

# 控制与决策

Control and Decision

## 基于深度学习的行人轨迹预测方法综述

孔玮, 刘云, 李辉, 王传旭, 崔雪红

引用本文:

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.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.1841>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 行人重识别中度量学习方法研究进展

A survey on metric learning in person re-identification

*控制与决策*. 2021, 36(7): 1547–1557 <https://doi.org/10.13195/j.kzyjc.2020.0801>

### 基于图卷积网络的行为识别方法综述

A survey of action recognition methods based on graph convolutional network

*控制与决策*. 2021, 36(7): 1537–1546 <https://doi.org/10.13195/j.kzyjc.2020.0514>

### 基于MCPDDPG的智能车辆路径规划方法及应用

The method and application of intelligent vehicle path planning based on MCPDDPG

*控制与决策*. 2021, 36(4): 835–846 <https://doi.org/10.13195/j.kzyjc.2019.0460>

### 一种基于深度学习的时间序列预测方法

A time series prediction method based on deep learning

*控制与决策*. 2021, 36(3): 645–652 <https://doi.org/10.13195/j.kzyjc.2019.0809>

### 考虑退化轨迹差异性与相似性的轴承RUL预测

Prediction of bearing remaining useful life involving difference and similarity of degradation trajectories

*控制与决策*. 2021, 36(11): 2832–2840 <https://doi.org/10.13195/j.kzyjc.2020.1028>

# 基于深度学习的行人轨迹预测方法综述

孔 玮, 刘 云, 李 辉<sup>†</sup>, 王传旭, 崔雪红

(青岛科技大学 信息科学技术学院, 山东 青岛 266061)

**摘 要:** 为了规划合理的路径以规避行人, 针对行人轨迹预测的研究具有广泛的应用价值. 基于手工特征的传统方法难以预测复杂场景下的行人轨迹. 深度学习以人工神经网络为架构, 具有强大的学习能力, 在各个领域取得了显著的效果. 基于深度学习的行人轨迹预测方法已逐渐发展为一种趋势. 为了宏观把握基于深度学习的行人轨迹预测的研究状况, 首先, 对不同方法进行组织与分类, 比较不同方法的优缺点, 讨论不同方法在行人轨迹预测领域的应用与发展; 其次, 根据行人轨迹预测模型的设计差异, 对比不同算法对模型性能产生的影响; 最后, 针对行人轨迹预测中存在的问题, 对基于深度学习的行人轨迹预测方法的未来发展进行了展望.

**关键词:** 行人轨迹; 轨迹预测; 神经网络; 预测方法; 深度学习; 综述

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.1841

开放科学(资源服务)标识码(OSID):



引用格式: 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.

## Survey of pedestrian trajectory prediction methods based on deep learning

KONG Wei, LIU Yun, LI Hui<sup>†</sup>, WANG Chuan-xu, CUI Xue-hong

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

**Abstract:** In order to plan a reasonable path to avoid pedestrians, the research on pedestrian trajectory prediction has a wide range of application value. Traditional methods based on manual features are difficult to predict pedestrian trajectory in complex scenes. Deep learning is based on artificial neural networks, which has strong learning ability and has achieved remarkable results in various fields. The pedestrian trajectory prediction method based on deep learning has gradually developed into a trend. In order to grasp the research status of pedestrian trajectory prediction based on deep learning, firstly, different methods are organized and classified, their advantages and disadvantages are compared, and the application and development of these methods in the field of pedestrian trajectory prediction are discussed. Then, according to the design differences of pedestrian trajectory prediction models, effects of different algorithms on the model performance are compared. Finally, in view of existing problems in pedestrian trajectory prediction, the future development of pedestrian trajectory prediction method based on deep learning is prospected.

**Keywords:** pedestrian trajectory; trajectory prediction; neural network; prediction methods; deep learning; survey

## 0 引言

随着社会自动化、智能化水平的提高, 现有的系统能够从视频中挖掘前所未有的丰富的视觉信息, 从而实现诸如事故避免和智能个人帮助等应用. 其中一个重要的分析是预测行人的未来路径. 例如, 自动驾驶汽车如果能精确预测周围行人的位置, 则可以避免交通事故; 有社交意识的机器人需要预测人类的轨迹, 以便优化它们的巡航路径; 用于城市规划的智能跟踪和监视系统必须了解人群将如何交互以更好地管理基础设施. 所以, 对于从自动驾驶汽车、社交机

器人的控制到安全监视等许多应用场景, 行人的未来轨迹预测已变得越来越重要, 在计算机视觉界受到广泛关注.

由于客观环境的影响, 行人之间、行人与环境的交互变得复杂抽象, 传统的模型基于手工函数不能模拟复杂情景的相互作用, 环境适应性差, 限制了模型的预测性能. 随着深度学习<sup>[1]</sup>的发展, 神经网络在语音识别、图像分类、物体检测等领域取得了重大成果, 其完备的理论体系和丰富的网络模型为深度学习应用到行人轨迹预测领域提供了必要的条件. 尤其

收稿日期: 2020-12-31; 修回日期: 2021-03-23.

基金项目: 国家自然科学基金项目(61702295, 61672305).

责任编辑: 薛建儒.

<sup>†</sup>通讯作者. E-mail: lihui@qust.edu.cn.

是,用于序列学习的递归神经网络(recurrent neural network, RNN)、生成对抗网络(generative adversarial networks, GAN)及图卷积网络(graph convolutional network, GCN)成为行人轨迹预测建模的主要网络.为了明确不同神经网络在行人轨迹预测中的应用与发展,需要对相关的文献进行综述和梳理.

文章的组织结构如下:第1部分简介了行人轨迹预测的发展和基于三大神经网络的行人轨迹预测方法;第2部分对比了基于不同神经网络的行人轨迹预测方法的优缺点和发展方向;第3部分设计了行人轨迹预测的通用模型,针对网络模型的关键差异,对方法进行归类;第4部分比较了不同算法对网络性能产生的影响;第5和第6部分阐述了目前行人轨迹预测方法中存在的问题以及未来的研究方向;最后一部分总结全文.

## 1 相关工作

### 1.1 行人轨迹预测

行人轨迹预测的核心任务是利用过去观察到的行人运动轨迹,通过建立模型,让机器学习由行为推理、与他人的交互、周围环境的影响等产生的一些规则<sup>[2]</sup>,理解复杂环境中的人类运动,从而在未来短时间内(例如5 s)预测行人的运动轨迹.

由于行人与行人、环境交互的复杂性和不确定性,行人的轨迹预测具有一定的难度.传统的方法通过社会力模型<sup>[3-6]</sup>、多模型方法和混合估计<sup>[7-8]</sup>对行人交互的研究取得了一些成果,但这些方法普适性差,存在固有的局限性.随着深度学习的发展,基于模式的方法可从数据中学习拟合不同的函数(如神经网络)来学习人的交互感知,提高了模型的灵活性和泛化性能.行人的运动轨迹不仅受周围行人的影响,还受到场景环境的干扰.基于场景交互的方法主要包括静态障碍物规避法<sup>[3,9]</sup>、地图感知法<sup>[10-11]</sup>以及语义图方法<sup>[12-18]</sup>,在一定程度上都可以提高行人轨迹的预测性能.

