

doi: 10.19562/j.chinasae.qcgc.2025.05.002

# 基于稀疏注意力的时空交互车辆轨迹预测\*

高凯<sup>1,2</sup>, 刘欣宇<sup>2</sup>, 胡林<sup>2</sup>, 黄向明<sup>1</sup>, 邹铁方<sup>2</sup>, 刘鹏<sup>3</sup>(1. 湖南大学机械与运载工程学院, 长沙 410082; 2. 长沙理工大学汽车与机械工程学院, 长沙 410114;  
3. 湖南星邦智能装备股份有限公司, 长沙 410600)

**[摘要]** 在混合交通环境中, 准确预测周边车辆轨迹对自动驾驶汽车安全至关重要。然而, 现有技术在长时预测方面仍存在精度低和计算量大的问题。本文提出了一种结合意图概率的时空交互稀疏注意力模型, 通过高效的编码-解码结构进行轨迹预测。模型首先构建位置掩码矩阵提取历史轨迹中的位置信息, 利用稀疏注意力机制筛选出关键特征, 并通过意图行为分析模块提高意图识别的准确率。最终将时空特征、位置特征和意图特征融合输入解码器, 以多任务学习方式训练模型。试验结果表明, 该模型在HighD和NGSIM数据集上相较于当前最优算法, 在3~5 s长时预测的均方根误差均有降低, 显著提升了预测效果。此外, 通过实车试验对模型在实际场景中的表现进行验证, 进一步展示了其在复杂交通环境中的应用潜力。

**关键词:** 交通工程; 轨迹预测; 稀疏注意力; 深度学习

## Vehicle Trajectory Prediction with Spatial-Temporal Interaction Based on Sparse Attention

Gao Kai<sup>1,2</sup>, Liu Xinyu<sup>2</sup>, Hu Lin<sup>2</sup>, Huang Xiangming<sup>1</sup>, Zou Tiefang<sup>2</sup> & Liu Peng<sup>3</sup>1. College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082;  
2. College of Automotive and Mechanical Engineering, Changsha University of Science & Technology, Changsha 410114;  
3. Hunan Sinoboom Intelligent Equipment Co., Ltd., Changsha 410600

**[Abstract]** In a mixed traffic ecosystem, accurately predicting the trajectories of surrounding vehicles is crucial for the safety of autonomous vehicles. However, existing technologies still face issues of accuracy and computational complexity in long-term prediction. A spatiotemporal interactive sparse attention model combined with intention probability is proposed in this paper, which predicts trajectories through an efficient encoder-decoder structure. The position mask matrix is first constructed to extract positional information from historical trajectories, and key features are selected using the sparse attention mechanism. The intention behavior analysis module is utilized to improve the accuracy of intention recognition. Finally, spatiotemporal features, positional features, and intention features are fused and input into the decoder, and the model is trained using a multi-task learning approach. The experimental results show that, compared to the optimal algorithm on the HighD and NGSIM datasets, the proposed model achieves a notable reduction in root mean square error (RMSE) in long-term prediction of 3 to 5 seconds, significantly enhancing prediction accuracy. In addition, the model's performance in real-world scenarios is validated through road tests, further demonstrating its application potential in complex traffic environment.

**Keywords:** traffic engineering; trajectory prediction; sparse attention; deep learning

\* 国家杰出青年科学基金(52325211)、国家自然科学基金(52172399)和湖南省自然科学基金(2024JJ5023)资助。

原稿收到日期为2024年08月14日, 修改稿收到日期为2024年11月27日。

通信作者: 胡林, 教授, 博士生导师, 博士, E-mail: hulin@csust.edu.cn;

黄向明, 教授, E-mail: h\_xiangming@aliyun.com。

## 前言

随着人工智能技术的迅速发展和应用,自动驾驶汽车商业应用加速,在相当长时间内与传统汽车混行的交通场景将成为常态。在智能系统中,车辆轨迹预测技术至关重要。通过分析周围环境、道路状态和其他车辆行为,预测车辆未来行驶轨迹,为自动驾驶提供精准的决策支持,有效避免交通事故。然而,自动驾驶汽车在实际运营中面临复杂的挑战,尤其在城市环境中,交通参与者的多样性增加了系统复杂性。自动驾驶汽车须快速准确地识别和适应不断变化的道路条件,从而对轨迹预测的实时性和精准性提出更高要求。

为了提高自动驾驶汽车在实际运行过程中的安全性,国内外学者在车辆轨迹预测技术方面进行了大量的研究和实践,车辆轨迹预测技术经历了从传统机器学习方法到深度学习方法的演进。初始的轨迹预测方法主要采用传统机器学习技术,例如高斯过程和回归模型<sup>[1]</sup>。Hubmann等<sup>[2]</sup>将车辆轨迹视为部分可观察的马尔科夫决策过程,通过基于点的求解器预测未来轨迹,为当时的方法提供了一种新的视角。然而,这些方法存在对手工设计特征和模型的依赖,对于复杂的交通场景表现出一定的局限性,难以在长时轨迹预测(3~5 s)方面取得突破性进展。随着深度学习技术的崛起,基于深度学习的轨迹预测方法逐渐成为主流,弥补了传统方法在处理复杂场景时的不足。Alahi等<sup>[3]</sup>首次提出了基于长短期记忆(LSTM)的轨迹预测方法,通过网格化周围位置允许目标共享隐藏状态。季学武等<sup>[4]</sup>还考虑了轨迹预测的不确定性,通过结合不同的驾驶意图信息预测周围车辆的轨迹。但由于交通环境的复杂性,车辆的运动还会受到周围环境的影响,只依赖LSTM无法充分考虑车辆间的动态交互信息<sup>[5-10]</sup>。因此,后续学者在此基础上将LSTM与CNN相结合<sup>[11-14]</sup>,但这些方法忽略了周围车辆与目标车辆之间的交互影响差异性。为了处理这些差异性引入了注意力机制,不少文献<sup>[15-18]</sup>利用注意力机制结合LSTM对车辆时间和空间特征进行建模以此表示车辆之间的动态变化。Cai等<sup>[19]</sup>提出了一个环境-注意网络模型,使用图注意力网络关注车辆与环境中的非欧氏相关结构(图结构)之间的相互作用,获取车辆与驾驶环境之间的全交互信息。然而,过分重视交通参与者的交互作用会限制模型的泛化能力不利于轨迹的

预测<sup>[20]</sup>。Giuliani等<sup>[21]</sup>利用Transformer的多头注意力机制,关注目标自身的运动状态,实现更为准确的轨迹预测。Hou等<sup>[22]</sup>提出了Transformer交互式轨迹预测方法,可以并行学习多个周围车辆之间的时空依赖性。在上述这些方法中,车辆的历史轨迹信息、状态信息以及交互信息都对轨迹预测至关重要,当把这些信息作为预测的输入特征时,模型可以提高预测精度做出更加合理的长时预测结果,但是也面临着复杂度和计算开销较大的问题,使得模型在实际场景中难以简化和推广<sup>[23]</sup>。在输入数据量较大时,神经网络模型会很难捕捉序列中长期的相互依赖性,产生梯度爆炸或梯度消失而影响模型的预测准确率。

鉴于此,本文提出一种结合驾驶意图的稀疏注意力机制轨迹预测方法,利用稀疏注意力机制捕获时序和空间交互关系,有效衡量目标车辆与周围车辆之间的特征相关性,进而提高模型的预测能力,减少模型的计算开销。同时进行多特征融合,将不同时间戳下时序特征、空间特征、环境特征聚合,并与意图特征进行组合,以全面提升模型对长时轨迹的预测效果。

