

基于仿射全注意力机制的车辆轨迹预测方法

吴林蓉

上海理工大学光电信息与计算机工程学院, 上海

收稿日期: 2025年8月16日; 录用日期: 2025年9月8日; 发布日期: 2025年9月16日

摘要

针对现有车辆轨迹预测方法在复杂交通场景中时空信息提取不充分、多智能体交互建模能力有限的问题, 本文提出一种基于仿射全注意力机制的车辆轨迹预测方法BiFSTNet。该方法采用动态邻接矩阵建模车辆间空间关系, 利用视觉Transformer从多车辆历史轨迹中提取全局上下文特征, 并设计仿射全注意力机制融合时间多头注意力和空间图注意力网络, 有效捕获时空交互模式和长期时间依赖关系。采用双向长短期记忆网络构建编码器-解码器架构, 实现端到端的轨迹预测。在NGSIM US-101和I-80数据集上的实验结果表明, BiFSTNet在US-101数据集上平均均方根误差为1.77m, 在I-80数据集上平均均方根误差为2.06m, 相比基线方法在各预测时间范围内均有性能提升。

关键词

车辆轨迹预测, 视觉Transformer, 仿射全注意力机制, LSTM, 长期预测

A Vehicle Trajectory Prediction Method Based on Affine Full Attention Mechanism

Linrong Wu

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Aug. 16th, 2025; accepted: Sep. 8th, 2025; published: Sep. 16th, 2025

Abstract

To address the insufficient spatiotemporal information extraction and limited multi-agent interaction modeling capability of existing vehicle trajectory prediction methods in complex traffic scenarios, this paper proposes BiFSTNet, a vehicle trajectory prediction method based on affine full attention mechanism. The proposed method employs dynamic adjacency matrices to model spatial relationships between vehicles and utilizes Vision Transformer to extract global contextual features

文章引用: 吴林蓉. 基于仿射全注意力机制的车辆轨迹预测方法[J]. 建模与仿真, 2025, 14(9): 128-137.

DOI: 10.12677/mos.2025.149590

from multi-vehicle historical trajectories. An affine full attention mechanism is designed to fuse temporal multi-head attention and spatial graph attention networks, effectively capturing spatio-temporal interaction patterns and long-term temporal dependencies. A bidirectional long short-term memory network is adopted to construct an encoder-decoder architecture for end-to-end trajectory prediction. Experimental results on the NGSIM US-101 and I-80 datasets demonstrate that BiFSTNet achieves an average root mean square error of 1.77 m on the US-101 dataset and 2.06 m on the I-80 dataset, showing performance improvements over baseline methods across all prediction time horizons.

Keywords

Vehicle Trajectory Prediction, Vision Transformer, Affine Full Attention Mechanism, Long Short-Term Memory, Long-Term Prediction

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着我国国务院办公厅于 2020 年底发布的《新能源汽车产业发展规划(2021~2035 年)》，明确提出力争在 2025 年前实现高度自动驾驶汽车在特定区域和场景下的商业化应用，自动驾驶技术迅速成为未来交通发展的核心方向之一[1]。为了实现安全高效地驾驶，自动驾驶车辆需要像人类一样检测和识别其他物体，并预测和反应这些物体在短期内的行为。因此，准确预测其他道路参与者的轨迹是自动驾驶车辆做出明智决策的基础。

当前，车辆轨迹预测方法主要分为基于物理运动学的方法、基于传统机器学习的方法和基于深度学习的方法。基于物理运动学的方法如卡尔曼滤波(Kalman Filter) [2]和切换卡尔曼滤波(Switched Kalman Filter, SKF)等[3]，虽然能够在一定程度上捕捉车辆的运动规律，但在处理非线性和多主体交互的复杂场景时表现不足。传统机器学习方法，如高斯过程回归[4]、支持向量机(SVM) [5]和隐马尔科夫模型(HMM) [6]，通过预测车辆的机动意图并生成轨迹，提升了预测性能，但其难以适应多变的实际驾驶环境，且通常忽略了车辆之间的复杂交互作用，导致预测精度有限[7]。

近年来，基于深度学习的方法因其强大的特征学习能力和对复杂场景的适应性，逐渐成为轨迹预测研究的主流方向。循环神经网络(RNN) [8]和长短期记忆网络(LSTM) [9]在捕捉时间序列数据的动态特征方面表现出色。卷积神经网络(CNN)和注意力机制(Attention) [9]的引入，进一步提升了轨迹预测的性能，能够提取更丰富的交互信息。例如，Chen 等人[10]提出的 LSTM 编解码模型通过卷积 - 社交池化层更好地学习车辆间的相互依赖关系，Pengbo [11]等人结合 LSTM 编码器与地图信息实现基于行为的轨迹预测，Dekai [12]等人引入注意力机制以强调邻近车辆对自车未来状态的重要性，Wang [13]等人则采用 CNN-LSTM 混合网络建模不同道路间的交互作用。这些方法在一定程度上提升了轨迹预测的精度，但在复杂动态环境下的实时性和鲁棒性仍需进一步加强。