### 1.2 行人轨迹预测方法介绍

基于RNN的行人轨迹预测. RNN专为序列建模而设计,具备递归组织结构,在时间分析和序列学习方面显示出强大的建模能力.但随着时序序列长度的增加, RNN不能实现对数据状态的长期记忆,导致网络层停止学习.其变体长短期记忆网络(long short-term memory, LSTM)、门循环单元(gated recurrent unit, GRU)通过门控机制调节信息的流动与选择,解决了长序列的信息传递问题.在行人轨迹预测领域,很多方法都专注于RNN的研究并取得了显著的成

功<sup>[19-22]</sup>.但是,基于RNN的模型存在很多限制.在有效性方面, RNN模型容易出现梯度消失和梯度爆炸问题而难以处理.在运行效率方面,由于RNN模型当前时刻的状态依赖于上一时刻的隐藏状态,不能实现并行处理,导致模型的训练和推理速度都不高.

基于GAN的行人轨迹预测.行人轨迹预测的分布本质上具有多模态的属性,基于RNN的方法,在给定多个观测轨迹时,可能会被简单地处理为平均输出.深度生成模型可以学习数据分布并产生新的样本,比较流行的方法有变分自编码器和GAN网络.在行人轨迹预测中,基于变分自编码器的模型<sup>[23-25]</sup>一般不能建立行人之间的深层交互,并且概率计算复杂. GAN网络通过生成模型和判别模型之间的博弈,克服生成概率难以计算的困难.同时,借鉴GAN网络在超分辨率、图像转换和图像合成等领域的成功经验,轨迹采样器<sup>[26]</sup>通过将GAN输入随机向量与其他行人轨迹的隐藏表示相结合来处理所有观察到的行人之间的交互作用.但基于GAN模型的神经网络容易出现收敛速度慢、模式崩塌等问题.

基于GCN的行人轨迹预测. RNN虽然具备显著的序列建模能力,但缺乏直观的高层时空结构.在行人运动过程中,行人的数量不确定,行人之间的交互没有规律,图结构是表示行人间交互行为的自然方法,它比基于聚合的RNN方法更直观、更有效. GCN对非欧氏空间的图数据处理<sup>[27]</sup>非常有效,核心思想主要是通过深度学习方法,将图结构中的节点或边映射到向量空间,然后进行聚类、分类等处理.许多方法<sup>[28-30]</sup>都将图结构作为基本组成部分.时空图是比较流行的工具,因为它可以同时捕获空间和时间关系.这些方法通常将行人表示为节点,将它们的交互表示为连接,并结合诸如长短期记忆网络等深层序列模型进行建模<sup>[31-32]</sup>. GCN的应用,虽然提高了预测性能,但仍存在浅层网络、结构不稳定、自适应能力不强等问题.

## 2 行人轨迹预测的方法对比

### 2.1 基于RNN的预测方法

递归神经网络RNN将传入的序列数据编码为固定大小的隐藏表示,然后使用另一个RNN解码此隐藏表示,以生成顺序的时间表示输出.尽管LSTM具有学习和复制长序列的能力,但它们不能捕获多个相关序列之间的依赖关系.为此, S-LSTM<sup>[20]</sup>的社交池化层被提出,启发式聚集邻域行人的信息. S-LSTM是递归神经网络在拥挤场景中模拟人与人交互的首次应用,已成为行人轨迹预测的基线.基于RNN的行

人轨迹预测方法的对比如表 1 所示。

表 1 基于 RNN 的行人轨迹预测方法对比

方法	优点	缺点	研究方向	损失函数
SS-LSTM <sup>[19]</sup>	场景布局	交互不足	注意机制	均方误差
S-LSTM <sup>[20]</sup>	社交池化	局部交互	交互优化	负对数似然
ST-RNN <sup>[21]</sup>	时空递归	交互缺失	感知交互	负对数似然
SR-LSTM <sup>[22]</sup>	意图细化	数据偏差	强适应性	均方误差
DESIRE <sup>[25]</sup>	逆优控制	空间约束	社交联合	交叉熵+回归+重建+KL
S-Attention <sup>[31]</sup>	社交注意	维护整图	优化推广	负对数似然
Su 等 <sup>[33]</sup>	隐式数据	算力开销	泛化性能	负对数似然
MX-LSTM <sup>[34]</sup>	联合预测	推广面窄	丰富场景	负对数似然
Su 等 <sup>[35]</sup>	人群场景	泛化不强	深层分析	对数函数
Varshneya 等 <sup>[36]</sup>	扩展池化	参数繁多	多类对象	交叉熵+负对数似然
SNS-LSTM <sup>[37]</sup>	多模输入	注意平均	应用推广	负对数似然
Fernando 等 <sup>[38]</sup>	软硬注意	局部交互	领域扩展	均方误差

根据获取邻居信息的方式, 可把预测模型分为根据当前结果(速度、位置等)的模型<sup>[13,22,25-26,31,33]</sup>和根据先前状态的模型<sup>[13,20,25-26,33-37,39]</sup>。为了区分行人周围邻居的重要性, 一些基于 RNN 的行人轨迹预测方法利用注意力机制为邻居的权值计算提供了不同方式。比如, 根据隐藏状态计算软注意分数<sup>[13-26,28-31]</sup>, 计算成对速度相关<sup>[33,35]</sup>, 采用了“软注意”和“硬注意”组合的注意模型<sup>[38]</sup>等。基于 LSTM 的位置-速度-时间-注意模型<sup>[40]</sup>设计两个时间注意机制分别应用于位置和速度 LSTM 层的隐藏状态向量的计算。注意力机制的应用增强了行人之间的交互, 提高了轨迹预测的可信度。

基于 RNN 的模型在训练过程中存在梯度消失和不易扩展的问题, 许多预测工作<sup>[19,41-42]</sup>结合了卷积神经网络(convolutional neural networks, CNN), 证实了 CNN 在轨迹预测方面具有竞争性。然而, 它无法模拟行人间的空间交互。除此之外, 有些模型将 RNN 与其他结构相结合。例如, 结合高斯过程<sup>[33]</sup>预测行人未来轨迹的完整分布。由于未来的不确定性, CGNS<sup>[39]</sup>使用 GRU 从概率的角度进行推断, 并提出了可行假设。但是, 行人间的互动无法理解情景语境的影响, 对此, 一些模型结合场景上下文进行定义<sup>[13,19,36-37,43-45]</sup>。

### 2.2 基于 GAN 的预测方法

基于 GAN 的行人轨迹预测方法通过组合序列预测模型和 GAN 网络, 将对抗思想引入行人轨迹预测的任务中, 展现出所有可能解的样本空间。GAN 网络的应用克服了以往方法中大多基于优化行人之间的距离, 只预测一条“平均好”轨迹的不足。

