

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

黄峻 田永林 戴星原 王晓 平之行

Deep learning-based multimodal trajectory prediction methods for autonomous driving: state of the art and perspectives

HUANG Jun, TIAN Yonglin, DAI Xingyuan, WANG Xiao, PING Zhixing

引用本文：

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.

Jun HUANG, Yonglin TIAN, Xingyuan DAI, et al. Deep learning-based multimodal trajectory prediction methods for autonomous driving: state of the art and perspectives[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(2): 180-199.

您可能感兴趣的其他文章

行人轨迹预测方法关键问题研究：现状及展望

Key problems and progress of pedestrian trajectory prediction methods: the state of the art and prospects

智能科学与技术学报. 2023. 5(2): 143-162

3D 目标检测方法研究综述

A survey of 3D object detection algorithms

智能科学与技术学报. 2023. 5(1): 7-31

基于视觉 Transformer 的多损失融合水下图像增强网络

Underwater image enhancement network based on visual Transformer with multiple loss functions fusion

智能科学与技术学报. 2022. 4(4): 522-532

城市交通路网动态短时推理与精准预测研究

Short-term traffic state reasoning and precise prediction in urban networks

智能科学与技术学报. 2022. 4(3): 380-395

基于典型相关分析的多视图学习方法综述

A survey on canonical correlation analysis based multi-view learning

智能科学与技术学报. 2022. 4(1): 14-26

行人轨迹预测方法综述

Review of pedestrian trajectory prediction methods

智能科学与技术学报. 2021. 3(4): 399-411

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

黄峻^{1,2,3}, 田永林^{1,4}, 戴星原^{1,5}, 王晓^{3,6,7,8}, 平之行^{4,5,9}

(1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190;

2. 澳门科技大学, 澳门 999078;

3. 青岛智能产业技术研究院, 山东 青岛 266000;

4. 北京怀柔平行传感智能研究院, 北京 101400;

5. 中国科学院自动化研究所北京市智能化技术与系统工程技术研究中心, 北京 100190;

6. 安徽大学人工智能学院, 安徽 合肥 230601;

7. 自主无人系统技术教育部工程研究中心, 安徽 合肥 230601;

8. 安徽省无人系统与智能技术工程研究中心, 安徽 合肥 230601;

9. 北方自动控制技术研究所, 山西 太原 030006)

摘要: 对周围车辆轨迹的精确预测可以辅助自动驾驶车辆做出合理的即时决策。虽然相比传统轨迹预测算法, 深度学习已取得较好效果, 但是自动驾驶车辆在异构高动态复杂变化环境下实现多模态高精度预测仍存在信息丢失、交互和不确定性难以建模、预测缺乏可解释性等问题。Transformer具备的长距离建模能力和并行计算能力使其不仅在自然语言处理领域取得巨大成功, 而且在扩展至自动驾驶多模态轨迹预测任务时也解决了以上问题。基于此, 对过去基于深度神经网络的方法, 特别是对基于Transformer的方法进行全面总结与回顾; 同时分析了Transformer相较于传统序列网络、图神经网络、生成模型的优势, 并结合现有难题进行针对性分析与分类。Transformer模型可以更好地应用于多模态轨迹预测任务, 此类模型具有更好的泛化性和可解释性。最后, 对多模态轨迹预测未来发展方向进行了展望。

关键词: Transformer; 序列网络; 图神经网络; 生成模型; 轨迹预测; 多模态

中图分类号: TP29

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202317

Deep learning-based multimodal trajectory prediction methods for autonomous driving: state of the art and perspectives

HUANG Jun^{1,2,3}, TIAN Yonglin^{1,4}, DAI Xingyuan^{1,5}, WANG Xiao^{3,6,7,8}, PING Zhixing^{4,5,9}

1. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation,
Chinese Academy of Sciences, Beijing 100190, China

2. Macau University of Science and Technology, Macau 999780, China

3. Qingdao Academy of Intelligent Industries, Qingdao 266000, China

4. Beijing Huairou Academy of Parallel Sensing, Beijing 101400, China

5. Beijing Engineering Research Center of Intelligent Systems and Technology, Institute of Automation,
Chinese Academy of Sciences, Beijing 100190, China

6. School of Artificial Intelligence, Anhui University, Hefei 230601, China

7. Engineering Research Center of Autonomous Unmanned System Technology, Ministry of Education, Hefei 230601, China

8. Anhui Provincial Engineering Research Center for Unmanned System and Intelligent Technology, Hefei 230601, China

9. North Automatic Control Technology Institute, Taiyuan 030006, China

收稿日期: 2023-05-19; 修回日期: 2023-05-26

基金项目: 国家自然科学基金项目 (No.62173329)

Foundation Item: The National Natural Science Foundation of China (No.62173329)

Abstract: Although deep learning methods have achieved better results than traditional trajectory prediction algorithms, there are still problems such as information loss, interaction and uncertainty difficulties in modelling, and lack of interpretability of predictions when implementing multimodal high-precision prediction for autonomous vehicles in heterogeneous, highly dynamic and complex changing environments. The newly developed Transformer's long-range modelling capability and parallel computing ability make it a great success not only in the field of natural language processing, but also in solving the above problems when extended to the task of multimodal trajectory prediction for autonomous driving. Based on this, the aim of this paper is to provide a comprehensive summary and review of past deep neural network-based approaches, in particular the Transformer-based approach. The advantages of Transformer over traditional sequential network, graphical neural network and generative model were also analyzed and classified in relation to existing challenges, simultaneously. Transformer models can be better applied to multimodal trajectory prediction tasks, and that such models have better generalisation and interpretability. Finally, the future directions of multimodal trajectory prediction were presented.

Key words: Transformer, sequential network, graph neural network, generative model, trajectory prediction, multimodal

0 引言

作为人工智能领域的主要研究方向之一，自动驾驶正受到学术界和产业界的重点关注^[1-2]，自动驾驶在解决交通事故、堵车、排放等问题方面具有广阔的应用前景^[3]。近几年来，自动驾驶的感知、决策和控制模块取得了快速的发展，然而全面部署自动驾驶需要先验证其安全性。在复杂交通场景中，即使驾驶员最终意图确定，但决策过程中其他车辆的动态行驶路径的即时变化会导致自车路径呈现多模态属性，即车辆的未来轨迹有多种可能性。对车辆的多模态轨迹进行预测，并保证预测的准确性与多样性，是当前自动驾驶领域关注的热点内容。类比人类驾驶时通过观察周围交通参与者来预测其未来可能的行为，为进一步提高安全性，在自动驾驶研究中关键的技术之一是像人类驾驶员一样可以在高动态的复杂多变的场景中预测周围环境的未来状态，其未来状态可以由未来轨迹表示，即多模态轨迹预测任务。多模态轨迹预测旨在为处于异构复杂高动态环境中的目标车辆生成多条可能的轨迹，由于不确定性的存在，目标车辆即使在相同场景下也有可能表现不同，这也是多模态轨迹预测面临的挑战。目前深度神经网络（deep neural network, DNN）因其优秀的性能，已成为解决多模态轨迹预测任务的主流模型之一^[4]，并发展出不同的网络结构及范式^[5-11]。DNN在多模态轨迹预测任务中的应用主要分为3个主流方向：基于序列网络（sequential network）的方法、基于图神经网络（graph neural network, GNN）的方法以及基于生

成模型（generative model）的方法。

经典的序列网络可分为卷积神经网络（convolutional neural network, CNN）^[12-14]和循环神经网络（recurrent neural network, RNN）^[15-20]。其中，CNN通过卷积层和池化操作处理图像数据，在轨迹预测任务中用于处理鸟瞰图（bird eye view, BEV）等图像数据；RNN通过循环单元处理序列或时序数据，在轨迹预测任务中用于处理车辆历史轨迹信息。除了对环境信息和历史轨迹的处理，轨迹预测中车辆之间的交互信息也是至关重要的，基于注意力机制（attention mechanism, AM）^[21-23]的方法用于建模车辆之间的交互行为，根据车辆之间的相关性建立联系，并提取出对目标车辆未来轨迹影响最大的特征。

Transformer^[24]作为一种新的神经网络结构，最初在自然语言处理（natural language processing, NLP）领域取得了突破性的效果，目前在计算机视觉（computer vision, CV）和多模态（multi-modal）等多个领域中展现出了极大的潜力^[5-8]。相比于CNN和RNN，Transformer是完全基于自注意力机制的非循环网络结构，通过编码器-解码器以及注意力机制打破序列顺序输入的限制，从而解决了长短期记忆（long short-term memory, LSTM）^[17-18]和门控循环单元（gate recurrent unit, GRU）^[19-20]中存在的无法并行训练的问题。同时Transformer模型有更好的记忆力，可以解决长距离依赖问题。在处理自动驾驶轨迹预测任务中，基于Transformer轨迹预测在Argoverse数据集集中的平均位移误差（ADE）和最终位移误差（FDE）性能指标已

达到了最优水平^[25]。在处理交互问题时, Transformer模型有更好的可解释性^[26]。因此本文将详细介绍基于Transformer的轨迹预测方法,并根据轨迹预测难点问题分类。

虽然传统的序列网络方法在提取欧氏空间数据的特征方面取得了巨大的成功,但是许多实际应用场景中的数据是从非欧氏空间生成的,基于序列网络的方法在处理非欧氏空间数据上的表现不尽如人意^[26]。在轨迹预测问题中,虽然CNN、RNN等方法可用于轨迹预测中的时间空间建模,但无法有效建模目标车辆间的交互。图神经网络^[27-28]是一种用于处理图数据的神经网络,其核心思想是学习映射函数,通过该映射函数提取交互感知特征,GNN通过将卷积运算从传统图像数据推广到图数据,从而捕获时空预测问题中利用非欧图结构展示的空间关联,在轨迹预测任务中用来对图结构和环境信息(contextual information)建模。同时可以通过该映射函数从图中节点的特征及其邻近节点的特征中建模目标车辆与环境的交互。由于每个节点的相邻节点数量不同,环境中的每个参与者都可以被视为图的节点,每个场景都是不规则图,因此在图中直接做卷积计算是不可行的。GNN可以通过节点间相连的边获取每个交通参与者之间的依赖关系,进而建模目标车辆与周围车辆的交互关系^[29-40]。但此类模型计算量过大,且生成的轨迹可解释性差。

在轨迹预测任务中,驾驶环境和驾驶策略存在不确定性,导致最终的预测轨迹是具有不确定性和多模态性的,生成模型主要用于解决不确定性和多模态性预测问题。现有的生成模型方法可分为两类:生成对抗网络(general adversarial network, GAN)^[41-45]和条件变分自编码器(conditional variational autoencoder, CVAE)。GAN由生成器和鉴别器组成。生成器用于生成与真实样本相似的随机样本,鉴别器用于确定样本是否为真实样本,通过生成器和鉴别器之间的相互博弈优化模型。在轨迹预测任务中,生成器用于生成预测轨迹,鉴别器用于判断轨迹是否为真实轨迹(ground truth, GT)。CVAE^[46-49]通过编码器将数据压缩为低维向量表示,并使用解码器解码低维向量以获得重构输出,然而此类模型存在模式崩溃问题。

在现有针对轨迹预测任务的综述文献中,

Lefèvre等人^[50]从基于物理的运动模型、基于机动性的运动模型和基于交互感知的运动模型3个层次调研了2014年以前的轨迹预测及风险评估方法。其中,基于物理的模型在低水平上运行,假设汽车的速度和方向不变,车辆的运动只取决于物理学规律,虽然对风险进行了有效的计算,但仅限于短时间碰撞预测。基于机动性的方法对长期运动和风险都有稳定估计,但忽略了车辆之间的依赖关系,导致预测结果并不是完全可靠的。基于交互运动的模型虽然增加了车辆之间的依赖性,但是计算复杂性太高,并不能很好地与实时风险兼容。同时,Lefèvre认为风险评估方法的选择与运动模型的选择是紧密相连的。González等人^[51]、Liu等人^[52]和Mozaffari等人^[53]都对轨迹预测方法进行了总结。其中,González等人从基于图搜索的方法、基于采样的方法、基于插值曲线的方法以及基于数值优化的方法概述了自动驾驶运动规划技术在超车和避障方面的具体实现方法。Liu等人从表征、建模和学习的角度对现有基于学习的轨迹预测方法进行了介绍,从输入模式、输出模式及模型建模方面进行了评述及分类,并从特征编码、交互建模、预测及生成模型4个角度对建模方法做了进一步介绍,详细介绍了TNT方法的实现过程。Mozaffari等人基于CNN、RNN、LSTM方法,从输入表示、输出类型及预测方法3个方面对现有的方法进行分类,总结各方法的优缺点,并使用均方根误差指标(root mean squared error, RMSE)评估了部分算法的性能。然而这4篇文章都不涉及最新发布的算法,对比的算法较早,且对比的指标不全面。

