第 54 卷 第 4 期
2026 年 4 月

华南理工大学学报（自然科学版）
Journal of South China University of Technology
（Natural Science Edition）

Vol. 54 No. 4
April 2026

doi：10. 12141/j. issn. 1000-565X. 250314

基于时序模式分解的环形交叉口车辆轨迹预测

张建华　李玮

（东北林业大学 土木与交通学院，黑龙江 哈尔滨 150040）

摘　要：为提升环形交叉口等复杂结构化场景下的车辆轨迹预测精度，提出一种结合了多头注意力机制（MHA）、简化图卷积（SGC）网络和TimeMixer机制的深度学习模型，构建了一个宏观-微观双层编码体系。在宏观层面，采用MHA捕捉由全局道路拓扑施加的长时程战略约束，即通过建模车辆完整历史轨迹和环岛结构（如出入口的深层关系）来推断其长期行驶意图；在微观层面，首先利用SGC网络提取车辆间的即时空间关系，随后引入TimeMixer机制，将一维交互时序映射为多尺度、多分辨率的二维时空图像，通过对周期性战术行为与趋势性战略意图的显式解耦和分层融合，实现对深层交互模式的精准捕捉。两类信息经门控网络融合后，由门控循环单元解码器生成最终的轨迹。实验结果表明，在5 s的预测时间内，该文提出的模型在公开数据集INTERACTION上的平均位移误差与最终位移误差分别为1. 19和1. 85 m，在公开数据集RounD上则分别为1. 16和1. 80 m，均优于基线模型，证明通过对宏观全局约束与微观时空交互进行分层建模，特别是对交互模式进行解耦分析，能有效提升模型在复杂场景下的轨迹预测性能。

关键词：环形交叉口；车辆轨迹预测；多头注意力；简化图卷积；TimeMixer

中图分类号：U461；TP39　　**文章编号：**1000-565X（2026）04-0132-12

随着智能交通系统和车辆技术的飞速发展，自动驾驶已成为当前的研究热点[1-2]。在自动驾驶系统中，对周围车辆未来运动轨迹的精准预测，是实现安全、可靠的自主决策与规划的关键前提[3-4]。然而，在环岛、交叉口等交通流交织的复杂场景中，车辆轨迹预测面临着巨大挑战[5-6]。特别是无信号环岛，其独特的道路结构和复杂的车辆交互规则，使得预测任务更加艰巨[7]。

传统的预测方法（如基于物理模型或经典机器学习的方法）在处理长期预测和复杂的交互场景时存在局限性[8]。近年来，基于深度学习的方法展现出了强大的能力，但仍面临诸多挑战[9]。首先，仅依赖车辆自身历史轨迹信息不足以做出准确预测，必须充分考虑高清地图提供的车道、路口等环境上下文信息以及与其他交通参与者的交互[10]。现有深度学习方法主要分为两类：基于栅格化的方法和基于矢量化的方法[11]。栅格化方法虽然直观，但基于卷积神经网络（CNN）的模型因其感受野有限，在提取长距离的几何特征方面存在困难[12]，因此，矢量化方法逐渐成为主流。但现有的矢量化方法在信息提取的全面性以及交互建模的有效性上仍有提升空间，尤其是在处理场景中众多实体间的复杂关系方面[13-14]。

收稿日期：2025-09-05

基金项目：黑龙江省重点研发计划项目（JD22A014）；黑龙江省自然科学基金项目（YQ2022E003）

Foundation items：Supported by the Key Research and Development Program of Heilongjiang Province（JD22A014）and the Natural Science Foundation of Heilongjiang Province（YQ2022E003）

作者简介：张建华（1973—），男，博士，副教授，主要从事智能交通研究。E-mail：jhzhang609@163. com

为应对上述挑战，本文提出了一种分层处理框架，即将预测问题分解为宏观和微观两个层面进行建模。在宏观层面，重点是建模环岛全局结构对车辆长时程战略意图的引导性影响。一辆车的路径不仅是对邻近车辆的即时反应，同时也受到其入口和预期出口等宏观拓扑的约束。多头注意力机制(MHA)尤其适用于此任务，因为它能够关联车辆历史轨迹中的早期位置信息和全局道路结构，从而推断出贯穿整个行驶过程的战略规划[15]。在微观层面，核心是建模车辆之间的局部时空交互。对于空间特征提取，基于图的神经网络能有效刻画车辆间的拓扑关系，其中简化图卷积(SGC)网络通过简化结构，突出了特征聚合的核心作用，兼具效率与性能[6]。对于微观层面的时间维度动态交互，现有研究多采用循环神经网络(RNN)或Transformer变体(如Informer)进行建模[17]。然而，这些方法在处理时间序列时存在固有挑战。例如，RNN难以捕捉长期依赖，而Transformer类模型虽然通过自注意力机制克服了这一缺点，但在处理像车辆交互这样包含多种重叠时间模式(如周期性的跟驰、趋势性的加速)的数据时，难以将这些模式有效表达为独立的“令牌”，从而限制了其捕捉复杂时间结构的能力[18-19]。

鉴于此，需要一种新的范式，它不仅能处理长期依赖，还能显式地分离和理解交织在一起的多种时间模式。为了突破这一瓶颈，文中引入文献[20]中关于TimeMixer机制的通用时间序列模式机核心思想。该思想主张将一维时间序列转化为多尺度、多分辨率的二维时间图像，并在隐空间内对其进行季节性和趋势性的显式分解与混合。这种方法能够从时–频联合域角度更深刻地理解时间序列的内在模式。基于此，文中提出了一种结合MHA、SGC网络和TimeMixer机制的深度学习模型(MHA-SGC-TimeMixer，简记为MST)，专用于环岛场景下的车辆轨迹预测。该模型由4个核心模块组成。

模块1　宏观编码器　利用多头注意力机制处理车辆与环岛全局结构信息的交互。

模块2　微观编码器　采用SGC网络提取车辆间的空间局部特征，并结合TimeMixer机制将交互的时序动态分解为周期性与趋势性模式，以捕捉深层交互规律。

模块3　宏–微观融合层　通过门控机制整合宏观和微观信息。

模块4　轨迹解码器　使用门控循环单元(GRU)生成最终的预测轨迹。

本研究旨在通过对全局路径约束和局部时空交互，特别是交互模式的分解与融合的有效建模，在复杂的环岛场景中实现更为精准、可靠的轨迹预测。

1　问题描述

本研究的核心问题是：在环形交叉口这一典型复杂场景下，对目标车辆的未来行驶轨迹进行精准、可靠的预测。此项任务是实现安全、可靠的自主决策和规划的关键前提。

形式上，将车辆轨迹预测问题定义为一个序列到序列的学习任务。给定观测时刻t之前一段时间内场景中所有交通参与者的历史状态信息及环境信息，目标是预测目标车辆i从时刻t+1到未来预测时域t_p的轨迹序列z_t^i：

$$z_t^i=[(\hat{x}_{t+1}^i,\hat{y}_{t+1}^i),(\hat{x}_{t+2}^i,\hat{y}_{t+2}^i),\cdots,(\hat{x}_{t+t_p}^i,\hat{y}_{t+t_p}^i)] \tag{1}$$

