

图模体矩阵增强下的扩散模型轨迹生成方法*

袁洲旅¹, 何兆成^{1,2}, 谢传智¹

1. 中山大学智能工程学院 / 广东省智能交通系统重点实验室, 广东 深圳 518107
2. 鹏城实验室, 广东 深圳 518055

摘要: 在实际智能交通系统中, 一旦物联网出现故障就会引发数据中断, 这凸显了轨迹数据生成算法的重要性。现有的两种轨迹生成方法均存在不足。基于物理模型的算法不能适应动态变化的交通场景, 而数据驱动方法容易忽略车辆之间的交互影响。针对交叉口这类复杂场景, 本文提出一种基于条件扩散模型的轨迹生成方法。该方法引入模体矩阵量化周围车辆对目标车辆的影响程度, 使用图注意力机制准确识别关键邻居车辆信息, 并将条件信息引入扩散模型的反向去噪过程中生成车辆轨迹。最后, 在 Nuscenes 数据集和自建轨迹数据集上开展了验证实验。结果显示: 本文方法在多个精度评价指标上均优于现有主流模型, 且更加符合运动学规律。

关键词: 智能交通; 车辆轨迹生成; 条件扩散模型

中图分类号: U12 **文献标志码:** A **文章编号:** 2097-0137(XXXX)XX-0001-10

Graph-Motif matrix enhanced conditional diffusion model for vehicle trajectory generation

Yuan Zhouli¹, He Zhaocheng^{1,2}, Xie Chuazhi¹

1. School of Intelligent Systems Engineering, Sun Yat-sen University / Guangdong Provincial Key Laboratory of Intelligent Transportation Systems, Shenzhen 518107, China
2. Pengcheng Laboratory, Shenzhen 518055, China

Abstract: In practical intelligent transportation systems, failures of IoT may lead to data interruption, underscoring the importance of trajectory data generation algorithms. Existing trajectory generation methods can be categorized into two types, both of which exhibit limitations. Physics-based algorithms are unable to adapt to dynamically changing traffic scenarios, whereas data-driven methods tend to overlook interactions among vehicles. To address complex scenarios such as intersections, this paper proposes a trajectory generation method based on a conditional diffusion model. The proposed method introduces a motif matrix to quantify the influence of surrounding vehicles on the target vehicle, and employs a graph attention mechanism to accurately identify key neighboring vehicle information, and the conditional information is incorporated into the reverse denoising process of the diffusion model to generate vehicle trajectories. Experiments conducted on the Nuscenes dataset and a self-collected trajectory dataset demonstrate that the proposed method outperforms existing mainstream models across multiple accuracy evaluation metrics and is more consistent with kinematic principles.

Key words: intelligent transportation; vehicle trajectory generation; conditional diffusion model

* 收稿日期: 2026-03-20

录用日期: 2026-05-04

网络首发日期: XXXX-XX-XX

基金项目: 国家重点研发计划(2023YFB4301900); 深圳市基础研究专项(JCYJ20240813151445059)

作者简介: 袁洲旅(2000年生), 男; 研究方向: 智能交通; E-mail: yuanzhlv@mail2.sysu.edu.cn

通信作者: 何兆成(1977年生), 男; 研究方向: 智能交通; E-mail: hezhch@mail.sysu.edu.cn

全文阅读



ZR20260066

近年来,路侧摄像头、雷达等物联网技术在智能交通系统中的应用能力明显提升(Sliwa et al., 2020)。通过物联网采集到的交通数据更加精准,但没有经过处理的原始数据对交通智能化的作用有限(Klein, 2024)。例如,大型车辆对小型车辆的遮挡会导致跟踪中断(Gellida-Coutiño et al., 2025)。因此需要先进的轨迹挖掘和生成算法来恢复缺失的轨迹(Song et al., 2021)。在城市交通中,交叉口中的交通主体之间冲突更多、行为复杂,这对生成算法提出了更高的要求(Zhao et al., 2023)。

目前主流的轨迹生成方法分为两种:1)物理模型方法。基于物理规则的模型基于运动学规律和概率建模,如跟驰模型(Newell, 1961)。这类方法的可解释性与可控性更高,但难以适应日益复杂的交通场景。2)数据驱动方法(Leon et al., 2021)。数据驱动方法依靠深度学习从已有的历史轨迹中拟合数据分布,生成未知的轨迹。当前,生成模型是最先进的深度学习方法,比如:对抗神经网络(GAN)、扩散模型等(Ho et al., 2020)。它们的特点是学习能力强但可解释性差(Leon et al., 2021)。且由于缺少物理规则的引导,数据驱动方法难以真正理解场景中的交互信息(Li et al., 2023),容易生成违反物理规则的轨迹。

因此,面向复杂的交叉口场景本文提出了一种基于条件扩散模型的改进型车辆轨迹生成方法。首先基于图模体矩阵增强的扩散模型,将目标车辆的轨迹和关键邻居车辆的信息聚合在一起。在识别关键邻居时,通过将模体矩阵加入图注意力机制,更精确地测算出了周围车辆的影响权重。最后,在公开的 nuScenes 数据集(Caesar et al., 2019)和自建车辆轨迹数据集上进行实验,验证了模型的可靠性和有效性。

1 轨迹生成方法

轨迹生成方法经历了基于物理规则的模型、深度学习模型、生成模型3个阶段。初期的物理模型主要依赖运动学假设和马尔可夫建模,但物理模型在应对交叉口这种多主体交互与冲突密集的场景时表现出明显的局限性。而 LSTM 和 AgentFormer (Altché et al., 2017; Yuan et al., 2021)等模型使用神经网络对轨迹进行序列建模与拟合,并通过池化、图结构或注意力机制来描述局部交互,显著提升了轨迹生成的精度,也让分布式输出成为了主流。生成模型中,生成对抗网络用对抗博弈的方式