Leon等人^[54]从跟踪、预测两个角度对基于深度神经网络的方法和随机技术(stochastic techniques)方法进行了回顾。其中跟踪方法主要以基于CNN、RNN、LSTM、GAN的方法为主,介绍了其在行人、车辆和障碍物识别任务中,解决噪声处理、鲁棒计算、多目标识别、点云融合问题的具体方法。轨迹预测方法主要以长尾问题、意图推理、计算效率问题为切入点,从基于模型的方法、基于数据驱动的方法两个角度进行分类并介绍。

Gomes等人^[55]分析了自2008年以来针对机动意图和相互作用如何提高估计预测性能的研究,对基于驾驶意图和交互的方法进行了总结。在这篇综述中,作者还从交通规则、多模态轨迹预测、交通

状况等方面对如何优化现有模型提出构想。Shirazi 等人^[56]将问题聚焦于路口场景，介绍了车辆、驾驶员和行人3类参与者在交叉口场景的安全性评估及行为分析。作者将车辆间隙、不确定性导致的风险程度以及碰撞风险作为安全性的评估指标，并基于该指标对阐述方法进行分类。Shirazi 等人认为基于贝叶斯框架的概率方法能够很好地处理不确定性问题。

上述文献从不同角度对早期自动驾驶轨迹预测方法进行了介绍。然而，这些文献都没有对基于Transformer的多模态轨迹预测方法进行详细介绍，其对数据集的分类及介绍也不全面。Huang 等人^[57]基于深度学习、强化学习^[58]的轨迹预测方法，将轨迹预测问题进行分类，该文章虽然包含了部分最新的文献，但是没有深入讨论多模态轨迹预测的热点问题及解决方案。

因此本文的主要贡献在于笔者将全面回顾10年内基于深度学习的多模态轨迹预测方法，并结合热点问题总结，分析各方法的优势，同时对基于Transformer的多模态轨迹预测方法进行详细介绍与分析；此外针对具体问题，对基于Transformer的方法进行分类与总结。

后文组织如下：第1节描述轨迹预测问题；第2节总结数据集及评估指标；第3~5节结合轨迹预测中存在的问题，分别介绍基于序列网络、图神经网络及生成模型的方法，并分析每种方法的优缺点；第6节对自动驾驶的问题与挑战进行分类总结；第7节总结全文并展望了自动驾驶轨迹预测的发展趋势。

1 轨迹预测问题描述

1.1 模型输入与输出

轨迹预测任务可以被表示为在某个场景中，给定一段预测时间范围 T_h 、智能体的历史轨迹 X 及环境信息 M ，预测其在未来一段时间范围 T_f 的轨迹。模型的输入一般为历史轨迹 X ，其由目标车辆的物理状态 S_0 及环境车辆的物理状态 $(S_1, S_2, S_3, \dots, S_N)$ 组成，其中车辆物理状态由速度、加速度、转向角，以及当前环境信息 M 组成。通常情况下，假设目标车辆周围有 N 个交通参与者（如周边车辆、自行车、行人）。

$$X = \{p^1, p^2, \dots, p^{T_h}\} \quad (1)$$

其中， $p^t (t \in 1, 2, \dots, T_h)$ 表示时间范围为 t 的历史轨迹， $p^t = (x_t, y_t, M, S_n)$ ，其中 x_t, y_t 表示当前车辆的坐标， p^{T_h} 表示当前目标车辆的状态。

在轨迹预测任务中，模型的输出大多为未来 T_f 个时间步的预测轨迹 Y ：

$$Y = \{p^{T_h+1}, p^{T_h+2}, \dots, p^{T_h+T_f}\} \quad (2)$$

1.2 轨迹预测主要研究问题

轨迹预测主要的挑战来自环境的不确定性、生成轨迹的多模态性以及模型的可解释性。这一部分将针对这些问题进行讨论。

(1) 多模态问题

在轨迹预测过程中，周围交通参与者的未来行为难以预测，其未来轨迹是多模态的（如图1所示），因此实际预测的未来轨迹不是唯一的，即在相同历史轨迹条件下，车辆的未来轨迹具有多种可能性。因此预测模块需要对不确定性进行建模^[59-60]，并预测其他车辆多条可能的轨迹。多模态轨迹预测就是预测模型根据每个未来轨迹设置相应的置信度，并为每个交通参与者生成多条未来轨迹。早期自动驾驶轨迹预测模型以单峰为主，即给定确定的历史轨迹，预测算法只输出一条确定的轨迹。然而单模态轨迹预测无法充分捕捉复杂场景中的可能性，且无法保证预测精度及后续与规划模块联合后的安全性。Deo 等人^[61]将混合密度网络（mixture density network）用于高速路轨迹预测任务，该模型将周围车辆的历史轨迹以及高速公路的拓扑结构作为输入，将置信度值分为6个机动类别，将每种驾驶动作定义为预测模式，其中每种运动模式都与一个高斯分量相匹配，进而输出未来估计的多模态分布。然而轨迹数据模式众多且复杂^[62]，因此参考文献^[61]仅通过预先定义的动作并不能保证轨迹多样性。因此在后续研究中，研究人员以不确定性为前提条件，生成符合真实驾驶场景的多模态轨迹预测。本文将重点介绍基于深度学习的多模态轨迹预测模型。

(2) 可解释性问题

基于深度学习的研究通常最终获得“黑盒”深度学习模型，导致模型太复杂且无法被人类理解或排除故障^[63]。特别是在医疗诊断、规划、控制问题的研究中，所提出的模型需要与其输出建立一定程度的信任关系，而信任关系的概念取决于人类对模型或机器工作的可见性^[64]，即深度神经网络应

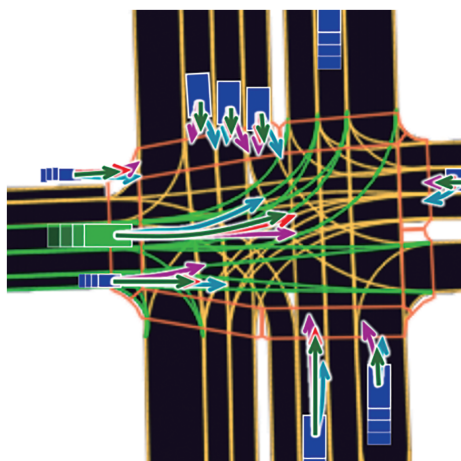


图1 多模态轨迹预测可视化结果^[26]

该为其输出人类可以理解的理由,从而使人类洞悉其内部工作原理,这种模型被称为可解释模型。可解释性AI(XAI)系统的目的是通过提供解释使其行为更容易被人类理解^[65]。因此笔者认为,可解释性在自动驾驶轨迹预测与规划任务中是必不可少的。

目前的研究将可解释性问题^[66]分为内在可解释性、事后可解释性。内在可解释性即训练一个简单的可解释的模型,如线性模型、决策树、基于规则的模型等;事后可解释性即训练一个复杂的优化模型,在后续研究中通过技术发展解释该模型。事后可解释性又分为局部可解释性(即理解每个决策背后的过程和基础)和全局可解释性(即理解模型背后的机制和内部逻辑)。

针对预测问题,大部分研究对因果关系的解释感兴趣。对于可解释的XAI系统,是指寻找模型输入和输出之间的因果关系。在现有研究中,严格的定义因果关系和识别潜在的因果关系往往是不全面且不符合实际情况的。对于此问题,通常通过检测容易识别的输入-输出对来获得其中的因果关系^[67]。Yuan等人^[68]提出了在不预先定义模式的情况下获得不同预测模式的解决方案,然而此方法改变了优化目标,不再使预测可能性最大化,导致模型可解释性差。基于注意力机制的方法在构建可解释性模型方面具有优势,因为此类模型在训练过程中自然地集中在输入特征的少数部分,可以计算不同部分的重要性分数,并通过热图(heat map)的方式突出现实重要的输入因素。可视化也是目前有助于直观理解高维输入的工具,例如Kim等人^[69]利用基于CNN的注意力模型进行额外的显著性过

滤,从而可视化对转向角有因果影响的区域。

然而,目前还没有合适的可解释性度量方法来判断XAI系统是否比非XAI系统更容易使用户理解,其中一些衡量标准是根据主观判断衡量的,对于输出解释的有效性通常通过模型完成任务的表现来客观衡量。

第3~5节将介绍针对此两类问题的多模态轨迹预测的方法。

2 数据集及评估指标

2.1 数据集

车辆轨迹在交通流分析、驾驶行为分析、决策模型设计、训练、测试,以及路径规划方面都有很重要的作用,因此选择合适的数据集对于模型性能的提升至关重要。早期NGSIM数据集中交通参与者只有高速路场景下的车辆,数据内容只有历史轨迹,将此类数据集作为各轨迹预测算法模型的输入显然是不全面的,环境复杂度低导致早期轨迹预测方法难以生成多模态的预测轨迹。后期出现的数据集中添加了环境信息更丰富的高清地图(HD map)、多种交通参与者(行人、自行车)及场景(路口、环岛等),使近期的模型在复杂高动态、不确定性高的场景中有较好的预测精度及泛化性。本节将介绍目前轨迹预测任务中常用的数据集,并列出了使用这些数据集的经典算法(见表1)。

(1) NuScenes^[70]

该数据集通过Renault Zoe迷你电动车采样,收集来自不同城市的1 000个场景。传感器包括6个相机、1个激光雷达、5个毫米波雷达、IMU和GPS,整个数据集包含了由人工挑选的84段记录,时长约15 h,距离约242 km,平均车速为16 km/h。场景覆盖了城市、住宅区、郊区、工业区等,也涵盖了白天、黑夜、晴天、雨天、多云等不同时段和不同天气状况。最终该数据集被分为1 000个片段,每个片段约20 s。

(2) Lyft^[71]

该数据集是目前最大的智能交通数据集合。该数据集由23辆车耗费超过1 000 h驾驶采集,具体内容主要分为4个部分:超过55 000个人工注释框架;由7个摄像头和3个激光雷达传感器收集的点云数据;可实现安全驾驶的地图;一个包含超过4 000个车道段、197条人行横道、60个停车标志、54个停车区的基础HD空间语义地图。驾驶轨迹包

表1 用于自动驾驶轨迹预测任务的数据集

数据集	年份	交通参与者	传感器	场景	数据内容	基于此数据集的算法
NuScenes ^[70]	2020年	车辆、行人、自行车	雷达、相机	城市	驾驶轨迹、高清地图	MHA-JAM
Lyft ^[71]	2020年	车辆、行人、自行车	雷达、相机	城市	驾驶轨迹、高清地图	Gatformer
Waymo ^[72]	2020年	车辆、行人、自行车	雷达、相机	城市	驾驶轨迹、高清地图	Wayformer、DenseTNT
Argoverse ^[73-74]	2019年	车辆	雷达、相机	无信号灯路口	驾驶轨迹、高清地图	mmTransformer、HOME、TrajFormer、VectorNet
INTERACTION ^[75]	2019年	车辆、行人	无人机、相机	高速公路、无信号灯路口、有信号灯路口、环岛	驾驶轨迹、高清地图	GOHOME、THOMAS、TNT
HighD ^[76]	2018年	车辆	无人机	城市路口	驾驶轨迹、车道	MHA-LSTM
ApolloScape ^[77]	2018年	车辆、行人、自行车	雷达、相机	城市 高速公路	驾驶轨迹	AI-TP、S2TNet、GRIP++
KITTI ^[78]	2013年	行人、自行车	雷达、相机	城市 高速公路	图像、点云	DESIRE
NGSIM	2006年	车辆	相机	高速公路路口	驾驶轨迹、车道	Maneuver based LSTM、GRIP++