在生成判别过程中, GAN 网络容易出现不可微的操作, 为了克服这种局限性, S-GAN 将社交池化扩展成多层感知机网络, 为行人预测了多种社交可接受

的轨迹。然而, S-GAN 在建模行人交互时不仅特征简单, 而且没有充分利用行人的深层交互信息。为此, 后续的方法通过利用注意力机制<sup>[13,46-48]</sup>、增加场景交互<sup>[13,49-51]</sup>、可行性约束<sup>[39]</sup>等方式, 综合行人交互模式, 探索行人轨迹影响因素。S-GAN 和 SoPhie 是具有高方差的单一行为模式, 受到社交行为方式的限制, 无法学习到行人真正的多模态分布。为此, 基于图结构生成对抗网络 Social-BiGAT<sup>[48]</sup>被提出, 构建学习多峰轨迹分布的生成模型。由于 GAN 网络容易发生模式崩溃和掉落, Social Ways<sup>[52]</sup>中的 Info-GAN 可以较好地改善多模态行人轨迹预测, 避免出现这些问题。基于 GAN 的方法通常是通过抽样一个潜在变量来处理未来的不确定性, 以往的研究很少对这些潜在变量进行深入细致的探讨。TPPO<sup>[53]</sup>却设计了一种潜在变量预测模型, 用于从观测和真实轨迹中估计潜在变量分布, 并取得了优异的预测性能。上面提出的行人轨迹预测方法主要是在二维图像空间学习和预测, 由于人的运动发生在三维世界中, 实验证明<sup>[54]</sup>在三维空间中学习和预测行人的轨迹更加有效, 这也为 GAN 网络在轨迹预测领域的推广和应用提供了新的方向。基于 GAN 的行人轨迹预测方法的对比如表 2 所示。其中: L2 是最小平方误差; KL 是相对熵, 又称为 KL 散度。

表 2 基于 GAN 的行人轨迹预测方法对比

方法	优点	缺点	研究方向	损失函数
SoPhie <sup>[13]</sup>	语义场景	手工排序	灵活性能	L2+ 对抗
S-GAN <sup>[26]</sup>	合理多样	采样较慢	丰富模式	L2+ 对抗
CGNS <sup>[39]</sup>	概率分布	复杂多变	广应用性	重建+KL+对抗
张志远等 <sup>[46]</sup>	特征丰富	泛化不足	场景交互	L2+ 对抗
孙亚圣等 <sup>[47]</sup>	社交注意	时间开销	信息融合	L2+ 对抗
STG-GAN <sup>[49]</sup>	全局交互	图间失联	优化扩展	L2+ 对抗
Sun 等 <sup>[51]</sup>	互反约束	拟合较慢	应用推广	对抗+互易
Social Ways <sup>[52]</sup>	多模分布	应用限制	决策优化	对抗+信息
TPPO <sup>[53]</sup>	潜在变量	不够稳健	改进控制	L2+ 对抗+KL
Zhong 等 <sup>[54]</sup>	三维空间	复杂度高	泛化性能	L2+ 对抗

### 2.3 基于 GCN 的预测方法

基于时空图的轨迹预测。基于图卷积网络在行为识别<sup>[55]</sup>、交通预测<sup>[56]</sup>、需求预测<sup>[57]</sup>等方面的广泛应用, 很多研究尝试将时空图<sup>[32,49,58-63]</sup>应用于行人轨迹预测任务, 并取得了不错的预测性能。预测任务中的时空图可以分为空间和时间两个维度, 空间维度建模目标行人与其邻居之间的交互, 而时间维度建模行人的历史轨迹。有些方法在此基础上进行拓展, 例如, 递归社交行为图<sup>[29]</sup>递归更新交互范围内的个体特征来提升交互关系。Zhang 等<sup>[30]</sup>在位置和速度方向上动态构建有向社交图, 有效捕捉行人的交互行为。STAR<sup>[63]</sup>基于 transformer 结构的增强注意力机

制处理图的时空建模. Liang 等<sup>[64]</sup>在空间图上设计 RNN, 编码行人运动模式的归纳偏差. 但是, 这些基于图网络的预测方法还没有多模态建模能力, 为此 Ivanovic 等<sup>[32,59]</sup>展示了一种高度多模态多人的场景, 保障了轨迹预测建模中的性能. 由于这些方法在每一个时间步中都会引入一个图, 它们能够处理在预测步骤之间发生变化的图. 然而, 这只是处理动态边的一种隐式能力, 不能显式地处理动态节点和边.

基于图注意力机制的轨迹预测. 行人的轨迹经常受周围行人及其障碍物(建筑物、人行道、草地等)的影响, 可能会改变或限制人类活动. 因此, 关注影响较大的因素, 对行人轨迹预测十分必要. 为了捕捉行人与邻居的交互作用, 一些方法根据行人之间的距离<sup>[19,26]</sup>计算行人的影响因子, 聚集建模的交互行为时, 通过池化、对称函数或者几何关系进行编码, 既不直观也不直接, 不能完整诠释行人之间的相互影响. 图注意力网络(graph attention networks, GAT)<sup>[65]</sup>利用软注意力或转移机制来区分邻居的重要性, 实现了节点之间有效的加权消息传递和更好的群体理解. 基于图注意力网络的轨迹预测<sup>[48,58,60-64]</sup>通过捕捉周围行人对目标行人的重要性, 打破了 RNN 网络的顺序依赖性, 为在共享空间中再现行人的拓扑结构提供了更直观的方法. 但现有的方法大多是基于轨迹的时空特性的研究, 空域信息表征单一, 忽略了行人运动过程中的深层运动特性, 比如运动速度、运动方向、运动状态等, 而这些多特征信息更符合真实场景中的行人运动状态, 更具有应用价值. 基于 GCN 的行人轨迹预测方法的对比见表 3. 表中 ADE 的解释见 4.2 节.

表 3 基于 GCN 的行人轨迹预测方法对比

方法	优点	缺点	研究方向	损失函数
RSBG <sup>[29]</sup>	监督递归	交互不足	场景交互	指数 L2
Zhang 等 <sup>[30]</sup>	时间潜码	不够稳定	文间关联	重建+KL
S-BiGAT <sup>[48]</sup>	多峰轨迹	特征不足	深层交互	对抗+分类+重建+L2+KL
STGAT <sup>[58]</sup>	序列 GAT	简化交互	监督注意	L2 损失
Ivanovic 等 <sup>[59]</sup>	多模多人	计算开销	建动态图	负对数似然
S-STGCNN <sup>[60]</sup>	图的应用	时域无关	应用拓展	负对数似然
GraphTCN <sup>[61]</sup>	门控适应	特征单一	灵活不足	ADE 损失
Haddad 等 <sup>[62]</sup>	动静交互	局部交互	类型推广	负对数似然
STAR <sup>[63]</sup>	新颖框架	场景缺失	应用推广	均方误差

### 3 行人轨迹预测的网络设计

基于深度学习背景下的行人轨迹预测网络模型的设计流程如图 1 所示. 通过观察图 1, 不同的预测模型设计的区别主要在于输入输出、编解码结构以及轨迹预测之后的判别与回归.

