A Hybrid Trajectory Prediction Framework for Automated Vehicles With Attention Mechanisms

Mingqiang Wang, Lei Zhang, *Member, IEEE*, Jun Chen, *Senior Member, IEEE*, Zhiqiang Zhang, Zhenpo Wang, *Member, IEEE*, and Dongpu Cao, *Senior Member, IEEE*

***Abstract*— The driving safety of automated vehicles is largely dependent on accurately predicting the motions of surrounding vehicles. However, the existing approaches invariably neglect the impact of the ego vehicle's future behaviors on the surrounding vehicles and lack model explainability for the prediction results. To tackle these issues, a hybrid trajectory prediction framework based on long short-term memory (LSTM) encoding is proposed. It introduces a reactive social convolution structure to model the planned trajectory of the ego vehicle with the historical trajectories of the surrounding vehicles to reduce uncertainty in potential trajectories. Furthermore, a spatio-temporal attention mechanism is presented to quantitatively describe the contributions of historical trajectories and interactions among the surrounding vehicles to the prediction results by appropriate weights setting. Finally, the proposed scheme is comprehensively evaluated based on the NGSIM and HighD datasets. The results demonstrate that the proposed approach can elucidate the prediction process from a spatio-temporal perspective and outperform other state-of-the-art methods under different traffic scenarios. The root-mean-square errors on the NGSIM and HighD datasets are reduced to less than 3.65 m and 2.36 m over a time horizon of 5 s, respectively. The qualitative analysis on the reliability and reactivity is also presented.**

***Index Terms*— Automated vehicles, interaction, long short-term memory (LSTM), trajectory prediction.**

I. Introduction

AUTOMATED vehicles (AVs) have emerged as a significant trend in the automotive industry. In order to operate safely and efficiently, AVs require accurate prediction of the future trajectories of the surrounding vehicles. Trajectory prediction acts as a bridge between the perception and decision modules, allowing vehicle to better understand its surroundings and make proper decisions about driving behaviors [1], [2]. Despite the increasing interest in predicting the trajectories of surrounding vehicles in autonomous driving applications, ensuring prediction accuracy and reliability still faces great challenges [3], [4].

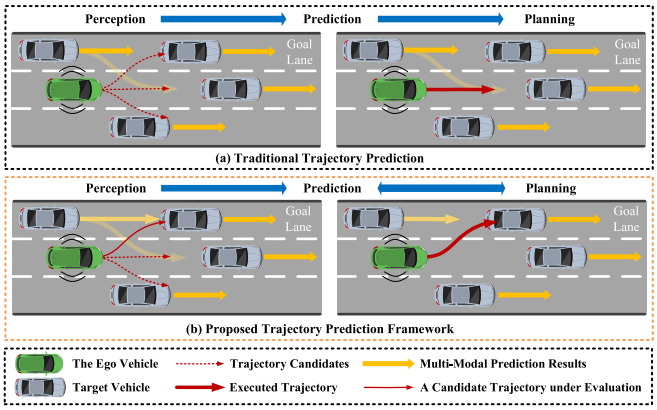

Fig. 1. Illustration of the traditional trajectory prediction and the proposed trajectory prediction approach: (a) and (b) show the corresponding frameworks and the corresponding driving performance in a dense traffic scenario, respectively. Specifically, the ego vehicle with the traditional method struggles to merge into the left lane since it cannot reason about how other agents would react to its candidate trajectory, while the proposed trajectory prediction method compresses the target agent's free space and reasons that the target agent would slow down, allowing it to perform a safe lane merge maneuver.

Many approaches adopt the "perception–prediction–planning" workflow as illustrated in Fig. 1(a). In this process, the perception module detects the motion states of surrounding vehicles, the prediction module forecasts their possible future paths, and the planning module creates a collision-free trajectory based on the information collected by the perception and prediction modules. However, this method assumes that the behaviors of the ego vehicle have no impact on other vehicles and neglects the interactions between vehicles, which would remarkably compromise its efficacy in dense traffic situations [5]. For instance, as shown in Fig. 1(a), it is difficult for the ego vehicle to merge into the left lane in heavy traffic conditions due to neglecting the influence of the ego vehicle's behaviors on the predicted paths of other vehicles. On the other hand, human drivers always anticipate potential actions of other drivers based on their maneuvers, as indicated in Fig. 1(b). Thereby, forecasting future trajectories can be regarded as a process of restricting the free driving space of a vehicle, which is closely related not only to the prediction accuracy [6], but also to the reactive quality of the prediction. In addition, vehicle trajectory variations are subject to the interactions with other surrounding vehicles in addition to the vehicle's historical trajectory. For instance, the correlation of different surrounding vehicles for prediction

Manuscript received 8 May 2023; revised 27 October 2023 and 5 December 2023; accepted 8 December 2023. Date of publication 25 December 2023; date of current version 19 September 2024. This work was supported in part by the Beijing Municipal Science and Technology Commission via the Beijing Nova Program under Grant Z201100006820007. *(Corresponding author: Lei Zhang.)*

Mingqiang Wang, Lei Zhang, Zhiqiang Zhang, and Zhenpo Wang are with the Collaborative Innovation Center for Electric Vehicles in Beijing and the National Engineering Research Center for Electric Vehicles, Beijing Institute of Technology, Beijing 100081, China (e-mail: lei_zhang@bit.edu.cn).

Jun Chen is with the Department of Electrical and Computer Engineering, Oakland University, Rochester, MI 48309 USA.

Dongpu Cao is with the School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China.

Digital Object Identifier 10.1109/TTE.2023.3346668

cannot be efficiently generalized. However, analyzing and quantifying such abstract concepts are lacking in the existing literature, and this leads to poor model interpretability. In a nutshell, the existing trajectory prediction methods still face great challenges in improving prediction accuracy and model interpretability.

To address these challenges, numerous studies have been directed to developing enabling trajectory prediction methods that consider the interactions among multiple vehicles. These can be generally categorized into three groups: model-based, maneuver-based, and learning-based approaches [7]. Model-based methods use vehicle kinematic or dynamic models in conjunction with filtering algorithms, such as Kalman filters, for trajectory prediction [8], [9], [10]. They are computationally efficient and typically effective in predicting trajectories in a time frame of one or two seconds. However, these methods exhibit reduced accuracy over long prediction horizons, which limits the autonomy of AVs in heavy traffic situations. Maneuver-based methods segment vehicle motion into longitudinal and lateral driving behaviors to achieve multimodal predictions. The well-established methods mainly include the Bayesian networks [9], Monte-Carlo method [10], and hidden Markov models [11]. As maneuvers are manually defined, they are often interpretable. However, the level of detail involved in maneuver classification increases substantially with the increasing complexity of traffic scenarios [12]. Consequently, it requires significant manual calibration and enormous computational resources.

Many deep-learning-based methods have been developed for vehicle motion prediction, including noninteractive prediction, multiagent interaction modeling, and reactive prediction methods [13], [14], [15], [16], [17]. Given that vehicle behaviors are interrelated especially in dense traffic scenarios, this study proposes deep learning pattern-based methods to account for these interrelationships. Noninteractive prediction methods use recurrent neural networks and their derivatives, such as long short-term memory (LSTM) networks, for sequence modeling and generation [18], [19]. These methods can extract long-term relationships between different features and model their mutual reliance [20], [21]. While such methods have been proven effective, they may struggle to capture the underlying interactions in complex traffic scenarios. Assuming that vehicles can influence each other's behaviors, the prediction models that consider intervehicle interactions treat vehicles as interactive agents to account for inherent motion uncertainty with multiple agents while simulating their interactions through proper neural network models. For instance, a social pooling mechanism with multimode distributions [22], [23], [24] was proposed to achieve high prediction accuracy. The interactions among multiple agents were simulated through graph neural networks in [25], [26], and [27]. Additionally, generative adversarial networks (GANs) have made great progress in capturing intractable high-dimensional probabilistic distribution. On top of this, the social GAN [28], [29], [30] manipulated social interactions among agents and the constraints from the scene context via convolutional fusion operation retain the spatial structure of agents. However, the game among multiple vehicles and the correlation between two trajectories need to be analyzed by using other methods. Hence, to fit the correlations of different interactions in traffic, attention-based mechanism schemes [31], [32], [33], [34] are extended based on the structure of convolutional social pooling to analyze traffic scenes by modeling interaction features among multiple vehicles and incorporating multisource information including soft and hard attention [35] and environment attention [36]. These structures are effective in improving prediction accuracy but cannot provide quantitative descriptions about interrelated trajectories and interactions among multiple vehicles, resulting in an insufficient model explanation. Reactive prediction methods present several end-to-end planning frameworks that integrate environmental perception, behavior prediction for the target vehicle, and trajectory planning for the ego vehicle in a comprehensive model to characterize their interactions while efficiently reducing prediction uncertainty. These methods include semantic occupancy graphs [38], behavioral cost graphs [39], and energy models [40]. Several studies [41], [42], [43], [44], [45] have tried to incorporate the behaviors of the ego vehicle into the trajectory prediction process. For example, Huang et al. [44] developed a decision-making framework using online learning for an interaction-aware motion prediction model, resulting in high success rate. Furthermore, Huang et al. [45] proposed a differentiable integrated prediction and planning framework to obtain reactive trajectories similar to those of a human driver. These findings imply that decision-making performance is strongly influenced by the ego vehicle's future plans.

Although the combination of learning-based methods and attention mechanisms have been periodically reported in the literature, the potential impact of the interactions between the ego vehicle and its surrounding vehicles on the prediction system has not been sufficiently accounted for. This may lead to ineffective assessment of different driving behaviors and significant prediction errors especially in dense traffic flows. Additionally, the existing methods rarely provide the quantitative descriptions about the long-term information embedded in historical trajectories, the impact of the ego vehicle's future plans, and the interactions among multiple vehicles. These limitations can curtail prediction accuracy and result in poor model interpretability. To address these challenges, this study proposes a hybrid trajectory prediction framework that uses a spatial-temporal attention mechanism and incorporates the planning information of the ego vehicle. The exclusive contributions of this study can be summarized as follows.

1) A convolutional pooling model is proposed to model the interaction between the planned trajectories of the ego vehicle and the historical trajectories of the surrounding vehicles. This can reduce the prediction uncertainty and thus improve trajectory prediction accuracy by integrating the planned information into the prediction process, especially in highly interactive traffic scenarios.
2) An attention mechanism is proposed to identify the contributions of historical trajectories, target vehicles, and surrounding vehicles at both temporal and spatial levels to improve the accuracy and interpretability of trajectory prediction. By emphasizing the key factors

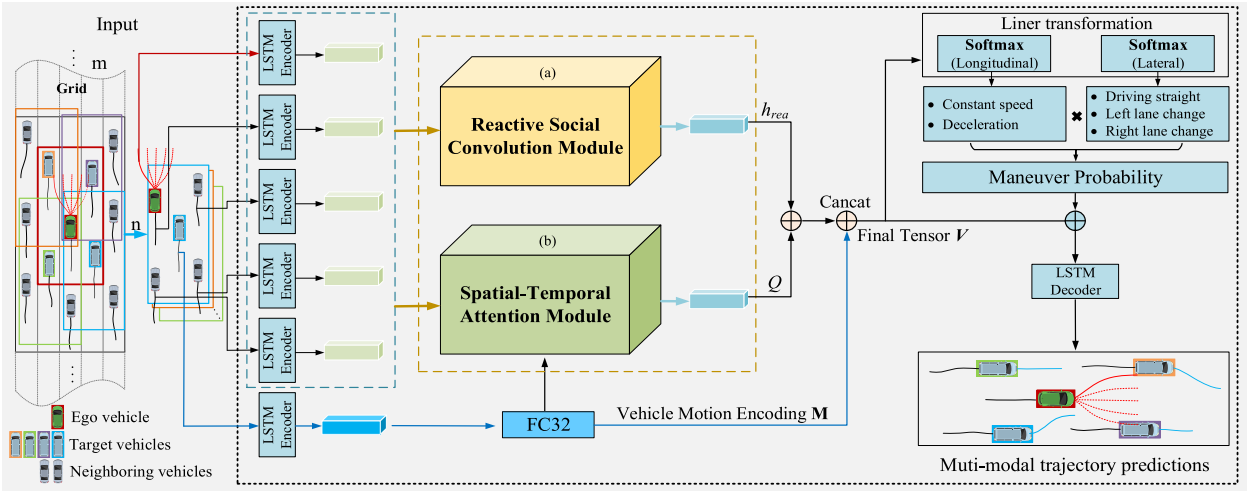

Fig. 2. Architecture of the proposed hybrid trajectory prediction method. A $n \times m$ grid is defined for the social interaction range among all vehicles associated with the ego vehicle, with n denoting the distance in the longitudinal direction of the lane and m denoting the lateral distance perpendicular to the direction of the lane. The target vehicle is set in the area centered on the ego vehicle (red square) during the input stage, followed by using the encoder to encode the vehicle in the target-centered region to obtain the encoding information of each vehicle motion. These hidden states are then transmitted simultaneously to two key modules: (a) reactive social convolution module, which learns the interactions of the ego vehicle with other vehicles and (b) spatial-temporal attention module, which captures spatio-temporal interactions among vehicles. Further, the kinematic feature of the target vehicle is extracted by the FC layer. All the encoding information is integrated into a final tensor V. Finally, the multimodal trajectory distribution is performed through an LSTM-based encoder-decoder framework.

with significant contributions, this approach can lead to more reasonable and accurate trajectory prediction.