式中，$(\hat{x},\hat{y})$为车辆在笛卡尔坐标系下的预测位置。

为了实现这一目标，模型必须充分理解并建模复杂的输入信息。在环岛场景中，车辆的行为不仅受邻近车辆的即时影响，更受到其入口、预定出口及环岛整体几何结构的全局引导。因此，将时刻t的观测信息集合C_t定义为一个由宏观状态与微观状态组成的集合：

$$C_t=[C_t^M,C_t^m]。$$

式中：$C_t^M=\{c_t^{M1},c_t^{M2},\cdots,c_t^{MN}\}$表示$N$个车辆在时刻$t$的宏观历史信息，$c_t^{Mi}$包含了车辆$i$在过去$t_h$时间步内与环岛全局结构的交互信息序列，用以推断其长时程战略意图；$C_t^m=\{c_t^{m1},c_t^{m2},\cdots,c_t^{mN}\}$表示$N$个车辆在时刻$t$的微观历史信息，$c_t^{mi}$包含了车辆$i$在过去时间步内的自身运动学信息以及与其他车辆的相对时空关系，用以捕捉其即时战术行为。

对宏观和微观信息的分层建模旨在使模型能够同时理解车辆的长期战略规划和短期战术决策，从而解决环岛场景下的轨迹预测难题。

2　车辆轨迹预测模型的构建

文中提出的结合多头注意力的轨迹预测基于编码器–解码器机制，其关键组件包括：基于MHA的宏观编码层，用于实现环岛结构与车辆之间的特征交互；基于SGC网络和TimeMixer机制的微观编

码层，用于捕获目标车辆与选定的重要实体之间的全局交互特征；宏-微观融合层，用于整合宏观和微观特征；基于GRU的轨迹解码器。所提出的模型框架如图1所示。

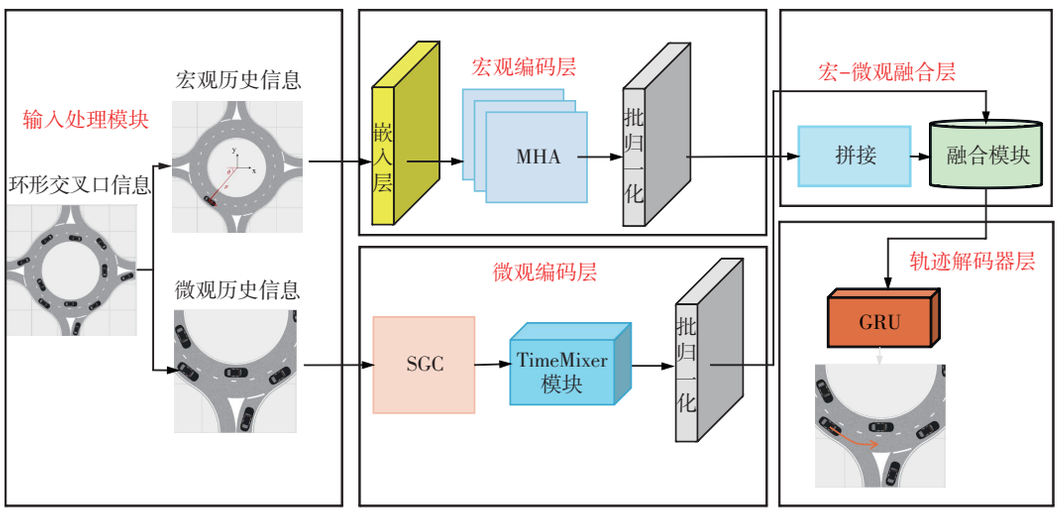


图1 基于时序模式分解的轨迹预测框架

Fig. 1 Trajectory prediction framework based on time series pattern decomposition

2.1 建模思路与框架设计

为应对环岛场景下车辆行为的复杂性，文中建模的核心思路是采用一种分层处理框架，将复杂的驾驶行为分解为由宏观全局路径引导的战略意图和由微观局部交互驱动的战术决策两个层面，并对其进行针对性建模。

从宏观层面战略意图的角度而言，车辆在环岛内的行驶路径并非随机，而是强力地受其入口和目标出口的约束。这种长时程的路径规划意图适合采用能捕捉序列中长距离依赖关系的模型来提取，因此，选用MHA作为宏观编码器。MHA能够将车辆当前状态与历史轨迹的早期位置信息、全局道路拓扑相关联，从而有效推断其贯穿全程的行驶战略。

从微观层面战术决策的解耦角度而言，车辆间的局部交互是高度动态和非线性的，因此，进一步将其分解为空间交互和时间动态两个维度进行建模。

对于空间交互，车辆间的邻接关系每个时刻都在变化，构成了一个动态图，文中采用SGC网络来高效地捕捉这种瞬时空间关系。选择SGC网络而非更复杂的图神经网络，旨在保证有效性的前提下追求计算效率。SGC网络通过将多层非线性图卷积网络(GCN)折叠为单次线性变换，能快速地将非结构化的场景信息转化为结构化的空间邻接关系，为后续时序分析提供清晰的输入。

对于时间动态，车辆间的交互(如跟驰、让行、博弈)包含多种交织的时间模式。传统时序模型(如RNN)难以捕捉长期依赖，而Transformer类模型也难以将这些重叠的模式有效分离。文中创新性地引入TimeMixer机制，其核心优势在于能将一维的时序交互序列映射为二维图像，并显式地将周期性的战术行为与趋势性的战略意图进行解耦。这种分解能够让模型更深刻地理解交互的内在逻辑，例如，区分开一次短暂的减速让行这一战术行为和一次为驶离环岛而进行的持续加速的战略意图。

2.2 输入处理模块

如图2所示，为更直观地反映环岛全局地图与车辆空间特征的关系，文中从鸟瞰视角重建环岛。环岛的每个入口由相应的二进制代码p_i表示。由于环岛内单向通行的规则，所选车辆在时刻t的位置可以通过其初始入口的二进制代码、经过的弧度θ以及与环岛中心的距离ρ来表示。对于环岛结构的静态信息，车辆驶入的初始入口定义为p_0，其余入口按逆时针方向定义为p_i。值得注意的是，车辆的

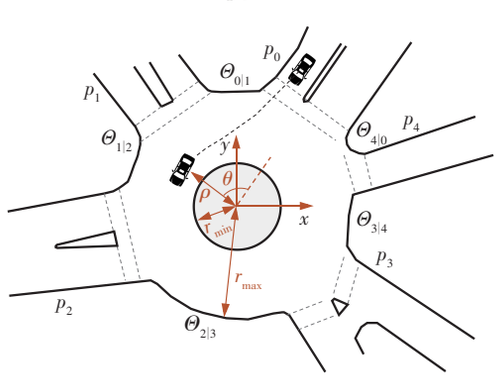


图2 环岛场景中宏观状态的定义规则

Fig. 2 Rules for defining the macro state in roundabout scenarios

宏观状态(I_i,θ_i,ρ_i)会随其在环岛内位置的变化而实时更新，其中θ_i和ρ_i为变量，而初始入口I_i作为该轨迹的静态属性保持不变。

此外，环岛的内外直径分别表示为$r_{\min}$和$r_{\max}$，p_i与p_0之间的角度表示如下：

$$\boldsymbol{\Theta}_i=\sum_{j=0}^{i-1}\boldsymbol{\Theta}_{j|j+1} \tag{2}$$

式中，$\boldsymbol{\Theta}_{j|j+1}$为相邻入口之间的角度。

基于上述规定，宏观编码器层的最终输入格式可以表示为

$$s_t^{Mi}=\left[p_0^i\left\|\rho_t^i\right\|\Delta\rho_t^i\left\|\theta_t^i\right\|\Delta\theta_t^i\right\|r_{\min}\left\|r_{\max}\right\|\boldsymbol{\Theta}_{0\text{-}S}\right] \tag{3}$$

