

行人轨迹预测方法关键问题研究：现状及展望

杜泉成 王晓 李灵犀 宁焕生

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

DU Quancheng, WANG Xiao , LI Lingxi , NING Huansheng

引用本文：

杜泉成, 王晓, 李灵犀, 等. 行人轨迹预测方法关键问题研究：现状及展望[J]. 智能科学与技术学报, 2023, 5(2): 143-162.

Quancheng DU, Xiao WANG, Lingxi LI, et al. Key problems and progress of pedestrian trajectory prediction methods: the state of the art and prospects[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(2): 143-162.

您可能感兴趣的其他文章

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

基于强化学习的数据驱动多智能体系统最优一致性综述

An overview of optimal consensus for data driven multi-agent system based on reinforcement learning

智能科学与技术学报. 2020, 2(4): 327-340

基于混合神经网络的光伏组件输出特性数据驱动建模方法

Research on data-driven modeling for photovoltaic characteristics based on hybrid neural network

智能科学与技术学报. 2020, 2(2): 169-178

行人轨迹预测方法关键问题研究：现状及展望

杜泉成¹, 王晓^{2,3}, 李灵犀⁴, 宁焕生¹

(1. 北京科技大学计算机与通信工程学院, 北京 100083;

2. 安徽大学人工智能学院, 安徽 合肥 230601;

3. 青岛智能产业技术研究院, 山东 青岛 266109;

4. 美国印第安纳大学-普渡大学印第安纳波利斯联合分校电子与计算机工程系, 美国 印第安纳州 IN 46204)

摘要: 行人轨迹预测旨在利用观察到的人类历史轨迹和周围环境信息来预测目标行人未来的位置信息, 该研究具有重要的应用价值, 可以降低自动驾驶车辆在社会交互下的碰撞风险。然而, 传统的模型驱动的行人轨迹预测方法难以在复杂高动态的场景下对行人进行轨迹预测。相比之下, 数据驱动的行人轨迹预测方法依靠大规模数据集平台, 可以更好地捕捉和建模更复杂的行人交互关系, 进而取得较精准的行人轨迹预测效果, 成为自动驾驶、机器人导航和视频监控等领域的研究热点。为了宏观把握行人轨迹预测方法的研究现状及关键问题, 以行人轨迹预测技术和方法分类为切入点, 首先, 详述行人轨迹预测已有方法的研究进展并归纳了目前存在的 key 问题与挑战; 其次, 根据行人轨迹预测模型的建模差异, 将现有方法分为模型驱动和数据驱动的行人轨迹预测方法, 同时总结了不同方法的优缺点及适用场景; 然后, 对行人轨迹预测任务中使用的主流数据集进行了归纳总结, 并对比了不同算法的性能指标; 最后, 针对行人轨迹预测的未来发展方向进行了展望。

关键词: 行人轨迹预测; 数据驱动; 社会交互; 自动驾驶

中图分类号: TP29

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202315

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

DU Quancheng¹, WANG Xiao^{2,3}, LI Lingxi⁴, NING Huansheng¹

1. School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

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

3. Qingdao Academy of Intelligent Industries, Qingdao 266109, China

4. Department of Electrical and Computer Engineering, Indiana University-Purdue University Indianapolis, Indianapolis IN 46204, USA

Abstract: Pedestrian trajectory prediction aims to use observed human historical trajectories and surrounding environmental information to predict the future position of the target pedestrian, which has important application value in reducing collision risks for autonomous vehicles in social interactions. However, traditional model-driven pedestrian trajectory prediction methods are difficult to predict pedestrian trajectories in complex and highly dynamic scenes. In contrast, data-driven pedestrian trajectory prediction methods rely on large-scale datasets and can better capture and model more complex pedestrian interaction relationships, thereby achieving more accurate pedestrian trajectory prediction results, and have become a research hotspot in fields such as autonomous driving, robot navigation and video surveillance. In order to macroscopically grasp the research status and key issues of pedestrian trajectory prediction methods, We started with the classification of pedestrian trajectory prediction technology and methods. First, the research progress of existing pedes-

收稿日期: 2023-03-01; 修回日期: 2023-03-24

通信作者: 王晓, x.wang@ia.ac.cn

基金项目: 国家自然科学基金项目 (No.U1811463, No.62173329)

Foundation Items: The National Natural Science Foundation of China (No.U1811463, No.62173329)

trian trajectory prediction methods were elaborated and the current key issues and challenges were summarized. Second, according to the modeling differences of pedestrian trajectory prediction models, existing methods were divided into model-driven and data-driven pedestrian trajectory prediction methods, and the advantages, disadvantages and applicable scenarios of different methods were summarized. Then, the mainstream datasets used in pedestrian trajectory prediction tasks were summarized and the performance indicators of different algorithms were compared. Finally, the future development direction of pedestrian trajectory prediction was prospected.

Key words: pedestrian trajectory prediction, data driven, social interaction, autonomous driving

0 引言

交通安全问题一直是当今社会关注的焦点问题,安全的道路环境是自动驾驶车辆在社会交互下行驶的必要条件。根据世界卫生组织(WHO)发布的《全球道路安全现状报告》,每年全球道路死亡人数已达到史无前例的135万人。其中,超过一半的事故受害者是弱势道路交通参与者(vulnerable road users, VRU)^[1],如道路行人、骑行者和摩托车驾驶者等。因此,对VRU群体的道路行驶安全问题需要重点关注。另外,行人作为交通场景中的重要参与者,是交通事故中最大的受害者,对其未来运动轨迹进行合理的推理及预测,对于自动驾驶决策规划和道路交通安全具有重要意义。

行人轨迹预测的应用受到多个领域的关注,其中包括自动驾驶^[2]、智能交通^[3]、机器人导航^[4]和视频监控^[5]等领域。例如,自动驾驶感知系统需要精准预测行人轨迹以提高车辆行驶安全性;在机器人导航应用中,机器人只有准确预测智能体的轨迹信息才能避免碰撞;在人流较多的公共领域,通过视频监控对行人轨迹进行预测,可以避免紧急状况下发生踩踏事件。因此,在上述应用背景下,对行人轨迹预测进行研究变得越发重要。

近年来,国内外学者对行人轨迹预测方法的研究与日俱增,该类方法从传统的基于模型驱动的方法逐渐转向基于数据驱动的深度学习方法。与此同时,学习社会交互关系也成为行人轨迹预测中最受关注的热点话题。在实际场景中,人们在行走时通常不是独立的,他们的行动可能受到周围其他人的干扰和制约。例如,当行人穿过拥挤的人群时,他们可能会调整速度和方向,以避免与其他人碰撞或干扰他们的行动。因此,考虑社会交互关系在行人轨迹预测中是必不可少的,可以提高预测准确性和真实性,为智能驾驶、机器人导航和其他相关领域的应用提供更好的决策基础。

为了全面了解行人轨迹预测领域的相关方法,需要进行综述和梳理。同其他行人轨迹预测相关的综述文献^[6-8]相比,本文的区别和主要贡献在于笔者以行人轨迹预测技术和方法分类为切入点,对行人轨迹预测方法在数据驱动和模型驱动两种方式下存在的关键问题进行研究,并综述了相关文章的解决方案和思路。而其他文献^[6-8]则更多只关注单一的驱动方式,对行人轨迹预测方法的关键问题解析以及归纳内容的整体性上不够全面。本文的组织架构如下:第1节介绍了行人轨迹预测的相关工作,包括行人轨迹预测方法介绍和关键问题分析;第2节对行人轨迹预测方法进行分类对比与优缺点总结,同时对不同方法所适用的场景进行分析;第3节介绍了行人轨迹预测领域涉及的数据集及评价指标,并比较了不同算法的网络性能;第4节总结了全文并展望了行人轨迹预测技术的发展趋势。

1 行人轨迹预测方法研究进展及其关键问题分析

行人轨迹预测是车辆与周边环境进行社会交互并采取安全高效决策规划的重要保障。自动驾驶车辆通过预测行人的具体位置进而规划出合理的路线,能够有效降低因行人轨迹突变而造成的碰撞风险^[9]。本节将介绍行人轨迹预测方法,并从建模角度将其分类为基于模型驱动的方法和基于数据驱动的方法。此外,考虑到行人的行动会受到周围其他行人和物体的交互影响,笔者还结合社会交互关系对行人轨迹预测的约束性质,对该领域中的关键问题进行了阐述。

1.1 行人轨迹预测方法研究进展概述

行人轨迹预测是指基于给定的行人历史轨迹信息预测行人未来几秒内的运动轨迹^[10]。根据建模方式的差异性,可以将行人轨迹预测方法分为基于模型驱动的方法和基于数据驱动的方法。传统的基于模型驱动的行人轨迹预测方法主要通过复杂的数

学统计模型来表征行人的运动特征^[11]。其中，社会力模型^[12-14]和运动学模型^[15-17]是两类常见的模型。早期的工作中，Helbing等人^[12]首次提出社会力模型，通过能量势场（排斥力与吸引力）来描述行人之间的交互关系。Kooij等人^[17]将贝叶斯滤波器和运动学模型相结合，建立基于上下文的动态贝叶斯网络用于预测行人轨迹。Schneider等人^[18]将运动学模型（如常量速度模型）和卡尔曼滤波算法结合起来对行人轨迹进行预测。传统的基于模型驱动的行人轨迹预测方法具有严格的数学证明与假设，需要依赖手工定义规则及能量势场对模型进行严谨的数学建模以提取社会交互信息，进而获取行人运动轨迹。因此，该类方法的应用局限性较强，需要依赖领域专家的知识，对复杂场景的适应性较差。

近年来，随着深度学习技术的不断发展，基于数据驱动的深度学习预测方法已经成为行人轨迹预测的研究热点^[19-21]。与传统的基于模型驱动的方法相比，深度学习方法不需要手工定义规则及能量势场等复杂模型，而是直接从数据中学习行人的社会交互关系。通过训练大量的数据样本，深度神经网络可以更新网络参数，并自适应地学习出更加合理的映射关系，从而实现精准的行人轨迹预测。总体而言，相比模型驱动的方法，基于数据驱动的深度学习方法具有更高的灵活性和泛化性，能够适应不同场景下的行人运动模式，并实现更加精准的轨迹预测，因此具有更广阔的研究前景。

目前，基于数据驱动的行人轨迹预测方法主要分为4类，包括基于长短时记忆网络（long short-term memory, LSTM）^[22]、基于生成对抗网络（generative adversarial network, GAN）^[21]、基于图卷积神经网络（graph convolution network, GCN）和基于

Transformer^[20]的方法。其中，基于LSTM的方法是最常见和经典的一种方法，其通过对历史轨迹序列进行编码，再将编码后的信息输入LSTM中进行学习和预测。而基于GAN的方法则是通过对抗博弈思想将生成的轨迹与真实轨迹进行区分和优化，提高轨迹预测的准确性。基于GCN的方法则是将行人与周围环境建立图网络结构，通过GCN对图上的节点和边进行卷积运算来提取行人运动时的上下文信息，以实现更精确的预测。基于Transformer的方法则是近年来迅速发展起来的一种新方法，通过引入多头自注意力机制来对历史轨迹进行建模，并利用解码器生成未来轨迹。总之，基于数据驱动的轨迹预测方法在完成行人轨迹预测任务时，能够充分利用大规模的数据样本来提高模型的预测性能，并且能够适应不同的场景和更复杂的环境。

基于数据驱动的轨迹预测方法对行人轨迹进行预测时，可以将整个预测过程分为以下5个阶段，即输入阶段、特征提取阶段、编码阶段、解码阶段和输出阶段，具体流程如图1所示。下面对这5个阶段分别进行介绍。

（1）输入阶段

该阶段的主要任务是获取输入数据，即行人历史轨迹数据。通常情况下，历史轨迹数据是以序列的形式输入深度学习模型中的，因此需要将轨迹数据转化为适合深度学习模型输入的格式，例如向量、矩阵或张量等。此外，在输入阶段，模型需要对接收的历史轨迹信息进行解析，并处理与行人运动相关的各种交互特征信息。以上所指信息包括行人坐标信息、场景交互信息、语义信息、速度信息和位姿信息等，这些信息可以被用来描述行人的运动行为和社会交互关系，对于