3) The proposed hybrid trajectory prediction framework explains the actions of other vehicles in response to the future plans of the ego vehicle from a spatio-temporal perspective, which allows for better understanding of the driver's decision-making process and informs the development of more efficient automated driving systems.

The remainder of this article is structured as follows. Section II introduces the proposed prediction method that includes the preliminaries, the reactive social convolution model, and the attention mechanism. Section III provides the primary experimental data and discusses the prediction results. Lastly, Section IV concludes this article and discusses future research directions.

II. Hybrid Trajectory Prediction With an Attention Mechanism

Most existing methods for predicting the trajectories of surrounding vehicles often assume that the ego vehicle is a static traffic participant. However, AVs are controlled dynamic participants in real traffic scenarios relative to the surrounding vehicles, and their planned trajectories can affect the future states of the surrounding vehicles, leading to significant errors for the trajectory prediction of the surrounding vehicles. To address this issue, a hybrid trajectory prediction framework is proposed to describe how the ego vehicle interacts with the surrounding vehicles during trajectory prediction when it has a motion trend. It is established based on the convolutional pooling model due to its inherent advantages in capturing spatial interdependencies among multiple vehicles. Additionally, a temporal-spatial attention mechanism is designed for interpretable prediction, which quantifies the interrelational features of trajectories and the interactions among multiple vehicles by introducing appropriate attention weights.

The schematic of the proposed prediction framework is shown in Fig. 2. It comprises of the ego vehicle v_{ego}, the target vehicle set $\boldsymbol{V}_{\text{tar}}$ (see red box in Fig. 2), and the neighboring vehicle set $\boldsymbol{V}_{\text{nbrs}}$ (see nonred box in Fig. 2). For the ego vehicle, the vehicle in its grid, whose trajectory we are predicting, is regarded as the target vehicle. For the target vehicle, the vehicles in its grid are considered as the neighboring vehicles (including the ego vehicle) that interact with it.

The vehicle's trajectories are fed into the model via the stack. The encoder first utilizes LSTM to obtain the hidden state of the ego vehicle's motion trend within the target-vehicle-centered region and the hidden states of the historical trajectories of the target vehicle and its neighboring vehicles. These hidden states are then transmitted simultaneously to: 1) the reactive social convolution module (Section II-B) for capturing the interactions among the planned trajectory of the ego vehicle, the predicted trajectory of the target vehicle, and the history trajectories of the neighboring vehicles and 2) the spatio-temporal attention module (Section II-C) for characterizing the spatio-temporal interaction features among the involved vehicles. They calculate the contributions of the ego vehicle's historical trajectories at different moments and the interactions among different vehicles to the current prediction. At the same time, the encoded information of the target vehicle is extracted by a fully-connected (FC) layer to represent its kinematic features. Then the information of each module is integrated into a final tensor $\boldsymbol{V}$, and a maneuver-based decoding is further utilized to generate multimodal trajectory distributions in the future horizon (Section II-D). Specifically, the final tensor is transformed into longitudinal and lateral driving behaviors using two linear transformations. The

driving behaviors are the outputs in probability form using a softmax function and their multiplication results in the maneuver probability. The final tensor is also concatenated with the maneuver probability to predict the trajectory for each maneuver through the decoder. Finally, the details of the model training and implementation are presented in Section II-E.

A. Preliminaries

1) Problem Formulation: The prediction framework assumes that the ego vehicle can accurately detect or measure the trajectories of its surrounding vehicles via sensors and storage them as discrete series of waypoints. Therefore, the objective of trajectory prediction is to forecast the future movements of other vehicles in close proximity, which is referred to as the target vehicle. Let

$$\boldsymbol{X}_i^{\text{tar}} = \left\{\left(x_i^{\text{tar}}, y_i^{\text{tar}}\right) \in \mathbf{R}^2 | i = -T_{\text{obs}}, \ldots, -1, 0\right\} \tag{1}$$

where $\boldsymbol{X}_i^{\text{tar}}$ denotes the historical trajectory of a target vehicle within a certain ego-vehicle-centered area O_{tar} and an observable time domain T_{obs}, including the longitudinal coordinate x_i^{tar} and the lateral coordinate y_i^{tar}. The motions of neighboring vehicles can also significantly influence the trajectory prediction for a specific target vehicle, which can be formulated as

$$\boldsymbol{X}_i^{\text{nbr}} = \left\{\left(x_i^{\text{nbr}}, y_i^{\text{nbr}}\right) \in \mathbf{R}^2 | i = -T_{\text{obs}}, \ldots, -1, 0\right\} \tag{2}$$

where $\boldsymbol{X}_i^{\text{nbr}}$ denotes the historical trajectories of the neighboring vehicles within a target-vehicle-centered region O_{nbrs}. As there are frequent interactions between the target and the ego vehicle, the influence of the ego vehicle's motion trend on other vehicles cannot be neglected [13]. Hence, the motion trend of the ego vehicle is used as an informed feature, which can be given by

$$\boldsymbol{X}_i^{\text{ego}} = \left\{\left(x_i^{\text{ego}}, y_i^{\text{ego}}\right) \in \mathbf{R}^2 \middle| i = 1, 2, \ldots, T_{\text{pred}}\right\} \tag{3}$$

where $\boldsymbol{X}_i^{\text{ego}}$ represents the vehicle's trajectory composed of longitudinal and lateral coordinates, which can also be regarded as the planned trajectory of the ego vehicle in the prediction horizon T_{pred}. The task of the prediction framework is to calculate the future trajectory for the target vehicle in the prediction horizon. Thus, the corresponding coordinates of the predicted future trajectory are given by

$$\hat{\boldsymbol{Y}}_i^{\text{tar}} = \left\{\left(x_i^{\text{tar}}, y_i^{\text{tar}}\right) \in \mathbf{R}^2 \middle| i = 1, 2, \ldots, T_{\text{pred}}\right\} \tag{4}$$

where $\hat{\boldsymbol{Y}}_i^{\text{tar}}$ denotes the future trajectory of the target vehicle in the prediction horizon.

Overall, the prediction framework is to simultaneously obtain the posterior distribution $P(\boldsymbol{Y}|\boldsymbol{X}_i^{\text{nbr}}, \boldsymbol{X}_i^{\text{tar}}, \boldsymbol{X}_i^{\text{ego}})$ of the future trajectories $\hat{\boldsymbol{Y}}_i^{\text{tar}}$ of multiple target vehicles, in which the historical trajectory information of target vehicles and neighboring vehicles and the planning information of the ego vehicle are considered in the prediction process.

2) LSTM-Based Encoder-Decoder Network: LSTM can effectively alleviate the long-term reliance on continuous sequences. It incorporates three mutually coordinated gate structures to propagate cell state, and thus avoids gradient vanishing or explosion when the information is back-propagated in a sequence. Specifically, the forget gate $\boldsymbol{f}_t$ retains the valuable state information in time series; the input gate $\boldsymbol{i}_t$ identifies the necessary information in the network propagation and constructs a candidate vector for the subsequent state; the output gate $\boldsymbol{o}_t$ outputs the relevant information valuable to the current state. An LSTM cell structure is shown in Fig. 3. Their formulations are given by

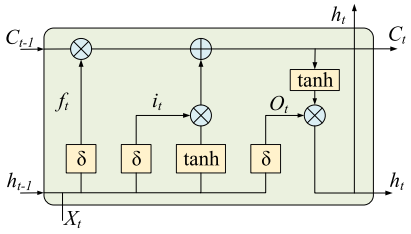

Fig. 3. Schematic and calculation process of LSTM.

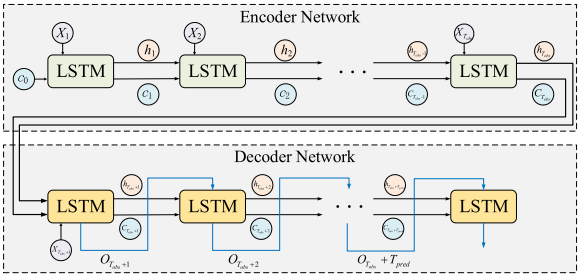

Fig. 4. LSTM-based encoder-decoder network architecture.

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_{Xf}\boldsymbol{X}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f) \tag{5}$$

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_{Xi}\boldsymbol{X}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i) \tag{6}$$

$$\tilde{\boldsymbol{c}}_t = \tanh(\boldsymbol{W}_{Xc}\boldsymbol{X}_t + \boldsymbol{W}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c) \tag{7}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \cdot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \cdot \tilde{\boldsymbol{c}}_t \tag{8}$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{Xo}\boldsymbol{X}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o) \tag{9}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \cdot \tanh(\boldsymbol{c}_t) \tag{10}$$

where σ is the sigmoid activation function; $\boldsymbol{f}_t$, $\boldsymbol{i}_t$, and $\boldsymbol{o}_t$ are the forget, input, and output gates, respectively; $\boldsymbol{X}_t$ and $\boldsymbol{h}_{t-1}$ are the current input and previous hidden states, respectively; and $\boldsymbol{c}_t$ is the cell state that enables the information to be forgotten or added in. The final output $\boldsymbol{h}_t$ is obtained by multiplying the elements of $\tanh(\boldsymbol{c}_t)$. The parameters to be learned are the weights and bias represented by $[\boldsymbol{W}_*, \boldsymbol{b}_*]$. LSTM has been widely adopted to solve practical problems via an encoder-decoder architecture due to the ability to retain temporal information. The encoder-decoder design extract intrinsic representational information from input sequences and maps them to unequal-length output sequences, making it easy to implement vehicle trajectory prediction on highways. As shown in Fig. 4, the encoder sends time sequence information to LSTM, which produces the corresponding tensor and calculates the encoded information $\boldsymbol{c}_t$ and $\boldsymbol{h}_t$. The decoder generates the predicted result $O_{T_{\text{obs}}+1}$ corresponding to the current input $\boldsymbol{X}_{\text{obs}}$ based on the configured node parameters, with the prediction process being recursive.

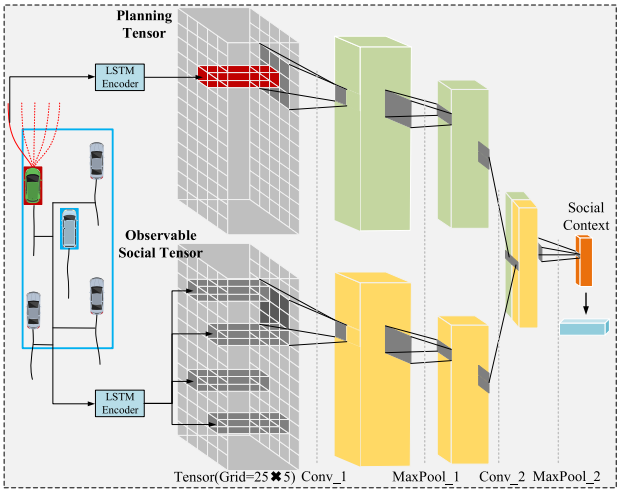

Fig. 5. Schematic of the proposed reactive social convolution module. The future trajectories of ego vehicles in the scope of social interaction are encoded as the planning tensor with the historical trajectories of other vehicles as social tensors. Both of them are placed in the corresponding grid cell to retain the spatial relationships of all vehicles. Then the convolutional pooling structure is further employed for fusion.

B. Reactive Social Convolution Module

For the purpose of tackling the interactions among multiple vehicles, a reactive social convolution mechanism is developed in this study. Specifically, the hidden states of all the involved vehicles, which mainly include the embedding information of $\boldsymbol{V}_{\text{tar}}$, the social interactions between $\boldsymbol{V}_{\text{tar}}$ and $\boldsymbol{V}_{\text{nbrs}}$, and the potential spatial occupation relationship between the planned trajectory of v_{ego} and the predicted trajectory of $\boldsymbol{V}_{\text{tar}}$, are comprehensively considered here to predict the waypoint sequence of the target vehicle. The overall architecture is depicted in Fig. 5, and the details are given as follows.

