

基于自适应群体掩码图卷积网络的行人轨迹预测

陈旺兴, 桑海峰*, 刘 晴

(沈阳工业大学信息科学与工程学院, 辽宁沈阳 110870)

摘要: 行人轨迹预测对于提升自动驾驶和服务机器人的决策能力以及降低未来潜在碰撞风险方面具有重要意义。然而,由于行人群体内外社会交互关系的差异性和复杂性,现有研究往往未能对群内和群外交互关系进行显式区分与独立建模,不同类型交互特征在模型学习过程中相互混淆,难以精准刻画行人在复杂场景下的真实运动模式,进而制约了模型预测性能的进一步提升。因此,本文提出了一种基于自适应群体掩码图卷积网络(Adaptive Group Masked Graph Convolution Network, AGMGCN)的行人轨迹预测模型,通过对群内和群外交互关系进行独立建模,从而提升模型轨迹预测的准确性。该模型首先构建社会图并采用自注意力机制进行处理,以获得注意力矩阵用于初步表示行人之间的交互关系。后续设计了时频域卷积模块,通过在时域和频域同时对注意力矩阵进一步处理,生成用于表征行人时空交互关系的时频交互矩阵,以实现对行人复杂动态交互更准确的刻画。为有效区分并独立建模群内和群外交互,模型设计了自适应群体掩码模块,根据行人之间的特征相似性自适应确定阈值,并通过阈值处理生成群内掩码矩阵和群外掩码矩阵,为后续群内和群外交互关系的独立建模提供支持。在此基础上,将时频交互矩阵与群内和群外掩码矩阵相结合,并分别应用图卷积捕捉群内交互特征和群外交互特征,从而实现行人群内和群外交互关系的独立建模。最后,模型设计了特征融合模块动态加权融合群内交互特征和群外交互特征,并通过时间卷积网络完成行人未来轨迹的预测。在ETH、UCY和SDD数据集上的实验结果表明,在仅使用23.9 K模型参数的条件下,本文提出的方法相较于DSTIGCN在平均位移误差和终点位移误差上分别降低了12%和20%,验证了所提方法在预测精度和计算成本方面的优势。

关键词: 行人轨迹预测;自适应群体掩码;图卷积网络;时频域卷积模块;时间卷积网络

基金项目: 辽宁省科技厅重点研发项目(No.2024JH2/102400028);辽宁省教育厅科研项目(No.YTMS20231210)

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2026)02-0774-11

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20251200

Adaptive Group Masked Graph Convolution Network for Pedestrian Trajectory Prediction

CHEN Wangxing, SANG Haifeng*, LIU Qing

(School of Information Science and Engineering, Shenyang University of Technology, Shenyang, Liaoning 110870, China)

Abstract: Pedestrian trajectory prediction is crucial for improving the decision-making capabilities of autonomous vehicles and service robots, as well as mitigating potential future collision risks. However, due to the differences and complexity of social interactions within and outside pedestrian groups, existing studies often fail to explicitly distinguish and independently model interactions within and outside groups. This leads to confusion of different types of interaction features during model learning, making it difficult to accurately depict the real motion patterns of pedestrians in complex scenarios, thus restricting further improvement in model prediction performance. Therefore, this paper proposes a pedestrian trajectory prediction model based on an adaptive group masked graph convolution network (AGMGCN). By independently modeling in-group and out-group interactions, the accuracy of model trajectory prediction is improved. The model first constructs a social graph and processes it using a self-attention mechanism to obtain an attention matrix that initially represents the interactions between pedestrians. Subsequently, a time-frequency domain convolution module is designed to further process the attention matrix in both the time and frequency domains, generating a time-frequency interaction matrix to characterize the spatial-temporal interactions of pedestrians, thus achieving a more accurate portrayal of complex dynamic interactions. To effectively distinguish and independently model in-group and out-group interactions, an adaptive group masking module is designed. This module adaptively determines a threshold based on the feature similarity between pedestrians and generates in-group and out-group masking matrices through threshold processing, providing support for the subsequent independent

modeling of in-group and out-group interactions. Building upon this foundation, the time-frequency interaction matrix is combined with out-group and in-group masking matrices, and graph convolution is applied to respectively capture in-group and out-group interaction features, thereby achieving independent modeling of pedestrian in-group and out-group interaction relationships. Finally, a feature fusion module is designed to dynamically weight and fuse in-group and out-group interaction features and then temporal convolution networks are used to predict the future trajectory of pedestrians. Experimental results on the ETH, UCY and SDD datasets demonstrate that, with only 23.9 K model parameters, the proposed method reduces the average displacement error by 12% and the final displacement error by 20% compared to DSTIGCN, validating the advantages of the proposed method in terms of prediction accuracy and computational cost.

Keywords: pedestrian trajectory prediction; adaptive group masking; graph convolution network; time-frequency domain convolution module; temporal convolution networks

Foundation Item(s): Key Research and Development Projects of Liaoning Provincial Department of Science and Technology (No.2024JH2/102400028); Liaoning Provincial Department of Education Research Project (No.YTMS20231210)

0 引言

准确预测行人运动轨迹对于提升自动驾驶车辆^[1]以及服务机器人^[2]的决策能力,降低潜在碰撞风险以及增强系统整体安全性方面具有重要意义。然而,由于行人之间的社会交互关系高度复杂且难以精确刻画,当前行人轨迹预测任务仍然面临诸多挑战。

传统轨迹预测方法^[3]采用人工设计函数模拟行人运动模式,但受限于固定函数的表达能力,难以捕捉真实场景中行人行为的多样性和复杂性。随着深度学习的发展,基于数据驱动的轨迹预测方法能够从历史轨迹中自主学习行人潜在运动模式,已成为轨迹预测的主流方法^[4]。早期研究 Social LSTM^[5]通过长短期记忆网络(Long Short-Term Memory, LSTM)和社交池化学习行人运动状态和社会关系,SR-LSTM^[6]和 Social GAN^[7]在此基础上分别引入状态细化模块和生成对抗网络,以提升模型预测性能。后续研究进一步引入注意力机制^[8-9]和 Transformer 架构^[10-11]用于学习行人社会交互关系和运动模式,进而降低模型预测误差。近年来,图神经网络(Graph Neural Networks, GNN)通过显式建模行人间的图结构关系^[12-14],在学习行人复杂社会动态方面展现出显著优势。后续基于 GNN 的轨迹预测方法进一步探索了冗余交互滤除^[15-16]、时空交互增强^[17]以及频域特征建模^[18-19],以提升模型建模能力,降低模型预测误差。为捕获群体层面的社会交互关系, GroupNet^[20]首次将超图神经网络(Hyper-Graph Neural Network, HGNN)引入行人轨迹预测,后续研究^[21]在此基础上进一步扩展了 HGNN 的建模能力。然而,这些基于 HGNN 的轨迹预测方法并未显式区分群体内部和群体外部的不同交互行为,这在一定程度上限制了其对行人真实社会动态的准确建模。

在实际场景中,行人往往以群体形式运动,行人社会交互可以分为群内交互和群外交互两种类型,这两类交互既共存又存在显著差异^[22],如图 1 所示。尽管现有基于 GNN 和 HGNN 的方法在行人社会交互建