## 1 问题定义

车辆轨迹预测的核心任务是根据目标车辆及其周围车辆的历史运动信息以及它们之间的相互作用,来预测目标车辆未来的运动轨迹及其可能的概率分布。在本文中,目标车辆被定义为自动驾驶车辆的临近车辆。图1展示了一个典型场景,其中目标车辆被标记为TV,目标车辆的正前方车辆标记为FV、左前方车辆标记为LF、右前方车辆标记为RF、左后方车辆标记为LB、右后方车辆标记为RB,在各个方向上距离目标车辆最近的车辆构成周围车辆。最终将目标车辆周围车辆标记为NV = {FV, LF, RF, LB, RB}。

### 1.1 模型的输入

为更精确地预测目标车辆轨迹,不仅考虑目标车辆的历史轨迹数据,引入周围车辆及其之间的交互信息。在时刻 $t(1 \leq t \leq T)$ ,输入特征表示为

$$\begin{aligned} X_t &= [I_t, O_t, N_t] \\ \left\{ \begin{aligned} I_t &= [x_t^{\text{TV}}, y_t^{\text{TV}}, v_{x,t}^{\text{TV}}, v_{y,t}^{\text{TV}}, a_{x,t}^{\text{TV}}, a_{y,t}^{\text{TV}}] \\ O_t &= [\Delta x_t, \Delta y_t, \Delta v_{x,t}, \Delta v_{y,t}] \\ N_t &= [y_t^n] \end{aligned} \right. \end{aligned} \quad (1)$$

式中:  $I_i$  为目标车辆自身物理状态信息矩阵;  $x_i^{TV}, y_i^{TV}, v_{x,i}^{TV}, v_{y,i}^{TV}, a_{x,i}^{TV}, a_{y,i}^{TV}$  分别为  $t$  时刻车辆在  $x$  和  $y$  方向上的位置、速度和加速度;  $O_i$  为相互作用信息矩阵;  $\Delta x_i, \Delta y_i, \Delta v_i$  分别是 TV 相对于 NV 的横向位置差、纵向位置差和速度差;  $N_i$  为周围车辆纵向位移矩阵;  $n$  为对应的邻居车辆编号。因此模型的输入由过去  $T$  时刻中的  $X=[X_1, X_2, \dots, X_T]$  组成。

### 1.2 模型输出

由于输入特征是非线性函数  $\hat{Y} = y_{unq}(X)$ , 为了使得目标车辆的预测轨迹  $\hat{Y} = \{(\hat{x}_{T+1}^{TV}, \hat{y}_{T+1}^{TV}), \dots, (\hat{x}_{T+F}^{TV}, \hat{y}_{T+F}^{TV})\}$  尽可能接近将来的真实轨迹  $Y = \{(x_{T+1}^{TV}, y_{T+1}^{TV}), \dots, (x_{T+F}^{TV}, y_{T+F}^{TV})\}$ , 模型的输出是  $\hat{Y}$  上的概率分布  $P(\hat{Y}|X)$ 。同时, 为获得不确定性估计, 所提模型的输出被参数化为二元高斯分布。对于时刻  $t' = T + F$ , 将  $(x_{T+F}^{TV}, y_{T+F}^{TV})$  参数化为双变量高斯分布, 其均值为  $(\mu_{T+F,x}^{TV}, \mu_{T+F,y}^{TV})$ , 方差为  $((\sigma_{T+F,x}^{TV})^2, (\sigma_{T+F,y}^{TV})^2)$ , 相关系数为  $\rho_{T+F}$ <sup>[24]</sup>。因此, 本

模型的输出转化为  $\delta = [\delta_{T+1}, \delta_{T+2}, \dots, \delta_{T+F}]$ , 其中  $\delta_{T+F} = (\mu_{T+F,x}, \mu_{T+F,y}, \sigma_{T+F,x}, \sigma_{T+F,y}, \rho_{T+F})$ 。

## 2 模型结构

为精确预测车辆的长时轨迹, 必须考虑自车与邻车之间复杂的时间和空间相关性, 且能够对轨迹的不确定性进行合理的概率分布输出。因此, 本文提出了结合意图的时空交互轨迹预测模型 (spatio-temporal estimation with intent, STEI), 如图 1 所示。该模型包含以下模块: (1) 位置掩码矩阵 (position masking, PM) 模块, 描述周围车辆存在状态, 帮助模型理解环境状态。(2) 时序交互 (temporal encoding, TE) 模块, 使用稀疏自注意力机制捕捉目标车辆自身的动态特性及其运动趋势。(3) 空间交互 (spatial encoding, SE) 模块, 捕捉自车与邻车之间在不同时刻的空间交互关系。(4) 意图行为分析 (intent analysis, IA) 模块, 基于不同的驾驶意图融合时间和空间特征, 以得到车辆的驾驶意图概率。

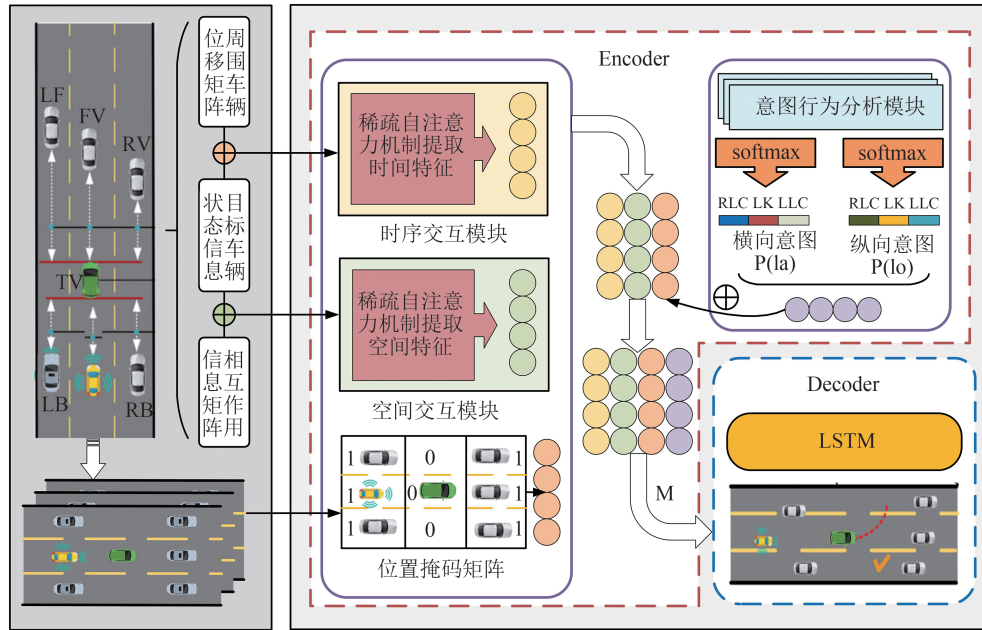


图 1 STEI模型整体框架

### 2.1 位置掩码矩阵的定义

目前研究主要采用车辆轨迹信息作为模型输入<sup>[25]</sup>, 但复杂交通环境使长时轨迹预测仍面临挑战。如果能获取周围车辆的状态信息, 预测结果可能会更准确, 因为附近车辆的位置会影响目标车辆的行驶轨迹。为此, 设计了利用 MASK 机制的位置掩码