First, the global coordinates of the planned and historical trajectories of v_{ego} and $\boldsymbol{V}_{\text{nbrs}}$ are preprocessed as relative coordinates with respect to the target vehicle for calculation simplification. To track the trajectory characteristics of the input data as accurately as possible, these trajectories are embedded by a nonlinear transformation to obtain independent motion embeddings. Then the embedded trajectory information is further encoded in the LSTM network to obtain the hidden state $\boldsymbol{h}(\cdot)$ that can be used as the motion encoding to fulfill the prediction network requirements. It is worth noting that the planned trajectories $\boldsymbol{X}^{\text{ego}}$ are processed in a reverse order to cooperate with the historical trajectory sequence. Therefore, independent LSTM models with different parameters are utilized to encode $\boldsymbol{X}^{\text{ego}}$, $\boldsymbol{X}^{\text{tar}}$, and $\boldsymbol{X}^{\text{nbr}}$ due to their different time-domain characteristics. This process can be formulated as

$$\begin{cases} \boldsymbol{h}_t^{\text{tar}} = \text{LSTM}_{\text{enc}}\left(\text{Emb}\left(\boldsymbol{X}_t^{\text{tar}}\right)\right) \\ \boldsymbol{h}_t^{\text{ego}} = \text{LSTM}_{\text{enc}}\left(\text{Emb}\left(\boldsymbol{X}_t^{\text{ego}}\right)\right) \\ \boldsymbol{h}_t^{\text{nbr}} = \text{LSTM}_{\text{enc}}\left(\text{Emb}\left(\boldsymbol{X}_t^{\text{nbr}}\right)\right) \end{cases} \tag{11}$$

where Emb() is the FC layer that embeds x and y coordinates into higher dimensions and $\text{LSTM}_{\text{enc}}()$ is the LSTM encoder employed in this model; $\boldsymbol{h}_t^{\text{tar}}$, $\boldsymbol{h}_t^{\text{nbr}}$, and $\boldsymbol{h}_t^{\text{ego}}$ denote the hidden states corresponding to the target vehicle, neighboring vehicle, and ego vehicle, respectively.

Second, the spatial interaction relationship is modeled between the planned and historical trajectories. As mentioned above, LSTM encoders are capable of capturing the temporal structure of sequential trajectories; but they are insufficient in characterizing the spatial interactions among participants in a defined traffic scenario. Motivated by Deo and Trivedi [23] and Hasan et al. [47], this study develops a hierarchical convolutional social pooling structure to model spatial interactions among multiple vehicles. Since the planned and historical trajectories belong to different time domains, this method constructs two target-centered spatial grids and then places the encoding vectors $\boldsymbol{h}_t^{\text{nbr}}$ of the historical trajectories in the corresponding grids of the lower branches, while the encoding vectors $\boldsymbol{h}_t^{\text{ego}}$ of the planned trajectories are placed in the grids of the upper branches (see Fig. 5), which are also labeled as the planning and the observable tensor, respectively. Then the kinematic information $\boldsymbol{h}_t^{\text{tar}}$ of the target vehicle is embedded by the FC layer network to obtain the target tensor M, which is given by

$$\boldsymbol{M}_t = \text{FC}\left(\boldsymbol{h}_t^{\text{tar}}\right). \tag{12}$$

Finally, the planning and the observable historical tensor are simultaneously combined to fully account for the motion trend of the ego vehicle in the prediction process. The convolution and pooling layers featuring retained local spatial structure are employed to handle the planned and observable tensors in parallel to obtain the social tensor, which can not only model multivehicle interactions in terms of spatial properties, but also concatenate the planned and observable tensors through a max-pooling layer, which is given by

$$\boldsymbol{h}_t^{\text{rea}} = \text{MaxP}\left(\text{Soc}\left(\boldsymbol{h}_t^{\text{nbr}}, \boldsymbol{h}_t^{\text{ego}}\right)\right) \tag{13}$$

where Soc() is a two-layer convolutional neural network; and the corresponding two-layer maximum pooling layer is labeled as MaxP().

Through the above operation, the merged social context vector $\boldsymbol{h}^{\text{rea}}$ can be combined with the motion encoding $\boldsymbol{M}$ to realize a rough trajectory prediction. However, it cannot intuitively explain the interaction differences between vehicles.

C. Spatial–Temporal Attention Mechanism Modeling

In practical driving conditions, a competent human driver can focus on essential information related to vehicle driving safety and perform proper actions in response to environmental changes. The attention mechanism is designed to imitate this feature by combining it with an encoder-decoder to quickly focus on the key information, which enables high-value features to be extracted from the dynamic interactions among multiple vehicles. That is to say, the prediction model can automatically focus on key information impacting trajectory changes for reasonable and accurate prediction [37]. As mentioned above, the LSTM-based trajectory prediction method can obtain the corresponding encoded information from the input trajectories of the involved vehicles. This allows the

attention module to evaluate the degree of interaction among multiple vehicles by correlating different encoded information for interpretable prediction. However, the decoupled attention model cannot sufficiently capture the spatio-temporal correlated information due to simplifying the embedding part of the process.

To achieve an explainable prediction for driving behaviors, this study incorporates a spatio-temporal-coupled attention mechanism into the prediction model. Specifically, for every vehicle $v_{i,\text{tar}}$ in $\boldsymbol{V}_{\text{tar}}$, the influence of historical and planned trajectories of its $\boldsymbol{V}_{\text{nbrs}}$ is adjusted by the spatio-temporal learned attention weights. In this work, the attention weights at the temporal level are utilized to analyze the impact of historical trajectories of $\boldsymbol{V}_{\text{tar}}$ on prediction. Meanwhile, the attention weights at the spatial level are capable of explaining the impact of the encoding information of $\boldsymbol{V}_{\text{tar}}$ and $\boldsymbol{V}_{\text{nbrs}}$ (including the planning information of v_{ego}) on prediction. Notably, the spatial grid size utilized here aligns with the reactive social convolution module to ensure that the attention weights can be provided at the correct spatial grid locations. The details are presented as follows.

1) Temporal-Level Attention: At time step t, the inputs of the LSTM encoder are the historical trajectories $\boldsymbol{X}_t^{\text{nbr}}$ and $\boldsymbol{X}_t^{\text{tar}}$ of the neighboring and target vehicles within T_{obs} steps, and the hidden state is generated by

$$\boldsymbol{E}_t^{v_i} = \left\{ \boldsymbol{h}_{t-T_{\text{obs}}+1}^{v_i}, \ldots, \boldsymbol{h}_j^{v_i}, \ldots, \boldsymbol{h}_t^{v_i} \right\} \tag{14}$$

where $\boldsymbol{E}_t^{v_i} \in \mathbf{R}^{d \times T_{\text{obs}}}$ denotes the hidden state of each vehicle v_i; and d is the length of the hidden state. Due to high nonlinearity, it is hard to directly obtain the correlation of trajectories through a specific measurement technique. Therefore, a linear transformation is employed to learn the interrelation of the trajectory, and then the *softmax* function is introduced to maximize the temporal attention of the high-contribution historical moments. The temporal attention weights can be expressed as

$$\boldsymbol{K}_t^{v_i} = \text{softmax}\left(\tanh\left(\boldsymbol{W}_\alpha \boldsymbol{E}_t^{v_i}\right)\right) \tag{15}$$

where $\boldsymbol{K}_t^{v_i} = \{\alpha_{t-T_{\text{obs}}+1}^{v_i}, \ldots, \alpha_j^{v_i}, \ldots, \alpha_t^{v_i}\}$ represents the temporal-level attention weights associated with each vehicle v_i in the historical horizon; and $\boldsymbol{W}_\alpha$ represents the learnable weight.

The context vectors containing temporal information can be obtained by multiplying the original hidden states and the corresponding weight coefficients, which can be indicated by

$$\boldsymbol{H}_t^{v_i} = \sum_{k=t-T_{\text{obs}}+1}^{t} \alpha_k^{v_i} \boldsymbol{h}_k^{v_i}. \tag{16}$$

Notably, the temporal weights are calculated from the corresponding hidden states, and thus the correlation of the trajectory can be explicitly obtained through the temporal context vector. As the temporal attention weights $\boldsymbol{K}_t^{v_i}$ can significantly affect the spatial attention weights [34], the value of the grid cell associated with v_i can also be represented by $\boldsymbol{H}_t^{v_i}$ to calculate the spatial-level attention weights and be further engaged in the trajectory prediction of the target vehicle.

2) Spatial-Level Attention: The presented framework is developed based on the target-vehicle-centered grid. Therefore, the values of all grid cells at time step t can be represented by

$$\boldsymbol{L}_t = \left\{ \boldsymbol{l}_t^1, \boldsymbol{l}_t^2, \ldots, \boldsymbol{l}_t^N \right\} \tag{17}$$

where N represents the total number of grid cells. It should be noted that when there is a vehicle in the corresponding grid cell, $\boldsymbol{l}_t^n$ needs to be placed on the corresponding grid; otherwise, 0 is assigned. Consequently, it can be calculated by

$$\boldsymbol{l}_t^n = \begin{cases} \boldsymbol{H}_t^{v_i}, & \text{if the vehicle } v_i \text{ locates at grid cell } n \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

For the grid cell of v_{ego}, a max-pooling layer is employed to merge all the information since it is capable of featuring both historical and planned trajectories. In this way, the planning information of the ego vehicle can be incorporated at this stage. To ensure that the weights scales of each neighboring vehicle and the target vehicle are the same, the spatial attention weight $\boldsymbol{S}_t$ is calculated by

$$\boldsymbol{S}_t = \text{softmax}(\tanh(\boldsymbol{W}_\beta \boldsymbol{L}_t)) \tag{19}$$

where $\boldsymbol{W}_\beta$ represents the learnable weight; and $\boldsymbol{S}_t = \{\varepsilon_t^1, \varepsilon_t^2, \ldots, \varepsilon_t^N\}$ represents the weights of the grid cells within the grid scope. Through the above definition, all the information from the target and neighboring vehicles are combined as a fused tensor, which is given by

$$\boldsymbol{Q}_t = \sum_{n=1}^{N} \varepsilon_t^n \boldsymbol{l}_t^n. \tag{20}$$

The fused tensor $\boldsymbol{Q}_t$ integrates the trajectory-related hidden states, which allows it to be utilized for implementing predictions through the decoder. Consequently, the fused spatial-temporal tensor $\boldsymbol{Q}_t$, the merged social context vector $\boldsymbol{h}^{\text{rea}}$ and the motion encoding $\boldsymbol{M}$ are fused as

$$\boldsymbol{V}_t = \boldsymbol{w}_Q \boldsymbol{Q}_t + \boldsymbol{w}_h \boldsymbol{h}^{\text{rea}} + \boldsymbol{w}_m \boldsymbol{M} \tag{21}$$

and it is fed into the prediction framework to obtain the predicted trajectory with the spatio-temporal characteristics. The whole process is illustrated in Fig. 6.