括与自动驾驶车辆在同一交通场景下的车辆、自行车、行人和交通参与者的历史轨迹，适合训练运动预测模型。

(3) Waymo^[72]

该数据集由装载5个LiDAR和5个高分辨率针孔相机的车辆收集完成。所有相机图像都进行下采样处理，并从原始图像中裁剪出来。其中，每个场景为20 s、10 Hz，共103 354个场景。驾驶轨迹结合地图信息被分成每个时间长度为9 s的窗口，包括1 s历史轨迹和8 s未来轨迹。

(4) Argoverse^[73-74]

该数据集是由Argo AI、卡内基梅隆大学、佐治亚理工学院发布的用于支持自动驾驶汽车3D目标跟踪和运动预测研究的数据集。该数据集收集了美国匹兹堡和迈阿密驾驶时长共1 006 h的驾驶数据，包括城市场景中的3D跟踪数据集和运动预测数据集，采样率为10 Hz，共有2亿多条轨迹。这些轨迹被分割成5 s的序列。

(5) INTERACTION^[75]

该数据集由无人机和交通摄像机采集。NGSIM和HighD数据集驾驶行为多样性低、环境复杂度有限、可获得的环境信息和交互信息不完整等问题，导致早期基于这两个数据集的模型无法实现预测轨迹多模态化。因此该数据集中增加了交互场景地多样性；它能准确再现不同国家各种驾驶场景中道路使用者（如车辆、行人）的大量交互性行为，并将激进的驾驶风格加入其中，驾驶场景不再是理想化的无碰撞/少碰撞场景，提高了场景复杂度。针对不同驾驶场景采集不同的车辆轨迹（如环岛场景记

录了来自5个不同位置的10 479辆车辆的轨迹，持续约365 min）。

(6) HighD^[76]

该数据集是德国亚琛工业大学汽车工程研究所提出的由无人机从空中采集的来自6个地点、11.5 h测量值、110 000台车辆的数据集。测量的车辆总行驶里程为45 000 km。该数据集还包括5 600条完整的变道记录。

(7) ApolloScape^[77]

该数据集是从中国北京城市街道收集的大型数据集，主要分为3个部分：仿真数据集、演示数据集和标注数据集。轨迹数据集的采样率为2 Hz，由53 min的训练序列和50 min的测试序列组成，从而产生了大约10万条轨迹。此数据集中涉及5种不同类型的交通参与者，包括车辆、行人、摩托车、自行车以及其他对象。

(8) KITTI^[78]

该数据集由从市区、乡村和高速公路等场景中采集的真实图像数据组成，其中每张图像最多包含15辆车、30个行人及各种程度的遮挡与截断。整个数据集由389对立体图像和光流图、39.2 km视觉测距序列以及超过200 k个3D标注物体的图像组成，以10 Hz的频率采样。

(9) NGSIM^[79]

该数据集是美国联邦公路管理局收集的美国高速公路行车数据，它包括US101、I-80等道路上的所有车辆在一个时间段的车辆行驶状况。数据集包含4个带有交通信号灯的路口，每个方向的道路包含3条或者4条车道。摄像头采集到的原始视频数

据时长约为9 h, 分为早晚两个高峰录制。数据集在45 min内以10 Hz(每秒10帧)的频率进行采样, 从而产生800多万条轨迹。这两个数据集被分为训练数据集和测试数据集, 其中四分之一的数据被用于测试。每个轨迹被分割成8 s的序列, 其中前3 s用作观察的历史轨迹, 剩余5 s作为地面真实轨迹。

2.2 评估指标

车辆轨迹预测方法通常采用以下评估指标。

(1) 平均位移误差 (average displacement error, ADE): 预测轨迹与实际轨迹之间的平均欧氏(L2)距离。

$$ADE = \frac{1}{N_p \times T} \sum_{i=1}^{N_p} \sum_{t=1}^T |Y_{\text{pred}}^t[i] - Y_{\text{GT}}^t[i]| \quad (3)$$

其中, N_p 表示所有预测对象, T 表示时间。对于多模态预测, 最小 ADE (minimum ADE, mAde) 表示 K 个预测中 ADE 的最小值。 Y_{pred}^t 、 Y_{GT}^t 分别表示在 t 个时间步长的预测结果和真实轨迹。

(2) 最终位移误差 (final displacement error, FDE): 最终预测轨迹与实际轨迹之间的 L2 距离。

$$FDE = \frac{1}{N_p} \sum_{i=1}^{N_p} |Y_{\text{pred}}^t[i] - Y_{\text{GT}}^t[i]| \quad (4)$$

其中, Y_{pred}^t 、 Y_{GT}^t 分别表示在 t 个时间步长的预测结果和真实轨迹。对于多模态轨迹预测, 最小 FDE (minimum FDE, mFDE) 表示 K 个预测中 FDE 的最小值。

(3) 丢失率 (miss rate, MR): 基于 L2 距离, 最终预测轨迹与真实轨迹偏差 2.0 m 的数量的比率。

(4) 均方根误差 (root mean squared error, RMSE): 预测误差平方平均值的平方根。RMSE 对较大的预测误差敏感, 是常用的轨迹预测指标之一。

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_{\text{pred}}^t - Y_{\text{GT}}^t)^2} \quad (5)$$

(5) 负对数似然 (negative log likelihood, NLL): 对于建模的轨迹分布 $f(Y)$, 有:

$$NLL = -\log(f(Y)) \quad (6)$$

其中, Y 表示真实轨迹; NLL 不是物理量, 相比用于计算模型平均误差的 RMSE, NLL 用于确定基于意图 (maneuver-based) 模型的轨迹预测的正确性。

(6) 预测范围 (prediction horizon): 指模型可以预测的未来时间步长。通常预测时间步越长, 预测的精度越低。为了满足后续规划和控制, 预测的时间步不宜太短, 一般模型将预测范围为 5 s 或 10 s 的预测轨迹作为输入, 并且各模块内时间步应保持一致。

(7) 计算时间 (computation time): 由于轨迹预测模型复杂度高, 因此需要大量的计算资源。为达到更高的自动驾驶水平, 每个模块的计算速度必须快, 以此减少延迟, 因此计算时间对于模型能否实现实车部署很重要。

(8) 可驱动区域顺应性 (drivable area compliance, DAC)^[79]: 将可驾驶区域内未来轨迹的数量与所有可能的轨迹数量的比率作为评估所提出解决方案的可行性的指标。

2.3 小结

本节介绍了常用的自动驾驶轨迹预测任务数据集以及在这些数据集下的经典方法。近期公开的数据集中交通参与者种类有明显增加, 同时引入了高清地图数据, 从而研究者可以从这些数据集中获取更复杂的环境信息。然而现有数据集并没有针对长尾问题提供相应的数据, 数据集场景过于理想化; 此外, 驾驶员驾驶风格单一, 对于不同驾驶员风格没有明确的分类, 使得训练后的轨迹预测方法难以准确预测在有激进驾驶风格的驾驶员存在的场景中的未来轨迹。

然后, 本节总结了评估指标。现有研究大多以预测距离与真实距离的偏差为评估指标, 同时预测范围和计算时间对于提高模型的性能以及最终能否应用于真实驾驶环境也非常重要。在现有预测范围的基础上, 目前没有太多对预测时间范围与预测精度之间因果关系的定量研究, 预测时间应控制在什么范围内可以使预测精度最高难以评估。对于预测模型的可解释性也缺乏定量评估标准。

3 基于序列网络的轨迹预测方法

本节介绍基于序列网络的多模态轨迹预测方法, 并分析这些方法的典型特征和优劣性。其中, Transformer^[24]作为一种新的神经网络结构, 因为其具有并行计算能力以及可以解决长距离依赖问题的能力, 在多模态轨迹预测任务中表现出了非常好的性能。本节将重点介绍并分析基于 Transformer 的方法, 同时基于问题对基于 Transformer 的方法进行分类。

3.1 基于卷积神经网络的轨迹预测方法

CNN 一般为由序列到序列 (sequence-to-sequence) 的结构, 在轨迹预测任务中通常被用于处理空间信息, 将历史轨迹作为输入, 通过将卷积层与全连接层叠加的方式, 输出具有强时空连续性的未来轨迹^[80-81]。早期基于 CNN 的方法^[82-84]将鸟瞰图作为输入, 通过卷积层, 在训练过程中提取场景上下文信息。在基于 CNN 模型的后续发展中, 为解决无法显示编码运动约束的问题, Phan-Minh 等人^[83]通过创建轨迹集的方法, 排除不符合物理约束的轨迹, 隐式估计物理特征, 并直接对可能的轨迹分类; 将输出定义为相对于初始状态的运动 (左转、右转、加速等), 确保在较大速度范围内, 输出轨迹的多样性。早期研究中没有明确编码车辆运动学, 而是依靠模型直接从数据中学习约束, 导致运动学不可行问题。Cui 等人^[84]为解决此问题, 将基于物理的车辆运动学模型^[85]与 CNN 结合, 将包含场景上下文信息的光栅图作为 CNN 的输入, 提出深度运动学模型, 以实现更精准的位置和航向角预测。Marzinotto 等人^[86]使用卷积社会池化 (convolutional social pooling, CSP) 方法鲁棒地学习车辆运动中的社会依赖性, 通过 LSTM 编码器-解码器学习车辆的运动动力学, 编码最近时刻的历史轨迹, 用于预测车辆及其周围所有车辆多模态未来轨迹。然而该方法只依靠车辆历史轨迹预测未来轨迹, 并没有结合视觉或地图信息, 导致预测精度不高, 而且该方法假设周围车辆数量是恒定的, 无法泛化到高动态复杂场景中。

针对长期轨迹预测问题, 现有基于 CNN 的方案使用其变体模型, 如 Zhang 等人^[87]使用时间卷积网络 (temporal convolution network, TCN) 预测长期变道轨迹和变道行为, TCN 能够准确预测长期轨迹, 计算时间更短。近期基于 CNN 的研究中, Gilles 等人^[88]提出了 HOME 算法, 使用高清地图的图形表示和系数投影作为输入, 使用 CNN 输出轨迹热图。基于 HOME, Gilles 等人后续提出了 GOHOME 算法^[89]和 THOMAS 算法^[90]: GOHOME 算法显著减少了内存负担, THOMAS 算法实现了多智能体轨迹预测。Ye 等人^[91]提出时态点云网络 (temporal point cloud networks, TPCN) 模型, 通过扩展点云学习及动态学习, 将轨迹预测分割为时间维度和空间维度来捕获时间信息和空间信息。其中, 在空间维度中, 车辆被视为无序点集, 可以直

接利用点云学习建模车辆位置, 进而预测车辆运动。

本节主要介绍了基于序列网络的轨迹预测方法, 现有研究中大多将鸟瞰图作为输入, 通过 CNN 输出预测轨迹。但是此类方法并不能很好地考虑环境不确定性及交互问题; 另外, CNN 难以捕捉到细粒度的场景特征, 如车道的几何形状和交通规则。

3.2 基于循环神经网络的轨迹预测方法

与 CNN 不同, RNN 在轨迹预测任务中能够处理序列变化的数据, 其主要用于处理时序信息。RNN 可以存储历史时间段内的信息, 通过将隐藏状态和输入相结合来输出未来轨迹。轨迹预测中使用 RNN 模型由早期使用单个 RNN 模型扩展到使用多个 RNN 模型。在处理长时间片段或长历史轨迹时, 使用单个 RNN 模型会出现梯度消失或梯度爆炸问题, 为解决此类问题, 现有方法通常使用将 LSTM^[17-18]或 GRU^[19-20]与 RNN 相结合的方法。本节将对基于 RNN-LSTM 和 RNN-GRU 的多模态轨迹预测方法进行介绍。

