

DOI: 10.19650/j.cnki.cjsi.J2514160

# 基于分层特征融合和终点诱导的车辆多模态轨迹预测\*

陈熙源, 聂姝涵, 刘炜焱, 经纬铭

(东南大学仪器科学与工程学院 南京 210096)

**摘要:** 车辆多模态轨迹预测作为感知和决策规划之间的桥梁, 在自动驾驶系统中发挥着重要作用。针对现有方法特征融合不充分和难以有效平衡预测精度和效率的问题, 提出了一种基于分层特征融合和终点诱导的车辆多模态轨迹预测模型 (HFF-EI)。首先, 利用一维残差卷积和特征金字塔网络 (FPN) 编码车辆历史轨迹信息, 充分提取特征; 其次, 构建了一种分层特征融合结构, 对车辆和地图分别进行特征融合后再进行全局特征融合, 实现场景各元素特征高效且全面融合; 然后, 引入基于动态权重模式的多层感知机 (MLP) 进行轨迹终点预测, 提高模型在不同交通场景下的自适应能力; 最后, 提出了一种基于终点信息交互的终点细化模块, 使用注意力机制在更长时空范围交互轨迹信息, 提高了车辆多模态轨迹预测的准确性。在公开数据集 Argoverse1 进行消融实验和对比实验, 消融实验结果表明: HFF-EI 模型的 3 个模块均有效提升了轨迹预测性能, 在最小平均位移误差、最小最终位移误差、丢失率和具有惩罚项的最小最终位移误差上分别降低 8.87%、13.52%、31.07% 和 8.93%; HFF-EI 模型在测试集上最小最终位移误差为 1.134 m, 具有惩罚项的最小最终位移误差为 1.773 m, 推理时间为 10.22 ms, 与 10 个基准模型相比综合性能优势显著, 证明了所提模型的有效性。

**关键词:** 轨迹预测; 注意力机制; 终点预测; 特征融合

**中图分类号:** TH89 **文献标识码:** A **国家标准学科分类代码:** 520.20 510.80

## Multi-modal vehicle trajectory prediction based on hierarchical feature fusion and endpoint induction

Chen Xiyan, Nie Shuhan, Liu Weiyan, Jing Weiming

(School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China)

**Abstract:** Multi-modal vehicle trajectory prediction, as a bridge between perception and decision planning, plays an important role in autonomous driving systems. Aiming at the problems of insufficient feature fusion and difficulty in balancing the prediction accuracy and efficiency of existing methods, a vehicle multimodal trajectory prediction model based on Hierarchical Feature Fusion and End-point Induction (HFF-EI) is proposed. Firstly, One-dimensional residual convolution and a feature pyramid network (FPN) are used to encode the vehicle historical trajectory information, thereby fully extracting the relevant features. Then a hierarchical feature fusion structure is constructed, and local feature fusion is carried out for the vehicle and the map, followed by global feature fusion, achieving efficient and comprehensive fusion of scene features across all elements. Secondly, a multi-layer perceptron (MLP) based on the dynamic weight model is introduced for trajectory endpoint prediction, enhancing the adaptive ability of the model under different traffic scenes. Finally, an endpoint refinement module based on endpoint information interaction is proposed, which uses the attention mechanism to interact trajectory information in a longer spatial and temporal ranges. Ablation and comparative experiments were conducted on the public dataset Argoverse1. Results of the ablation experiments show that the three modules of the HFF-EI model effectively improve the performance of trajectory prediction, and reduce the minimum average displacement error, minimum final displacement error, loss rate and minimum final displacement error with penalty by 8.87%, 13.52%, 31.07%, and 8.93%, respectively. On the test set, the minimum final displacement error is 1.134 m, the minimum final displacement error with penalty term is 1.773 m and the inference time is 10.22 ms, which proves the effectiveness of the proposed model by its comprehensive performance

收稿日期: 2025-06-15 Received Date: 2025-06-15

\* 基金项目: 国家自然科学基金 (61873064)、江苏省重点研发计划 (BE2022139) 项目资助

advantages compared with the 10 benchmark models.

**Keywords:** trajectory prediction; attention mechanism; endpoint prediction; feature fusion

## 0 引 言

随着自动驾驶技术的发展,轨迹预测作为感知和决策规划之间的桥梁,其准确性和高效性对自动驾驶系统的安全至关重要。轨迹预测任务是利用感知到的周围环境和车辆信息来预测交通参与者在未来一段时间内的轨迹和行驶意图<sup>[1]</sup>。在复杂的交通场景中,驾驶员的行驶意图是难以确定的,其他车辆行驶路径的即时变化会导致自车轨迹呈现多模态属性<sup>[2]</sup>。因此,实现准确且高效的多模态轨迹预测对自动驾驶技术具有重大意义。

现有的车辆轨迹预测方法主要可以分为三大类:基于物理模型的预测方法、基于机器学习的预测方法以及基于深度学习的预测方法。基于物理模型的车辆轨迹预测是利用车辆动力学模型推算车辆未来的状态,这种方法虽然模型简单,但忽略了道路结构和环境信息等先验知识的影响,仅适用于短时域内的轨迹预测<sup>[3]</sup>。基于机器学习的方法首先判断驾驶行为,再结合行为对未来运动轨迹预测,忽略了驾驶场景中其他车辆对自车的影响。基于深度学习的轨迹预测方法可以对场景中所有智能体和道路结构进行交互,增强对场景信息的理解,使得长时域的预测更准确。因此近年来,深度学习模型成为解决轨迹预测问题的一种有效方法<sup>[4]</sup>。

在场景特征提取和交互方面,文献[5]提出将车道信息和车辆轨迹信息矢量化后,利用图神经网络(graph neural network, GNN)对特征进行交互,但在十字路口等复杂的道路场景中预测精度降低;文献[6]提出使用图卷积神经网络(graph convolutional network, GCN)来捕获车道图的复杂拓扑结构和长时依赖关系,但在面对大型图多个节点的时候效率低下;文献[7]采用分层Transformer对场景特征进行编码,来学习复杂的时空交互,减少了模型参数,但在预测场景中所有车辆的轨迹时的计算复杂度和车辆数量呈线性关系;文献[8]提出利用紧凑且高效的折线来描述道路特征,基于动力学控制和连续时间多项式对轨迹特征进行建模,但会预测出不符合驾驶舒适度的轨迹;文献[9]提出双层交互模型实现更好的场景元素动态交互;文献[10]提出了一种融合多特征的车辆轨迹预测模型,实现了不同场景下的泛化性。

在多模态轨迹预测方面,文献[11]提出通过基于场景采样选择目标终点,补全轨迹再对轨迹进行评分和筛选,但预测性能在很大程度上依赖于对目标终点的准确预测;文献[12]提出了基于无锚点的预测模型,直接从密集的目标候选集中输出一组轨迹,但在编码阶段忽略

了车辆运动状态的信息;文献[13]提出利用堆叠式Transformer架构来整合上下文信息的多个通道用于多模运动行为预测,有效地减轻了运动预测的复杂性;文献[14]提出以查询为中心的方法,重复使用窗口之间重叠的特征输出多模态预测轨迹,虽然提高了预测精度,但模型参数量大;文献[15]提出一种自适应头来适应不同场景多模态轨迹预测,降低了模型的复杂度,但模型的泛化能力不够;方鑫等<sup>[16]</sup>提出了一种基于Transformer轨迹预测模型,实现端到端轨迹预测;文献[17]引入小波变换进行目标的轨迹预测,实现大机动目标的准确预测;卢一凡等<sup>[18]</sup>引入不变风险最小化生成轨迹终点,提高轨迹预测的泛化性。周正阳等<sup>[19]</sup>设计了一种改进樽海鞘优化算法对纯追踪控制中的前视距离进行实时调整优化,提高轨迹控制准确性。陈熙源等<sup>[20]</sup>提出了变分贝叶斯的自适应因子图优化算法为轨迹预测提供准确定位信息。

上述的轨迹预测模型利用深度学习有效实现了多模态轨迹预测,但仍然存在不足,具体体现在:1)在复杂高动态交通场景中,地图和车辆轨迹的特征交互没有聚焦于对目标车辆未来轨迹更重要的特征,效率较低;2)在终点预测时,没有针对不同交通场景的特征差异进行调整;3)在多模态轨迹预测时,没有充分考虑场景中所有车辆终点信息的交互;4)现有的模型大多数在提升准确性的同时忽略了效率问题,难以有效平衡轨迹预测的精度和效率。针对以上问题,本文提出了基于分层特征融合和终点诱导的车辆多模态轨迹预测(hierarchical feature fusion and end-point induction, HFF-EI)模型,具体工作为:1)实现地图和轨迹特征的高效融合:采用分层特征融合,局部特征融合后进行全局特征融合,使得场景各元素特征全面交互;2)提升终点预测的鲁棒性和准确性:引入对不同场景具有自适应能力的终点预测模块;3)实现轨迹终点信息交互:采用注意力机制对预测的轨迹终点特征进行交互,利用更长时空范围内的信息细化终点,使得预测的车辆未来轨迹合理且准确;4)平衡轨迹预测的精度和效率:使用Transformer结构和基于动态权重模式的多层感知机(multilayer perceptron, MLP)在保证轨迹预测精度的同时提高预测效率,降低模型复杂度。