学习数据分布(Gupta et al., 2018; Dendorfer et al., 2021),而变分自编码器则把输入数据编码到低维的潜在空间,在采样完成后再解码(Choi et al., 2022; Zhong et al., 2022)。这些研究先天性存在训练缺乏稳定或难以充分捕捉轨迹数据中离散的多模态分布等问题,导致生成的轨迹呈现模糊平均。而扩散模型以其更加稳定的训练与强大的拟合能力弥补了前述模型的不足。

去噪扩散概率模型(DDPMs)最初被应用于图像生成领域(Nichol et al., 2021),它通过模拟和学习非平衡统计物理的扩散过程来重建原来的数据分布。后续学者将扩散模型用作概率性的时间序列预测(Tashiro et al., 2021)。在轨迹生成方面, MID 模型首次将扩散模型用于行人轨迹预测(Gu et al., 2022);LED 模型改进了扩散模型的去噪过程,直接生成一种粗粒度的行人轨迹分布,省去了大量的降噪步骤(Mao et al., 2023)。这些传统的扩散模型只将轨迹数据作为输入,缺少对于车辆间深层次交互行为的学习机制。针对这一问题,受 Li et al.(2023)的条件扩散模型启发,本文提出了一种基于图注意力机制与模体共同作用的条件扩散模型,加强了对车辆间交互行为的深度学习。

2 轨迹预测建模

2.1 问题建模

首先定义轨迹生成问题。设 t 时刻交叉口正在运行的车辆集合为 $V_t = \{1, 2, \dots, M_t\}$ 。对于其中任意车辆 m , 其状态向量为

$$\mathbf{s}_{m,\tau} = (x_{m,\tau}, y_{m,\tau}, v_{m,\tau}, a_{m,\tau}, \theta_{m,\tau})^T,$$

其中 (x, y) 表示空间位置, v 表示速度, a 表示加速度, θ 表示航向角。设历史观测窗口长度为 L , 生成窗口长度为 H 。车辆 m 的历史轨迹片段为

$$X_{m,t-L+1:t} = (\mathbf{s}_{m,t-L+1}, \dots, \mathbf{s}_{m,t}).$$

生成的未来轨迹为

$$Y_{m,t+1:t+H} = (\mathbf{s}_{m,t+1}, \dots, \mathbf{s}_{m,t+H}).$$

此外,定义场景条件集合 $C_t = E_t$, 这里 E_t 表示车辆之间的交互冲突情况。综上,轨迹生成任务是指在已知历史轨迹和交互条件下学习未来轨迹的条件分布,并生成符合约束的轨迹样本。即:

$$p(Y_{t+1:t+H} | X_{t-L+1:t}, C_t).$$

模型输出的不是一条确定的轨迹,而是符合条件分布的多个轨迹序列。因此,与传统的回归任务不同,本文模型更注重对时序信息的编码能力和对场景信息的捕捉能力。

2.2 模型结构

如图1所示,整个模型由时空编码器、模体计算模块、引导条件生成以及条件扩散模型组成。时空编码器负责提取历史轨迹的特征,将目标车辆和周围车辆的历史轨迹 X 与未来的真实轨迹 Y 编码为输入向量。然后,使用模体计算机制遍历整个图结构,得到模体矩阵,它反映了目标车辆和周围车辆

之间的连接强度。模体矩阵和历史轨迹特征一起被送进条件融合模块中,图神经网络以目标车的轨迹作为中心节点,结合模体矩阵得到关键邻居的注意力权重,并将邻居信息加权融合,最终生成场景条件 C_i 。通过门控机制把场景信息引入扩散模型的反向去噪过程,就可以动态调整场景条件在不同生成阶段的影响力大小。