D. Multimodal Future Trajectory Prediction

A maneuver-based decoder is developed to describe the uncertainty and multimodal properties of different driving behaviors, which characterizes the distribution of six predefined maneuver classes $\boldsymbol{M} = \{m_k | k = 1, 2, \ldots, 6\}$. The maneuver classes consist of three lateral behaviors (driving straight, left lane change, and right lane change) and two longitudinal behaviors (constant speed and deceleration). The encoding information $\boldsymbol{V}_t$ of the target vehicle is mapped with the probabilities of the lateral and longitudinal maneuvers via two *softmax* functions, which are further multiplied to obtain the probability $P(m_k|\boldsymbol{X})$ of each complete maneuver. Furthermore, the involved hidden states of the target vehicle are concatenated with the one-hot vectors corresponding to the

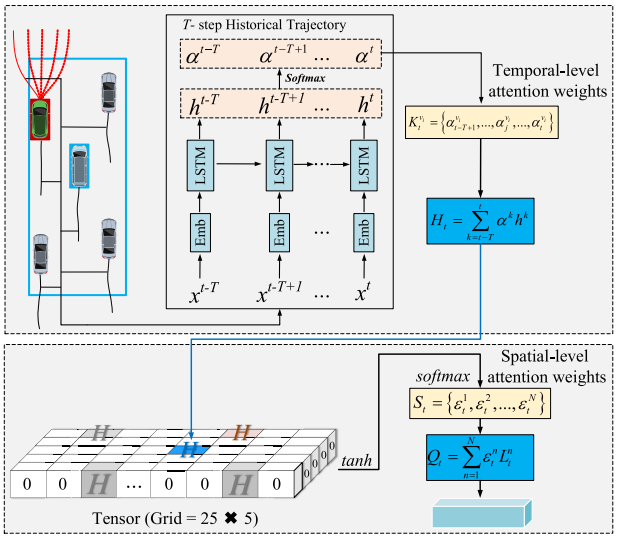

Fig. 6. Schematic of the proposed spatio-temporal attention module.

driving behaviors, and then the new combined hidden states are passed through an LSTM decoder to obtain the predicted trajectories for each maneuver class. To best describe the motion uncertainties of the target vehicles in the prediction horizon, the bivariate Gaussian distribution is employed to output the probability distribution of the predicted trajectories instead of the absolute accurate positions, which can be defined as

$$\hat{\boldsymbol{Y}}_t^{\text{tar}} = \left(\hat{x}_t^{\text{tar}}, \hat{y}_t^{\text{tar}}\right) \sim N\left(\boldsymbol{\mu}_t^{\text{tar}}, \boldsymbol{\sigma}_t^{\text{tar}}, \boldsymbol{\rho}_t^{\text{tar}}\right) \tag{22}$$

where $\hat{\boldsymbol{Y}}_t^{\text{tar}}$ is the predicted trajectory coordinates of the target vehicle; $\boldsymbol{\mu}_t^{\text{tar}} \in \mathbf{R}^2$, $\boldsymbol{\sigma}_t^{\text{tar}} \in \mathbf{R}^2$, and $\boldsymbol{\rho}_t^{\text{tar}} \in \mathbf{R}$ represent the mean value, the standard deviation, and the correlation coefficient of the Gaussian distribution at timestamp $t \in (1, 2, \ldots, T_{\text{pred}})$, respectively. Consequently, the posterior probability for all target vehicles' future trajectories can be deduced as

$$\begin{aligned} P\left(\hat{\boldsymbol{Y}}_t^{\text{tar}} \mid \boldsymbol{X}\right) &= P\left(\left(\hat{x}_t^{\text{tar}}, \hat{y}_t^{\text{tar}}\right) \mid \boldsymbol{X}\right) \\ &= \sum_{k=1}^{|M|} P_{\boldsymbol{\Theta}}\left(\left(\hat{x}_t^{\text{tar}}, \hat{y}_t^{\text{tar}}\right) | m_k, \boldsymbol{X}\right) P(m_k \mid \boldsymbol{X}) \end{aligned} \tag{23}$$

where $\boldsymbol{\Theta} = [\boldsymbol{\Theta}^{t+1}, \boldsymbol{\Theta}^{t+2}, \ldots, \boldsymbol{\Theta}^{t+T_{\text{pred}}}]$ denotes the Gaussian parameters over all future time steps for the target vehicles, corresponding to the means and variances of future locations and velocities.

E. Training and Implementation Details

Following the strategy to compute loss in [23], the proposed algorithm is trained by minimizing the negative log-likelihood (NLL) between the predicted and the ground-truth trajectories for all target vehicles, which can be given by

$$-\log(P_{\boldsymbol{\Theta}}(\hat{\boldsymbol{Y}}^{\text{tar}} | m_{\text{real}}, \boldsymbol{X}) P(m_{\text{real}} | \boldsymbol{X})) \tag{24}$$

where m_{real} denotes the true class of driving maneuvers with $\boldsymbol{X} = \{\boldsymbol{X}^{\text{nbr}}, \boldsymbol{X}^{\text{tar}}, \boldsymbol{X}^{\text{ego}}\}$. The complete process is summarized as Algorithm 1.

Algorithm 1 Training Algorithm

Input:

$n_{\text{e}} \leftarrow$ epoch_number; $n_{\text{b}} \leftarrow$ batch_size;
$\boldsymbol{X}^{\text{tar}}, \boldsymbol{X}^{\text{nbr}} \leftarrow$ historical trajectories of the target vehicle and neighboring vehicles within the grid scope;
$\boldsymbol{X}^{\text{ego}} \leftarrow$ the future trajectories of the ego vehicle in the prediction horizon; $\boldsymbol{Y}^{\text{tar}} \leftarrow$ ground truth of the target vehicle's future trajectory;
for $i = 1$, $i \leq n_{\text{e}}$ **do**
  for $j = 1$, $j \leq n_{\text{b}}$ **do**
    Input $\boldsymbol{X}^{\text{nbr}}$, $\boldsymbol{X}^{\text{tar}}$, $\boldsymbol{X}^{\text{ego}}$ to the LSTM encoder for acquiring historical trajectory encodings $\boldsymbol{h}(\boldsymbol{X}^{\text{nbr}})$, $\boldsymbol{h}(\boldsymbol{X}^{\text{tar}})$ and motion feature $\boldsymbol{h}(\boldsymbol{X}^{\text{ego}})$ of the planned trajectory of the ego vehicle via Eq.(11);
    Input $[\boldsymbol{h}(\boldsymbol{X}^{\text{nbr}}), \boldsymbol{h}(\boldsymbol{X}^{\text{ego}})]$ to the reactive social convolution module to acquire $\boldsymbol{h}^{\text{rea}}$ according to the Eq.(13);
    Input $[\boldsymbol{h}(\boldsymbol{X}^{\text{nbr}}), \boldsymbol{h}(\boldsymbol{X}^{\text{tar}}), \boldsymbol{h}(\boldsymbol{X}^{\text{ego}})]$ to the attention module to acquire tensor $\boldsymbol{Q}_t$ according to the Eq.(20);
    Get the dynamics encoding $\boldsymbol{M}$ of the target vehicle by the FC layer through the Eq.(12);
    Get the final fused tensor $\boldsymbol{V}_t$ by performing concatenation operation $\boldsymbol{h}^{\text{rea}}$, $\boldsymbol{Q}_t$ and $\boldsymbol{M}$ via Eq.(21), and feed it into the maneuver-based decoder;
    Output the maneuver probabilities $P(m_k|\boldsymbol{X})$ and the corresponding future trajectory $\hat{\boldsymbol{Y}}^{\text{tar}}$, and calculate temporal weights $\boldsymbol{K}_t$ and spatial attention weights $\boldsymbol{S}_t$ according to Eq.(23);
  Update the NLL of the future trajectories by descending its gradient for all the target vehicles: $\nabla_{\boldsymbol{\Theta}}[-\log(P_{\boldsymbol{\Theta}}(\hat{\boldsymbol{Y}}^{\text{tar}}|m_{\text{real}}, \boldsymbol{X}) P(m_{\text{real}}|\boldsymbol{X}))]$

For data configuration, each data in the public dataset specifies a vehicle as v_{ego}, and thus the vehicles within the ego-vehicle-centered areas O_{tar} are defined as the target vehicles. Similarly, the target-vehicle-centered area O_{nbrs} of each target vehicle gets the same definition as O_{tar}. For the grid size, a larger grid size can provide more effective information, but it may also compromise computational efficiency [15]. Following similar settings proposed by previous studies [23], [42], the spatial grid in this study is discretized as a 25 × 5 grid with an actual size of 200 × 35 ft. In practical application of AVs, trajectory prediction can only be carried out based on current planning cycle. Hence, a limited number of waypoints from the actual future trajectory of the ego vehicle can be applied to represent the planning input $\boldsymbol{X}^{\text{ego}}$. In the training process, the planning input $\boldsymbol{X}^{\text{ego}}$ is derived from its downsampled actual trajectories for better understanding of the ego vehicle's motion trends [13]. Furthermore, a fitted quintic polynomial profile based on actual future trajectory is implemented for evaluation, which contributes to its deployment in intelligent transportation systems. Notably,

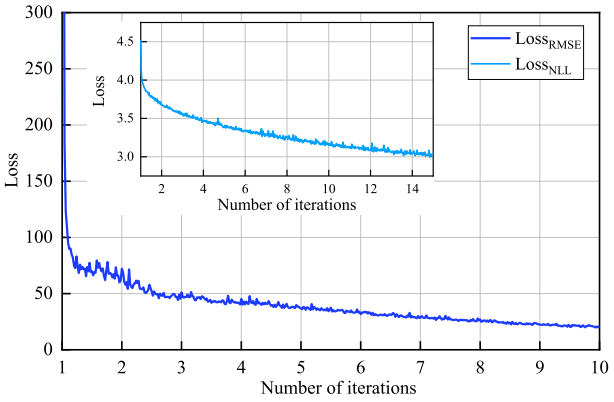

Fig. 7. Convergence process of the loss function during training. The two curves indicate the change in the mean square and NLL errors, respectively.

TABLE I
PARAMETERS AND SIZE IN MODEL TRAINING

Description	Symbol	Function & Parameter
Adam	-	0.001
Embedded trajectory	$Emb()$	Linear(2,32)
Hidden state	h_t	LSTM(32,64,1)
Kinematics of target vehicle	M_t	Linear(64,32)
Convolutional layer_1	-	Conv2d(64,64,3)
Pooling layer_1	-	MaxPool2d((3,3),2)
Convolutional layer_2	-	Conv2d(64,32,(3,1))
Pooling layer_2	-	MaxPool2d((2,2),(1,0))
Temporal Attention weight	k_t	Softmax(15,1)
Spatial Attention weight	S_t	Softmax(125,1)
Lateral & longitudinal behavior	m_k	Linear(64,3)
Decoder	-	LSTM(70,128,1)

rel-pos is adopted as input in the training process. It is a relative position coordinate referencing the last frame in the historical trajectory. *rel-pos* has smaller values than *abs-pos* (absolute position coordinates), making it numerically stable and enabling the model to converge to a smaller range. It also implies the vehicle's velocity information, and helps the model learn hidden features related to velocity [42].

Adam optimizer [48] is used to train the model in an end-to-end fashion. The batch size is set as 128. Furthermore, the prediction model is built in the Python environment with a PyTorch backend and with an Intel Core i9-12900KS CPU and an NVIDIA GeForce GTX 3090 GPU. The convergence process of the loss function in the model training process is shown in Fig. 7, in which the overall downward trend of the two loss functions demonstrates the convergence of the proposed model. Some key parameters are listed in Table I.

III. EXPERIMENTAL EVALUATIONS

The effectiveness of the proposed scheme is evaluated based on the public driving datasets described in Section III-A. The quantitative comparison results between the proposed scheme and the state-of-the-art methods (Section III-B) are presented in Section III-C. Different driving maneuvers of target vehicles in diverse traffic scenarios are predicted, and comprehensive driving scenarios are simulated and evaluated to demonstrate the efficacy of the proposed method. Moreover, a conventional model-based approach [49] is utilized to generate diverse planned trajectories in the prediction horizon, with more results provided in Section III-D for further analysis.

A. Datasets and Metrics

Two public benchmark datasets that record vehicle trajectories in real-world scenarios are employed to examine the accuracy and generalization ability of the proposed method.

1) NGSIM: The NGSIM dataset [50] includes the US-101 and I-80 freeway traffic data recorded by multiple overhead cameras in 2005 and has been widely used in trajectory prediction studies. It is worth noted that each dataset contains abundant interaction scenarios including mild, moderate, and congested traffic with a sampling frequency of 10 Hz.

2) HighD: The HighD dataset [51] includes the real-world vehicle data recorded by camera-equipped unmanned aerial vehicles at a frequency of 25 Hz on German freeways in 2017 and 2018. It includes more than 110 000 vehicles with a total traveling distance of about 45 000 km.

The two datasets were divided into 70% training, 10% validation, and 20% testing subsets, respectively [23]. The trajectories of each vehicle were composed of 8-s segments, each of which consisted in 3 s of historical and 5 s of predicted trajectories. Moreover, each segment was downsampled to obtain 5 frames/s for complexity reduction.

3) Evaluation Metrics: To quantitatively analyze the accuracy of the prediction model, two common error metrics are introduced [24].

a) RMSE: The RMSE error between the predicted trajectories and the ground truth is used to evaluate the prediction accuracy, which is given by

$$\text{RMSE} = \sqrt{\frac{1}{T_{\text{pred}}}\sum_{t=1}^{T_{\text{pred}}}\left|\hat{\boldsymbol{Y}}_t^{\text{tar}} - \boldsymbol{Y}_t^{\text{tar}}\right|^2} \tag{25}$$

where $\hat{\boldsymbol{Y}}_t^{\text{tar}}$ and $\boldsymbol{Y}_t^{\text{tar}}$ are the predicted and ground-truth location at time step t within a prediction horizon of 5 s.

b) NLL error: The NLL error is adopted to measure the similarity between the two probability distributions [35].