矩阵, 以帮助目标车辆全面了解周围车辆的状态。在时刻  $t$ , 定义矩阵  $P_t$ :

$$P_t = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \quad (2)$$

如图 1 所示的换道场景下周围车辆集 NV, 将该

场景划分为九宫格网格的形式,按照图上的行驶方向定义 $P_i$ 中行列代表的含义。其中 $p_{22}$ 代表目标车辆, $p_{11}$ 代表目标车辆的左前方位置, $p_{12}$ 代表目标车辆的正前方位置,其他位置含义依此类推。由于目标车辆在换道过程中一侧通常处于无车状态,在换道行为已发生的情况下,另一侧车辆即使存在,其影响相对较小,因此可适当降低对该区域车辆状态的关注度,以简化模型计算。将 $p_{21}$ 、 $p_{22}$ 、 $p_{23}$ 的数值设置为0,对于其他位置的数值设置1。在不换道场景中若存在车辆则设置为1,不存在设置为0。从而获取 $t$ 时刻下的位置矩阵 $P = [P_1, P_2, \dots, P_r]$ ,将其传递至分别包含16和32个卷积核的一维卷积神经网络中以提取到不同尺度的特征,因此最终输出为特征 $\xi_p$ ,如式(3)所示。

$$\xi_p = 1DConv_{32}(1DConv_{16}(P)) \quad (3)$$

式中 $1DConv(\cdot)$ 代表一维卷积层。

## 2.2 稀疏自注意力机制

Vaswani等<sup>[26]</sup>定义的自注意力机制是基于元组输入,其计算复杂度较高,需要进行二次时间的点积计算。为了降低复杂度,研究发现自注意力的概率分布具有潜在稀疏性,可以有效减少计算量<sup>[27]</sup>。在实际算法中,只有少数点积对在注意力函数中起重要作用,其他点积对的权重可适当降低。意味着序列中的某个元素一般只与少数几个元素具有较高相似性。因此将 $query$ 稀疏性度量定义为

$$M(q, K) = \max_j \left\{ \frac{q_i K_j^T}{\sqrt{d}} \right\} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i K_j^T}{\sqrt{d}} \quad (4)$$

式中: $q_i$ 、 $k_i$ 、 $v_i$ 分别表示元组中的第 $i$ 行的查询、键和值的对应部分; $L_K$ 为 $Q$ 的长度。式(4)的等号右侧第1项对应最大点积,而第2项为算术平均值。为了进一步优化计算效率,若第 $i$ 个 $Q$ 的 $M(q, K)$ 取值较大,则说明其注意力权重影响力较高,因此可以选择排名最靠前的 $u$ 个 $Q$ 作为SoftMax函数的输入。对于每个 $key$ 而言,只关注 $u$ 个主要的 $query$ ,采用随机抽样的方式引入采样因子 $c$ ,设定 $u = c \ln L_Q$ 。同时,对其他点积对进行零填充,即可实现选择稀疏的 $Top-u$ 作为 $\bar{Q}$ 。稀疏注意力的运算机制就可以表示为

$$A(Q, K, V) = \text{Softmax} \left( \frac{\bar{Q}K^T}{\sqrt{d}} \right) V \quad (5)$$

## 2.3 基于稀疏注意力机制的时序交互特征学习模块

在复杂交通环境中,目标车辆的运动路径受周

围车辆相对距离的影响,导致轨迹多样性。因此,在预测目标车辆行驶轨迹时,不仅须考虑自身的历史运动状态,还须综合周围车辆的互动信息。为了获取时序交互模块的输入特征数据,构建如图2所示的时间交互稀疏注意力模块,首先将目标车辆的状态信息 $I_i$ 和相互作用信息 $O_i$ 通过全连接层嵌入到高维空间,以提取丰富的特征信息。

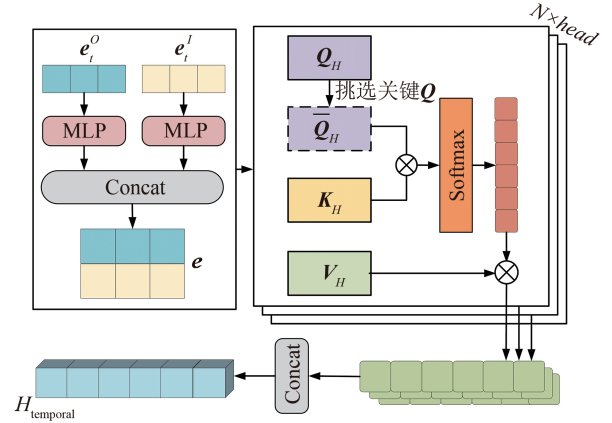


图2 时间交互稀疏注意力模块

$$e_i^e = \text{MLP}(I_i, W_1) \quad (6)$$

$$e_i^o = \text{MLP}(O_i, W_2) \quad (7)$$

式中: $e_i^e$ 、 $e_i^o$ 分别代表状态信息和相互作用信息的特征向量;MLP代表全连接层; $W_1$ 、 $W_2$ 是可训练的权重矩阵。同时为了让模型学习到这种交互信息,将 $e_i^e$ 、 $e_i^o$ 两个特征向量连接为一个整体,采用自适应激活函数APReLU减少计算量<sup>[28]</sup>。

$$e = \text{Concat}(e_i^e, e_i^o) \quad (8)$$

将得到的 $e$ 作为稀疏自注意力的输入,使用3个可学习的线性变换去获得3个不同的向量 $Q_H$ 、 $K_H$ 、 $V_H$ ,通过式(4)选取排名最靠前的若干个 $\bar{Q}$ 来进行点积计算得到不同时间戳的注意力权重 $W_{\text{temporal}}$ 。

$$\begin{cases} Q_H = \text{MLP}(e; W_{Q_H}) \\ K_H = \text{MLP}(e; W_{K_H}) \\ V_H = \text{MLP}(e; W_{V_H}) \end{cases} \quad (9)$$

$$W_{\text{temporal}} = \text{Softmax} \left( \frac{\bar{Q}_H K_H^T}{\sqrt{d}} \right) \quad (10)$$

随后,利用 $W_{\text{temporal}}$ 与值 $V_H$ 相乘得到时序交互特征矩阵。最后,使用 $n$ 个独立的头部去获得对特征的不同表示,Concat层聚合 $n$ 个特征子空间以提高模型对特征的代表能力。

$$h_{\text{temporal}}^i = W_{\text{temporal}}^i \otimes V_U, 1 \leq i \leq n \quad (11)$$

$$H_{\text{temporal}} = \text{Concat} \left[ h_{\text{temporal}}^1, h_{\text{temporal}}^2, \dots, h_{\text{temporal}}^n \right] W_{\text{temporal}}^o \quad (12)$$

TE模块充分挖掘了观测时段内各时刻历史状态特征向量与相互作用信息特征向量的时间依赖性,更有效地利用了历史信息。

#### 2.4 基于稀疏注意力机制的空间交互特征学习模块

在复杂交通场景中,目标车辆与周围车辆的空间关系至关重要。换道车辆通过观察相邻车辆的速度、位置和航向偏差来调整自身轨迹。尽管之前的模块提取了时间交互关系,但未考虑车辆与邻车之间的空间关系。为此,引入了空间交互特征学习模块(SE),如图3所示。该模块与TE模块类似,采用稀疏注意力机制,结合目标车辆的状态信息 $I_t$ 和周围车辆纵向位移矩阵 $N_t$ ,以捕捉周围车辆与目标车辆特征向量之间的交互关系。计算过程如下所示:

$$U = \text{Concat}(I_t, N_t) \quad (13)$$

$$\begin{cases} Q_U = \text{MLP}(U; W_{Q_U}) \\ K_U = \text{MLP}(U; W_{K_U}) \\ V_U = \text{MLP}(U; W_{V_U}) \end{cases} \quad (14)$$