## 1 模型概述

本文提出一种基于分层特征融合和终点诱导的车辆多模态轨迹预测模型HFF-EI,模型总体框架如图1所示,其中主要包括3个主要模块:特征提取模块、分层特征融合模块和轨迹预测模块。

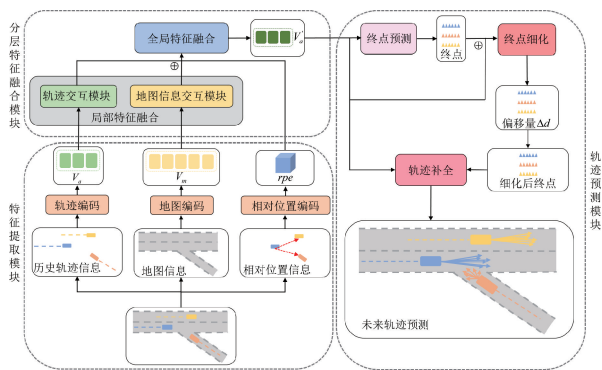


图1 模型整体框架

Fig. 1 Overall framework of the model

在特征提取模块中,利用轨迹编码器、地图编码器和相对位置编码器分别对车辆历史轨迹信息、静态地图信息和相对位置信息分别编码,将其转换成高维特征向量;在分层特征融合模块中,使用注意力机制和 Transformer 架构分别实现局部特征融合和全局特征融合,增强车辆在复杂驾驶环境的交互;在轨迹预测模块中,首先利用动态权重模块预测轨迹终点,然后引入注意力机制对融合终点信息的轨迹特征进行交互,实现终点细化,最后基于细化后的终点信息进行轨迹补全,从而实现车辆的多模态轨迹预测。

### 1.1 问题表述

由于在轨迹预测过程中需要考虑不确定性因素,车辆未来的轨迹存在多种可能,预测模型就是要对未来行为进行合理的描述,并给出合理的概率分布。所以多模态轨迹预测是根据车辆状态和驾驶环境来预测未来一段时间内的车辆多条可能的轨迹,并给出每条轨迹相应的概率。其中,多模态表示输出的轨迹条数,在本文中选取模态数为6。模型的输入为静态地图信息、所预测车辆以及周围车辆的历史轨迹信息和车辆之间的位置关系。其中地图在观测时间内是保持静态的,地图信息定义为  $M = \{M_0, M_1, \dots, M_{N_t}\}$ , 其中  $M_0$  为场景中的车道中心线,  $N_t$  表示观测时间内地图中车道中心线的数量。车辆的历史轨迹信息是观测时间内场景中所有车辆的轨迹信息,可以定义为  $X = \{X_0, X_1, \dots, X_{N_a}\}$ ,  $N_a$  表示观测时间内的车辆数量,其中  $X_i = \{X_{i,-H+1}, \dots, X_{i,0}\}$  表示第  $i$  个车辆在过去  $H$  时间步长上的历史轨迹信息。车辆间的位置关系可以通过航向角差  $\alpha_{i-j}$ , 相对方位角  $\beta_{i-j}$  和距离  $\|d_{i-j}\|$  这3个量<sup>[21]</sup>来描述。模型的输出为所有车辆的在预测时间步长的未来轨迹及其概率,其中所有车辆的未来轨迹定义为  $Y = \{Y_0, Y_1, \dots, Y_{N_a}\}$ , 第  $i$  个车辆的未来  $K$  条轨迹定义为  $Y_i = \{Y_i^1, \dots, Y_i^K\}$ , 其中  $Y_i^K = \{y_{i,1}^K, \dots, y_{i,T}^K\}$  表示第  $i$  个车辆在未来时间步长  $T$  中第  $K$  条轨迹分布,第

$i$  个车辆的  $K$  条预测轨迹的概率定义为  $\alpha_i = \{\alpha_i^1, \dots, \alpha_i^K\}$ 。因此,多模态轨迹预测问题可以表述为估计一个混合概率分布如式(1)所示。

$$P(Y_i | X, M) = \sum_{k=1}^K \alpha_i^k P(Y_i^k | X, M) \quad (1)$$

### 1.2 特征提取模块

#### 1) 车辆轨迹编码

因为车辆轨迹信息和地图信息是向量化形式,所以需要对这些信息进行编码,获得高维的特征张量。由于鸟瞰图(bird's eye view, BEV)可以提供更全面的交通场景信息,有助于车辆感知周围环境,所以场景中所有车辆的坐标都定义在 BEV 中。为了降低特征提取的复杂度,采用以场景为中心的表示方法,即特征提取基于自身时空坐标系,不依赖于全局时空坐标系。车辆历史轨迹编码器如图2所示。

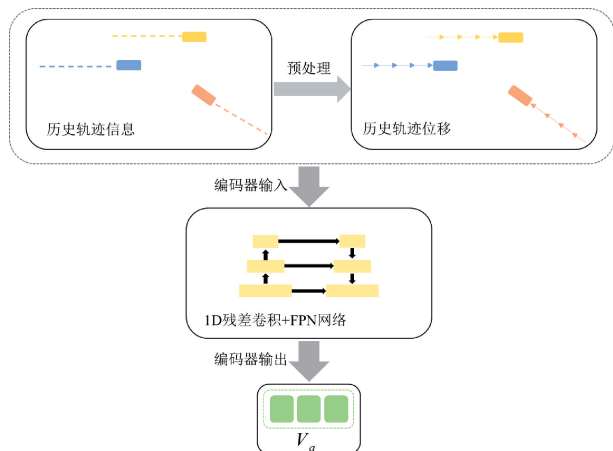


图2 车辆轨迹编码器

Fig. 2 Vehicle track encoder

由于轨迹信息是包含车辆位置的时序序列,所以为了实现模型在不同场景的泛化性和鲁棒性,车辆轨迹编码器输入均采用位移表示。首先对场景中所有车辆的历史轨迹数据进行预处理,将第  $i$  个车辆在  $H$  时间步长上历史轨迹信息表示为位移序列  $\Delta X_i = \{\Delta X_{i,-H+1}, \dots, \Delta X_{i,0}\}$ , 其中  $\Delta X_{i,t}$  表示第  $i$  个车辆从  $t-1$  时刻到  $t$  时刻的位移。为了提高多尺度特征提取的充分性和效率,使用一维残差卷积和特征金字塔网络(feature pyramid network, FPN)处理历史轨迹输入。其中一维残差卷积在处理时序数据方面优势明显,FPN 虽然小幅度增加了计算量,但是显著提升特征表达的尺度鲁棒性,使得特征图在尺度上更加丰富和全面。轨迹特征编码器中一维残差卷积网络中的所有卷积层的卷积核大小为3,输出通道数为128。因此,编码器输出的车辆特征张量  $V_a$  大小为  $[N_a, 128]$ , 其中  $N_a$  表示历史步长内的车辆数量。



## 2) 地图信息编码

地图信息主要包括每个车道段的空间位置、方向和长度。在地图编码中,用车道中心线中点的位置和方向来表示每个车道段的位置和方向。考虑到在观测时间内车道段是静态的,所以通过简单的两层 MLP 对地图信息编码来提高运行效率。编码器输出的地图特征张量  $V_m$  大小为  $[N_l, 128]$ , 其中  $N_l$  表示观测时间内地图中车道中心线的数量。

## 3) 相对位置信息编码

在复杂的交通场景中,为了避免碰撞和预测出不合理的轨迹,各个元素之间的相对位置关系编码非常重要。在相对位置编码中,使用航向角差  $\alpha_{i \rightarrow j}$ , 相对方位角  $\beta_{i \rightarrow j}$  和距离  $\|d_{i \rightarrow j}\|$  这3个量来描述元素  $i$  和  $j$  之间的相对位置和姿态,如图3所示。为了增强数值稳定性,航向差和相对方位角使用正弦值和余弦值表示。