式中，S为环岛入口的数量。

2.3　宏观、微观编码层

2.3.1　宏观编码层

考虑到广泛使用的RNN方法无法提取每一帧对未来轨迹的影响，文中采用MHA作为宏观时间历史编码器来捕获长距离时间依赖性。相应的编码器设计如下：

$$f_t^i=\mathrm{MHA}_{M_\mathrm{temp}}(c_t^{Mi}) \tag{4}$$

式中，f_t^i表示车辆i在时刻t的宏观特征，$\mathrm{MHA}_{M_\mathrm{temp}}$表示其对应的编码器。

MHA的核心组件包括若干查询Q_i、键K_i和值V_i，其中Q_i和V_i对应当前帧，K_i对应整个历史轨迹的帧。

之后，基于缩放点积注意力机制计算头i的注意力Z_i：

$$Z_i=\mathrm{Softmax}\left(\frac{Q_iK_i^{\mathrm{T}}}{\sqrt{d_k}}\right)V_i \tag{5}$$

式中，$d_k=d_{\mathrm{model}}/h$，为K_i的维度，d_{model}为宏观历史信息c_t^{Mi}的维度，h为注意力头的数量。

最后，这些独立的注意力将被连接起来并乘以权重矩阵$\boldsymbol{W}^o$以获得基于多头注意力的输出f：

$$f=\mathrm{MHA}(Q,K,V)=\mathrm{Concat}(H_1,\cdots,H_h)\boldsymbol{W}^o \tag{6}$$

式中，H_i=Z_i，Concat表示连接操作。

2.3.2　微观编码层

微观编码器层包含微观时间历史编码器和微观空间交互编码器。

微观空间交互编码器采用SGC网络，这是一种简单而有效的数据编码方法。具体而言，为了减轻GCN的过度复杂性，SGC网络迭代地消除了GCN层之间的非线性，并将生成的函数折叠成单个线性变换。这种简化显著加快了处理时间，同时仍能保持与GCN相当的性能。如图3所示，SGC网络将整个过程简化为一个简单的特征传播步骤，随后经标准逻辑回归。在文中框架内，场景图是在每个历史时间步上动态构建的。为优先保证计算效率，采用了一种简化的同构图结构。具体而言，如果任意两个实体(如车辆–车辆、车辆–车道)之间的欧氏距离低于预设阈值，就在它们之间创建一条无向边。因此，邻接矩阵$\boldsymbol{A}$是对称且无权的，仅表示实体间是否存在空间邻近关系。该设计简化了模型，没有显式定义不同类型的边，也不引入交通规则或先行权等方向性或语义信息，其核心在于实现模型内部的功能解耦。在此架构中，SGC模块被定位为一个高效的前端场景解析器，其首要职责是在每个时间步将非结构化的场景信息转化为结构化的空间邻接关系。相较于能编码复杂语义先验的异构图，这种简化的图结构已足以向下游的时序动态分析模块提供清晰的交互序列。这使得TimeMixer模块能够完全专注于其核心任务——即对复杂的时序模式进行多尺度分解与融合，从而在保证特征提取有效性的前提下，避免模型复杂度的冗余增加。在明确了图的拓扑结构后，需要进一步定义图节点所承载的时

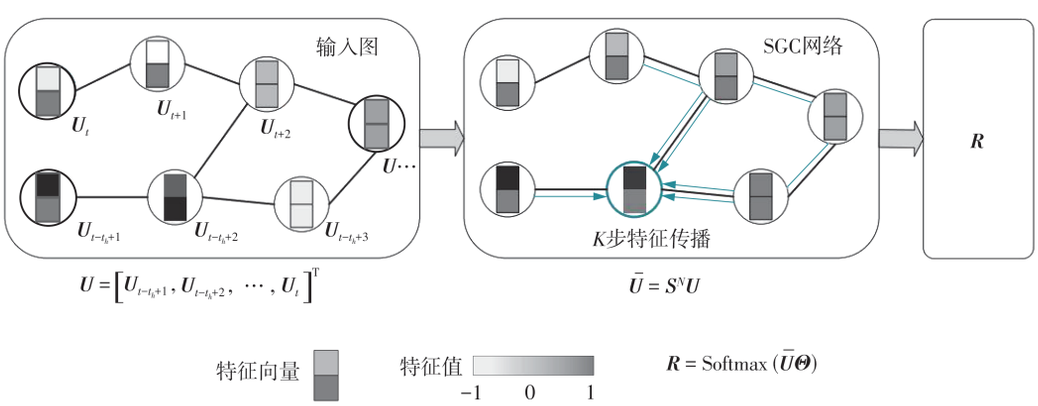


图3　基于SGC网络的局部特征提取过程

Fig. 3　local feature extraction process based on SGC network

空信息。具体而言，场景图中各实体在t时刻的输入特征定义为

$$\boldsymbol{U}=[\boldsymbol{U}_{t-t_h+1},\boldsymbol{U}_{t-t_h+2},\cdots,\boldsymbol{U}_t]^{\mathrm{T}} \tag{7}$$

式中，$\boldsymbol{U}_t$为实体在t时刻的特征向量，即目标车辆的历史轨迹U_t^{TV}、周围车辆的历史轨迹U_t^{SV}以及人行横道信息m_t^{CW}、车道信息m_t^{Lane}的向量化表示。

而后，SGC网络第N层的传播模式可以表示如下：

$$\bar{\boldsymbol{U}}=\boldsymbol{S}^N\boldsymbol{U} \tag{8}$$

式中，$\boldsymbol{S}^N$为归一化邻接矩阵$\boldsymbol{S}$重复多次相乘形成的单个矩阵，

$$\boldsymbol{S}=\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tag{9}$$

其中，$\tilde{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{J}$，$\boldsymbol{A}$为图的对称邻接矩阵，$\boldsymbol{J}$为单位矩阵。此外，$\tilde{\boldsymbol{D}}$为$\tilde{\boldsymbol{A}}$的度矩阵。

最后，实体$\boldsymbol{U}$的局部特征向量$\boldsymbol{R}$表示如下：

$$\boldsymbol{R}=\mathrm{Softmax}(\bar{\boldsymbol{U}}\boldsymbol{\Theta}) \tag{10}$$

式中，$\boldsymbol{\Theta}$为一个可学习的权重矩阵，用于将聚合后的特征进行线性变换。

Softmax函数在这里的作用是对特征进行归一化，最终得到的$\boldsymbol{R}$是每个实体的局部特征向量。

此外，局部特征提取模块的输出$\boldsymbol{P}$是每个实体局部特征的集合：

$$\boldsymbol{P}=\{R^{\mathrm{TV}},R^{\mathrm{SV}_1},\cdots,R^{\mathrm{SV}_n},R^{\mathrm{CW}_1},\cdots,R^{\mathrm{CW}_i},R^{\mathrm{Lane}_1},\cdots,R^{\mathrm{Lane}_i}\} \tag{11}$$

式中，R^{TV}、R^{SV_n}、R^{CW_i}和R^{Lane_i}分别为目标车辆、第n个周围车辆、第i个人行横道和第i条车道的局部特征，n和i分别表示周围车辆和道路的数量。

在微观编码层，文中首先在每个历史时间步t上构建场景图，并利用SGC网络提取该时刻的空间交互特征，将所有历史帧的空间特征整合成一个特征序列。

为深度挖掘车辆间交互时序中隐含的复杂动态模式，文中创新性地引入TimeMixer机制作为微观时序动态编码器。该编码器旨在将SGC网络输出的一维交互特征序列$\boldsymbol{P}_{\mathrm{hist}}=\{R_{t-t_h+1},\cdots,R_t\}$从传统的时域提升至一个多尺度、多分辨率的时-频联合域进行分析，并对其内在模式进行显式解耦与分层融合。该过程遵循一个精密的4阶段架构，如图4所示。