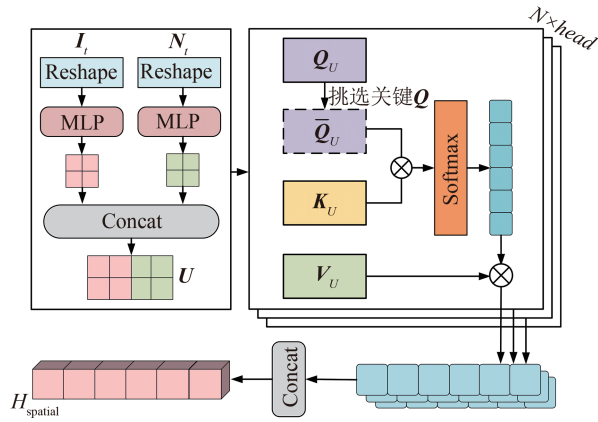


图3 空间交互稀疏注意力模块

同样地,通过将原始隐藏状态值 $V_U$ 与权重 $W_{\text{spatial}}$ 相乘得到空间交互特征矩阵,利用矩阵包含空间信息的上下文向量以提取到不同的空间交互特征表示,如式(16)所示。

$$W_{\text{spatial}} = \text{Softmax} \left( \frac{\bar{Q}_U K_U^T}{\sqrt{d}} \right) \quad (15)$$

$$h_{\text{spatial}}^i = W_{\text{spatial}}^i \otimes V_U, 1 \leq i \leq n \quad (16)$$

$$H_{\text{spatial}} = \text{Concat} \left[ h_{\text{spatial}}^1, h_{\text{spatial}}^2, \dots, h_{\text{spatial}}^n \right] W_{\text{spatial}}^o \quad (17)$$

SE模块利用稀疏注意力机制有效捕获了目标车辆状态与周围车辆纵向位移矩阵的交互关系,获取车辆间动态特征,提高了预测的精度。

#### 2.5 意图行为分析模块

车辆在驾驶过程中会产生多种行为。在多车道高速公路场景下,横向行为包括左变道(LLC)、不变道(LK)和右变道(RLC),纵向行为则包括加速(ACC)、减速(DEC)和匀速(CON)。本文考虑了车辆的横向和纵向行为,如图4所示。具体而言,在获得时序交互特征、空间交互特征和位置掩码矩阵特征后,将它们连接以获取不同驾驶意图的概率。

$$M_{\text{link}} = \text{Concat}(\text{Head}_{\text{temporal}}, \text{Head}_{\text{spatial}}, \xi_P) \quad (18)$$

在获取拼接特征矩阵 $M_{\text{link}}$ 后, $M_{\text{link}}$ 经过如图4所示的流程,最终获得未来时刻车辆发生横向行为概率 $P(la)$ 和纵向行为概率 $P(lo)$ ,其中 $la \in \{\text{RLC}, \text{LK}, \text{LLC}\}$ ;  $lo \in \{\text{ACC}, \text{DEC}, \text{CON}\}$ , GELU为高斯误差线性单元激活函数。

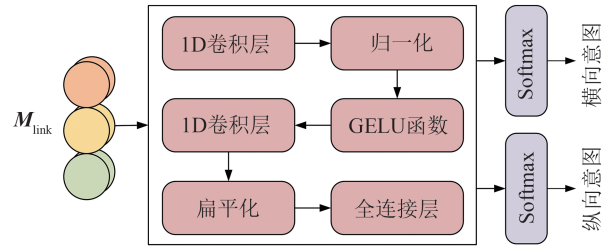


图4 意图行为分析模块

为了更好地理解车辆未来的运动趋势,将获得横向和纵向意图概率与输入的时间步进行拼接,获得时刻 $t' = T + F$ 下不同意图的未来轨迹概率分布 $\delta_{la,t'}$ 、 $\delta_{lo,t'}$ 。意图行为分析模块融合了先前模块的输出特征,形成新的特征图 $M$ ,如式(19)所示。

$$M = \text{Concat}(H_{\text{temporal}}, H_{\text{spatial}}, \xi_P, \delta_{la,t'}, \delta_{lo,t'}) \quad (19)$$

由于车辆预测轨迹在时间上具有连续性,同时为减少计算量利用LSTM对不同时刻下的特征图 $M$ 进行解码,最终再经MLP层输出所需的预测双变量高斯分布 $\delta_{t'}$ 。

$$h_{t'} = \text{LSTM}(M, h_{t'-1}, W^{\text{dec}}) \quad (20)$$

$$\delta_{t'} = \text{MLP}(h_{t'}, W^{\text{pre}}) \quad (21)$$

式中: $W^{\text{dec}}$ 为LSTM单元中的权重; $W^{\text{pre}}$ 将隐藏状态 $h_{t'}$ 映射为描述二元高斯分布参数的权重。

### 3 试验结果与分析

#### 3.1 数据准备

本文使用 HighD<sup>[29]</sup>和 NGSIM<sup>[30]</sup>数据集对所提模型进行评估。HighD数据集详细记录了德国6条不同类型高速公路的数据,采样频率为25 Hz,NGSIM数据集源于美国联邦公路局发起的下一代交通仿真计划,采样频率为10 Hz。为了适应试验场景并降低计算成本,两个数据集的采样频率设定为5 Hz,每个车辆序列的步长为8 s,其中历史序列为3 s,预测序列为5 s,筛选出满足条件的换道和不换道车辆数据。以目标车辆及其周围车辆构建每一个样本。为使深度学习网络易于训练,在筛选轨迹序列时,首先确定初始时刻 $t=1$  s时的坐标值 $(x_1, y_1)$ ,后续时刻的坐标值 $(x_t, y_t)$ 均与其相减。

为应对可能存在的车辆缺失情况,引入虚拟车辆的概念来模拟其影响<sup>[22]</sup>。虚拟车辆的位置分为两种情况:如果目标车辆的前方存在虚拟车辆,那么虚拟车辆的速度设置为 $v=1000$  m/s,与目标车辆的距离设置为 $d=1000$  m。反之,如果目标车辆的后方存在虚拟车辆,那么 $-v=0$  m/s,  $d=-1000$  m。

通过真实轨迹序列获取横向行为和纵向行为标签,以更好地根据意图行为进行轨迹预测,横向行为标签按照道路代号的改变划分,纵向行为标签按照式(22)划分。

$$\begin{cases} \text{ACC}, \beta \geq 1.2 \\ \text{DEC}, \beta \leq 0.8 \\ \text{CON}, \beta \in (0.8, 1.2) \end{cases} \quad (22)$$

其中 $\beta = \frac{(y_T - y_1)/F}{(y_{T+F} - y_T)/T}$ ,RLC、LK、LLC的标签代号分别为1、0、-1,ACC、CON、DEC标签代号分别为1、0、-1。