针对上述问题，本文提出了一种名为 BiFSTNet 的新型车辆轨迹预测模型。在空间特征提取方面，引入视觉 Transformer 处理多车辆历史轨迹。与卷积操作相比，ViT 的自注意力机制能直接建模场景中所有车辆对的全局交互，能够更有效地从由大量智能体构成的复杂场景中提取统一的上下文特征表示。在时空交互建模方面，本文设计了仿射全注意力机制，采用一种时序优先的序贯处理流程。车辆的行为首先

由其自身的历史轨迹(时间维度)所主导,再与周围环境(空间维度)发生交互。这种设计允许模型先独立地强化每个车辆轨迹内部的长程时间依赖关系,再在每一时间步的语义空间中精确计算车辆间的空间交互权重,避免了时空信号过早混合带来的混淆,实现了更清晰、更有效的时空模式捕获。最终,通过双向 LSTM 编码器-解码器架构整合上述特征,实现端到端的未来轨迹预测。

2. 问题描述

轨迹预测问题可描述为:基于观测场景中所有物体的历史轨迹,预测它们在未来时刻的轨迹。考虑到与预测物体位置相比,预测速度相对更容易,因此输入包括历史位置与速度,模型负责预测未来的速度。最终位置的预测通过累加预测的速度与最近观测到的位置获得。

设当前时刻为 t , 历史观测窗口长度为 t_h , 则在时间区间 $[t-t_h, t]$ 内, 每个离散时刻 τ 均可观测到场景中 N 个目标。第 i 个目标在时刻 τ 的状态向量用式(1)表示。

$$s_i(\tau) = (x_i(\tau), y_i(\tau), u_i(\tau), v_i(\tau)) \quad (1)$$

其中, $(x_i(\tau), y_i(\tau))$ 为目标在平面坐标系中的位置坐标, $(u_i(\tau), v_i(\tau))$ 为目标在该时刻的速度分量。

将历史时间窗口内所有目标的观测数据集合表示为式(2)。

$$X = \{(x_i(\tau), y_i(\tau), u_i(\tau), v_i(\tau)) | i = 1, \dots, N; \tau = t-t_h, \dots, t\} \quad (2)$$

设定预测时长为 t_f , 模型需对未来时刻 $t+1$ 至 $t+t_f$ 内所有目标的速度进行预测。将这些未来速度预测结果集合表示为式(3)。

$$Y = \{(u_i(\tau), v_i(\tau)) | i = 1, \dots, N; \tau = t+1, \dots, t+t_f\} \quad (3)$$

在获得未来速度预测 Y 后, 利用最后一次观测到的位置 $(x_i(t), y_i(t))$ 对预测的速度进行逐步积分与累加, 得到各目标在未来每个时间步的预计位置轨迹。

3. 网络模型

图 1 显示了 BiFSTNet 的模型架构, 该模型由四个组件组成: 输入预处理模块、ViT 特征提取模块、仿射全注意力机制编码模块、轨迹预测模块。

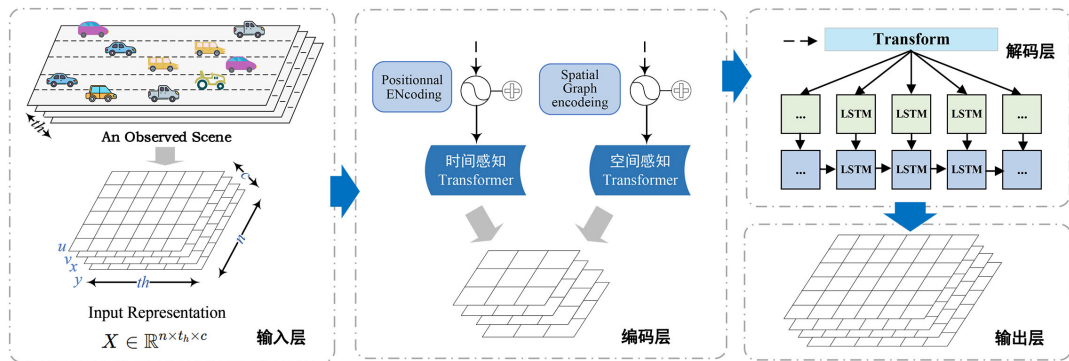


Figure 1. Workflow of the proposed BiFSTNet algorithm
图 1. BiFSTNet 算法流程图

3.1. 输入预处理模块

3.1.1. 输入表示

在交通场景中, 每个时间步 t 的车辆轨迹数据由 n 辆车的状态组成。经过归一化处理后, 输入数据表

示为式(4)。

$$X^t \in R^{n \times 4}, \quad X^t = \left[\left(x^t, y^t, u^t, v^t \right) \right] \quad (4)$$