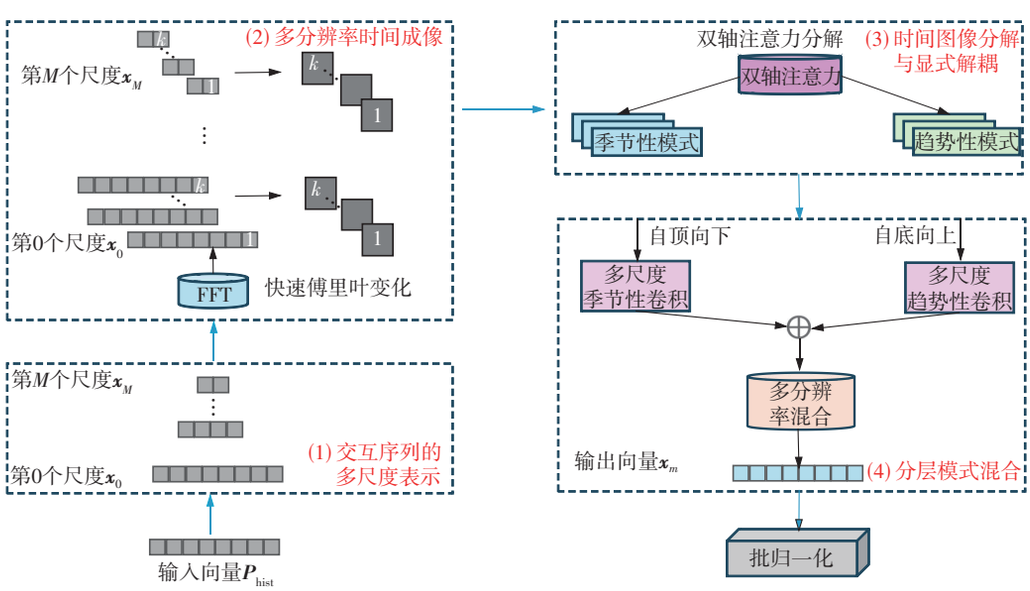


图4 TimeMixer机制架构图

Fig. 4 Architecture of TimeMixer mechanism

(1)交互序列的多尺度表示

时间序列的真实世界过程具有多尺度特性，例如，车辆交互既包含秒级的快速反应，也包含跨越数秒的策略性调整。为捕捉不同观测粒度下的动态特性，将SGC网络输出的交互序列$\boldsymbol{P}_{\mathrm{hist}}$(视为初始尺度$\boldsymbol{x}_0$)通过一系列步长为2的一维卷积进行递归下采样，从而构建出一个多尺度的特征金字塔$\boldsymbol{X}_{\mathrm{init}}=\{\boldsymbol{x}_0,\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M\}$，其中，$\boldsymbol{x}_m\in\mathbf{R}^{\lfloor T/2^m\rfloor\times C}$代表在第$m$个尺度上的交互特征序列。这种多尺度设计使得模型能够同时在宏观(粗尺度)和微观(细尺度)层面审视交互动态。

(2)多分辨率时间成像

为了将隐含在时序中的周期性规律显式化，文中借鉴了多分辨率分析的思想。首先，对最具全局

视野的最粗尺度序列$\boldsymbol{x}_M$应用快速傅里叶变换(FFT)，以识别其频谱中能量最集中的Top-K个主导频率$\{f_1,\cdots,f_K\}$，这些频率对应了交互行为中最核心的几种周期模式$\{P_1,\cdots,P_K\}$。随后，对于每一个尺度的交互序列$\boldsymbol{x}_m$，都依据这K个核心周期，通过填充与重塑操作，将其转换为K个二维的时间图像$\boldsymbol{z}_m^{(k)}\in\mathbf{R}^{p_k\times f_{m,k}\times d_{\text{model}}}$。在此二维表示中，行代表跨周期的同一时间步，而列代表一个周期内的连续时间步。通过多分辨率时间成像这一过程，一维非结构化的时序交互被转化为二维结构化的图像，这不仅使得周期性模式在空间上变得明确，也为后续应用强大的二维卷积算子进行特征提取奠定了基础。

(3)时间图像分解与显式解耦

车辆间的交互本质上是多种驾驶行为模式的叠加。TimeMixer机制的核心创新，在于借鉴经典时间序列分解理论，通过一种数据驱动的方式在深度特征空间内实现显式解耦。通过在第2步构建的结构化时间图像上应用双轴注意力机制，文中将混合的交互模式强制分离为2个正交的、可解释的组成部分——周期性战术行为与趋势性战略意图，其数学本质是通过施加正交的结构性先验，强迫模型学习解耦的特征表示：

$$s_m^{(l,k)}=Z_{\text{col}}(Q_{\text{col}},K_{\text{col}},V_{\text{col}}) \tag{12}$$

$$t_m^{(l,k)}=Z_{\text{row}}(Q_{\text{row}},K_{\text{row}},V_{\text{row}}) \tag{13}$$

式中，沿时间图像列的注意力Z_{col}捕捉周期内部的高频重复性规律，在物理上对应车辆间的周期性战术行为，如观察-等待-通行，即季节性模式$s_m^{(l,k)}$。沿行的注意力Z_{row}则捕捉长期的、平滑的低频演化趋势，对应车辆的趋势性战略意图(如为驶离环岛而持续加速)，即趋势性模式$t_m^{(l,k)}$。这种基于正交注意力轴的架构设计施加了强大的归纳偏置，将交互行为在特征层面强制分解，为后续更精准的预测提供了清晰、无干扰的特征输入。

(4)分层模式混合

最后，分解出的多尺度、多分辨率模式需要被有效地整合，以形成一个全面的交互理解。为此，设计了一种分层混合策略来融合不同尺度间的模式信息。

①多尺度混合

对于季节性战术模式$\{s_m^{(l,k)}\}_{m=0}^{M}$，模型采用自底向上的信息流，通过2D卷积网络将细粒度尺度的特征逐层聚合到粗粒度尺度上，其过程可形式化为

$$s_m^{(l,k)}=s_m^{(l,k)}+\text{2D-Conv}(s_{m-1}^{(l,k)}),\quad m=1,\cdots,M \tag{14}$$

对于趋势性战略意图$\{t_m^{(l,k)}\}_{m=0}^{M}$，模型则采用自顶向下的信息流，通过2D转置卷积网络将粗粒度尺度的全局意图逐层向下传递，以指导和修正细粒度尺度的行为预测：

$$t_m^{(l,k)}=t_m^{(l,k)}+\text{2D-TransConv}(t_{m+1}^{(l,k)}),\quad m=M-1,\cdots,0 \tag{15}$$

②多分辨率混合

在每个尺度内部，不同周期的模式对最终预测的贡献是不同的。模型利用FFT分析得到的各频率振幅A_{f_k}作为其重要性的自然度量，通过Softmax函数归一化后，对K个分辨率的聚合特征进行自适应加权求和，得到该尺度最终的特征表示x_m^l：

$$\{\hat{A}_{f_k}\}_{k=1}^{K}=\text{Softmax}(\{A_{f_k}\}_{k=1}^{K}) \tag{16}$$

$$x_m^l=\sum_{k=1}^{K}\hat{A}_{f_k}\circ z_m^{(l,k)} \tag{17}$$

经过上述4个阶段的处理，模型最终输出一个深度融合了多尺度、多周期交互模式的微观动态特征向量$\boldsymbol{E}$，该向量为后续的轨迹解码器提供了前所未有的丰富信息。

2.4　融合层与解码器

文中应用基于门控机制的融合层来平衡宏观特征$\mathcal{F}_t$和微观空间特征$\mathcal{U}_t$、微观时间特征$\mathcal{Q}_t$的影响。具体来说，目标车辆的宏观和微观输出将分别乘以权重矩阵$\boldsymbol{W}_M$和$\boldsymbol{W}_m$，即

$$[r_M,r_m]=\text{Softmax}\left\{\left[(\mathcal{F}_t\|\mathcal{Q}_t)\boldsymbol{W}_M+b_M\right]\|\left[\mathcal{U}_t\boldsymbol{W}_m+b_m\right]\right\} \tag{18}$$