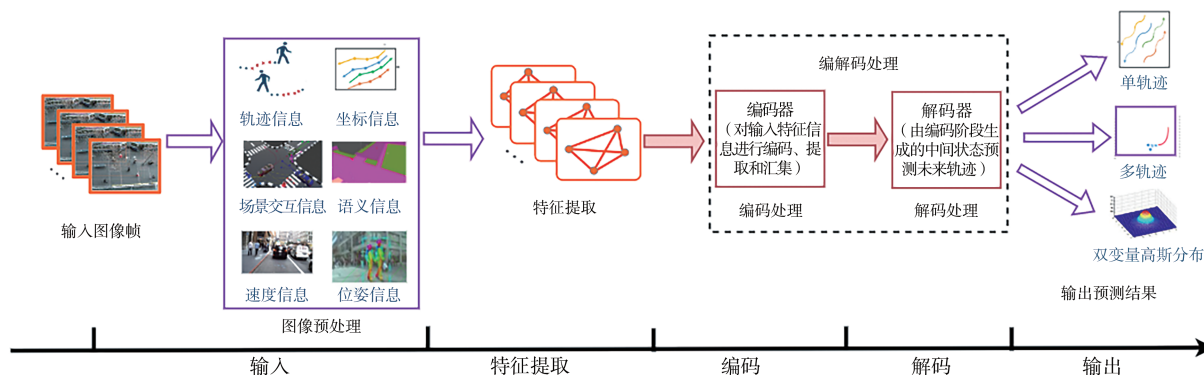


图1 基于数据驱动的行人轨迹预测方法流程

后续的特征提取和编码阶段的进行至关重要。

(2) 特征提取阶段

该阶段的主要任务是从输入数据中提取特征,以便模型能够更好地理解轨迹数据。在行人轨迹预测中,常用的特征提取模型包括卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)等。由于CNN能够从图像中提取空间信息,而RNN能够从序列数据中提取时间信息,因此这两种方法也常常被结合使用。

(3) 编码阶段

该阶段的主要任务是将特征信息编码成一个中间状态,以便后续解码阶段生成轨迹时使用。编码阶段的实现方式不尽相同,例如可以使用LSTM、GRU等循环神经网络,也可以使用GAN、Transformer等网络进行编码处理。

(4) 解码阶段

该阶段的主要任务是根据编码阶段生成的中间状态进行解析,进而预测未来的轨迹。解码阶段通常使用类似于自回归模型的方式进行轨迹生成,即利用前一个时刻生成的点作为输入,随后再生成下一个点。解码阶段也可以使用LSTM、GRU等循环神经网络或者Transformer等结构进行解码处理。

(5) 输出阶段

该阶段的主要任务是将解码阶段生成的轨迹输出为最终结果。输出的形式可以是轨迹的坐标点(包括单轨迹和多轨迹)形式,也可以是概率分布(如双变量高斯分布)形式。在输出阶段,通常会使用一些评估指标来衡量模型的预测性能,例如平均位移误差(average displacement error, ADE)和最终位移误差(final displacement error, FDE)等^[22]。

1.2 行人轨迹预测的关键问题

行人轨迹预测问题的本质可以被看作一个序列决策问题,即通过数据集中行人位置信息、历史序列信息等来推理(决策)出行人在未来固定时刻的位置坐标及轨迹^[23]。在已有的基于模型驱动和基于数据驱动的轨迹预测工作中,通常会面临预测结果不准确的问题,这主要是由以下3个关键因素造成的。

(1) 行人与目标之间动态的社会交互关系难以建模

在行人轨迹预测问题中,行人与目标之间的动

态社会交互关系是一个关键因素,因为它会在很大程度上影响行人的运动轨迹。在同一场景中,不同行人与目标行人之间的交互关系是动态变化的,这种交互力量被称为“社会力”^[12]。对目标行人进行轨迹预测的同时也会受到其他行人的运动趋势影响,因此在算法中需要考虑行人之间的社会交互关系和相互影响,以增加预测的准确性。此外,动态的社会交互关系往往具有复杂性和抽象性,仅仅通过简单的规则建模来体现行人之间的交互关系并不容易,因为这种交互关系难以量化和建模,进而增加了轨迹预测的难度。因此,研究人员需要探索更加先进的模型来捕捉这种复杂的动态交互关系,从而提高轨迹预测的准确性。

(2) 行人与场景之间静态的社会交互关系难以描述

行人的运动轨迹不仅受到行人个体的时序特征影响,也受到周围场景中其他实体的影响。在复杂场景中,行人在前行的过程中遇到障碍物时,会本能地选择不同的策略来改变自己的运动方向。场景中的障碍物可以分为静态障碍物和动态障碍物两类。静态障碍物包括建筑物、草坪、路边停放的汽车等,动态障碍物则包括行驶而来的汽车、骑行者等。在遇到静态障碍物时,目标行人会本能地选择绕行避开障碍物,而在遇到动态障碍物时,行人会预先估算出动态障碍物的行驶速度以及是否会对自身前行路径造成影响,进而会选择原地停止运动或者减速慢行^[11]。因此,如何准确地描述场景变化特征,进而深度刻画行人的时序特征信息与交互特征信息,是复杂场景下行人轨迹预测的一个关键问题。

(3) 行人意图及其与环境中其他主体的变化关系难以确定

在现实中,相对于自行车、汽车等运动学模型,行人运动方式更加灵活多变,预测其轨迹更加困难。行人的运动轨迹通常会受到行人意图、心理状态等情感因素的影响。因此,推理行人意图及确定其与环境中其他主体的变化关系也是行人轨迹预测的关键问题之一。现有的行人轨迹预测模型通常使用客观测量数据集进行训练实验,即模型只关心行人运动的最终结果,很少关注行人意图推理的层面,因此模型对行人意图把握不准确。例如,正在加速奔跑的行人可能会突然停下或者突然掉头再跑,而此时的预测模型只关心行人最终的动作是停

在原地还是跑到另一个位置，而不是推理行人为什么会这样做。这使得在极端场景（corner case）下，行人轨迹预测模型经常出现预测不准确的情况^[24]。

2 行人轨迹预测方法分类

行人轨迹预测方法可以从建模角度分为两大类：基于模型驱动的行人轨迹预测方法和基于数据驱动的行人轨迹预测方法。其中，基于模型驱动的行人轨迹预测方法主要基于社会力模型、运动学模型等对行人运动规律和社会交互关系进行数学建模，然后使用预测模型进行行人轨迹预测。这种方法借助于复杂的数学模型，可以较为准确地描述行

人的运动规律和社会交互关系，但需要提前进行模型建立和参数调整，且对场景的适应性较差。而基于数据驱动的行人轨迹预测方法则使用大规模数据集进行训练，通过学习数据集中的行人运动行为和社会交互关系建立预测模型，进而实现行人轨迹预测。这种方法以深度学习为主，不需要进行模型建立和参数调整，对场景具有很强的泛化性，但会对数据集的质量和数量要求较高等问题。综上所述，以上两种方法各有优缺点，具体的应用需要根据场景和数据情况进行选择。基于前文内容，本文对行人轨迹预测方法进行了详细分类，分类结果如图2所示，下面对以上分类涉及的文献、方法及思路分别进行详述。

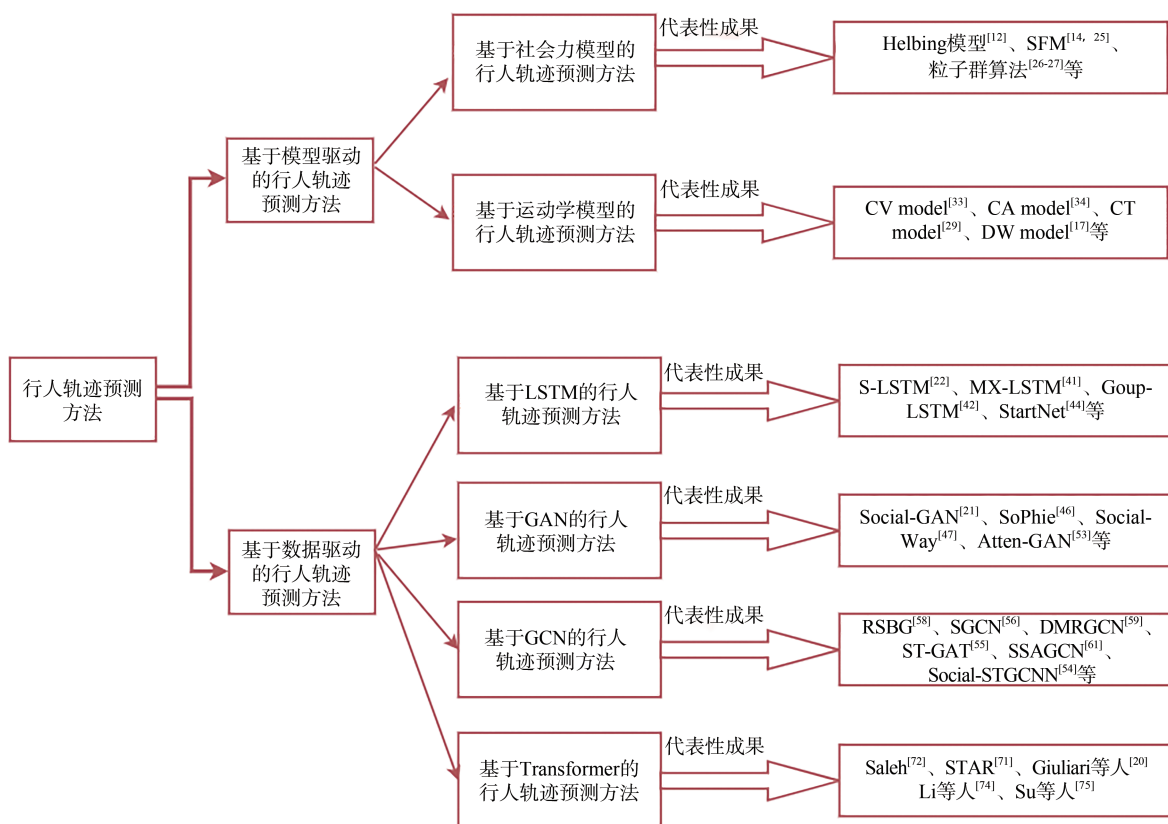


图2 行人轨迹预测方法分类

2.1 基于模型驱动的行人轨迹预测方法

2.1.1 基于社会力模型的行人轨迹预测方法

基于社会力模型的行人轨迹预测方法是一种基于人类社会行为学的模型，它模拟了行人在行走过程中的相互作用和群体行为。该模型通过描述行人之间的社会力和群体动力学行为来预测行人的未来动作和轨迹。具体来说，社会力模型将每个行人视

为一个受力体，考虑行人之间的相互作用和周围环境的影响，通过计算合力来对行人轨迹进行预测。常用的社会力模型包括 Helbing 模型^[12]、Social Force 模型^[14,25]、粒子群算法^[26-27]等。