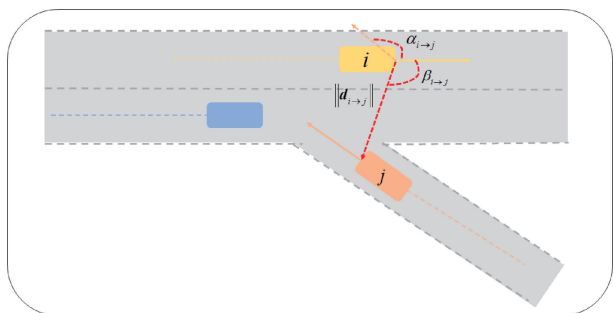


图3 相对位置编码

Fig. 3 Relative position encode

所以,场景元素  $i$  和元素  $j$  之间的相对位置信息  $r_{i \rightarrow j}$  可以由五维向量表示,如式(2)所示。

$$r_{i \rightarrow j} = \begin{bmatrix} \sin(\alpha_{i \rightarrow j}) \\ \cos(\alpha_{i \rightarrow j}) \\ \sin(\beta_{i \rightarrow j}) \\ \cos(\beta_{i \rightarrow j}) \\ \|d_{i \rightarrow j}\| \end{bmatrix} \quad (2)$$

对于在观测时间内具有  $N$  个元素的场景,各种元素之间的相对位置信息可以由一个大小为  $[N, N, 5]$  的数组表示。最后使用 MLP 对相对位置信息进行编码得到大小为  $[N, N, 128]$  的相对位置张量  $rpe$ 。

## 1.3 分层特征融合模块

复杂的交通场景中交通参与者的行为往往是动态且复杂的,为了充分且高效的融合模型特征,采用分层特征融合结构。首先对车辆和地图分别进行各自的局部特征融合,利用注意力机制实现给车辆和地图特征赋予不同权重,给对被预测车辆未来轨迹影响大的特征赋予高权重,反之赋予低权重;然后对场景中所有元素进行全局特征融合,全面交互各元素时空关系。

## 1) 局部特征融合

为了让模型聚焦于更重要的特征,给与车辆未来轨迹有关的特征分配更高的权重,在全局特征融合前先进行局部特征融合。局部特征融合主要包括轨迹特征融合和地图信息特征融合。轨迹特征融合通过交互车辆的位置和行驶方向等信息,增强了车辆对周围交通环境的感知能力,使得车辆能够高效且快速的预测并响应其他车辆的行为。地图特征融合通过交互场景的车道信息使得车辆能够更好地理解驾驶场景中道路信息,包括车道段位置、方向和长度。

在局部特征融合中,采用单头注意力机制分别实现轨迹特征融合和地图特征融合,既能提高轨迹预测精度又能提高计算效率,网络结构如图4所示。

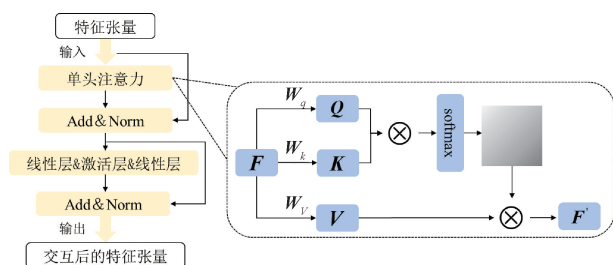


图4 局部特征融合网络结构

Fig. 4 Local feature fusion network architecture

将特征提取模块输出的轨迹特征张量和地图特征张量分别输入局部特征融合网络,经过单头自注意力机制给特征分配不同的权重,捕捉车辆轨迹中的关键点和细节变化,其中  $F$  表示输入的特征张量,  $F'$  表示经过注意力机制后的特征张量如式(3)和(4)所示:

$$\begin{cases} Q = W_q F \\ K = W_k F \\ V = W_v F \end{cases} \quad (3)$$

$$F' = \text{softmax}\left(\frac{QK^T}{\sqrt{\dim_k}}\right) V \quad (4)$$

式中:  $W_q$ 、 $W_k$ 、 $W_v$  是权重矩阵。

最后经过层归一化网络和前馈网络得到交互后的特征张量。局部特征融合丰富了车辆轨迹和地图特征,为全局特征融合奠定了基础。

## 2) 全局特征融合

局部特征融合增强了车辆和地图各自的特征,但是没有构建场景中各元素的相对位置关系之间的交互。Transformer 模型被广泛应用于计算机视觉和特征融合领域,它可以高效处理和融合不同尺度的特征,减少模型的参数量,有助于减少过拟合的风险并提高模型的泛化能力。为此,采用对称融合 Transformer (symmetric fusion Transformer, SFT)<sup>[22]</sup> 进行全局特征融合,该特征融合方



法以对称方式执行定向信息传递,使网络能够在一次前馈传递中预测所有道路使用者的未来运动,不仅能进行有效的交互还能提高模型效率。该网络结构借鉴了Transformer架构,包括多头注意力机制、前馈神经网络、连接层和层归一化。网络结构如图5所示。

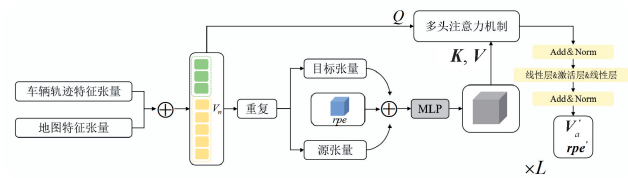


图5 全局特征融合网络结构

Fig. 5 Global feature fusion network architecture

首先对局部融合后的车辆轨迹特征张量和地图特征张量拼接得到张量  $V_n$ , 张量大小为  $[N, 128]$ , 其中  $N$  为场景中车辆和车道段之和; 其次对  $V_n$  沿着两个不同维度重复后堆叠, 得到目标张量和源张量, 大小均为  $[N, N, 128]$ ; 然后将目标张量、源张量和相对位置特征张量拼接后经过一个MLP得到多头注意力的输入  $K, V$ , 多头注意力的  $Q$  为张量  $V_n$ ; 最后经过多头注意力机制、连接层和层归一化以及前馈神经网络后输出更新后的车辆轨迹特征张量  $V'_a$  和相对位置特征张量  $rpe'$ , 其中注意力头数为8。张量  $V'_a$  大小为  $[N_a, 128]$ , 张量  $rpe'$  大小为  $[N, N, 128]$ , 其中  $N_a$  为场景中车辆数。将上述全局特征融合过程重复  $L$  次, 其中  $L$  为该网络结构层数。

全局特征融合模块有效的融合了场景所有交通参与者信息和道路结构信息, 提升了轨迹预测任务对动态交通环境的鲁棒性, 确保了轨迹预测的准确性和安全性。

#### 1.4 轨迹预测模块

为了抑制预测轨迹的发散, 提高预测轨迹的合理性, 本文模型采用具有终点诱导的多模态轨迹预测解码器。受到自适应轨迹预测 (efficient multi-agent trajectory prediction with adaptation, ADAPT) 方法的启发, 该模块将轨迹预测任务分成3个小任务: 首先利用基于动态权重模式的MLP模块预测轨迹的终点, 确定多模态轨迹的大致分布; 再对融合终点信息的轨迹特征进行交互得到终点偏移量, 实现轨迹终点的细化; 最后补全轨迹。

##### 1) 终点预测

不同的交通场景特征各不相同, 车辆行驶轨迹和驾驶人行车意图是不确定的。为了适应这种差异, 采用动态权重MLP模块针对各种场景生成不同权重的特征从而实现准确的多模态轨迹的终点预测, 网络结构如图6所示。该网络架构利用两个动态权重参数来自适应不同场景之间的差异。

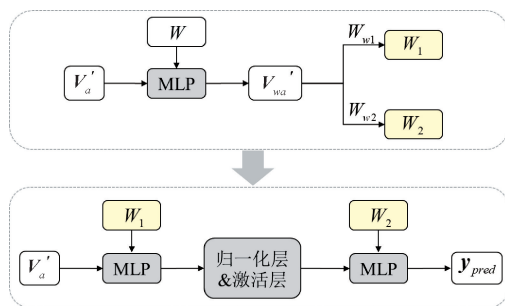


图6 终点预测网络结构

Fig. 6 Endpoint prediction network structure

首先对分层特征融合模块输出的车辆轨迹特征张量  $V'_a$  使用一个MLP层生成张量  $V'_wa$ , 然后引入两个MLP层中的可训练权重参数  $W_{w1}$  和  $W_{w2}$ , 使  $V'_wa$  分别经过这两个MLP层生成两个动态权重参数  $W_1$  和  $W_2$ , 如式(5)所示。

$$\begin{cases} W_1 = W_{w1} V'_{wa} \\ W_2 = W_{w2} V'_{wa} \end{cases} \quad (5)$$

最后将轨迹终点预测过程转化为先经过权重为  $W_1$  的MLP层, 再通过归一化层和激活层, 最后经过权重为  $W_2$  的MLP层生成多模态终点  $y_{pred}$ 。

##### 2) 终点细化及轨迹补全

分层特征融合模块只考虑了车辆历史轨迹的交互, 忽略了车辆终点意图的交互。自注意力机制可以高效捕捉复杂的依赖关系, 能够增强终点特征的表达力。为了使得预测出的轨迹分布更合理更准确, 引入注意力机制对包含终点信息的车辆轨迹特征进行交互预测, 利用预测出的不同运动意图辅助实现轨迹终点的细化, 网络结构如图7所示。