RNN 是链式结构的神经网络, 其中长短期记忆网络 (LSTM) 是一种时间循环神经网络, 其可以记忆历史输入, 因此 LSTM 可用于解决 RNN 存在的长期依赖问题。在轨迹预测任务中, LSTM 基于连续处理序列和存储的潜在状态来建模关于交通智能体的运动信息^[92], 将目标智能体的历史轨迹作为基于 LSTM 的编码器-解码器模型的输入, 将最后一个单元的隐藏状态反馈到输出层以预测未来位置。Altché 等人^[93]使用单个 LSTM 层抽象出输入数据中有意义的时间序列表示, 进而预测高速公路车辆未来的横向和纵向轨迹, 解决行驶速度不平衡问题。Zyner 等人^[94]使用 3 层 LSTM 获得加权高斯混合模型 (GMM) 参数, 使用概率最高的模态对多条预测轨迹进行聚类。Park 等人^[95]使用 LSTM 编码器对历史轨迹进行编码, 使用 LSTM 解码器搜索算法, 使解码器输出的未来轨迹保持局部最优解, 从而在占用图中生成多条最有可能的未来轨迹。然而仅依靠单层 LSTM 的方法无法实现车辆间的交互建模。因此 Deo 等人^[61]通过将卷积社会池与 LSTM 组成的编码器结合的方法, 使用 CSP 提取交互信息和空间信息, 并将其作为 LSTM 的输入, 解码器结合 CSP 和 LSTM 的输出联合预测车辆的未来轨迹。Song 等人^[96]在此基础上, 在输入中添加规划轨迹,

优化最终的预测轨迹。Xin 等人^[97]使用两种 LSTM，通过横向特征评估每一次机动的可能性，然后将此可能性作为预测未来轨迹的依据。Shirazi 等人^[98]设计了 7 层堆叠的 LSTM 和密集层组成的深度 LSTM 模型，预测车辆在路口的机动。Xing 等人^[99]基于 LSTM 提出联合时间序列建模的方法——JTSM，使用高斯混合模型 GMM 区分驾驶风格，并通过 LSTM 和全连接回归层区分序列数据和驾驶风格预测未来轨迹。Zhang 等人^[100]通过双层 LSTM 推测路口场景驾驶员意图与多模态轨迹预测。

针对交互问题，现有方法采用交互感知方法估计先验轨迹，然后使用不同方法进行细化。Bae 等人^[101]将问题聚焦到交通繁忙情况下的匝道汇入场景，通过 RNN 预测交互运动，并将其纳入安全约束中，基于实时卷积的启发式算法找到局部最优轨迹。Mo 等人^[102]提出交互感知方法，将周围 8 辆车之间的交互作用引入模型中。Huang 等人^[103]设计了基于 LSTM 和注意力机制的编码器-解码器结构预测骑车人的未来轨迹。Wang 等人^[104]将知识推理与多层 LSTM 相结合，通过交通规则约束实现轨迹预测的合法性、安全性以及通过本体对场景建模。Ju 等人^[105]提出交互感知-卡尔曼神经网络（IaKNN），交互感知层将高维环境感知解析为加速度，运动层将加速度转换为感知轨迹，通过卡尔曼滤波网络输出未来轨迹。GRU 是 RNN 神经网络中用于处理交互问题的主要方法，在现有研究中，GRU 通常与注意力机制结合。Liu 等人^[106]基于 LSTM 设计时空注意力模块，根据周围车辆对自动驾驶车辆的影响程度进行排序，并识别最重要的特征点。Shi 等人^[107]提出的 TrajGRU 方法可以主动学习递归连接的位置变量结构。Ding 等人^[108]通过 GRU 方法从交互车辆的潜在状态和连接特征中获取交互。

本节主要介绍了基于 RNN 及其扩展的方法，然而这些方法在处理交互问题时，存在梯度消失及长时间依赖问题，导致模型预测精度受到影响。在现有交互问题研究中，学者更倾向于使用 LSTM 与注意力机制结合的方法。

3.3 基于注意力机制的轨迹预测方法

与处理空间-时间问题的 CNN、RNN 方法不同，基于注意力机制的方法是受到人类在处理大量信息时能够快速聚焦于高价值因素的认知系统的启发所提出的方法。注意力模型能够在需要的时间和地点选择性地集中于输入特征中最感兴趣的部分，

忽略其他可感知的特征，因此将注意力机制应用于解释多模态轨迹预测成为目前研究的热点方向。Transformer 作为一种新的基于编码器-解码器结构的神经网络，在多模态轨迹预测任务中可以将历史轨迹与环境信息结合。本节首先介绍基于注意力机制的轨迹预测方法，然后介绍 Transformer 一般性框架，并对基于 Transformer 的轨迹预测方法进行梳理。

Messaoud 等人^[21-23]是最早使用多头注意机制建模车辆间交互的团队。Messaoud 等人最初^[21]使用多头注意机制计算周围车辆轨迹特征的加权和，选择性地建模周围车辆间的交互特征，基于车辆间的相关性，通过多头注意机制将交互关系分类，以此捕获更高阶的关系。由于环境不确定性的存在，基于此方法，Messaoud 等人提出能够预测多模态有限轨迹集的方法^[22]，此方法使用了基于 LSTM 的编码器-解码器预测目标车辆的轨迹，这些轨迹对应于周围车辆的注意程度的预测轨迹。Mercat 等人^[109]使用多头自注意机制建立可解释所有智能体之间的完整交互的模型，该方法仅依赖历史轨迹，不依赖空间网格图表示的场景，并在 NLL 指标上取得了最好的效果。Messaoud 等人^[23]通过使用智能体-场景交互地图以及多头自注意机制生成多模态未来轨迹。Zhang 等人^[110]提出基于注意的交互感知方法 AI-TP，采用编码器-解码器结构，使用图注意神经网络建模车辆间的交互，并使用 GRU 预测未来轨迹。Tang 等人^[111]针对多智能体行为建模和多模态轨迹预测中使用 CVAE 导致的模式崩溃问题（即模型在预测智能体未来轨迹时会忽略历史社会背景，导致显著的预测错误和较差的泛化性），提出系数图注意力信息传递层（sparse graph attention message-passing layer, Sparse-GAMP layer），用于检验实验中的严重后验崩溃问题，提高模型泛化性。基于注意力机制和编码器-解码器结构的 Transformer 和双向 Transformer（bidirectional encoder representation from Transformers, BERT）^[112]被证明在长期预测方面有较好的性能，并能够处理真实传感器中普遍存在的输入数据缺失问题。后文将对 Transformer 模型进行详细介绍，并介绍基于此模型的多模态轨迹预测方法。

3.3.1 Transformer 框架

2017 年，Vaswani 等人^[24]提出 Transformer 模型，这是一种完全基于自注意力机制的模型。自注意力

机制是一种捕捉向量之间相关性的方法，既可以考虑全局，又可以聚焦重点，其在捕获车辆之间交互信息方面有非常好的性能，自注意力机制使得模型的预测轨迹有很好的可解释性。Transformer框架主要包含编码器、解码器、多头自注意力机制3个重要部分，如图2所示。

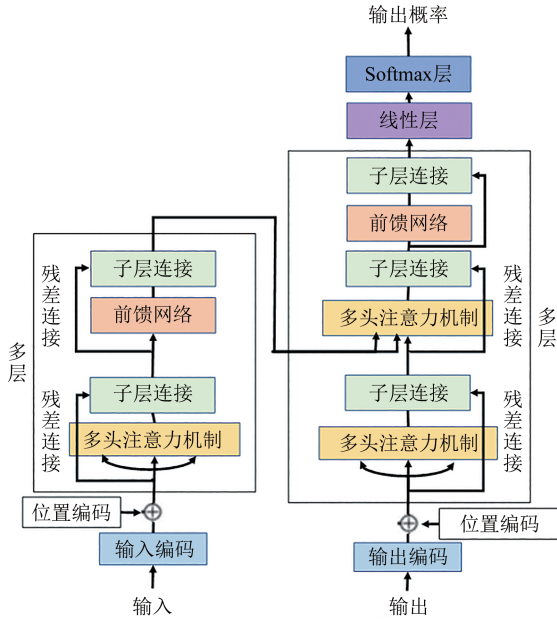


图2 Transformer框架 [24]

（1）编码器-解码器

编码器用于将历史轨迹和环境信息嵌入上下文信息中，并输入Transformer中，其输入为车道信息、历史轨迹、车辆交互信息等，输出为具有这些信息的特征。编码器由多个独立层组成，每层有两个子层，分别是多头注意力和全连接前馈网络，子层通过残差连接后进行归一化输出。

解码器用于生成预测轨迹，其输入为编码器的输出，输出为预测轨迹。除了多头注意力和全连接前馈网络，还插入第3个子层——掩码多头注意力（masked multi-head attention），用于对编码器堆栈的输出执行多头注意，掩码用于未来时刻进行掩码处理，确保当前位置的预测不会依赖于未来位置，同时可确保预测的多模态性。

（2）Transformer的注意力机制

注意力机制用于建模车辆间的交互关系。注意力机制将查询向量 Q （query）、键向量 K （key）和值向量 V （value）映射到输出，输出值的加权和，权重则是通过 Q 和 K 相似度计算得到。Transformer

框架主要由缩放点积注意力机制和多头注意力机制组成。缩放点积注意力机制中输入由向量 $query(d_k)$ 、 $key(d_k)$ 以及 $value(d_v)$ 组成， Q 、 K 通过点积处理计算相似度，通过比例因子 $\sqrt{d_k}$ （用来求 d_k 的平方根）处理避免 Q 、 K 内积方差太大导致的难以学习的情况，应用softmax函数获得value的权重。掩码处理避免解码器在训练时获取未来的信息影响预测。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (7)$$

多头注意机制通过将 Q 、 K 、 V 分别线性投影到缩放点积注意机制中，投影 h 次后做 h 次注意力函数运算，通过并行计算，生成 d_v 维输出value，将每一个输出值连接后再做一次投影，得到最终value。通过多头注意机制，Transformer模型可以联合注意来自不同位置的不同子空间信息。

（3）小结

本节主要介绍了Transformer框架中的3个主要部分：编码器、解码器、注意力机制的输入输出及其在轨迹预测中的用途。后文将对基于Transformer的多模态轨迹方法进行介绍。

3.3.2 基于Transformer的多模态轨迹预测方法

本节将介绍近年来基于Transformer框架的可随场景变化的自适应调整的多模态方法。实现多模态预测的另一个挑战在于如何用有限的训练样本覆盖给定场景中所有可能的结果。多智能体轨迹预测需要在两个关键维度建模。①时间维度：将历史信息对智能体未来状态的影响进行建模；②社会维度：对每个智能体之间的交互关系进行建模。在时间维度层面，现有基于经典深度学习的模型CNN、RNN无法建模长时间序列，会导致时间信息丢失问题，Transformer模型可以通过将位置编码、时间编码的形式保存长历史轨迹的信息。在社会维度层面，Transformer模型可以通过注意力机制建模人-车、车-车、车-环境之间的交互关系，通过分配权重的方式选择影响力最大的交互，以此为基础，Transformer可扩展到多智能体交互环境中。本节将轨迹预测难点问题分类，并对相应问题的解决办法进行介绍，见表2。

现有基于概率的方法^[113]和基于建议的启发式的方法^[114]虽然可以通过添加规则的方式输出概率分布或通过添加具有强约束的锚点，实现多模态轨

表2

多模态轨迹预测问题及基于 Transformer 模型的解决办法

研究问题	解决办法
时间、空间维度编码问题： 时间维度、空间维度信息如何编码？ 是否会有信息丢失？	mmTransformer: 通过使用3个堆叠Transformer对环境信息、历史轨迹及车辆特征信息分别编码； LAformer: 时间密集型多模态轨迹预测模型，通过两阶段方法生成随时间变化、可调整的多模态轨迹并建模交互； AgentFormer: 时间编码器代替位置编码器，通过时间戳保留时间信息
交互问题(自注意力计算)： 智能体间交互如何建模？ 有什么样的依赖关系？ 是否有交互信息丢失问题？如何解决？ 自注意力计算效率如何提升？	mmTransformer: 对交互信息编码，对车辆间依赖关系建模； AgentFormer: Agent-Aware 机制分别计算智能体间和智能体自身内在注意力； TrajFormer: 通过自注意力解决提取局部特征受限问题，更好地提取环境交互信息； MPTP: 通过补丁合并和移位窗口方法降低自注意力计算复杂性； Gatformer: 对智能体-智能体、智能体-环境交互建模 Spatio-Channel TF: 通过基于挤压激励网络的Channel模块捕获相邻Channel之间的相互作用，进而捕获多智能体间的交互信息； GatFormer: 采用灵活的图结构，降低计算复杂度，通过交互权重分配提高性能和模型的可解释性，同时保证鲁棒性
多智能体交互问题： 多智能体交互信息如何捕获？	
可解释性问题： 可解释性差的原因？ 如何提高可解释性？	MPTP: Swim Transformer编码环境信息，GAR编码历史轨迹，生成可解释性高的预测轨迹
规模化问题： 规模化难的原因？ 编码模块如何优化？	Wayformer: 分析前融合、后融合、分层融合的利弊，为TF如何扩展到大型多维序列提供解决方案

