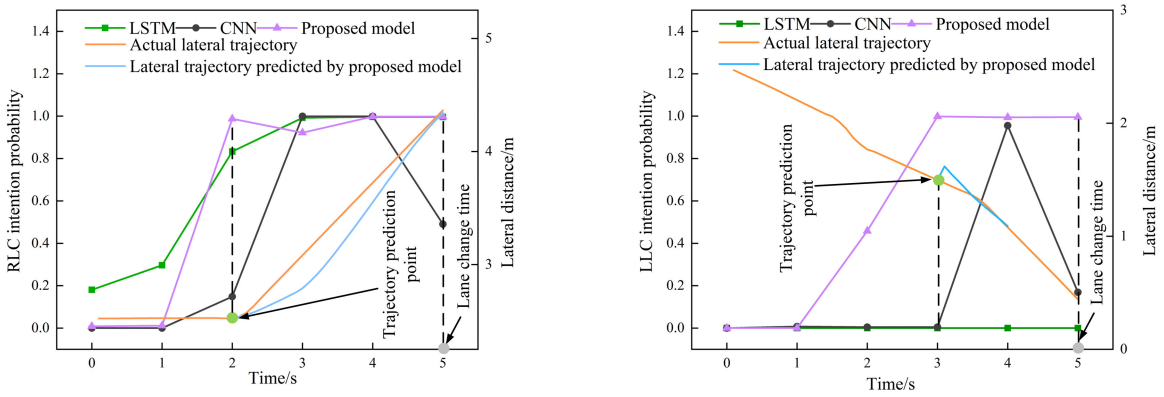
As shown in Fig. 11 (a), both LSTM and the proposed model make an accurate prediction, but the accuracy of LSTM is lower than that of the proposed model. The CNN model is unstable. Specifically, the vehicle's lateral position does not change of the trajectory sequence in $T1 = 0 \sim 2s$. Then, the LSTM model predicts the RLC probability of 83.31% at $T1 = 2s$, while the CNN model only has 14.86%. More importantly, the proposed model makes a more accurate prediction based on the interactive information 3s before the lane change occurs. The probability of predicting the target vehicle to switch right lanes is over 90%. At $T1 = 3s$, there is a slight decrease in the RLC probability of the proposed model, which the increase in lateral position may cause. Finally, the RLC probability gradually increases reaches a plateau after $T1 = 3s$. Then, the proposed model outputs the predicted lateral trajectory in Fig. 11 (a), combining the intention probability vector output with the historical lateral trajectory input at $T1 = 2s$. In $T1 = 2 \sim 4s$, the predicted lateral position deviates from the real lateral position, while prediction and the actual close match in $T1 = 4s, 5s$. From a practical point of view, most vehicle lane changes, in reality, are smooth lane changes. Hence, the lateral trajectory predicted by the proposed model is more moderate and smooth, which is consistent with the actual lateral trajectory of the lane change.

Looking at the whole process in Fig.11 (b), CNN and LSTM's prediction results could be better when faced with an uneven trajectory sequence. At the same time, the proposed model can make an accurate prediction. However, the CNN model only makes an accurate prediction at the fourth moment and does not predict the intention of the LLC at other moments. And the LSTM model does not recognize LLC's intention. After $T2 = 1s$, the predicted probability of LLC increases as the vehicle's lateral position gradually increases, which is shown in Fig. 11 (b). At $T2 = 3s$, the probability of the model predicting the intention of LLC is close to 100%. The reason is that the vehicle's lateral position deviates from the driving lane at this time. The current interactive information may not meet the lane change requirement, unlike Fig. 11 (a). The model can only be accurate according to the change of lateral position. Subsequently, the proposed model predicts the trajectory in the horizon of 1s, and the result is closer to the actual trajectory.

It can be learned that, in terms of intention prediction, the proposed model can identify the RLC intention of the target vehicle 3s in advance and the LLC intention 2s in advance in Fig. 11 (a),(b). Compared to LSTM and CNN, the proposed model can accurately identify the chemical lane change intentions of the vehicle. In terms of trajectory prediction, the predicted trajectory of the proposed model is more in line with the actual trajectory, which may be due to noise in the data. Therefore, from the practical point of view, the predicted trajectory of this model is more in line with the actual driving trajectory of the target vehicle.

TABLE IV
TIME-CONSUMING COMPARISON OF EACH MODEL AND TIME-CONSUMING OF COMPLETE PREDICTION

	Trajectory prediction			Intention prediction		
Model	LSTM	KF	Proposed model	LSTM	CNN	Proposed model
Reasoning time (s)	0.055	0.007	0.068	0.048	0.051	0.062



(a) Intention and trajectory predictions for RLC of the target vehicle (b) Intention and trajectory predictions for LLC of the target vehicle

Fig. 11. Intention and trajectory predictions of different lane change types. Over time and with changes in the lateral position of the target vehicle, each model identifies changes in the probability of changing lanes and the prediction results of the lateral trajectory of the proposed model.