Helbing 模型^[12]是最早的基于社会力模型的方法之一，它是一种基于微观力学的模型，用于描述行人运动的动态行为，其中行人之间相互影响的力

被建模为排斥力和吸引力。在模型中,行人试图避免相互碰撞并保持一定的速度和方向,而模型的主要目标是预测行人流的行为,例如拥堵和瓶颈等。**Social Force**模型则是对**Helbing**模型的改进,它将行人的运动视为受到一个包括人与人之间和人与环境之间相互作用的合力的影响,并通过建立相互作用模型来预测行人的轨迹。在过去的几十年中,上述开创性工作已经通过多种方法进行了扩展和改进^[28-30]。**Yan**^[13]基于社会力模型和局部行为模型来预测人类社交行为,并应用于行人跟踪任务中。其中,社会力模型主要用于建模行人与行人之间的交互作用,通过估计每个行人对周围环境的影响来预测其未来的移动轨迹。该方法结合了运动学模型和社会交互模型的优点,能够更准确地预测行人的轨迹和行为,实现较好的跟踪效果。**Rudenko**等人^[31]提出一种基于规划的社会力模型方法来实现人体运动的长期预测任务。该方法将社会力模型和规划方法相结合,首先使用社会力模型来预测每个行人的运动轨迹,然后将这些轨迹用于规划行人之间的交互,通过联合预测方式来长期捕获人体运动。**Trautman**等人^[32]提出了一种用于机器人导航的动态交互人群模型,旨在通过在人群中使用交互模型来预测和规划机器人行动,从而实现在密集人群中安全、有效地导航。该模型基于人-人交互作用力的概念,通过将人与机器人之间的交互建模视为相互作用的动态力场来描述人群的行为。作者验证了该方法的有效性和准确性,并展示了该模型在密集人群中机器人导航方面的优越性。

具体来说,基于社会力模型的行人轨迹预测方法将每个行人视为一个受力体,考虑行人之间的相互作用和周围环境的影响,通过计算合力对行人进行轨迹预测。该方法的优点在于它能够较好地捕捉到行人之间的协同行为,如群体移动、分散、聚集等,并且能够应对复杂的场景,如拥挤的人群、突发事件等。不过,基于社会力模型的轨迹预测方法也存在一些缺点。首先,该模型需要大量的参数调整,因此在实际应用中可能存在一定的难度;其次,社会力模型的计算复杂度较高,需要考虑大量的变量,导致计算效率低下,因此无法应用到对实时性要求很高的无人驾驶场景;此外,该模型在处理行人个体之间的交互关系时可能存在一定的误差,无法完全捕捉到行人的个体特征和行为习惯。综上所述,基于社会力模型的行人轨迹预测方法具

有一定的优点和缺点,需要根据具体场景进行选择和应用。

2.1.2 基于运动学模型的行人轨迹预测方法

基于运动学模型的行人轨迹预测方法是一种基于物理学原理的模型驱动方法。它通常采用行人运动学模型来描述行人的运动过程,以推导出行人的运动轨迹。运动学模型通常包括行人的位置、速度和加速度等运动状态,以及行人与周围环境之间的交互作用。这些模型可以基于行人的运动规律和行人在场景中的行为模式来构建。

在行人轨迹预测领域,基于运动学模型的方法通常需要先通过观察行人的历史轨迹数据来估计其当前的状态(位置、速度和加速度等),然后通过行人运动学模型来预测其未来的运动轨迹。常用的预测方法包括基于常量速度模型(CV model)^[33]、常量加速度模型(CA model)^[34]、常量转向率模型(CT model)^[29]、动态窗口模型(DW model)^[17]等。这些方法可以通过对行人未来运动的方向、速度、加速度等进行估计,进而预测行人的未来位置和轨迹。

在早期的工作中,**Karamouzas**等人^[34]提出一种基于长加速度模型的行人避碰方法,用于模拟行人在拥挤环境中的运动。该模型结合了运动学和社会力模型,可以预测行人的运动方向和速度,并避免与其他行人和障碍物的碰撞。**Zhou**等人^[35]提出了一种基于运动学模型的行人-智能体混合模型(mixture model of dynamic pedestrian-agents, MDA)来理解集体人群交互行为。该模型包括3个变量:行人轨迹、行人个体属性和行人之间的交互关系。MDA基于贝叶斯框架进行推断,并且使用了一个粒子滤波器来对模型进行训练和预测。实验结果表明,该模型能够准确地预测行人的轨迹和行为,同时也能够帮助人们更好地理解行人群体的行为模式,比如行人的拥挤程度、走路速度等。**Kooij**等人^[17]将贝叶斯滤波器和运动学模型相结合来预测行人路径。该模型将姿态感知、事件紧急程度和场景上下文空间布局作为潜在状态嵌入切换线性动力系统(switched linear dynamical system, SLDS)^[36]模型顶部,进而控制SLDS的切换状态,以实现更精准的行人路径预测。**Best**等人^[37]提出一种基于贝叶斯推理的意图推断模型,将观察到的行动轨迹和先验知识融合在一起,从而推断行人的目标位置和未来行动。该方法通过将人类行动过程建

模为马尔可夫决策过程来实现预测。实验结果表明,该模型可以自动识别行人的行为意图,具有较高的预测精度,并且可以应用于复杂的人类运动场景。但是该模型需要大量的训练数据进行模型的训练,且计算复杂度较高。Xie等人^[38]提出了一种基于深度神经网络和运动学模型相结合的方法,用于学习和推断视频中人类行为的“暗物质”,以预测人类在复杂场景中的运动轨迹和意图。论文中模型结构采用基于循环神经网络(RNN)和卷积神经网络(CNN)的方法来对人类行为进行建模,进而在人类行为预测任务上实现较高的准确性。但是该模型比较复杂,需要大量的训练数据和计算资源,并且在场景中存在大量的遮挡、重叠、交叉等情况下预测效果可能会

降低。

综上所述,基于运动学模型的行人轨迹预测方法是一种基于物理学和运动学原理的预测方法,利用行人运动的速度、加速度和方向等信息,通过数学模型对行人未来的运动轨迹进行预测。这种方法可以准确地预测行人的运动轨迹,适用于人流密集的公共场所,如地铁站、商场、车站等室内场景,可以用于人员管理、安全监控等领域。但是,该方法存在的一些局限性阻碍了其进一步发展,例如模型假设不准确、建模相对简单、预测范围有限等问题,这些问题可能会导致预测的精度下降。因此,需要在应用中根据实际情况进行调整和优化。以下是基于模型驱动的行人轨迹预测方法的相关文献对比,详见表1。

表1 基于模型驱动的行人轨迹预测方法对比

方法	输入信息	网络结构	优缺点
Helbing等人 ^[12]	行人自身状态和周围环境信息	基于社会力模型结构	优点:模型简单有效、适用性强 缺点:参数敏感、缺少对行人不确定性建模
Karamouzas等人 ^[34]	行人位置、速度、加速度等信息以及场景信息	基于运动学模型和社会力模型的混合模型	优点:长期预测 缺点:训练成本较高、推理速度较慢
Trautman等人 ^[32]	历史轨迹信息	基于社会力模型和高斯模型结合	优点:泛化性强 缺点:实时性较低、模型复杂度较高
Zhou等人 ^[35]	行人历史轨迹数据	基于混合高斯模型的无监督模型结构	优点:高准确性、实时性 缺点:依赖大量数据、可解释性较差
Kooij等人 ^[17]	历史轨迹数据和环境上下文信息	基于贝叶斯滤波器和运动学模型	优点:全局建模、多模态输入、精度高 缺点:计算量大、需要大量训练数据
Yan等人 ^[13]	历史轨迹数据信息	基于运动学模型和社会力模型结合	优点:实时性高、灵活性好、可解释性强 缺点:网络结构复杂、训练难度大
Best等人 ^[37]	行人历史轨迹信息、场景上下文信息	基于运动学模型和贝叶斯模型结合	优点:可预测性强、可解释性好 缺点:计算复杂、依赖先验知识
Xie等人 ^[38]	行人位置、姿态、方向信息以及场景上下文信息	基于循环神经网络和卷积神经网络以及运动学模型结合	优点:建模精准、多模态融合 缺点:模型复杂、需要大量的训练数据和计算资源
Rudenko等人 ^[31]	地图、行人速度、位置以及环境信息	基于社会力模型和循环神经网络模型结合	优点:多模型组合、长期预测、扩展性好 缺点:计算复杂

2.2 基于数据驱动的行人轨迹预测方法

基于数据驱动的行人轨迹预测方法是近年来发展较快的一类方法,该类方法通过收集和分析历史轨迹数据,利用机器学习和深度学习等数据驱动技术建立模型,以预测未来行人轨迹。相较于基于模型驱动的方法,基于数据驱动的方法可以更好地处理复杂环境和行人交互行为,该方法不需要对环境进行特定的建模,而是直接从数据中学习模型。基于数据驱动的行人轨迹预测方法主要包括基于LSTM的行人轨迹预测方法、基于GAN的行人轨迹预测方法、基于GCN的行人轨迹预测方法和基

于Transformer的行人轨迹预测方法。以上类比方法的模型在训练完成后对同类型数据输入具有很好的适用性,拥有较高的轨迹预测精度,成为近些年来行人轨迹预测任务的研究热点。下面对这4类行人轨迹预测方法进行详细介绍。

2.2.1 基于LSTM的行人轨迹预测方法

基于LSTM的行人轨迹预测方法是一种常见的基于数据驱动的方法,可用于对行人轨迹进行建模和预测。LSTM是一种递归神经网络,能够处理序列数据和捕捉长期依赖关系,比较适用于行人运动轨迹预测任务。在行人轨迹预测领域,基于LSTM

的轨迹预测方法利用 LSTM 模型来建模行人轨迹的时间序列数据,并根据历史轨迹数据来预测未来轨迹的运动方向和位置。通常,模型的输入数据包括历史位置和速度等信息,输出数据为未来的位置等信息。模型的训练通常采用监督学习方法,即通过比较预测轨迹和真实轨迹之间的误差来更新模型参数。鉴于 LSTM 模型方法的巨大优势,大量研究者将其应用在行人轨迹预测的任务上。

Alahi 等人^[22]首次将 LSTM 模型应用到行人轨迹预测任务中,提出一种 S-LSTM 网络 (social long short-term memory, S-LSTM) 用于社会交互空间下的行人轨迹预测。相较于以往工作中存在的两个难点——手工构造能量函数而不能通过数据驱动方式来捕获社会交互关系,以及只考虑近距离情景而不能涵盖更远距离可能发生的社会交互场景, S-LSTM 模型通过采用数据驱动方式建模行人之间复杂的社会交互行为,进而能够更好地捕捉行人之间的社会交互关系。具体地,模型首先为场景中的每个行人配备一个独立的 LSTM 来提取各自的运动特征信息。其次,通过社交池 (social-pooling) 层相互连接来提取空间中隐藏的交互特征信息,以此来计算周围行人交互产生的影响,最后生成无冲突的行人轨迹。S-LSTM 模型中基于社交池化机制来共享空间信息的想法为基于数据驱动方式进行行人运动交互建模奠定了基础,大量的研究者在随后的工作中提出了类似结构,在公用的行人轨迹预测数据集上测试结果不断提升^[24,39-40]。

Hasan 等人^[41]提出一种 MX-LSTM 模型,通过对已有的社交池化机制进行优化来获取行人预测轨迹。该方法中,作者考虑到行人在交互过程中只能注意到人头朝向的扇形区域的其他行人的事实,因此模型通过社交池对扇形区域的可视化范围内行人进行池化处理,借助标准头部姿势估计器和基于注意力的社交池来提高模型长期轨迹预测的能力。文献[42-43]提出通过人为规定邻域范围的方法来选取参与交互的对象或者通过定义一个函数式的群组进而筛选交互对象的模型,通过社交池化操作来融合池化特征的方式获得行人社会交互特征。以上方式能够成功获取行人的运动轨迹,但是在复杂的动态场景下,人为设定的规则不能总是涵盖全部的场景状况,并且人为计算交互方式在很大程度上增加了模型计算量。为此,Zhu 等人^[44]提出了一种 StarNet 神经网络模型,通过建模行人之间的全局交互来实

现高效的行人轨迹预测。该模型中心网络节点将参与交互的所有行人共同计算为一个集合交互特征,随后再分配给场景中的所有行人,从而节省了网络对交互方式的计算量。模型中的 Hub Network 模块是基于 LSTM 的全局时序交互计算网络,用于获取所有行人的观测轨迹。而 Host Network 是基于 LSTM 的轨迹预测网络,每个 Host Network 对应一个行人,通过参考描述信息对未来轨迹进行预测。该模型在行人轨迹预测主流数据集 ETH/UCY^[29,45]上的实验表明,在 80% 的场景下该模型的效果优于其他算法,并具有较高的实时性。