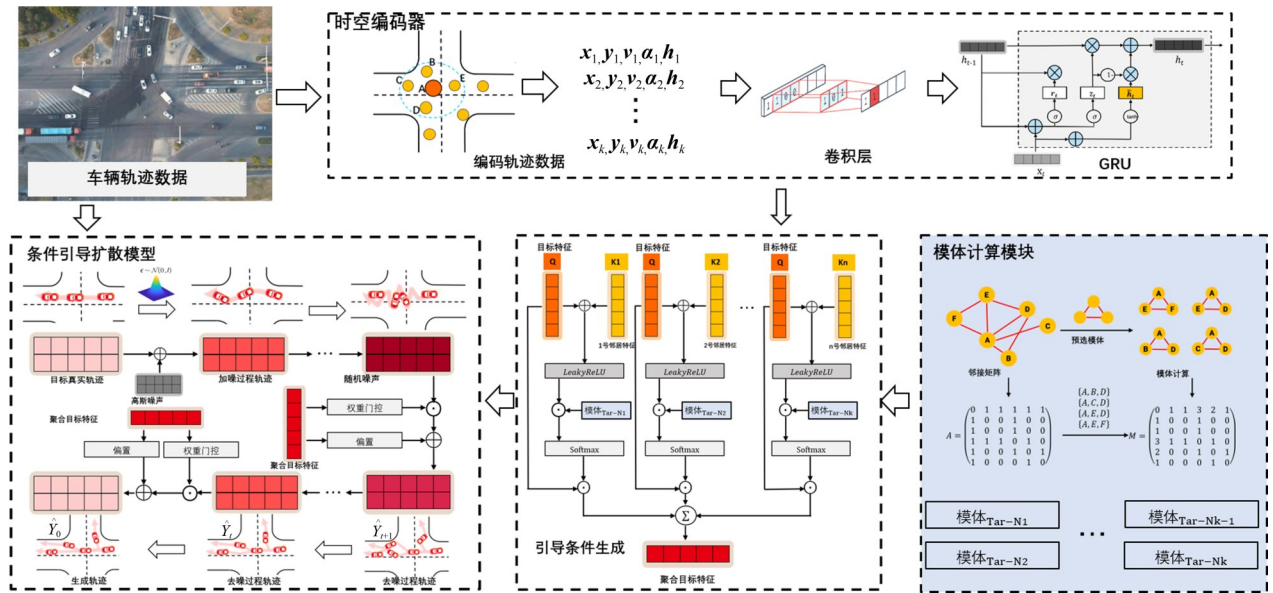


图1 条件扩散模型

Fig. 1 Conditional diffusion model

2.3 时空编码器

特征提取模块对每一辆车的历史轨迹 X 和目标车辆的未来真实轨迹 Y 进行特征编码,输入数据包含多个时间步长的历史轨迹,涉及相对坐标、速度、加速度、偏航角等动态信息。输入数据首先经过卷积层沿着时间维度滑动,对每个时间步长进行加权求和与激活函数处理;输出高维特征后,进入门控循环单元 GRU。GRU 是循环神经网络 (RNN) 的一种变体,它将当前输入与上一个时间步的隐藏状态结合,对卷积输出做更新和重置处理;通过控制历史信息 and 当前输入比重,最终得到整个时间序列的全局特征。

2.4 引导条件生成

在复杂的交通场景里,目标车辆的轨迹受到多个车辆的潜在影响,这种影响仅使用简单的邻接矩阵难以捕捉。本文引入了模体机制来深化对交互行为的理解。模体是复杂网络的基本构成单元,这种子图结构在统计上具有明确意义。在交通场景中,图结构记为 $G_v = (v, E_v)$,其中, v 表示车辆集合, E_v 表示边集合。建立邻接矩阵 A ,矩阵元素 A_{ij} 表示

车辆 i 和车辆 j 处于影响半径之内。

为了描述车辆之间的相关性强弱,在邻接矩阵基础上计算得到模体矩阵 M 。模体矩阵中的元素 M_{ij} 定义为

$$M_{ij} = \sum_{v \in \{V_{ij}\}} \delta(A_v, g),$$

其中 g 表示车辆网络中的某一种特定子图, V_{ij} 是在整个图结构中同时包含车辆 i 和 j 的子图 g 的集合。 A_v 表示车辆 i 和 j 共同出现的特定子图的类型。当 A_v 和 g 的类型一致时,函数 $\delta(A_v, g)$ 记为 1,否则记为 0。

在生成模体矩阵时,对整个图结构进行遍历,计算出多个时间步下的矩阵 $\{M^t\}_{t \in \{0:t_n\}}$ 。对于历史时间步 t 下的模体矩阵 M^t ,它的元素 M_{ij}^t 定义为时间步 t 车辆 i 和 j 共同出现的模体结构数量。对于多个时间步,有:

$$\hat{M} = \sum_t M^t,$$

最后对矩阵做归一化处理。

图注意力网络先计算每个节点相对于中心节

点的注意力权重,再用权重将特征信息聚合起来。有:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\alpha^T[\mathbf{W}h_i|\mathbf{W}h_j]\right)\right)}{\sum_{k \in N_i} \exp\left(\text{LeakyReLU}\left(\alpha^T[\mathbf{W}h_i|\mathbf{W}h_k]\right)\right)},$$

其中 α_{ij} 为注意力系数; h_i 和 h_j 分别代表节点 i 和节点 j 的初始特征向量; \mathbf{W} 是一个可学习的变换矩阵,为每个节点生成更高维度的特征表示。经过激活函数和归一化处理之后, α_{ij} 表示节点 i 在聚合信息时应赋予它的邻居节点 j 的权重大小。

更新之后,节点 i 的特征向量 h'_i 可以表示为

$$h'_i = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} \mathbf{W}h_j\right).$$

图注意力网络接收所有车辆 N 的编码历史轨迹特征和车辆之间的交互信息两部分输入,在计算中加入节点 i 和 j 对应的模体元素,得到基础注意力得分:

$$e_{ij} = M_{ij} \times \text{LeakyReLU}\left(\alpha^T[\mathbf{W}h_i|\mathbf{W}h_j]\right).$$

归一化后,得到最终的注意力系数:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})}.$$

通过加权聚合,可以得到目标车辆的场景特征

$$C = h'_{\text{Target}} = h'_i = \sigma\left(\sum_{j \in N(i)} \alpha_{ij} \mathbf{W}h_j\right).$$

在注意力计算过程中,通过模体矩阵元素与注意力分数相乘,对邻近车辆的影响进行了过滤与增强。生成的引导条件本质上是目标车辆与关键邻居车辆的聚合历史信息。

2.5 条件引导扩散模型

扩散模型主要由正向扩散过程和反向去噪过程组成。本文在去噪过程中加入了聚合历史信息引导轨迹生成。

2.5.1 正向扩散过程 正向加噪是逐步向目标车辆的真实轨迹里添加随机噪声,直到未来的轨迹分布接近高斯分布,变成一堆完全无序的轨迹点。单步的正向扩散为

$$q(Y_t|Y_{t-1}) = N\left(\sqrt{\alpha_t} Y_{t-1}, \beta_t I\right),$$

其中 $\alpha_t = 1 - \beta_t$ 。在时间 $t-1$ 的状态上添加高斯随机噪声 I 获得时间 t 的状态,通过递归和重参数化技术,可由初始轨迹 Y_0 推导出任意时刻 t 的噪声轨迹

$$q(Y_t|Y_0) = N\left(\sqrt{\alpha_t} Y_0, \sqrt{1 - \alpha_t} I\right),$$

其中 $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ 。整个正向过程可以表示成一个马尔可夫链 $q(Y_{1:T}|Y_0) = \prod_{i=1}^T q(Y_i|Y_{i-1})$ 。