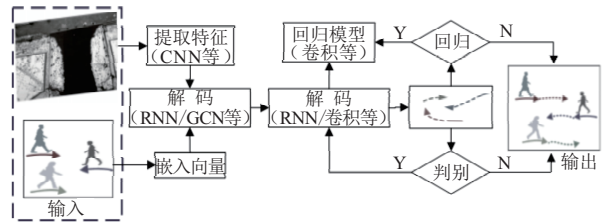


图 1 行人轨迹预测网络模型的设计流程

#### 3.1 按照输入输出特征区分

在预测行人轨迹时, 按照网络模型的不同输入, 行人轨迹预测模型分为场景输入模型<sup>[28,36,42,66]</sup>、轨迹输入模型<sup>[21-22,26,29-32,35,40,52-53,58,60-63,67-68]</sup>以及两者结合的输入模型<sup>[13,25,37,39,44-45,48,51,64,69]</sup>. 轨迹输入表示为不同行人在不同时刻的二维运动坐标, 场景输入包括场景图片、语义图、导航地图, 如图 2 所示, 还有经过处理的深度图<sup>[51]</sup>等.

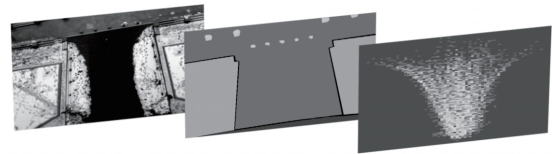


图 2 场景图、语义图和导航图

根据输出轨迹的多样性, 可以分为单一轨迹模型和多模态轨迹模型. 单一轨迹模型主要侧重于预测每个行人的单个未来轨迹. 然而, 当两个行人交互时, 有多种未来轨迹可以避免碰撞. 多模态就是指未来不同行为的可能性, 交互主体可以遵循几种可行的运动轨迹. 因此, 多模态轨迹模型是对可能轨迹的分布预测, 主要分为基于变分自动编码器的模型<sup>[25,32,39,61,67]</sup>和通过抽样潜在变量的生成对抗网络模型<sup>[26,39,48,51-53]</sup>. 除此之外, 还有引入噪声的方法<sup>[13,58,61,69]</sup>、生成热图的方法<sup>[45,66]</sup>、概率密度模型<sup>[64,70]</sup>以及时间随机模型<sup>[30]</sup>等.

#### 3.2 按照编解码结构区分

在编码阶段, 按照行人交互或者场景交互的建模结构区分, 可以把模型分为基于图结构的方法和非图结构的方法. 基于图结构的方法主要是应用 GCN 或者 GAT 来建模行人之间的交互. 基于 GCN 的方法<sup>[29,60,63,71]</sup>利用邻域聚合的思想融合信息, 但此方法无法捕捉周围行人对中心行人的分类贡献大小, 所以, 基于 GAT 的方法<sup>[48,58,60-62]</sup>提出了辨别邻节点贡献能力的注意力机制, 并取得了不错的实验效果. 非图结构的方法在建模行人与行人、场景交互时, 一般使用池化的方法<sup>[20,26,36-37,51-53,69,72]</sup>、卷积的方法和注意力的方法<sup>[13,31,40,68]</sup>. 基于池化的方法是建模行人与行人之间的交互, 为了体现场景(例如街道、障碍物、

建筑物)对行人轨迹的影响,在基于卷积的方法中,通常使用卷积神经网络CNN提取行人特征<sup>[25,28,66]</sup>或场景的特征<sup>[13,36,39,51,66]</sup>并进行特征融合.基于注意力的方法通常是根据行人之间的距离<sup>[13,40]</sup>、亲密度<sup>[31,73]</sup>、速度<sup>[40]</sup>、相对运动<sup>[68]</sup>等计算行人之间的相互作用.

在解码阶段,广泛应用的模型是RNN及其变体LSTM、GRU<sup>[25,39]</sup>,使用LSTM居多.但是,LSTM存在参数多、耗时长等问题,很多研究利用时间卷积网络TCN解码<sup>[60-61]</sup>,不仅缓解了基于RNN的序列预测的误差累积问题,而且比RNN体积小.另外,还有一些研究使用了反卷积<sup>[66]</sup>、全连接网络<sup>[63]</sup>、图注意力网络<sup>[64]</sup>进行解码,虽然应用比较少,但为以后的解码研究提供了新的思路和方法.

### 3.3 按照轨迹预测之后的判别与回归区分

轨迹预测后,为了减小与真实轨迹的差距,有些模型进行了轨迹的判别与回归.基于此,可把模型结构分为判别模型与回归模型.判别模型主要是基于GAN网络<sup>[13,26,39,48,51-53,74]</sup>的生成器,学习识别哪些样本(轨迹)是真实的以及生成了哪些样本(轨迹),从而改善生成器的性能.回归模型<sup>[25,28,45,64,66]</sup>通过结合场景特征、划分网格、设计精细化网络等方式进行预测轨迹的拟合与回归.例如,DESIR融合场景信息,从回归预测中接收迭代反馈.Scene-LSTM增加场景数据过滤器,结合场景特征调整目标轨迹.PIF将场景划分为曼哈顿网格并链接语义场景,计算分类和回归.卷积RNN解码器<sup>[64]</sup>回归计算网格的偏移向量.空间精化网络<sup>[66]</sup>对生成的热图分布进行位置回归.

## 4 行人轨迹预测的算法分析

### 4.1 数据集

在基于深度学习的行人轨迹预测方法中,涉及到的数据集主要有ETH/UCY、SDD、ActEV/VIRAT、Town Center、PETS09S2、Edinburgh、Interaction以及拥挤场景的数据集Osaki Station、Grand Central、CUHK.

ETH和UCY组合数据集是广泛使用的用于评估行人轨迹预测方法的公共数据集.数据集包含5个场景,2206条人体运动轨迹.详细介绍如表4所示.

表4 ETH/UCY数据集

场景	帧数	人数	分组数	障碍物数
ETH	1448	360	243	44
HOTEL	1168	390	326	25
UNIV	541	434	297	16
ZARA1	866	148	91	34
ZARA2	1052	204	140	34

数据集SDD是使用无人机在斯坦福大学校园上

空以俯视的角度拍摄的视频,涵盖了20个不同的场景,场景中既涉及动态对象,也包含静态障碍物和物理地标.ActEV/VIRAT是用于行为检测的公共数据集,涉及12个场景.视频总长超过12h,视频中含有丰富的注释,注释时长高达4.5h.Town Center(1个视频)和PETS09S2(3个视频)、Grand Central数据集(1个视频)最初是分别用于目标跟踪、人群行为分析的数据集,这3个数据集视频时间比较短,但含有非常多的社交互动.Edinburgh数据集是在爱丁堡大学信息论坛上采集的行人运动轨迹,采集时间高达数月,共有9.2万多条轨迹.Interaction数据集由无人机和车载激光雷达共同采集,观测轨迹通过视觉检测技术获取.

Osaki station数据集是使用二维激光传感器在日本东京大阪车站收集的轨迹数据,激光传感器的优势是即使在行人相互遮挡的情况下,也能较准确地获得行人的位置.CUHK数据集是在很多场景拍摄的不同密度、不同角度的人群数据集,还包括目标的真实轨迹、组群状态以及人群视频分类.