H. Performance Discussion

1) Different Results: The above experimental results show that the prediction error of the highD dataset is smaller than that of the NGSIM dataset, and there are differences in the results. The first possible reason is that the motion trajectories, such as the position and velocity of the highD dataset, are more accurate than the NGSIM dataset [41]. In addition, the size of the highD dataset is about 16 times that of the NGSIM dataset, which means that the highD dataset covers more abundant scenes and has a broader data distribution [41]. Thus, there are enough samples to train the network. Moreover, there is a similar situation in recent trajectory prediction research [32], [50], [51].

2) Computation Load: Reasoning time is an essential metric to measure real-time performance. Thus, to validate the computation performance of the proposed model, we use the computer platform of the experimental operation described in Section IV. The computational power of the platform used in this paper is greatly smaller than that of mainstream autonomous vehicle platforms. For example, the computational power of Tesla's dual-chip FSD autonomous driving computing platform is 144 Tera Operations Per Second (TOPS), while the computational power of the platform used in this paper is 5 TOPS. We compare the reasoning time of each model in Table IV. In our calculations, the inference time of intention prediction and trajectory prediction is 0.062s and 0.068s, respectively, which means that the proposed model can infer the prediction results in real time.

3) Implementation: The prediction model is trained on historical data from NGSIM and highD datasets. As long as the reasoning time of the prediction model is small enough, the prediction model can be implemented in real scenario applications. Based on our experimental results in Table IV, it takes 0.13s for the proposed model to make the complete prediction, which is much smaller than the typical human decision-making time is 0.3-1.35s [48]. It is worth mentioning that the mainstream autonomous vehicle platforms have superior computing power than the computer used in this paper. When the proposed model is implemented in real autonomous vehicles, the computational time of the proposed model could be smaller, which verifies its effectiveness in real-time application scenarios.

4) Application Scenarios: According to Section II, the NGSIM and highD datasets depict a scenario where vehicles are driving under a highway, so the proposed model is suitable for prediction in highway scenarios. For urban roads and other roads, the applicability of the proposed model is not known, which is a future research direction.

V. CONCLUSION

This paper proposes an integrated prediction model based on Transformer for lane change intention prediction and trajectory prediction of target vehicles. An intention prediction model is first established based on Transformer. Then, combined with the intention probability vector as a part of the input, a trajectory prediction model based on Transformer is established. The results show that by changing the input form, the proposed model's performance is better than that of existing models on both NGSIM and highD datasets. In terms of the intention prediction performance, the evaluation metrics of the proposed model on the NGSIM dataset can reach 93.40-93.57% at $T = 1s$, and their improvement are 7.40-10.09% and 8.86-9.59% compared with CNN and LSTM, respectively. Meanwhile, the evaluation metrics of the proposed model on the highD dataset can achieve 98.98% at $T = 1s$, and they increase by 2.17-2.69%. Moreover, in the two datasets, the evaluation metrics also have improvement at all T. For the performance of the trajectory prediction, the intention probability vector is used as a part of the trajectory prediction input, which is compared with the situation without the intention probability

vector. The RMSE value is significantly reduced by adding an intention probability vector, which can improve the quality of the trajectory prediction. The proposed model is superior to existing models both in the short horizon and long horizon prediction, which provides a new idea for designing a novel algorithm for trajectory prediction. At the same time, the proposed model can meet the real-time requirements of autonomous vehicles. The model can provide a reference for designing advanced risk-warning algorithms for autonomous vehicles.

At present, this article has only verified the effect of the prediction model in the highway scenarios. In the future, the application of the model should be expanded to a wider range of applications.

REFERENCES

[1] A. Sarker et al., "A review of sensing and communication, human factors, and controller aspects for information-aware connected and automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 7–29, Jan. 2020.

[2] D. Dolgov. (Sep. 2016). *Google Self-Driving Car Project Monthly Report*. Accessed: Jan. 20, 2017. [Online]. Available: https://static.googleusercontent.com/media/www.google.com/deselfdrivingcar/files/reports/report-0916.pdf

[3] L. Hu, H. Li, P. Yi, J. Huang, M. Lin, and H. Wang, "Investigation on AEB key parameters for improving car to two-wheeler collision safety using in-depth traffic accident data," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 113–124, Jan. 2023.

[4] 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, doi: 10.1016/j.trc.2021.103414.

[5] N. Lyu, J. Wen, Z. Duan, and C. Wu, "Vehicle trajectory prediction and cut-in collision warning model in a connected vehicle environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 966–981, Feb. 2022.