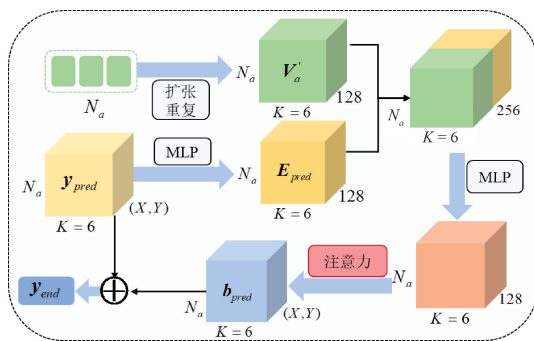


图7 终点细化网络结构

Fig. 7 Endpoint refinement network structure

首先对终点预测模块输出的轨迹终点进行编码得到轨迹终点特征张量  $E_{pred}$ , 然后将拼接后的车辆轨迹特征张量  $V'_a$  和轨迹终点特征张量  $E_{pred}$  输入注意力机制中, 输出轨迹终点偏移量  $b_{pred}$ 。具体计算流程如式(6)所示。

$$\begin{cases} \mathbf{E}_{pred} = MLP(\mathbf{y}_{pred}) \\ \mathbf{b}_{pred} = MHA(\text{concat}(\mathbf{V}'_a, \mathbf{E}_{pred})) \\ \mathbf{y}_{end} = \mathbf{y}_{pred} + \mathbf{b}_{pred} \end{cases} \quad (6)$$

式中:  $\mathbf{y}_{pred}$  为终点预测模块输出的多模态轨迹终点;  $\mathbf{E}_{pred}$  为轨迹终点特征;  $\mathbf{b}_{pred}$  为预测的终点偏移量;  $\mathbf{y}_{pred}$  为细化后的轨迹终点;  $MHA$  表示多头注意力机制;  $\text{concat}$  表示特征张量拼接。最后使用一个简单 MLP 补全剩余轨迹点并且预测各个模态的概率。

### 1.5 损失函数

本文模型通过先预测终点来实现轨迹预测任务, 所以损失函数包括 3 项: 全部轨迹的回归损失、分类损失和终点回归损失的加权和如式(7)所示。

$$L = aL_{reg} + bL_{cls} + cL_{end} \quad (7)$$

式中:  $L_{reg}$  是全部轨迹的回归损失;  $L_{cls}$  是分类损失;  $L_{end}$  是终点的回归损失。其中全部轨迹的回归损失能优化轨迹的整体形状和连续性, 使预测轨迹更接近真实轨迹; 分类损失能提高模型对车辆运动意图的识别准确性, 从而为轨迹预测提供更可靠的决策依据; 终点回归损失能优化轨迹的终点位置, 提高轨迹预测的准确性。  $a, b, c \in [0, 1]$  是平衡三项损失的权重。为了在训练过程中重点关注回归损失, 选取  $a = 0.8, b = 0.2, c = 1$ 。在处理多模态时使用赢者通吃(winner takes all, WTA)策略<sup>[23]</sup>, 该策略可以实现信息的选择性传递, 从而提高模型的效率和泛化性能。通过在多个模态中选择最终位移误差最小的模态作为最佳模态计算损失。在分类损失中, 使用交叉熵损失函数对概率得分进行优化如式(8)所示。

$$L_{cls} = L_{CE}(\bar{p} - p) \quad (8)$$

式中:  $\bar{p}$  为真实轨迹概率分布;  $p$  为预测轨迹的概率分布。

全部轨迹的回归损失包括车辆位置坐标回归损失  $L_{pos}$  和航向角损失  $L_{yaw}$ , 引入航向角损失使得预测轨迹具有更高的平滑性和可行性, 如式(9)所示。

$$\begin{cases} L_{pos} = \frac{1}{T} \sum_{t=1}^T \| \mathbf{y}_t^{K^*} - \bar{\mathbf{y}}_t \|_2 \\ L_{yaw} = [1 - \text{CosSim}(\mathbf{Y}_{yaw}^{K^*}, \bar{\mathbf{Y}}_{yaw})] / 2 \\ L_{reg} = L_{pos} + L_{yaw} \end{cases} \quad (9)$$

式中:  $T$  为预测时间步长;  $\mathbf{y}_t^{K^*}$  为最佳模态的轨迹在  $t$  时刻的位置坐标;  $\bar{\mathbf{y}}_t$  为真实轨迹在  $t$  时刻的位置坐标;  $\text{CosSim}$  为余弦相似度度量, 两个向量同向时取值为 1, 反向时取值为 -1;  $\mathbf{Y}_{yaw}^{K^*}$  为最佳模态时车辆的航向角;  $\bar{\mathbf{Y}}_{yaw}$  为车辆真实的航向角。

终点回归损失如式(10)所示。

$$L_{end} = \| \mathbf{y}_T^{K^*} - \bar{\mathbf{y}}_T \|_2 \quad (10)$$

式中:  $\mathbf{y}_T^{K^*}$  表示最佳模态的轨迹终点的位置坐标;  $\bar{\mathbf{y}}_T$  为真实轨迹终点的位置坐标。

## 2 实验分析

### 2.1 数据集及评价指标

为了验证模型的有效性, 使用 Argoverse1<sup>[24]</sup> 运动预测数据集进行实验。该数据集包含 323 557 个场景, 其中包含 205 942 个训练集样本, 39 472 个验证集样本和 78 143 个测试集样本, 表现均为 2 D 鸟瞰图, 每个场景时长 5 s, 采样率为 10 Hz, 数据集场景大多是在十字路口, 转弯和变道这种交通较为复杂的地方。数据集中的车辆包括自动驾驶车辆、被预测车辆和其他感兴趣对象。该数据集提供了两张高精地图, 覆盖了迈阿密和匹兹堡的特定区域, 总车道覆盖里程分别为 204 和 86 km。地图信息包括车道中心线、交通方向和地面高度。在该数据集上轨迹预测任务是根据前 2 s 的轨迹信息和地图信息预测出车辆后 3 s 的多模态轨迹。

本文轨迹预测模型中模态数为 6, 即  $K = 6$ 。采用 4 个指标对多模态轨迹预测任务进行评价, 分别是最小平均位移误差(minimum average displacement error,  $\text{minADE}_6$ )、最小最终位移误差(minimum final displacement error,  $\text{minFDE}_6$ )、丢失率(miss rate,  $\text{MR}_6$ )和具有惩罚项的最小最终位移误差(brier minimum final displacement error,  $\text{b-minFDE}_6$ )这 4 个指标。 $\text{minADE}_6$  表示真实轨迹和最佳模态轨迹的平均欧氏距离;  $\text{minFDE}_6$  表示真实轨迹终点和最佳模态轨迹终点的欧氏距离;  $\text{MR}_6$  表示预测轨迹中轨迹终点超出真实轨迹终点 2 m 之外的轨迹比例;  $\text{b-minFDE}_6$  表示结合了预测概率权重的最小最终位移误差。计算公式如式(11)所示。

$$\begin{cases} \text{minADE}_6 = \min_{j \in \{1, \dots, K\}} \frac{1}{T} \sum_{t=1}^T \| \mathbf{y}_t^j - \bar{\mathbf{y}}_t \|_2 \\ \text{minFDE}_6 = \min_{j \in \{1, \dots, K\}} \| \mathbf{y}_T^j - \bar{\mathbf{y}}_T \|_2 \\ \text{MR}_6 = \frac{\sum_{n=1}^N M}{N}, \quad M = \begin{cases} 0, & \min_{j \in \{1, \dots, K\}} \| \mathbf{y}_T^j - \bar{\mathbf{y}}_T \|_2 \leq 2 \\ 1, & \text{其他} \end{cases} \\ \text{b-minFDE}_6 = \min_{j \in \{1, \dots, K\}} ( \| \mathbf{y}_T^j - \bar{\mathbf{y}}_T \|_2 + (1 - p_j)^2 ) \end{cases} \quad (11)$$

式中:  $K$  为预测的模态数;  $\mathbf{y}_t^j$  为第  $j$  模态的轨迹在  $t$  时刻的位置坐标;  $\bar{\mathbf{y}}_t$  为真实轨迹在  $t$  时刻的位置坐标;  $\mathbf{y}_T^j$  表示第  $j$  模态的轨迹终点的位置坐标;  $\bar{\mathbf{y}}_T$  为真实轨迹终点的位置坐标;  $N$  为预测样本的总数;  $p_j$  为第  $j$  条预测轨迹的概率。