### 4.2 评价指标

评估模型性能的指标主要有两个:平均位移误差(ADE)和最终位移误差(FDE).具体来说,ADE评价的是模型沿轨迹的平均预测性能,而FDE只考虑最后的预测精度.两个指标的数值越小,网络的表现性能越优.两个指标的定义如下:

$$ADE = \frac{\sum_{i \in N} \sum_{t \in T_{end}} \|\hat{p}_i^t - p_i^t\|_2}{NT_{end}}, \quad (1)$$

$$FDE = \frac{\sum_{i \in N} \|\hat{p}_i^t - p_i^t\|_2}{N}, \quad t = T_{end}. \quad (2)$$

其中: $N$ 代表行人数量, $T_{end}$ 是最后的时刻, $\hat{p}_i^t$ 是预测轨迹, $p_i^t$ 是真实轨迹.另外,在ADE和FDE的基础上,还有一些评估指标,例如Best-of-N(BoN)、负对数似然(NLL)和平均非线性位移误差(NDE).BoN是从 $N$ 个随机抽样轨迹中得出的最低ADE和FDE.NLL测量了真值样本与预测分布的拟合度.NDE是非线性预测轨迹和真实轨迹所有位置的均方根误差.

### 4.3 算法性能比较

在行人轨迹预测领域,ETH/UCY数据集是被广泛应用的数据集,不同的行人轨迹预测算法在指标ADE和FDE上的性能如图3和图4所示.通过观察图3和图4的趋势可以发现,产生多条轨迹或多模态的不确定模型的性能要优于生成确定的单一轨迹的模型(星号标记).值得注意的是PECNet模型<sup>[67]</sup>,它之所以能够取得优异的性能,是因为它不仅密切模仿多模态人体轨迹规划过程,还能根据观察轨迹模拟行人

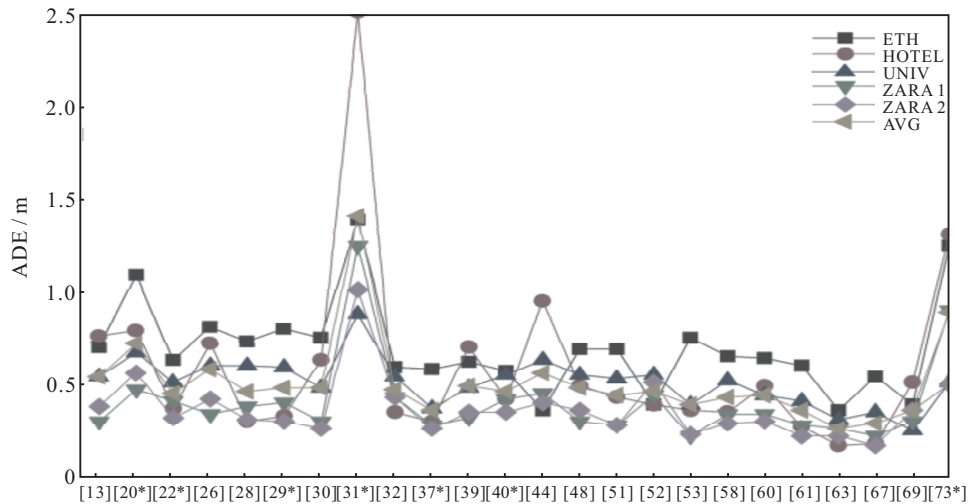


图3 ETH/UCY上的ADE对比

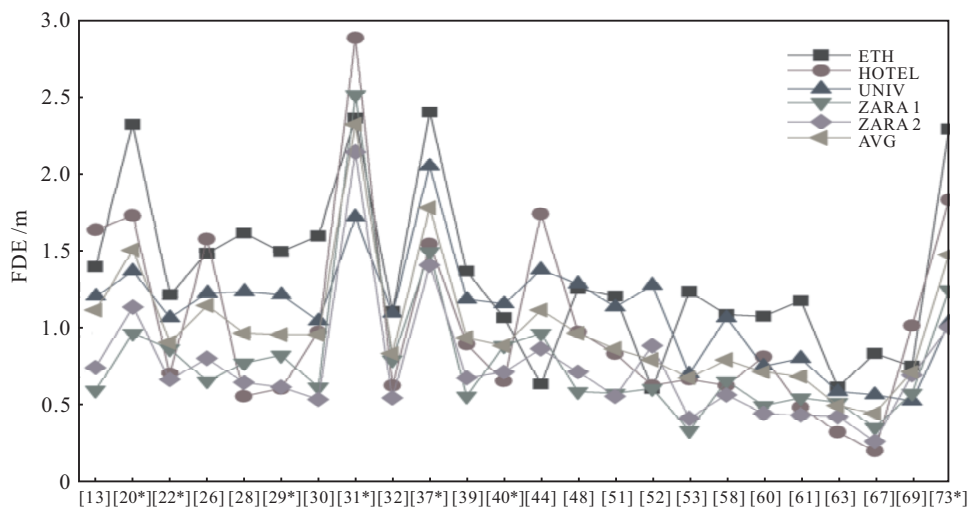


图4 ETH/UCY上的FDE对比

的可能终点. 目标导向的路径规划为基于多模态轨迹的研究开辟了新的道路. 时空图轨迹预测框架 STAR<sup>[63]</sup> 应用 transformer 模型摒弃了 RNN 的重复性, 实现长期依赖建模和大批量并行训练, 成为行人轨迹预测中的亮点. 其中, 基于 transformer 的空间图卷积 TGConv, 能够捕捉更复杂的社交互动, 倾向于在行人密度较高的数据集 (UNIV、ZARA1、ZARA2) 上进行更多改进. 然而, transformer 模型很难推广到复杂的结构, 如图结构. Social Ways<sup>[52]</sup> 和 TPPO<sup>[53]</sup> 依靠生成对抗网络的轨迹采样器提出合理的未来轨迹. GAN 网络自然地包含了行人转向决策的不确定性和潜在的多模态, 这为模型的更高层决策提供了重要指导. 另外, 观察图3和图4中的数据可以发现, 结合场景信息的模型, 例如 Scene LSTM<sup>[44]</sup> 和 Zhu 等<sup>[69]</sup> 能预测更加合理的轨迹, 因为高层场景的上下文 (如人体运动), 可以显著提高人体轨迹预测的准确性.

在 SDD 数据集上, PECNet 的 ADE/FDE 值为 9.96/15.88, 与 SimAug<sup>[75]</sup> (10.27/19.71)、CGNS (15.6/28.2)、

SoPhie (16.27/29.38)、DESIRE (19.25/34.05)、S-GAN (27.23/41.44)、S-LSTM (31.2/57.0) 相比取得了更好的性能, 比 S-GAN 的测试性能提高了 60% 左右, 强调了未来终点条件在轨迹预测中的重要性. 以上方法的测试数据都是在采样数量等于 20 的情况下的取值 (除了 DESIRE 取 5). 在 ActEV/VIRAT 数据集中, PIF<sup>[28]</sup> 中的多任务学习框架对在视频中执行未来路径和活动联合预测进行了探索性研究, 并首次展示了联合学习的优势. PIF 的 ADE 和 FDE 的值为 16.00/32.99, 与 S-GAN (21.16/38.05)、SimAug (21.7/42.2) 和 S-LSTM (23.10/44.27) 相比, ADE 降低 24 到 30 个百分点, FDE 降低 13 到 25 个百分点.