其中, (x, y) 表示车辆的位置坐标, (u, v) 表示车辆的速度信息。通过归一化, 将数据值限定在特定范围内, 以便后续处理。

3.1.2. 空间图构建

在交通场景中, 车辆之间的相互关系对运动建模具有重要意义。为刻画这种关系, 针对每个时间步 t , 构建无向图 $G_t = (V_t, E_t)$, 其中:

节点集: $V_t = \{v_i^t \mid i = 1, 2, \dots, n\}$, 每个节点表示一辆车辆;

边集: E_t 表示满足特定交互条件的车辆对。

具体而言, 若两车辆间的距离满足 $\text{distance}(v_i^t, v_j^t) \leq T_{\text{close}}$, 且车道差满足 $|\text{lane}_i^t - \text{lane}_j^t| \leq 1$, 则认为车辆 v_i^t 和 v_j^t 存在交互关系, 用邻接矩阵 $A_t \in R^{n \times n}$ 表示该关系, 定义如式(5)。

$$A_t[i, j] = \begin{cases} 1 & \text{if } \text{distance}(v_i^t, v_j^t) \leq T_{\text{close}} \text{ and } |\text{lane}_i^t - \text{lane}_j^t| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

上述邻接矩阵 A_t 随时间步 t 变化, 形成邻接矩阵序列 $\{A_t\}_{t=1}^{t=t_h}$ 。与车辆状态序列 $\{X^t\}_{t=1}^{t=t_h}$ 共同构成模型输入, 用于刻画车辆间的动态空间关系。

3.2. ViT 特征提取模块

在获得车辆时序状态序列 $\{X^t\}_{t=1}^{t_h}$ 及其对应的邻接矩阵序列 $\{A_t\}_{t=1}^{t_h}$ 后, 需要进一步提取高层特征以捕捉复杂的时空交互关系。为此, 可将每一时间步的数据表示为一组适合 Transformer 结构处理的 token 序列, 并利用 ViT 从中提取全局上下文特征。

首先, 对每个时间步 t 的节点数据 $X^t \in R^{n \times 4}$ 进行特征映射, 将每个节点 $(x_i^t, y_i^t, u_i^t, v_i^t)$ 映射至 d 维特征空间。定义节点嵌入函数 $\phi: R^4 \rightarrow R^d$, 见式(6)。

$$z_i^t = \phi(x_i^t, y_i^t, u_i^t, v_i^t) \in R^d \quad (6)$$

由此, 可获得时间步 t 的节点特征序列, 见式(7)。

$$Z^t = [z_1^t, z_2^t, \dots, z_n^t] \in R^{n \times d} \quad (7)$$

考虑到 Transformer 对输入序列敏感, 于是对序列加入位置嵌入(Positional Embedding) $P \in R^{n \times d}$ 以编码节点的空间排列和场景结构信息, 见式(8)。

$$H^t = Z^t + P \quad (8)$$

将 H^t 输入至 ViT 编码器(ViT Encoder)中。ViT 编码器由多层 Multi-Head Self-Attention 和前馈网络组成。经过 L 层编码器后得到输出, 见式(9)。

$$H^{\text{ViT}} = \text{ViTEncoder}(H^t) \in R^{n \times t_h \times d_{\text{model}}} \quad (9)$$

n 表示节点(车辆)的数量, t_h 表示用于预测的历史时间步数长度, d_{model} 表示 ViT 编码器输出特征的模型隐含维度。此时, H^{ViT} 为处理后得到的高层语义特征序列, 能够较好地表示车辆间的全局时空交互关系。

3.3. 仿射全注意力机制模块

仿射全注意力机制模块由时间多头注意力机制、空间多头注意力机制、残差连接、归一化层等组成,

如图 2 所示。给定从 ViT 模块获得的输入数据 $H^{ViT} \in R^{n \times t_h \times d_{model}}$ ，执行以下操作。