综上所述,基于 LSTM 的方法是一种常见的行人轨迹预测方法。该方法通过手工定义规则对池化对象进行更细化的定义,并利用社交池化层对交互对象进行融合,以获取社会交互特征。这类方法主要解决行人本身对轨迹的影响问题,然后根据社会交互模型的结论来修正目标预测轨迹,从而获得更准确的预测结果。然而,基于 LSTM 的行人轨迹预测方法中池化机制无法显式地获取场景中所有行人对交互场景的影响程度。此外,该方法主要通过隐藏状态函数来存储历史时刻的特征信息,当历史信息数据量较大时,可能会出现数据丢失的现象,从而导致预测结果不准确。最后,由于 LSTM 是一种 Seq2Seq 的时序结构,网络无法实现并行计算,这也大大降低了网络的性能。因此,随着时间的推移,研究者们逐渐发展了基于 GAN、基于 GCN 以及基于 Transformer 的新型行人轨迹预测方法。表 2 对基于 LSTM 的行人轨迹预测方法进行了汇总。

2.2.2 基于 GAN 的行人轨迹预测方法

基于 GAN 的行人轨迹预测方法利用 GAN 的生成能力,从历史轨迹数据中学习行人移动的规律并生成未来的轨迹。该网络通常包含生成器和判别器两个部分。生成器从历史轨迹数据中学习行人交互行为并生成未来的轨迹,判别器则判断生成的轨迹数据是否与真实轨迹数据相似,以判断真假性。相较于基于 LSTM 等循环神经网络的方法,基于 GAN 的方法能更好地处理轨迹中的不确定性和多模态性,并生成具有多样性的轨迹结果。在行人轨迹预测领域,基于 GAN 的代表性工作包括 Social GAN^[21]、SoPhie^[46]、Social Way^[47]、STI-GAN^[48]和 AEE-GAN^[49]等方法。

Gupta 等人^[21]首次将 GAN 方法引入行人轨迹预测任务中,提出一种基于生成对抗网络 (Social

表2

基于LSTM的行人轨迹预测方法

方法	输入信息	网络结构	数据集
S-LSTM ^[22]	行人历史轨迹信息	基于LSTM网络和社交池机制结合	ETH/UCY
MX-LSTM ^[41]	行人历史轨迹和姿态信息	基于LSTM网络、注意机制和社交池化机制	ETH/UCY、KITTI
Group LSTM ^[42]	历史轨迹信息	基于LSTM模型于社交池化机制	ETH/UCY
Social-Grid LSTM ^[43]	社会交互信息、时序上下文信息	基于社交池化与LSTM结合	ETH/UCY
SS-LSTM ^[39]	历史轨迹、上下文场景信息	基于LSTM编解码结构	ETH/UCY
Scene-LSTM ^[40]	场景上下文、行人轨迹点信息	基于LSTM网络和CNN网络结合	ETH/UCY
Shi等人 ^[24]	历史轨迹信息	基于社交池化机制	GCDC/MOT17
StarNet ^[44]	行人历史轨迹信息	基于LSTM和社交池化机制	ETH/UCY
SNS-LSTM ^[22]	场景上下文、社会交互信息	基于LSTM网络和池化机制	ETH/UCY

GAN)的行人轨迹预测方法。与传统方法不同的是,该方法利用了GAN的生成对抗思想,通过LSTM构建生成对抗网络模型,对行人轨迹进行预测。具体来说,该方法首先利用“运动编码器模块”处理时间信息,其次通过“位置编码器模块”对空间交互信息进行建模,最后通过生成对抗网络的生成器和判别器相互博弈来不断更新模型参数,进而生成更加规范的行人轨迹。该方法将行人抽象为地图上的点,并利用GAN的特征实现了行人多模态轨迹预测,但它未充分考虑目标周围的视觉特征和场景上下文等语义信息,仅考虑了行人之间的交互关系,这会导致模型缺少可解释性,且可能产生违背生活规律的预测结果。因此,该方法还需要进一步改进以提高预测准确性和可解释性。近期的研究中,Sadeghian等人^[46]设计了一种基于生成对抗网络的可解释架构SoPhie。该架构在模型中添加了社会注意力和物理注意力机制模块,用于为周围的行人和场景分配不同的注意力权重,进而提取与路径相关的图像显著特征信息(深层特征信息);最后,利用GAN生成样本来捕捉行人行为轨迹的不确定性。实验结果表明,该方法能够实现更加鲁棒的行人轨迹预测效果,相比之前的方法具有更好的可解释性。Social-BiGAT^[50]采用了类似文献[21,46]中基于行人社会交互机制的想法,设计了一种基于图的生成对抗网络模型。该模型通过图注意力网络^[51]对场景上下文信息和行人的多模态交互行为进行编码,利用LSTM的隐藏状态和图注意力网络来建模社会交互关系,并通过对抗训练的方式生成更加真实的行人运动轨迹。为了验证该方法的有效性,作者在当前主流数据集(ETH/UCY)上进行了实验。实验结果表明,该方法能够有效地预测行人的运动轨迹。

传统的GAN存在模式崩溃和模式下降等问题,这极大地影响了网络的稳定性和训练效果。为了解决这些问题,Amirian等人^[47]提出了一种Social Way网络架构,该架构通过引入Info-GAN^[52]结构来改进多模式轨迹预测,从而避免了GAN出现的模式崩溃和模式下降问题。该方法在Social LSTM和Social GAN模型的基础上进一步提升,通过引入注意力机制使模型能够自主分配交互信息的注意权重。在模型结构方面,Social Way舍弃了L2代价函数,引入基于互信息的信息丢失(information loss),从而使网络在多模态行人轨迹预测任务上具有更好的预测效果。Fang等人^[53]提出了一种基于注意力机制的生成对抗网络模型(Atten-GAN)来进行行人轨迹预测。该模型利用注意力机制来提取输入轨迹的时空特征,然后在GAN的框架下进行轨迹生成。实验结果表明,该方法相对于其他方法具有更好的预测性能。

综上所述,相比于基于LSTM等循环神经网络的方法,基于GAN的方法通过生成器与判别器之间的相互博弈来训练模型,能够更好地处理轨迹中的不确定性和多样性,并且能够生成具有多样性的轨迹结果。然而,该方法也存在一些缺点影响其进一步运用的空间。其主要缺点如下:①网络训练不稳定,训练过程中存在模式崩塌问题,即生成器网络只能生成部分数据的情况;②GAN训练过程相对较慢,需要更长的时间和更多的计算资源;③网络难以处理长期依赖关系,预测结果可能存在短期内的偏差;④GAN模型的可解释性较差,难以理解模型是如何生成轨迹的。这也是近年来基于GAN的方法很少应用在轨迹预测领域的原因。表3对基于GAN的行人轨迹预测方法进行了汇总。

表3 基于GAN的行人轨迹预测方法对比

方法	输入信息	网络结构	数据集
Social-GAN ^[21]	行人历史轨迹信息	基于LSTM的编解码器结构	ETH/UCY
Sophie ^[46]	历史轨迹和上下文场景信息	基于GAN网络和注意机制结合	ETH/UCY和SDD
Social-BiGAT ^[50]	行人动态特征、场景上下文信息	基于GCN和GAN结合	ETH/UCY
Social Way ^[47]	历史轨迹信息	基于Info-GAN和注意机制结合	ETH/UCY
AEE-GAN ^[49]	场景上下文、历史轨迹	基于InfoGAN网络和LSTM网络架构结合	ETH/UCY和SDD
STI-GAN ^[48]	历史轨迹信息	基于GAN和图注意机制网络结合	ETH/UCY
Atten-GAN ^[53]	历史轨迹信息、场景图信息	基于GAN和双向循环神经网络结合	ETH/UCY

2.2.3 基于GCN的行人轨迹预测方法

基于GCN的行人轨迹预测方法是近年来提出的一种方法。该方法利用GCN对行人之间的空间关系进行建模，从而捕捉社会交互信息和环境特征，进而实现行人轨迹预测。具体来说，该方法首先将行人视为节点，并将节点之间的关系表示为图网络结构；其次，通过GCN对图网络结构进行卷积操作，从而学习节点之间的相互作用；然后，将每个节点的位置、速度等信息作为节点特征，通过GCN进行特征学习和传播；最后，预测出每个行人的未来位置。目前，已经有许多学者提出了基于GCN的行人轨迹预测方法，如Social STGCNN^[54]、STGAT^[55]和SGCN^[56]等。这些方法在不同的数据集和场景下进行了验证，并取得了较好的轨迹预测效果。

Mohamed等人^[54]在STGCN模型^[57]的基础上进行改进，提出的社会时空图卷积神经网络（social spatio-temporal graph convolutional neural network, Social STGCNN）是在行人轨迹预测领域中使用GCN的代表性工作。该网络通过将行人之间的社会交互关系嵌入邻接矩阵中，建模为图网络结构，并通过核函数对邻接矩阵进行处理，以捕捉空间和时间信息。这种方法能够优化社会交互行为模型，减少网络所需参数并提高网络计算速度，进而实现更高效的轨迹预测效果。Sun等人^[58]提出一种递归社交行为图（recursive social behavior graph, RSBG）的方法来进行行人轨迹预测。其将具有相似行为、目的地的行人分组，并对组内行人进行关系标注，进而对整个场景中的行人关系进行详尽的抽取。RSBG网络通过递归的方式对交互范围内的个体特征进行不断更新，进而获取更好的实时交互关系，最后通过LSTM解码器生成准确的预测结果。Shi等人^[56]提出了一种用于行人轨迹预测的稀疏图卷积网络（sparse graph convolution network, SGCN）模型，解决了已有工作中行人密集无向交互中存在

的建模冗余以及忽略行人轨迹运动趋势的问题。该模型使用稀疏有向空间图对行人交互关系进行建模。使用稀疏有向时间图来建模运动趋势，便于对观测方向进行预测。最后将上述两种稀疏图进行融合，推理出用于轨迹预测的双变量高斯分布参数进行行人轨迹预测。SGCN模型在ETH/UCY数据集上实现了精准的行人轨迹预测。但是该方法需要固定参数进行阈值设定，无法适应不同的交互场景，从而缺乏对复杂场景中行人动态交互关系的捕捉能力，因此也很难在无人驾驶场景中展开使用。在行人社会交互过程中存在的过度避碰问题上，Bae等人^[59]提出一种解耦多关系图卷积网络（disentangled multi-relational graph convolutional network, DMRGCN），其利用一组子图来表示行人复杂的社会交互关系，进而构建出基于社会关系的图卷积网络来提取场景中复杂的社会交互行为。此外，模型通过时间卷积网络（temporal convolutional network, TCN）设计了一个全局的时间聚合^[51]函数用于补偿因过度避碰造成的轨迹累积误差。实验结果表明，该模型能够纠正过度避碰导致的预测误差，实现了较好的轨迹预测结果。此外，在行人穿行意图推理上，Cadena等人^[60]提出一种Pedestrian Graph+模型用于估计行人穿越马路的意图，该模型将行人位姿信息和上下文场景信息作为输入数据，通过GCN建模其交互关系，进而在行人预测数据集上得到了更快、更准确的性能。

在行人轨迹预测领域，基于注意力机制的图卷积神经网络也得到了广泛的应用。在行人交互过程中，容易受到静态障碍物（例如墙壁、树木和建筑物等）和动态障碍物（例如其他行人、自行车和汽车等）的影响而改变行人的运动轨迹。因此，考虑到不同物体的影响力，将其分配不同的权重并参与行人轨迹预测是必要的。先前的研究工作中，社会力模型^[12]和基于池化机制^[21-22]的方法通过获取行人

之间的欧氏距离来计算交互关系的强度,但这些方法不能直观地解释不同物体在行人交互过程中的影响力。相比之下,图注意力网络^[51]作为一种新型的网络架构,通过在图结构上运行网络,能够隐式地为图网络邻域中的不同节点分配不同的权重。图注意力网络利用自注意层的掩码计算不同物体的影响力,能够更加准确地捕捉到物体之间的交互关系,进而提高行人轨迹预测的准确性。