## 5 行人轨迹预测的问题与挑战

行人运动轨迹和模式受规则常识、相互作用、步态特征<sup>[76]</sup> 等影响, 研究中仍存在一些问题.

1) 交互缺乏可解释性. 网络模型在进行训练时, 使用的数据都是能客观测量的数据, 对行人运动意图把握不准, 缺少依赖于人的主观判断来训练算法的

数据. 例如,在自动驾驶场景中,自动驾驶辅助系统关心的是此时的行人是否会过马路,而不是行人是否会这样做. 有的模型利用头部姿势<sup>[34]</sup>,结合行人行为预测<sup>[28]</sup>进行了一些尝试,但获取数据的方式单一,行人的主观意图研究的少. 所以,目前的模型对计算到的交互缺乏可解释性,仍然依赖于数据驱动.

2) 动态图缺少时序相关. 基于图结构的网络架构,在时序构建动态图的过程中,对不同时刻的目标的相关信息缺少跟踪与更新. 换言之,模型在各个时间点能清楚地获得目标(例如障碍物)的位置,但目前算法没有对目标在时序上进行关联<sup>[31,60,62]</sup>,网络无法理解两个时刻目标的对应关系,降低了交互的性能,导致图网络结构不稳定.

3) 预测算法环境适应能力不强. 现有的社交感知方法假设所有被观察到的行人行为相似,并且他们的运动可以用相同的模型和特征来预测,对高层社会属性的捕捉和推理不强. 大多数模型都是针对特定的场景、任务或运动而设计. 这些方法在空间结构具体、运动模式固定时表现良好,例如,当环境中运动模式显著、空间结构和行人目标已知时,而在未定义的、不断变化的情况下性能较差.

4) 语义场景了解不够深入. 场景的语义理解就是让计算机如同人的大脑能正确理解自然场景及内容. 很多现有的模型或方法<sup>[37,42,44-45,64]</sup>对上下文信息的建模只局限于较小范围,只能学习局部特征,导致场景的语义理解受到阻碍.

## 6 行人轨迹预测的未来展望

1) 拓展获取数据途径,增强数据融合. 结合摄像头、激光雷达和其他硬件传感器的功能应用,级联完成数据融合. 通过软件识别算法或者面部识别技术,精确识别行人姿态,判断行人的主观意图,增强模型的可解释性能,从而使得模型的感知和决策更为有效.

2) 设置额外存储模块,强化时域约束. 为保证时序序列在长期预测中的时间一致性,可以通过设置额外的存储模块,对时域信息进行持续更新,实现预测期间连续地对嵌入进行平滑处理. 另外,结合注意力机制,对不同时间步的特征分配权重,重点关注关键时间点和关键信息,强化时间维度特征信息的利用.

3) 学习归纳多种算法,提升模型适应性. 重视迁移学习和方法推广,结合多预测框架的不同优点,实现更可靠的预测. 在新的环境下,学习归纳模型,挖掘和推理行人运动模式和避免碰撞的规则和规范. 同时,将预测与控制相结合,提高系统的整体鲁棒性.

4) 增强上下文特征分析,深入理解语义场景. 上

下文和周围环境的结合可以实现长期预测并有利于解释行人的运动意图. 而动静环境特征及其语义的上下文理解对于更好的轨迹预测仍然是一个相对未探索的领域. 对于场景理解而言,针对局部上下文的局限性等问题,需要充分利用全局的上下文信息和高级语义特征对场景内容进行描述. 同时,博弈论方法可能更适合于模拟人类行为.

## 7 结语

近几年来,基于深度学习的行人轨迹预测的研究取得了重大的进展. 但是,由于行人与行人、环境交互的复杂性和不确定性,行人的轨迹预测具有一定的难度,受到越来越多的关注. 本文主要总结了深度学习下的不同神经网络在行人轨迹预测领域的应用与发展. 行人轨迹预测的研究内容涉及自动驾驶、行为识别、目标跟踪、机器人导航等领域,不仅是计算机视觉界的一项重大挑战而且对人工智能的探索起到了推波助澜的作用.

## 参考文献(References)

- [1] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.  
(Zhang S, Gong Y H, Wang J J. The development of deep convolution neural network and its applications on computer vision[J]. Chinese Journal of Computers, 2019, 42(3): 453-482.)
- [2] 苏璐敏. 基于机器学习的轨迹预测方法研究[D]. 北京: 北京邮电大学, 2019: 1-65.  
(Su L M. Research on trajectory prediction method based on machine learning[D]. Beijing: Beijing University of Posts and Telecommunications, 2019: 1-65.)
- [3] Helbing D, Molnár P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.
- [4] Ferrer G, Sanfeliu A. Behavior estimation for a complete framework for human motion prediction in crowded environments[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Hong Kong: IEEE, 2014: 5940-5945.
- [5] Yan X, Kakadiaris I A, Shah S K. Modeling local behavior for predicting social interactions towards human tracking[J]. Pattern Recognition, 2014, 47(4): 1626-1641.
- [6] 叶汉荣, 刘梦源, 郑伟诗, 等. 基于惯常速率的多类目标轨迹预测[J]. 华中科技大学学报: 自然科学版, 2017, 45(10): 100-104.  
(Ye H R, Liu M Y, Zheng W S, et al. Trajectory prediction for multi-class target based on preferred speed[J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2017, 45(10): 100-104.)
- [7] Schneider N, Gavrila D M. Pedestrian path prediction with recursive bayesian filters: A comparative study[C].

