

# 基于注意力机制和稀疏图卷积的行人轨迹预测

陈敏, 曾凯\*, 沈韬, 朱艳

昆明理工大学信息工程与自动化学院云南省计算机技术应用重点实验室, 云南 昆明 650500

**摘要** 行人轨迹预测能够有效降低行人轨迹突变造成的碰撞风险,在智能交通及监控系统等领域有着广泛应用。目前已有的研究大多利用无向图卷积网络对行人间的社会交互关系进行建模,这种方法缺少对行人隐藏状态关联性的考虑,容易产生行人间的冗余交互。针对这一问题,提出一种基于注意力机制和稀疏图卷积的行人轨迹预测模型(DASGCN),通过构建深度注意力机制,捕捉行人运动隐藏状态的关联性,从而准确地提取行人运动状态特征。进一步提出自调节稀疏方法,减小冗余信息带来的运动轨迹偏差,解决行人密集无向交互的问题。将所提模型在ETH和UCY数据集上进行验证,其平均位移误差(ADE)和最终位移误差(FDE)分别达到0.36和0.63。实验结果表明,DASGCN对行人轨迹的预测能力要优于传统算法。

**关键词** 图像处理; 深度注意力; 图卷积神经网络; 轨迹预测; 稀疏交互

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP213274

## Pedestrian Trajectory Prediction Based on Attention Mechanism and Sparse Graph Convolution

Chen Min, Zeng Kai\*, Shen Tao, Zhu Yan

Yunnan Key Laboratory of Computer Technologies Application, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, Yunnan, China

**Abstract** Pedestrian trajectory prediction can effectively reduce the collision risk caused by sudden changes in pedestrian trajectory, which has been widely used in intelligent transportation and monitoring systems. At present, most of the existing researches use undirected graph convolution network to model the social interaction between pedestrians. This method lacks the consideration of the relevance of the hidden state of pedestrians, and is prone to generate redundant interactions between pedestrians. To solve this problem, a pedestrian trajectory prediction model based on attention mechanism and sparse graph convolution (DASGCN) is proposed. By constructing a deep attention mechanism, the association of motion hiding states among pedestrians is captured, and the pedestrian motion state features are accurately extracted. Self-adjusting sparse method is further proposed to reduce the motion trajectory deviation caused by redundant information and solve the problem of dense and undirected pedestrian interaction. The proposed model was verified on ETH and UCY datasets, and the average displacement error (ADE) and final displacement error (FDE) reached 0.36 and 0.63 respectively. The experimental results show that DASGCN is superior to traditional algorithms in predicting pedestrian trajectory.

**Key words** image processing; deep attention; graph convolution neural network; trajectory prediction; sparse interaction

## 1 引言

行人轨迹预测能够有效降低行人轨迹变化造成的潜在碰撞事故风险,在智能交通系统及高级监控系统等领域有着广泛的应用<sup>[1-6]</sup>。行人轨迹预测的复杂性

来自于不同的社会行为,如与他人并行行走、群体内行走和避免碰撞等<sup>[7]</sup>。另一方面,运动随机性同样增加预测的复杂性。行人预期路径未知性往往放大行人轨迹预测的偏差。当前基于社会属性交互的行人轨迹预测研究仍存在以下两个问题:

收稿日期: 2021-12-17; 修回日期: 2022-01-18; 录用日期: 2022-02-28; 网络首发日期: 2022-03-18

基金项目: 国家自然科学基金(61971208)、云南省中青年学术技术带头人后备人才(2019HB005)、云南省万人计划青年拔尖人才(2018 73)、云南省重大科技专项(202002AB080001-8)

通信作者: \*zengkailink@sina.com

1) 目前主流方法通过行人的潜在状态捕捉其运动轨迹,进一步合并近邻行人潜在状态模拟行人间交互。社会池化<sup>[8-10]</sup>通过池化合并潜在状态,但这种方式平等对待邻里区域内的行人,无法重点突出不同行人间的关系。注意机制<sup>[7,11-12]</sup>能够突出相邻行人在轨迹预测中的不平等重要性,可在一定程度上缓解上述问题。然而,当前使用的注意机制仅简单地融合行人运动隐藏状态,无法深度模拟人与人之间的互动,缺少行人轨迹在隐藏状态内获取关联性的考虑。

2) 目前有关行人轨迹预测的研究主要在图卷积网络(GCN)上开展。GCN将行人交互关系嵌入图邻接矩阵中对行人间的社会交互进行有效建模。然而通过无向图建模行人间的社会关系,会导致密集交互问题即行人之间的冗余交互。Shi等<sup>[13]</sup>提出的稀疏有向图卷积能缓解冗余交互问题,但其需要固定参数进行阈值设定,无法适应不同场景的切换,从而缺乏对复杂场景中行人交互关系的动态捕捉。

为了克服上述的两个问题,本文提出一种基于注意力机制和稀疏图卷积的行人轨迹预测模型(DASGCN)。设计深度注意力机制捕捉行人社会关系的内在关联,从而获得行人隐藏状态的深层特征,该机制促使行人间的社会交互关系得到准确表达。进一步提出自调节稀疏方法,通过行人的密集交互关系自动调整参数,这能够减小冗余信息带来的运动轨迹偏差,更适应复杂场景下多行人交互关系的捕捉。