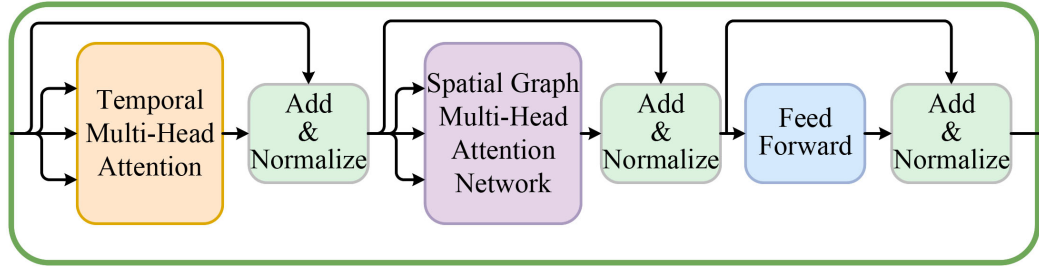


Figure 2. Component layout of the affine full-attention module

图 2. 仿射全注意力机制组成模块

3.3.1. 时间多头注意力模块

时间多头注意力模块用于从车辆轨迹序列中提取时间依赖性，并通过掩蔽多头注意力机制，确保当前时间步只访问历史时间步的信息，满足时间因果性。输入为从 ViT 模块获得的嵌入特征 $H^{ViT} \in R^{n \times t_h \times d_{model}}$ ，其中 n 表示车辆数量， t_h 表示时间序列长度， d_{model} 表示嵌入特征的维度。

对于车辆 i 在时间步 t 的嵌入特征记为 $h_{i,t} \in R^{d_{model}}$ 。通过线性映射函数 f_Q, f_K, f_V ，分别计算查询矩阵 $Q_{i,t} \in R_k$ 、键矩阵 $K_{i,t} \in R^{d_k}$ 和值矩阵 $V_{i,t} \in R^{d_v}$ ，具体表示为式(10)。

$$Q_{i,t} = f_Q(h_{i,t}), \quad K_{i,t} = f_K(h_{i,t}), \quad V_{i,t} = f_V(h_{i,t}), \quad (10)$$

其中 d_k 和 d_v 分别表示查询/键的维度和值的维度。 f_Q, f_K, f_V 是共享的线性变换函数。

在时间步 t ，车辆 i 对所有历史时间步 $s \leq t$ 的注意力分数通过查询和键计算得到式(11)。

$$m_{t \leftarrow s} = \frac{Q_{i,t} K_{i,s}^T}{\sqrt{d_k}} \quad (11)$$

其中 $m_{t \leftarrow s}$ 表示时间步 t 与时间步 s 之间的相关性分数， $\sqrt{d_k}$ 用于数值归一化，防止注意力权重的数值过大。为了满足时间因果性，引入掩蔽机制，规范化注意力分数以生成注意力权重如式(12)。

$$a_{t \leftarrow s} = \frac{\exp(m_{t \leftarrow s})}{\sum_{s' \leq t} \exp(m_{t \leftarrow s'})}, \quad (12)$$

其中 $a_{t \leftarrow s}$ 表示时间步 t 对时间步 s 的注意力权重，并通过掩蔽确保仅考虑的时间步。基于注意力权重，时间步 t 的上下文表示可通过值向量 $V_{i,s}$ 加权求和得到式(13)。

$$\text{Attn}_{i,t} = \sum_{s \leq t} a_{t \leftarrow s} V_{i,s} \quad (13)$$

$\text{Attn}_{i,t}$ 表示从历史轨迹中提取的时间信息。

为增强模型表达能力，引入多头注意力机制。设共有 k 个注意力头，每个头独立计算注意力，如式(14)。

$$\text{head}_j = \text{Attn}_{i,t}^{(j)}, \quad j = 1, 2, \dots, k \quad (14)$$

其中 $\text{Attn}_{i,t}^{(j)}$ 是第 j 个注意力头的输出。将所有头的输出拼接后，通过全连接层 f_o 进行映射，得到最终时间多头注意力的输出式(15)。

$$T_{i,t} = f_o([\text{head}_1; \text{head}_2; \dots; \text{head}_k]) \quad (15)$$

其中 $T_{i,t} \in R^{d_{\text{model}}}$ 是车辆 i 在时间步 t 的最终特征表示。

经过时间多头注意力模块, 所有车辆的特征表示被更新为 $T \in R^{n \times t_h \times d_{\text{model}}}$, 其中包含从轨迹中提取的时间依赖性信息。

3.3.2. 空间图多头注意力网络

基于获得的 $T \in R^{n \times t_h \times d_{\text{model}}}$, 应用空间图多头注意力网络来提取观测车辆之间的空间交互。