- Proceedings of the German Conference on Pattern Recognition. Saarbrücken: Springer, 2013: 174-183.
- [8] Schulz A T, Stiefelhagen R. A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction[C]. Proceedings of the IEEE International Conference on Intelligent Transportation Systems. Gran Canaria: IEEE, 2015: 173-178.
- [9] Luo Y F, Cai P P, Hsu D, et al. Gamma: A general agent motion prediction model for autonomous driving[EB/OL]. (2019-10-23)[2021-01-23]. <https://arxiv.org/pdf/1906.01566v3.pdf>.
- [10] Rehder E, Kloeden H. Goal-directed pedestrian prediction[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops. Santiago: IEEE, 2015: 139-147.
- [11] Chen Y F, Liu M, Everett M, et al. Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017: 285-292.
- [12] Karasev V, Ayvaci A, Heisele B, et al. Intent-aware long-term prediction of pedestrian motion[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Stockholm: IEEE, 2016: 2543-2549.
- [13] Sadeghian A, Kosaraju V, Sadeghian A, et al. Sophie: An attentive gan for predicting paths compliant to social and physical constraints[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 1349-1358.
- [14] Sun L, Yan Z, Mellado S M, et al. 3D of pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data[C]. Proceedings of the IEEE International Conference on Robotics and Automation. Brisbane: IEEE, 2018: 5942-5948.
- [15] Molina S, Cielniak G, Krajník T, et al. Modelling and Predicting rhythmic flow patterns in dynamic environments[C]. Proceedings of the Annual Conference on Towards Autonomous Robotic Systems. Bristol: Springer, 2018: 135-146.
- [16] Kuhnt F, Schulz J, Schamm T, et al. Understanding interactions between traffic participants based on learned behaviors[C]. Proceedings of the IEEE Intelligent Vehicles Symposium. Gotenburg: IEEE, 2016: 1271-1278.
- [17] Keller C G, Gavrila D M. Will the pedestrian cross? A study on pedestrian path prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(2): 494-506.
- [18] Hong J, Sapp B, Philbin J. Rules of the road: Predicting driving behavior with a convolutional model of semantic interactions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 8454-8462.
- [19] Xue H, Huynh D Q, Reynolds M. Ss-lstm: A hierarchical lstm model for pedestrian trajectory prediction[C]. Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe: IEEE, 2018: 1186-1194.
- [20] Alahi A, Goel K, Ramanathan V, et al. Social lstm: Human trajectory prediction in crowded spaces[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE CS, 2016: 961-971.
- [21] Liu Q, Wu S, Wang L, et al. Predicting the next location: A recurrent model with spatial and temporal contexts[C]. Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix: AAAI, 2016: 194-200.
- [22] Zhang P, Ouyang W L, Zhang P F, et al. SR-LSTM: State refinement for lstm towards pedestrian trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 12077-12086.
- [23] Sohn K, Yan X C, Lee H. Learning structured output representation using deep conditional generative models[C]. Proceedings of the Annual Conference on Neural Information Processing Systems. Montreal: NIPS, 2015: 3483-3491.
- [24] Ivanovic B, Schmerling E, Leung K, et al. Generative modeling of multimodal multi-human behavior[C]. Proceedings of the IEEE International Conference on Intelligent Robots and Systems. Madrid: IEEE, 2018: 3088-3095.
- [25] Lee N, Choi W, Vernaza P, et al. Desire: Distant future prediction in dynamic scenes with interacting agents[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2165-2174.
- [26] Gupta A, Johnson J, Li F F, et al. Social gan: Socially acceptable trajectories with generative adversarial networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE CS, 2018: 2255-2264.
- [27] 孔玮, 刘云, 李辉, 等. 基于图卷积网络的行为识别方法综述[J]. 控制与决策, 2021, 36(7): 1537-1546. (Kong W, Liu Y, Li H, et al. A survey of action recognition methods based on graph convolutional network[J]. Control and Decision, 2021, 36(7): 1537-1546.)
- [28] Liang J W, Jiang L, Niebles J C, et al. Peeking into the future: predicting future person activities and locations in videos[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE CS, 2019: 2960-2963.
- [29] Sun J H, Jiang Q H, Lu C W. Recursive social behavior graph for trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE CS, 2020: 657-666.
- [30] Zhang L D, She Q, Guo P. Stochastic trajectory prediction with social graph network[EB/OL]. (2019-07-24)[2021-01-23]. <https://arxiv.org/pdf/1907.10233.pdf>.
- [31] Vemula A, Muelling K, Oh J. Social attention: Modeling attention in human crowds[C]. Proceedings of the IEEE

- International Conference on Robotics and Automation. Brisbane: IEEE, 2018: 4601-4607.
- [32] Ivanovic B, Pavone M. The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs[C]. Proceedings of the IEEE International Conference on Computer Vision. Seoul: CV/IEEE, 2019: 2375-2384.
- [33] Su H, Zhu J, Dong Y P. Forecast the plausible paths in crowd scenes[C]. Proceedings of the International Joint Conference on Artificial Intelligence. Melbourne, Australia, 2017: 2772-2778.
- [34] Hasan I, Setti F, Tsesmelis T, et al. MX-LSTM: Mixing tracklets and vislets to jointly forecast trajectories and head poses[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE CS, 2018: 6067-6076.
- [35] Su H, Dong Y P, Zhu J, et al. Crowd scene understanding with coherent recurrent neural networks[C]. Proceedings of the International Joint Conference on Artificial Intelligence. New York: IJCAI, 2016: 3469-3476.
- [36] Varshneya D, Srinivasaraghavan G. Human trajectory prediction using spatially aware deep attention models[EB/OL]. (2017-05-26)[2021-01-23]. <https://arxiv.org/pdf/1705.09436.pdf>.
- [37] Lisotto M, Coscia P, Ballan L. Social and scene-aware trajectory prediction in crowded spaces[C]. Proceedings of the IEEE International Conference on Computer Vision Workshop. Seoul: IEEE, 2019: 2567-2574.
- [38] Fernando T, Denman S, Sridharan S, et al. Soft+hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection[J]. Neural Networks, 2018, 108: 466-478.
- [39] Li J C, Ma H B, Tomizuka M. Conditional generative neural system for probabilistic trajectory prediction[EB/OL]. (2019-07-28)[2021-01-23]. <https://arxiv.org/pdf/1905.01631.pdf>.
- [40] Xue H, Huynh D Q, Reynolds M. A location-velocity-temporal attention lstm model for pedestrian trajectory prediction[J]. IEEE Access, 2020, 8: 44576-44589.
- [41] Nikhil N, Morris B T. Convolutional neural network for trajectory prediction[C]. Proceedings of the European Conference on Computer Vision. Munich: Springer, 2018: 186-196.
- [42] Chandra R, Bhattacharya U, Bera A, et al. Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 8475-8484.
- [43] Zhang W L, Sun L T, Wang X, et al. Seabig: A deep learning-based method for location prediction in pedestrian semantic trajectories[J]. IEEE Access, 2019, 7: 109054-109062.
- [44] Huynh M, Alaghband G. Trajectory prediction by coupling scene-lstm with human movement lstm[C]. Proceedings of the International Symposium on Visual Computing. Nevada: Springer, 2019: 244-259.
- [45] Manh H, Alaghband G. Scene-lstm: A model for human trajectory prediction[EB/OL]. (2018-08-12)[2021-01-23]. <https://arxiv.org/ftp/arxiv/papers/1808/1808.04018.pdf>.
- [46] 张志远, 刁英华. 结合社会特征和注意力的行人轨迹预测模型[J]. 西安电子科技大学学报, 2020, 47(1): 10-17.  
(Zhang Z Y, Diao Y H. Pedestrian trajectory prediction model with social features and attention[J]. Journal of Xidian University, 2020, 47(1): 10-17.)
- [47] 孙亚圣, 姜奇, 胡洁, 等. 基于注意力机制的行人轨迹预测生成模型[J]. 计算机应用, 2019, 39(3): 668-674.  
(Sun Y S, Jiang Q, Hu J, et al. Attention mechanism based pedestrian trajectory prediction generation model[J]. Journal of Computer Applications, 2019, 39(3): 668-674.)
- [48] Kosaraju V, Sadeghian A, Martín-Martín R, et al. Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks[C]. Proceedings of the Annual Conference on Neural Information Proceedingsessing Systems. Vancouver: NeurIPS, 2019: 1-10.
- [49] Zou X Y, Sun B, Zhao D, et al. Multi-modal pedestrian trajectory prediction for edge agents based on spatial-temporal graph[J]. IEEE Access, 2020, 8: 83321-83332.
- [50] Zhao T Y, Xu Y F, Monfort M, et al. Multi-agent tensor fusion for contextual trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 12118-12126.
- [51] Sun H, Zhao Z Q, He Z H. Reciprocal learning networks for human trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE CS, 2020: 7414-7423.
- [52] Amirian J, Hayet J B, Pettre J. Social ways: Learning multi-modal distributions of pedestrian trajectories with gans[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach: IEEE CS, 2019: 2964-2972.
- [53] Yang B, Yan G C, Wang P, et al. Tppo: A novel trajectory predictor with pseudo oracle[EB/OL]. (2020-02-04)[2021-01-23]. <https://arxiv.org/pdf/2002.01852.pdf>.
- [54] Zhong J Q, Sun H, Cao W M, et al. Pedestrian motion trajectory prediction with stereo-based 3d deep pose estimation and trajectory learning[J]. IEEE Access, 2020, 8: 23480-23486.
- [55] Yan S J, Xiong Y J, Lin D H. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018: 7444-7452.
- [56] Yao H X, Tang X F, Wei H, et al. Modeling spatial-temporal dynamics for traffic prediction[EB/OL]. (2018-11-03)[2021-01-23]. <https://arxiv.org/pdf/1803.01254.pdf>.