根据上述规则,从所有筛选的样本中选取70%作为训练集,20%作为测试集,10%作为验证集,其中HighD数据集中训练集样本数量为35 597个,测试集样本数量为10 170个,验证集样本数量为5 086个;NGSIM数据集中训练集样本数量为5 205个,测试集样本数量为1 487个,验证集样本数量为745个。

#### 3.2 模型设置

采用多任务学习的方法,以端到端的方式训练模型。对于意图识别,使用交叉熵损失函数来优化模型,如式(23)和式(24)所示。

$$NLL_{lat} = -\sum_{lat \in \{\text{LK, LLC, RLC}\}} y_{lat} \log P(lat|X) \quad (23)$$

$$NLL_{lon} = -\sum_{lon \in \{\text{ACC, DEC, CON}\}} y_{lon} \log P(lon|X) \quad (24)$$

对于轨迹预测,由于输出是二元高斯分布,所以损失计算公式为式(25)所示。

$$NLL_{traj} = \left\{ \sum_{t'=T+1}^{T+F} \log (2\pi\sigma_{t',x}\sigma_{t',y}\sqrt{1-\rho_{t'}^2}) + \frac{1}{2(1-\rho_{t'}^2)} \left[ \frac{(\mu_{t',x} - x_{t'})^2}{\sigma_{t',x}^2} - \frac{(\mu_{t',x} - x_{t'})(\mu_{t',y} - y_{t'})}{\sigma_{t',x}\sigma_{t',y}} + \frac{(\mu_{t',y} - y_{t'})^2}{\sigma_{t',y}^2} \right] \right\} \quad (25)$$

因此,模型的总损失为两个分类损失和回归损失的加权和,如式(26)所示。

$$NLL = \alpha_1 NLL_{traj} + \alpha_2 NLL_{lat} + \alpha_3 NLL_{lon} \quad (26)$$

在本文中,轨迹预测为主要任务,意图分类为次要任务,因此设置损失权重以区分两者的重要性,使得 $\alpha_1 = 1, \alpha_2 = \alpha_3 = 0.5$ 。使用Tensorflow<sup>[31]</sup>深度学习框架在NVIDIA Tesla P100GPU上以端到端方式训练网络。在训练中,当模型的验证损失在100 Epoch都不再下降时,停止训练网络,保存验证集损失最低的模型作为测试集推理模型。主要参数设置见表1。

表1 STEI模型参数

参数	数值
输入长度	15
输出长度	25
Batch_size	128
学习率	0.001
Epoch	2 000
Dropout	0.1
头部数目( $h$ )	5
特征数( $n_{features}$ )	26
优化算法	Adam
停止迭代	100 patience
损失函数	NLL

#### 3.3 模型对比

本文采用均方根误差(RMSE)来评估轨迹预测方法的预测性能<sup>[22,32-33]</sup>。公式如下:

$$RMSE = \sqrt{\frac{1}{NF} \sum_{n=1}^N \sum_{t'=T+1}^F \left\{ (\mu_{t',x}^n - x_{t'}^n)^2 + (\mu_{t',y}^n - y_{t'}^n)^2 \right\}} \quad (27)$$

式中: $N$ 是测试集的样本数量; $F$ 是预测时域,在本节

中设置其范围  $1 s \leq F \leq 5 s$ 。

以下是对比模型的简要介绍。

**S-LSTM<sup>[34]</sup>**:使用共享 LSTM 对每个车辆的轨迹数据编码,通过全连接的社会池化层提取特征,生成单峰输出分布。

**CS-LSTM<sup>[35]</sup>**:与 S-LSTM 不同,引入堆叠卷积层和池化层捕捉邻近车辆信息,并考虑基于预测策略的多模态。

**NLS-LSTM<sup>[36]</sup>**:基于 LSTM 的编解码器结构,结合本地和非本地操作捕获社会交互,但缺乏多模态预测。

**S-GAN<sup>[37]</sup>**:结合循环序列到序列模型和生成对抗网络,来聚合不同车辆信息,产生多个可信预测轨迹。

**PiP<sup>[33]</sup>**:利用自主车辆规划的多条轨迹信息,将其传递给预测模块,更准确地生成多个交通参与者的未来轨迹。

**STDAN<sup>[38]</sup>**:应用时空动态注意网络进行车辆轨迹预测,通过分层综合捕捉时间和社会关系,深入理解复杂交通场景。

**iNATran<sup>[39]</sup>**:基于 Transformer 的创新方法,具备多注意力学习和意图感知能力,实现多模态车辆轨迹预测,提高了效率和实时性。

**STEI**:本节所提模型。

将以上模型在同等条件下进行对比,结果如图 5 所示,数值统计如表 2 所示。模型在 1~2 s 的均方根误差(RMSE)仅低于 iNATran 模型,基本满足短期预测要求,在 3~5 s 内 STEI 的均方根误差对比其他模型都有所提升,与表中长期预测效果最佳的 iNATran 模型相比, RMSE 值在 HighD 数据集上分别降低了 14% (0.03)、59% (0.32)、74% (0.82),在

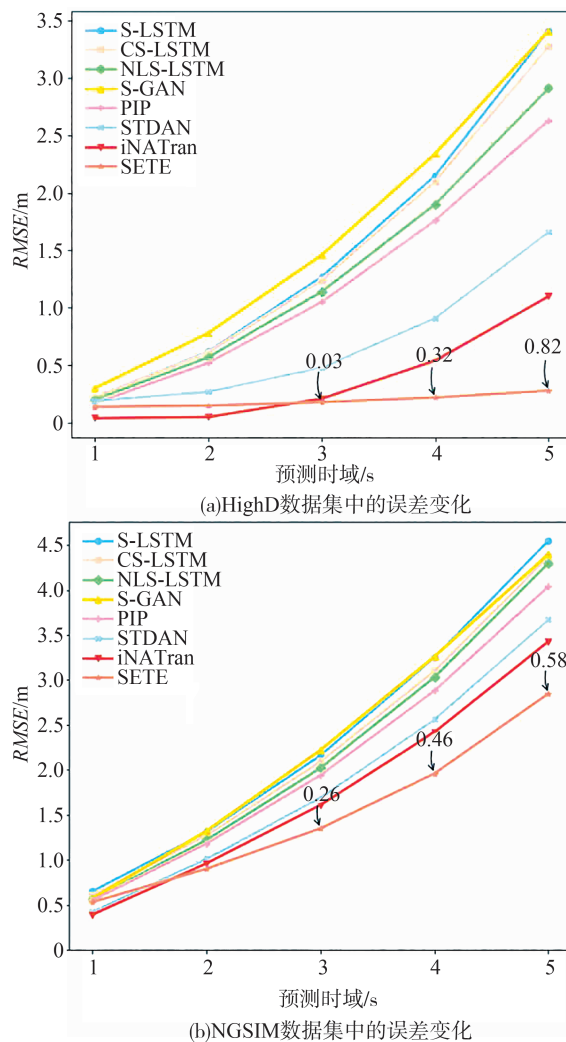


图 5 不同模型在公开数据集预测误差随时间变化情况

NGSIM 数据集上分别降低了 16% (0.26)、19% (0.46)、17% (0.58),表明该模型在长时轨迹预测中具有较好的改善效果。此外,随着预测时间的增加,其他模型的预测误差也随之增大,而所提模型在长