自注意力机制可以被视为在无向完全连接图上的信息传递。对于某一时间步 t , 从 T 中获取 n 辆车的特征 $\{h_i^t\}_{i=1}^n \in R^{n \times d_{\text{model}}}$, 并将其对应的查询向量、键向量和值向量分别表示为式(16)。

$$q_i^t = f_Q(h_i^t), \quad k_i^t = f_K(h_i^t), \quad v_i^t = f_V(h_i^t) \quad (16)$$

计算 Q_t 、 K_t 、 V_t 如式(17)所示。

$$Q_t = f_Q\left(\{h_i^t\}_{i=1}^n\right), \quad K_t = f_K\left(\{h_i^t\}_{i=1}^n\right), \quad V_t = f_V\left(\{h_i^t\}_{i=1}^n\right) \quad (17)$$

并将车辆 j 到车辆 i 在完全连接图中的信息传递定义为式(18)。

$$m_{j \rightarrow i} = (q_i^t)^T k_i^t \quad (18)$$

然后, 时间步 t 的注意力计算如式(19)所示。

$$\text{Attention}(Q_t, K_t, V_t) = \text{Softmax}\left(\left[m_{j \rightarrow i}\right]_{i,j=1}^n\right) \frac{1}{\sqrt{d_k}} \left[v_i^t\right]_{i=1}^n \quad (19)$$

然而, 将车辆之间的空间交互视为完全连接图是不高效的。因此, 使用邻接矩阵 A 来替代完全连接图, 这确保了车辆 j 到车辆 i 的信息传递仅在当前两辆车之间的距离小于阈值 T_{close} 且两辆车位于相邻车道时才发生。然后, 重写车辆 i 在时间步 t 的注意力计算如式(20)所示。

$$\text{Attention}(i) = \text{Softmax}\left(\left[m_{j \rightarrow i}\right]_{j \in N(i)}\right) \frac{1}{\sqrt{d_k}} \left[v_j^t\right]_{j \in N(i)} \quad (20)$$

其中, $N(i) = \{j | A[i, j] = 1, j \in [1, n]\}$ 表示车辆 i 的邻居集合。类似地, 车辆 i 在时间步 t 计算多头注意力(共 k 个头), 见式(21)、式(22)。

$$S_i^t = f_O\left([head_1; head_2; \dots; head_k]\right) \quad (21)$$

其中,

$$head_j = \text{Attention}_j(i) \quad (22)$$

f_O 是一个全连接层, 用于合并 k 个头的信息。在为每辆车 $i \in [1, n]$ 和每个时间步 $t \in [1, t_h]$ 计算完多头注意力 S_i^t 后, 获得 $S \in R^{n \times t_h \times d_{\text{model}}}$, 它包含了从历史轨迹中提取的车辆间交互信息。堆叠六个仿射全注意力机制层, 以捕捉更复杂和抽象的时间和空间信息。

3.4. 轨迹预测模块

将上一节得到的时空特征序列 $S \in R^{n \times t_h \times d_{\text{model}}}$ 按时间送入双向 LSTM, 得到第 i 辆车的上下文向量 $h_i^{\text{enc}} \in R^{2h}$, 其中 $h_i^{\text{enc}} = [\tilde{h}_{i,t_h}; \tilde{h}_{i,1}]$ 同时融合先行与滞后信息; 随后以 $h_{i,0} = h_i^{\text{enc}}$ 作为隐藏状态、以零向量 $c_{i,0} = 0$ 作为记忆单元, 并以最后观测速度 $s_{i,0} = v_{i,t_h} \in R^2$ 作为解码器首帧输入, 启动单层 LSTM-Decoder; 在预测步 $t = 1, \dots, t_p$ 内, 递推关系见式(23)~(25)。

$$(h_{i,t}, c_{i,t}) = \text{LSTM}(s_{i,t-1}, h_{i,t-1}, c_{i,t-1}) \quad (23)$$

其中上一步输入 $s_{i,t-1}$ 直接取模型预测速度 $\hat{v}_{i,t-1}$ ；随后通过线性映射获得二维速度估计：

$$\hat{v}_{i,t} = W_o h_{i,t} + b_o \quad (W_o \in R^{2 \times h}, b_o \in R^2) \quad (24)$$