B. Baseline Models

The proposed model is compared with the state-of-the-art methods that mainly include the kinetic-based prediction models and the prediction models considering multiagent interactions.

1) *CV:* A representative constant velocity Kalman filter [7] is employed for trajectory prediction.
2) *S-LSTM:* The social LSTM [22] utilizes a social model that considers interactive agents to output the uni-modal distribution of future locations.
3) *CS-LSTM:* The convolutional social LSTM [23] constructs a vehicle-centric grid with a convolutional pooling model to output multimodal prediction results.
4) *MATF:* The multiagent tensor fusion GAN [30] handles the social interactions among multiple agents and scene context constraints through a spatial feature map.

TABLE II
QUANTITATIVE RESULTS OF THE PROPOSED AND BASELINE METHODS ON THE NGSIM AND HIGHD DATASETS. ALL THE RESULTS ARE REPORTED IN RMSE OVER A 5-S PREDICTION HORIZON IN METERS. ALL THE MODELS TAKE AS AN INPUT 3 S. NOTE THAT THE BEST RESULTS ARE MARKED BY BOLD NUMBERS

Metric	Dataset	Prediction Horizon (s)	CV	S-LSTM	CS-LSTM	MATF	NLS-LSTM	MHA-LSTM	PiP-LSTM	TS-GAN	Proposed Method
RMSE(m)	NGSIM	1s	0.73	0.65	0.61	0.66	0.56	0.56	0.55	0.60	**0.49**
		2s	1.78	1.31	1.27	1.34	1.22	1.22	1.18	1.24	**1.09**
		3s	3.13	2.16	2.09	2.08	2.02	2.01	1.94	1.95	**1.78**
		4s	4.78	3.25	3.10	2.97	3.03	3.00	2.88	2.78	**2.62**
		5s	6.68	4.55	4.37	4.13	4.30	4.25	4.04	3.72	**3.65**
	HighD	1s	-	0.22	0.22	-	0.20	0.19	0.17	-	**0.16**
		2s	-	0.62	0.61	-	0.57	0.55	0.52	-	**0.47**
		3s	-	1.27	1.24	-	1.14	1.10	1.05	-	**0.94**
		4s	-	2.15	2.10	-	1.90	1.84	1.76	-	**1.58**
		5s	-	3.41	3.27	-	2.91	2.78	2.63	-	**2.36**

Abbreviations: -: No results or not given.

TABLE III
QUANTITATIVE RESULTS ON THE NGSIM AND HIGHD DATASETS USING NLL METRICS

Metric	Dataset	Prediction Horizon (s)	CV	S-LSTM	CS-LSTM	PiP-LSTM	Proposed Method
NLL	NGSIM	1s	3.72	2.28	1.91	1.72	**1.20**
		2s	5.37	3.86	3.44	3.30	**2.94**
		3s	6.40	4.69	4.31	4.17	**3.89**
		4s	7.16	5.33	4.94	4.80	**4.54**
		5s	7.76	5.89	5.48	5.32	**5.07**
	HighD	1s	-	0.42	0.37	0.20	**0.13**
		2s	-	2.58	2.43	2.28	**2.05**
		3s	-	3.93	3.65	3.53	**3.23**
		4s	-	4.87	4.51	4.39	**4.04**
		5s	-	5.57	5.17	5.05	**4.69**

5) *MHA-LSTM:* The multiagent attention LSTM [31] combines the partial and global attention paid to the surrounding vehicles to generate a multimodal solution.
6) *NLS-LSTM:* The nonlocal social pooling LSTM [46] employs a nonlocal multiheaded attention mechanism to model the interactions among vehicles.
7) *PiP-LSTM:* The planning-informed prediction LSTM [42] treats the planning information as an informed condition to generate maneuver-based multimodal trajectories.
8) *TS-GAN:* The spatio-temporal GAN [15] introduces a multivehicle collaborative learning framework with a spatio-temporal tensor fusion mechanism for vehicle trajectory prediction.

C. Quantitative Evaluation

The proposed approach is trained and evaluated based on the NGSIM and HighD datasets, respectively. The evaluation results are listed and compared to the baseline methods in Table II. For the models with multimodal prediction distributions, RMSE is calculated from the predicted trajectories with the highest probability maneuver $P(m_k)$. The results are presented in Table II. Observing the results, the proposed scheme exhibits higher accuracy at different time steps on the HighD and NGSIM datasets. In comparison, the model-based *CV* method exhibits the largest error. This is because it utilizes the current speed of the model to predict future trajectories, which makes it difficult to obtain reliable results in long-term predictions. *S-LSTM* focuses more on pedestrian trajectory prediction and thus cannot sufficiently address vehicle trajectory prediction problem. *CS-LSTM* improves *S-LSTM* by using convolutional pooling layers to achieve higher prediction accuracy and to enable multimodal prediction. As a stochastic model, *MATF* exhibits high accuracy due to its generator and discriminator mechanisms. *NLS-LSTM* and *MHA-LSTM* introduce different attention mechanisms to achieve high prediction accuracy by capturing spatial interactions between vehicles. *PiP-LSTM* achieves competent prediction accuracy by considering the motion trend of the ego vehicle. *TS-GAN* uses the social convolution and spatio-temporal mechanisms to model the interactions among multiple vehicles. In contrast, the proposed method not only considers the future planning information of the ego vehicle, but also introduces the attention mechanisms to capture the interrelated features among trajectories. It exhibits excellent potential in improving prediction accuracy and model explainability. Moreover, the prediction error can be effectually limited within a certain range even with elongated prediction horizon. In addition, RMSE may fail to reflect accuracy for distinct maneuvers due to its limitations in evaluating multimodal prediction. Thereby, this study utilizes the NLL of the true trajectories under the prediction results generated by either uni-modal or multimodal distributions, which includes the most related baseline methods reported by NLL, i.e., *S-LSTM*, *CS-LSTM*, and *PiP-LSTM*, as shown in Table III.

It can be seen in Tables II and III that the RMSE values obtained based on the NGSIM dataset are larger than those based on the HighD dataset due to their different numbers of involved vehicles. To be specific, the larger number of vehicles and the smaller data noise of the HighD dataset contribute to improving the performance of the proposed scheme. Furthermore, the quantitative results under different grid sizes are represented in Table IV. It can be seen that the grid size of 25×5 demonstrates the best prediction accuracy.

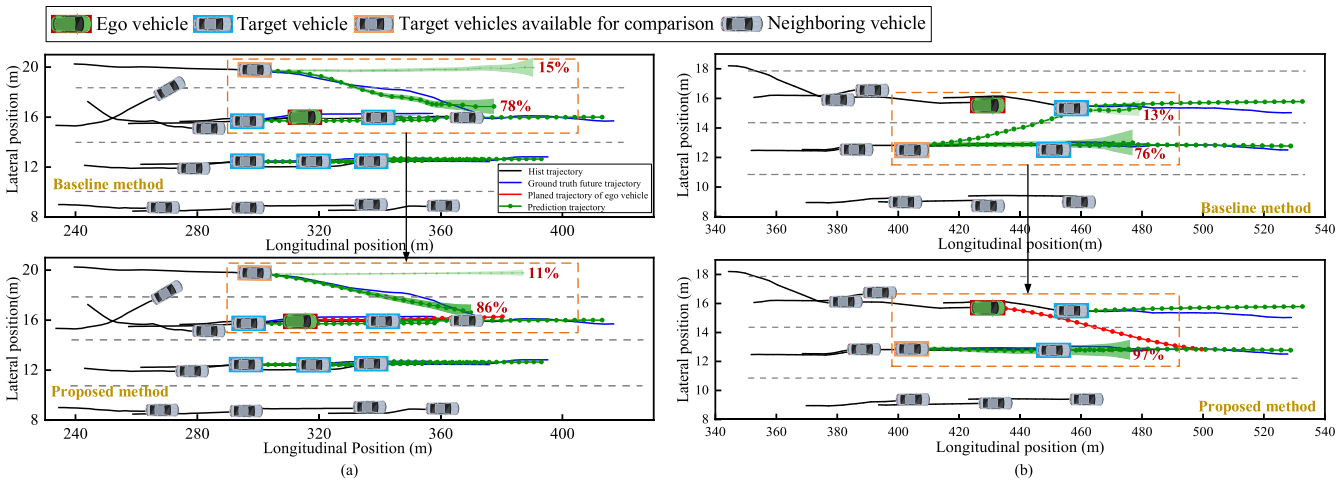

Fig. 8. Visualized results on NGSIM predicted by the baseline method [23] and the proposed method. The green shaded area represents the variance of the distribution, and the trajectory color is proportional to the maneuver probability of the corresponding trajectory. (a) Predicted results for the ego vehicle scheduled to go straight. (b) Predicted results for the ego vehicle scheduled to turn right.

TABLE IV
QUANTITATIVE EXPERIMENTAL ANALYSIS ABOUT DIFFERENT GRID SIZES USING RMSE IN METERS ON THE NGSIM DATASET

Grid Size	1s	2s	3s	4s	5s	Average
13x3_grid	0.52	1.10	1.79	2.66	3.78	1.97
25x5_grid	0.49	1.09	1.78	2.62	3.65	1.93

D. Qualitative Analysis

The performance of the proposed algorithm in terms of prediction accuracy, general applicability, reactive prediction, attention mechanism, and computation time is further discussed in this subsection. In addition, the prediction results on real-world data are also evaluated.

1) Comparison With Benchmark: To comprehensively examine the performance of the proposed method in various traffic scenarios, the typical method in [23] is compared with the proposed method in terms of prediction distribution in the same traffic situation, as they both generate multimodal trajectories through maneuver-based decoding. As shown in Fig. 8, the historical trajectories are represented by black solid lines, and the real trajectories are represented by blue solid lines. The vehicles with different background colors represent the ego vehicle and target vehicles, respectively, and the vehicles without color backgrounds represent neighboring vehicles. Further, the ground truth, planned and predicted trajectories are displayed as sets of waypoints with a time interval of 0.2 s. Note that only the predicted trajectories with maneuver probabilities larger than 10% are presented here.

In Fig. 8(a), the target vehicle (the vehicle with orange backdrop) in the left-rear position of the ego vehicle generates two predicted trajectories with different probabilities [see top of Fig. 8(a)]. In contrast, the proposed method [see bottom of Fig. 8(a)] focuses more on the lane-changing maneuver based on the larger probability. This is because the planned trajectory of the ego vehicle over the future time domain is specified as keeping straightforward driving in its original lane, which means the target vehicle is informed that it has sufficient space to make a right-turning maneuver. In Fig. 8(b), the target vehicle under the conventional prediction method generates two types of maneuvers (more than 10% probability), i.e., straightforward driving and left-turning, while the proposed method accurately predicts that the target vehicle would keep straightforward driving due to the negligible probability of other maneuvers. This is because the ego vehicle will gently merge into the right lane, which would create a potential collision risk for the target vehicle to make a left-turning operation in the future time domain. That is, the planned trajectory of the ego vehicle compresses the feasible space of the target vehicle, thus making it closer to the real trajectory. Such results show the superiority of the proposed method in terms of multimodal uncertainty reduction and prediction accuracy enhancement due to the consideration of future states of the ego vehicle.