迹预测。但是基于概率的方法过度依赖先验分布和损失函数，容易出现优化不稳定或模式崩溃的现象；基于建议的启发式方法过度依赖于锚点质量，不能保证生成多模态情况。基于Transformer的方法可以避免在设计先验分布和损失函数的过程中的大量人工工作，同时可以更好地捕捉到轨迹预测的多模态性质，实现多模态轨迹预测。

Liu等人^[25]针对如何实现多模态轨迹预测，提出mmTransformer框架，该方法在Argoverse基准排行榜中排名第一，框架由3个独立的堆叠式的Transformer模型组成，分别聚合历史轨迹、道路信息以及交互信息。mmTransformer整体框架可由两部分组成，第一部分仅由运动提取器和地图聚合器分别对车辆的信息及环境信息进行编码，不考虑交互信息；第二部分通过社会构造函数对临近信息进行聚合，并对车辆之间的依赖关系进行建模，整个过程依照逻辑顺序，即社会关系是基于每个车辆特征构建的。该方法还提出基于区域的训练策略，在初始化建议后，将建议路径分为空间群组，通过路径分配计算路径回归损失和分类损失，以确保生成多模态预测轨迹。为解决多模态预测轨迹随时间可调整的问题，Liu等人^[115]提出LAformer时间密集型轨迹预测模型。该方法分为两阶段，第一阶段将道路随时间步长进行分段，并结合历史轨迹，通过自注意机制选择多条得分高的预测轨迹，使用交叉注意力机制，将目标智能体过去的轨迹作为查询

向量，将车道编码作为键和价值向量，然后使用多模态条件解码器优化预测轨迹精度；第二阶段对预测轨迹进行二次优化，并学习转向角。Wang等人^[60]针对路口场景不确定性的问题，提出了适用于该场景的两阶段联合轨迹预测-规划方法，利用专家数据混合法从人类驾驶员的轨迹数据中学习，构建了一个多模态运动规划器，该规划器通过明确考虑风格，使用Transformer框架对车辆间的相互作用建模，确保集成网络实现场景一致的多模态轨迹预测和候选轨迹生成。

Yuan等人^[116]针对时间和社会维度上独立特征编码信息丢失的问题，提出AgentFormer允许一个智能体在某个时间的状态直接影响另一个智能体未来的状态，而不是通过在一个维度上编码的中间特征；AgentFormer可以同时学习时序信息和交互关系，智能体当前时刻的关系可以通过不同时刻的关系体现，解决了传统Transformer注意力中各个输入元素权重平等造成的时间和智能体信息损失。该模型采用时间编码减少时间信息损失，通过独特的Agent-Aware注意力机制编码智能体和时间的关系，采用CVAE形式，以概率形式描述，确保了生成轨迹的多模态性。Pini等人^[117]使用Transformer框架，通过编码器编码驾驶场景中所有车辆的动态信息、静态信息，输出预测轨迹，作为下一阶段基于模仿学习的轨迹规划模块的专家数据。Chen等人^[118]为解决LSTM难以处理复杂的时间依赖性的

问题,提出S2TNet模型,通过空间自注意机制获取道路网络中所有交通参与者的交互关系,通过TCN提取连续帧的时间依赖关系。Bhat等人^[119]针对提取场景上下文面临局限性的问题,提出基于自注意编码器的TrajFormer模型,在给定场景下更好地表征智能体行为和社交礼仪,显著改进了在minADE、minFDE方面的性能。

Zhao等人^[120]针对传统注意力机制无法捕获多智能体之间交互的问题,提出空间通道Transformer(spatial-channel Transformer),该模型在Transformer框架的基础上,插入了一个通道注意力(channel-wise attention)模块,即挤压激励(SE)网络^[121],并将SE网络用于捕获相邻通道之间的相互作用。Huang等人^[122]针对如何编码多智能体交互问题,使用TF编码器建模智能体与周围车辆的交互关系,多头注意机制可以帮助提取智能体交互的不同信息。通过矢量地图表示和基于车道集的地图结构提取地图与目标智能体之间的关系。Zhang等人^[123]针对多智能体轨迹预测问题,提出了Gatformer,其采用灵活的图结构,相比基于图神经网络的方法,降低了全连通图造成的计算复杂性。基于稀疏图,Gatformer可以预测多智能体未来的轨迹,同时考虑智能体之间的相互作用。Gatformer注意机制通过对车-车、车-环境的交互进行权重分配,提高了预测精度并减少了推理时间,已在多种不同场景下验证其鲁棒性,然而该方法缺乏可解释性。针对可解释性问题,Zhang等人提出基于Swin Transformer^[124]的多模态轨迹预测模型MPTP^[26],通过下采样(patch merging)和滑动窗口方法,对输入的交通场景信息建模,使用GRU方法处理历史轨迹,根据融合后的输入特征输出可解释性强的多模态轨迹预测。

复杂驾驶环境通常包含静态基础设施及高动态交通参与者,针对多源信息融合(如车道几何形状、车道连通性、时变交通信号灯状态、其他交通参与者历史轨迹及交互信息等)问题,现有方法通常设计多样性特征提取模块处理多源数据,但该方法难以规模化,同时建模质量和效率无法同时保证。针对此问题,Nayakanti等人^[125]提出WayFormer,在Transformer框架的基础上,研究了3种输入模式(前融合、后融合和分层融合)的利弊,对于每种融合类型,探索通过分解注意或潜在查询注意来权衡效率和质量的策略。后融合中每种特征

都有与之相对应的编码器;前融合不是将注意编码器专用于每个模态,而是减少特定模态的参数到投影层;分层融合是前融合、后融合折中的模型,将场景信息分别通过注意编码器编码后聚合,将聚合特征输入最终的注意机制交叉模型中,有效地将场景编码器的深度在模态特定编码器和跨模态编码器之间平均。该文还为如何将Transformer扩展到大型多维序列中提供了解决方案,减少了每个块的注意分量和位置前馈网络的计算成本。

4 基于图神经网络的轨迹预测方法

车辆为实现安全自动驾驶,必须准确预测其他交通参与者的未来运动轨迹。为解决涉及车辆交互的轨迹预测问题,高清地图为运动预测提供了必需的几何信息和语义信息。早期实验中将HD地图作为启发式方法的输入数据,将交通参与者与车道相关联,然后基于地图拓扑生成候选运动路径。虽然通过这种方式有可能实现安全的自动驾驶,但是通过这种方式生成的预测结果会受到地图约束,且无法捕捉到极端的驾驶行为。基于CNN的方法在光栅化过程会导致信息丢失,且很难捕获复杂的地图拓扑结构,通常需要扩大覆盖范围,这将导致计算量过大。此外,同向车道与对向车道具有完全不同的语义信息和依赖关系,且此依赖关系与车道间距离无关,这导致前期处理需要耗费大量人力。为解决上述问题,图神经网络^[27,126]通过向量化方法,将环境中的每个对象视为一个节点,将节点连接形成图,节点间由边连接,用于表示关系,通过图形的数据处理的方法,学习可以从节点特征及相邻特征中提取交互感知特征的函数,由此建模对象间的依赖关系。因此图神经网络非常适合处理基于交互的轨迹预测问题。

在现有基于图卷积网络(graph convolutional network, GCN)的方法中,Liang等人^[127]提出了LaneGCN模型,首先从矢量地图数据中构建车道图,然后使用车道图卷积网络捕获车道图的复杂拓扑及长距离依赖关系,通过融合网络表示车对车、车道对车道、车对车道、车道对车4种交互模式,生成多模态轨迹。在基于图神经网络的联合学习方法^[128]中,自动驾驶车辆需要在具有复杂交互关系的潜在场景中服从动力学要求,并学习未来轨迹的联合分布。现有方法假设交通参与者未来轨迹具有独立性^[129-130],但存在无法获得与场景一致的未来

轨迹等问题。此外,利用自回归方法的模型需要顺序采样,导致训练速度慢,效率低。针对以上问题 Casas 等人^[131]提出 ILVM 模型,通过潜在变量模型恢复轨迹联合分布潜在空间,将场景建模为交互图,通过图神经网络将场景编码到潜在空间中,并将潜在样本解码为未来轨迹。Li 等人^[132]在基于空间的图形卷积网络研究中,提出由卷积层以及图形操作组成的 GRIP 模型。该模型使用图形表示提取特征,使用 LSTM 进行预测。为提高模型精度, Li 等人在后续研究中提出 GRIP++ 模型^[133],该方法使用固定图和动态图预测不同类型的智能体轨迹,使车辆可以避免交通事故。Chandra 等人^[134]使用动态几何图建模车辆间的交互关系,将 DCN 与双流 LSTM 结合,通过频谱正则化方法预测未来轨迹。

与基于 CNN 的方法相比,基于 GNN 的方法将环境信息渲染成二维图,并将其作为模型的输入,对每一时刻的场景进行渲染,此方法需要消耗大量的计算资源,并且 CNN 提取特征的能力有限,很难保存长距离的交互关系。针对此问题, Gao 等人^[135]提出分级神经网络 VectorNet,将车道信息、车辆历史轨迹抽象为几何形状,如点或曲线,然后将其转换为向量的矢量格式作为输入。该模型首先抽取行人轨迹、车道等特征,然后通过基于注意力机制的全连通图建模车辆间的交互关系。在 VectorNet 的基础下, Gao 等人在后续研究中提出 TNT^[114],将每条轨迹的模态定义在每条轨迹的终点上,将轨迹预测问题变成终点预测问题,基于终点补充完整的轨迹。然而基于锚点的方法对于锚点质量有很强的依赖性,锚点的选择会对预测结果产生很大的影响。为避免启发式地设置锚点, Gu 等人^[136]提出 DenseTNT 方法,使用更好的策略选择锚点,在尽可能准确的前提下覆盖更多的模态。为解决近似误差问题及不确定性问题, Chai 等人^[137]提出 MultiPath (MAP) 方法,通过预定义锚的方法,在 BEV 视角下对不同目标做特征提取及状态分析,在多个预定义锚点上预测该目标的输出轨迹及相应的不确定性。

基于图神经网络的方法除了通过学习映射函数的图卷积网络^[30-31],还有:①图注意力网络 (graph attention networks, GAT)^[32-34],此类模型将注意力机制引入基于空间域的图神经网络来放大数据重要特征。Diehl 等人^[29]对 GNN、GAN 进行评估,使用 GAT 方法进行问题归纳^[110],节点-邻近点

之间可以实现并行计算,计算效率高,通过加权邻域的方法可以对不同程度的节点分类。Huang 等人^[138]将遗传算法应用于轨迹预测,通过 LSTM 对输入轨迹进行编码,通过 GAT 计算每个交通参与者的权重,通过 LSTM 解码器输出多模态轨迹。②图自编码器模型 (graph autoencoders)^[35-36],此类模型通过编码器-解码器获取图中节点,并将其作为输入;③图生成网络^[37-38],其由生成器、鉴别器组成,用于生成多模态预测轨迹;④时间-空间依赖图网络 (graph spatial-temporal networks)^[39-40],此类模型用于捕捉时空图信息。

本节主要介绍了基于图神经网络的方法。虽然图神经网络在处理自动驾驶交互问题中取得了较好的效果,但是此类模型将整个预测过程置于“黑盒”中,导致模型可解释性差。

5 基于生成模型的轨迹预测方法

在轨迹预测任务中,驾驶环境和驾驶策略存在不确定性,导致最终预测轨迹是多模态的,生成模型 (generator model) 主要用于生成多模态预测轨迹。现有的生成模型方法可分为两类:生成对抗网络 (GAN)^[41-44]、条件变分自编码器 (CVAE)。