最后以固定步长 $\Delta t = 0.2 \text{ s}$ 累积速度得到位置

$$\hat{p}_{i,t} = p_{i,t_h} + \Delta t \sum_{k=1}^t \hat{v}_{i,k} \quad (25)$$

其中 p_{i,t_h} 为最后观测坐标；迭代 t_p 步即可输出未来 5 s 内每辆车完整、物理一致的速度 - 位置轨迹

$$\{(\hat{v}_{i,t}, \hat{p}_{i,t})\}_{t=1}^{t_p}。$$

3.5. 算法整体实施步骤

首先，输入层接收历史轨迹张量 $X \in R^{n \times t_h \times c}$ ，其中 n 为车辆数、 t_h 为历史帧数 ($3 \text{ s} \times 5 \text{ Hz} = 15$)、 $c = 4$ 对应位置与速度 (x, y, v_x, v_y) ，并同步读取邻接矩阵序列 $A \in \{0, 1\}^{N \times N \times t_h}$ 以刻画每帧车辆间的空间关联；随后编码层将 X 经位置编码与空间图编码注入时序与拓扑信息先后送入 ViT 与仿射全注意力机制，输出统一的时空特征 $S \in R^{n \times t_h \times d}$ ($d = 128$)；解码层利用 Bi-LSTM 先汇总历史上下文得到隐藏状态，再按自回归方式逐帧输出未来速度序列 $\hat{V} \in R^{N \times t_p \times 2}$ (预测窗口 $t_p = 25$ 对应 5 s)；最后输出层以 $\Delta t = 0.2 \text{ s}$ 对速度积分，得到位置预测 $\hat{P} \in R^{N \times t_p \times 2}$ ，并据此计算各时刻 RMSE，实现从“历史轨迹 + 动态邻接”到“未来轨迹”端到端的预测映射。

4. 实验与结果分析

4.1. 数据集

所提出的模型使用公开的 NGSIM US-101 和 I-80 数据集进行训练和评估。这两个数据集均以 10 Hz 的采样率捕捉了 45 分钟的交通数据，并分为三个 15 分钟的时段。这些时段分别代表了轻度、中度和拥挤的交通状况。这两个数据集包含了真实高速公路交通中的车辆轨迹。每辆车的轨迹被分割为 8 秒的片段，其中前 3 秒作为观测轨迹历史，剩余 5 秒作为预测时间范围[14]。按照 Deo [15] 等人的方法，轨迹数据从 10Hz 下采样至 5Hz，即每秒五帧。上述两个数据集被合并为一个数据集，随机打乱后按 7:1:2 的比例划分为训练集、验证集和测试集。

4.2. 评估指标

采用与其他方法相同的评价指标，以预测未来轨迹在 5 秒预测范围内每个时间步的均方根误差 (RMSE) 作为评价结果。时间步 t 的 RMSE 计算公式如式(23)所示。

$$\text{RMSE}_t = \sqrt{\frac{1}{m} \sum_{i=1}^m (Y_{t,i}^{\text{pred}} - Y_{t,i}^{\text{gold}})^2} \quad (23)$$

其中， m 表示测试数据集中车辆的数量， $Y_{t,i}^{\text{pred}}$ 和 $Y_{t,i}^{\text{gold}}$ 分别表示时间步 t 时的预测位置和真实位置。

4.3. 对比实验

4.3.1. 基线模型

DSCAN [16]: 该方法使用注意力机制判断哪些周围车辆对目标车辆更为重要，并结合约束网络捕捉周围环境信息，综合动态和静态上下文进行轨迹预测。

GAT-LSTM: 此方法基于动作生成多模态轨迹分布, 选择概率最高的模式进行评估, 与 V-LSTM 不同, 它能够更好地模拟复杂的驾驶行为。

V-LSTM: 此方法仅利用目标车辆的轨迹特征, 不考虑周围车辆的交互, 采用基于 LSTM 的编码器-解码器结构进行轨迹预测。

TSMHA: 此方法采用卷积社交池化层, 通过空间网格捕捉目标车辆与周围车辆的交互关系, 输出单模态轨迹分布。

S-LSTM [17]: 与 V-LSTM 不同, 此方法通过全连接层将目标车辆及其周围车辆的历史轨迹特征进行融合, 从而在解码过程中建模车辆间交互。

CV: 该方法利用恒速卡尔曼滤波器, 仅依赖目标车辆的速度信息进行轨迹预测。

S-TF: 该方法采用基于 Transformer 的框架, 将目标车辆与周围车辆的时空特征作为序列输入进行编码解码, 从而对车辆轨迹进行预测。

4.3.2. 对比结果

表 1 和表 2 分别展示了各方法在 NGSIM 数据集 US-101 和 Peachtree Street 子数据集上的 RMSE 性能对比。实验结果表明, 仅依赖单车轨迹的基线方法(CV 和 V-LSTM)表现较差, 平均 RMSE 超过 3.4 m, 而融合多车辆交互信息的方法能够提升预测精度。其中, S-LSTM、GAT-LSTM 和 TSMHA 等方法通过建模邻近车辆动态关系, 将平均误差降低至 2.0~2.4 m 范围。所提出的 BiFSTNet 方法在两个数据集上均实现最优性能, 在 US-101 数据集上平均 RMSE 为 1.77 m, 在 Peachtree Street 数据集上为 2.06 m, 相比次优方法分别提升 9.7%和 7.6%。特别是在长期预测场景(4 s 和 5 s)中, BiFSTNet 表现出显著优势, 验证了仿射全注意力机制在建模多车辆长期交互关系方面的有效性。