基于此,Huang等人^[55]在行人轨迹预测领域引入了时空图注意网络(spatial-temporal graph attention network, STGAT)用于聚合场景内不同智能体的运动特征。STGAT是一种Seq2Seq序列架构,通过图注意力机制捕获每个时刻行人的空间交互行为信息,并采用LSTM对行人的时间交互信息进行编码,从而实现较好的多模态轨迹预测结果。实验结果表明,图注意网络中的注意力机制能够隐式地为节点分配不同权重,并根据邻居的运动状态为其分配合理的重要性,进而提高网络的预测精度。Lyu等人^[61]提出一种社会软注意力图卷积网络(social soft attention graph convolution network, SSAGCN)模型。该模型能够同时处理行人之间的社会交互关

系以及行人与周围环境之间的物理场景交互信息。对于社会交互信息,在社交软注意函数和GCN作用下聚合社交特征。对于物理交互信息,使用坐标构建的初始图与顺序场景进行合并。最终通过TCN获得预测的行人轨迹。受Mohamed等人^[54]的启发,Zhou等人^[62]提出一种AST-GNN模型用于交互感知行人的轨迹预测。该模型在Social STGCNN中的时空图卷积层中加入了注意力机制,提升了网络对行人社交行为的时空推理能力。具体来说,AST-GNN模型中S-GNN和T-GNN替换原有的时空图卷积网络,通过在图神经网络中引入注意力机制,在一定程度上提升了行人轨迹的预测精度。

综上所述,基于GCN的行人轨迹预测方法具有许多优点,如能够建模复杂的空间和时间关系、具有良好的可扩展性、能够自适应地为每个节点分配不同的权重及具有较高的预测准确度等。然而,该方法也存在一些缺点,如计算复杂度高、对数据质量的要求较高、可解释性较差且难以应用于实时场景等。总体来说,该方法在行人轨迹预测领域具有广泛的应用前景,但仍需要进一步研究和改进。表4对基于GCN的行人轨迹预测方法进行了汇总。

表4

基于GCN的行人轨迹预测方法对比

方法	输入信息	网络结构	数据集
STGAT ^[55]	历史轨迹信息、场景图像信息	基于GCN网络、LSTM网络和GAT网络结合	ETH/UCY
RSBG ^[58]	历史轨迹信息	基于GCN网络和LSTM网络、CNN网络结合	ETH/UCY
Social-STGCNN ^[54]	历史轨迹信息	基于GCN网络和CNN网络结合	ETH/UCY
SGCN ^[56]	历史轨迹信息、速度、加速度信息	基于稀疏有向图建模时空关系	ETH/UCY
DMRGCN ^[59]	历史轨迹信息、行人速度信息	基于GCN和注意机制结合	ETH/UCY
SSAGCN ^[61]	历史轨迹信息	基于GCN、TCN和注意机制结合	ETH/UCY、SDD
AST-GNN ^[62]	历史轨迹信息	基于GCN和注意机制结合	ETH/UCY
Pedestrian Graph + ^[60]	行人位姿、场景上下文信息、车辆速度信息	基于GCN网络和卷积神经网络结合	JAAD/PIE

2.2.4 基于Transformer的行人轨迹预测方法

基于Transformer的方法已经成为深度学习任务发展的一个趋势^[63]。该方法最初被用于自然语言处理(natural language processing, NLP)^[64-66]任务中,之后在计算机视觉领域迅速发展^[67-69]。在行人轨迹预测领域,基于Transformer的行人轨迹预测方法将行人轨迹看作时间序列数据,利用Transformer模型进行建模和预测。与基于LSTM、GAN和GCN等的模型相比,基于Transformer的模型具有并行计算和长程依赖建模的优势,能够更好地处理轨迹数据中的长程依赖和多模态性。具体来说,

该方法将行人轨迹中的位置、速度等信息视为输入序列,利用Transformer模型对序列进行编码和解码,从而预测行人的未来位置。在编码过程中,该方法通过多头自注意力机制学习序列之间的相互关系,并使用位置编码对序列中的位置信息进行建模。在解码过程中,该方法利用预测过程中得到的信息对序列进行逐步生成,以此得到未来的轨迹。基于Transformer的行人轨迹预测方法可分为单模态预测和多模态预测两类。单模态预测是指在预测行人轨迹时,只考虑行人当前的位置、速度等单一模态的信息,不考虑多种可能的情况。多模态预测

则考虑了不同模态（上下文信息、空间关系和位姿信息等）之间的关系，通过对可能的模态进行建模，得到多个不同的预测结果，并通过模型的可信度进行融合，提高预测的准确性。

在单模态轨迹预测方法上，Giuliani 等人^[20]利用 Transformer 模型实现了单模态行人轨迹预测任务。模型基于 Transformer 的编码器和解码器结构加入当前和过去时刻的位置信息进行训练和测试，用于预测未来的行人轨迹。该方法较好地实现了单模态轨迹预测效果，但是未考虑解码器输出的时序性会造成预测结果级联错误的问题。为此，Yao 等人^[70]提出了一种基于 Transformer 的端到端的轨迹预测算法，通过在解码器的训练集中加入适当的偏差（噪声），迫使网络实现自纠错的能力，避免了文献[20]中出现的轨迹预测结果级联错误的问题。Yu 等人^[71]提出了一种基于空间-时间图 Transformer 网络（spatio-temporal graph transformer networks）的单模态行人轨迹预测方法。该方法使用空间-时间图来表示行人轨迹数据，然后利用图卷积神经网络和 Transformer 结构对行人轨迹进行建模和预测。其中，图卷积神经网络用于提取行人轨迹数据中的空间关系，Transformer 用于学习序列中的时间关系，从而更好地对行人轨迹数据进行建模和预测。Saleh^[72]提出一种基于上下文增强 Transformer 网络的单模态行人轨迹预测方法。该方法使用历史轨迹数据以及环境上下文信息（例如地图）来预测未来的行人轨迹。其中，历史轨迹数据通过位置编码和时间编码输入 Transformer 网络中，而环境上下文信息则通过卷积神经网络编码，并与历史轨迹数据的编码进行结合。最后，通过 Transformer 解码器输出未来的轨迹预测。该方法在多个数据集（ETH/UCY）上进行了实验验证，取得了较好的预测性能。

在多模态轨迹预测方法上，Yin 等人^[73]考虑到 CNN 或 RNN 在捕捉行人和自行车之间的高动态运动交互特征方面存在的缺陷，设计了一种多模态 Transformer 网络架构（multimodal transformer network, MTN），通过引入光流来补偿自行车和行人之间的高动态运动特征，进而预测行人多模态轨迹。实验结果表明，该方法能够很好地应用在无人驾驶场景中。Li 等人^[74]结合基于图的空间 Transformer 和记忆重放机制实现对空间交互信息的全面利用，并且纠正了轨迹预测中的时间不一致问题。该方法

在多个行人轨迹预测数据集上进行了实验，并与多种基准模型进行了比较，实验结果表明该方法能够在多模态轨迹预测任务中取得优秀的性能。Su 等人^[75]提出了一种基于交叉模态 Transformer 的生成框架来进行行人轨迹预测。该模型将视觉、语义和时空信息集成到同一模型中，实现了对多模态数据的处理。实验结果表明，该模型在多个数据集上均取得了最优或接近最优的效果，证明了其在多模态轨迹预测任务中的有效性。

综上所述，基于 Transformer 的行人轨迹预测方法可以根据预测输入的信息分为单模态预测和多模态预测两类。在单模态预测中，只考虑行人自身的历史轨迹信息进行预测，而在多模态预测中，还会考虑环境信息和其他行人的信息等多种输入。总体而言，无论是单模态还是多模态预测，基于 Transformer 的行人轨迹预测方法能够显现出强大的长程依赖建模能力、并行计算能力、位置编码能力，以及具有多头自注意力机制等模型优点。但是，该方法同样也存在着对短序列的处理效果可能不如传统模型以及模型本身结构较为复杂等缺点。此外，基于 Transformer 的方法对大规模的数据集及模型依赖性较强，模型受限于正常的数据序列，很难将其推广到结构化数据（序列）中，例如图序列等，这也是该方法在该领域未来需要研究的问题。表5对基于 Transformer 的行人轨迹预测方法进行了汇总。

2.3 行人轨迹预测方法适用场景分析

在以上所综述的论文中，虽然涉及不同神经网络的行人轨迹预测方法，但是由于考虑的应用场景不同，它们在适用范围上也会存在一些不同之处，而并非所有方法都是针对无人驾驶安全性需求场景所产生的。考虑到基于数据驱动的行人轨迹预测方法是当前比较主流的轨迹预测方法，接下来将对以上涉及的基于数据驱动的行人轨迹预测方法适用的场景进行分析。

（1）基于 LSTM 的行人轨迹预测方法

该方法通常适用于需要考虑长时依赖关系的场景，例如行人在复杂交通情境中的移动场景。基于 LSTM 的行人轨迹预测模型可以对过去的历史轨迹数据进行建模，并使用该模型来预测未来的行人轨迹。由于 LSTM 可以处理序列数据的长时依赖关系，因此它在预测长期行人轨迹方面表现良好。然而，LSTM 模型具有较高的计算复杂度，因此可能

表5 基于Transformer的行人轨迹预测方法

方法	输入信息	网络结构	数据集
Saleh等人 ^[72]	历史轨迹、上下文信息	基于Transformer的编解码结构	ETH/UCY
STAR ^[71]	历史轨迹信息	基于图卷积(GCN)和Transformer结构结合	ETH/UCY
Giuliani等人 ^[20]	历史轨迹信息	基于Transformer的编解码结构	ETH/UCY、SDD
Yin等人 ^[73]	历史轨迹、上下文信息	基于Transformer的编解码结构	JAAD/PIE
Yao等人 ^[70]	历史轨迹信息	基于端到端的Transformer编解码结构	ETH/UCY
Li等人 ^[74]	历史轨迹信息、RGB图像信息、场景语义信息	基于图卷积网络和Transformer网络结合	ETH/UCY、VIRAT/ActEV
Su等人 ^[75]	历史轨迹信息、行人速度和加速度信息	基于交叉模态的Transformer网络架构	JAAD/PIE

不适合在实时性要求较高的无人驾驶安全性场景中使用。但是在视频监控及城市交通规划等场景中，该方法具有很好的适用性。

(2) 基于GAN的行人轨迹预测方法

该方法使用GAN进行对抗博弈来迫使网络生成具有真实性质的行人轨迹，可以很好地模拟真实世界中的行人运动。在无人驾驶场景中，行人的运动轨迹需要实时预测和更新，以便更好地规划自动驾驶车辆的行动路线。基于GAN的方法可以通过不断地学习和优化，实现实时的行人轨迹预测和更新。此外，行人的运动轨迹往往具有多样性，例如在交叉路口、拐角等场景下，行人的运动轨迹可能会发生较大的变化。基于GAN的方法可以生成多样化的轨迹，从而提高预测的适应性和灵活性。然而，GAN需要大量的数据来训练模型，但在无人驾驶场景下获取大规模的行人轨迹数据可能比较困难，特别是在复杂的城市环境中，需要考虑到行人与车辆的交互作用，以及道路拥堵等情况。因此，这也是该方法需要重点关注的一个问题。此外，GAN生成的轨迹是通过学习真实数据中的模式来生成的，因此可能会出现与真实世界不一致的情况，需要在实际无人驾驶应用场景中进行安全评估。

(3) 基于GCN的行人轨迹预测方法

基于GCN的行人轨迹预测方法适用于需要对行人未来行为进行预测的场景，其中最常见的是在智能交通、智能视频监控和自动驾驶场景。例如，在城市智能交通系统中，该方法可以利用车辆和行人的位置和移动信息来预测未来的交通流量和拥堵情况，从而帮助交通管理部门优化交通流。在公共场所，如机场、商场和公园等场景中，使用智能视频监控系统能够监测人群活动。基于GCN的方法能够预测行人未来的移动轨迹，以提前识别和预防可能出现的危险事件。在自动驾驶场景，自动驾驶车辆需要能够准确预测行人的行为，从而做出正确的