在基于 GAN 的多模态轨迹预测研究中, Gupta 等^[139]提出的 SGAN (social GAN) 模型,首次提出多样化损失 (variety loss),使模型可以生成多模态轨迹分布。其中,生成器使用 LSTM 编码器、池化模块和 LSTM 解码器生成未来轨迹,池化模块融合全局信息并建模交互,通过损失函数在多条生成轨迹中选出最优轨迹,通过鉴别器优化轨迹。该方法是最早基于 GAN 多模态轨迹的研究。基于 SGAN, Li 等人^[140]将生成轨迹的概率作为条件,提出 CGNS 模型,利用深度特征提取器提取静态上下文和交互信息,并与环境注意机制 (EAM) 结合,通过 GRU 输出预测轨迹。Yang 等人^[141]为有效提取交互信息并生成多模态可行轨迹,在 SGAN 的基础上添加潜在变量预测器,用于估计潜在变量。Sadeghian 等人^[142]提出基于 GAN 的可解释框架 Sophie,利用历史轨迹和场景上下文信息解决了多智能体交互情况下的路径规划问题。Liu 等人^[143]提出基于注意力机制的 Col-GAN 方法,用于建模低碰撞的多模态社会轨迹,其中使用卷积神经网络作为鉴别器,以生成碰撞较少的轨迹。为解决不同驾驶员驾驶风格对轨迹预测的影响, Choi 等人^[144]提出

DSA-GAN 模型, 该模型分为两个阶段。第一阶段识别驾驶风格, 通过递归图将序列数据转换为图像, 并通过 CNN 将转换后的图像处理为驾驶风格; 第二阶段轨迹预测阶段, 通过条件生成网络根据分布生成多模态真实驾驶轨迹, 生成的轨迹与驾驶风格有较高的依赖性, 因此可以根据不同驾驶风格生成多模态轨迹。

自动编码器 (autoencoder) 方法作为无监督学习方法, 可学习如何表示数据分布。在自动驾驶轨迹预测任务中, 自编码器由编码器解码器结构组成, 其中编码器将输入映射为数据表示, 解码器将数据表示输出到模型。条件自编码器 CVAE^[46-49]将编码器解码器调节为条件变量, 基于条件变量每个输出都有概率分布, CVAE 允许对数据生成过程进行更大的控制。在轨迹预测任务中, CVAE 方法可以用于预测未来车辆的位置。通过编码器将数据压缩为低维向量表示, 并使用解码器解码低维向量以获得重构输出。Kingma 等人^[145]使用 CVAE 框架对变分推理中的分布进行参数化, 以提高生成能力。Basldi 等人^[146]针对社会后验失效问题提供了解决方案, Gómez-Huélamo 等人^[147]基于此方案将变分自动编码器 (variational autoencoder, VAE) 与 GAN 模型融合, 使用对抗性框架训练轨迹生成器, 生成具有物理约束的符合现实情况的轨迹。

本节介绍了基于 GAN 的轨迹预测方法。在交互任务中, GAN 模型有较好的性能, 然而由于 GAN 的参数优化问题并不是一个凸优化问题, 虽然 Kosaraju 等人^[45]针对模式崩溃问题将 GAN 与 GAT 结合, 提出了 Social-BiGAT 解决方案, 但是当 GAN 进入某一个纳什均衡状态时, 损失函数表现为收敛, 其仍可产生模式崩溃。

6 问题与潜在研究方向

现有基于交互式预测的方法存在以下难题: ①交互式方法通常包括多个递归模型, 因此预测速度不够快, 现有方法需要使用一个并行模型来降低计算成本; ②Transformer 模型平等对待每个节点, 因此在输入多个 SV 时很难训练; ③当输入多个 SV 时, Transformer 模型无法将每个 SV 与其他 SV 区分开来。目前基于深度学习的方法虽然可以实现多模态轨迹预测, 但是无法考虑到驾驶员在规划范围内与某些自我行为的交互作用, 虽然 Social GAN^[101]模型可以通过周围车辆预测自动驾驶车辆

的行动, 但是没有考虑到司机的驾驶风格以及意图的变化, 因此需要足够的交通数据进行训练。此外, 对驾驶风格与意图推理的研究不足, 现有预测模型在高动态复杂场景中泛化性欠佳, 难以实车部署, 对于模型可解释性研究缺乏定量分析。

本文总结了过去 10 年基于深度学习的轨迹预测方法, 并对未来的潜在研究方向进行了预测, 具体如下。

- 数据集需要包含更多的信息: 现有数据集中多车交互场景不足^[148], 基于深度学习的轨迹预测方法需要大量的训练数据。真实驾驶环境中车辆的轨迹不仅受环境影响, 还受到周围驾驶员驾驶风格影响, 因此需要增加数据集社会交互信息。除此之外, 现有方法大多不受交通规则约束, 与实际驾驶情况不符, 数据集中红绿灯、交通标志等都应作为预测的输入。

- 其他领域先进的算法也可以在自动驾驶中使用: Transformer 算法最早在 NLP 领域取得非常显著的效果, 研究人员将此算法引入轨迹预测任务中同样取得了非常卓越的表现, 现有算法局限于使用高清地图、交互信息、历史轨迹等进行轨迹预测, 虽然取得了较好的效果, 但是为实现符合实际的预测估计, 应参考其他领域的优秀算法, 提高预测精度。

- 模型解释性需要增强: 现有算法忽略了轨迹预测的可解释性, 虽然保证了安全性要求, 但是生成的轨迹在可解释性方面并没有突破性的成就, 虽然 Zhang 等人^[26]在可解释性方面取得了较好的进展, 但是现有对于可解释性方面的研究缺少定量分析与进一步探索。

- 提高自动驾驶闭环框架的有效性: 轨迹预测的研究是为后续轨迹规划及控制提供基础, 安全的轨迹预测可以使轨迹规划及控制取得更好的表现, 目前运动系统大多将其他交通参与者视为匀速运动, 与真实驾驶情况不符。当集成到实车时, 周围环境变化会导致车辆安全性受到严重的影响, 因此提高整个系统的有效性也是轨迹预测未来的方向。

- 提高模型鲁棒性: 现有方法使用的数据集是半自动注释的, 地面真实轨迹是具有噪声的, 在真实驾驶环境中, 感知模块无法提供精准预测导致的噪声会影响整个模型的预测精度, 因此需要研究模型的鲁棒性及抗干扰能力。

• 智慧交通部署问题：复杂交通系统存在交通网络难构建、数据难获取、系统难分析等问题。智慧交通的部署一方面依赖于交通基础设施升级，另一方面需要自动驾驶技术升级。平行驾驶集成物理实体车辆和描述，预测及引导3类虚拟车辆^[11,149-151]，可以通过虚实结合的方法，在人工交通环境中积累知识经验，通过虚实互动引导真实世界的车辆驾驶过程。

7 总结与展望

综上所述，现阶段基于深度学习的多模态轨迹预测方法的整体框架已经成型，都是由编码器+交互+解码器组成，通过CNN、RNN、GNN等神经网络提取空间交互信息及地图信息，训练后的模型在复杂场景及长时间范围内具有较好的性能，并可以生成与车辆机动性以及场景一致的多模态轨迹。但是基于深度学习的方法需要大量的实验数据。数据方面，目前针对轨迹预测任务的数据集内容越来越丰富，场景复杂度越来越高，数据集的发展提高了轨迹预测算法的性能。然而目前数据集中周围车辆驾驶风格单一，没有包含长尾问题，使目前的轨迹预测算法难以在由不同驾驶风格车辆组成的驾驶场景中应用。展望未来数据集发展，在丰富场景复杂度的同时，应添加不同驾驶风格车辆共同存在的场景，进而增强轨迹预测模型在高动态不确定性场景下的泛化性。现有评估度量大多以预测轨迹与真实轨迹偏差为主，模型计算时间与预测范围也应作为衡量模型能否应用到真实驾驶环境的重要度量。然而，现有方法只是基于先验知识选择一定时间段作为预测范围，并没有方法特别研究时间长度与预测精度之间如何平衡，因此短时-长时轨迹预测问题也可以作为未来研究的问题之一。

针对多模态轨迹预测目前具有的挑战性问题，基于Transformer轨迹预测在Argoverse数据集的平均位移误差和最终位移误差性能指标上取得了最优水平。基于Transformer的模型处理交互问题，特别是与其他交通参与者、障碍物交互效果相比，CNN与RNN方法有明显的提升，Transformer可以解决长历史轨迹信息丢失问题，同时依靠注意力机制捕获车辆之间交互信息。除此之外，基于自注意力机制的Transformer方法有较好的可解释性，研究人员可以根据模型输出结果更好地判断可改进的参数。

然而，Transformer模型虽然在自然语言处理

及视觉领域均取得了非常显著的成果，但是在自动驾驶轨迹预测方向的研究还较少。目前还无法确定Transformer算法能否应用到更加复杂多变的环境中。在现实环境中，感知模块在真实驾驶环境存在局限性，如果有其他交通参与者被遮挡，或者出现缺失/过时/不准确的道路基础设施信息等情况，无法获得实验阶段输入的理想数据，会导致预测轨迹出现偏差。同时如何进一步提升可解释性也是基于Transformer模型面临的主要问题之一，现有方法中对预测轨迹的置信度难以解释，导致模型可解释性低。这些问题也是未来使用Transformer做多模态轨迹预测可继续深入的方向。此外，现有方法对于多模态的研究还不充分，相信在未来的发展中，基于Transformer的多模态轨迹预测方法会更加完善。

参考文献：

- [1] LEVINSON J, ASKELAND J, BECKER J, et al. Towards fully autonomous driving: systems and algorithms[C]//Proceedings of 2011 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2011: 163-168.
- [2] MAURER M, GERDES J C, LENZ B, et al. Autonomous driving: technical, legal and social aspects[M]. Springer Berlin: Nature, 2016.
- [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] 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报, 2017, 43(8): 1289-1305.
ZHANG H, WANG K F, WANG F Y. Advances and perspectives on applications of deep learning in visual object detection[J]. Acta Automatica Sinica, 2017, 43(8): 1289-1305.
- [5] 田永林, 王雨桐, 王建功, 等. 视觉Transformer研究的关键问题: 现状及展望[J]. 自动化学报, 2022, 48(4): 957-979.
TIAN Y L, WANG Y T, WANG J G, et al. Key problems and progress of vision transformers: the state of the art and prospects[J]. Acta Automatica Sinica, 2022, 48(4): 957-979.
- [6] 卢经纬, 郭超, 戴星原, 等. 问答ChatGPT之后: 超大预训练模型的机遇和挑战[J]. 自动化学报, 2023, 49(4): 705-717.
LU J W, GUO C, DAI X Y, et al. The ChatGPT after: opportunities and challenges of very large scale pre-trained models[J]. Acta Automatica Sinica, 2023, 49(4): 705-717.
- [7] 王飞跃, 缪青海, 张军平, 等. 探讨AI for Science的影响与意义: 现状与展望[J]. 智能科学与技术学报, 2023, 5(1): 1-6.
WANG F Y, MIAO Q H, ZHANG J P, et al. The DAOs to AI for Science by DeSci: the state of the art and perspective[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(1): 1-6.
- [8] 李娟娟, 秦蕊, 丁文文, 等. 基于Web3的去中心化自治组织与运营新框架[J]. 自动化学报, 2023, 49(5): 985-998.
LI J J, QIN R, DING W W, et al. A new framework for Web3-powered decentralized autonomous organizations and operations[J]. Acta Automatica Sinica, 2023, 49(5): 985-998.