Table 1. Experimental results on NGSIM US-101 sub-dataset

表 1. NGSIM 中 US-101 子数据集实验结果

模型	RMSE/m					平均
	1 s	2 s	3 s	4 s	5 s	
V-LSTM	0.71	1.82	3.23	4.86	7.02	3.53
S-LSTM	0.65	1.52	2.13	3.18	4.47	2.39
GAT-LSTM	0.74	1.29	2.02	2.96	4.18	2.24
TSMHA	0.56	1.18	1.94	2.78	3.76	2.05
S-TF	0.86	1.47	1.92	2.22	3.33	1.96
DSCAN	0.58	1.26	2.03	2.98	4.13	2.12
CV	0.73	1.78	3.13	4.78	6.68	3.42
BiFSTNet (Ours)	0.53	1.13	1.79	2.16	3.25	1.77

Table 2. Experimental results on NGSIM Peachtree Street sub-dataset

表 2. NGSIM 中 Peachtree Street 子数据集实验结果

模型	RMSE/m					平均
	1 s	2 s	3 s	4 s	5 s	
V-LSTM	0.81	1.96	4.27	6.41	8.34	4.36
S-LSTM	0.76	1.67	2.69	3.44	5.15	2.74
GAT-LSTM	0.93	1.59	2.47	3.19	4.64	2.56
TSMHA	0.65	1.38	2.34	2.95	4.37	2.34

续表

S-TF	0.94	1.52	2.15	2.57	3.98	2.23
DSCAN	0.64	1.47	2.36	3.18	4.45	2.42
CV	0.81	1.89	3.36	5.28	6.96	3.66
BiFSTNet (Ours)	0.61	1.33	2.06	2.46	3.75	2.06

4.4. 消融实验

为验证 BiFSTNet 各模块的有效性, 本文在 US-101 数据集上进行了消融实验, 结果如表 3 和表 4 所示。实验结果表明, 去除 ViT 模块、时间嵌入和空间图嵌入后, 模型平均 RMSE 分别从 1.772 m 上升至 2.552 m、3.020 m 和 3.086 m, 性能显著下降, 验证了各组件在多尺度特征提取和时空关系建模中的关键作用。解码器对比实验显示, 采用 RNN 和 GRU 替换 BiLSTM 后, 平均 RMSE 分别增加至 2.420 m 和 2.486 m, 特别是在长期预测(5 s)中, 误差从 3.25 m 增至 4.43 m 和 4.56 m, 表明双向 LSTM 在捕获长期时序依赖方面具有提升作用。

Table 3. Ablation study results on NGSIM US-101 sub-dataset
表 3. NGSIM 中 US-101 子数据集的消融实验结果

模型	RMSE/m					
	1 s	2 s	3 s	4 s	5 s	平均
去除 ViT 模块	0.64	1.36	2.61	3.47	4.68	2.552
去除时间嵌入	0.68	1.54	2.74	4.21	5.93	3.020
去除空间图嵌入	0.69	1.56	2.81	4.24	6.13	3.086
BiFSTNet (Ours)	0.53	1.13	1.79	2.16	3.25	1.772

Table 4. Performance comparison of different decoders on NGSIM US-101 sub-dataset
表 4. 采用不同的解码器在 US-101 子数据集对性能的影响

模型	RMSE/m					
	1 s	2 s	3 s	4 s	5 s	平均
RNN	0.59	1.29	2.47	3.32	4.43	2.420
GRU	0.60	1.32	2.56	3.39	4.56	2.486
BiLSTM (Ours)	0.53	1.13	1.79	2.16	3.25	1.772