2.5.2 反向去噪过程 反向去噪过程是正向加噪的逆过程。它通过学习和逼近逐步加入的噪声,再从高斯分布中去噪、重建数据。反向去噪过程用马尔可夫链表示为

$$p_\theta(Y_{0:T}) = p(Y_T) \prod_{i=1}^T p_\theta(Y_{i-1}|Y_i, c),$$

这里, θ 是扩散模型要学习的参数, C 是获得的受周围车辆影响的目标车辆的轨迹特征。

在训练去噪时,使用多层感知器(MLP)来生成噪声和轨迹。输入包括真实的未来轨迹 Y_0 、条件信息 C 、噪声 $\epsilon \sim N(0, I)$ 和时间嵌入 t 。通过逐步向 Y_0 加入噪声,得到干扰后的随机轨迹点

$$Y_T = \sqrt{\alpha_T} Y_0 + \sqrt{1 - \alpha_T} \epsilon,$$

其中 α_T 为超参数,且 $\bar{\alpha}_T = \prod_{i=1}^T \alpha_i$ 。

在去噪过程中,用一个线性变换层来融合条件特征,从而生成扩散模型所需的噪声。具体做法是先将时间嵌入 t 和条件信息 C 拼在一起,这样 MLP 就能根据不同的生成阶段,动态调整条件输入和当前轨迹输入的比重。单步噪声的生成过程如图 1 条件引导扩散模型中所示。 t 时刻的噪声输出可以表示为

$$\epsilon_t = (\mathbf{W}\hat{Y}_{t+1}) \odot \sigma(G(C)) + B(C),$$

其中 \hat{Y}_{t+1} 是上一时间步生成的轨迹,门控函数 $G(C) = W_g C$,偏置项 $B(c) = W_b C$ 。当前时间步的轨迹为 $\hat{Y}_t = \hat{Y}_{t+1} - \epsilon_t$ 。

去噪过程开始时,模型生成的轨迹失真严重。随着去噪过程推进,轨迹序列质量逐步提高,模型逐渐提高前序生成轨迹的影响权重。逐步从前序轨迹中减去预测噪声,得到当前轨迹。重复以上过程,直至时间步长为 0,最终生成目标轨迹。

2.5.3 损失函数 训练过程中的损失函数是生成噪声与预测噪声之间的均方误差,即:

$$L = E_{x_0, \epsilon, t} \left[\left\| \epsilon - \hat{\epsilon}(x_t, t, C) \right\|^2 \right].$$

3 实验

为了验证模型的性能,我们在 2 个数据集上开展了多个实验。通过与多个基线模型进行对比,消融实验、参数敏感性分析等全面评估了模型性能。最终选取生成轨迹样本进行可视化展示与分析。

3.1 实验设置

3.1.1 数据集 我们使用 nuScenes 数据集 (Caesar et al., 2019) 和自行采集的某大型交叉口车辆轨迹数据来训练和验证生成模型。nuScenes 数据集包含 4 个街区的 1 000 个真实驾驶场景, 每个场景时长 20 s。该数据集提供了超过 20 万条 8 s 时长的邻近目标轨迹序列样本, 采样频率为 2 Hz。

第 2 个数据集来自激光雷达、高清摄像机和边缘计算器等组成的路侧智慧基站, 如图 2 所示。智慧基站在高密度交叉口采集的轨迹数据信息如表 1 所示, 包括了轨迹点位置、车辆平均速度、航向角等信息。数据集的采样频率为 10 Hz, 同样构建了 2 万条 8 s 时长的轨迹序列样本。实验时间为 2025 年 5 月的某一工作日的早高峰时段。

3.1.2 评估指标 使用常用的轨迹生成评估指标来验证模型的精度。平均位移误差 (ADE) 是每个时间点真实轨迹和预测轨迹的平均误差。即:

$$ADE = \frac{1}{T} \sum_{t=1}^T |\hat{Y}_t - Y_t|_2,$$

其中 \hat{Y}_t 是模型在时间 t 生成的位置, Y_t 是该时间点的真实位置。

最终位移误差 (FDE) 是最后一个时间点真实轨迹和预测轨迹的误差。

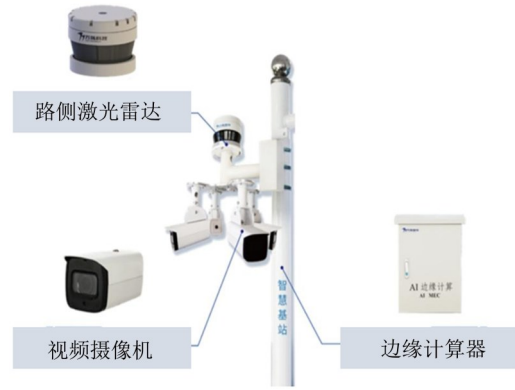


图 2 路侧智慧基站

Fig. 2 Roadside intelligent base station

$$FDE = |\hat{Y}_T - Y_T|_2,$$

其中 \hat{Y}_T 是模型在最终时刻 T 生成的位置, Y_T 是该时间点的真实位置。

误差率是真实轨迹与生成轨迹的误差超过阈值的比率, 阈值设置为 2.0 m。对于模型生成轨迹, 分别用 $\min ADE_K$ 、 $\min FDE_K$ 和 MissRate_K 作为评估和分析指标。这些值表示 K 条轨迹中的指标平均值, 数值越小表示生成模型的精度越高。

3.2 定量分析

在 2 个数据集上与多个先进基线模型进行精度对比, 结果如表 2~3 所示。

表 1 轨迹采集的基础信息

Table 1 Basic information on trajectory acquisition

列名	示例	含义
Global Frame No	74530	特定时间段内车辆唯一编号
Global Time Stamp	2025-05-30 07:00:00:27	车辆轨迹捕捉时间戳
longitude	118.740507	车辆轨迹点经度
latitude	30.9719645	车辆轨迹点纬度
speed	30 m/s	车辆平均速度
courseAngle	1199	车辆航向角
participantNum	16	同时跟踪车辆数量
laneNum	8	车辆所在车道编号