模型在 GeForce GTX 4090D 上进行了 50 个周期的

训练,训练的批次大小(batch size)为 16,使用 Adam 优化器进行优化。采用 warm-up 余弦退火动态调整学习率,预热周期为 5,初始学习率为  $1 \times 10^{-4}$ ,最小学习率为  $1 \times 10^{-5}$ 。前 25 个周期,每训练两个周期进行一次验证;后 25 个周期,每训练一个周期进行一次验证。训练损失和验证损失随周期变化图像如图 8 所示。由于在训练过程中增加了正则化等操作,所以在初期训练损失会高于验证损失。在 50 个周期训练过程中 4 项评价指标随周期变化曲线如图 9 所示。由训练和验证损失以及 4 项评价指标随周期变化曲线可知,训练 50 个周期可以确保模型充分学习特征,保证轨迹预测的精度,同时可以节省训练时间和资源。

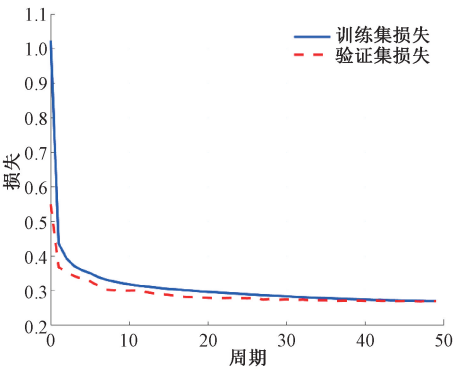


图 8 损失变化曲线  
Fig. 8 Loss variation curve

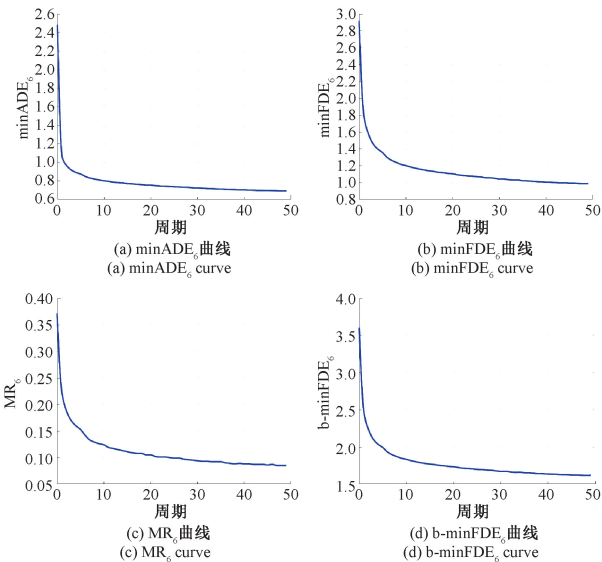


图 9 评价指标变化曲线

Fig. 9 Evaluation of indicator change curves

2.2 消融实验

为了验证 HFF-EI 模型中各个模块的有效性,采用消融实验评估分层特征融合模块、终点预测模块和终点细

化模块的有效性。并且对于分层特征融合模块中的局部特征融合和全局特征融合分别证明其有效性。其中“√”表示使用该模块,“×”表示不使用该模块。消融实验的 4 项评价指标如表 1 所示。

表 1 在 Argoverse1 预测数据集(验证集)上的消融实验  
Table 1 Ablation experiments on the argoverse1 prediction dataset (validation set)

局部融合	全局融合	终点预测	终点细化	minADE <sub>6</sub> /m	minFDE <sub>6</sub> /m	MR <sub>6</sub> /%	b-minFDE <sub>6</sub> /m
×	×	×	×	0.706	1.043	10.3	1.680
√	√	×	×	0.685	0.996	9.0	1.613
×	×	√	√	0.665	0.952	8.0	1.566
√	√	√	×	0.654	0.937	7.6	1.557
×	√	√	√	0.651	0.942	7.5	1.544
√	×	√	√	0.658	0.947	7.7	1.556
√	√	√	√	0.644	0.902	7.1	1.530

根据表 1 的数据可得,本文 HFF-EI 模型中的分层特征融合、终点预测和终点细化 3 个模块均有效提升了多模态轨迹预测的性能,包括分层特征融合模块中的局部特征融合和全局特征融合。

分层特征融合模块对模型性能提升较大,在丢失率(MR<sub>6</sub>)上降低了 11.25%,证明分层特征融合可以全面且有效的交互复杂场景下各种元素特征,提高模型对环境的感知能力,其中全局特征融合更能对场景元素进行全面交互,提升模型性能,在丢失率(MR<sub>6</sub>)上降低了 6.25%,局部特征融合也能进一步提升模型预测的准确性,在丢失率(MR<sub>6</sub>)上降低了 3.75%;终点预测模块在 4 个评价指标上均有较大提升,其中最小最终位移误差(minFDE<sub>6</sub>)上降低了 5.92%,证明该模块不仅能准确预测驾驶员的决策行为,还能抑制预测轨迹发散;终点细化模块对性能也有提升,有效降低了最小最终位移误差(minFDE<sub>6</sub>),证明终点细化模块可以在更长时空范围内实现终点信息的交互,使得预测的车辆轨迹符合舒适性和安全性原则。

2.3 对比实验

为了验证 HFF-EI 多模态轨迹预测模型的有效性,将其与 10 个基准先进模型的单一模型进行对比实验,对比实验包括了 minADE<sub>6</sub>、minFDE<sub>6</sub>、MR<sub>6</sub> 和 b-minFDE<sub>6</sub> 这 4 个指标的对比结果,如表 2 所示。

由表 2 结果表明,由于 HFF-EI 模型基于终点信息进行了更长时空范围内的交互,所以在 minFDE<sub>6</sub>、b-minFDE<sub>6</sub> 两个评价指标上显著优于其他基准模型,其中在 minFDE<sub>6</sub> 上最大降低了 21.20%,在 b-minFDE<sub>6</sub> 上最大降低了 13.68%,证明了模型的终点预测和终点细化



表 2 在 Argoverse1 预测数据集 (测试集) 上的对比实验

Table 2 Comparative experiments on the argoverse1 prediction dataset (test set)

模型	minADE <sub>6</sub> /m	minFDE <sub>6</sub> /m	MR <sub>6</sub> /%	b-minFDE <sub>6</sub> /m
HiVT <sup>[7]</sup>	0.774	1.169	12.7	1.842
DenseTNT <sup>[12]</sup>	0.882	1.282	12.6	1.976
mmTrans <sup>[13]</sup>	0.844	1.338	15.4	2.033
ADAPT <sup>[15]</sup>	0.786	1.169	12.2	1.796
SIMPL <sup>[22]</sup>	0.793	1.179	12.3	1.809
LaneGCN <sup>[23]</sup>	0.870	1.362	16.2	2.054
THOMAS <sup>[25]</sup>	0.942	1.439	10.4	1.974
TPCN <sup>[26]</sup>	0.815	1.244	13.3	1.929
GANet <sup>[27]</sup>	0.806	1.160	11.8	1.790
SceneTrans <sup>[28]</sup>	0.803	1.232	12.6	1.887
HFF-EI	0.779	1.134	10.9	1.773

模块可以准确的预测轨迹终点,对预测轨迹进行有效约束。HFF-EI 模型在 minADE<sub>6</sub> 指标上略高于分层向量 Transformer(hierarchical vector Transformer, HiVT),但在其他 3 个指标上都远远低于 HiVT,特别是在 MR<sub>6</sub> 上降低了 14.17%。热力图轨迹预测模型<sup>[25]</sup>(trajectory heatmap output with multi-agent sampling, THOMAS)模型由于采用热力图输出的方式预测轨迹,所以其 MR<sub>6</sub> 指标最低,但是 HFF-EI 在 minADE<sub>6</sub>、minFDE<sub>6</sub> 和 b-minFDE<sub>6</sub> 这 3 个指标上显著优于 THOMAS,分别降低了 17.30%,21.20%和 10.18%。HFF-EI 模型的 4 项指标均优于时空点云轨迹预测模型(temporal point cloud networks, TPCN)<sup>[26]</sup>、目标区域网络模型(goal area network, GANet)<sup>[27]</sup>和场景 transformer 轨迹预测模型(scene transformer, SceneTrans)<sup>[28]</sup>。结合 4 个评价指标说明 HFF-EI 模型在多模态轨迹预测方面性能优势显著。

为了证明所提出的 HFF-EI 模型有效平衡了多模态轨迹预测的精度和效率,通过 model.parameters() 返回模型中所有参数的迭代器,使用 sum 函数和 for 循环遍历所有参数并计算总参数量。在 Argoverse1 数据集上对比了参数量和推理时间两个指标,如表 3 所示。同时绘制了评价指标 b-minFDE<sub>6</sub> 与参数量和推理时间的关系图,如图 10 所示。

在参数量方面,与车道图神经网络模型(lane graph convolutional network, LaneGCN)、多模态运动预测 transformer 模型(multimodal motion transformer, mmTrans)、HiVT 和 SceneTrans 这 4 个基准模型相比, HFF-EI 以最少的参数量实现了最低的轨迹预测误差; HFF-EI 模型虽然在参数量上比密集终点轨迹预测模型(dense end-to-end trajectory prediction, DenseTNT)略多