[6] R. Izquierdo, A. Quintanar, I. Parra, D. Fernandez-Llorca, and M. A. Sotelo, "Experimental validation of lane-change intention prediction methodologies based on CNN and LSTM," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3657–3662.

[7] H. Q. Dang, J. Furnkranz, A. Biedermann, and M. Hoepfl, "Time-to-lane-change prediction with deep learning," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–7.

[8] J. Wang, Z. Zhang, and G. Lu, "A Bayesian inference based adaptive lane change prediction model," *Transp. Res. C, Emerg. Technol.*, vol. 132, Nov. 2021, Art. no. 103363.

[9] K. Saleh, M. Hossny, and S. Nahavandi, "Contextual recurrent predictive model for long-term intent prediction of vulnerable road users," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3398–3408, Aug. 2020.

[10] D. Frossard, E. Kee, and R. Urtasun, "DeepSignals: Predicting intent of drivers through visual signals," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 9697–9703.

[11] 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.

[12] J. Hong, B. Sapp, and J. Philbin, "Rules of the road: Predicting driving behavior with a convolutional model of semantic interactions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8454–8462.

[13] S. Casas, W. Luo, and R. and Urtasun, "IntentNet: Learning to predict intention from raw sensor data," in *Proc. 2nd Conf. Robot Learn.*, Oct. 2018, pp. 947–956.

[14] Z. Shou, Z. Wang, K. Han, Y. Liu, P. Tiwari, and X. Di, "Long-term prediction of lane change maneuver through a multilayer perceptron," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 246–252.

[15] 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, doi: 10.1109/IVS.2016.7535559.

[16] 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.

[17] T. Yagi, K. Mangalam, R. Yonetani, and Y. Sato, "Future person localization in first-person videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7593–7602.

[18] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "STGAT: Modeling spatial–temporal interactions for human trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6272–6281.

[19] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1468–1476.

[20] H. Zhang, Y. Wang, J. Liu, C. Li, T. Ma, and C. Yin, "A multi-modal states based vehicle descriptor and dilated convolutional social pooling for vehicle trajectory prediction," 2020, *arXiv:2003.03480*.

[21] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder–decoder architecture," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1672–1678.

[22] G. Chen, L. Hu, Q. Zhang, Z. Ren, X. Gao, and J. Cheng, "ST-LSTM: Spatio-temporal graph based long short-term memory network for vehicle trajectory prediction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 608–612, doi: 10.1109/ICIP40778.2020.9191332.

[23] X. Ji, C. Fei, X. He, Y. Liu, and Y. Liu, "Intention recognition and trajectory prediction for vehicles using LSTM network," *China J. Highway Transp.*, vol. 32, no. 6, pp. 34–42, Jun. 2019.

[24] O. Scheel, N. S. Nagaraja, L. A. Schwarz, N. Navab, and F. Tombari, "Attention-based lane change prediction," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 8655–8661.

[25] 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, doi: 10.1109/TIV.2022.3161785.

[26] H. Han and T. Xie, "Lane change trajectory prediction of vehicles in highway interweaving area using Seq2Seq-attention network," *China J. Highway Transp.*, vol. 33, no. 6, pp. 106–118, May 2020.

[27] J. Wang, Q. Zhang, and D. Zhao, "Highway lane change decision-making via attention-based deep reinforcement learning," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 3, pp. 567–569, Mar. 2022.

[28] J. Wang, P. Wang, C. Zhang, K. Su, and J. Li, "F-Net: Fusion neural network for vehicle trajectory prediction in autonomous driving," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 4095–4099.

[29] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Relational recurrent neural networks for vehicle trajectory prediction," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 1813–1818.

[30] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Non-local social pooling for vehicle trajectory prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 975–980.

[31] M. Fu, T. Zhang, W. Song, Y. Yang, and M. Wang, "Trajectory prediction-based local spatio-temporal navigation map for autonomous driving in dynamic highway environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6418–6429, Jul. 2022, doi: 10.1109/TITS.2021.3057110.

[32] Y. Cai et al., "Environment-attention network for vehicle trajectory prediction," *IEEE Trans. Veh. Technol.*, vol. 70, no. 11, pp. 11216–11227, Nov. 2021.

[33] A. Vaswani, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 5998–6008.

[34] J. Guo et al., "CMT: Convolutional neural networks meet vision transformers," 2021, *arXiv:2107.06263*.

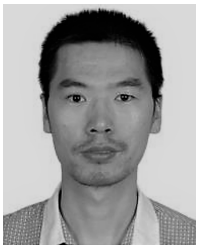
[35] Y. Liu, J. Zhang, L. Fang, Q. Jiang, and B. Zhou, "Multimodal motion prediction with stacked transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 7577–7586.