模方面取得了进展,但仍缺乏对群内和群外交互进行显式区分与独立建模。

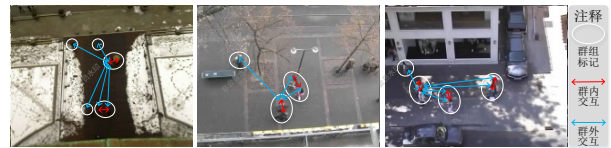


图1 群体内和群体外交互的示意图

Figure 1 Diagram illustrating interactions within and outside groups

为此,本文提出了一个自适应群体掩码图卷积网络(Adaptive Group Masked Graph Convolution Network, AGMGCN),通过分别屏蔽群内和群外行人,对群内紧密关系与群外潜在影响独立建模,以更准确地反映行人真实社会动态。具体而言,AGMGCN 首先基于行人历史轨迹信息构建社会图,并利用自注意力机制初步建模行人社会交互关系。随后,通过时频域卷积模块同时在时域和频域进行进一步处理,得到时频交互矩阵,以提升对行人复杂交互模式的表征能力。为分别建模群内和群外交互,本文设计了一个自适应群体掩码模块,基于行人速度和位置信息自适应划分群体内外行人并生成对应的群内和群外交互掩码矩阵。后续将时频交互矩阵与群内和群外交互掩码矩阵进行融合,并利用图卷积操作捕获群内交互特征和群外交互特征,从而显式建模群内和群外交互关系。最后,使用特征融合模块整合群体内外交互特征,并通过时间卷积网络实现行人未来轨迹的预测。

本工作的主要贡献可以概括为四点:(1)提出了一种自适应群体掩码图卷积网络,独立建模群体内和群体外社会交互关系,从而提升行人轨迹预测的准确性;(2)设计了一种自适应群体掩码模块,基于行人速度和位置信息实现场景中群体内外行人的自适应划分并生成相应的掩码矩阵,为群内交互和群外交互的独立建模提供支持;(3)为进一步准确描述行人复杂时空交互关系,设计了一个时频域卷积模块,同时

在时域和频域中采用非对称三维卷积进行处理,以弥补单一时域建模对复杂时空交互刻画不足;(4)在多个公开数据集上的实验表明,所提出方法在预测误差方面优于现有基线方法,后续消融实验进一步验证了模型各组成部分的有效性。

1 方法

1.1 问题定义

行人轨迹预测任务通过构建轨迹预测模型,根据场景中 N 个行人前一段时间 $T_{\text{obs}} = \{1, 2, \dots, t_{\text{obs}}\}$ 的历

史轨迹 $X = \{p_i^t = (x_t^i, y_t^i) | i \in N, t \in T_{\text{obs}}\}$, 预测行人未来一段时间 $T_{\text{pred}} = \{t_{\text{obs}+1}, t_{\text{obs}+2}, \dots, t_{\text{pred}}\}$ 的运动轨迹 $\hat{Y} = \{\hat{p}_i^t = (\hat{x}_t^i, \hat{y}_t^i) | i \in N, t \in T_{\text{pred}}\}$ 。为保证模型预测的准确性,生成的预测轨迹 \hat{Y} 应尽可能接近真实轨迹 $Y = \{p_i^t = (x_t^i, y_t^i) | i \in N, t \in T_{\text{pred}}\}$ 。

1.2 模型描述

为有效建模行人群内和群外社会交互关系,本文提出了一个基于自适应群体掩码图卷积网络的轨迹预测模型,其整体架构如图2所示。

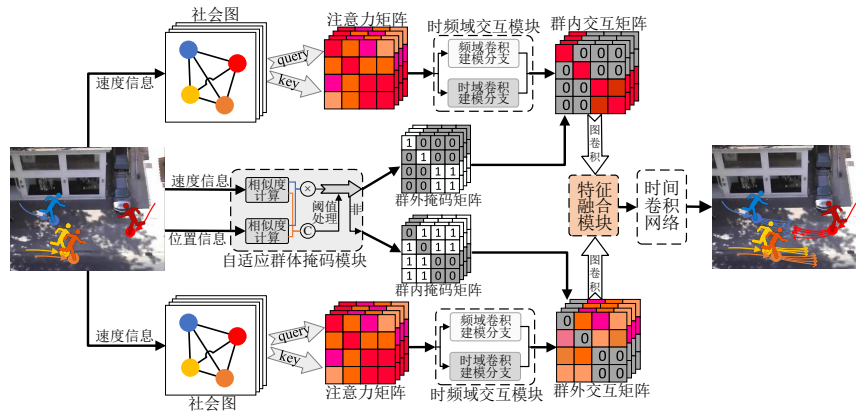


图2 基于自适应群体掩码图卷积网络的轨迹预测模型整体框图

Figure 2 Overall architecture diagram of the trajectory prediction model based on adaptive group masked graph convolution network

具体而言,模型首先根据行人速度信息构建社会图,并利用自注意力机制生成注意力矩阵,初步表征行人之间的社会交互关系。随后,通过时频域交互模块对注意力矩阵进一步处理,以提升其对行人复杂时空依赖关系的刻画能力。与此同时,模型将行人速度和位置信息输入到自适应群体掩码模块,生成分别用于屏蔽群体内和群体外行人的群内掩码矩阵和群外掩码矩阵。后续,将群外和群内掩码矩阵与时频域交互模块输出矩阵相结合,得到分别用于表征群内紧密互动和群外潜在影响的群内和群外交互矩阵。模型后续应用图卷积捕获群体内外社会交互特征,并通过特征融合模块进行融合。最后,通过时间卷积网络完成行人未来轨迹的预测。

1.3 社会图的构建和初步处理

模型基于行人速度信息 $v_t^i = p_t^i - p_{t-1}^i$ 构建以行人为节点,行人之间的社会关系为边的社会图 $G_S = (V_t, U_t)$, $V_t = \{v_t^i | i = 1, 2, \dots, N\}$ 和 $U_t = \{u_t^{ij} | i, j = 1, 2, \dots, N\}$ 分别表示社会图 G_S 的节点和边。 U_t 中的 u_t^{ij} 用于表示行人 i 和 j 之间是否存在交互。若 $u_t^{ij} = 1$, 表示行人 i 和 j 之间存在交互;若 $u_t^{ij} = 0$, 则表示无交互。 U_t 中的所有元素在初始化时均设置为 1, 以确保模型能够考虑所

有行人之间的社会交互。

为初步表征行人社会交互关系,模型通过线性变换 ϕ 处理社会图 G_S , 用于获得行人高维特征 $H_t = \phi(G_S, W_t)$, 其中, W_t 表示线性变换的权重。随后,分别通过两个不同的线性变换 ϕ , 将高维特征 H_t 进一步转化为对应的查询向量 $\text{Query} = \phi(H_t, W_t^q)$ 和关键向量 $\text{Key} = \phi(H_t, W_t^k)$, 其中, W_t^q 和 W_t^k 表示对应的线性变换权重。后续将查询向量 Query 和关键向量 Key 进行相乘,再通过 Softmax 函数生成注意力矩阵 R_S , 用于初步表征行人之间的社会交互关系,如式(1)所示。其中, T 表示矩阵转置,而缩放因子 $\sqrt{d_k}$ 用于保证训练过程的数值稳定性。

$$R_S = \text{Softmax} \left(\frac{\text{Query} \times \text{Key}^T}{\sqrt{d_k}} \right) \quad (1)$$