- [9] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. *Neural Computation*, 1989, 1(4): 541-551.
- [10] YE P J, WANG X, ZHENG W B, et al. Parallel cognition: hybrid intelligence for human-machine interaction and management[J]. *Frontiers of Information Technology & Electronic Engineering*, 2022, 23(12): 1765-1779.
- [11] WANG X, YANG J, HAN J P, et al. Metaverses and DeMetaverses: from digital twins in CPS to parallel intelligence in CPSS[J]. *IEEE Intelligent Systems*, 2022, 37(4): 97-102.
- [12] XIE G, SHANGGUAN A Q, FEI R, et al. Motion trajectory prediction based on a CNN-LSTM sequential model[J]. *Science China Information Sciences*, 2020, 63(11): 1-21.
- [13] GIRSHICK R. Fast R-CNN[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*. Piscataway: IEEE Press, 2015: 1440-1448.
- [14] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*. Piscataway: IEEE Press, 2017: 2980-2988.
- [15] GRAVES A. Generating sequences with recurrent neural networks[J]. *arXiv preprint*, 2013, arXiv:1308.0850.
- [16] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. *arXiv preprint*, 2014, arXiv:1409.0473.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [18] FU R, ZHANG Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction[C]//*Proceedings of 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. Piscataway: IEEE Press, 2016: 324-328.
- [19] DEY R, SALEM F M. Gate-variants of Gated Recurrent Unit (GRU) neural networks[C]//*Proceedings of 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*. Piscataway: IEEE Press, 2017: 1597-1600.
- [20] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. *arXiv preprint*, 2014, arXiv:1412.3555.
- [21] MESSAOUD K, YAHIAOUI I, VERROUST-BLONDET A, et al. Non-local social pooling for vehicle trajectory prediction[C]//*Proceedings of 2019 IEEE Intelligent Vehicles Symposium (IV)*. Piscataway: IEEE Press, 2019: 975-980.
- [22] MESSAOUD K, YAHIAOUI I, VERROUST-BLONDET A, et al. Attention based vehicle trajectory prediction[J]. *IEEE Transactions on Intelligent Vehicles*, 2021, 6(1): 175-185.
- [23] MESSAOUD K, DEO N, TRIVEDI M M, et al. Trajectory prediction for autonomous driving based on multi-head attention with joint agent-map representation[C]//*Proceedings of 2021 IEEE Intelligent Vehicles Symposium (IV)*. Piscataway: IEEE Press, 2021: 165-170.
- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. New York: ACM, 2017: 6000-6010.
- [25] LIU Y C, ZHANG J H, FANG L J, et al. Multimodal motion prediction with stacked transformers[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2021: 7573-7582.
- [26] ZHANG K, LI L. Explainable multimodal trajectory prediction using attention models[J]. *Transportation Research Part C: Emerging Technologies*, 2022, 143: 103829.
- [27] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4-24.
- [28] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. *IEEE Transactions on Neural Networks*, 2009, 20(1): 61-80.
- [29] DIEHL F, BRUNNER T, LE M T, et al. Graph neural networks for modelling traffic participant interaction[C]//*Proceedings of 2019 IEEE Intelligent Vehicles Symposium*. Piscataway: IEEE Press, 2019: 695-701.
- [30] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. *arXiv preprint*, 2016, arXiv:1609.02907.
- [31] ZHANG S, TONG H H, XU J J, et al. Graph convolutional networks: a comprehensive review[J]. *Computational Social Networks*, 2019, 6(1): 1-23.
- [32] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. *arXiv preprint*, 2017, arXiv:1710.10903.
- [33] JI H Y, WANG X, SHI C, et al. Heterogeneous graph propagation network[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(1): 521-532.
- [34] WANG X, HE X N, CAO Y X, et al. KGAT: knowledge graph attention network for recommendation[C]//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 2019: 950-958.
- [35] KIPF T N, WELING M. Variational graph auto-encoders[J]. *arXiv preprint*, 2016, arXiv:1611.07308.
- [36] PAN S, HU R, LONG G, et al. Adversarially regularized graph auto-encoder for graph embedding[J]. *arXiv preprint*, 2018, arXiv:1802.04407.
- [37] ZHOU D W, ZHENG L C, HAN J W, et al. A data-driven graph generative model for temporal interaction networks[C]//*Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 2020: 401-411.
- [38] BONGINI P, BIANCHINI M, SCARSELLI F. Molecular generative graph neural networks for drug discovery[J]. *Neurocomputing*, 2021, 450: 242-252.
- [39] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[J]. *arXiv preprint*, 2018, arXiv:1801.07455.
- [40] GUO S N, LIN Y F, FENG N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 922-929.
- [41] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [42] CRESWELL A, WHITE T, DUMOULIN V, et al. Generative adversarial networks: an overview[J]. *IEEE Signal Processing Magazine*, 2018, 35(1): 53-65.
- [43] HO J, ERMON S. Generative adversarial imitation learning[C]//*Proceedings of the 30th International Conference on Neural Information Processing Systems*. New York: ACM, 2016: 4572-4580.
- [44] SONG J M, REN H Y, SADIGH D, et al. Multi-agent generative adversarial imitation learning[C]//*Proceedings of the 32nd International Conference on Neural Information Processing Systems*. New York: ACM, 2018: 7472-7483.
- [45] KOSARAJU V, SADEGHIAN A, MARTÍN-MARTÍN R, et al. Social-BiGAT: multimodal trajectory forecasting using bicycle-GAN and

- graph attention networks[J]. arXiv preprint, 2019, arXiv: 1907.03395.
- [46] ZHAO T, ZHAO R, ESKENAZI M. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders[J]. arXiv preprint, 2017, arXiv:1703.10960.
 - [47] BAO J M, CHEN D, WEN F, et al. CVAE-GAN: fine-grained image generation through asymmetric training[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2764-2773.
 - [48] IVANOVIC B, LEUNG K, SCHMERLING E, et al. Multimodal deep generative models for trajectory prediction: a conditional variational autoencoder approach[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 295-302.
 - [49] HONG J, SAPP B, PHILBIN J. Rules of the road: predicting driving behavior with a convolutional model of semantic interactions[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 8446-8454.
 - [50] LEFÈVRE S, VASQUEZ D, LAUGIER C. A survey on motion prediction and risk assessment for intelligent vehicles[J]. ROBOMECH Journal, 2014, 1(1): 1-14.
 - [51] GONZÁLEZ D, PÉREZ J, MILANÉS V, et al. A review of motion planning techniques for automated vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4): 1135-1145.
 - [52] LIU J B, MAO X Y, FANG Y Q, 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 (ROBIO). Piscataway: IEEE Press, 2021: 978-985.
 - [53] MOZAFFARI S, AL-JARRAH O Y, DIANATI M, et al. Deep learning-based vehicle behavior prediction for autonomous driving applications: a review[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(1): 33-47.
 - [54] LEON F, GAVRILESCU M. A review of tracking and trajectory prediction methods for autonomous driving[J]. Mathematics, 2021, 9(6): 660.
 - [55] GOMES I, WOLF D. A review on intention-aware and interaction-aware trajectory prediction for autonomous vehicles[J]. 2022.
 - [56] SHIRAZI M S, MORRIS B T. Looking at intersections: a survey of intersection monitoring, behavior and safety analysis of recent studies[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(1): 4-24.
 - [57] 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.
 - [58] KAEHLING L P, LITTMAN M L, MOORE A W. Reinforcement learning: a survey[J]. Journal of Artificial Intelligence Research, 1996, 4: 237-285.
 - [59] NAYAK A, ESKANDARIAN A, DOERZAPH Z. Uncertainty estimation of pedestrian future trajectory using Bayesian approximation[J]. IEEE Open Journal of Intelligent Transportation Systems, 2022, 3: 617-630.
 - [60] WANG X, TANG K, DAI X Y, et al. Safety-balanced driving-style aware trajectory planning in intersection scenarios with uncertain environment [J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(4): 2888-2898.
 - [61] DEO N, TRIVEDI M M. Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs[C]//Proceedings of 2018 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2018: 1179-1184.
 - [62] SHOUNO O. Deep unsupervised learning of a topological map of vehicle maneuvers for characterizing driving styles[C]//Proceedings of 2018 21st International Conference on Intelligent Transportation Systems (ITSC). Piscataway: IEEE Press, 2018: 2917-2922.
 - [63] LIPTON Z C. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery[J]. Queue, 2018, 16(3): 31-57.
 - [64] CHAKRABORTY S, TOMSETT R, RAGHAVENDRA R, et al. Interpretability of deep learning models: A survey of results[C]//2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation. Piscataway: IEEE Press, 2017: 1-6.
 - [65] GUNNING D, STEFIK M, CHOI J, et al. XAI-explainable artificial intelligence[J]. Science Robotics, 2019, 4(37): eaay7120.
 - [66] DU M N, LIU N H, HU X. Techniques for interpretable machine learning[J]. Communications of the ACM, 2019, 63(1): 68-77.
 - [67] HUME D. An inquiry concerning human understanding[M]. New York: Oxford University Press, 2003.
 - [68] YUAN Y, KITANI K. Diverse trajectory forecasting with determinantal point processes[J]. arXiv preprint, 2019, arXiv:1907.04967.
 - [69] KIM J, CANNY J. Interpretable learning for self-driving cars by visualizing causal attention[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2961-2969.
 - [70] CAESAR H, BANKITI V, LANG A H, et al. nuScenes: a multimodal dataset for autonomous driving[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 11618-11628.
 - [71] HOUSTON J, ZUIDHOF G, BERGAMINI L, et al. One thousand and one hours: Self-driving motion prediction dataset[C]//Proceedings of Conference on Robot Learning. New York: PMLR, 2021: 409-418.
 - [72] SUN P, KRETZSCHMAR H, DOTIWALLA X, et al. Scalability in perception for autonomous driving: waymo open dataset[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 2443-2451.
 - [73] CHANG M F, LAMBERT J, SANGKLOY P, et al. Argoverse: 3D tracking and forecasting with rich maps[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 8740-8749.
 - [74] WILSON B, QI W, AGARWAL T, et al. Argoverse 2: next generation datasets for self-driving perception and forecasting[J]. arXiv preprint, 2023, arXiv:2301.00493.
 - [75] ZHAN W, SUN L, WANG D, et al. Interaction dataset: an international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps[J]. arXiv preprint, 2019, arXiv:1910.03088.
 - [76] KRAJEWSKI R, BOCK J, KLOEKER L, et al. The highD dataset: a drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems[C]//Proceedings of 2018 21st International Conference on Intelligent Transportation Systems (ITSC). Piscataway: IEEE Press, 2018: 2118-2125.
 - [77] HUANG X Y, CHENG X J, GENG Q C, et al. The ApolloScape dataset for autonomous driving[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2018: 954-960.
 - [78] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11): 1231-1237.
 - [79] FANG L J, JIANG Q H, SHI J P, et al. TPNet: trajectory proposal net-