表 2 不同模型在 HighD 和 NGSIM 数据集上的 RMSE 对比

数据集	预测时长/s	S-LSTM	CS-LSTM	NLS-LSTM	S-GAN	PIP	STDAN	iNATran	STEI
HighD	1	0.22	0.22	0.2	0.3	0.17	0.19	<b>0.04</b>	0.14
	2	0.62	0.61	0.57	0.78	0.52	0.27	<b>0.05</b>	0.15
	3	1.27	1.24	1.14	1.46	1.05	0.48	0.21	<b>0.18</b> ↓
	4	2.15	2.1	1.9	2.34	1.76	0.91	0.54	<b>0.22</b> ↓
	5	3.41	3.27	2.91	3.41	2.63	1.66	1.10	<b>0.28</b> ↓
NGSIM	1	0.65	0.61	0.56	0.57	0.55	0.42	<b>0.39</b>	0.53
	2	1.31	1.27	1.22	1.32	1.18	1.01	0.96	0.90
	3	2.16	2.09	2.02	2.22	1.94	1.69	1.61	<b>1.35</b> ↓
	4	3.25	3.10	3.03	3.26	2.88	2.56	2.42	<b>1.96</b> ↓
	5	4.55	4.37	4.30	4.40	4.04	3.67	3.43	<b>2.85</b> ↓

时预测(3~5 s)误差增长较慢。造成这一现象的原因可能有:首先,所提模型融合了多种特征,为模型提供丰富的先验知识,从而提升整体性能;其次,以往研究在结合意图的轨迹预测<sup>[39-40]</sup>中通常忽略了意图分类准确性对结果的影响,而该模型采用多任务学习策略和合理的权重设置,进一步增强了性能。

同时,通过式(28)将测试集中每个样本的误差进行计算得到测试集中所有样本的误差分布。

$$RMSE_{\text{sing}} = \sqrt{\frac{1}{F} \sum_{t'=T+1}^F \left\{ (\mu_{t',x} - x_{t'})^2 + (\mu_{t',y} - y_{t'})^2 \right\}} \quad (28)$$

如图6所示,随着预测时域的增加,总体的误差呈非线性持续上升,预测的过程是递归的,后1 s的预测依赖于前1 s的结果,越接近当前时刻的预测值越精确,从而会导致预测误差的传播,造成预测结果的不确定性。

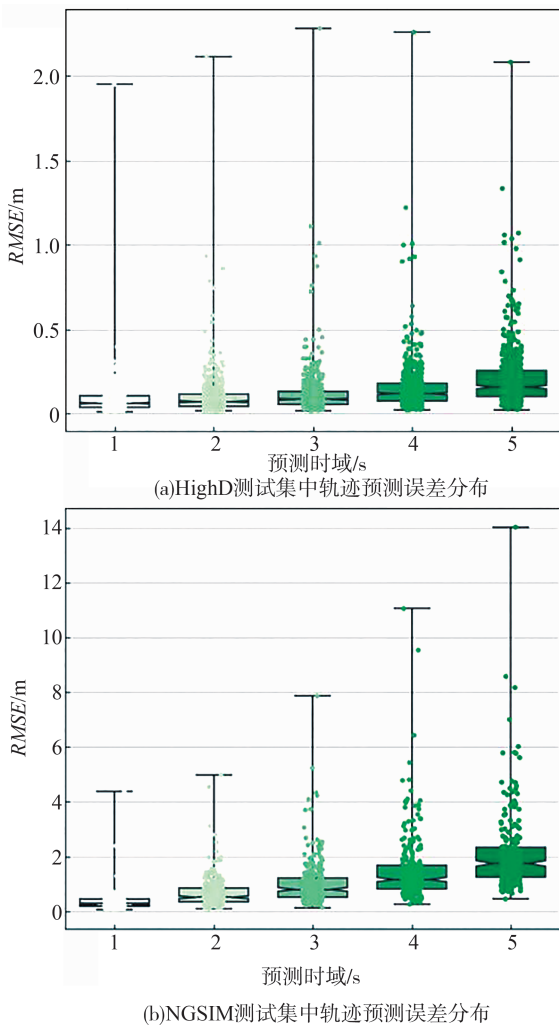


图6 本文模型在测试集中的轨迹预测误差分布

### 3.4 模型复杂度对比

为更全面评估稀疏注意力模块在模型性能中的影响,在试验中分别使用稀疏注意力和多头自注意力机制提取车辆的时空特征进行对比。对于稀疏注意力机制,其时间复杂度和空间复杂度均为  $O(L \log L)$ ,当  $L$  增加时,算法的执行时间和额外存储需求都会以  $L \log L$  的速度增长,而多头注意力机制的时间复杂度和空间复杂度为  $O(L^2)$ ,增长速度为  $L^2$ ,其执行时间和存储需求增长得更快。根据表3中数据发现,在采用稀疏注意力机制提取特征时,对比于使用多头注意力Flops减少了约0.21 M,Params减小了0.12 M,这对于减小模型复杂度、提升模型计算效率有较好作用。

表3 模型复杂度对比

模型	Time	Memory	Flops/M	Params/M
稀疏注意力	$O(L \log L)$	$O(L \log L)$	3.32	0.53
多头注意力	$O(L^2)$	$O(L^2)$	3.53	0.65

### 3.5 消融试验

为了全面评估各模块在模型中的作用,本文通过消融试验来对比各模块对模型预测性能的影响,结果见表4。试验结果显示,各模块对车辆长期预测性能均产生一定影响,导致预测精度下降,如图7所示。特别值得注意的是,移除SE模块后,模型在1~5 s的预测性能显著下降,尤其在后3 s表现得尤为明显。这表明在长期预测任务中,充分考虑周围车辆与目标车辆的空间关系对模型性能至关重要。此外,在移除TE和SE模块后,模型的推理时间明显减少,表明在模型运行过程中,时间和空间交互模块可能对计算资源造成较大负担。

表4 移除模型各模块后RMSE和推理时间对比

模块	1 s	2 s	3 s	4 s	5 s	推理时间/ms
移除IA	0.21	0.22	0.24	0.28	0.36	0.18
移除PM	0.18	0.20	0.24	0.28	0.35	0.19
移除TE	0.20	0.21	0.23	0.27	0.33	0.15
移除SE	0.28	0.30	0.31	0.36	0.44	0.16
STEI	0.14	0.15	0.18	0.22	0.28	0.21

### 3.6 预测轨迹可视化

为对不同场景下模型性能进行定性分析,图8可视化了几种典型的预测场景,包括车道保持、向左变道和向右变道。图中以目标车辆为中心, $x$ 轴表示车辆行驶方向,采样时间为0.2 s,因此目标车辆在