1.4 自适应群体掩码模块

为区分真实场景中群内外行人,生成相应的掩码矩阵以支持后续群体内外交互关系的建模,本文提出了一种自适应群体掩码模块,如图3所示。该模块首先基于行人的位置和速度信息,采用余弦相似度分别计算其位置相关矩阵和速度相关矩阵,用于量化行人

之间的关联程度。随后,通过轻量化卷积模块生成阈值矩阵,并对位置与速度相关矩阵的乘积进行二值化处理,以生成群外掩码矩阵和群内掩码矩阵。自适应群体掩码模块能够根据行人速度信息和位置信息,自适应动态生成两类掩码矩阵:群外掩码矩阵用于屏蔽群外行人,以实现群体内部行人社会交互关系的独立建模;群内掩码矩阵则用于屏蔽群内行人,从而支持对群体外行人社会交互关系的独立建模。

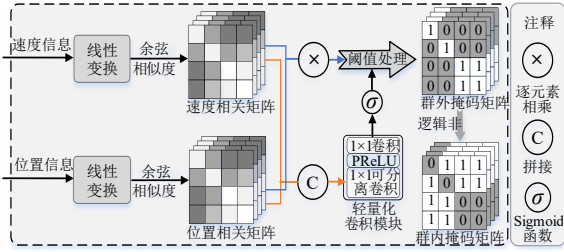


图3 自适应群体掩码模块架构

Figure 3 Architecture of the adaptive group masking module

自适应群体掩码模块首先通过线性变换 ϕ 对行人位置信息 p_i^t 进行映射,以获得更具表达力的位置高维特征 $p_i^t = \phi(p_i^t, W_p)$ 。后续采用余弦相似度计算位置相关矩阵 $A_p = \{A_{p_i^t, p_j^t} | i \in N, j \in N, t \in T_{\text{obs}}\}$, 其中, $A_{p_i^t, p_j^t}$ 用于表示在时间帧 t 下行人 i 和行人 j 之间空间位置信息的关联程度。

$$A_{p_i^t, p_j^t} = \frac{p_i^t p_j^t}{|p_i^t| |p_j^t|} \quad (2)$$

同时,自适应群体掩码模块使用线性变换 ϕ 处理行人速度信息 v_i^t ,以获得速度高维特征 $v_i^t = \phi(v_i^t, W_v)$ 。并同样通过余弦相似度来获得行人速度相关矩阵 $A_v = \{A_{v_i^t, v_j^t} | i \in N, j \in N, t \in T_{\text{obs}}\}$, 其中, $A_{v_i^t, v_j^t}$ 用于表示在时间帧 t 下行人 i 和行人 j 之间速度信息的关联程度。

$$A_{v_i^t, v_j^t} = \frac{v_i^t v_j^t}{|v_i^t| |v_j^t|} \quad (3)$$

自适应群体掩码模块后续采用逐元素相乘融合位置相关矩阵 A_p 和速度相关矩阵 A_v ,生成行人相关矩阵 $A_g = A_p \times A_v$ 。为自适应确定阈值,该模块将位置与速度相关矩阵进行拼接,并通过轻量化卷积模块和 Sigmoid 函数 σ 生成阈值矩阵 T 。轻量化卷积模块由 1×1 可分离卷积, PReLU 激活函数和 1×1 卷积依次构成。 1×1 可分离卷积能够独立处理行人位置和速度相关矩阵; PReLU 激活函数引入非线性变换,增强模型表达能力;而 1×1 卷积用于特征整合。其中, QCM 表示轻量化卷积模块。

$$T = \sigma\left(\text{QCM}\left(\left[A_p, A_v\right]\right)\right) \quad (4)$$

后续将行人相关矩阵 A_g 与阈值矩阵 T 进行比较,通过阈值化处理生成二值化的群外掩码矩阵 M_o 。由于阈值化处理过程中所使用的单位阶跃函数 $U(\cdot)$ 不可微,本文借鉴了直通估计器 (Straight-Through Estimator, STE)^[23] 对梯度进行近似反向传播,以确保模型的端到端训练。

$$M_o = U(A_g - T), U(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

群外掩码矩阵 M_o 用于屏蔽群体外部的行人,从而使模型专注于学习群体内部行人社会交互关系。为进一步捕获群体外部行人社会交互关系,模型对群外掩码矩阵 M_o 进行逻辑非操作,以生成相应的群内掩码矩阵 $M_i = 1 - M_o$ 。群内掩码矩阵 M_i 能够屏蔽群体内部行人,使模型专注于建模群体外部行人间的交互。

1.5 时频域卷积模块

为增强模型对行人社会交互关系的表征能力,本文设计了一个时频域卷积模块,同时在时域和频域对注意力矩阵进行处理,弥补以往单一时域建模对复杂动态交互刻画不足,以提升模型对复杂社会交互的捕捉能力。时频域卷积模块的整体架构如图4所示,具体来说,该模块同时在时域和频域中执行非对称三维卷积操作,进一步挖掘潜在时空依赖关系,以提升模型对于行人社会交互关系的建模能力。

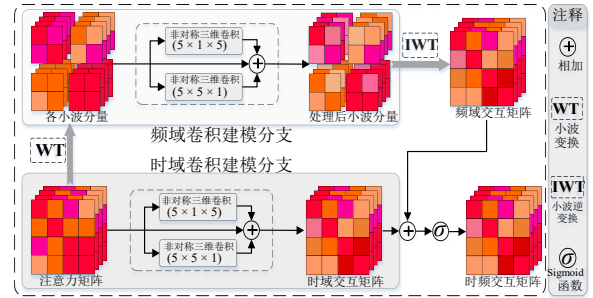


图4 时频域卷积模块架构

Figure 4 Architecture of the time-frequency domain convolution module

时频域卷积模块首先在时域中采用非对称三维卷积对注意力矩阵 R_s 进行处理,建模行人之间的主被动交互关系及其时空依赖特性,以获得时域交互矩阵 R_s^T ,如式(6)所示。非对称三维卷积^[24]的非对称架构能够捕获行人之间的主被动关系,而三维卷积能够同时挖掘空间和时间维度信息,有助于模型学习更具判别力的时空社会交互特征。其中, $C_{5 \times 1 \times 5}$ 和 $C_{5 \times 5 \times 1}$ 分别表示卷积核为 $5 \times 1 \times 5$ 和 $5 \times 5 \times 1$ 的非对称三维卷积。

$$R_s^T = C_{5 \times 1 \times 5}(R_s) + C_{5 \times 5 \times 1}(R_s) + R_s \quad (6)$$

同时,时频域卷积模块利用小波变换对注意力矩

阵 \mathbf{R}_S 进行多尺度分解,得到对应不同频率成分的小波分量 $\{\mathbf{R}_S^{c_0c_1} | c_0, c_1 \in \{L, H\}\}$,进而从不同角度进一步建模行人交互关系,如式(7)所示。其中, $(\downarrow 2)$ 和 $*$ 表示下采样和卷积操作, $\{f_{c_0c_1} | c_0, c_1 \in \{L, H\}\}$ 用于表示小波变换中的滤波器。

$$\mathbf{R}_S^{c_0c_1} = (\downarrow 2)(f_{c_0c_1} * \mathbf{R}_S), c_0, c_1 \in \{L, H\} \quad (7)$$