- work for motion prediction[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 6796-6805.
- [80] NIKHIL N, MORRIS B T. Convolutional Neural Network for Trajectory Prediction[C]//European Conference on Computer Vision. Cham: Springer, 2019: 186-196.
- [81] CHOU F C, LIN T H, CUI H G, et al. Predicting motion of vulnerable road users using high-definition maps and efficient ConvNets[C]//Proceedings of 2020 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2020: 1655-1662.
- [82] CUI H G, RADOSAVLJEVIC V, CHOU F C, et al. Multimodal trajectory predictions for autonomous driving using deep convolutional networks[C]//Proceedings of 2019 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2019: 2090-2096.
- [83] PHAN-MINH, GRIGOREE C, BOULTONF A, et al. CoverNet: multimodal behavior prediction using trajectory sets[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 14062-14071.
- [84] CUI H G, NGUYEN T, CHOU F C, et al. Deep kinematic models for kinematically feasible vehicle trajectory predictions[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2020: 10563-10569.
- [85] DEO N, TRIVEDI M M. Convolutional social pooling for vehicle trajectory prediction[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2018: 1468-1476.
- [86] MARZINOTTO A, COLLEDANCHISE M, SMITH C, et al. Towards a unified behavior trees framework for robot control[C]//Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2014: 5420-5427.
- [87] ZHANG Y, ZOU Y, TANG J, et al. A lane-changing prediction method based on temporal convolution network[J]. arXiv preprint, 2020, arXiv:2011.01224.
- [88] GILLES T, SABATINI S, TSISHKOU D, et al. HOME: heatmap output for future motion estimation[C]//Proceedings of 2021 IEEE International Intelligent Transportation Systems Conference (ITSC). Piscataway: IEEE Press, 2021: 500-507.
- [89] GILLES T, SABATINI S, TSISHKOU D, et al. GOHOME: graph-oriented heatmap output for future motion estimation[C]//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2022: 9107-9114.
- [90] GILLES T, SABATINI S, TSISHKOU D, et al. Thomas: trajectory heatmap output with learned multi-agent sampling[J]. arXiv preprint, 2021, arXiv:2110.06607.
- [91] YE M S, CAO T Y, CHEN Q F. TPCN: temporal point cloud networks for motion forecasting[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 11313-11322.
- [92] GIULIARI F, HASAN I, CRISTANI M, et al. Transformer networks for trajectory forecasting[C]//Proceedings of 2020 25th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE Press, 2020: 10335-10342.
- [93] ALTCHÉ F, DE LA FORTELLE A. An LSTM network for highway trajectory prediction[C]//Proceedings of 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). Piscataway: IEEE Press, 2017: 353-359.
- [94] ZYNER A, WORRALL S, NEBOT E. Naturalistic driver intention and path prediction using recurrent neural networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(4): 1584-1594.
- [95] PARK S H, KIM B, KANG C M, et al. Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture[C]//Proceedings of 2018 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2018: 1672-1678.
- [96] SONG H, DING W, CHEN Y, et al. Pip: planning-informed trajectory prediction for autonomous driving[C]//Proceedings of Computer Vision-ECCV 2020: 16th European Conference. [S.l.:s.n.], 2020: 598-614.
- [97] XIN L, WANG P, CHAN C Y, et al. Intention-aware long horizon trajectory prediction of surrounding vehicles using dual LSTM networks [C]//Proceedings of 2018 21st International Conference on Intelligent Transportation Systems (ITSC). Piscataway: IEEE Press, 2018: 1441-1446.
- [98] SHIRAZI M S, MORRIS B T. Trajectory prediction of vehicles turning at intersections using deep neural networks[J]. Machine Vision and Applications, 2019, 30(6): 1097-1109.
- [99] XING Y, LV C, CAO D P. Personalized vehicle trajectory prediction based on joint time-series modeling for connected vehicles[J]. IEEE Transactions on Vehicular Technology, 2020, 69(2): 1341-1352.
- [100] ZHANG T, SONG W J, FU M Y, et al. Vehicle motion prediction at intersections based on the turning intention and prior trajectories model [J]. IEEE/CAA Journal of Automatica Sinica, 2021, 8(10): 1657-1666.
- [101] BAE S, SAXENA D, NAKHAEI A, et al. Cooperation-aware lane change maneuver in dense traffic based on model predictive control with recurrent neural network[C]//Proceedings of 2020 American Control Conference (ACC). Piscataway: IEEE Press, 2020: 1209-1216.
- [102] MO X Y, 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. Piscataway: IEEE Press, 2020: 5057-5062.
- [103] HUANG Z, WANG J, PI L, et al. LSTM based trajectory prediction model for cyclist utilizing multiple interactions with environment[J]. Pattern Recognition, 2021, 112: 107800.
- [104] WANG S B, ZHAO P, YU B, et al. Vehicle trajectory prediction by knowledge-driven LSTM network in urban environments[J]. Journal of Advanced Transportation, 2020, 2020: 1-20.
- [105] JU C, WANG Z, LONG C, et al. Interaction-aware Kalman neural networks for trajectory prediction[C]//Proceedings of 2020 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2020: 1793-1800.
- [106] LIN L, LI W Z, BI H K, et al. Vehicle trajectory prediction using LSTMs with spatial-temporal attention mechanisms[J]. IEEE Intelligent Transportation Systems Magazine, 2022, 14(2): 197-208.
- [107] SHI X J, GAO Z H, LAUSEN L, et al. Deep learning for precipitation nowcasting: a benchmark and a new model[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 5622-5632.
- [108] DING W C, CHEN J, SHEN S J. Predicting vehicle behaviors over an extended horizon using behavior interaction network[C]//Proceedings of 2019 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2019: 8634-8640.
- [109] MERCAT J, GILLES T, EL ZOGHBY N, et al. Multi-head attention for multi-modal joint vehicle motion forecasting[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2020: 9638-9644.
- [110] ZHANG K P, ZHAO L, DONG C X, et al. AI-TP: attention-based

- interaction-aware trajectory prediction for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(1): 73-83.
- [111] TANG C, ZHAN W, TOMIZUKA M. Exploring social posterior collapse in variational autoencoder for interaction modeling[J]. Advances in Neural Information Processing Systems, 2021, 34: 8481-8494.
- [112] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, 2018, arXiv:1810.04805.
- [113] LEE N, CHOI W, VERNAZA P, et al. DESIRE: distant future prediction in dynamic scenes with interacting agents[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 2165-2174.
- [114] ZHAO H, GAO J, LAN T, et al. Tnt: target-driven trajectory prediction[C]//Proceedings of Conference on Robot Learning. New York: PMLR, 2021: 895-904.
- [115] LIU M, CHENG H, CHEN L, et al. LAformer: trajectory prediction for autonomous driving with lane-aware scene constraints[J]. arXiv preprint, 2023, arXiv:2302.13933.
- [116] YUAN Y, WENG X S, OU Y L, et al. AgentFormer: agent-aware transformers for socio-temporal multi-agent forecasting[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 9793-9803.
- [117] PINI S, PERONE C S, AHUJA A, et al. Safe Real-World Autonomous Driving by Learning to Predict and Plan with a Mixture of Experts[J]. arXiv preprint, 2022, arXiv:2211.02131.
- [118] CHEN W, WANG F, SUN H. S2tnet: spatio-temporal transformer networks for trajectory prediction in autonomous driving[C]//Proceedings of Asian Conference on Machine Learning. New York: PMLR, 2021: 454-469.
- [119] BHAT M, FRANCIS J, OH J. Traformer: trajectory prediction with local self-attentive contexts for autonomous driving[J]. arXiv preprint, 2020, arXiv:2011.14910.
- [120] ZHAO J, LI X, XUE Q, et al. Spatial-channel transformer network for trajectory prediction on the traffic scenes[J]. arXiv preprint, 2021, arXiv:2101.11472.
- [121] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.
- [122] HUANG Z Y, MO X Y, LV C. Multi-modal motion prediction with transformer-based neural network for autonomous driving[C]//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2022: 2605-2611.
- [123] ZHANG K P, FENG X L, WU L, et al. Trajectory prediction for autonomous driving using spatial-temporal graph attention transformer[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 22343-22353.
- [124] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 9992-10002.
- [125] NAYAKANTI N, AL-RFOU R, ZHOU A, et al. Wayformer: motion forecasting via simple & efficient attention networks[J]. arXiv preprint, 2022, arXiv:2207.05844.
- [126] CHAMI I, ABU-EL-HAJJA S, PEROZZI B, et al. Machine learning on graphs: A model and comprehensive taxonomy[J]. Journal of Machine Learning Research, 2022, 23(89): 1-64.
- [127] LIANG M, YANG B, HU R, et al. Learning Lane Graph Representations for Motion Forecasting[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2020: 541-556.
- [128] BEHBAHANI F, SHIARLIS K, CHEN X, et al. Learning from demonstration in the wild[C]//Proceedings of 2019 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2019: 775-781.
- [129] CASAS S, GULINO C, LIAO R J, et al. SpAGNN: spatially-aware graph neural networks for relational behavior forecasting from sensor data[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2020: 9491-9497.
- [130] RHINEHART N, MCALLISTER R, KITANI K, et al. PRECOG: PREDiction conditioned on goals in visual multi-agent settings[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2020: 2821-2830.
- [131] CASAS S, GULINO C, SUO S, et al. Implicit Latent Variable Model for Scene-Consistent Motion Forecasting[C]//Proceeding of European Conference on Computer Vision. Cham: Springer, 2020: 624-641.
- [132] LI X, YING X W, CHUAH M C. GRIP: graph-based interaction-aware trajectory prediction[C]//Proceedings of 2019 IEEE Intelligent Transportation Systems Conference (ITSC). Piscataway: IEEE Press, 2019: 3960-3966.
- [133] LI X, YING X, CHUAH M C. Grip++: enhanced graph-based interaction-aware trajectory prediction for autonomous driving[J]. arXiv preprint, 2019, arXiv:1907.07792.
- [134] CHANDRA R, GUAN T R, PANUGANTI S, et al. Forecasting trajectory and behavior of road-agents using spectral clustering in graph-LSTMs[J]. IEEE Robotics and Automation Letters, 2020, 5(3): 4882-4890.
- [135] GAO J Y, SUN C, ZHAO H, et al. VectorNet: encoding HD maps and agent dynamics from vectorized representation[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 11522-11530.
- [136] GU J R, SUN C, ZHAO H. DenseTNT: end-to-end trajectory prediction from dense goal sets[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 15283-15292.
- [137] CHAI Y, SAPP B, BANSAL M, et al. Multipath: multiple probabilistic anchor trajectory hypotheses for behavior prediction[J]. arXiv preprint, 2019, arXiv:1910.05449.
- [138] HUANG Y F, BI H K, LI Z X, et al. STGAT: modeling spatial-temporal interactions for human trajectory prediction[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 6271-6280.
- [139] GUPTA A, JOHNSON J, LI F F, et al. Social GAN: socially acceptable trajectories with generative adversarial networks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2255-2264.
- [140] LI J C, MA H B, TOMIZUKA M. Conditional generative neural system for probabilistic trajectory prediction[C]//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE Press, 2019: 6150-6156.
- [141] YANG B, HE C, WANG P, et al. TPPO: a novel trajectory predictor with pseudo oracle[J]. arXiv preprint, 2020, arXiv:2002.01852.
- [142] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. SoPhic: an attentive GAN for predicting paths compliant to social and physical constraints[C]//Proceedings of 2019 IEEE/CVF Conference on Com-

puter Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 1349-1358.

- [143] LIU S H, LIU H B, BI H K, et al. CoL-GAN: plausible and collisionless trajectory prediction by attention-based GAN[J]. IEEE Access, 2020, 8: 101662-101671.
- [144] CHOI S, KWEON N, YANG C, et al. DSA-GAN: driving style attention generative adversarial network for vehicle trajectory prediction[C]//Proceedings of 2021 IEEE International Intelligent Transportation Systems Conference (ITSC). Piscataway: IEEE Press, 2021: 1515-1520.
- [145] KINGMA D P, WELING M. Auto-encoding variational bayes[J]. arXiv preprint, 2013, arXiv:1312.6114.
- [146] BALDI P. Autoencoders, unsupervised learning and deep architectures [C]//Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning workshop. New York: ACM, 2011: 37-50.
- [147] GÓMEZ-HUÉLAMO C, CONDE M V, ORTIZ M, et al. Exploring attention GAN for vehicle motion prediction[C]//Proceedings of 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC). Piscataway: IEEE Press, 2022: 4011-4016.
- [148] WANG W, LIU C, ZHAO D. How much data are enough? A statistical approach with case study on longitudinal driving behavior[J]. IEEE Transactions on Intelligent Vehicles, 2017, 2(2): 85-98.
- [149] 王晓, 杨林瑶, 胡斌, 等. 平行推理: 一种基于 ACP 方法的虚实互动的知识协同框架[J]. 智能科学与技术学报, 2023, 5(1): 69-82.
WANG X, YANG L Y, HU B, et al. Parallel reasoning: a virtual-real interactive knowledge collaboration framework based on ACP approach[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(1): 69-82.
- [150] 田永林, 陈苑文, 杨静, 等. 元宇宙与平行系统: 发展现状、对比及展望[J]. 智能科学与技术学报, 2023, 5(1): 121-132.
TIAN Y L, CHEN Y W, YANG J, et al. Metaverses and parallel systems: the state of the art, comparisons and prospects[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(1): 121-132.
- [151] MIAO Q H, LV Y S, HUANG M, et al. Parallel learning: overview and perspective for computational learning across Syn2Real and Sim2Real [J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(3): 603-631.

[作者简介]



黄峻 (1998-), 男, 中国科学院自动化研究所复杂系统管理与控制国家重点实验室轨迹规划工程师, 主要研究方向为智能交通系统、深度学习、强化学习。



田永林 (1994-), 男, 博士, 中国科学院自动化研究所博士后, 主要研究方向为计算机视觉、智能交通。



戴星原 (1993-), 男, 博士, 中国科学院自动化研究所助理研究员, 主要研究方向为人工智能、智能交通系统、认知自动驾驶。



王晓 (1988-), 女, 博士, 安徽大学人工智能学院教授, 青岛智能产业技术研究院院长, 主要研究方向为社会交通、动态网络组织、平行智能和社交网络分析。

平之行 (1961-), 男, 北方自动控制技术研究所资深研究员, 主要研究方向为平行系统、平行智能、社会计算、知识自动化。