表 3 参数量和推理时间与 b-minFDE<sub>6</sub> 的关系

Table 3 Relationship between b-minFDE<sub>6</sub> and the number of parameters and inference time

模型	b-minFDE <sub>6</sub>	参数量/M	推理时间/ms
HiVT <sup>[7]</sup>	1.842	2.5	64.45
DenseTNT <sup>[12]</sup>	1.976	1.1	444.66
mmTrans <sup>[13]</sup>	2.033	2.6	7.66
SIMPL <sup>[22]</sup>	1.809	1.8	—
LaneGCN <sup>[23]</sup>	2.054	3.7	38.37
SceneTrans <sup>[28]</sup>	1.887	15.3	—
HFF-EI	1.794	2.0	10.22

一些,但是在 b-minFDE<sub>6</sub> 上降低了 9.21%,提高了轨迹预测的精度;HFF-EI 与简单高效轨迹预测模型(SIMPL)参数量相当的同时降低了轨迹预测误差。

在推理时间方面,HFF-EI 的推理时间为 10.22 ms,可以满足实时性要求。与 LaneGCN、DenseTNT 和 HiVT 这 3 个基准模型相比,HFF-EI 在推理时间上分别降低了 28.16、434.44 和 54.23 ms,提升了模型的预测效率;与 mmTrans 相比,HFF-EI 虽然牺牲了 2.56 ms 的少量推理时间但是大幅度地提高了预测精度,在 b-minFDE<sub>6</sub> 上降低了 11.76%。所以,HFF-EI 模型在预测精度和预测效率方面都具有较好的性能。

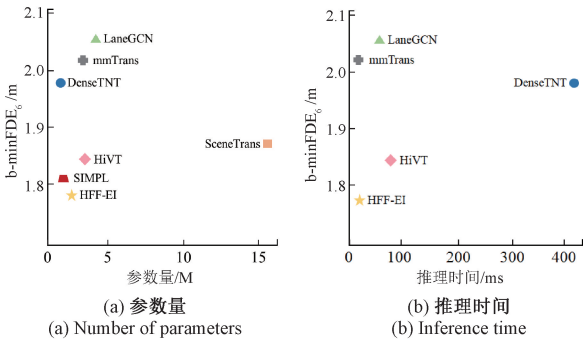


图 10 b-minFDE<sub>6</sub> 与参数量和推理时间的关系

Fig. 10 Relationship between b-minFDE<sub>6</sub> and the number of parameters and inference time

2.4 可视化实验

考虑到对比实验和消融实验都是基于评价指标的数据层面上说明本文模型的有效性,不易清晰直观的表明本文模型在轨迹预测任务上的具体表现,因此将数据集众多场景中的几个场景进行可视化。在可视化中,将被预测车辆所处的车道段信息、真实轨迹信息以及多模态预测轨迹信息进行图形化展示,主要包括典型交通场景的可视化和消融实验的可视化。可视化过程中剔除了概率<0.05 的预测轨迹。

1) 典型交通场景可视化

为了更加直观表述本文多模态轨迹预测模型的效果,选取多车道直行、变道和转弯等典型的交通场景进行预测轨迹和真实轨迹的可视化,如图 11 和 12 所示。其中,虚线代表被预测车辆的真实轨迹,三角形代表被预测车辆的真实轨迹终点;实线代表被预测车辆的多模态预测轨迹,星形代表被预测车辆的多模态轨迹终点。

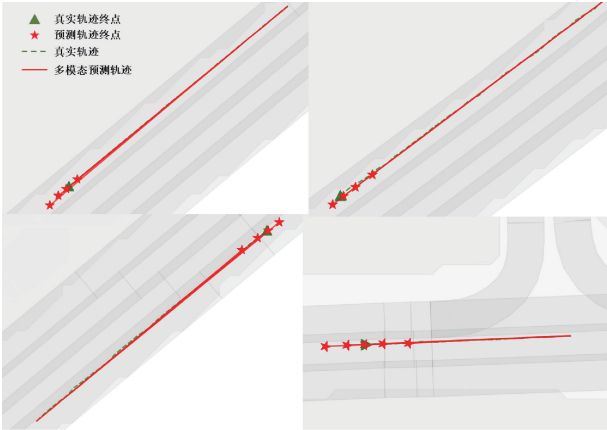


图 11 多车道直行场景可视化

Fig. 11 Visualisation of multi-lane straight ahead scenes

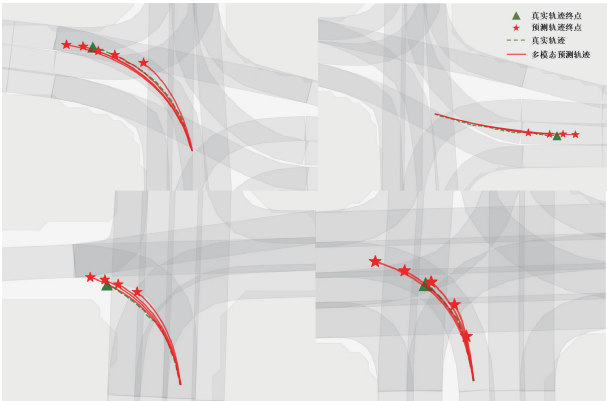


图 12 变道转弯场景可视化

Fig. 12 Visualisation of lane change turn scenes

由上述可视化结果可知,在多车道直行和变道转弯驾驶场景中,模型均能较为准确的预测未来轨迹和轨迹终点,符合数据集上对比实验的性能结果。

2) 消融实验可视化

为了直观清晰的说明本文模型各个模块有效性,选取数据集中场景序号为 10031、10219 和 10220 的 3 个场景进行消融实验的可视化,如图 13~15 所示。3 个场景都是位于交通情况复杂的十字路口区域,具有一定的代表性,在一定程度上更能体现本文模型中每个模块的有效性。同时,3 个场景消融实验的 4 项评价指标数据分别见表 4~6,旨在从数据和可视化两个方面充分说明消