## 2 相关工作

本小节对轨迹预测在行人间交互的相关研究进行简要回顾。Social LSTM<sup>[8]</sup>是较早关注行人轨迹预测的深度模型之一,其通过循环网络捕捉行人运动状态,使用池机制聚合循环输出来预测轨迹。peeking into the future (PIF)<sup>[14]</sup>通过视觉特征和新的池机制扩展提高了预测精度。SR-LSTM<sup>[15]</sup>通过加权机制来衡量每个行人对其他人的贡献。Social-BiGAT<sup>[16]</sup>使用图注意力机制提取行人间交互作用的轨迹特征相对权重。Social GAN<sup>[9]</sup>基于行人轨迹遵循多模态分布的假设,将Social LSTM<sup>[8]</sup>扩展为基于递归神经网络(RNN)的生成模型。GI-GAN<sup>[17]</sup>基于生成对抗网络(GAN)及注意力模块获取行人个体运动信息和群体交互信息,从而预测行人未来轨迹。SoPhie<sup>[18]</sup>使用卷积神经网络(CNN)从整个场景中提取特征,对每个行人使用双向注意机制,将注意力输出与视觉CNN输出连接起来,最后使用基于长短期记忆神经网络(LSTM)的自动编码器生成模型生成未来轨迹。CGNS<sup>[19]</sup>在结构上与SoPhie相似,但其使用门控循环单元(GRU)捕捉行人轨迹特征,使得预测轨迹相比LSTM更加精确。综上所述,之前的大多数研究都使用循环网络搭建行人运动模型。

注意力模型近年来被广泛应用于图像处理、语音识

别和自然语言处理等不同类型任务中,其目的是从大量信息中选择出当前任务的关键信息。注意力模型对轨迹预测偏差有所改善。Vemula等<sup>[20]</sup>采用注意力模块为行人的邻居隐藏状态分配权重。朱江平等<sup>[21]</sup>表明注意力机制基于特征信息可以在不同维度获取上下文信息。Fernando等<sup>[22]</sup>提出组合注意力模块,该模块使用软注意力和硬连线注意力来捕获来自邻居的轨迹信息。Li等<sup>[23]</sup>采用时间域的注意力模型来衡量目标在不同时刻的位置,并对不同目标的时间间隔进行对齐。Social-BiGAT<sup>[16]</sup>使用图注意力机制提取行人间交互作用的轨迹特征相对权重。本研究设计深度关系注意力模块来衡量目标行人与周围行人交互关系的内在关联。

Social-BiGAT<sup>[16]</sup>依赖于图形注意力网络来模拟行人间的社交互动。Social-STGCNN<sup>[24]</sup>直接将行人轨迹建模为一个图,从图中提取空间和时间信息,并创建一个合适的嵌入,从嵌入操作来预测行人轨迹。STAR<sup>[25]</sup>通过时空图Transformer捕捉行人间交互信息,为轨迹预测任务提供一个简洁高效的解决方案。SGCN<sup>[13]</sup>在图卷积神经网络基础上提出稀疏定向相互作用和运动趋势模型,以建模行人交互。图卷积网络是用于图的机器学习的神经网络体系结构,其利用节点和图结构从本地图邻居聚合特征信息,并通过相邻矩阵图节点之间的消息传递来捕获图的依赖性。PointNet<sup>[26]</sup>采用图卷积根据邻域点的信息有效提取局部特征。在轨迹预测问题中,将场景中行人看作图上的节点,而行人间复杂的交互作用看作边,行人的运动特征决定边的权值。本研究采用自调节稀疏方法去掉冗余的行人交互信息,其涉及的图卷积网络根据稀疏邻接矩阵对节点进行聚合。

## 3 模型

### 3.1 行人轨迹预测问题描述

行人轨迹预测根据行人前几帧的历史轨迹预测出后几帧的轨迹,行人位置用2D笛卡尔坐标表示,即 $(x_n^t, y_n^t)$ 。历史轨迹和预测轨迹的表达式分别为

$$t_n^o = \left\{ (x_n^t, y_n^t) \mid t \in (1, 2, \dots, T_{\text{obs}}) \right\}, \quad (1)$$

$$t_n^p = \left\{ (x_n^t, y_n^t) \mid t \in (T_{\text{obs}} + 1, T_{\text{obs}} + 2, \dots, T_{\text{pred}}) \right\} \quad (2)$$

式中: $t_n^o$ 表示观察轨迹; $t_n^p$ 表示预测轨迹; $n$ 表示行人; $T_{\text{obs}}$ 表示观察时间帧; $T_{\text{pred}}$ 表示预测时间帧; $t$ 表示时间。

### 3.2 模型架构

图1为所提模型总体框架图。首先,利用行人轨迹构建时间图和空间图作为输入,通过深度关系注意力机制得到时间和空间的嵌入向量以及时间和空间的密集交互图,以此获取行人间隐藏状态内的关联性信息;其次,通过自调节稀疏模块可以得到时间和空间的稀疏矩阵,其能够有效表示行人交互关系,从而去掉冗余的行人交互信息,同时嵌入向量和关系交互图通过