决策。基于GCN的方法能够准确预测行人未来的移动轨迹，帮助车辆识别行人可能的行为，避免潜在的交通事故。

(4) 基于Transformer的行人轨迹预测方法

基于Transformer的行人轨迹预测方法在处理长序列数据方面具有很好的性能，尤其在捕捉序列中的长距离依赖关系方面表现出色。因此，它适用于需要对行人行为进行长时间预测的场景，例如对人流密集的商场、火车站、机场和体育场馆等室内公共场所的人流管控。此外，该方法还适用于自动驾驶场景、机器人的导航与避障等场景。例如，通过利用Transformer模型对行人轨迹进行预测，可以帮助自动驾驶汽车预测行人的移动方向和速度，从而更好地规划车辆的行驶路线和速度。在机器人的导航和避障中，Transformer模型也可以用于预测行人的移动轨迹，帮助机器人更好地规划移动路线和避开行人。

除了以上适用场景范围之外，在面向无人驾驶安全性需求的行人轨迹预测场景中，可能会存在更加严格的要求和挑战。下面对相关的场景特点进行介绍。

- 复杂的行人运动：行人可能会在道路上快速穿行、交叉或折返，而这些运动往往是不可预测的。此外，行人在交通流中的位置和速度也可能受到各种因素的影响，例如人口密度、行人意图和环境因素等。

- 多样的场景环境：行人轨迹预测需要考虑不同的场景环境，例如城市道路、公共场所、停车场等。每个场景都可能有不同的行人流量、空间限制、路面条件等因素，这些因素可能会对行人的行动产生影响。

- 不同行人之间的交互：在行人轨迹预测中，不同行人之间的交互也需要考虑。例如，行人可能会相互避让或者穿过彼此的路径，这些行为会直接

影响行人的行进方向和速度。

- 数据集的标注质量：在许多行人轨迹预测数据集中，由于标注的困难性，可能会存在一定的标注误差。这些误差可能会影响模型的训练和预测效果，因此需要针对数据集中的标注问题进行特别的处理。

- 模型的鲁棒性：在真实的无人驾驶场景中，可能会存在各种各样的干扰因素，例如天气变化、道路状况等。这些因素可能会对行人的运动轨迹产生影响，从而影响预测结果。因此，需要设计出鲁棒性较强的模型来应对这些干扰因素。

总之，以上场景特点可能会使行人轨迹预测变得更加困难，增加了无人驾驶的难度，因此需要更加复杂和先进的算法来解决。例如，可以通过深度学习、计算机视觉和机器人学等技术相结合来提高预测精度和鲁棒性，以确保无人驾驶车辆在道路上的安全性。

3 数据集及性能比较

3.1 数据集介绍

数据集是神经网络模型训练或测试必不可少的组成部分。为了研究和推进行人轨迹预测算法的发展，许多主流的数据集应运而生。这些数据集涵盖了各种不同的场景和情况，例如室内和室外环境、不同时间段和不同人群等。本文接下来将介绍行人轨迹预测方法所涉及的数据集，它们被广泛应用于行人轨迹预测算法的研究和评估。

ETH^[29]和UCY^[45]数据集是开源的公共数据集，广泛用于行人轨迹的预测及评估。其中，ETH数据集包含ETH和Hotel两个场景；UCY数据集包含3个场景，分别是UNIV、ZARAE1和ZARAE2。以上5个场景视频中共有5 075帧，涉及超过1 600条行人运动轨迹，并每隔0.4 s对行人具体位置的坐标进行标注。每个场景都包括一个或多个视频序列，涵盖了不同的时间段、天气条件和人流量。每个视频序列都提供了高质量的注释数据，包括行人的位置、速度、方向和行走方式等。该数据集被广泛应用于计算机视觉和机器学习领域的研究，成为许多行人轨迹预测算法的基准数据集之一。

斯坦福无人机数据集（Stanford drone dataset, SDD）^[76]是一个用于行人轨迹预测与目标追踪的大规模数据集。该数据集由斯坦福大学计算机科学系开发，包括视频、图像和注释数据。它通过

无人机以鸟瞰图的角度获取斯坦福大学校园的一系列不同的场景，包括街道、公园、广场等。该数据集被分为训练集和测试集，其中测试集仅提供行人的初始位置和出现时间，用于预测智能体未来的运动轨迹。此外，数据集中提供了大学校园区域内大约19 000条agent轨迹，并且行人的位置每隔0.4 s被注释一次。该数据集中的标注信息包括目标类别（如行人、车辆、自行车等）、位置和尺寸等信息，可以用于训练目标检测和跟踪模型。SDD已经成为无人机视觉领域中广泛使用的数据集之一，对于研究无人机视频分析算法和应用具有重要意义。

ActEV/VIRAT^[77]是两个用于视频活动事件检测和识别的基准数据集。其中，ActEV数据集包含超过180 h的视频，共包括接近3 000个视频剪辑，涵盖了许多不同的场景和活动类型，包括室内和室外环境中的人类活动、车辆行驶、运动员比赛等。VIRAT数据集包括超过100 h的视频和相应的元数据和注释，包括时间戳、GPS位置、对象类别和行为注释等。该数据集涵盖了许多活动类型，如人类活动、车辆、船只、飞机等。它还包括一些任务，如目标跟踪、行为分析、事件检测等。这些数据集的发布推动了视频中活动识别和相关领域的研究发展，也为开发和评估视频分析算法提供了重要的基础。

Caltech Pedestrian数据集^[78]是一个用于行人检测和行人跟踪任务的数据集。该数据集由美国加州理工学院开发，通过车载相机获取实际场景中的行人图像。Caltech Pedestrian数据集包含大约10 h的高分辨率视频，视频中每个图像都被标注为行人或非行人，并且行人的历史位置也都被标记。此外，数据集还包含了行人运动轨迹的注释，使该数据集非常适合进行行人跟踪和行人轨迹预测的研究。

PETS 2009（performance evaluation of tracking and surveillance）^[79]数据集是用于计算机视觉中视频跟踪和监视性能评估的一个具有挑战性的数据集，该数据集由6个不同的视频序列组成，其中包括在室内和室外拍摄的行人和车辆等不同类型的目标。这些序列具有不同的挑战性，例如包含光照变化、目标重叠和部分遮挡等情况。该数据集还包含了手动标注的目标位置，以供跟踪算法评估使用。PETS 2009数据集是学术界和工业界广泛使用的数据集之一，用于评估和比较行人跟踪和监视算法的

性能。

CrowdHuman^[80]是一个大规模的自然场景下的行人检测数据集，由香港中文大学多位研究人员于2019年发布。该数据集包含超过15 000张图像和超过470 000个标注的行人框。这些图像是在各种自然场景中拍摄的，如街道、商场和火车站等。数据集对人群中的行人进行了精细的标注，包括对行人遮挡、多个行人之间的相互干扰等场景的标注。数据集的发布旨在推动计算机视觉领域对自然场景下行人检测算法的研究和发展。

随着传感器技术的不断发展，大量新型的数据集不断涌现。近些年来，最新出现的数据集中，InD (intersection drone dataset, InD)^[81]数据集是一个大规模的空中视频和图像数据集，用于目标检测和跟踪。数据集中视频时长10 h，共计约13 500条轨迹。数据集包括来自6个城市交叉路口的高分辨率图像和视频。每个交叉路口包含多个摄像机。由于拍摄的角度和距离不同，因此数据集具有丰富的场景内容。数据集中的目标类别包括行人、自行车、汽车、公交车和货车等。数据集中还包含各种天气条件下的场景，例如晴朗、阴天和雨天等。该数据集的发布推动了无人机自动驾驶技术的发展，为无人机应用提供了基础研究和评估工具。

JAAD^[82]和PIE^[83]数据集是用于研究交通场景中行人行为的两个大型公共自然数据集。JAAD数据集是一个包含了视频和图像序列的数据集，其中包括了各种不同的交通场景和行人行为，例如行人

穿越街道、十字路口等待和穿越斑马线等。该数据集包含了超过400个视频和100 000帧图像，其中每个视频长度约为5 min。此外，数据集还提供了基于视频中的车辆和行人的注释标签。PIE数据集是一个由北京理工大学提供的行人图像和视频数据集，其中包含超过40 000张图片和1 200个视频。该数据集涵盖了各种不同的行人行为，包括行走、奔跑、上下楼梯、穿行马路等。此外，数据集还提供了对每个图像和视频的详细注释，包括行人的位置、姿态、行为和属性等。以上两个数据集都提供了行人穿越马路意图的图像信息和注释信息，其中JAAD数据集更关注行人穿行意图信息，PIE数据集主要提供在马路旁所有人的注释，标识出谁可能会或可能不会试图在此车辆前面穿过马路。JAAD数据集和PIE数据集为行人行为识别和理解的研究提供了有价值的资源，并且被广泛应用于行人姿态估计和行人重识别等领域。

STCrowd^[84]是一个拥挤场景的大型多模态数据集，主要用于行人检测和目标跟踪任务。该数据集中共有219 000个行人，视频中平均每帧20人，在极度拥挤的场景下，每帧包括30人，并且每一帧图像具有不同程度的遮挡。该数据集采集了9个不同场景的数据，涵盖了不同的天气、光照条件和路况。凭借丰富的注释信息，STCrowd数据集能够适用于不同的任务，包括仅LiDAR、仅图像和基于传感器融合的行人检测与跟踪任务。表6呈现了数据集相关对比信息。

表6

数据集预览

数据集	Agent	场景数量	传感器	时长	位置
UCY ^[29]	行人	3	相机	29.5 min	校园、城市街道
ETH ^[45]	行人	2	相机	25 min	大学入口、人行道、酒店入口
PETS 2009 ^[79]	行人	10	相机	20 h	公共场所、室外场景
Caltech Pedestrian ^[78]	行人	7	车载相机	10 h	城市道路
SDD ^[76]	车辆、行人	20	相机	—	校园
JAAD ^[82]	行人	346	车载相机	240 h	城镇地区、城乡地区
ActEV/VIR ^[77]	行人、车辆、船只等	12	高清相机	280 h	路口场景、街道
Crowd Human ^[80]	行人	15 000	相机	—	城市路口、商场街道
InD ^[81]	车辆、行人、自行车	4	相机	10 h	城市路口
PIE ^[83]	行人、车辆	—	车载相机	6 h	城镇地区、城乡地区
STCrowd ^[84]	行人	9	激光雷达、相机	—	交通路口

3.2 评价指标

行人轨迹预测模型的评价指标主要用于估计未

来预测轨迹与真实轨迹的偏差，常用的评价指标概括如下。

• 平均位移误差 (average displacement error, ADE) 和最终位移误差 (final displacement error, FDE)。ADE 指每个预测位置和每个真值位置之间的平均欧氏距离差值, 代表了模型在轨迹上的平均预测性能。FDE 代表终点预测位置和终点真值位置之间的平均欧氏距离差值。以上两个指标的数值越小, 预测精度越高。相应的计算式如式 (1) 和式 (2) 所示:

$$ADE = \frac{\sum_{i=1}^{N_t} \sum_{t=T_{obs}+1}^{T_{pred}} \|o_t^i - o_t^{-i}\|_2}{N^t(T_{pred} - T_{obs})} \quad (1)$$

$$FDE = \frac{\sum_{i=1}^{N_t} \|o_{pred}^i - o_{pred}^{-i}\|_2}{N_t} \quad (2)$$

其中, N_t 代表时间 t 时刻目标轨迹中的行人总数, T_{obs} 是观测的时刻, T_{pred} 是预测的时刻。 o_t^i 和 o_t^{-i} 是真实坐标, o_{pred}^i 是真实轨迹, o_{pred}^{-i} 是预测轨迹。