式中，$\boldsymbol{W}_M$、$\boldsymbol{W}_m$、b_M、b_m为融合层中的学习参数，r_M、r_m为门控系数，将在下面的公式中使用。

最终得到车辆i在时刻t的融合特征h_t^i：

$$h_t^i=r_M\odot\left(\mathcal{F}_t\|\mathcal{Q}_t\right)+r_m\odot\mathcal{U}_t \tag{19}$$

为简化模型，在预测器建模中应用基于GRU的解码器来生成未来轨迹。考虑到车辆异构性对轨迹预测的影响，具体的输入输出关系如下式所示：

$$z_t^i=\text{GRU_Dec}_{\text{pred}}(h_t^i) \tag{20}$$

3　数据与实验

3.1　数据集介绍

为验证所提模型的有效性，采用2个大型的、公开的、以环形交叉口为核心场景的车辆轨迹数据集(INTERACTION和RounD)进行实验。

INTERACTION数据集是一个大规模、国际化的运动数据集，专注于研究自动驾驶中复杂的多智能体交互行为。其核心特点在于捕捉了高密度、

高度交互性的交通流，其中不仅包含了常规驾驶行为，还特别关注了具有挑战性和对抗性的驾驶行为(如抢道、不规范超车)和协作性驾驶互动(如礼让、交替通行)。此外，INTERACTION数据集提供了高精度的语义地图，为轨迹预测、行为理解、决策规划等下游研究任务提供了丰富的环境上下文信息。

RounD数据集是另一个专门针对环岛场景的大型自然驾驶行为数据集。该数据集利用无人机在德国的多个真实环岛上空进行航拍采集，获得了高分辨率的鸟瞰视角视频，再通过先进的计算机视觉算法，从中提取出了精确的车辆轨迹、速度、加速度以及尺寸等详细数据。RounD数据集的独特价值在于其真实性和场景的专注性，它完整地记录了无信号控制下各类车辆在进入、环绕和驶出环岛过程中的真实交互行为，是评估环岛场景下轨迹预测模型性能的理想基准。

INTERACTION和RounD数据集的共同特点在于捕捉了真实世界中从交通流畅到因车辆交织而趋于饱和的多种交通状态，这些富含复杂博弈行为的数据为模型训练提供了坚实基础，也决定了文中所提MST模型的最适应用场景——交通流密度中等偏高至接近饱和的环形交叉口。在这一场景下，车辆间交互频繁且行为决策链条复杂，恰好能最大程度地发挥模型对短期战术行为(如让行)与长期战略意图(如选择出口)进行解耦的预测能力。至于极端稀疏的自由流或完全拥堵的状态，由于车辆交互模式简化，因此其并非本模型旨在解决的核心问题。

研究中对2个数据集均进行了统一处理。参考前人工作，将原始数据集进行降采样后分割，划分为训练集(占比70%)、验证集(占比10%)和测试集(占比20%)。实验中统一选择2 s的历史轨迹信息来预测未来1~5 s的目标车辆轨迹。

3.2 参数配置

实验在NVIDIA GeForce RTX 2080 Ti上进行，代码在PyTorch深度学习框架下实现模型的主体框架，与图相关的具体细节则使用PyTorch Geometric实现。模型的输入是多智能体的历史状态信息序列，输出则是包含预测时间步长在内的未来轨迹。实验中的超参数设置如下：①宏观时间历史编码器中，MHA内的嵌入大小设置为64，注意力头设置为4；②微观空间交互编码器中，SGC网络层数设为3，隐藏单元固定为64个；③微观时间历史编码器中，对于TimeMixer机制，MixerBlock层数设为2，尺度数设为3，Top-K分辨率设为3；④基于GRU的未来轨迹解码器中，隐藏单元设为64，层数设为4。ReLU作为激活函数。

此外，该模型共训练50个epoch，使用64个批大小。使用AdamW优化器并初始化学率，学习率为0.000 5。

3.3 模型误差对比

根据现有工作惯例，使用平均位移误差(ADE，以E_{AD}表示)和最终位移误差(FDE，以E_{FD}表示)作为模型预测结果的衡量标准，其定义分别如式(21)和式(22)所示。

$$E_{\mathrm{AD}i}=\frac{1}{t_{\mathrm{p}}}\sum_{t=t_{\mathrm{obs}}+1}^{t_{\mathrm{obs}}+t_{\mathrm{p}}}\sqrt{(\hat{x}_t^i-x_t^i)^2+(\hat{y}_t^i-y_t^i)^2} \tag{21}$$

式中，t_{p}为预测的总时间步数，$(\hat{x}_t^i,\hat{y}_t^i)$为智能体$i$在时间步$t$的预测位置，$(x_t^i,y_t^i)$为智能体$i$在时间步$t$的真实位置，$t_{\mathrm{obs}}$为观测结束的时间步。

$$E_{\mathrm{FD}i}=\sqrt{(\hat{x}_{t_{\mathrm{obs}}+t_{\mathrm{p}}}^i-x_{t_{\mathrm{obs}}+t_{\mathrm{p}}}^i)^2+(\hat{y}_{t_{\mathrm{obs}}+t_{\mathrm{p}}}^i-y_{t_{\mathrm{obs}}+t_{\mathrm{p}}}^i)^2} \tag{22}$$

式中，$(\hat{x}_{t_{\mathrm{obs}}+t_{\mathrm{p}}}^i,\hat{y}_{t_{\mathrm{obs}}+t_{\mathrm{p}}}^i)$为智能体$i$在预测终点时刻$t_{\mathrm{obs}}+t_{\mathrm{p}}$的预测位置，$(x_{t_{\mathrm{obs}}+t_{\mathrm{p}}}^i,y_{t_{\mathrm{obs}}+t_{\mathrm{p}}}^i)$为智能体$i$在预测终点时刻$t_{\mathrm{obs}}+t_{\mathrm{p}}$的真实位置。

为了验证所提出的MST模型的有效性，计算该模型和其他6个经典模型(CVM、LR、P-LSTM、S-GAN、Trajectron、HiVT)在1~5 s的预测时间内的预测坐标与真实坐标之间的平均位移误差和最终位移误差。表1和表2分别示出了不同模型在

表1 不同模型在INTERACTION数据集上的平均位移误差与最终位移误差对比

Table 1 Comparison of ADE and FDE for different models on INTERACTION dataset

预测时间/s	不同模型的E_{AD}与E_{FD}/m						
	CVM	LR	P-LSTM	S-GAN	Trajectron	HiVT	MST
1	0.16/0.29[1)]	0.24/0.32	0.23/0.28	0.22/0.28	0.22/0.27	0.20/0.25	0.16/0.23
2	0.59/0.78	0.58/0.92	0.47/0.60	0.42/0.58	0.38/0.41	0.29/0.38	0.24/0.35
3	1.21/1.92	1.43/2.28	0.84/1.53	0.81/1.54	0.62/1.09	0.55/0.95	0.49/0.90
4	2.94/3.98	3.85/4.73	1.27/1.51	1.28/1.87	0.81/1.26	0.70/1.20	0.65/1.19
5	4.28/6.12	5.89/6.91	1.78/2.21	1.65/2.68	1.36/2.02	1.22/1.85	1.19/1.85