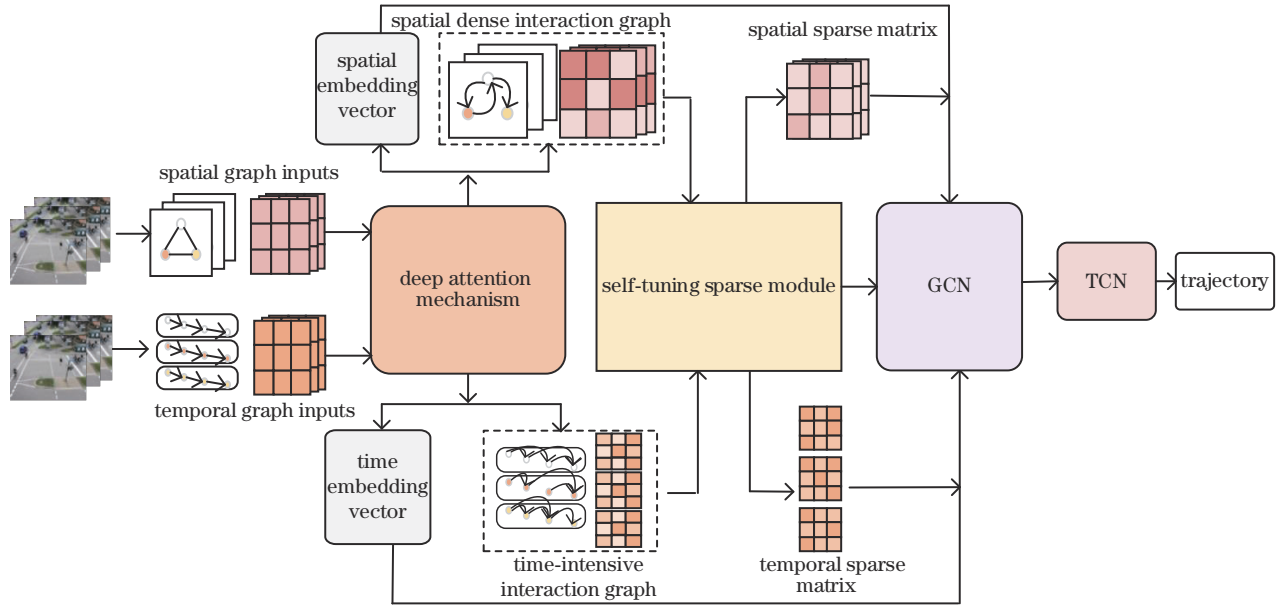


图 1 所提模型总体框架

Fig. 1 Overall framework of proposed model

图卷积后可以得到行人的运动轨迹特征表示;最后,通过时间神经网络中的估计双高斯分布参数对行人的未来轨迹进行预测。

### 3.3 运动信息编码

受 SGCN<sup>[13]</sup> 启发,实际场景中行人行走的趋势不仅受周围相邻行人影响还与行人自身运动趋势有关系,空间图建模行人与行人的交互,时间图表示行人运动趋势。因此,将轨迹以时间和空间进行建模比单纯进行空间建模的效果好。本研究将行人运动轨迹构造成时间图  $G_t$  和空间图  $G_s$ ,定义时间图  $G_s = (V_s, E_s)$ ,空间图  $G_t = (V_t, E_t)$ ,在图上进行信息传递。

$$V_s = \{v_i | i \in \{1, \dots, N\}\}, \quad (3)$$

$$E_s = \{e_s^{ij} | i, j \in \{1, \dots, N\}\}, \quad (4)$$

$$V_t = \{v_t | t \in \{1, \dots, T_{\text{obs}}\}\}, \quad (5)$$

$$E_t = \{e_t^{ij} | i, j \in \{1, \dots, T_{\text{obs}}\}\}, \quad (6)$$

式中:  $V_s$  表示空间图  $G_s$  的节点即行人;  $E_s$  表示空间  $G_s$  的边即行人与行人之间的关系;  $V_t$  表示时间图  $G_t$  的节点即行人行走的时间帧;  $E_t$  表示时间图  $G_t$  的边即当前行人的帧与帧之间的关系。

### 3.4 深度注意力机制

深度注意力机制是在自注意力机制的基础上改进而来的,其作用主要包括两点:1)获取当前行人与周围行人的交互关系信息,交互关系程度影响当前行人下一帧行走的方向;2)获取当前行人前一帧与后一帧之间的信息关系,由帧之间的信息可以得到行人的运动趋势。

深度注意力机制的主要结构如图 2 所示,由时间图和空间图的输入分别得到相应的嵌入向量,从嵌入向量中获取行人间的交互关系与行人时间帧之间的信

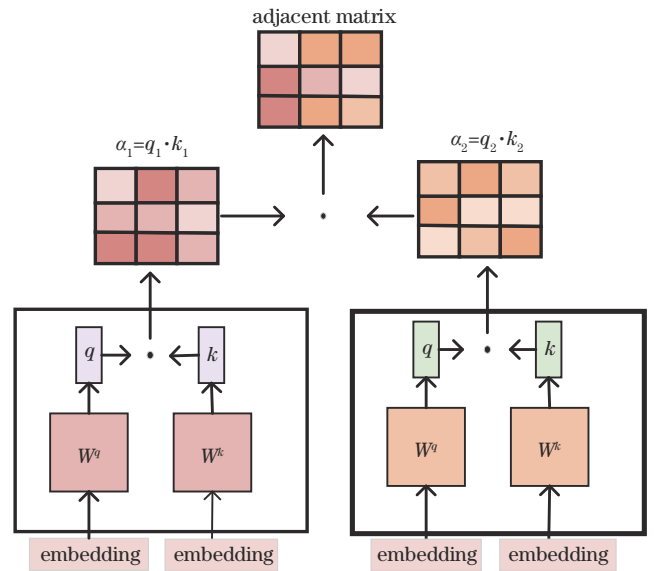


图 2 深度注意力机制

Fig. 2 Deep attention mechanism

息表示特征,最后通过两层自注意力机制获取行人交互之间的关联信息。具体公式如下:

$$\begin{cases} \mathbf{E} = \phi(\mathbf{G}, \mathbf{W}_E) \\ \mathbf{Q} = \phi(\mathbf{E}, \mathbf{W}_Q), \\ \mathbf{K} = \phi(\mathbf{E}, \mathbf{W}_K) \end{cases} \quad (7)$$

$$\mathbf{R} = \frac{(\mathbf{Q}_1 \mathbf{K}_1^T)(\mathbf{Q}_2 \mathbf{K}_2^T)}{\sqrt{d}}, \quad (8)$$

$$\mathbf{R}' = \text{Softmax} \left[ \mathbf{R} \frac{(\mathbf{Q}_3 \mathbf{K}_3^T)}{\sqrt{d}} \right], \quad (9)$$

式中:  $\phi(\cdot, \cdot)$  表示线性变化;  $\mathbf{E}$  表示图嵌入;  $\mathbf{Q}$  和  $\mathbf{K}$  表示自注意力机制的查询与关键字;  $\mathbf{W}_E, \mathbf{W}_Q, \mathbf{W}_K$  表示线



性变化的权重矩阵;  $d$  是保持数值稳定的比例因子。

### 3.5 自调节稀疏模块

当前行人受周围人影响的关系为交互关系。在模型中, 周围行人出现在定义范围内就会被纳入当前行人的交互关系中。所提自调节稀疏模块能合适选取行人间的交互关系和行人在时间帧之间的信息关系, 该模块主要根据行人密集交互矩阵计算出合适阈值, 通过阈值筛选出行人之间冗余的交互信息。因此, 对行人运动轨迹的预测不会出现冗余信息带来的轨迹预测偏差。相应公式如下:

$$\lambda = \text{mean}\left(\frac{1}{1+x^2}\right), \quad (10)$$

$$\mathbf{M} = \Pi\{\sigma(\mathbf{F}) \geq \lambda\} \quad (11)$$

$$\mathbf{A} = (\mathbf{M} + \mathbf{I}) \odot \mathbf{R}^{s-1} \quad (12)$$

式中:  $\lambda$  表示稀疏阈值;  $\text{mean}(\cdot)$  表示平均计算;  $x$  表示密集矩阵元素;  $\mathbf{M}$  表示稀疏掩码;  $\Pi$  表示过滤操作;  $\sigma$  表示 Sigmoid 激活函数;  $\mathbf{F}$  指行人间密集交互特征;  $\mathbf{I}$  表示单位矩阵;  $\odot$  表示元素计算, 对邻域内的元素值进行加权平均并重新赋值;  $\mathbf{R}^{s-1}$  表示邻接矩阵;  $\mathbf{A}$  表示稀疏邻接矩阵。

### 3.6 轨迹预测模块

时间卷积神经网络(TCN)以堆叠的顺序数据作为输入, 并将预测序列作为一个整体。这可以改善循环神经网络在连续预测中累积错误的问题, 而且时间神经网络更加节省计算机算力资源。受时间卷积神经网络的启发, 可由图卷积网络得到行人轨迹特征表示。本研究采用时间卷积神经网络来预测行人未来轨迹, 选择时间卷积网络的原因在于其能解决传统循环卷积神经网络中存在的梯度消失和计算成本高的问题。Social LSTM 将负对数似然损失最小化来进行训练, 所提方法也选用负对数似然损失函数。

$$f(\mathbf{V}^l, \mathbf{A}) = \sigma(\hat{\mathbf{A}}_l \mathbf{V}^l \mathbf{W}^l), \quad (13)$$

$$f(\mathbf{V}^l, \mathbf{A}) = \sigma(\hat{\mathbf{A}}_l \mathbf{V}^l \mathbf{W}^l), \quad (14)$$

式中:  $\mathbf{W}^l$  表示第  $l$  层的训练参数权重矩阵;  $\mathbf{V}^l$  表示第  $l$  层的行人节点;  $\sigma$  表示激活函数。负对数似然损失的表达式为

$$L^n(W) = - \sum_{t=T_{\text{obs}}+1}^{T_{\text{pred}}} \log P[(x_n^t, y_n^t) | \hat{\mu}_n^t, \hat{\sigma}_n^t, \hat{\rho}_n^t], \quad (15)$$

式中: 行人  $n$  在时间  $t$  的坐标  $(x_n^t, y_n^t)$  遵循双变量高斯分布  $N(\hat{\mu}_n^t, \hat{\sigma}_n^t, \hat{\rho}_n^t)$ ;  $\hat{\mu}_n^t$  表示平均数;  $\hat{\sigma}_n^t$  表示标准方差;  $\hat{\rho}_n^t$  表示相关系数;  $W$  表示模型可训练的参数;  $P[\cdot]$  表示概率值。