- [57] Yao H X, Wu F, Ke J T, et al. Deep multi-view spatial-temporal network for taxi demand prediction[C]. Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018: 2588-2595.
- [58] Huang Y F, Bi H K, Li Z X, et al. STGAG: Modeling spatial-temporal interactions for human trajectory prediction[C]. Proceedings of the IEEE International Conference on Computer Vision. Seoul: CV/IEEE, 2019: 6271-6280.
- [59] Ivanovic B, Schmerling E, Leung K, et al. Generative modeling of multimodal multi-human behavior[C]. Proceedings of the IEEE International Conference on Intelligent Robots and Systems. Madrid: IEEE, 2018: 3088-3095.
- [60] 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 the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE CS, 2020: 14412-14420.
- [61] Wang C X, Cai S F, Tan G. GraphTcn: Spatio-temporal interaction modeling for human trajectory prediction[EB/OL]. (2020-09-29)[2021-01-23]. <https://arxiv.org/pdf/2003.07167.pdf>.
- [62] Haddad S, Wu M Q, Wei H. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model[EB/OL]. (2019-02-13)[2021-01-23]. <https://arxiv.org/pdf/1902.05437.pdf>.
- [63] Yu C J, Ma X, Ren J W, et al. Spatio-temporal graph transformer networks for pedestrian trajectory Prediction[C]. Proceedings of the European Conference on Computer Vision. Glasgow: Springer, 2020: 507-523.
- [64] Liang J W, Jiang L, Murphy K, et al. The garden of forking paths: Towards multi-future trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual: IEEE CS, 2020: 10505-10515.
- [65] Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks[C]. Proceedings of the International Conference on Learning Representations. Vancouver: ICLR, 2018: 566-577.
- [66] Chiho C, Bariush B. Looking to relations for future trajectory forecast[C]. Proceedings of the IEEE International Conference on Computer Vision. Seoul: CV/IEEE, 2019: 921-930.
- [67] Mangalam K, Girase H, Agarwal S, et al. It is not the journey but the destination: Endpoint conditioned trajectory prediction[EB/OL]. (2020-07-18)[2021-01-23]. <https://arxiv.org/pdf/2004.02025.pdf>.
- [68] Shi X D, Shao X W, Guo Z L, et al. Pedestrian trajectory prediction in extremely crowded scenarios[J]. Sensors, 2019, 19(5): 1-18.
- [69] Zhu Y L, Ren D C, Fan M Y, et al. Robust trajectory forecasting for multiple intelligent agents in dynamic scene[EB/OL]. (2020-05-25)[2021-01-23]. <https://arxiv.org/pdf/2005.13133.pdf>.
- [70] 程媛, 迟荣华, 黄少滨, 等. 基于非参数密度估计的不确定轨迹预测方法[J]. 自动化学报, 2019, 45(4): 787-798.
- (Cheng Y, Chi R H, Huang S B, et al. Uncertain trajectory prediction method using non-parametric density estimation[J]. Acta Automatica Sinica, 2019, 45(4): 787-798.)
- [71] 周于涛, 吴华意, 成洪权, 等. 结合自注意力机制和结伴行为特征的行人轨迹预测模型[J]. 武汉大学学报: 信息科学版, 2020, 45(12): 1989-1996.
- (Zhou Y T, Wu H Y, Cheng H Q, et al. Pedestrian trajectory prediction model based on self-attention mechanism and group Behavior Characteristics[J]. Geomatics and Information Science of Wuhan University, 2020, 45(12): 1989-1996.)
- [72] 毛琳, 巩欣飞, 杨大伟, 等. 空时社交关系池化行人轨迹预测模型[J]. 计算机辅助设计与图形学学报, 2020, 32(12): 1918-1925.
- (Mao L, Gong X F, Yang D W, et al. Space-time social relationship pooling pedestrian trajectory prediction model[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(12): 1918-1925.)
- [73] Xu Y Y, Piao Z X, Gao S H. Encoding crowd interaction with deep neural network for pedestrian trajectory prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE CS, 2018: 5275-5284.
- [74] 李琳辉, 周彬, 连静, 等. 基于社会注意力机制的行人轨迹预测方法研究[J]. 通信学报, 2020, 41(6): 175-183.
- (Li L H, Zhou B, Lian J, et al. Research on pedestrian trajectory prediction method based on social attention mechanism[J]. Journal on Communications, 2020, 41(6): 175-183.)
- [75] Liang J W, Jiang L, Hauptmann A. Simaug: learning robust representations from simulation for trajectory prediction[EB/OL]. (2020-07-1)[2021-01-23]. <https://arxiv.org/pdf/2004.02022.pdf>.
- [76] 贲晔焯, 徐森, 王科俊. 行人步态的特征表达及识别综述[J]. 模式识别与人工智能, 2012, 25(1): 71-81.
- (Ben X Y, Xu S, Wang K J. Review on pedestrian gait feature expression and recognition[J]. Pattern Recognition and Artificial Intelligence, 2012, 25(1): 71-81.)

## 作者简介

孔玮(1986—), 女, 博士生, 从事计算机视觉的研究, E-mail: kongwei1902@126.com;

刘云(1962—), 男, 教授, 博士生导师, 从事数字计算机视觉、图像处理等研究, E-mail: lyun-1027@163.com;

李辉(1984—), 男, 副教授, 博士, 从事计算机视觉、深度学习的目标检测与跟踪等研究, E-mail: lihui@qust.edu.cn;

王传旭(1968—), 男, 教授, 博士, 从事计算机视觉、图像处理与模式识别等研究, E-mail: wangchuanxu\_qd@163.com;

崔雪红(1978—), 女, 副教授, 博士, 从事计算机视觉的研究, E-mail: cuixuehongzhe@163.com.

(责任编辑: 孙艺红)