在 nuScenes 数据集上, 所提模型在 $K = 1, 5, 10$ 时的指标均优于对比方法。其中, $\min ADE_K$ 平均下降了 6.6%, $\min FDE_K$ 平均下降了 5.5%。随着 K 的增大, 模型误差下降得更明显, 这说明生成轨迹与真实轨迹分布相近。与同类方法相比, 本文模型在终点误差上改善更加显著, 这意味着它对车辆的长时间预测更准确, 这可能得益于模型能够合理推断出

多个车辆之间的交互信息。

表 3 展示了本文方法在自采真实数据集上的精度结果。在 $K = 1, 5, 10$ 时, 本文方法的整体表现不错。具体来看, $\min ADE_1$ 、 $\min ADE_5$ 、 $\min ADE_{10}$ 分别为 2.45、1.01 和 0.63, 而且 $\min ADE_1$ 与 $\min ADE_{10}$ 分别比次优基线下降了 11.6% 和 40.4%; $\min FDE_K$ 则分别是 6.63、2.26 和 1.82。

表2 在 nuScenes 数据集上的精度指标

Table 2 Accuracy metrics on the nuScenes dataset

模型	minADE ₁	minADE ₅	minADE ₁₀	minFDE ₁	minFDE ₅	minFDE ₁₀
CoverNet(Phan-Minh et al. , 2020)	3.90	1.97	1.48	9.26	4.20	2.87
PreTraM (Xu et al. , 2022)	3.76	1.70	1.45	9.01	4.15	3.22
DAMM (Wen et al. , 2024)	2.84	1.39	1.02	6.99	3.14	2.12
AgentFormer (Yuan et al. , 2021)	3.12	1.86	1.45	8.01	3.89	2.86
CXX(Luo et al. , 2020)	3.25	1.63	1.29	8.86	3.93	3.10
Real-Time(Li et al. , 2024)	3.56	1.60	1.23	8.63	3.34	2.32
ContextVAE (Xu et al. , 2023)	3.54	1.59	1.33	8.24	3.28	2.87
PGP(Deo et al. , 2022)	2.91	1.30	1.00	7.63	3.25	2.63
本文	2.61	1.24	0.93	6.69	2.92	2.01

表3 在自采数据集上的精度指标

Table 3 Accuracy metrics on the self-collected dataset

模型	minADE ₁	minADE ₅	minADE ₁₀	minFDE ₁	minFDE ₅	minFDE ₁₀
CoverNet (Phan-Minh et al. , 2020)	3.69	2.02	1.26	8.26	4.68	3.10
PreTraM (Xu et al. , 2022)	2.95	1.77	1.22	8.75	4.52	3.16
DAMM (Wen et al. , 2024)	2.86	1.33	1.12	6.54	3.42	2.13
AgentFormer(Yuan et al. , 2021)	3.16	1.88	1.32	8.10	3.94	2.88
CXX (Luo et al. , 2020)	2.82	1.56	1.17	7.66	3.87	2.91
Real-Time(Li et al. , 2024)	3.72	1.64	1.12	7.52	3.32	2.58
ContextVAE (Xu et al. , 2023)	3.60	1.53	1.21	8.28	3.22	2.67
PGP(Deo et al. , 2022)	2.77	1.22	1.05	7.67	3.01	2.97
本文 (nuScenes)	2.61	1.24	0.93	6.69	2.92	2.01
本文	2.45	1.01	0.63	6.63	2.26	1.82

虽然,与 PGP、DAMM 模型相比,本文模型的 minADE₅、minFDE₁ 相差不大;但随着 K 的增大,本文模型的误差下降更明显,这说明本文模型的候选轨迹覆盖能力更强,更容易命中真实的未来轨迹分支。尤其值得注意的是,DAMM 和本文模型都使用了模体先验来捕捉交互行为,但本文模型在核心指标上更优。这说明在相近的交互先验条件下,条件扩散生成框架能学到更稳定的未来行为分布。

对比在两个数据集上的生成精度,发现自建数据集上的精度指标比公开数据集上的更好。这主要是因为:自建数据集的车辆密度更大,对车辆间交互的捕捉更准;而公开数据集更多的是根据目标车自身的动力学来做轨迹预测,不确定性更大。

3.3 消融实验与性能评估

3.3.1 消融实验 设置了4组消融变量:去掉模体机制、去掉 GAT、同时去掉模体机制和 GAT、完整模型。基线是一个基于历史轨迹和场景信息的扩散模型,用来生成未来轨迹。实验结果如表4所示。

和无条件的扩散模型相比,GAT对预测精度的提升比较有限;但 GAT 和模体矩阵协作时,生成轨迹的精度明显提升。这说明采用多种方法协同的交互式特征提取策略,有助于在轨迹预测任务中达到更高的精度。

3.3.2 条件特征维度的影响 模型利用条件信息的能力依赖于条件特征的维度,条件特征的维度决定了模型能从条件门中得到多少场景特征以及对

表 4 在自采数据集上的消融实验

Table 4 Abandonment experiments on the self-collected dataset

模型	minADE ₅	minFDE ₅	minADE ₁₀	minFDE ₁₀
w/o Motif	2.78	7.47	0.97	1.91
w/o GNN	2.91	7.57	0.88	1.99
w/o Motif & GNN	2.99	7.99	1.03	2.04
完整模型	2.61	7.23	0.93	1.81