2) Traffic Flow Prediction: Considering the nonlinear relationship between prediction accuracy and scene complexity, the performance of the proposed prediction model in various traffic flows should be equally noteworthy. The US101 traffic flow data in NGSIM are determined for trajectory prediction. These data were collected from 07:50 to 08:05 A.M., 08:05 to 08:20 A.M., and 08:20 to 08:35 A.M., corresponding to traffic flows of 2169, 2017, and 1915 vehicles. The more complex the scene, the more quickly the prediction error would increase. Consequently, the time period of 08:05–08:20 A.M., during which the traffic is in between uncongested and congested conditions during the time period and is termed as moderate congestion, was first considered. Fig. 9(a) and (b) illustrates the overall prediction results when the planning module turns off and on, respectively, where the parts indicated by the orange dashed boxes are presented in Fig. 9(c) and (d). The definition of each part in Fig. 9 is the same as that in Fig. 8. Therefore, the predicted trajectories generated by the LSTM-based social convolutional network model without the planning module and attention mechanism are denoted by purple dotted lines, which can also be described as LSTM-only

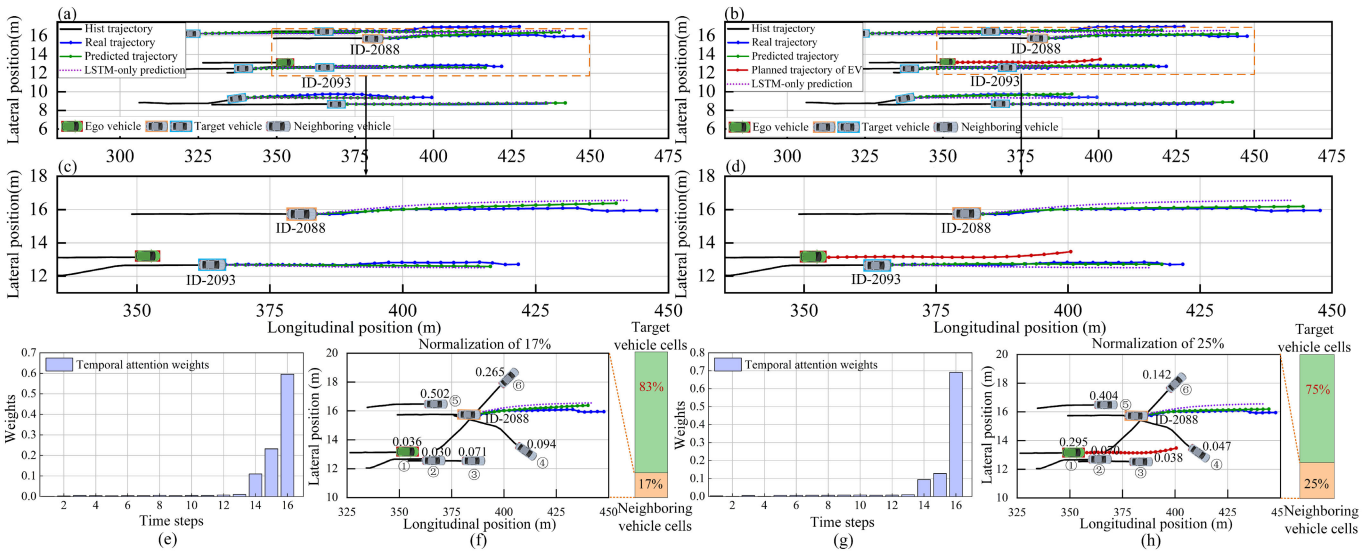

Fig. 9. Prediction results and attention distribution during the time period of 08:05–08:20 A.M. on highway US101. (a) and (c) Prediction results for the neighboring vehicle trajectory and the further enlargement trajectory when the planning module is turned off. (b) and (d) Corresponding results when turning on the planning module. (e) and (g) Changing trend of the temporal attention weight. (f) and (h) Spatial attention distribution of the target vehicle in the two cases.

prediction. The trajectories predicted by the proposed method are shown by green solid lines. Notably, all the neighboring vehicles that interacted with the target vehicle are not presented together in Fig. 9(a). For convenience, one of the target vehicles, vehicle ID-2088, is chosen to show the detailed attention weight distribution. The temporal attention distribution of the target vehicle ID-2088 in the current prediction period is given in Fig. 9(e) and (g). The spatial attention weight distribution of the target vehicle and the neighboring vehicles' contribution to the predicted trajectory in the prediction process are represented in Fig. 9(f) and (h).

According to Fig. 9(a) and (c), the proposed method has better prediction accuracy in predicting the trajectories of target vehicles, with the overall prediction trend remaining consistent with the real trajectory. This is because the spatio-temporal attention mechanism can effectively divide the influence weights of vehicles around the target vehicle and capture the key information affecting the predicted trajectories, as shown in Fig. 9(e) and (f). Consequently, the attention mechanism can capture the features between the trajectory data of interacting vehicles through the trained network model and provide a more reasonable and interpretable interactive prediction method. In contrast, the prediction results of the LSTM-only network have a significant bias in the lateral direction, especially for the target vehicles ID-2088 and ID-2093, as shown in Fig. 9(b); the predicted trajectory of the target vehicle ID-2088 gradually deviates to the left with time, while the predicted trajectory of the target vehicle ID-2093 demonstrates a tendency to move to the right, both of them exhibiting large prediction biases. As shown in Fig. 9(f), the weight of the target vehicle ID-2088 in the prediction process reaches 83% while the weight of the neighboring vehicles is 17%, which reveals that the future trajectory of the target vehicle strongly relies on its own state. To better visualize the distribution of neighboring vehicles, the weights of neighboring vehicles are normalized and reconstructed in a 25 × 5 grid, and it can be seen that the vehicle with the largest weight on the trajectory prediction of the target vehicle ID-2088 is its left-rear vehicle, which is over 50%, while the weight of the ego vehicle, as one of the neighboring vehicles for the target vehicle, is negligible. Since the planned trajectory of the ego vehicle cannot be obtained in advance, it cannot provide further information to act on the predicted trajectory of the target vehicle. As shown in Fig. 9(b) and (d), by cooperating with the planning module of the ego vehicle, the proposed method can generate more accurate predicted trajectories for the target vehicles, especially for ID-2088 and ID-2093. Moreover, the predicted trajectories are almost the same as the actual trajectories in the 3-s prediction horizon, and only a small deviation occurs in longer prediction horizons. This can be attributed to the fact that the attention distribution of both the target vehicle itself and the surrounding vehicles has been changed due to the engagement of the planning information of the ego vehicle, which is shown in Fig. 9(h). The predicted trajectory of the target vehicle with the assistance of the planning module is closer to the actual trajectory.

For LSTM-only prediction, since the spatio-temperal attention distribution and the planning information of the ego vehicle are not considered, it exhibits unsatisfactory performance when the external environment changes. For the temporal attention weight distribution, as indicated in Fig. 9(e) and (g), the contribution of the historical trajectory at different timestamps to the prediction decreases nonlinearly with the increasing distance from the current position. In particular, the trajectory at the last time step has the largest influence on the prediction with more than 55%, except for the trajectories within 0.5 s, and the historical trajectories away from the current moment have an approximate weight of 0. Overall, the historical trajectories within 0.5 s have large

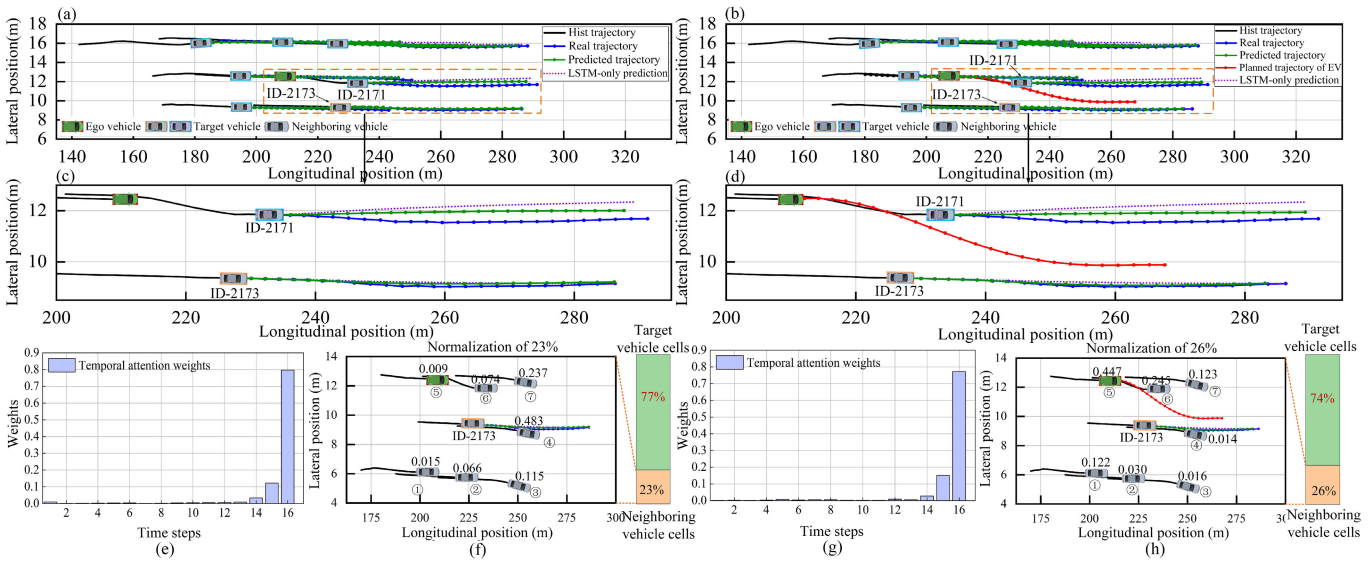

Fig. 10. Prediction results and attention distribution during the time period of 07:50–08:05 A.M. on highway US101. (a) and (c) Prediction results for the neighboring vehicle trajectory and the further enlargement of the trajectory when the motion trend module is turned off. (b) and (d) Corresponding results of turning on the planning module. (e) and (g) Changing trends in the temporal attention weight. (f) and (h) Spatial attention distribution of the target vehicle in the two cases.

influence on the prediction result, which aligns well with the intuition. For the spatial attention weight distribution, the two predicted spatial attention distributions for the target vehicle ID-2088 are shown in Fig. 9(f) and (h). It can be observed that the weight of the influence of the target vehicle's own state on the predicted trajectory decreases after the planning module is turned on, while the weights of the neighboring vehicles are increased. To better show the distribution of attention weights of the neighboring vehicles, their weights are normalized. It can be observed that the proportion of the ego vehicle in the spatial attention weight of the target vehicle increases from close to 0% to 29.5%. This change in spatial attention weight also reflects the real traffic situation, which illustrates that when the ego vehicle has driving intention, the surrounding vehicles would pay attention to the changes and have the tendency to compromise or compete. Collectively, the introduction of the attention mechanism and the planning information of the ego vehicle reduce the uncertainty of prediction in a certain degree, resulting in improved prediction accuracy and better explainability.

3) Prediction in Mix Traffic Flow: To illustrate the generalization ability of the proposed method in different traffic flows, the prediction results for the data segments from the congestion situations from 07:50 to 08:05 A.M. are presented in Fig. 10. The definitions of the parts in Fig. 10 are the same as those in Fig. 9. From Fig. 10(a) and (c), it can be observed that when the planning module turns off, there are significant lateral prediction errors for the target vehicles, especially for the target vehicles ID-2171 and ID-2173. The LSTM-only prediction method retains a similar trend with the real only with a slightly larger error. Observing the spatial attention distribution in Fig. 10(f), the weight of the target vehicle still holds a decisive role in the weight distribution for the current prediction of the target vehicle, which reveals that the future trajectory of the target vehicle largely depends on its own driving status. In addition, it can be found from Fig. 10(f) that the attention weights of the vehicles on both sides of the target vehicle are almost the same, which has little effect on the prediction result, and the target vehicle pays more attention to the front neighboring vehicles. As shown in Fig. 10(c) and (d), when the planning module turns on, the magnitude of the errors has been significantly reduced. From the perspective of spatial attention distribution, as shown in Fig. 10(h), the incorporation of the planning information increases the interaction between the ego vehicle and the target vehicle. At this point, the weight of the target vehicle itself is reduced from 77% to 74%, while the weight of the neighboring vehicle is increased to 26%. The reason for the change may be that the driver needs to pay more attention to the motion states of the neighboring vehicles in a more dense traffic flow scenario to avoid collision at any unexpected situation. As can be observed in the normalized weight distribution of neighboring vehicles, the weight of the ego vehicle is greatly increased to 44%, and this further compresses the prediction space of the target vehicles ID-2171 and ID-2173, especially for the target vehicle ID-2173. This can be attributed to the right lane change tendency of the ego vehicle, which causes the target vehicle ID-2173 to avoid ego vehicles and thus the predicted trajectory has a tendency of right-turning motion. The proposed method has higher prediction accuracy compared with the results of turning off the motion trend module in Fig. 10(a). As shown in Fig. 10(e) and (g), the temporal attention distribution in the congested scenario maintains a consistent trend compared to the moderate congested scenario, where the future trajectory of the vehicle is more dependent on the decisions near the last moment. This may be due to the fact that in denser traffic flows, drivers must pay increased attention to the current moment for safe driving.