[36] F. Giuliari, I. Hasan, M. Cristani, and F. Galasso, "Transformer networks for trajectory forecasting," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 10335–10342.

[37] M. Zhu et al., "TransFollower: Long-sequence car-following trajectory prediction through transformer," 2022, *arXiv:2202.03183*.

[38] H. Zhou, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. Assoc. Advance Artif. Intell. (AAAI)*, Feb. 2021, pp. 11106–11115.

[39] S. Zhou, "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2019, pp. 5244–5254.
[40] *Afederal Highway Administration, NGSIM-Next Generation Simulation*. Accessed: Oct. 11, 2021. [Online]. Available: https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm
[41] 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.
[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[43] P. B. M. Abadi and A. Agarwal. (2015). *TensorFlow: Large- Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: http://tensorflow.org/
[44] M. N. Azadani and A. Boukerche, "Driving behavior analysis guidelines for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6027–6045, Jul. 2022, doi: 10.1109/TITS.2021.3076140.
[45] T. Han, J. Jing, and U. Ozguner, "Driving intention recognition and lane change prediction on the highway," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 957–962.
[46] M. Schreier, V. Willert, and J. Adamy, "An integrated approach to maneuver-based trajectory prediction and criticality assessment in arbitrary road environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2751–2766, Oct. 2016.
[47] C. Fei, X. He, and X. Ji, "Multi-modal vehicle trajectory prediction based on mutual information," *IET Intell. Transp. Syst.*, vol. 14, no. 3, pp. 148–153, Feb. 2020.
[48] S. Doroudgar, H. M. Chuang, P. J. Perry, K. Thomas, K. Bohnert, and J. Canedo, "Driving performance comparing older versus younger drivers," *Traffic Injury Prevention*, vol. 18, no. 1, pp. 41–46, Jan. 2017.
[49] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.
[50] X. Chen, H. Zhang, F. Zhao, Y. Hu, C. Tan, and J. Yang, "Intention-aware vehicle trajectory prediction based on spatial–temporal dynamic attention network for Internet of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19471–19483, Oct. 2022, doi: 10.1109/TITS.2022.3170551.
[51] Q. Dong, T. Jiang, T. Xu, and Y. Liu, "Graph-based planning-informed trajectory prediction for autonomous driving," in *Proc. 6th CAA Int. Conf. Veh. Control Intell. (CVCI)*, Oct. 2022, pp. 598–614.



Kai Gao was born in Baoding, Hebei, China, in 1985. He received the B.S. and Ph.D. degrees from Central South University, Changsha, China, in 2008 and 2014, respectively. He joined the Changsha University of Science and Technology in 2015. His research interests include connected and automated vehicles, and intelligent transportation systems theory and applications.

Xunhao Li is currently pursuing the M.S. degree with the Changsha University of Science and Technology. His research interests include deep learning and intelligent vehicles.

Bin Chen received the B.S. and Ph.D. degrees from the School of Automation, Central South University, China, in 2013 and 2021, respectively. He is currently a Lecturer with the College of Automotive and Mechanical Engineering, Changsha University of Science and Technology. His current research interests include model-free control, train braking control, and cooperative control.

Lin Hu was born in Hunan, China, in 1978. He received the Ph.D. degree in engineering from the College of Mechanical and Automotive Engineering, Hunan University, in 2008. He is currently a Professor and an Associate Dean with the College of Automotive and Mechanical Engineering, Changsha University of Science and Technology, China. His research interests include intelligent vehicle, vehicle dynamics, and traffic safety. He is the winner of Hunan Outstanding Youth Fund in 2018. He received the second prize of Scientific and Technological Progress in Hunan Province in February 2017 and February 2019.

Jian Liu was born in Loudi, Hunan, China. He received the B.S. degree from Shaoyang University, Shaoyang, China, in 2020. He is currently pursuing the M.S. degree with the Changsha University of Science and Technology. His research interests include intelligent transportation and machine learning.

Ronghua Du was born in Hunan in 1973. He received the Ph.D. degree from the College of Computer Science and Technology, National University of Defense Technology, China. He has been the Director of the Institute of Intelligent Transportation and Vehicle-Road Collaborative Technology, Changsha University of Science and Technology, since March 2011. He is currently a Professor with the Changsha University of Science and Technology. His research interests include cooperative vehicle-infrastructure systems, ITS, vehicle dynamics, and connected automated vehicles.

Yongfu Li (Senior Member, IEEE) received the Ph.D. degree in control science and engineering from Chongqing University, Chongqing, China, in 2012. He is currently a Full Professor of control science and engineering, computer science and technology, and information and communication engineering with the Chongqing University of Posts and Telecommunications. His research interests include connected and automated vehicles, intelligent transportation systems, and cooperative control theory and applications.