## 4 实验和分析

### 4.1 评估数据集

为了验证所提方法的有效性, 选用广泛应用的公共数据集 ETH<sup>[27]</sup> 和 UCY<sup>[12]</sup> 进行实验。其中, ETH 数

据集包含 ETH 和 HOTEL 场景, 而 UCY 数据集包含 UNIV、ZARA1 和 ZARA2 等 3 个不同的场景。观察 8 帧(3.2 s)的轨迹, 并预测接下来的 12 帧(4.8 s)的轨迹。

在实验中, 注意机制的嵌入维数和图嵌入维数均设置为 64, 注意机制层数为 2。采用 PReLU 作为非线性激活函数。采用 Adam 优化器对该方法进行 300 个 epoch 的训练, 数据批大小为 128。初始学习率设置为 0.001, 衰减系数为 0.1。

### 4.2 评价指标

采用平均位移误差(ADE)和最终位移误差(FDE)两个指标来评价预测结果。ADE 表示预测轨迹点与真实未来轨迹点之间的平均距离; FDE 表示最终预测目的地与实际未来轨迹点目的地之间的距离:

$$E_{\text{ADE}} = \frac{\sum_{n \in N} \sum_{t \in T_p} \|\hat{\mathbf{p}}_i^n - \mathbf{p}_i^n\|_n}{N \times T_p}, \quad (16)$$

$$E_{\text{FDE}} = \frac{\sum_{n \in N} \|\hat{\mathbf{p}}_i^n - \mathbf{p}_i^n\|_n}{N}, \quad t = T_p, \quad (17)$$

式中:  $N$  为行人的个数;  $T_p$  为预测帧数;  $\hat{\mathbf{p}}_i^n$  和  $\mathbf{p}_i^n$  分别为时间步长为  $t$  时的预测结果和实际结果。

### 4.3 定量分析

为了评估所提算法的准确性, 将其与多种算法进行对比, 分别是 Social LSTM<sup>[8]</sup>、SGAN<sup>[9]</sup>、SoPhie<sup>[18]</sup>、PITF<sup>[14]</sup>、GAT<sup>[11]</sup>、Social-BIGAT<sup>[16]</sup>、Social-STGCNN<sup>[24]</sup>、STAR<sup>[25]</sup> 及 SGCN<sup>[13]</sup>, 上述所有算法均抽取 20 个样本。所提算法与其他算法在 5 个场景上的预测精度比较结果如表 1 所示, 其中, 越小的数据越接近真实轨迹, 黑色加粗的表示最好的预测结果。

由表 1 可知, 所提算法在 HOTEL、ZARA1 和 ZARA2 这 3 个场景上的两个指标优于其他算法。在 HOTEL 场景上, 所提模型的 ADE 和 FDE 比 SGCN 分别下降 18.8% 和 23.6%; 在 ZARA1 场景上, 所提模型的 ADE 和 FDE 比 SGCN 分别下降 3.5% 和 7.4%; 在 ZARA2 场景上, 所提模型的 ADE 和 FDE 比 SGCN 分别下降 8% 和 6.7%; 在 UNIV 场景上的两个指标与 SGCN 一样优于其他模型。5 个场景的平均 ADE 和 FDE 表明所提模型性能最好。所提模型使用深度注意力模块获得隐藏状态的关联, 加强行人之间的深层交互, 利用自调节的有向稀疏图来过滤掉冗余交互信息, 避免冗余信息形成预测偏差, 从而在数据集中表现出良好的鲁棒性。所提模型在 HOTEL、ZARA1、ZARA2 和 UNIV 等 4 个场景中表现出很好的稳定性, 总体预测精度较高。

### 4.4 消融实验

为了评估所提模型各个部分的作用, 进行了消融实验, 结果如表 2 所示, 其中, DA 表示深度注意力模块, SP 表示自调节稀疏模块, 第 1 行表示在 DASGCN

表1 不同方法在ETH/UCY数据集场景下的ADE/FDE  
Table 1 ADE/FDE of different methods in ETH/UCY data set scenario

Model	ETH	HOTEL	UNIV	ZARA1	ZARA2	Average
Social LSTM <sup>[8]</sup>	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
SGAN <sup>[9]</sup>	0.87/1.62	0.67/1.37	0.76/1.52	0.35/0.68	0.42/0.84	0.61/1.21
SoPhie <sup>[18]</sup>	0.70/1.43	0.76/1.67	0.54/1.24	0.30/0.63	0.38/0.78	0.51/1.15
PITF <sup>[14]</sup>	0.73/1.65	0.30/0.59	0.60/1.27	0.38/0.81	0.31/0.68	0.46/1.00
GAT <sup>[11]</sup>	0.68/1.29	0.68/1.40	0.57/1.29	0.29/0.60	0.37/0.75	0.52/1.07
Social-BIGAT <sup>[16]</sup>	0.69/1.29	0.49/1.01	0.55/1.32	0.30/0.62	0.36/0.75	0.48/1.00
Social-STGCNN <sup>[24]</sup>	0.64/1.11	0.49/0.85	0.44/0.79	0.34/0.53	0.30/0.48	0.44/0.75
STAR <sup>[25]</sup>	<b>0.56</b> /1.11	<b>0.26</b> /0.50	0.52/1.15	0.41/0.90	0.31/0.71	0.41/0.87
SGCN <sup>[13]</sup>	0.63/ <b>1.03</b>	0.32/0.55	0.37/0.70	0.29/0.53	0.25/0.45	0.37/0.65
Proposed model	0.65/1.12	<b>0.26/0.42</b>	<b>0.37/0.70</b>	<b>0.28/0.49</b>	<b>0.23/0.42</b>	<b>0.36/0.63</b>