5. 结语

本文针对复杂交通场景下车辆轨迹预测的挑战, 提出了一种基于 ViT 与仿射全注意力机制的车辆轨迹预测模型。模型通过 ViT 编码器有效提取了场景中多车辆轨迹的全局空间特征, 并通过仿射全注意力机制序贯地捕获了长程时间依赖和动态空间交互, 克服了现有方法在全局上下文提取和时空联合建模方面的不足。在 NGSIM 公开数据集上的实验结果表明, BiFSTNet 的性能优于多种基线模型。在 US-101 和 I-80 数据集上的平均 RMSE 分别达到 1.77 米和 2.06 米, 尤其在 4 s、5 s 等长期预测范围内保持了较高的精度。消融实验进一步证实了 ViT 模块、时间嵌入、空间图嵌入以及双向 LSTM 解码器各个组件对提升模型预测性能的关键作用。

参考文献

- [1] 左世全. 解读《新能源汽车产业发展规划(2021-2035)》[J]. 智能网联汽车, 2020(6): 21-23.
- [2] 张志勇, 王宇翔, 黄彩霞, 等. 融合灰色预测和卡尔曼滤波的车辆侧向碰撞预警[J]. 机械工程学报, 2024, 60(20): 240-250.
- [3] 龙皓明, 薛振锋, 陈卓, 等. 智能驾驶车辆轨迹预测方法综述[J]. 兵工自动化, 2024, 43(5): 43-52.
- [4] 李秦, 何洪文, 胡满江. 基于高斯过程回归的无人车辆轨迹跟踪模型预测控制[J/OL]. 兵工学报, 1-10. <http://kns.cnki.net/kcms/detail/11.2176.TJ.20241125.1512.002.html>, 2024-12-23.
- [5] 杨杨. 基于隐马尔科夫预测算法的无人驾驶车辆换道策略研究[D]: [硕士学位论文]. 十堰: 湖北汽车工业学院, 2021.
- [6] 龙皓明, 薛振锋, 陈卓, 等. 智能驾驶车辆轨迹预测方法综述[J]. 兵工自动化, 2024, 43(5): 43-52.
- [7] 张智阳. 基于循环神经网络的车辆轨迹预测算法研究[D]: [硕士学位论文]. 长春: 吉林大学, 2024.
- [8] Zhang, J., Wang, Z., Li, L., Yang, K. and Lu, Y. (2024) Trajectory Tracking Control of Autonomous Vehicles Based on Event-Triggered Model Predictive Control. *IET Intelligent Transport Systems*, **18**, 2856-2868. <https://doi.org/10.1049/itr2.12589>
- [9] Ge, L., Wang, S. and Wang, G. (2024) Rapid Vehicle Trajectory Prediction Based on Multi-Attention Mechanism for Fusing Multimodal Information. *Electronics*, **13**, Article 4806. <https://doi.org/10.3390/electronics13234806>
- [10] Chen, Y. and Yuhui, F. (2022) Prediction of Ship Track Based on ARIMA-LSTM. Dalian Maritime University.
- [11] Xiao, P., Xie, H. and Yan, L. (2024) Vehicle Motion Trajectory Prediction Fusion Algorithm with Driver Adventurousness Correction Factor Based on CS-LSTM. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **238**, 3541-3552. <https://doi.org/10.1177/09544070231188783>
- [12] Zhu, D., Khan, Q. and Cremers, D. (2024) Multi-Vehicle Trajectory Prediction and Control at Intersections Using State and Intention Information. *Neurocomputing*, **574**, Article 127220. <https://doi.org/10.1016/j.neucom.2023.127220>
- [13] Wang, J., Liu, K. and Li, H. (2024) LSTM-Based Graph Attention Network for Vehicle Trajectory Prediction. *Computer Networks*, **248**, Article 110477. <https://doi.org/10.1016/j.comnet.2024.110477>
- [14] Ratre, S.U. and Joshi, B. (2024) Trajectory and Motion Prediction of Autonomous Vehicles Driving Assistant System Using Distributed Discriminator Based Bi-Directional Long Short Term Memory Model. *International Journal of Intelligent Transportation Systems Research*, **23**, 245-258. <https://doi.org/10.1007/s13177-024-00447-8>
- [15] Deo, N. and Trivedi, M.M. (2018). Convolutional Social Pooling for Vehicle Trajectory Prediction. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, 18-22 June 2018, 1549-15498. <https://doi.org/10.1109/cvprw.2018.00196>
- [16] Yu, J., Zhou, M., Wang, X., Pu, G., Cheng, C. and Chen, B. (2021) A Dynamic and Static Context-Aware Attention Network for Trajectory Prediction. *ISPRS International Journal of Geo-Information*, **10**, Article 336. <https://doi.org/10.3390/ijgi10050336>
- [17] Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L. and Savarese, S. (2016) Social LSTM: Human Trajectory Prediction in Crowded Spaces. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 961-971. <https://doi.org/10.1109/cvpr.2016.110>