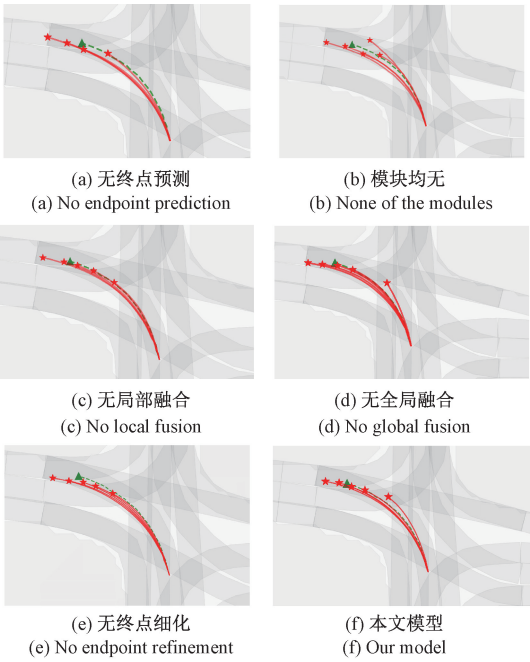


图 13 场景 10031 可视化

Fig. 13 Visualisation of scene 10031

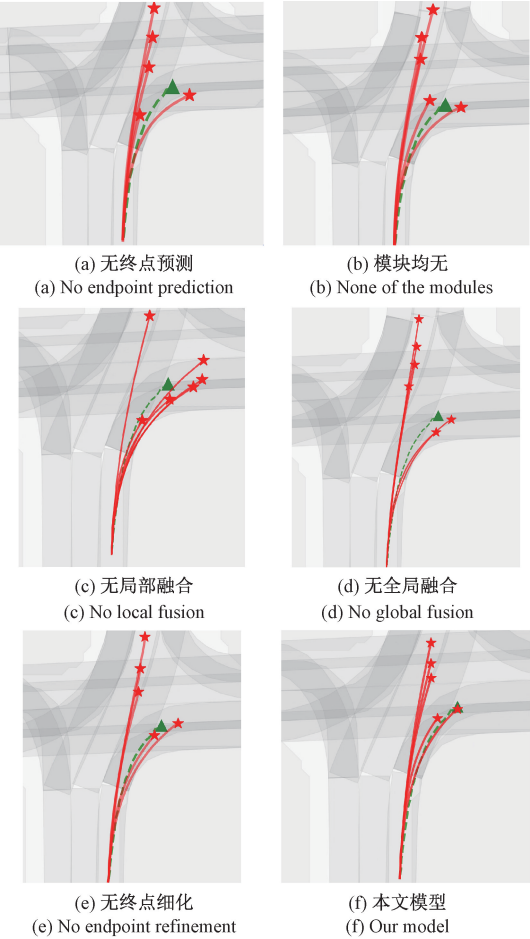


图 14 场景 10219 可视化

Fig. 14 Visualisation of scene 10219

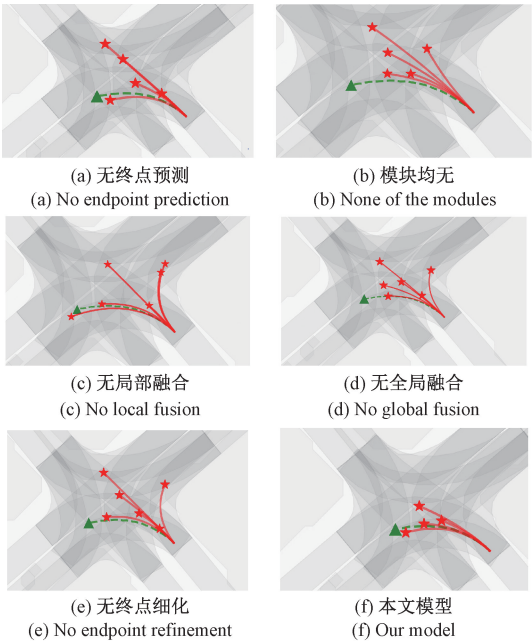


图 15 场景 10220 可视化  
Fig. 15 Visualisation of scene 10220

表 4 10031 场景消融实验评价指标  
Table 4 Evaluation metrics for scene 10031 ablation experiments

局部融合	全局融合	终点预测	终点细化	minADE <sub>6</sub> /m	minFDE <sub>6</sub> /m	MR <sub>6</sub> /%	b-minFDE <sub>6</sub> /m
×	×	×	×	0.890	1.452	0.0	2.094
√	√	×	×	1.082	1.075	0.0	1.523
×	×	√	√	0.703	1.023	0.0	1.617
√	√	√	×	0.753	1.012	0.0	1.538
×	√	√	√	0.683	1.015	0.0	1.351
√	×	√	√	0.695	1.020	0.0	1.446
√	√	√	√	0.436	0.339	0.0	0.789

表 5 10219 场景消融实验评价指标  
Table 5 Evaluation metrics for scene 10219 ablation experiments

局部融合	全局融合	终点预测	终点细化	minADE <sub>6</sub> /m	minFDE <sub>6</sub> /m	MR <sub>6</sub> /%	b-minFDE <sub>6</sub> /m
×	×	×	×	0.854	2.825	100.0	3.571
√	√	×	×	0.775	2.281	100.0	3.133
×	×	√	√	1.550	1.957	0.0	2.374
√	√	√	×	1.658	1.653	0.0	2.363
×	√	√	√	1.388	1.678	0.0	2.129
√	×	√	√	1.432	1.790	0.0	2.334
√	√	√	√	0.756	1.249	0.0	1.814

表 6 10220 场景消融实验评价指标  
Table 6 Evaluation metrics for scene 10220 ablation experiments

局部融合	全局融合	终点预测	终点细化	minADE <sub>6</sub> /m	minFDE <sub>6</sub> /m	MR <sub>6</sub> /%	b-minFDE <sub>6</sub> /m
×	×	×	×	1.591	3.833	100.0	4.313
√	√	×	×	0.677	1.913	0.0	2.394
×	×	√	√	0.473	0.962	0.0	1.872
√	√	√	×	1.069	2.726	100.0	3.476
×	√	√	√	0.338	0.878	0.0	1.332
√	×	√	√	0.452	0.922	0.0	1.545
√	√	√	√	0.053	0.084	0.0	0.771

融实验结果的真实性和有效性。其中,虚线代表被预测车辆的真实轨迹,三角形代表被预测车辆的真实轨迹终点;实线代表被预测车辆的多模态预测轨迹,星形代表被预测车辆的多模态轨迹终点。

由场景 10031 中的可视化结果可知,本文模型在该场景下终点预测明显更加接近真实轨迹的终点。

如表 4 所示,根据 4 项评价指标可知,在场景 10031 中,本文模型中的 3 个模块以及分层特征融合中的局部融合和全局融合均具有一定的有效性,轨迹预测的误差均显著降低。

由可视化可知,在场景 10219 中本文模型在完整轨迹和轨迹终点两个方面都更加接近真实轨迹。

如表 5 所示,由评价指标可知,在场景 10219 中,本文模型中的 3 个模块均具有有效性,模型在 4 项评价指标上均显著降低。

由可视化可知,在场景 10220 中本文模型预测的整条轨迹几乎贴合真实轨迹,同时可以使得预测的多模态轨迹更加分散,更好的模拟驾驶行为的不确定性。如表 6 所示,由评价指标可知,在场景 10220 中,本文模型中的预测精度得到了显著提升。

根据选取两类典型场景的可视化以及 3 个场景消融实验的可视化结果可知,本文模型能有效提升多模态轨迹预测的准确性。

3) 失败场景分析

为了进一步探索模型在实际场景中应用的不足,明确后续改进的方向,选取数据集的 4 个失败场景进行分析,场景序号分别为 1002、10014、10019、10080。如图 16 所示,4 个场景误差如表 7 所示,其中,虚线代表被预测车辆的真实轨迹,三角形代表被预测车辆的真实轨迹终点;实线代表被预测车辆的多模态预测轨迹,星形代表被预测车辆的多模态轨迹终点。



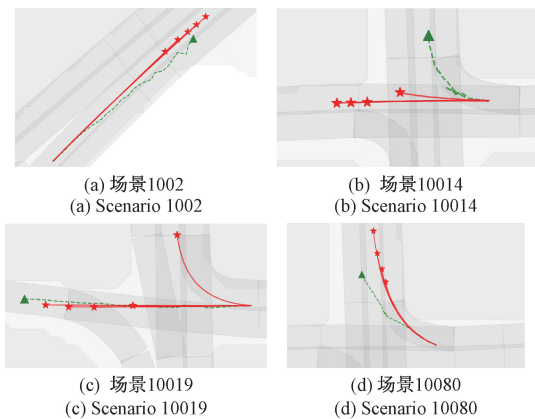


图 16 失败场景可视化

Fig. 16 Visualisation of failure scenarios

表 7 失败场景评价指标

Table 7 Evaluation metrics for failure scenarios				
场景序号	minADE <sub>6</sub> /m	minFDE <sub>6</sub> /m	MR <sub>6</sub> /%	b-minFDE <sub>6</sub> /m
1002	1.647	1.573	0.0	2.150
10014	0.677	1.913	0.0	2.394
10019	0.941	0.889	0.0	1.569
10080	1.355	2.613	100.0	3.160

由可视化结果可知,本文模型对于车辆行驶轨迹方向变化突然、高动态的场景准确率不足,从模型输出的预测轨迹来看,本文模型更偏向于生成较为平滑的预测结果,未能准确捕捉到真实轨迹中那些突发性的方向变化,如上述 4 个场景。

3 结 论

针对现有车辆轨迹预测方法特征融合不充分和难以有效平衡预测精度和效率的问题,提出了一种基于分层特征融合和终点诱导的车辆多模态轨迹预测模型(HFF-EI)。HFF-EI 模型构建分层结构实现特征充分且高效融合,并且将轨迹预测任务分解为终点预测、终点细化和轨迹补全 3 个部分,有效平衡预测精度和效率。在分层特征融合中,先利用局部特征融合筛选出更重要的特征赋予更高权重,再进行基于位置关系的全局特征融合;在终点预测中引入动态权重 MLP,实现了轨迹的高效预测;在终点细化中利用注意力机制进行轨迹终点信息的交互,提高了轨迹预测的精度。在 Argoverse1 数据集上进行消融实验和对比实验,消融实验结果充分证明 HFF-EI 模型中各个模块的有效性,与其他 11 个基准模型的对比实验结果显示 HFF-EI 模型在 4 个评价指标上具有良好性能,可以同时兼顾预测精度和效率。此外,本文还存在一定

的不足,比如在轨迹特征提取部分,采用了较为基础的一维残差卷积和 FPN 结构。在未来的工作中,可以考虑使用 LSTM 或者 Transformer 进行轨迹时序特征的提取,并且增加自建数据集或实际应用场景测试结果。

参考文献

[ 1 ] 王红霞, 聂振凯, 钟强, 等. 融合距离阈值和双向 TCN 的时空注意力行人轨迹预测模型[J]. 计算机应用研究, 2024, 41(11): 3303-3310.

WANG H X, NIE ZH K, ZHONG Q, et al. Fusion of distance threshold and Bi-TCN for spatio-temporal attention pedestrian trajectory prediction model [J]. Application Research of Computers, 2024, 41(11): 3303-3310.

[ 2 ] 黄峻, 田永林, 戴星原, 等. 基于深度学习的自动驾驶多模态轨迹预测方法: 现状及展望[J]. 智能科学与技术学报, 2023, 5(2): 180-199.