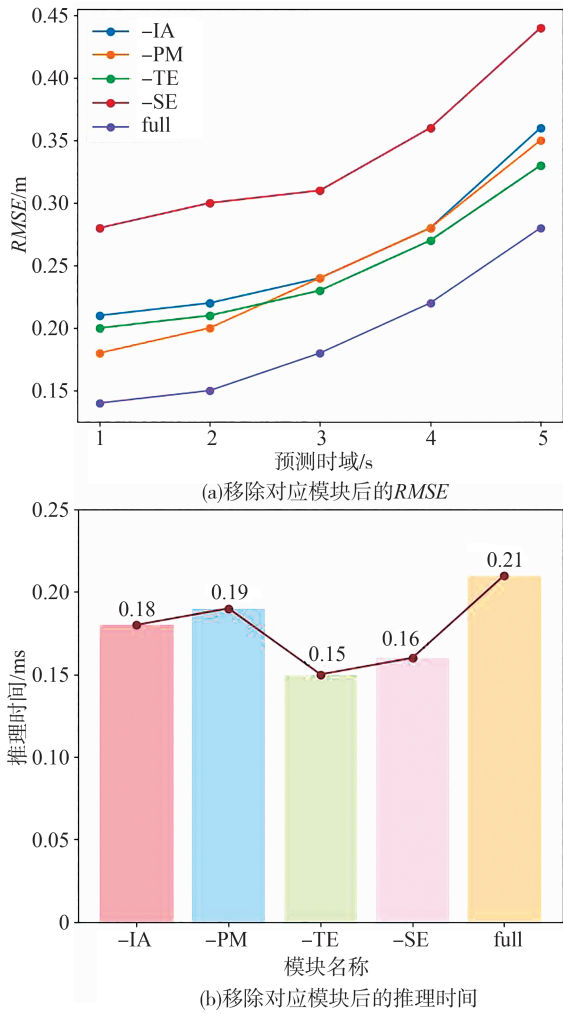


图7 消融试验影响分析

未来5 s内有25个轨迹分布点,轨迹长度越长表示车辆速度越大。可以看出,在1~5 s的预测时间范围内,预测轨迹与真实轨迹几乎一致,仅有少量偏差。然而,预测轨迹末端逐渐偏离真实轨迹,可能是因为目标车辆速度越快,未来轨迹走向的不确定性增加。总体而言,所提模型能够成功预测出合理的未来轨迹。

### 3.7 实车试验

本次试验采用两辆行深智能小车,模拟左换道、右换道和不换道3种典型驾驶场景,如图9所示。试验设置一辆小车作为目标车辆,另一辆作为周围车辆。目标车辆在左换道场景中通过遥控控制顺利并入相应车道,右换道和不变道场景遵循相似流程。行驶数据由智能硬件控制器、32线激光雷达、360°超声波雷达(探测距离5 cm至5 m)和4台以上摄像机组成的视觉系统获取。为进行轨迹预测,目标车辆在3种场景下的连续行驶数据被切分为8 s的样本

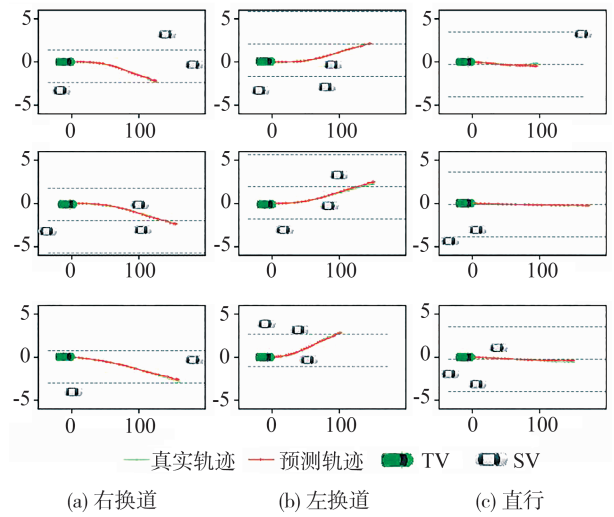


图8 模型预测的目标车辆轨迹结果可视化



图9 试验场景设置

数据,用于相应场景的分析。数据经过预处理后,输入到训练后的STEI模型,以生成未来5 s内的轨迹和行为预测概率。如表5所示,STEI模型在3种驾驶场景下的行为意图预测表现出极高的准确性。在右换道、左换道和不变道场景中,模型的横向意图预

表5 实车试验结果

场景	时间/s	轨迹预测		意图预测	
		Lateral	Longitudinal	Lateral	Longitudinal
右换道	1 s	0.16	3.89	RLC 预测值	CON 预测值
	2 s	0.18	4.44	99.93%	100%
	3 s	0.20	5.01	真实标签	真实标签
	4 s	0.23	5.59	RLC	CON
	5 s	0.27	6.16	√	√
左换道	1 s	0.031	4.66	LLC 预测值	DEC 预测值
	2 s	0.040	5.18	99.98%	55.43%
	3 s	0.053	5.72	真实标签	真实标签
	4 s	0.069	6.30	LLC	DEC
	5 s	0.084	6.83	√	√
不变道	1 s	0.075	4.54	LK 预测值	CON 预测值
	2 s	0.070	5.01	98.69%	99.99%
	3 s	0.061	5.47	真实标签	真实标签
	4 s	0.067	5.97	LK	CON
	5 s	0.056	6.47	√	√

测准确率接近 100%,能够准确识别目标车辆的未来行为。对于轨迹预测,横向误差随着预测时域的增加有所增大,但整体误差保持在较低水平;纵向误差则显著增加,主要原因可能是传感器精度限制导致数据噪声增加。此外,外界环境干扰和数据预处理过程中的不确定性也可能对纵向预测的准确性产生负面影响。

## 4 结论

(1) 研究创新性地采用稀疏注意力机制提取目标车辆和周围车辆的运动状态及历史轨迹特征,深入理解车辆行为。不仅挖掘了目标车辆的时间依赖性,准确反映了其与周围车辆的博弈关系,还通过关键特征的选择降低了模型复杂度,提高了资源利用效率。

(2) 在模型设计中,引入了位置掩码矩阵和意图行为分析模块。位置掩码矩阵通过网格划分和状态描述,有效捕捉周围车辆的存在状态,增强了对交通场景的理解。意图行为分析模块则深度分析目标车辆的意图,为模型提供更高层次的语境理解,从而提高预测的准确性和鲁棒性。试验结果表明,与现有研究相比,提出的模型在 3~5 s 的长期预测中表现良好。