表2 DA与SP验证结果(ADE/FDE)  
Table 2 Verification results of DA and SP (ADE/FDE)

DA	SP	ETH	HOTEL	UNIV	ZARA1	ZARA2	Average
		0.94/1.53	0.62/1.21	0.68/1.43	0.47/0.79	0.43/0.87	0.63/1.17
	✓	0.76/1.29	0.49/0.83	0.52/1.13	0.35/0.57	0.34/0.64	0.49/0.89
✓		0.71/1.23	0.45/0.92	0.49/0.98	0.34/0.58	0.33/0.56	0.46/0.85
✓	✓	0.65/1.12	0.26/0.42	0.37/0.70	0.28/0.49	0.23/0.42	0.36/0.63

中去除DA和SP,仅用自注意力机制建模密集交互的模型,并记为基准模型,第2行表示在基准模型的基础上增加DA进行行人之间的关系交互的模型,第3行表示在基准模型的基础上增加SP去除行人冗余交互信息的模型,第4行表示在基准模型的基础上增加DA和SP的模型,即所提DASGCN。

从表2可以看出,增加DA可有效利用行人间交互的关系,使得模型有更好的预测精度。基准模型不加DA和SP时与DASGCN相比ADE上升0.27,FDE上升0.54,这表明所提深度注意力机制和自调节稀疏模块对行人轨迹预测的最终准确性具有提升作用。此外,基准模型加DA时与DASGCN相比ADE和FDE分别上升0.10和0.22,这说明行人交互关系的关联对

于行人轨迹预测的重要性。基准模型加SP时与DASGCN相比ADE和FDE分别上升了0.13和0.26,这说明自调节稀疏方法可有效缓解冗余交互信息带来的轨迹偏差。

#### 4.5 可视化

为了精确且直观地显示出所提自调节稀疏模块和深度注意力模块在提升轨迹预测准确性方面的效果,对DASGCN与SR-LSTM的轨迹进行可视化,结果如图3所示,DASGCN模拟人群社会交互关系,预测效果不论是在行人稀疏还是密集场景下都优于SR-LSTM。

从图3第1列可以看出,在行人受邻人影响弱时,DASGCN与SR-LSTM的预测轨迹虽有差距却预测

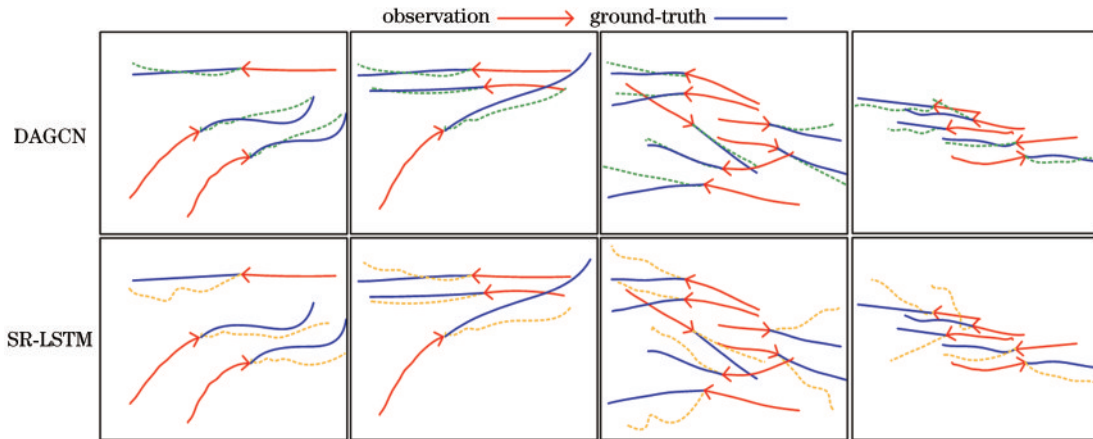


图3 DAGCN与SR-LSTM的轨迹可视化  
Fig. 3 Track visualization of DAGCN and SR-LSTM



方向相近,说明所提模型在稀疏的行人场景中预测结果准确。从图 3 第 2 列可以看出,从左向右移动的行人受相向的两个行人影响时,SR-LSTM 的预测结果与实际轨迹差距明显,而 DASGCN 相比 SR-LSTM 其预测结果更好。即深度注意力可有效捕捉行人间的关联性。从图 3 第 3、4 列可以看出,密集行人场景中 DASGCN 的预测结果与 SR-LSTM 相比更好。DASGCN 的自调节稀疏模块能够在密集场景中准确建模行人间的交互关系。