• 最小平均位移误差 (min average displacement error, minADE_k) 和最小最终位移误差 (min final displacement error, minFDE_k)。它们是 ADE 和 FDE 的变体形式, 用于模型在生成多模态轨迹预测时评估模型性能。minADE_k 用于计算每个行人生成 k 条轨迹样本与地面真实轨迹之间的最小欧氏距离。minFDE_k 用于计算每个行人生成 k 条轨迹样本中最小终点预测位置和终点真值位置之间的平均欧氏距离差值。用式 (3) 和式 (4) 表示如下:

$$\min ADE_k = \frac{\sum_{i=1}^{N_t} \min_{k=1}^k \sum_{t=T_{obs}+1}^{T_{pred}} \|o_t^i - o_t^{-i}\|_2}{N^t(T_{pred} - T_{obs})} \quad (3)$$

$$\min FDE_k = \frac{\sum_{i=1}^{N_t} \min_{k=1}^k \|o_{pred}^i - o_{pred}^{-i}\|_2}{N_t} \quad (4)$$

3.3 性能比较

在以往的研究工作中, ETH/UCY 数据集被广泛应用于行人轨迹预测领域的性能测试。因此, 本文重点比较了基于该数据集的不同行人轨迹预测算法在 ADE 和 FDE 两个指标上的表现, 具体结果如图 3 和图 4 所示。通过分析图 3 和图 4 的趋势可以看出, 基于数据驱动的行人轨迹预测方法 (S-LSTM^[22]、Group-LSTM^[42]、SGAN^[21]、Social-STGCNN^[54]、SSAGCN^[61] 和 STAR^[71] 等) 能够实现一个较好的预测结果, 其预测精度远高于基于模型驱动的行人轨迹预测方法 (LSTM (linear)^[22])。基于 LSTM 的行人轨迹预测方法 (S-LSTM^[22]、

Group-LSTM^[42] 和 StarNet^[44] 等) 通过 LSTM 本身特有的网络结构, 能够处理序列数据和捕捉长期依赖关系, 因此能够获取较好的轨迹预测结果。值得注意的是, 所对比的方法中, SSAGCN^[61] 在 ADE 和 FDE 上取得了最好的性能指标。它之所以能够取得优异的性能, 是因为它不仅能够生成多条轨迹样本, 更重要的是模型将图卷积网络和软注意机制结合起来进行学习。图卷积网络能够对社交网络中的节点和边进行建模, 软注意机制可以动态地学习和调整节点和边的重要性。该模型同时考虑了行人之间的社交互动和行人与环境之间的场景互动, 从而实现了较好的行人轨迹预测效果。基于 GAN 的行人轨迹预测方法 (SGAN^[21]、Sophie^[46] 和 Social Way^[47] 等) 通过 GAN 的对抗博弈思想来促使网络生成更加真实的轨迹样本, 进而使行人轨迹预测的准确度有了很大的提升。此外, 基于 Transformer 的方法则是近年来迅速发展起来的一种新方法, 通过图中的趋势可以看出, 基于 Transformer 的模型 (STAR^[71] 等) 在处理行人密度较高的数据集场景 (Univ、ZARA1 和 ZARA2) 上有更好的表现。与 RNN 不同, Transformer 模型能够处理长期依赖关系, 并且能够进行大批量并行训练, 这也是该方法在行人轨迹预测中的亮点之一。总体而言, 基于数据驱动的轨迹预测方法借助于大规模数据集和深度神经网络的结构特征, 能够精确地预测行人未来的运动轨迹。

4 总结与展望

行人轨迹预测是计算机视觉和人工智能领域的研究热点, 目前已经在自动驾驶、智能交通、机器人导航、智慧城市以及智能安防等领域得到了广泛应用。近年来, 随着深度学习技术的不断发展, 行人轨迹预测技术也得到了广泛关注和研究。本文对近年来行人轨迹预测方法进行了研究, 以分类任务为切入点, 对目前已有工作中的行人轨迹预测方法进行了综述, 对轨迹预测中现存的关键问题进行了分析, 同时对不同方法的优缺点进行了总结与归纳。结合行人轨迹预测技术的发展趋势可以看出, 伴随着大量数据集平台的推陈出新, 基于数据驱动的行人轨迹预测算法能够实现相对于基于模型驱动方法更好的轨迹预测效果。未来, 随着技术的不断进步和人工智能算法的不断优化, 行人轨迹预测技术将呈现以下几个发展趋势。

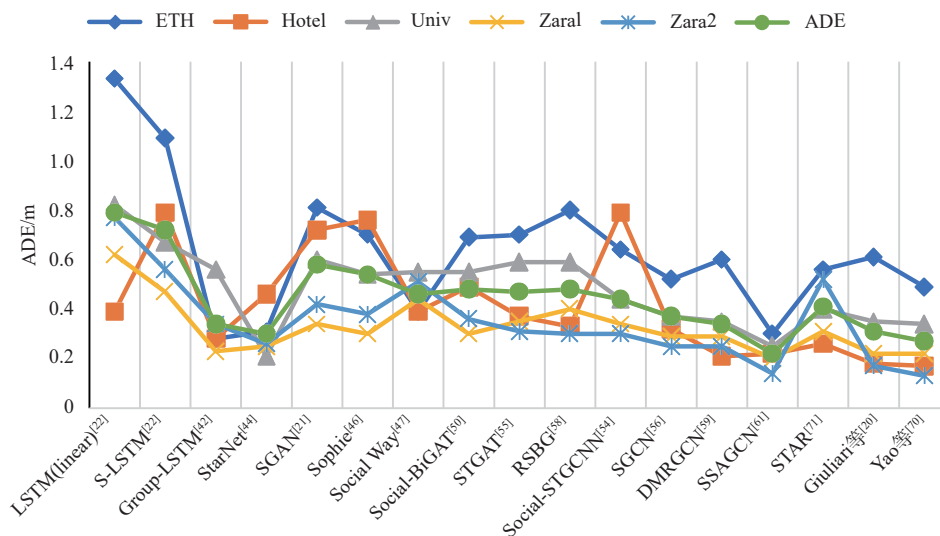


图3 ETH/UCY上的ADE对比

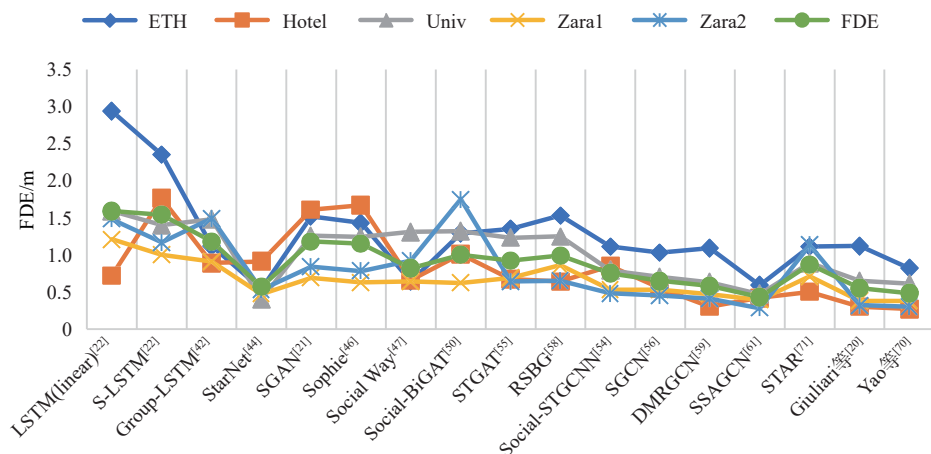


图4 ETH/UCY上的FDE对比

• 更高的预测准确度：随着深度学习算法的不断优化和硬件的提升，行人轨迹预测的准确度将不断提高。同时，越来越多的数据集和模型将被开发，以帮助改进行人轨迹预测算法的性能。

• 更高的智能化：未来行人轨迹预测技术将更加智能化。例如，基于大数据和人工智能算法，系统可以分析行人的历史行为，以预测未来的行为。这将为行人的行动提供更多的细节和上下文信息，并帮助系统更准确地预测其行动轨迹。

• 更广泛的应用场景：未来行人轨迹预测算法将在更广泛的应用场景中发挥作用。例如，在智慧城市领域，系统可以预测人群在不同时间和地点的分布，以帮助城市规划者更好地管理交通和城市设施。在机器人导航领域，行人轨迹预测算法可以帮助机器人更好地避让行人，进而使其更安全地与

人类交互。

• 更多的交互式应用：未来，行人轨迹预测将变得更加交互式。例如，在智能安防领域，系统可以根据行人的行为预测来警告安保人员；在游戏和虚拟现实领域，如在元宇宙中，行人轨迹预测可以用于创建更逼真的虚拟人物，并提高用户沉浸式体验。

总之，随着技术的不断发展和应用场景的不断扩大，行人轨迹预测方法将在未来得到更广泛的应用和更深度的发展，为人们的生产和生活带来更大的便利和安全。

参考文献：

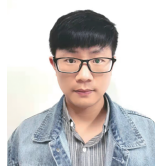
- [1] SHARMA N, DHIMAN C, INDU S. Pedestrian intention prediction for autonomous vehicles: a comprehensive survey[J]. Neurocomput-