动态修正主线性层权值和偏置水平的影响。一般来说, 维度越大, 模型能捕捉到的上下文信息就越准确。条件特征维度为 32、64、128、256 时, 生成轨迹的精度如图 3 所示。

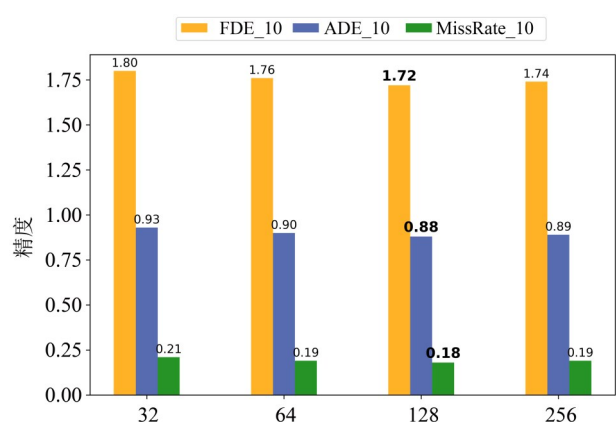


图 3 条件特征维度对模型的影响

Fig. 3 The impact of condition feature dimensions on the model

当维度从 32 增加到 128 时, minADE₁₀、minFDE₁₀、missrate₁₀ 单调下降, 这说明模型的性能

有明显提升。条件特征维度从 128 增加到 256 时, 各类指标的变化趋于平稳, 甚至有少量回升。这表明维度太小会导致交互信息的表达能力不足, 因模型不能充分利用这些信息, 导致错误率和遗漏率都比较高。因此, 提高维度能增强上下文信息的表达能力, 让模型更有效地动态调整权重, 但条件信息过于丰富就会导致性能提升放缓甚至轻微下降, 甚至可能导致模型出现一些冗余或者轻微过拟合。

3.3.3 效率分析 模型效率主要涉及计算成本和存储成本。为了评估各个模块的对模型效率的影响, 采用了预测精度、参数数量、浮点运算量 (FLOPs) 和推理延迟 4 个核心指标。设置了 3 种配置, 第 1 种是无条件信息的扩散模型, 第 2 种是加了 GAT 但没有“躯干”的条件扩散模型, 第 3 种是完整的条件交互扩散模型。对比结果如表 5 所示。

从表 5 中可以发现, 完整的条件交互扩散模型虽然参数数量更多, 但它的精度是最高的, 预测效果也最好。与无 GAT 的条件扩散模型相比, 完整模型在浮点运算量和推理延迟上明显更低, 计算开销也更小。总体而言, 本文模型在准确性和计算效率之间做到了较好的平衡。

表 5 不同组件对模型效率的影响¹⁾

Table 5 The impact of different components on model efficiency

扩散模型	GAT	模体矩阵	minADE ₅	Param/M	FLOPs/G	Latency/ms
✓	-	-	1.71	2.10	1.10	12.50
✓	✓	-	1.25	3.10	2.56	20.30
✓	✓	✓	0.93	3.60	1.78	15.20

1) “✓”代表有该组件;“-”无该组件。

表 6 在自采数据集上的运动学可行性实验

Table 6 Kinematics feasibility experiments on the self-collected dataset

模型	minADE ₁₀	minFDE ₁₀	TTC 违规率	加速度超限率
w/o Motif	0.89	1.41	9/200	7/200
w/o GNN	0.88	1.42	11/200	10/200
w/o Motif & GNN	0.93	1.55	12/200	11/200
完整模型	0.77	1.22	6/200	5/200

3.3.4 运动学可执行性实验 本文设计了一个运动学可执行性实验,在自建数据集上对比了精度指标 $\min ADE_{10}$ 、 $\min FDE_{10}$ 和 TTC 违规率、加速度超限率两个物理一致性指标。TTC 违规率是碰撞时间(TTC)低于安全阈值的样本占比,用来衡量多车交互过程中的安全风险。这个比例越低,说明模型在交互场景里越能生成符合安全边界的行为。实验中,安全阈值设为 2 s。加速度超限率则是生成轨迹中加速度超过车辆可接受范围的比例。比例越低,说明轨迹越符合基本的动力学约束。将加速度阈值设为 2 m/s^2 ,实验结果如表 6 所示。

完整模型在 $\min ADE_{10}$ 、 $\min FDE_{10}$ 和物理一致性指标上都取得了最佳结果,这说明模体机制和图注意力网络相互配合可以提高轨迹预测的精度和行为合理性。

在物理一致性方面,完整模型的 TTC 违规率为 3.0%,加速度超限率为 2.5%,明显低于另外 3 个消融模型。与去掉模体(w/o Motif)的模型相比,完整模型的 TTC 违规率下降了 33.3%,加速度超限率下降了 28.6%;与去掉图神经网络(w/o GNN)的模型比,完整模型的 TTC 违规率和加速度超限率分别下降了 45.5% 和 50.0%;与两个模块都去掉(w/o Motif 和 GNN)的模型比, TTC 违规率和加速度超限率分别下降了 50.0% 和 54.5%。这说明模体筛选能帮助模型更好地识别关键交互模式,而 GNN 则进一步强化了邻域交互的传播。两者结合共同抑制了不合理的轨迹变化和危险行为的生成。

3.4 可视化评估

本文基于自采集的现实数据集呈现了车辆轨迹生成结果。图 4 使用一个直行轨迹生成案例对交叉口场景做出展示,该交叉口为大型四相位多车道交叉口,设有专用转弯道和掉头道。

左转、右转、掉头和直行 4 种驾驶场景下的预测结果,如图 5 所示。每个场景中包含了目标车辆的真实未来轨迹、本文方法生成的轨迹、周围车辆的历史轨迹以及生成轨迹的分布带。周围车辆是模型在预测时通过过滤机制选出来的。所有车辆的历史轨迹都对齐到了同一个时间点。为了将轨迹点密集处表述地更清楚,使用等距采样的方式进行了可视化。