小波分量中的低频分量 \mathbf{R}_S^L 用于表征行人之间的全局交互模式,而高频分量 $\mathbf{R}_S^LH, \mathbf{R}_S^{HL}, \mathbf{R}_S^{HH}$ 用于刻画行人细节交互模式。随后,对各小波分量 $\{\mathbf{R}_S^{c_0c_1} | c_0, c_1 \in \{L, H\}\}$ 分别采用非对称三维卷积进行建模,以实现全局交互和细节交互模式的捕获,有

$$\mathbf{C}_S^{c_0c_1} = C_{5 \times 1 \times 5}(\mathbf{R}_S^{c_0c_1}) + C_{5 \times 5 \times 1}(\mathbf{R}_S^{c_0c_1}) + \mathbf{R}_S^{c_0c_1}, c_0, c_1 \in \{L, H\} \quad (8)$$

模型后续采用小波逆变换对卷积处理后的小波分量 $\{\mathbf{C}_S^{c_0c_1} | c_0, c_1 \in \{L, H\}\}$ 进行重构,用于获得频域交互矩阵 \mathbf{R}_S^F ,如式(9)所示。其中, $(\uparrow 2)$ 和 $*$ 分别表示上采样和卷积操作。

$$\mathbf{R}_S^F = \sum_{c_0, c_1 \in \{L, H\}} f_{c_0c_1} * (\uparrow 2) \mathbf{C}_S^{c_0c_1} \quad (9)$$

在获得时域交互矩阵 \mathbf{R}_S^T 和频域交互矩阵 \mathbf{R}_S^F 后,模型对二者进行相加并使用 Sigmoid 函数 σ 生成时频交互矩阵 \mathbf{R}_S^{TF} ,用于描述行人社会交互关系,即

$$\mathbf{R}_S^{TF} = \sigma(\mathbf{R}_S^T + \mathbf{R}_S^F) \quad (10)$$

1.6 图卷积

模型随后将时频交互矩阵 \mathbf{R}_S^{TF} 分别与群外掩码矩阵 \mathbf{M}_O 和群内掩码矩阵 \mathbf{M}_I 相乘,从而得到用于表征群内紧密互动的群内交互矩阵 \mathbf{A}_I 以及用于表征群外潜在影响的群外交互矩阵 \mathbf{A}_O 。群内交互矩阵和群体外交互矩阵能够在结构层面显式区分不同类型交互,从而提升模型对复杂交互模式的表征能力,即

$$\mathbf{A}_I = \mathbf{R}_S^{TF} \times \mathbf{M}_O \quad (11)$$

$$\mathbf{A}_O = \mathbf{R}_S^{TF} \times \mathbf{M}_I \quad (12)$$

为增强节点本身的影响力,在群内和群外交互矩阵中添加了一个额外的单位矩阵 \mathbf{I} 。后续采用图卷积对社会图 G_S 与群内交互矩阵 \mathbf{A}_I 和群外交互矩阵 \mathbf{A}_O 进行处理,以获得群内交互特征 F_I 和群外交互特征 F_O ,分别用于描述行人群内外交互关系,如式(13)(14)所示。其中, W_G^I 和 W_G^O 分别表示对应图卷积的权重。

$$F_I = \phi((\mathbf{A}_I + \mathbf{I})G_S, W_G^I) \quad (13)$$

$$F_O = \phi((\mathbf{A}_O + \mathbf{I})G_S, W_G^O) \quad (14)$$

1.7 特征融合模块

为加权融合群内交互特征 F_I 和群外交互特征

F_O ,本文设计了特征融合模块,如图5所示。特征融合模块首先分别使用 1×1 卷积 $C_{1 \times 1}$ 对群内交互特征和群外交互特征进行映射,以对齐特征维度并增强特征表达能力。随后,通过 Softmax 函数生成各特征对应权重 $[W_I, W_O]$,以便后续对群内与群外交互特征进行加权融合,即

$$[W_I, W_O] = \text{Softmax}([C_{1 \times 1}(F_I), C_{1 \times 1}(F_O)]) \quad (15)$$

与此同时,该模块使用另一组 1×1 卷积 $C_{1 \times 1}$ 并结合残差连接对群内外交互特征进行处理,进而与对应权重 $[W_I, W_O]$ 进行加权融合。最后通过相加操作融合群内与群外交互特征,获取行人轨迹表示特征 F_P ,即

$$F_P = W_I(C_{1 \times 1}(F_I) + F_I) + W_O(C_{1 \times 1}(F_O) + F_O) \quad (16)$$

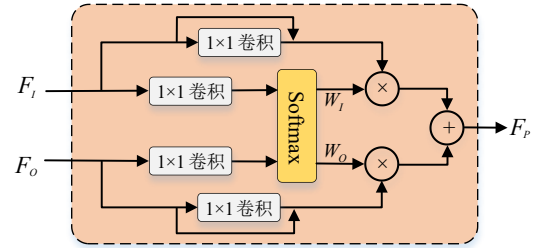


图5 特征融合模块架构

Figure 5 Architecture of the feature fusion module

1.8 轨迹预测和损失函数

在获得行人轨迹表示特征之后,模型采用时间卷积网络(Temporal Convolutional Network, TCN)预测行人未来运动轨迹。因 Uhlemann 等人^[25]验证了恒定速度模型(Constant Velocity Model, CVM)能够实现可靠的单条行人未来轨迹,受此启发,本文将 CVM 预测的轨迹 p_{CVM} 作为基线轨迹。随后,模型采用时间卷积神经网络预测基线轨迹与真实轨迹之间的偏差 $\{o_t^i | i \in N, t \in T_{pred}\}$,并将该偏差与基线轨迹 p_{CVM} 进行相加,以生成最终预测结果 $\{\hat{p}_t^i = o_t^i + p_{CVM} | i \in N, t \in T_{pred}\}$ 。此外,为实现多模态轨迹预测,模型能够预测 S 个偏差 $\{o_t^{(s)i} | i \in N, t \in T_{pred}, s \in S\}$,再结合基线轨迹 p_{CVM} ,以实现 S 条行人未来运动轨迹 $\{\hat{p}_t^{(s)i} = o_t^{(s)i} + p_{CVM} | i \in N, t \in T_{pred}, s \in S\}$ 的预测。

为优化模型训练,本文采用了一种结合了赢者通吃损失 L_w 和总体损失 L_o 的复合损失函数 $L = L_w + 0.01L_o$ 。赢者通吃损失 L_w 选择所有预测轨迹中与真实轨迹最接近的那条轨迹,与真实轨迹计算模型的平均和最终位移误差,用于增强模型预测结果的多样性。而总体损失 L_o 计算所有预测轨迹与真实轨迹之间的平均位

称误差,用于确保所有预测轨迹的准确性,具体公式如下:

$$L_w = \frac{1}{N(t_{\text{pred}} - t_{\text{obs}})} \sum_{i=1}^N \sum_{t=t_{\text{obs}}+1}^{t_{\text{pred}}} \min_{s \in [1,2,\dots,S]} \left\| \hat{p}_t^{(s)i} - p_t^i \right\|^2 \quad (17)$$

$$+ \frac{1}{N} \sum_{i=1}^N \min_{s \in [1,2,\dots,S]} \left\| \hat{p}_{t_{\text{pred}}}^{(s)i} - p_{t_{\text{pred}}}^i \right\|^2$$

$$L_o = \frac{1}{N(t_{\text{pred}} - t_{\text{obs}})_S} \sum_{i=1}^N \sum_{t=t_{\text{obs}}+1}^{t_{\text{pred}}} \sum_{n=1}^S \left\| \hat{p}_t^{(n)i} - p_t^i \right\|^2 \quad (18)$$

$$+ \frac{1}{NS} \sum_{i=1}^N \sum_{n=1}^S \left\| \hat{p}_{t_{\text{pred}}}^{(n)i} - p_{t_{\text{pred}}}^i \right\|^2$$