- ing, 2022, 508: 120-152.
- [2] 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.
 - [3] RUDENKO A, PALMIERI L, HERMAN M, et al. Human motion trajectory prediction: a survey[J]. The International Journal of Robotics Research, 2020, 39(8): 895-935.
 - [4] RÖSMANN C, OELJEKLAUS M, HOFFMANN F, et al. Online trajectory prediction and planning for social robot navigation[C]//Proceedings of 2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM). Piscataway: IEEE Press, 2017: 1255-1260.
 - [5] BALLAN L, CASTALDO F, ALAHI A, et al. Knowledge transfer for scene-specific motion prediction[M]//Computer Vision - ECCV 2016. Cham: Springer International Publishing, 2016: 697-713.
 - [6] SIGHENCEA B I, STANCIU R I, CĂLEANU C D. A review of deep learning-based methods for pedestrian trajectory prediction[J]. Sensors, 2021, 21(22): 7543.
 - [7] 孔玮, 刘云, 李辉, 等. 基于深度学习的行人轨迹预测方法综述[J]. 控制与决策, 2021, 36(12): 2841-2850.
KONG W, LIU Y, LI H, et al. Survey of pedestrian trajectory prediction methods based on deep learning[J]. Control and Decision, 2021, 36(12): 2841-2850.
 - [8] KORBMACHER R, TORDEUX A. Review of pedestrian trajectory prediction methods: comparing deep learning and knowledge-based approaches[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 24126-24144.
 - [9] 陈敏, 曾凯, 沈韬, 等. 基于注意力机制和稀疏图卷积的行人轨迹预测[J]. 激光与光电子学进展, 2023, 60(10): 1010013.
CHEN M, ZENG K, SHEN T, et al. Pedestrian trajectory prediction based on attention mechanism and sparse graph convolution[J]. Laser and Optoelectronics Progress, 2023, 60(10): 1010013.
 - [10] XU Y, WANG L C, WANG Y Z, et al. Adaptive trajectory prediction via transferable GNN[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 6510-6521.
 - [11] 裴绍, 邱文涛, 王淼, 等. 基于Transformer动态场景信息生成对抗网络的行人轨迹预测方法[J]. 电子学报, 2022, 50(7): 1537-1547.
PEI Z, QIU W T, WANG M, et al. Pedestrian trajectory prediction method using dynamic scene information based transformer generative adversarial network[J]. Acta Electronica Sinica, 2022, 50(7): 1537-1547.
 - [12] HELBING D, MOLNÁR P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.
 - [13] YAN X. Modeling local behavior for predicting social interactions towards human tracking[J]. Pattern Recognition, 2014, 47(4): 1626-1641.
 - [14] HELBING D, FARKAS I, VICSEK T. Simulating dynamical features of escape panic[J]. Nature, 2000, 407(6803): 487-490.
 - [15] SCHÖLLER C, ARAVANTINOS V, LAY F, et al. What the constant velocity model can teach us about pedestrian motion prediction[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 1696-1703.
 - [16] FOX E, SUDDERTH E B, JORDAN M I, et al. Bayesian nonparametric inference of switching dynamic linear models[J]. IEEE Transactions on Signal Processing, 2011, 59(4): 1569-1585.
 - [17] KOIJ J F P, SCHNEIDER N, FLOHR F, et al. Context-based pedestrian path prediction[M]//Computer Vision - ECCV 2014. Cham: Springer International Publishing, 2014: 618-633.
 - [18] SCHNEIDER N, GAVRILA D M. Pedestrian path prediction with recursive Bayesian filters: a comparative study[M]//Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 174-183.
 - [19] DENDORFER P, ELFLEIN S, LEAL-ITAIXÉ L. MG-GAN: a multi-generator model preventing out-of-distribution samples in pedestrian trajectory prediction[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2022: 13138-13147.
 - [20] 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, 2021: 10335-10342.
 - [21] 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.
 - [22] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: human trajectory prediction in crowded spaces[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 961-971.
 - [23] 李琳辉, 周彬, 任威威, 等. 行人轨迹预测方法综述[J]. 智能科学与技术学报, 2021, 3(4): 399-411.
LI L H, ZHOU B, REN W W, et al. Review of pedestrian trajectory prediction methods[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(4): 399-411.
 - [24] SHI X D, SHAO X W, GUO Z L, et al. Pedestrian trajectory prediction in extremely crowded scenarios[J]. Sensors, 2019, 19(5): 1223.
 - [25] SEYFRIED A, STEFFEN B, KLINGSCH W, et al. The fundamental diagram of pedestrian movement revisited[J]. Journal of Statistical Mechanics: Theory and Experiment, 2005, 2005(10): 10002.
 - [26] KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proceedings of ICNN'95 - International Conference on Neural Networks. Piscataway: IEEE Press, 2002: 1942-1948.
 - [27] JIA H F, LIN Y, LUO Q Y, et al. Multi-objective optimization of urban road intersection signal timing based on particle swarm optimization algorithm[J]. Advances in Mechanical Engineering, 2019, 11(4): 168781401984249.
 - [28] MEHRAN R, OYAMA A, SHAH M. Abnormal crowd behavior detection using social force model[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 935-942.
 - [29] PELLEGRINI S, ESS A, SCHINDLER K, et al. You'll never walk alone: modeling social behavior for multi-target tracking[C]//Proceedings of 2009 IEEE 12th International Conference on Computer Vision. Piscataway: IEEE Press, 2010: 261-268.
 - [30] CHOI W, SAVARESE S. A unified framework for multi-target tracking and collective activity recognition[M]//Computer Vision - ECCV 2012. Heidelberg: Springer Berlin Heidelberg, 2012: 215-230.
 - [31] RUDENKO A, PALMIERI L, ARRAS K O. Joint long-term prediction of human motion using a planning-based social force approach[C]//Proceedings of 2018 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2018: 4571-4577.
 - [32] TRAUTMAN P, KRAUSE A. Unfreezing the robot: navigation in dense, interacting crowds[C]//Proceedings of 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2010: 797-803.

- [33] PELLEGRINI S, ESS A, VAN GOOL L. Improving data association by joint modeling of pedestrian trajectories and groupings[M]//Computer Vision - ECCV 2010. Heidelberg: Springer Berlin Heidelberg, 2010: 452-465.
- [34] KARAMOZAS I, HEIL P, VAN BEEK P, et al. A predictive collision avoidance model for pedestrian simulation[M]//Motion in Games. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009: 41-52.
- [35] ZHOU B L, WANG X G, TANG X O. Understanding collective crowd behaviors: learning a Mixture model of Dynamic pedestrian-Agents [C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012: 2871-2878.
- [36] FOX E, SUDDERTH E B, JORDAN M I, et al. Bayesian nonparametric inference of switching dynamic linear models[J]. IEEE Transactions on Signal Processing, 2011, 59(4): 1569-1585.
- [37] BEST G, FITCH R. Bayesian intention inference for trajectory prediction with an unknown goal destination[C]//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE Press, 2015: 5817-5823.
- [38] XIE D, SHU T M, TODOROVIC S, et al. Learning and inferring "dark matter" and predicting human intents and trajectories in videos[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(7): 1639-1652.
- [39] XUE H, HUYNH D Q, REYNOLDS M. SS-LSTM: a hierarchical LSTM model for pedestrian trajectory prediction[C]//Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE Press, 2018: 1186-1194.
- [40] MANH H, ALAGHBAND G. Scene- lstm: a model for human trajectory prediction[J]. arXiv preprint, 2018, arXiv:1808.04018.
- [41] HASAN I, SETTI F, TSESMELIS T, et al. MX-LSTM: mixing tracklets and vislets to jointly forecast trajectories and head poses[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6067-6076.
- [42] BISAGNO N, ZHANG B, CONCI N. Group LSTM: group trajectory prediction in crowded scenarios[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 213-225.
- [43] CHENG B, XU X, ZENG Y J, et al. Pedestrian trajectory prediction via the Social-Grid LSTM model[J]. The Journal of Engineering, 2018 (16): 1468-1474.
- [44] ZHU Y L, QIAN D H, REN D C, et al. StarNet: pedestrian trajectory prediction using deep neural network in star topology[C]//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE Press, 2020: 8075-8080.
- [45] LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example[J]. Computer Graphics Forum, 2007, 26(3): 655-664.
- [46] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. SoPhie: an attentive GAN for predicting paths compliant to social and physical constraints[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 1349-1358.
- [47] AMIRIAN J, HAYET J B, PETTRÉ J. Social ways: learning multi-modal distributions of pedestrian trajectories with GANs[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2020: 2964-2972.
- [48] HUANG L, ZHUANG J H, CHENG X M, et al. STI-GAN: multi-modal pedestrian trajectory prediction using spatiotemporal interactions and a generative adversarial network[J]. IEEE Access, 2021, 9: 50846-50856.
- [49] LAI W C, XIA Z X, LIN H S, et al. Trajectory prediction in heterogeneous environment via attended ecology embedding[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 202-210.
- [50] KOSARAJU V, SADEGHIAN A, MARTÍN-MARTÍN R, et al. Social-BiGAT: Multimodal Trajectory Forecasting using Bicycle-GAN and Graph Attention Networks[J]. Advances in Neural Information Processing Systems, 2019: 32.
- [51] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[J]. arXiv preprint, 2017, arXiv: 1710.10903.
- [52] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. New York: ACM, 2016: 2180-2188.
- [53] FANG F, ZHANG P P, ZHOU B, et al. Atten-GAN: pedestrian trajectory prediction with GAN based on attention mechanism[J]. Cognitive Computation, 2022, 14(6): 2296-2305.
- [54] MOHAMED A, QIAN K, ELHOSEINY M, et al. Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 14412-14420.
- [55] 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, 2020: 6271-6280.
- [56] SHI L S, WANG L, LONG C J, et al. SGCN: sparse graph convolution network for pedestrian trajectory prediction[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 8990-8999.
- [57] YAN S J, XIONG Y J, LIN D H. Spatial temporal graph convolutional networks for skeleton-based action recognition[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 12328.
- [58] SUN J H, JIANG Q H, LU C W. Recursive social behavior graph for trajectory prediction[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 657-666.
- [59] BAE I, JEON H G. Disentangled multi-relational graph convolutional network for pedestrian trajectory prediction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(2): 911-919.
- [60] CADENA P R G, QIAN Y Q, WANG C X, et al. Pedestrian graph: a fast pedestrian crossing prediction model based on graph convolutional networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 21050-21061.
- [61] LYU P, WANG W, WANG Y, et al. SSAGCN: social soft attention graph convolution network for pedestrian trajectory prediction[J]. arXiv preprint, 2021, arXiv:2112.02459.
- [62] ZHOU H, REN D C, XIA H X, et al. AST-GNN: an attention-based spatio-temporal graph neural network for Interaction-aware pedestrian trajectory prediction[J]. Neurocomputing, 2021, 445: 298-308.
- [63] 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.

- [64] 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.
- [65] RADFORD A, NARASIMHAN K. Improving language understanding by generative pre-training[Z]. 2018.
- [66] WANG A, SINGH A, MICHAEL J, et al. GLUE: a multi-task benchmark and analysis platform for natural language understanding[J]. arXiv preprint, 2018, arXiv:1804.07461.
- [67] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[M]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 213-229.
- [68] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale [J]. arXiv preprint, 2020, arXiv:2010.11929.
- [69] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 3141-3149.
- [70] YAO H Y, WAN W G, LI X. End-to-end pedestrian trajectory forecasting with transformer network[J]. ISPRS International Journal of Geo-Information, 2022, 11(1): 44.
- [71] YU C J, MA X, REN J W, et al. Spatio-temporal graph transformer networks for pedestrian trajectory prediction[M]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 507-523.
- [72] SALEH K. Pedestrian trajectory prediction using context-augmented transformer networks[J]. arXiv preprint, 2020, arXiv:2012.01757.
- [73] YIN Z, LIU R, XIONG Z, et al. Multimodal transformer network for pedestrian trajectory prediction [C]//IJCAI. [S. l.: s. n.], 2021: 1259-1265.
- [74] LI L H, PAGNUCCO M, SONG Y. Graph-based spatial transformer with memory replay for multi-future pedestrian trajectory prediction [C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 2221-2231.
- [75] SU Z X, HUANG G, ZHANG S Y, et al. Crossmodal transformer based generative framework for pedestrian trajectory prediction[C]//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2022: 2337-2343.
- [76] ROBICQUET A, SADEGHIAN A, ALAHI A, et al. Learning social etiquette: human trajectory understanding in crowded scenes[M]//Computer Vision - ECCV 2016. Cham: Springer International Publishing, 2016: 549-565.
- [77] OH S, HOOGS A, PERERA A, et al. A large-scale benchmark dataset for event recognition in surveillance video[C]//Proceedings of CVPR. Piscataway: IEEE Press, 2011: 3153-3160.
- [78] GRIFFIN G, HOLUB A, PERONA P. Caltech-256 object category dataset[Z]. 2007.
- [79] ELLIS A, FERRYMAN J. PETS2010 and PETS2009 evaluation of results using individual ground truthed single views[C]//Proceedings of 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway: IEEE Press, 2010: 135-142.
- [80] SHAO S, ZHAO Z, LI B, et al. Crowdhuman: a benchmark for detecting human in a crowd[J]. arXiv preprint, 2018, arXiv:1805.00123.
- [81] BOCK J, KRAJEWSKI R, MOERS T, et al. The InD dataset: a drone dataset of naturalistic road user trajectories at German intersections [C]//Proceedings of 2020 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2021: 1929-1934.
- [82] KOTSERUBA I, RASOULI A, TSOTSOS J K. Joint attention in autonomous driving (JAAD)[J]. arXiv preprint, 2016, arXiv:1609.04741.
- [83] RASOULI A, KOTSERUBA I, KUNIC T, et al. PIE: a large-scale dataset and models for pedestrian intention estimation and trajectory prediction[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2020: 6261-6270.
- [84] CONG P, ZHU X, QIAO F, et al. Stercrowd: a multimodal dataset for pedestrian perception in crowded scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 19608-19617.

[作者简介]



杜泉成 (1994-), 男, 北京科技大学计算机与通信工程学院博士生, 主要研究方向为轨迹预测和车辆规划决策。



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, 主要研究方向为社交网络分析、社交交通、网络运动组织和多智能体建模。



李灵犀 (1977-), 男, 博士, 美国印第安纳大学-普渡大学印第安纳波利斯联合分校电子与计算机工程系教授, 主要研究方向为复杂系统的建模、分析、控制与优化, 互联与自动化车辆, 智能交通系统, 智能车辆, 离散事件动态系统和人机交互。



宁焕生 (1975-), 男, 博士, 北京科技大学计算机与通信工程学院教授, 主要研究方向为网络空间及物联网研究。