所提模型在不同场景下的行人预测可视化如图 4 所示,其中实线箭头表示行人的观察轨迹,实线表示行

人真实轨迹,虚线表示行人的预测轨迹。图 4 中,第 1~5 行的场景分别是 ZARA1、HOTEL、ETH、UVI 和 ZARA2。从图 4 可以看出,所提模型对多个行人轨迹的预测结果也很贴近真实轨迹。目标行人可以根据周围多个行人的社会交互关系来确定下一步走向,即所提模型可以很好提取行人间交互关系,并依据内在关联预测出轨迹。在稀疏的行人情况下,目标行人并未受到周围行人影响而维持其原有的走向。所提模型可以去掉周围行人冗余信息从而更加准确预测出行人运动轨迹。实验结果表明,所提轨迹预测模型适用于多目标交互的场景。

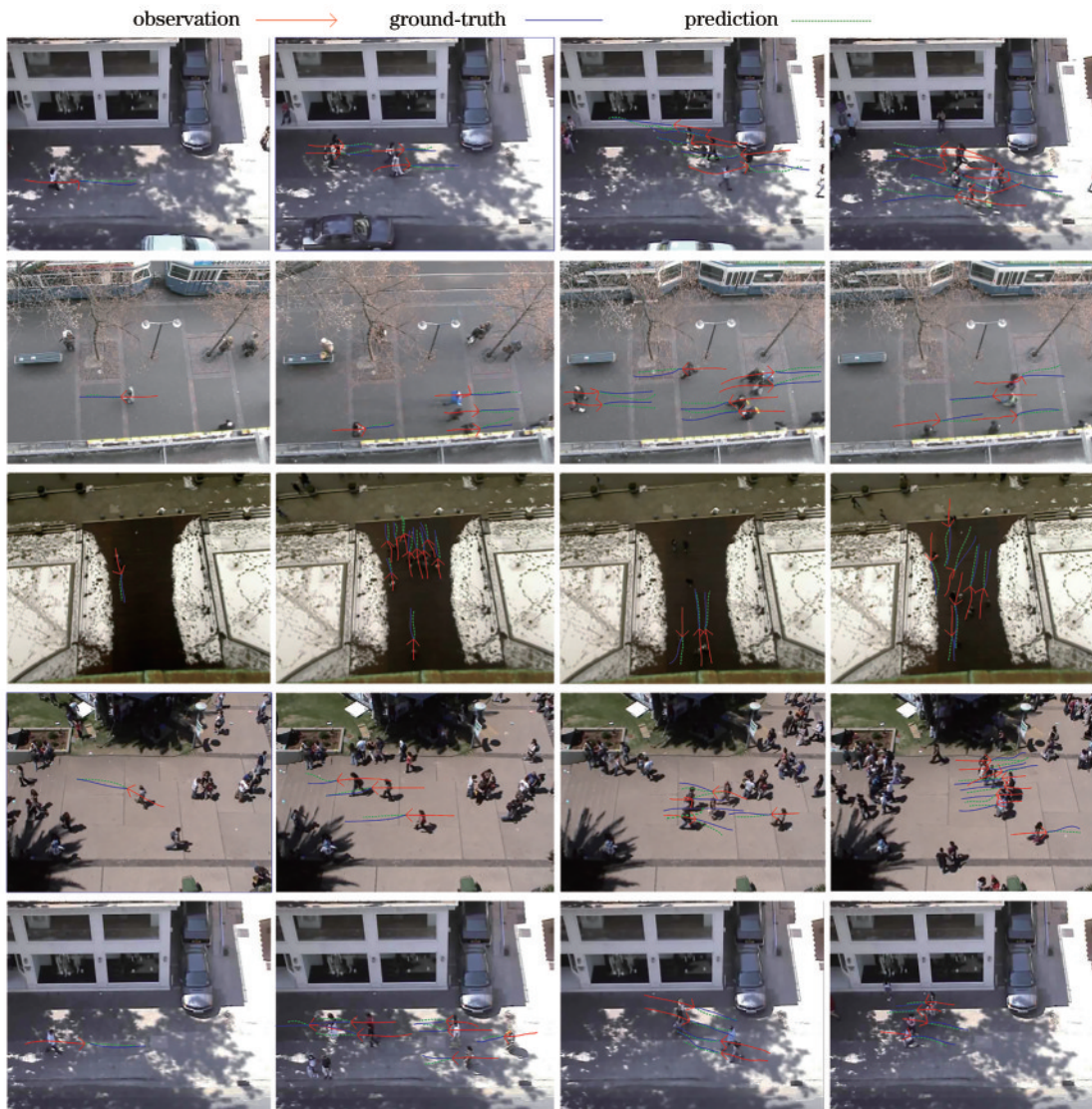


图 4 不同场景行人轨迹预测可视化

Fig. 4 Visualization of pedestrian trajectory prediction in different scenes

## 5 结 论

所提基于注意力机制和稀疏图卷积的行人轨迹预测模型考虑行人信息利用不足和多行人场景中的密集交互带来的轨迹偏差问题,从行人间社会关系的内在

关联性和交互的稀疏性角度出发,设计深度注意力机制和自调节稀疏模块,进而构建行人轨迹预测模型 DASGCN。在 ETH 和 UCY 数据集上对比 DASGCN 和已有模型的轨迹预测效果,验证了所提模型的有效性。