(3) 研究没有考虑交通场景下道路的拓扑结构对于轨迹预测的影响,今后的研究须考虑如何更好地集成道路信息,并提高模型的鲁棒性。

### 参考文献

- [1] GOLIS A, FAR B H, FAPOJUWO A O. Vehicle trajectory prediction with Gaussian process regression in connected vehicle environment[C]. 2018 IEEE Intelligent Vehicles Symposium (IV). New York: IEEE, 2018: 550-555.
- [2] HUBMANN C, SCHULZ J, BECKER M, et al. Automated driving in uncertain environments: planning with interaction and uncertain maneuver prediction[J]. IEEE Transactions on Intelligent Vehicles, 2018, 3(1): 5-17.
- [3] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: human trajectory prediction in crowded spaces[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 961-971.
- [4] 季学武,费聪,何祥坤,等. 基于 LSTM 网络的驾驶意图识别及车辆轨迹预测[J]. 中国公路学报,2019,32(6): 34-42.
- [5] JI X W, FEI C, HE X K, et al. Intention recognition and trajectory prediction for vehicles using LSTM network[J]. China Journal of Highway and Transport, 2019, 32(6): 34-42.
- [6] FEI C, HE X K, JI X. Multi-modal vehicle trajectory prediction based on mutual information[J]. IET Intelligent Transport Systems, 2020, 14(3): 148-153.
- [7] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: human trajectory prediction in crowded spaces[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 961-971.
- [8] XIN L, WANG P, CHAN C, et al. Intention-aware long horizon trajectory prediction of surrounding vehicles using dual LSTM networks[C]. 21st International Conference on Intelligent Transportation Systems. New York: IEEE, 2018: 1441-1446.
- [9] HOU L, XIN L, LE S E, et al. Interactive trajectory prediction of surrounding road users for autonomous driving using structural LSTM network[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(11): 4615-4625.
- [10] KIM B, KANG C M, KIM J, et al. Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network[C]. IEEE 20th International Conference on Intelligent Transportation Systems. New York: IEEE, 2017: 399-404.
- [11] WU Y, CHEN G, LI Z, et al. HSTA: a hierarchical spatio-temporal attention model for trajectory prediction[J]. IEEE Transactions on Vehicular Technology, 2021, 70(11): 11295-11307.
- [12] MO X Y, XING Y, LV C. Interaction-aware trajectory prediction of connected vehicles using CNN-LSTM networks[C]. 46th Annual Conference of the IEEE-Industrial Electronics Society (IECON). New York: IEEE, 2020: 5057-5062.
- [13] XIE G, SHANGGUAN A, FEI R, et al. Motion trajectory prediction based on a CNN-LSTM sequential model[J]. Science China Information Sciences, 2020, 63(11): 1-21.
- [14] MUKHERJEE S, WANG S, WEALLACE A. Interacting vehicle trajectory prediction with convolutional recurrent neural networks[C]. IEEE International Conference on Robotics and Automation. New York: IEEE, 2020: 4336-4342.
- [15] 高凯,李勋豪,胡林,等. 基于多头注意力的 CNN-LSTM 的换道意图预测[J]. 机械工程学报,2022,58(22):369-378.
- [16] GAO K, LI X H, HU L, et al. Lane change intention prediction of CNN-LSTM based on multi-head attention[J]. Journal of Mechanical Engineering, 2022, 58(22): 369-378.
- [17] 金立生,高铭,郭柏苍,等. 基于时空融合 LSTM 网络的驾驶视角轨迹预测[J]. 中国公路学报,2022,35(4):325-332.
- [18] JIN L S, GAO M, GUO B C, et al. Driver perspective trajectory prediction based on spatiotemporal fusion LSTM network[J]. China Journal of Highway and Transport, 2022, 35(4): 325-332.
- [19] 曹昊天,施惠杰,宋晓琳,等. 基于多特征融合的行人意图以及行人轨迹预测方法研究[J]. 中国公路学报,2022,35(10): 308-318.
- [20] CAO H T, SHI H J, SONG X L, et al. Prediction of pedestrian intention and trajectory based on multi-feature fusion[J]. China Journal of Highway and Transport, 2022, 35(10): 308-318.
- [21] LI C, LIU Z W, ZHANG J Y, et al. Two-stream LSTM network with hybrid attention for vehicle trajectory prediction[C]. 2022 IEEE 25th International Conference on Intelligent Transportation

- Systems (ITSC). New York: IEEE, 2022: 1927–1934.
- [18] FU M, ZHANG T, SONG W, et al. Trajectory prediction-based local spatio-temporal navigation map for autonomous driving in dynamic highway environments [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 23(7): 6418–6429.
- [19] CAI Y, WANG Z, WANG H, et al. Environment-attention network for vehicle trajectory prediction [J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(11): 11216–11227.
- [20] BECKER S, HUG R, HUBNER W, et al. An evaluation of trajectory prediction approaches and notes on the trajnet benchmark [J/OL]. *Computer Vision and Pattern Recognition*, 2022. <https://arxiv.org/abs/1805.07663>.
- [21] GIULIARI F, HASAN I, CRISTANI M, et al. Transformer networks for trajectory forecasting [C]. 2020 25th International Conference on Pattern Recognition (ICPR). New York: IEEE, 2021: 10335–10342.
- [22] HOU L, LI S E, YANG B, et al. Structural transformer improves speed-accuracy trade-off in interactive trajectory prediction of multiple surrounding vehicles [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(12): 24778–24790.
- [23] 刘古文,李文倩,林杉,等. 基于稀疏权重共享的多模态轨迹预测 [J]. *中国公路学报*, 2023, 36(9): 244–256.
- LIU Z W, LI W Q, LIN S, et al. Multimodal trajectory prediction based on sparse weight sharing [J]. *China Journal of Highway and Transport*, 2023, 36(9): 244–256.
- [24] GOLI S A, FAR B H, FAPOJUWO A O. Vehicle trajectory prediction with Gaussian process regression in connected vehicle environment [C]. 2018 IEEE Intelligent Vehicles Symposium (IV). New York: IEEE, 2018: 550–555.
- [25] 李文礼,韩迪,石晓辉,等. 基于时-空注意力机制的车辆轨迹预测 [J]. *中国公路学报*, 2023, 36(1): 226–239.
- LI W L, HAN D, SHI X H, et al. Vehicle trajectory prediction based on spatial-temporal attention mechanism [J]. *China Journal of Highway and Transport*, 2023, 36(1): 226–239.
- [26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. *Advances in Neural Information Processing Systems (NIPS)*. New York: NIPS, 2017: 5998–6008.
- [27] ZHOU H, ZHANG S, PENG J, et al. Informer: beyond efficient transformer for long sequence time-series forecasting [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(12): 11106–11115.
- [28] ZHAO M, ZHONG S, FU X, et al. Deep residual networks with adaptively parametric rectifier linear units for fault diagnosis [J]. *IEEE Transactions on Industrial Electronics*, 2020, 68(3): 2587–2597.
- [29] KRAJEWSKI R, BOCK J, KLOEKER L, et al. The hight dataset: a drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems [C]. 2018 21st International Conference on Intelligent Transportation Systems (ITSC). New York: IEEE, 2018: 2118–2125.
- [30] COLYAR J, HALKIAS J. Us highway 101 dataset [R]. Federal Highway Administration (FHWA), Tech. Rep. FHWA-HRT-07-030, 2007: 27–69.
- [31] ABADI M, AGARWAL A, BARHAM P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems [J]. *arXiv preprint arXiv:1603.04467*, 2016.
- [32] LI Z, WANG Y, ZUO Z. Interaction-aware prediction for cut-in trajectories with limited observable neighboring vehicles [J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(3): 2148–2161.
- [33] SONG H, DING W, CHEN Y, et al. Pip: planning-informed trajectory prediction for autonomous driving [C]. *Computer Vision-ECCV 2020: 16th European Conference*. Berlin: Springer International Publishing, 2020: 598–614.
- [34] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: human trajectory prediction in crowded spaces [C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 961–971.
- [35] DEO N, TRIVEDI M M. Convolutional social pooling for vehicle trajectory prediction [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. New York: IEEE, 2018: 1468–1476.
- [36] MESSAOUD K, YAHIAOUI I, VERROUST-BLONDET A, et al. Non-local social pooling for vehicle trajectory prediction [C]. 2019 IEEE Intelligent Vehicles Symposium (IV). New York: IEEE, 2019: 975–980.
- [37] GUPTA A, JOHNSON J, FEI-FEI L, et al. Social GAN: socially acceptable trajectories with generative adversarial networks [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2018: 2255–2264.
- [38] CHEN X, ZHANG H, ZHAO F, et al. Intention-aware vehicle trajectory prediction based on spatial-temporal dynamic attention network for internet of vehicles [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(10): 19471–19483.
- [39] CHEN X, ZHANG H, ZHAO F, et al. Vehicle trajectory prediction based on intention-aware non-autoregressive transformer with multi-attention learning for internet of vehicles [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1–12.
- [40] DEO N, WOLFF E, BEIJBOM O. Multimodal trajectory prediction conditioned on lane-graph traversals [C]. *Conference on Robot Learning*. New York: PMLR, 2022: 203–212.