1) 斜线前为E_{AD}，斜线后为E_{FD}，表2和表3同。

表2　不同模型在RounD数据集上的平均位移误差与最终位移误差对比

Table 2　Comparison of ADE and FDE for different models on RounD dataset

预测时间/s	不同模型的E_{AD}与E_{FD}/m						
	CVM	LR	P-LSTM	S-GAN	Trajectron	HiVT	MST
1	0. 27/0. 38	0. 36/0. 42	0. 34/0. 41	0. 31/0. 38	0. 27/0. 31	0. 26/0. 32	0. 26/0. 29
2	0. 65/0. 87	0. 64/1. 00	0. 57/0. 73	0. 47/0. 58	0. 40/0. 47	0. 35/0. 47	0. 30/0. 46
3	1. 27/1. 99	1. 48/2. 30	0. 90/1. 63	0. 81/1. 56	0. 52/0. 99	0. 49/0. 90	0. 47/0. 89
4	3. 01/4. 05	3. 88/4. 77	1. 29/1. 52	1. 26/1. 85	0. 84/1. 24	0. 70/1. 22	0. 67/1. 18
5	4. 35/6. 29	5. 95/6. 87	1. 83/2. 24	1. 69/2. 77	1. 30/1. 95	1. 16/1. 82	1. 16/1. 80

INTERACTION数据集和RounD数据集上的平均位移误差和最终位移误差。

1）恒定速度模型（CVM）　该模型将智能体的速度设为常数。

2）线性回归（LR）模型　该模型使用线性回归方法预测智能体的状态信息。

3）概率性LSTM（P-LSTM）模型　该模型结合了长短期记忆网络（LSTM）与概率分布输出，能够预测多智能体未来轨迹的概率分布，从而量化预测中的不确定性。

4）社交GAN（S-GAN）模型　该模型通过引入生成对抗网络机制，能够考虑交通参与者之间的社交交互，从而生成符合社会规范的多模态未来轨迹。

5）Trajectron模型　该模型是一种图结构化的递归模型，它通过结合智能体的运动和异构的场景数据来预测轨迹。

6）HiVT模型　该模型是一个分层的矢量化Transformer模型，它通过“先局部后全局”的注意力机制，在保证精度的同时极大地提高了运算效率。

实验结果清晰地表明，文中提出的MST模型在预测时间（1~5 s）内的ADE和FDE均显著优于基线模型，展现了卓越的预测精度和鲁棒性。传统的基于规则的CVM和LR模型由于其线性和匀速假设，无法捕捉环岛场景中复杂的非线性交互与变速行为，因此在长期预测中误差迅速累积，性能最差，这凸显了高度动态交互场景中基于学习的方法的不可替代的优势。主流深度学习模型P-LSTM、S-GAN等虽引入了深度学习，但仍存在局限性，特别是S-GAN，其采用的基于欧氏距离的社交池化机制，难以有效表达环岛内由道路曲率和交通规则共同决定的非各向同性空间关系，因此性能提升有限。先进基线模型Trajectron、HiVT代表了当前处理此类问题的先进水平，但文中提出的MST依然实现了更优的性能。这一核心优势源于两者在微观时序动态建模范式上的根本差异。传统时序模型将交互序列视为离散令牌，在单一时域内进行注意力计算；而MST模型通过多分辨率时间成像，将交互动态提升至时–频联合域进行结构化分析，使潜在的周期性模式得以显式表达。更重要的是，时间图像分解模块能够将战术性避让（季节性）与战略性加速（趋势性）等不同性质的驾驶行为进行有效解耦，避免了单一注意力机制下不同意图的相互干扰。最终，通过多尺度混合的分层聚合，模型得以构建对短期反应和长期规划的综合理解，从而在长时程预测上展现出卓越的鲁棒性和准确性。

3.4　消融实验

为系统性验证MST框架中各核心模块的不可或缺性，并探究其架构设计的内在逻辑与协同效应，设计了一系列严格的消融实验，通过选择性地移除或单独评估模型的关键组件，定量剖析它们各自对最终预测性能的贡献。实验在INTERACTION及RounD数据集上进行，结果如表3所示。

表3　消融实验结果

Table 3　Results of ablation experiments

模型	不同时间下INTERACTIONS上的E_{AD}与E_{FD}/m			不同时间下RounD上的E_{AD}与E_{FD}/m		
	1 s	3 s	5 s	1 s	3 s	5 s
只使用宏观时间编码器MHA	0. 38/0. 42	1. 01/1. 75	2. 36/2. 95	0. 40/0. 45	0. 95/1. 63	2. 15/2. 84
只用微观历史编码器TimeMixer	0. 31/0. 36	0. 78/1. 40	1. 61/2. 26	0. 33/0. 39	0. 75/1. 36	1. 54/2. 18
MHA-TimeMixer	0. 25/0. 31	0. 68/1. 28	1. 53/2. 20	0. 28/0. 33	0. 66/1. 26	1. 49/2. 15
SGC-TimeMixer	0. 16/0. 22	0. 53/1. 05	1. 33/1. 99	0. 24/0. 27	0. 53/1. 05	1. 34/1. 87
MHA-SGC-TimeMixer	0. 16/0. 23	0. 49/0. 90	1. 19/1. 85	0. 26/0. 29	0. 47/0. 89	1. 16/1. 80

表3结果揭示了文中模型宏观引导、微观感知、时空解耦的有效性，其性能并非源于单一组件的堆砌，而是源于一个功能高度特性化且互补的层次化推理体系。

首先，实验结果表明，直接对车辆间的动态博弈进行建模的微观信息，是整个轨迹预测任务的核心驱动力。一个完全忽略局部交互、仅依赖全局上下文的纯宏观模型，如只使用宏观时间编码器MHA的模型，其性能在所有变体中最差。与之相反，一个仅包含微观编码器的纯交互模型SGC-TimeMixer，其性能远优于前者，这证明对高频、非线性的车辆间动态进行精确捕捉是预测问题的关键所在。

然而，对微观交互的有效建模，其前提在于对每一时刻的场景进行显式的空间结构化。在完整框架中，SGC模块作为一种高效的场景解析器，其核心价值在于实现了时空特征的解耦，它在每个时间切片上将非结构化的智能体状态转化为一个结构化的动态关系图。移除SGC模块后模型性能的严重衰退验证了这一点。由于缺少SGC模块，TimeMixer将被迫从高维的时空数据中同时学习瞬时空间关系和长期时间动态，这不仅大幅增加了模型的学习难度，也极大地限制了其性能上限。

与此同时，一个精确的微观时空交互模型必须被锚定在一个符合物理与交通规则的全局坐标系中才能产生有意义的输出。移除MHA模块后模型性能的下降凸显了宏观约束不可替代的基石作用。微观编码器本质上是在一个以自我为中心的、相对的交互空间中进行推理，但它本身对于环岛这一宏观概念是不可知的。MHA提供的正是这种自上而下的全局路径先验与语义约束，它有效地剪除了在交互层面看似合理但在环岛拓扑结构下绝无可能发生的轨迹。缺少这一宏观引导，模型预测将处于一种语义真空状态，极易产生逻辑上不自洽的严重错误。

综上所述，MST模型的卓越性能是由体系化设计带来的，它并非源于单一模块的强大，而是源于宏观全局约束模块、微观空间感知模块与微观时序推理模块三者之间功能互补、紧密耦合的有机协同。SGC负责理解每一帧的场景布局，TimeMixer机制负责思考这些布局如何随时间演化，MHA则负责规范所有思考结果必须遵守全局规则，最终形成了一个从感知、推理到约束的、高度鲁棒的完整预测闭环。

3.5 计算效率分析

为进一步评估所提出的MST模型的实际应用潜力，本节对模型复杂度与计算效率进行量化分析。选取先进的图结构模型Trajectron与兼顾精度与运算效率的模型HiVT，在模型参数量和平均单样本推理时间2个关键指标上进行对比。所有模型的推理时间均在NVIDIA GeForce RTX 2080 Ti GPU上进行测试，结果如表4所示。