## 参 考 文 献

- [1] Luo Y F, Cai P P, Bera A, et al. PORCA: modeling and planning for autonomous driving among many pedestrians[J]. *IEEE Robotics and Automation Letters*, 2018, 3(4): 3418-3425.
- [2] Raksincharoensak P, Hasegawa T, Nagai M S. Motion planning and control of autonomous driving intelligence system based on risk potential optimization framework[J]. *International Journal of Automotive Engineering*, 2016, 7 (AVEC14): 53-60.
- [3] Ridel D, Rehder E, Lauer M, et al. A literature review on the prediction of pedestrian behavior in urban scenarios [C]//2018 21st International Conference on Intelligent Transportation Systems (ITSC), November 4-7, 2018, Maui, HI, USA. New York: IEEE Press, 2018: 3105-3112.
- [4] Musleh B, García F, Otamendi J, et al. Identifying and tracking pedestrians based on sensor fusion and motion stability predictions[J]. *Sensors*, 2010, 10(9): 8028-8053.
- [5] Yasuno M, Yasuda N, Aoki M. Pedestrian detection and tracking in far infrared images[C]//2004 Conference on Computer Vision and Pattern Recognition Workshop, June 27-July 2, 2004, Washington, DC, USA. New York: IEEE Press, 2004: 125.
- [6] Luber M, Stork J A, Tipaldi G D, et al. People tracking with human motion predictions from social forces[C]//2010 IEEE International Conference on Robotics and Automation, May 3-7, 2010, Anchorage, AK, USA. New York: IEEE Press, 2010: 464-469.
- [7] 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.
- [8] Alahi A, Goel K, Ramanathan V, et al. Social LSTM: human trajectory prediction in crowded spaces[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 961-971.
- [9] Gupta A, Johnson J, Li F F, et al. Social GAN: socially acceptable trajectories with generative adversarial networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2255-2264.
- [10] 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.
- [11] Huang Y F, Bi H K, Li Z X, et al. STGAT: modeling spatial-temporal interactions for human trajectory prediction[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6271-6280.
- [12] Ivanovic B, Pavone M. The trajectron: probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 2375-2384.
- [13] Shi L S, Wang L, Long C J, et al. SGCN: sparse graph convolution network for pedestrian trajectory prediction [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 8990-8999.
- [14] Liang J W, Jiang L, Niebles J C, et al. Peeking into the future: predicting future person activities and locations in videos[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 5718-5727.
- [15] Zhang P, Ouyang W L, Zhang P F, et al. SR-LSTM: state refinement for LSTM towards pedestrian trajectory prediction[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 12077-12086.
- [16] Kosaraju V, Sadeghian A, Martín-Martín R, et al. Social-BiGAT: multimodal trajectory forecasting using bicycle-GAN and graph attention networks[EB/OL]. (2019-06-04)[2022-02-01]. <https://arxiv.org/abs/1907.03395>.
- [17] 欧阳俊, 史庆伟, 王馨心, 等. 基于 GAN 和注意力机制的行人轨迹预测[J]. *激光与光电子学进展*, 2020, 57(14): 141016.
- Ouyang J, Shi Q W, Wang X X, et al. Pedestrian trajectory prediction based on GAN and attention mechanism[J]. *Laser & Optoelectronics Progress*, 2020, 57(14): 141016.
- [18] Sadeghian A, Kosaraju V, Sadeghian A, et al. SoPhie: an attentive GAN for predicting paths compliant to social and physical constraints[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1349-1358.
- [19] Li J C, Ma H B, Tomizuka M. Conditional generative neural system for probabilistic trajectory prediction[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), November 3-8, 2019, Macao, China. New York: IEEE Press, 2019: 6150-6156.
- [20] Vemula A, Muelling K, Oh J. Social attention: modeling attention in human crowds[C]//2018 IEEE International Conference on Robotics and Automation, May 21-25, 2018, Brisbane, QLD, Australia. New York: IEEE Press, 2018: 4601-4607.
- [21] 朱江平, 王睿珂, 段智涓, 等. 基于多尺度注意力机制相位展开的三维人脸建模[J]. *光学学报*, 2022, 42(1): 0112005.
- Zhu J P, Wang R K, Duan Z J, et al. Three-dimensional face modeling based on multi-scale attention phase unwrapping[J]. *Acta Optica Sinica*, 2022, 42(1): 0112005.

- [22] 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.
- [23] Li X S, Liu Y T, Wang K F, et al. A recurrent attention and interaction model for pedestrian trajectory prediction [J]. *IEEE/CAA Journal of Automatica Sinica*, 2020, 7 (5): 1361-1370.
- [24] Mohamed A, Qian K, Elhoseiny M, et al. Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 14412-14420.
- [25] Yu C J, Ma X, Ren J W, et al. Spatio-temporal graph transformer networks for pedestrian trajectory prediction [M]//Vedaldi A, Bischof H, Brox T, et al. *Computer vision-ECCV 2020. Lecture notes in computer science*. Cham: Springer, 2020, 12357: 507-523.
- [26] 徐田野, 丁海勇. 基于融合图卷积的深度学习点云分类方法[J]. *激光与光电子学进展*, 2022, 59(2): 0228005.
- Xu T Y, Ding H Y. Deep learning point cloud classification method based on fusion graph convolution [J]. *Laser & Optoelectronics Progress*, 2022, 59(2): 0228005.
- [27] Pellegrini S, Ess A, Schindler K, et al. You'll never walk alone: modeling social behavior for multi-target tracking[C]//2009 IEEE 12th International Conference on Computer Vision, September 29-October 2, Kyoto, Japan. New York: IEEE Press, 2009: 261-268.