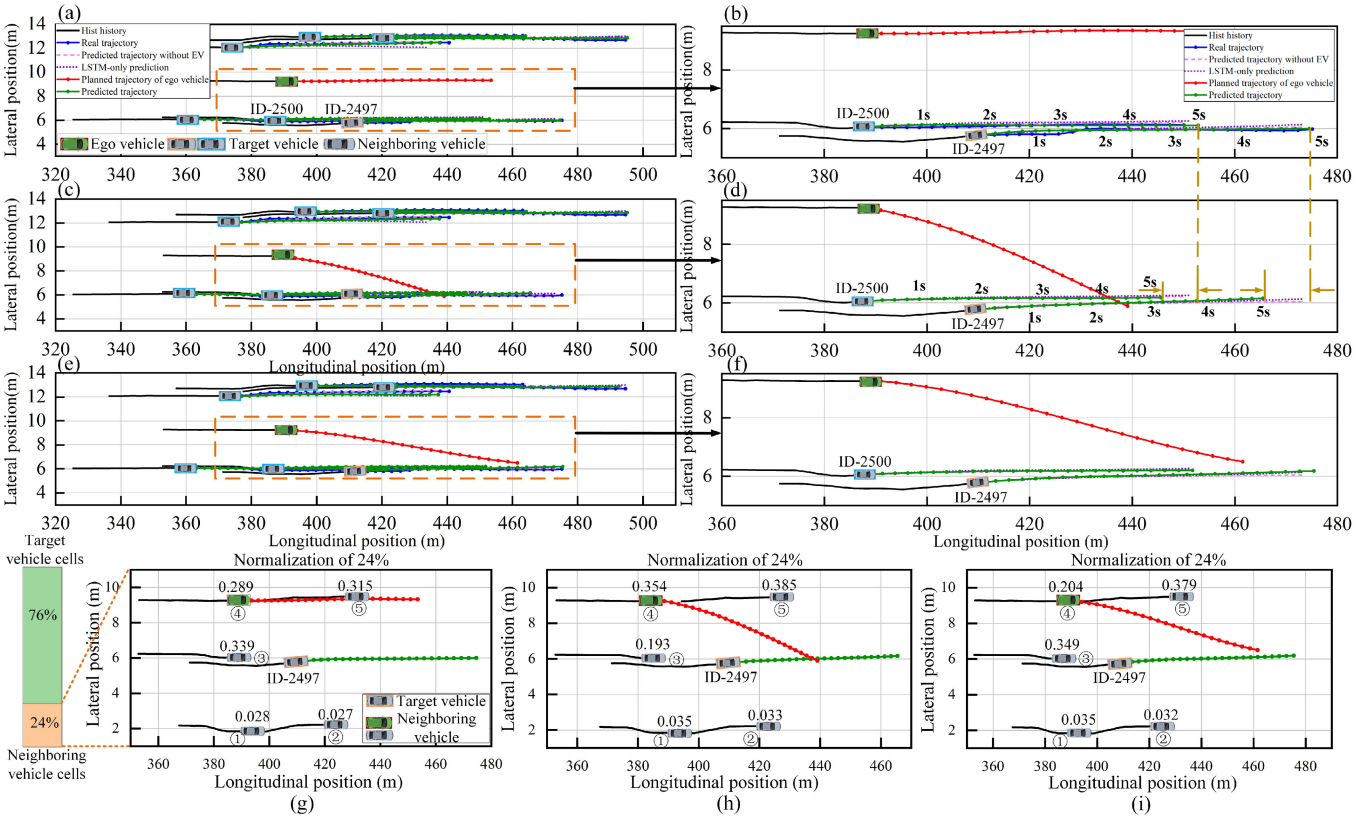

Fig. 11. Reactive prediction results reasoned from the ego vehicle's behavior. (a) and (b) Prediction results for the target vehicle and the further enlargement of the trajectory when the planning module is turned on. (c) and (d) Corresponding results when the ego vehicle is turning right in an aggressive manner. (e) and (f) Corresponding results when the ego vehicle is turning right in a mild manner. (g)–(i) Spatial attention distributions of the target vehicle with different plans of the ego vehicle.

In summary, the proposed method combines a trajectory prediction method with a planning module and an attention mechanism. It can significantly improve the prediction accuracy and quantify the impact of the target vehicle on its own and neighboring vehicles on the predicted trajectory. Specifically, the planning module is integrated to improve the prediction performance under congestion conditions. The attention mechanism can explicitly analyze the temporal and spatial distribution information that dominates the current prediction, and extract the driving characteristics of the target and neighboring vehicles under different operating conditions.

4) Reactive Trajectory Prediction: The trajectory planning module usually generates multiple future planning trajectories. To illustrate the flexibility of the proposed scheme to deal with different planning behaviors, the prediction results derived based on the NGSIM dataset are presented in Fig. 11. The definition of each part in Fig. 11 is the same as that in Fig. 10. As seen from Fig. 11(a) and (b), it achieves higher prediction accuracy when the planning module turns on compared to the LSTM-only approach, in which the planning information of the ego vehicle is obtained based on its future trajectory. Considering more significant noise with the NGSIM data, the quintic polynomial curves are generated for comparison by fitting the real trajectories of the ego vehicles with different driving behaviors, which are simplified in this study as mild or aggressive considering the dynamic characteristic. As shown in Fig. 11(c) and (d), it is assumed that the ego vehicle performs a right-turning maneuver aggressively, which would disturb the normal driving of the target vehicles ID-2497 and ID-2500. It can be seen that the predicted trajectory of the target vehicle located in the right lane decelerates compared to the predicted trajectory of the ego vehicle's normal driving behavior, and the gap is illustrated with a yellow dashed line in Fig. 11(d). The reason is that the ego vehicle's maneuver is captured and reacted by the planning module of the proposed scheme, which allows the target vehicle to keep driving in its lane while slowing down to avoid possible collision. As shown in Fig. 11(e) and (f), it is assumed that the ego vehicle performs a right-turning maneuver in a mild manner. As the ego vehicle leaves enough space for the target vehicle, the collision risk between the target and the ego vehicle is significantly reduced, and thus the predicted trajectory is basically consistent with the actual driving behavior. In contrast, the LSTM-only method demonstrates the same results against different planning trajectories of the ego vehicle because it fails to consider the reasonable information including the planning information of the ego vehicle. The evolutions of the spatial attention weight of the ego vehicle in three cases are shown in Fig. 11(g)–(i) while the weight of the target vehicle remains around 76%. In addition, it can be seen that the attention mechanism captures the feature when the planned trajectory of the ego vehicle is obtained based on the real trajectory with the weight of 28.9%. However, when the ego vehicle turns right in an aggressive manner, its weight increases rapidly to

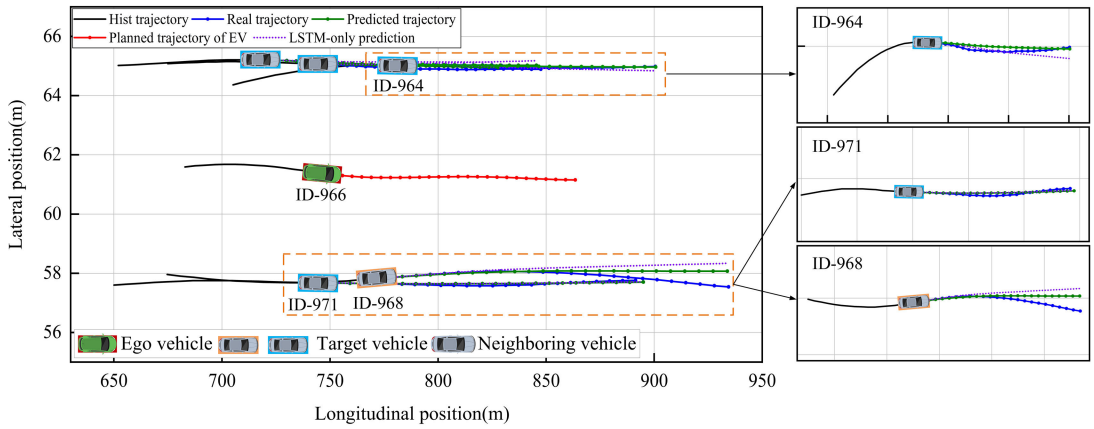

Fig. 12. Trajectory prediction results based on the HighD dataset.

around 35% due to the potential collision risk between the target and the ego vehicle, as shown in Fig. 11. Furthermore, the weight of the ego vehicle declines to the normal level when the ego vehicle turns right in a mild manner. In a nutshell, the proposed approach can respond to various future behaviors of the ego vehicle and demonstrates reasonable maneuver intentions while quantifying the influence of the neighboring vehicles on the target vehicle. This effectively contributes to efficient decision-making to improve driving safety.

5) Generalization Ability: To demonstrate the generalization ability of the proposed method, the results under dynamic traffic flow obtained from the HighD dataset are shown in Fig. 12. The line shapes and colors are consistent with those in Section III-D1. The orange boxes mark the two regions corresponding to the enlarged trajectories of the target vehicles on both sides of the ego vehicle, namely, ID-964, ID-971, and ID-968. As shown in Fig. 12, the proposed scheme still achieves good prediction performance for the target vehicles on the HighD dataset. The predicted trajectories generated by the proposed model outperform those generated by the LSTM-only method. For simplicity, Fig. 13 shows the temporal and spatial attention distributions of the target vehicle ID-968. For the temporal attention distribution, the temporal attention distribution is consistent with that in the aforementioned case, and the minute difference is that the temporal attention is more biased toward the weight at the last moment, accounting for approximately 80%. It indicates that the driver would greatly reduce the influence of the historical trajectory and pay more attention to the trajectory at the last moment to ensure smooth driving when the target vehicle is driving at high speeds. As shown in Fig. 13(b), the weight of the target vehicle itself is 75%, which is consistent with the general case. For the neighboring vehicles, the proportion of the ego vehicle in the spatial attention weight of the target vehicle exceeds 80%. This may be due to the fact that the target vehicle ID-968 has a tendency to move to the right, and the planning information of the ego vehicle ID-966 compresses the feasible space. This also reflects the real traffic scene information and characterizes the compromise or competition relationship of target vehicles when the ego vehicle has a driving intention.

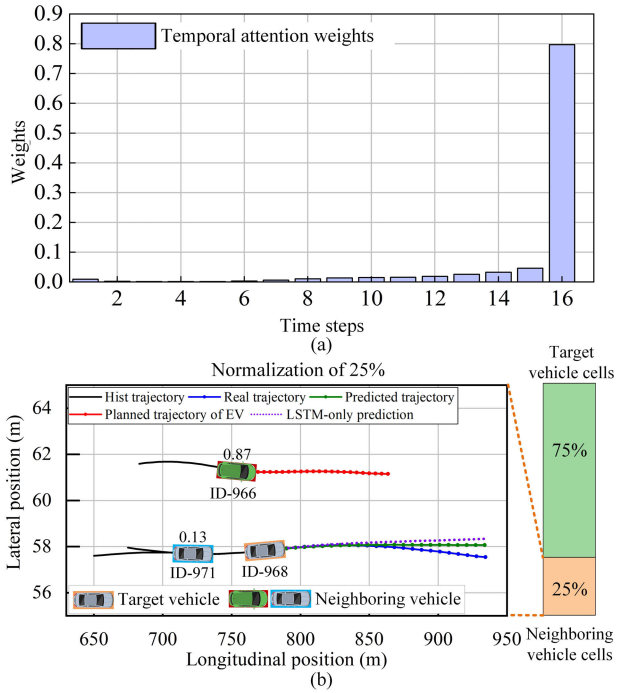

Fig. 13. Temporal and spatial attention distributions on the HighD dataset: (a) temporal attention distribution; (b) spatial attention distribution.

6) Discussion on Computation Efficiency: Computation efficiency is a crucial performance indicator for AVs. The computation time of the proposed algorithm is evaluated and listed in Table V. For a fair comparison, the code of *CS-LSTM* [23][1] is downloaded and run on the workstation to calculate its computation time. Table V shows that *CS-LSTM* takes 0.15 s to predict the trajectories of 1000 vehicles when the batch size is 128, while the proposed algorithm takes 0.33 s to perform the same task. In the scenario of autonomous driving applications, the resources are limited and the batch size can only be set to 1. The last column of Table V shows that the proposed method runs 2.8 times faster than *CS-LSTM*, which is attributed to the fact that the proposed method can simultaneously predict the trajectories of all observed target vehicles within the ego

[1]https://github.com/nachiket92/conv-social-pooling

TABLE V
COMPUTATION TIME

Method	Predicted objects	Time(s) 128 batch	Time(s) 1 batch
CS-LSTM	1000	0.15	7.41
The proposed	1000	0.33	2.68

vehicle-centered region while *CS-LSTM* can only predict one object. Such results reveal that the proposed method can meet the real-time deployment requirements of autonomous driving systems.

IV. CONCLUSION

This paper presents a hybrid trajectory prediction framework for automated vehicles. The framework combines the historical trajectories with the planned trajectories of the ego vehicle to improve prediction accuracy and reduce uncertainty. It can also deduce the reasonable interactive reactions among multiple traffic participants to minimize collision risk. Moreover, the proposed scheme is combined with a spatial-temporal attention mechanism to visualize the correlation between the historical trajectories of the ego vehicle and the target and neighboring vehicles on the prediction results by fitting the interactive features of massive trajectory data through the network. This expands the influence scale of the key elements and provides interpretable prediction results. Experiments based on the HighD and NGSIM datasets verify the effectiveness of the proposed scheme.