表4 不同模型的计算效率对比

Table 4 Comparison of computational efficiency among different models

模型	参数量/10^6	单个样本的平均推理时间/ms
Trajectron	3.8	68
HiVT	2.0	36
MST	1.7	32

从表4可以看出，文中提出的MST模型在计算效率上具有显著优势，这一效率上的飞跃主要归功于其创新的体系架构设计，具体包括以下两点。

1) 高效的空间编码 微观编码器采用的SGC网络通过将多层非线性GCN折叠为单次线性变换，极大地降低了空间交互建模的参数量和计算复杂度。这与Trajectron模型为处理异构场景数据而采用的复杂图递归网络形成了鲜明对比。

2) 并行的时序分析 MST模型创新性地引入TimeMixer机制，其核心操作(如卷积、FFT)在GPU上具有高度并行性，替代了传统模型中用于时序编码的、速度较慢的串行递归网络，从而显著加快了特征提取速度。

综上所述，MST模型不仅在预测精度上实现了领先，更在模型轻量化和推理速度上表现卓越。它成功地在性能与效率之间取得了更优的平衡，证明了该架构的先进性，并展现了其在资源受限的车载计算平台上的巨大部署潜力。

3.6 可视化结果分析

为深入探究MST模型在真实复杂场景下的决策逻辑与推理能力，选取了3个包含显著交互模糊性的典型案例进行定性的深度剖析。

案例1(见图5)展示了一个在环岛外侧车道行驶的红色目标车辆为超越前方慢车而执行换道操作的决策过程。初始状态下，红色车辆的战略意图是高效地沿环岛行驶，但受到了前方慢速车辆的阻碍。在此场景下，MST模型的微观编码器开始发挥作用。首先，SGC网络将前方慢车识别为阻碍，

并同时捕捉到外侧绿色车辆的相对位置，构建出动态的交互图。随后，核心的TimeMixer机制介入进行时序解耦。它将前方慢车的缓行识别为一个持续的、低频的趋势性信号，而将红色车辆自身的超车意图也视为一个趋势。当两个趋势冲突时，模型必须进行战术决策。TimeMixer机制通过分析与外侧绿色车辆的交互，判断出存在安全的换道窗口。最终，解码器在MHA提供的车道拓扑约束下，综合了超车这一高优先级的战略意图和存在安全窗口的战术判断，从而高置信度地生成了向内侧变道并加速超越的轨迹。

案例2(见图6)描绘了一个更具对抗性的交互行为，红色目标车辆从外侧车道强行切入内侧车道。此案例是检验模型对非合作式博弈行为理解能力的关键。红色车辆的战略意图可能是为了更快地接近出口或超越外侧车辆。SGC网络首先感知到外侧车道空间狭小，绿色车辆与红色车辆的汇入点存在直接冲突。此时，TimeMixer机制通过其双轴注意力分析红色车辆的历史轨迹模式。若模型在训练数据中学到过此类驾驶风格，它会将红色车辆保持高速、不主动减速的行为识别为一个强烈的趋势性信号，这个信号的优先级高于常规的、礼让性质的周期性行为。MHA则确保了即使是如此激进的变道，其轨迹终点也必须落在合规的车道内。最终，解码器权衡后认为，该车辆的战略意图压倒了常规的战术性避让，因此会生成一个侵略性地切入绿色车辆前方的预测轨迹，精准复现了这种高风险的驾驶行为。

案例3(见图7)展示了一个经典的环岛入口让行场景，红色目标车辆已抵达让行线前，其战略意图是汇入环岛主路，而环岛内行驶的绿色车辆拥有先行权。SGC网络将绿色车辆识别为即将进入冲突区域的关键交互对象。随后，TimeMixer机制对时序动态进行解耦。它将绿色车辆稳定通过的行为识别为

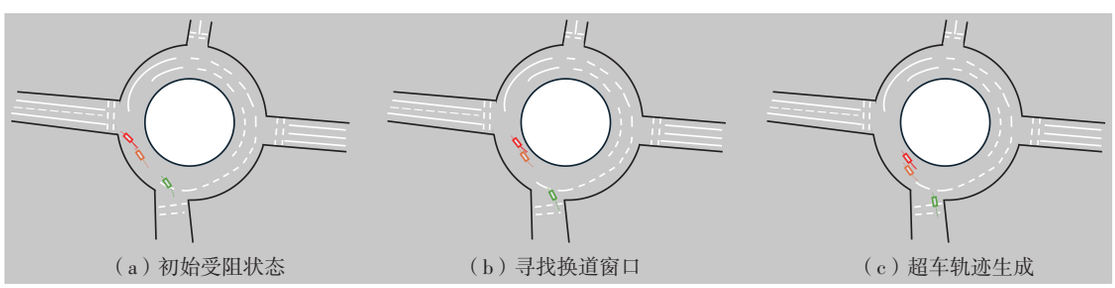


图5　案例1车辆轨迹预测图

Fig. 5　Vehicle trajectory prediction diagrams for case 1

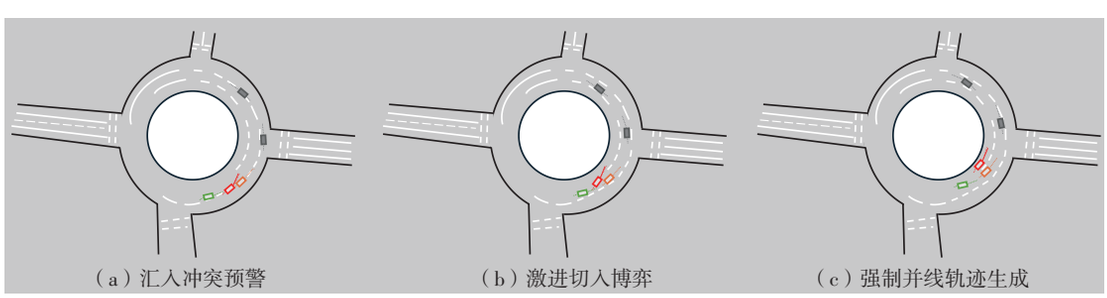


图6　案例2车辆轨迹预测图

Fig. 6　Vehicle trajectory prediction diagrams for case 2

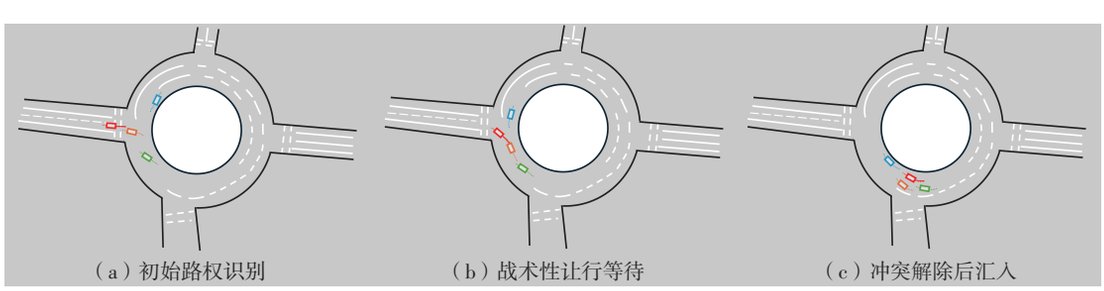


图7　案例3车辆轨迹预测图

Fig. 7　Vehicle trajectory prediction diagrams for case 3

一个需要被遵守的、强烈的趋势性信号；同时，对于红色车辆，当前的等待被理解为一个必要的、暂时的周期性行为。MHA的宏观约束则明确了红色车辆最终必须汇入，而不是无限期等待。解码器在综合所有信息后，做出了符合人类驾驶逻辑的推断。执行短期战术性等待，是实现长期战略意图的前提，因此，模型能够高置信度地生成一条先静止等待、待冲突解除后平滑加速汇入车流的未来轨迹，展现了模型对路权规则和车辆间合作博弈的深刻理解。