HUANG J, TIAN Y L, DAI X Y, et al. Deep learning-based multimodal trajectory prediction methods for autonomous driving: Current status and perspectives[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(2): 180-199.

[ 3 ] LEFÈVRE S, VASQUEZ D, LAUGIER C, et al. A survey on motion prediction and risk assessment for intelligent vehicles [J]. ROBOMECH Journal, 2014, 1(1): 1-14.

[ 4 ] HUANG Y J, DU J T, YANG Z R, et al. A survey on trajectory-prediction methods for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2022, 7(3): 652-674.

[ 5 ] GAO J Y, SUN CH, ZHAO H, et al. VectorNet: Encoding HD maps and agent dynamics from vectorized representation [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11522-11530.

[ 6 ] ZENG W Y, LIANG M, LIAO R J, et al. LaneRCNN: Distributed representations for graph-centric motion forecasting[C]. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2021: 532-539.

[ 7 ] ZHOU Z K, YE L Y, WANG J P, et al. HiVT: Hierarchical vector transformer for multi-agent motion prediction [C]. IEEE Conference on Computer Vision and Pattern Recognition, NJ: IEEE Press, 2022: 8813-

- 8823.
- [ 8 ] VARADARAJAN B, HEFNY A, SRIVASTAVA A, et al. Multipath + +: Efficient information fusion and trajectory aggregation for behavior prediction [ C ]. 2022 International Conference on Robotics and Automation, 2022: 7814-7821.
- [ 9 ] 胡杰, 吴作伟, 张志凌, 等. 基于结构化道路的车辆多模态轨迹预测方法 [ J ]. 中国公路学报, 2025, 38(2): 286-295.
- HU J, WU Z W, ZHANG ZH L, et al. Multi-modal vehicle trajectory prediction method based on structured road [ J ]. China Journal of Highway and Transport, 2025, 38(2): 286-295.
- [ 10 ] 王庆荣, 郝福乐, 朱昌锋, 等. 基于多特征融合的车辆轨迹预测研究 [ J/OL ]. 计算机工程, 1-14 [ 2025-06-10 ].
- WANG Q R, HAO F L, ZHU CH F, et al. Research on vehicle trajectory prediction based on multi-feature fusion [ J/OL ]. Computer Engineering, 1-14 [ 2025-06-10 ].
- [ 11 ] ZHAO H, GAO J Y, LAN T, et al. TNT: Target-driven trajectory prediction [ C ]. Conference on Robot Learning. PMLR, 2021: 895-904.
- [ 12 ] GU J R, SUN CH, ZHAO H. DenseTNT: End-to-end trajectory prediction from dense goal sets [ C ]. 2021 IEEE/CVF International Conference on Computer Vision, 2021: 15283-15292.
- [ 13 ] LIU Y CH, ZHANG J H, FANG L J, et al. Multimodal motion prediction with stacked Transformers [ C ]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7573-7582.
- [ 14 ] ZHOU Z K, WANG J P, LI Y H, et al. Query-centric trajectory prediction [ C ]. IEEE Conference on Computer Vision and Pattern Recognition, 2023: 17863-17873.
- [ 15 ] AYDEMIR G, AKAN A K, GUNEY F. ADAPT: Efficient multi-agent trajectory prediction with adaptation [ C ]. IEEE International Conference on Computer Vision, 2023: 8261-8271.
- [ 16 ] 方鑫, 何敏, 黄大荣, 等. 毫米波雷达微弱行人轨迹跟踪-预测一体化方法 [ J ]. 仪器仪表学报, 2023, 44(11): 300-309.
- FANG X, HE M, HUANG D R, et al. Integrated trajectory tracking and prediction method for weak pedestrian with millimeter wave radar [ J ]. Chinese Journal of Scientific Instrument, 2023, 44(11): 300-309.
- [ 17 ] 王江, 史元浩, 郭正玉, 等. 融合小波分解和 LSTM 的目标轨迹预测 [ J ]. 电子测量与仪器学报, 2023, 37(1): 204-211.
- WANG J, SHI Y H, GUO ZH Y, et al. Target trajectory prediction by fusing wavelet decomposition and LSTM [ J ]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1): 204-211.
- [ 18 ] 卢一凡, 李焯鹏, 薛启凡. 面向未知域场景的车辆轨迹预测模型 [ J ]. 智能系统学报, 2024, 19(5): 1238-1247.
- LU Y F, LI X P, XUE Q F. A vehicle trajectory prediction model for unknown domain scenarios [ J ]. CAAI Transactions on Intelligent Systems, 2024, 19(5): 1238-1247.
- [ 19 ] 周正阳, 潘树国, 蔚保国, 等. 基于 ABMSSA 的 PP 算法分布式自动驾驶轨迹跟踪控制策略 [ J ]. 电子测量与仪器学报, 2024, 38(6): 50-57.
- ZHOU ZH Y, PAN SH G, WEI B G, et al. Distributed automatic driving trajectory tracking control strategy based on PP algorithm based on ABMSSA [ J ]. Journal of Electronic Measurement and Instrumentation, 2024, 38(6): 50-57.
- [ 20 ] 陈熙源, 周云川, 钟雨露, 等. 基于变分贝叶斯的鲁棒自适应因子图优化组合导航算法 [ J ]. 仪器仪表学报, 2024, 45(1): 120-129.
- CHEN X Y, ZHOU Y CH, ZHONG Y L, et al. Robust adaptive factor graph optimization integrated navigation algorithm based on variational Bayesian [ J ]. Chinese Journal of Scientific Instrument, 2024, 45(1): 120-129.
- [ 21 ] CUI A, CASAS S, WONG K, et al. GoRela: Go relative for viewpoint-invariant motion forecasting [ C ]. IEEE International Conference on Robotics and Automation, 2023: 7801-7807.
- [ 22 ] ZHANG L, LI P L, LIU S K, et al. SIMPL: A simple and efficient multi-agent motion prediction baseline for autonomous driving [ J ]. IEEE Robotics and Automation Letters, 2024, 9(4): 3767-3774.
- [ 23 ] LIANG M, YANG B, HU R, et al. Learning lane graph representations for motion forecasting [ C ]. European

Conference on Computer Vision, 2020: 541-556.

[24] CHANG M F, LAMBERT J, SANGKLOY P, et al. Argoverse: 3D tracking and forecasting with rich maps[C]. Conference on Computer Vision and Pattern Recognition, 2019: 8748-8757.

[25] GILLES T, SABATINI S, TSISHKOU D, et al. THOMAS: Trajectory heatmap output with learned multi-agent sampling[C]. In Proceedings of the International Conference on Learning Representations, 2022.

[26] YE M SH, CAO T Y, CHEN Q F. TPCN: Temporal point cloud networks for motion forecasting [ C ]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11318-11327.

[27] WANG M K, ZHU X G, YU CH Q, et al. GANet: Goal area network for motion forecasting [ C ]. In IEEE International Conference on Robotics and Automation, 2023: 1609-1615.

[28] NGIAM J, VASUDEVAN V, CAINE B, et al. Scene Transformer: A unified architecture for predicting future trajectories of multiple agents[C]. In Proceedings of the International Conference on Learning Representations, 2021.

作者简介



E-mail: chxiyuan@seu.edu.cn

陈熙源(通信作者),1990 年于兰州理工大学获得学士学位,1995 年于合肥工业大学获得硕士学位,1998 年毕业于东南大学获得博士学位,现为东南大学教授,主要研究方向为惯性技术、软件接收机及多源组合导航。

Chen Xiuyan (Corresponding author) received his B. Sc.

degree from Lanzhou University of Technology in 1990, received his M. Sc. degree from Hefei University of Technology in 1995, received his Ph. D. degree from Southeast University (SEU) in 1998. He is currently a professor at SEU. His main research interests include inertial technologies, software defined receiver and multi-source integrated navigation.



聂姝涵,2023 年于东南大学获得学士学位,现为东南大学硕士研究生,主要研究方向为自动驾驶轨迹预测。

E-mail: 1016994079@qq.com

Nie Shuhan received her B. Sc. degree from Southeast University in 2023. She is currently a master student at Southeast University. Her main research interest is autonomous trajectory prediction.



刘炜焱,2021 年于重庆邮电大学获得学士学位,现为东南大学硕士研究生,主要研究方向为自动驾驶决策规划。

E-mail: 973661447@qq.com

Liu Weiyan received her B. Sc. degree from Chongqing University of Posts and Telecommunications in 2021. She is currently a master student at Southeast University. Her main research interests include autonomous driving perception and decision planning.



经纬铭,2023 年于三峡大学获得学士学位,现为东南大学硕士研究生,主要研究方向为多模态融合感知。

E-mail: weiming0120@163.com

Jing Weiming received his B. Sc. degree from Three Gorges University in 2023. He is currently a master student at Southeast University. His main research interest includes multimodal fusion perception.