2 实验和分析

2.1 数据集和评价指标

为客观评估所提出方法的预测性能,本文在 ETH^[26]、UCY^[27]和 SDD^[28]三个广泛使用的轨迹预测数据集上进行了实验。ETH 和 UCY 是行人轨迹预测研究中的经典数据集,其中 ETH 数据集包含 ETH 和 HOTEL 两个场景,UCY 数据集则包含 UNIV、ZARA1 和 ZARA2 三个场景。遵循以往研究的评估策略,实验采用留一法进行模型预测性能的评估,即在四个场景上训练模型,然后在剩余一个场景上进行模型预测性能的测试。SDD 数据集是由斯坦福大学发布,通过无人机以鸟瞰视角采集多个真实校园场景中的行人轨迹数据。与 DAG-Net^[29]和 Social-Implicit^[30]的设置一致,本文将训练集分为训练、测试和验证集,以进行全面且一致的模型评估。在上述轨迹预测数据集中,行人轨迹均以 0.4 s 的时间间隔进行采样,以方便进行模型预测性能的对比如。

在评价指标方面,本文采用行人轨迹预测领域广泛使用的两项指标即平均位移误差(Average Displacement Error, ADE)和终点位移误差(Final Displacement Error, FDE)。ADE 计算整个预测时段内预测轨迹点与真实轨迹点之间的平均 L2 距离,用于反映整体轨迹预测的精确度;而 FDE 则计算预测轨迹终点与真实终点之间的 L2 距离,用于衡量模型对最终位置的预测准确性,具体公式如下:

$$\text{ADE} = \frac{\sum_{i=1}^N \sum_{t=t_{\text{obs}}+1}^{t_{\text{pred}}} \left\| \hat{p}_t^i - p_t^i \right\|^2}{N(t_{\text{pred}} - t_{\text{obs}})} \quad (19)$$

$$\text{FDE} = \frac{\sum_{i=1}^N \left\| \hat{p}_{t_{\text{pred}}}^i - p_{t_{\text{pred}}}^i \right\|^2}{N} \quad (20)$$

2.2 实验设置

在本研究中,模型所有线性变换的输出维度 D 统一设置为 16。轨迹预测部分采用五层时间卷积网络

进行未来轨迹预测。时频域卷积模块选用 db1 小波基,非对称 3D 卷积核的尺寸分别设置为 $5 \times 1 \times 5$ 和 $5 \times 5 \times 1$ 。在自适应群体掩码模块中,逻辑非操作会额外加入一个极小值 e^{-6} ,以防止梯度异常并确保训练过程的稳定性。训练阶段采用 Adam 优化器,共进行 150 个训练轮次,初始学习率设置为 0.01,并采用分步衰减策略,每 50 个训练轮次将学习率衰减为原来的 0.1。与现有行人轨迹预测方法保持一致,模型在所有数据集上均使用 8 帧(3.2 s)行人轨迹作为输入,进而预测行人后续 12 帧(4.8 s)的运动轨迹。测试阶段遵循以往研究的设置,模型一共预测 20 条未来轨迹,并使用最接近真实轨迹的预测轨迹计算模型 ADE 和 FDE。所有实验均在 PyTorch 深度学习框架下实现,并在搭载 Intel i9-13900KF CPU 和 NVIDIA GeForce RTX 4090 GPU 的计算平台上完成。

2.3 定量分析

为全面评估所提出方法的预测性能,本文首先在 ETH 和 UCY 数据集上将其与现有模型在 ADE 和 FDE 评价指标上进行定量比较,预测误差对比结果如表 1 所示。由表 1 可知,本文提出的 AGMGCN 相比于现有模型在多个场景下实现了更低的预测误差。

与经典模型 SR-LSTM 和 STGAT 相比,AGMGCN 显著降低了轨迹预测误差。而相比于基于 Transformer 的 TF,所提出的方法在 ADE 和 FDE 评价指标上分别实现了 29% 和 35% 的预测性能提升。与同样学习群体内外交互的方法 GCHGAT 相比,AGMGCN 通过自适应群体掩码和时频域卷积协同建模大幅降低了模型的 ADE 和 FDE。与基于 HGNN 的 GroupNet 相比,AGMGCN 通过显式区分群内与群外互动,更准确地刻画了真实场景中行人社会动态,从而减少了模型 12% 的 ADE 和 18% 的 FDE。相较于群体动态交互建模方法 Dyn-GroupNet,AGMGCN 凭借其自适应群体划分与时频域协同学习,依然保持了领先的预测性能。相比于最新基于 GNN 的方法 DSTIGCN,本文提出的 AGMGCN 同样实现了更准确的行人未来轨迹预测,进一步验证 AGMGCN 显式区分并独立建模群内外交互的有效性。在复杂的 UNIV 场景下,AGMGCN 在 ADE 和 FDE 评价指标上取得了最好的预测效果,这表明所提出的方法在复杂人群场景下具有更强的泛化能力和鲁棒性。

为进一步验证所提出方法的预测性能,本文在 SDD 数据集上对模型预测误差进行了定量评估,实验结果如表 2 所示。由表 2 可知,所提出方法相比于现有模型在 SDD 数据集上实现了更低的 ADE 和 FDE。与基于小波变换图卷积方法 WTGCN 相比,本方法通过在时域和频域进行社会交互的进一步增强以及群体内外社会交互的独立建模,在 ADE 和 FDE 评价指标上大幅提升了模型轨迹预测的准确度。本文提出

表1 AGMGCN与现有方法在ETH和UCY数据集上的ADE/FDE指标(以米为单位)

Table 1 Comparison of ADE/FDE metrics (in meters) for AGMGCN and existing methods on the ETH and UCY datasets

	年份	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均值
Social GAN ^[7]	2018	0.81/1.52	0.72/1.61	0.60/1.26	0.34/0.69	0.42/0.84	0.58/1.18
SR-LSTM ^[6]	2019	0.63/1.25	0.37/0.74	0.51/1.10	0.41/0.90	0.32/0.70	0.45/0.94
STGAT ^[13]	2019	0.65/1.12	0.35/0.66	0.52/1.10	0.34/0.69	0.29/0.60	0.43/0.83
STAR ^[9]	2020	0.36/0.65	0.17/0.36	0.31/0.62	0.26/0.55	0.22/0.46	0.26/0.53
Social-STGCNN ^[14]	2020	0.64/1.11	0.49/0.85	0.44/0.79	0.34/0.53	0.30/0.48	0.44/0.75
TF ^[10]	2021	0.61/1.12	0.18/0.30	0.35/0.65	0.22/0.38	0.17/0.32	0.31/0.55
SGCN ^[15]	2021	0.63/1.03	0.32/0.55	0.37/0.70	0.29/0.53	0.25/0.45	0.37/0.65
GCHGAT ^[22]	2022	0.63/1.10	0.38/0.73	0.55/1.16	0.33/0.66	0.30/0.64	0.44/0.86
GroupNet ^[20]	2022	0.46/0.73	0.15/0.25	0.26/0.49	0.21/0.39	0.17/0.33	0.25/0.44
RDGCN ^[24]	2023	0.58/0.94	0.30/0.45	0.35/0.65	0.28/0.48	0.25/0.44	0.35/0.59
IMGCN ^[16]	2024	0.61/0.82	0.31/0.45	0.37/0.67	0.29/0.51	0.24/0.42	0.36/0.57
WTGCN ^[19]	2024	0.60/0.95	0.25/0.37	0.36/0.65	0.27/0.46	0.23/0.39	0.34/0.56
DynGroupNet ^[21]	2024	0.42/0.66	0.13/0.20	0.24/0.44	0.19/0.34	0.15/0.28	0.23/0.38
DSTIGCN ^[17]	2025	0.43/0.70	0.22/0.41	0.25/0.45	0.20/0.37	0.17/0.32	0.25/0.45
AGMGCN	—	0.41/0.63	0.16/0.26	0.22/0.40	0.17/0.30	0.13/0.23	0.22/0.36