4 结语

本研究聚焦于环岛等结构化复杂场景下的车辆轨迹预测难题，提出并验证了一种结合了多头注意力机制、简化图卷积网络和TimeMixer机制的深度学习框架，其核心贡献在于：首次将通用时间序列模式引入交通轨迹预测领域，将传统的一维时域分析创新性地提升至时–频联合域内的模式分解与融合；通过将微观交互时序映射为二维时空图像，并对其内在的周期性战术行为(如跟车、让行)与趋势性战略意图(如加速出弯、准备变道)进行显式解耦，使模型得以更深刻、更鲁棒地洞察驾驶员的潜在意图。

在大型公开数据集INTERACTION及RounD上的全面实验评估，证明了文中模型相较于主流基线模型的卓越预测精度；详尽的消融实验也系统验证了文中模型宏观引导、微观感知、时空解耦架构中各组件的必要性和强大的协同效应。研究结果表明，文中模型为理解和建模复杂场景中的多智能体动态交互提供了一个行之有效的全新范式。

尽管文中模型已被证明具有良好的效果，但仍存在进一步的研究空间。未来工作的关键，首先，是将现有确定性预测框架与生成模型或扩散模型结合，输出多模态概率性轨迹，以更好地应对驾驶行为的内在随机性；其次，可通过在无信号路口、高速汇流区等更多样化的交通环境中进行测试与迁移学习，拓展模型的跨场景泛化能力；此外，还可采用异构图神经网络等方式来深化交互建模的精细度，以融入更丰富的交通语义；最后，可着力提升模型的可解释性，利用TimeMixer机制的显式解耦特性，将模型的中间特征与具体的驾驶行为进行语义关联，为自动驾驶系统的安全验证提供直观、可靠的物理解释。

参考文献：

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

[2] LIU J, MAO X, FANG Y, et al. A survey on deep-learning approaches for vehicle trajectory prediction in autonomous driving [C]//Proceedings of 2021 IEEE International Conference on Robotics and Biomimetics. Sanya: IEEE, 2021: 978-985.

[3] YURTSEVER E, LAMBERT J, CARBALLO A, et al. A survey of autonomous driving: common practices and emerging technologies [J]. IEEE Access, 2020, 8: 58443-58469.

[4] SUN Y, XU T, LI J, et al. MMH-STA: a macro-micro-hierarchical spatio-temporal attention method for multi-agent trajectory prediction in unsignalized roundabouts [J]. IEEE Transactions on Vehicular Technology, 2023, 72(9): 11237-11250.

[5] 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): 212207/1-21.

[6] MA L, TIAN S. A hybrid CNN-LSTM model for aircraft 4D trajectory prediction [J]. IEEE Access, 2020, 8: 134668-134680.

[7] MO X, XING Y, LV C. Interaction-aware trajectory prediction of connected vehicles using CNN-LSTM networks [C]//Proceedings of IECON 2020, the 46th Annual Conference of the IEEE Industrial Electronics Society. Singapore: IEEE, 2020: 5057-5062.

[8] SU L, LI L. Trajectory prediction based on machine learning [C]//Proceedings of IOP Conference Series: Materials Science and Engineering. Guangzhou: IOP Publishing, 2020: 012032/1-9.

[9] MESSAOUD K, YAHIAOUI I, VERROUST-BLONDET A, et al. Attention based vehicle trajectory prediction [J]. IEEE Transactions on Intelligent Vehicles, 2020, 6(1): 175-185.

[10] LIANG M, YANG B, HU R, et al. Learning lane graph representations for motion forecasting [C]//Proceedings of European Conference on Computer Vision. Glasgow: Springer, 2020: 541-556

[11] BAHRA N, PIERRE S. RNN-based user trajectory prediction using a preprocessed dataset [C]//Proceedings of 2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications. Thessaloniki: IEEE, 2020: 1-6.

[12] SHI L, WANG L, ZHOU S, et al. Trajectory unified transformer for pedestrian trajectory prediction [C]//

Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023: 9675-9684.

[13] SUI Z, ZHOU Y, ZHAO X, et al. Joint intention and trajectory prediction based on transformer [C]// Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems. Prague: IEEE, 2021: 7082-7088.

[14] ZHOU Z, YE L, WANG J, et al. HiVT: hierarchical vector transformer for multi-agent motion prediction [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 8823-8833.

[15] BAO K, BI J, GAO M, et al. An improved ship trajectory prediction based on AIS data using MHA-BiGRU [J]. Journal of Marine Science and Engineering, 2022, 10(6): 804/1-18.

[16] HE R, MENG Z, ZHANG S, et al. The SGC-Informer-based deep learning framework: managing the efficiency-accuracy trade-off in vehicle trajectory prediction [J]. IEEE Transactions on Vehicular Technology, 2025, 74(9): 13492-13507.

[17] ZHOU H, ZHANG S, PENG J, et al. Informer: beyond efficient transformer for long sequence time-series forecasting [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual Event: AAAI Press, 2021: 11106-11115.

[18] NGIAM J, CAINE B, VASUDEVAN V, et al. Scene transformer: a unified architecture for predicting multiple agent trajectories [C]//Proceedings of the 10th International Conference on Learning Representations. Virtual Event: OpenReview.net, 2022.

[19] YU C, MA X, REN J, et al. Spatio-temporal graph transformer networks for pedestrian trajectory prediction [C]//Proceedings of European Conference on Computer Vision. Glasgow: Springer, 2020: 507-523.

[20] WANG S, WU H, SHI X, et al. TimeMixer: decomposable multiscale mixing for time series forecasting [C]//Proceedings of the 12th International Conference on Learning Representations. Vienna: OpenReview.net, 2024.

Vehicle Trajectory Prediction at Roundabouts Based on Time Series Pattern Decomposition

ZHANG Jianhua　LI Wei

(School of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, Heilongjiang, China)

Abstract: To enhance the vehicle trajectory prediction accuracy in complex structured scenarios such as roundabouts, a deep learning model combining multi-head attention (MHA) mechanism, simplified graph convolution (SGC) network and TimeMixer mechanism, namely MST, is proposed. The model is built upon a macro-micro dual-encoder architecture. At the macro level, an MHA mechanism is employed to capture the long-term guiding constraints imposed by global road topology. That is, by modeling the complete historical trajectory of the vehicle and the structure of the roundabout (such as the deep relationship of the entrances and exits) to infer vehicle's long-term driving intention. At the micro level, first, an SGC network is used to extract instantaneous spatial relationships among vehicles. Subsequently, TimeMixer mechanism is introduced to map the one-dimension interaction sequence into multi-scale, multi-resolution 2D spatio-temporal images. By explicitly decoupling and hierarchically fusing periodic tactical behaviors and trend-oriented strategic intentions, a precise capture of deep interaction patterns is achieved. The information streams from both levels are integrated via a gated fusion network and then fed into a gated recurrent unit decoder to generate the final trajectory. Experiments on the public INTERACTION and RounD datasets demonstrate that, within a 5 s prediction period, the proposed model achieves an average displacement error and a final displacement error of 1.19 m and 1.85 m on the INTERACTION dataset, and 1.16 m and 1.80 m on the RounD dataset, respectively, outperforming all baseline models. The results indicate that hierarchically modeling macro-level global constraints and micro-level spatio-temporal interactions, particularly through the decoupling analysis of interaction patterns, can significantly improve the trajectory prediction performance in complex scenarios.

Key words: roundabout; vehicle trajectory prediction; multi-head attention; simplified graph convolution; TimeMixer