图 4~5 清晰地说明了本文模型既能捕捉目标车的历史轨迹,也能学习周围车辆的交互行为。从图

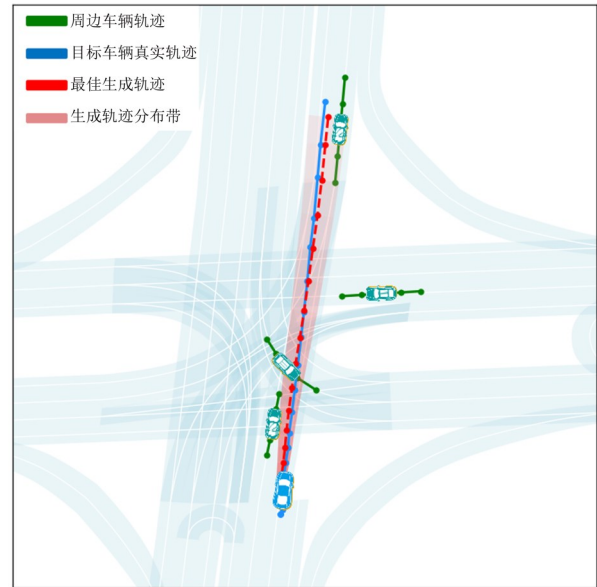


图 4 直行轨迹生成及交叉口场景展示

Fig. 4 Generation of straight-line trajectories and display of intersection scenarios

5a 和图 5b 可以看出,目标车在交叉口左转时受到了同向左转的外侧车辆的挤压,也受到对向车流的影响。另外,轨迹终点处因车辆占用车道,导致生成的路径严格沿着车道线走、没有变道。这些信息都作为环境条件被纳入了轨迹生成过程。从图 5c 和图 5d 可以看出,目标车受限于周围轨迹,只能出现在两辆车的空隙中。这一过程在生成轨迹里得到了很精确的体现。在图 5e 和图 5f 中,模型通过捕捉周围车的变道行为,能准确识别目标车的变道动作和驾驶意图,从而生成了很精准的轨迹。此外,虽然图 5a 和图 5e 中目标车的历史轨迹点分布不太规则,但模型还是能从周围车的历史轨迹中提取场景约束,最终准确生成了目标车的轨迹。

4 总结与展望

本文提出了一个同时考虑先验知识和生成模型的车辆轨迹预测模型,并通过一系列实验对生成的轨迹进行了评价。首先,模型将模体矩阵和图注意力网络结合为一种条件信息。条件信息既包含历史轨迹,也包含周围车辆的影响,并能通过门控机制融入到扩散模型的每一个去噪过程中,以动态调整不同时间步中轨迹生成信息来源比例,最终生成高质量的轨迹。

在主流公开数据集和自采的真实城市交叉口轨迹数据集上开展了验证实验。实验结果显示,所

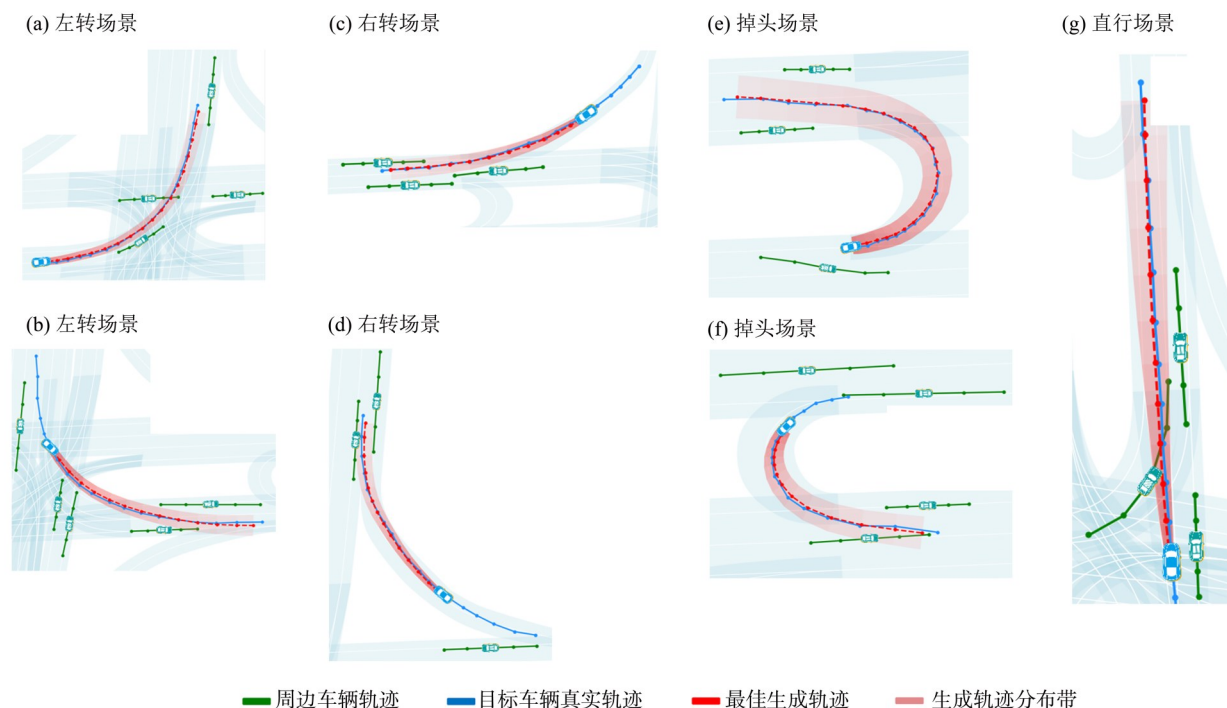


图5 多场景生成轨迹

Fig. 5 Multi-scenario trajectory generation

提方法能够适应复杂的多车交互场景,生成的轨迹兼具精度高和符合动力学规律的优势。未来,拟将方法扩展到多目标同步轨迹生成任务上。并将基

于大批量生成的轨迹,进一步推演交叉口内的宏观交通状态。

参考文献:

- Alché F, de la Fortelle A, 2017. An LSTM network for highway trajectory prediction [C]//2017 IEEE 20th International Conference on Intelligent Transportation Systems. Yokohama, Japan: 353-359.
- Caesar H, Bankiti V, Lang A H, et al, 2019. nuScenes: A multimodal dataset for autonomous driving [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, Washington, USA: 11618-11628.
- Choi D, Min K, et al, 2022. Hierarchical Latent Structure for Multi-modal Vehicle Trajectory Forecasting [M]// Computer Vision-ECCV 2022. Cham: Springer Nature Switzerland: 129-145.
- Dendorfer P, Elflein S, Leal-Taixé L, 2021. MG-GAN: A multi-generator model preventing out-of-distribution samples in pedestrian trajectory prediction [C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: 13138-13147.
- Deo N, Wolff E, Beijbom O, 2022. Multimodal Trajectory Prediction Conditioned on Lane-Graph Traversals [C]// Proceedings of the 5th Conference on Robot Learning. PMLR: 203-212.
- Gellida-Coutiño C, Rios-Cabrera R, Maldonado-Ramirez A, et al, 2025. Real-time multi-camera tracking for vehicles in congested, low-velocity environments: A case study on Drive-Thru Scenarios [J]. Electronics, 14(13): 2671.
- Gu T, Chen G, Li J, et al, 2022. Stochastic trajectory prediction via motion indeterminacy diffusion [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: 17092-17101.
- Gupta A, Johnson J, Fei-Fei L, et al, 2018. Social GAN: Socially acceptable trajectories with generative adversarial networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: 2255-2264.
- Ho J, Jain A, Abbeel P, 2020. Denoising diffusion probabilistic models [J]. Adv Neural Inf Process Syst, 33: 6840-6851.
- Klein L A, 2024. Roadside Sensors for Traffic Management [J]. IEEE Intell Transp Syst Mag, 16(4): 21-44.
- Leon F, Gavrilescu M, 2021. A Review of Tracking and

- Trajectory Prediction Methods for Autonomous Driving [J]. *Mathematics*, 9(6): 660.
- Li L, Wang X, Yang D, et al, 2024. Real-time heterogeneous road-agents trajectory prediction using hierarchical convolutional networks and multi-task learning[J]. *IEEE Trans Intell Veh*, 9(2): 4055-4069.
- Li Z, Liang H, Wang H, et al., 2023. A multi-modal vehicle trajectory prediction framework via conditional diffusion model: A coarse-to-fine approach [J]. *Knowl Based Syst*, 280: 110990.
- Luo C, Sun L, Dabiri D, et al, 2020. Probabilistic multi-modal trajectory prediction with lane attention for autonomous vehicles [C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2370-2376.
- Mao W, Xu C, Zhu Q, et al, 2023. Leapfrog diffusion model for stochastic trajectory prediction [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition . Vancouver, BC, Canada: 5517-5526.
- Newell G F, 1961. Nonlinear effects in the dynamics of car following[J]. *Oper Res*, 9(2): 209-229.
- Nichol A Q, Dhariwal P, 2021. Improved denoising diffusion probabilistic models [C]//Proceedings of the 38th International Conference on Machine Learning. PMLR: 8162-8171.
- Phan-Minh T, Grigore E C, Boulton F A, et al, 2020. CoverNet: Multimodal behavior prediction using trajectory sets [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 14062-14071.
- Sliwa B, Piatkowski N, Wietfeld C, 2020. The channel as a traffic sensor: Vehicle detection and classification based on radio fingerprinting [J]. *IEEE Internet Things J*, 7(8): 7392-7406.
- Song X, Pi R, Lv C, et al, 2021. Augmented multiple vehicles' trajectories extraction under occlusions with roadside LiDAR data[J]. *IEEE Sens J*, 21(19): 21921-21930.
- Tashiro Y, Song J, Song Y, et al, 2021. CSDI: Conditional score-based diffusion models for probabilistic time series imputation[J]. *Adv Neural Inf Process Syst*, 34: 24804-24816.
- Wen D, Xu H, He Z, et al, 2024. Density-Adaptive model based on motif matrix for multi-agent trajectory prediction[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: 14822-14832.
- Xu C, Li T, Tang C, et al, 2022. PreTraM: Self-supervised pre-training via connecting trajectory and map [M]//Avidan S, Brostow G, Cissé M, et al. *Computer Vision-ECCV 2022*: 13699. Cham: Springer Nature Switzerland: 34-50.
- Xu P, Hayet J B, Karamouzas I, 2023. Context-aware timewise VAEs for real-time vehicle trajectory prediction [J]. *IEEE Robot Autom Lett*, 8(9): 5440-5447.
- Yuan Y, Weng X, Ou Y, et al, 2021. AgentFormer: Agent-aware transformers for socio-temporal multi-agent forecasting [C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: 9793-9803.
- Zhao J, Knoop V L, Sun J, et al, 2023. Unprotected left-turn behavior model capturing path variations at intersections [J]. *IEEE Trans Intell Transp Syst*, 24(9): 9016-9030.
- Zhong Z, Luo Y, Liang W, 2022. STGM: Vehicle trajectory prediction based on generative model for spatial-temporal features [J]. *IEEE Trans Intell Transp Syst*, 23(10): 18785-18793.

(责任编辑 王海蓉)