注:值越低越好,最佳值以粗体显示。

表2 AGMGCN与现有方法在SDD数据集上的ADE/FDE指标(以米为单位)

Table 2 Comparison of ADE/FDE metrics (in meters) for AGMGCN and existing methods on the SDD datasets

	STGAT ^[13]	DAG-Net ^[29]	SGCN ^[15]	Social-Implicit ^[30]	WTGCN ^[19]	DSTIGCN ^[17]	AGMGCN
ADE	0.58	0.53	0.46	0.47	0.43	0.37	0.31
FDE	1.11	1.04	0.75	0.89	0.72	0.61	0.52

注:值越低越好,最佳值以粗体显示。

的AGMGCN能够分别捕捉群体内紧密的社交关联与群体外行人相互影响,更贴合真实场景中行人社会交互。与最新基于GNN方法DSTIGCN相比,AGMGCN在ADE和FDE评价指标上分别实现了16%和15%模型预测性能提升。

为评估所提方法的实时性能,本文在模型参数和推理时间上与现有可用方法进行了对比,对比结果如表3所示。表3中所有方法的测试均在统一的软硬件环境下进行,其中推理时间为多个单步预测时间的平均值。由表3可知,本文提出的AGMGCN在模型参数和推理时间上优于对比方法。与基于RNN的SR-LSTM和Social GAN相比,AGMGCN得益于其并行训练能力,在推理速度上表现出显著优势。相较于基于Transformer的TF和STAR,AGMGCN大幅降低了模型参数量并缩短了推理时间。与基于HGNN的GroupNet相比,本文提出的AGMGCN能够以更少的计算成本实现群体交互的有效建模。此外,与基于GNN的RDGCN和WTGCN相比,AGMGCN不仅取得了更优的预测性能,同时具备更少的模型参数以及更快的推理速度。这些实时性能对比结果表明,本文提出的方法能够在计算复杂度和模型预测性能之间实现良好的平衡。

表3 不同模型的模型参数和推理时间对比

Table 3 Comparison of model parameters and inference time for different models

模型	类型	模型参数/K	推理时间/s
SR-LSTM ^[6]	RNN	64.9	0.0708
Social GAN ^[7]	RNN	46.3	0.0551
TF ^[10]	Transformer	33082.8	0.0532
STAR ^[9]	Transformer	964.9	0.0214
GroupNet ^[20]	HGNN	3139.7	0.0416
RDGCN ^[24]	GNN	28.0	0.0025
WTGCN ^[19]	GNN	541.0	0.0416
AGMGCN	GNN	23.9	0.0023

2.4 消融实验

2.4.1 组件消融实验

为评估模型中各关键组件的贡献,本文构建了组件消融实验,通过分别移除不同模块以分析其对模型预测性能的影响,实验结果如表4所示。在表4中,w/o AGM表示移除自适应群体掩码模块,不再对群体内外交互进行显式区分和独立建模;w/o TFC表示移除时频域卷积模块,不再同时在时域与频域采用非对称三维卷积增强交互建模;w/o FFM表示模型采用相加替代特征融合模块。表4实验结果表明,

AGMGCN 的各模型组件均能够为模型带来预测性能的提升。自适应群体掩码模块能够显式建模群体内部与群体外部的差异化社会交互关系,有效提升模型对行人复杂社会交互关系的刻画能力。自适应群体掩码模块能够进一步降低模型 8% 的 ADE 和 10% 的 FDE。时频域卷积模块在时域和频域分别采

用非对称三维卷积进行处理,进一步挖掘行人潜在时空依赖关系,从而为模型在 ADE 和 FDE 评价指标上带来 8% 的综合预测性能提升。特征融合模块能够自适应融合群内交互特征和群外交互特征,避免简单特征叠加带来的信息冗余和相互干扰。特征融合模块的使用能够进一步降低模型 4% 和 5% 的 ADE 和 FDE。

表 4 模型组件消融实验

Table 4 Model component ablation experiments

	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均值
w/o AGM	0.45/0.70	0.17/0.28	0.23/0.42	0.19/0.34	0.14/0.25	0.24/0.40
w/o TFC	0.45/0.69	0.17/0.28	0.23/0.42	0.19/0.33	0.14/0.24	0.24/0.39
w/o FFM	0.43/0.68	0.17/0.27	0.23/0.42	0.17/0.31	0.14/0.24	0.23/0.38
AGMGCN	0.41/0.63	0.16/0.26	0.22/0.40	0.17/0.30	0.13/0.23	0.22/0.36

注:值越低越好。

2.4.2 自适应群体掩码模块消融实验

自适应群体掩码模块能够生成群内和群外掩码矩阵,进而实现行人群内交互关系和群外交互关系的独立建模。为评估自适应群体掩码模块中各组成部分的贡献,本文构建了自适应群体掩码模块消融实验,结果如表 5 所示。在表 5 中,w/o IM 和 w/o OM 分别表示移除群内掩码矩阵和群外掩码矩阵,而 w/o V 和 w/o P 分别表示去除速度和位置相似度建模。群内掩码矩阵通过屏蔽群体内部行人,使模型能够独立建模群外行人之间的社交交互关系。因此,移除群内掩码矩阵会削弱模型对群外交互的刻画能力,导致模型

ADE 和 FDE 分别增加 4% 和 8%。而群外掩码矩阵能够屏蔽群外行人,有助于模型独立学习群体内部行人之间的交互关系,从而更准确地判断群体整体运动趋势。群外掩码矩阵的使用能够进一步提升模型在 ADE 和 FDE 评价指标上 4% 和 3% 的预测性能。此外,模型联合使用行人速度和空间位置信息的相似度学习行人之间的关联程度,有助于更加准确地地区分群体内外行人,为后续群体内外交互关系的差异化建模提供支持。实验结果表明,额外使用行人速度和空间位置信息进行相似度建模,可有效区分群体内外行人,从而进一步提升模型的整体预测精度。