In future work, we will consider incorporating more detailed information, such as map information and other types of traffic participants, into the prediction model, and design a generic framework for environment representation to further improve prediction performance and generalization ability in diverse traffic scenarios.

REFERENCES

[1] G. Chen, K. Chen, L. Zhang, L. Zhang, and A. Knoll, "VCANET: Vanishing-point-guided context-aware network for small road object detection," *Automot. Innov.*, vol. 4, no. 4, pp. 400–412, Nov. 2021.

[2] C. Huang, P. Hang, Z. Hu, and C. Lv, "Collision-probability-aware human-machine cooperative planning for safe automated driving," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 9752–9763, Oct. 2021.

[3] X. Qu, D. Pi, L. Zhang, and C. Lv, "Advancements on unmanned vehicles in the transportation system," *Green Energy Intell. Transp.*, vol. 2, no. 3, Jun. 2023, Art. no. 100091.

[4] L.-L. Wang, Z.-G. Chen, and J. Wu, "Vehicle trajectory prediction algorithm in vehicular network," *Wireless Netw.*, vol. 25, no. 4, pp. 2143–2156, May 2019.

[5] W. Zeng, S. Wang, R. Liao, Y. Chen, B. Yang, and R. Urtasun, "DSDNet: Deep structured self-driving network," in *Proc. Eur. Conf. Comput. Vis.*, vol. 12366, 2020, pp. 156–172.

[6] M. Schreier, "Bayesian environment representation, prediction, and criticality assessment for driver assistance systems," *Automatisierungstechnik*, vol. 65, no. 2, pp. 151–152, Feb. 2017.

[7] M. Gulzar, Y. Muhammad, and N. Muhammad, "A survey on motion prediction of pedestrians and vehicles for autonomous driving," *IEEE Access*, vol. 9, pp. 137957–137969, 2021.

[8] G. Yu, H. Li, Y. Wang, P. Chen, and B. Zhou, "A review on cooperative perception and control supported infrastructure-vehicle system," *Green Energy Intell. Transp.*, vol. 1, no. 3, Dec. 2022, Art. no. 100023.

[9] J. Li, B. Dai, X. Li, X. Xu, and D. Liu, "A dynamic Bayesian network for vehicle maneuver prediction in highway driving scenarios: Framework and verification," *Electronics*, vol. 8, no. 1, p. 40, Jan. 2019.

[10] Y. Wang, Z. Liu, Z. Zuo, Z. Li, L. Wang, and X. Luo, "Trajectory planning and safety assessment of autonomous vehicles based on motion prediction and model predictive control," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 8546–8556, Sep. 2019.

[11] S. Zhang, Y. Zhi, R. He, and J. Li, "Research on traffic vehicle behavior prediction method based on game theory and HMM," *IEEE Access*, vol. 8, pp. 30210–30222, 2020.

[12] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Driver behavior classification at intersections and validation on large naturalistic data set," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 724–736, Jun. 2012.

[13] H. Guo, Q. Meng, D. Cao, H. Chen, J. Liu, and B. Shang, "Vehicle trajectory prediction method coupled with ego vehicle motion trend under dual attention mechanism," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.

[14] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, "SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1349–1358.

[15] Y. Wang, S. Zhao, R. Zhang, X. Cheng, and L. Yang, "Multi-vehicle collaborative learning for trajectory prediction with spatio-temporal tensor fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 236–248, Jan. 2022.

[16] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1179–1184.

[17] 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.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[19] P. Han, W. Wang, Q. Shi, and J. Yue, "A combined online-learning model with K-means clustering and GRU neural networks for trajectory prediction," *Ad Hoc Netw.*, vol. 117, Jun. 2021, Art. no. 102476.

[20] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified LSTM models for trajectory prediction," *IEEE Access*, vol. 7, pp. 38287–38296, 2019.

[21] M. Khakzar, A. Rakotonirainy, A. Bond, and S. G. Dehkordi, "A dual learning model for vehicle trajectory prediction," *IEEE Access*, vol. 8, pp. 21897–21908, 2020.

[22] 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.

[23] 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. 1549–15498.

[24] M. Shahverdy, M. Fathy, R. Berangi, and M. Sabokrou, "Driver behavior detection and classification using deep convolutional neural networks," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113240.

[25] S. Casas, C. Gulino, R. Liao, and R. Urtasun, "SpAGNN: Spatially-aware graph neural networks for relational behavior forecasting from sensor data," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 9491–9497.

[26] R. Greer, N. Deo, and M. Trivedi, "Trajectory prediction in autonomous driving with a lane heading auxiliary loss," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 4907–4914, Jul. 2021.

[27] G. Xiong et al., "Online measurement error detection for the electronic transformer in a smart grid," *Energies*, vol. 14, no. 12, p. 3551, 2021.

[28] P. Dendorfer, A. Osep, and L. Leal-Taixé, "Goal-GAN: Multi-modal trajectory prediction based on goal position estimation," in *Proc. Asian Conf. Comput. Vis.*, 2021, pp. 405–420.

[29] S. Eiffert, K. Li, M. Shan, S. Worrall, S. Sukkarieh, and E. Nebot, "Probabilistic Crowd GAN: Multimodal pedestrian trajectory prediction using a graph vehicle-pedestrian attention network," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5026–5033, Oct. 2020.

[30] T. Zhao et al., "Multi-agent tensor fusion for contextual trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12118–12126.

[31] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Attention based vehicle trajectory prediction," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 1, pp. 175–185, Mar. 2021.

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

[33] Y. Wu et al., "HSTA: A hierarchical spatio-temporal attention model for trajectory prediction," *IEEE Trans. Veh. Technol.*, vol. 70, no. 11, pp. 11295–11307, Nov. 2021.

[34] L. Lin, W. Li, H. Bi, and L. Qin, "Vehicle trajectory prediction using LSTMs with spatial–temporal attention mechanisms," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 2, pp. 197–208, Mar. 2022.

[35] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft+hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection," *Neural Netw.*, vol. 108, pp. 466–478, Dec. 2018.

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

[37] B. Liu and Y. Liang, "Optimal function approximation with ReLU neural networks," *Neurocomputing*, vol. 435, pp. 216–227, May 2021.

[38] A. Sadat, S. Casas, M. Ren, X. Wu, P. Dhawan, and R. Urtasun, "Perceive, predict, and plan: Safe motion planning through interpretable semantic representations," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 414–430.

[39] W. Zeng et al., "End-to-end interpretable neural motion planner," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8652–8661.

[40] J. Liu, W. Zeng, R. Urtasun, and E. Yumer, "Deep structured reactive planning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 4897–4904.

[41] N. Rhinehart, R. Mcallister, K. Kitani, and S. Levine, "PRECOG: PREdiction conditioned on goals in visual multi-agent settings," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2821–2830.

[42] H. Song, W. Ding, Y. Chen, S. Shen, M. Y. Wang, and Q. Chen, "PiP: Planning-informed trajectory prediction for autonomous driving," in *Proc. Eur. Conf. Comput. Vis.*, vol. 12366, 2020, pp. 598–614.

[43] L. Yang, C. Lu, G. Xiong, Y. Xing, and J. Gong, "A hybrid motion planning framework for autonomous driving in mixed traffic flow," *Green Energy Intell. Transp.*, vol. 1, no. 3, Dec. 2022, Art. no. 100022.

[44] Z. Huang, H. Liu, J. Wu, W. Huang, and C. Lv, "Learning interaction-aware motion prediction model for decision-making in autonomous driving," 2023, *arXiv:2302.03939*.

[45] Z. Huang, H. Liu, J. Wu, and C. Lv, "Differentiable integrated motion prediction and planning with learnable cost function for autonomous driving," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, 2023, doi: 10.1109/TNNLS.2023.3283542.

[46] 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.

[47] M. Hasan, A. Solernou, E. Paschalidis, H. Wang, G. Markkula, and R. Romano, "Maneuver-aware pooling for vehicle trajectory prediction," 2021, *arXiv:2104.14079*.

[48] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2015, *arXiv:1412.6980*.

[49] Z. Zhang, L. Zhang, J. Deng, M. Wang, Z. Wang, and D. Cao, "An enabling trajectory planning scheme for lane change collision avoidance on highways," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 1, pp. 147–158, Jan. 2023, doi: 10.1109/TIV.2021.3117840.

[50] C. James and H. John, *U.S. Highway 101 Dataset*, WHO, Geneva, Switzerland, 2007, p. 3.

[51] 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.

Mingqiang Wang is currently pursuing the Ph.D. degree in mechanical engineering with the National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing, China.

His research interests include trajectory prediction, motion trajectory planning, and active safety control for intelligent electric vehicles.

Lei Zhang (Member, IEEE) received the Ph.D. degree in mechanical engineering and electrical engineering from the Beijing Institute of Technology, Beijing, China, and the University of Technology, Sydney, NSW, Australia, in 2016.

He is currently an Associate Professor with the School of Mechanical Engineering, Beijing Institute of Technology. His research interests lie in the area of control theory and engineering applied to electrified vehicles with emphases on battery management techniques, vehicle dynamics control, and autonomous driving technology.

Dr. Zhang is a member of China Society of Automotive Engineers (CSAE). He serves on the Technical Committee on Vehicle Control and Intelligence and the Technical Committee on Parallel Intelligence in Chinese Association of Automation (CAA). He has served as a Guest Editor for several journals, including *International Journal of Vehicle Design*, *Chinese Journal of Mechanical Engineering*, and *China Journal of Highway and Transport*. He serves as an Associate Editor for *Proceedings of the Institution of Mechanical Engineers Part C: Journal of Mechanical Engineering Science*, *SAE International Journal of Connected and Autonomous Vehicles*, and *SAE Journal of Electrified Vehicles*.

Jun Chen (Senior Member, IEEE) received the bachelor's degree in automation from Zhejiang University, Hangzhou, China, in 2009, and the Ph.D. degree in electrical engineering with minor in computer science from Iowa State University, Ames, IA, USA, in 2014.

He was with the Idaho National Laboratory, Detroit, Michigan, USA, from 2014 to 2016, and with General Motors, Idaho Falls, ID, USA, from 2017 to 2020. He joined Oakland University, Rochester, MI, USA, in 2020, where he is currently an Assistant Professor with the ECE Department. His research interests include advanced control and optimization, model-predictive control, machine learning, and stochastic hybrid systems, with applications in automotive and energy systems.

Dr. Chen was a recipient of the Best Paper Award from IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, the Publication Achievement Award from the Idaho National Laboratory, the Research Excellence Award from Iowa State University, and the Outstanding Student Award from Zhejiang University. He is/was an Associate Editor or an Editorial Board Member of *Journal of Control and Decision*, *Energy Systems*, IFAC Symposium on Advances in Automotive Control, and IEEE International Conference on Robotics and Automation. He is currently a member of SAE.

Zhiqiang Zhang is currently pursuing the Ph.D. degree in mechanical engineering with the School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China.

His research interests include vehicle dynamics and active safety control for intelligent electric vehicles.

Zhenpo Wang (Member, IEEE) received the Ph.D. degree in automotive engineering from the Beijing Institute of Technology, Beijing, China, in 2005.

He is currently a Professor with the Beijing Institute of Technology, and also the Director of the National Engineering Laboratory for Electric Vehicles. He has published four monographs and translated books as well as more than 80 technical articles. His current research interests include pure electric vehicle integration, packaging and energy management of battery systems, and charging station design.

Prof. Wang was a recipient of numerous awards including the Second National Prize for Progress in Science and Technology, the First Prize for Progress in Science and Technology from the Ministry of Education, China, and the Second Prize for Progress in Science and Technology from the Beijing Municipal, China.

Dongpu Cao (Senior Member, IEEE) received the Ph.D. degree from Concordia University, Montreal, QC, Canada, in 2008.

He is currently a Professor with Tsinghua University, Beijing, China. He has contributed more than 200 articles and three books. His current research focuses on driver cognition, automated driving, and cognitive autonomous driving.

Prof. Cao received the SAE Arch T. Colwell Merit Award in 2012, IEEE VTS 2020 Best Vehicular Electronics Paper Award, and six best paper awards from international conferences. He has served as the Deputy Editor-in-Chief for *IET Intelligent Transport Systems Journal* and an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE/ASME TRANSACTIONS ON MECHATRONICS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE/CAA JOURNAL OF AUTOMATICA SINICA, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, and *ASME Journal of Dynamic Systems, Measurement, and Control*. He is an IEEE VTS Distinguished Lecturer.