表 5 自适应群体掩码模块消融实验

Table 5 Ablation study of the adaptive group masking module

	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均值
w/o IM	0.43/0.70	0.17/0.27	0.22/0.40	0.18/0.33	0.13/0.24	0.23/0.39
w/o OM	0.42/0.64	0.17/0.27	0.23/0.42	0.18/0.30	0.13/0.24	0.23/0.37
w/o V	0.42/0.68	0.18/0.30	0.22/0.40	0.18/0.31	0.13/0.23	0.23/0.38
w/o P	0.42/0.68	0.17/0.27	0.23/0.41	0.17/0.30	0.13/0.23	0.22/0.38
AGMGCN	0.41/0.63	0.16/0.26	0.22/0.40	0.17/0.30	0.13/0.23	0.22/0.36

注:值越低越好。

2.4.3 时频域卷积模块消融实验

时频域卷积模块分别在时域和频域中采用非对称三维卷积,进一步学习行人时空交互关系,以增强模型交互建模能力。为评估时频域卷积模块中各关键组件的贡献,本文构建了相应的消融实验。时频域卷积模块消融实验结果如表 6 所示。表 6 中,w/o TC 和 w/o FC 分别表示移除时域卷积建模和频域卷积建模,w/o AC 和 w/o 3DC 分别表示不采用非对称架构和三维卷积建模。时域卷积建模能够学习行人主被动行为和时空交互关系,提升模型在 ADE 和 FDE 评价指标上 4% 和 5% 的预测性能。而频域卷积建模通过

对各小波分量进行处理,进一步从频域层面学习行人全局和细节时空交互关系,有效弥补了单一时域建模对复杂动态交互刻画的不足,提升模型预测性能。因此,移除频域卷积建模会导致模型的 FDE 增加 5%。非对称架构采用不同卷积核的非对称三维卷积,分别作用于邻接矩阵的行切片和列切片,从而帮助模型学习行人主被动交互关系,而三维卷积则在时间维和空间维上同时进行建模,能够帮助模型学习行人时空交互关系。因此,移除非对称架构和三维卷积都会造成模型在 ADE 和 FDE 评价指标上 4% 和 5% 预测性能的损失。

表 6 时频域卷积模块消融实验

Table 6 Ablation study of the time-frequency domain convolution module

	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均值
w/o TC	0.43/0.67	0.17/0.28	0.22/0.40	0.18/0.31	0.14/0.24	0.23/0.38
w/o FC	0.41/0.66	0.16/0.27	0.23/0.41	0.18/0.31	0.13/0.23	0.22/0.38
w/o AC	0.43/0.68	0.16/0.28	0.23/0.42	0.18/0.31	0.13/0.22	0.23/0.38
w/o 3DC	0.43/0.67	0.17/0.28	0.23/0.41	0.17/0.30	0.13/0.23	0.23/0.38
AGMGCN	0.41/0.63	0.16/0.26	0.22/0.40	0.17/0.30	0.13/0.23	0.22/0.36

注:值越低越好。

2.5 定性分析

本文将所提出方法与现有方法 SGCN、RDGCN 和 DSTIGCN 进行预测轨迹可视化对比,如图 6 所示。在图 6①中,场景中群体与个体共同移动。SGCN 与 RDGCN 未能对群体内部交互进行有效建模,其在群体运动判断上出现明显偏差。DSTIGCN 通过引入可变形时空卷积处理行人运动,但仍未充分考虑群体内行人的交互关系。而本文方法通过引入群外掩码有效捕捉群体内部交互,因此本文方法的预测轨迹相比于 DSTIGCN 更接近真实轨迹。在图 6②中,一组行人以群体形式移动,其路径前方存在静止行人干扰。该场景需要模型能够准确区分群体内外交互,从而减弱静止行人的干扰并准确学习行人群体运动趋势。SGCN、RDGCN 和 DSTIGCN 均因未独立建模群体内外交互关系而出现明显的预测偏差,本文方法凭借对群体内外交互的分离建模,实现了更准确的轨迹预测。图 6③展示了群体与个体相遇情况。SGCN、RDGCN 与 DSTIGCN 因无法有效处理群体内外交互关系,模型预测轨迹出现明显偏差。相比之下,本文所提出的方法通过群内与群外掩码矩阵实现对群体内外交互关系的独立建模,能够更准确地判断群体与个体的未来运动轨迹,因此取得更准确的轨迹预测结果。图 6④给出了多人群混合运动的复杂场景,本文提出方法通过对群内与群外交互进行差异化建模,在多人群混合运动场景中仍能保持稳定的预测性能。

为验证模型能否准确区分群体内外行人,本文对不同场景下模型生成的群内和群外掩码矩阵进行了可视化分析,如图 7 所示,其中矩阵中的数字表示对应的行人,例如,1 表示行人 P1。群外掩码矩阵用于屏蔽群体外部行人的影响,群内掩码矩阵则用于屏蔽群体内部行人的影响。在图 7①中,模型生成的群外掩码矩阵能够有效屏蔽来自群外行人 P1 的影响,而群内掩码矩阵准确地对群体内部行人 P2 和 P3 之间的关系进行了屏蔽。在图 7②中,所提出方法能够有效区分运动群体和前方静止群体,并据此生成相应掩码矩阵,分别学习群体内外交互关系。图 7③和④进

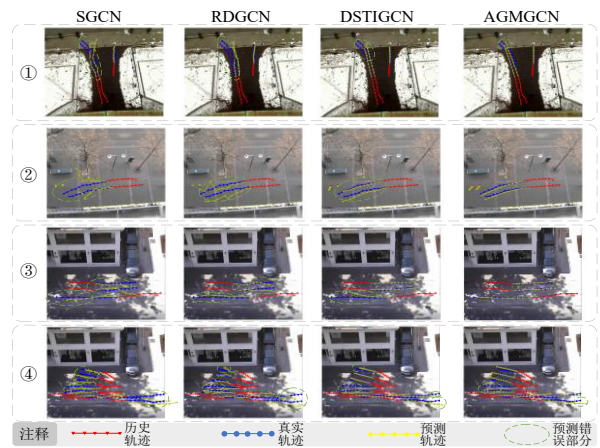


图 6 不同方法的轨迹预测可视化

Figure 6 Trajectory prediction visualization using different methods

一步表明,所提出的 AGMGCN 能够准确识别行人和群体相遇情况,并构建恰当的掩码矩阵。这些可视化结果表明,所提出方法能够准确识别群体内外行人,并生成对应掩码矩阵精确屏蔽群体内外交互,进而提升了群体行为建模的合理性。

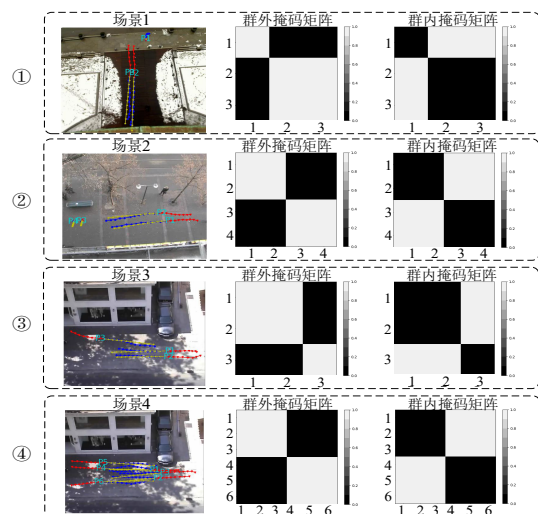


图 7 群内和群外掩码矩阵可视化

Figure 7 Visualization of in-group and out-group masking matrices

3 结束语

本文提出了一种基于自适应群体掩码图卷积网络的行人轨迹预测模型 AGMGCN。该模型能够独立建模行人群内交互和群外交互,准确描述群体复杂交互关系,完成行人未来轨迹的准确可靠预测。为准确划分群体内外行人,本文设计了一个自适应群体掩码模块,基于行人速度和位置信息计算个体间的特征相似度,并通过轻量化卷积模块生成阈值,进而通过阈值处理生成群内和群外掩码矩阵,为后续群内和群外交互的独立建模提供支持。同时,模型构建社会图并通过自注意机制和时频域卷积模块获得时频交互矩阵,用于描述行人社会交互关系。随后,将群内外掩码矩阵与时频交互矩阵相结合,并采用图卷积完成群体内部紧密交互和群体外部潜在交互的捕获。最后,通过特征融合模块整合群体内外的交互特征,并采用时间卷积网络完成行人未来轨迹的预测。定量分析和消融实验验证了所提方法的优越性以及其各组成部分的有效性。后续定性分析进一步表明,AGMGCN 比现有方法更适合群体运动的预测。

由于 AGMGCN 专注于群体交互的建模,忽略了场景物理结构对行人运动的约束。未来研究将进一步探索社会交互与场景上下文信息的联合建模方法,如通过引入表示场景语义信息或障碍物的特殊节点,与行人节点共同构建异构交互图,以实现行人社会交互与场景物理约束的联合学习,从而进一步提升模型在复杂真实场景中的预测准确度与合理性。

参考文献

- [1] 李勇明,李文正,张小恒,等. 基于分级包络域适应的行人轨迹预测模型[J]. 电子学报, 2025, 53(4): 1308-1321.
Li Yongming, Li Wenzheng, Zhang Xiaoheng, et al. The pedestrian trajectory prediction model based on hierarchical envelope domain adaptation[J]. Acta Electronica Sinica, 2025, 53(4): 1308-1321. (in Chinese)
- [2] Bhaskara R, Viswanath H, Bera A. Trajectory prediction for robot navigation using flow-guided Markov neural operator[C]//2024 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2024: 15209-15216.
- [3] 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.
- [4] 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 Wei, Liu Yun, Li Hui, et al. Survey of pedestrian trajectory prediction methods based on deep learning[J]. Control and Decision, 2021, 36(12): 2841-2850. (in Chinese)
- [5] 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. Piscataway: IEEE, 2016: 961-971.
- [6] 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. Piscataway: IEEE, 2019: 12077-12086.
- [7] 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. Piscataway: IEEE, 2018: 2255-2264.
- [8] Vemula A, Muelling K, Oh J. Social attention: Modeling attention in human crowds[C]//2018 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2018: 4601-4607.
- [9] Yu C J, Ma X, Ren J W, et al. Spatio-temporal graph transformer networks for pedestrian trajectory prediction[C]//Computer Vision - ECCV 2020. Cham: Springer, 2020: 507-523.
- [10] Giuliari F, Hasan I, Cristani M, et al. Transformer networks for trajectory forecasting[C]//2020 25th International Conference on Pattern Recognition. Piscataway: IEEE, 2021: 10335-10342.
- [11] Yuan Y, Weng X S, Ou Y L, et al. AgentFormer: Agent-aware transformers for socio-temporal multi-agent forecasting[C]//2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 9793-9803.
- [12] Sang H F, Chen W X, Zhao Z S. Review of pedestrian trajectory prediction based on graph neural networks[J]. Information Fusion, 2026, 127: 103727.
- [13] 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. Piscataway: IEEE, 2019: 6271-6280.
- [14] 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. Piscataway: IEEE, 2020: 14412-14420.
- [15] 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 Pat-

- tern Recognition. Piscataway: IEEE, 2021: 8990-8999.
- [16] Chen W X, Sang H F, Wang J Y, et al. IMGCN: Interpretable masked graph convolution network for pedestrian trajectory prediction[J]. *Transportmetrica B: Transport Dynamics*, 2024, 12: 2389896.
- [17] Chen W X, Sang H F, Wang J Y, et al. DSTIGCN: Deformable spatial-temporal interaction graph convolution network for pedestrian trajectory prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2025, 26(5): 6923-6935.
- [18] Cao D F, Li J C, Ma H B, et al. Spectral temporal graph neural network for trajectory prediction[C]//2021 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2021: 1839-1845.
- [19] Chen W X, Sang H F, Wang J Y, et al. WTGCN: Wavelet transform graph convolution network for pedestrian trajectory prediction[J]. *International Journal of Machine Learning and Cybernetics*, 2024, 15(12): 5531-5548.
- [20] Xu C X, Li M S, Ni Z Y, et al. GroupNet: Multiscale hypergraph neural networks for trajectory prediction with relational reasoning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 6488-6497.
- [21] Xu C X, Wei Y X, Tang B H, et al. Dynamic-group-aware networks for multi-agent trajectory prediction with relational reasoning[J]. *Neural Networks*, 2024, 170: 564-577.
- [22] Zhou L, Zhao Y L, Yang D Y, et al. GCHGAT: Pedestrian trajectory prediction using group constrained hierarchical graph attention networks[J]. *Applied Intelligence*, 2022, 52(10): 11434-11447.
- [23] Bengio Y, Léonard N, Courville A. Estimating or propagating gradients through stochastic neurons for conditional computation[PP/OL]. V1. arXiv. (2013-08-15) [2026-01-16]. <https://arxiv.org/abs/1308.3432>.
- [24] Sang H F, Chen W X, Wang J Y, et al. RDGCN: Reasonably dense graph convolution network for pedestrian trajectory prediction[J]. *Measurement*, 2023, 213: 112675.
- [25] Uhlemann N, Fent F, Lienkamp M. Evaluating pedestrian trajectory prediction methods with respect to autonomous driving[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(10): 13937-13946.
- [26] 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. Piscataway: IEEE, 2009: 261-268.
- [27] Lerner A, Chrysanthou Y, Lischinski D. Crowds by example[J]. *Computer Graphics Forum*, 2007, 26(3): 655-664.
- [28] 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.
- [29] Monti A, Bertugli A, Calderara S, et al. DAG-net: Double attentive graph neural network for trajectory forecasting[C]//2020 25th International Conference on Pattern Recognition. Piscataway: IEEE, 2021: 2551-2558.
- [30] Mohamed A, Zhu D, Vu W, et al. Social-implicit: Rethinking trajectory prediction evaluation and the effectiveness of implicit maximum likelihood estimation[M]//Computer vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 463-479.

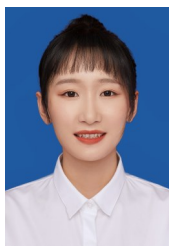
作者简介



陈旺兴 男,1998年3月出生于江西省抚州市。现为沈阳工业大学博士研究生。主要研究方向为轨迹预测、智能驾驶。
E-mail: chenwangxing@smail.sut.edu.cn



桑海峰 男,1978年1月出生于辽宁省沈阳市。现为沈阳工业大学教授、博士生导师。主要研究方向为机器视觉检测技术和智能视频分析技术。
E-mail: sanghaif@163.com



刘晴 女,1994年3月出生于辽宁省沈阳市。现为沈阳工业大学讲师。主要研究方向为机器视觉检测与图像识别、人工智能与医药交叉领域。
E-mail: Lqing0304@126.com