doi:10.12068/j.issn.1005－3026.2024.11.002

基于图卷积和卷积的行人轨迹预测算法

冯 昂[1]，宫 俊[1]，王 念[2]，王景龙[3]
（1.东北大学 信息科学与工程学院，辽宁 沈阳 110819；2.东风汽车集团有限公司 技术中心，湖北 武汉 430000；
3.燕山大学 电气工程学院，河北 秦皇岛 066004）

摘 要：行人轨迹预测取得了重要的进展，但现有的方法大多会受限于有限的车载计算资源，如何在自动驾驶车辆上实现高效的行人轨迹预测仍然存在着不足．针对该问题，提出了一种轻量化的行人轨迹预测算法，使用卷积神经网络（convolutional neural network，CNN）和图卷积神经网络（graph convolutional neural network，GCN）来处理和融合多模态信息．首先基于CNN设计了多尺度特征处理模块，使用多个卷积模块捕获行人轨迹和场景信息在不同时间和空间尺度上的特征；然后基于GCN构造特征融合模块，用于高效地建立轨迹和场景特征之间的时空关系并获得多个预测表示，最后融合多个预测表示以获得行人轨迹预测结果．在PIE和JAAD数据集上的实验表明，所提方法在仅用最少网络参数的情况下取得了最佳的预测性能，验证了所提方法的有效性；对比先前最轻量化的方法，参数优化了73%.

关 键 词：自动驾驶；行人轨迹预测；图卷积；多尺度；轻量化模型
中图分类号：TP 391 **文献标志码**：A **文章编号**：1005-3026(2024)11-1529-09

Pedestrian Trajectory Prediction Algorithm Based on Graph Convolution and Convolution

FENG Ang[1], *GONG Jun*[1], *WANG Nian*[2], *WANG Jing-long*[3]
(1. School of Information Science & Engineering, Northeastern University, Shenyang 110819, China; 2. Technology Center, Dongfeng Motor Group Co., Ltd., Wuhan 430000, China; 3. College of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China. Corresponding author: GONG Jun, E-mail: gongjun@ise.neu.edu.cn)

Abstract: Significant progress has been made in pedestrian trajectory prediction, but most of the existing methods are constrained by the limited on-board computing resources. How to achieve efficient pedestrian trajectory prediction in autonomous vehicles is still insufficient. To solve this problem, a lightweight pedestrian trajectory prediction algorithm is proposed, which uses convolutional neural network (CNN) and graph convolutional neural network (GCN) to process and integrate multimodal information. Firstly, a multi-scale feature processing module is designed based on CNN. Multiple convolution modules are used to capture the features of pedestrian tracks and scene information at different time and spatial scales. Then, a feature integration module is constructed based on GCN, which is used to efficiently integrate the spatial-temporal relationship between trajectory and scene features and obtain multiple prediction representations. Finally, multiple prediction representations are integrated to obtain pedestrian trajectory prediction results. Experiments on PIE and JAAD datasets show that the proposed method achieves competitive and optimal prediction performance with the least network parameters, respectively, which verifies the effectiveness of the proposed method. Compared with the previous lightest method, the parameters are optimized by 73%.

Key words: autonomous driving; pedestrian trajectory prediction; graph convolution; multi-scale features; lightweight model

自动驾驶领域中存在着许多挑战性的任务，如车辆轨迹预测[1]、目标分类[2]、行人行为分

收稿日期：2023-06-05
基金项目：国家自然科学基金资助项目(61871106).
作者简介：冯 昂(1999-)，男，海南三亚人，东北大学硕士研究生；宫 俊(1972-)，男，江苏徐州人，东北大学副教授，硕士生导师.

割[3]、行人和骑行者轨迹预测等.其中,行人轨迹预测需要通过考虑潜在的影响因素来推断出行人未来的轨迹.早期的行人轨迹预测研究[4-6]主要使用第三人称视角数据集,具有观测视角固定、场景信息多等特点,但这些研究难以直接应用在自动驾驶车辆上.随着近年来自动驾驶技术的发展,从以移动车辆为中心的视角去预测行人未来轨迹的研究被广泛关注[7],这能够直接有效地帮助车辆提前了解行人未来的意图并采取有效的措施,避免造成无法挽回的损失.

道路上的行人轨迹变化是复杂且难以确定的,这是由于轨迹运动会受到许多因素的影响,如当行人正在马路上行走时,可能因为与其他行人位置冲突而停止;当行人在人行道上穿越马路时,可能因为车辆转弯而停止等.因此,先前的一些研究仅仅考虑行人历史轨迹的变化是不合理的,还需要结合实际场景中的各种信息,如车辆速度、行人位置、行人数量等.近年来,许多学者不断地丰富移动车辆平台上的传感器,基于这些传感器捕获到的实际场景信息,提出更加丰富的多模态数据集[8-10],然后基于多模态信息完成行人轨迹预测任务,从而使行人轨迹预测任务更加的可靠.Cao等[11]结合骑行者相关的运动、自我车辆运动和环境特征,同时又考虑骑行者意图和环境约束的在线轨迹预测方法,通过考虑现实中的多模态因素能够准确、快速和合理地完成预测任务.这意味着该预测领域未来的发展趋势是通过多模态信息预测行人未来的轨迹.但是由于移动车辆上的计算资源十分有限,计算复杂度高的方法在车辆平台上需要过多的运行时间,这在实际场景中可能已经发生不可挽回的交通事故.因此如何搭建一个高效、轻量化的模型来处理多模态信息并实现准确的行人轨迹预测是该领域一个重要的研究问题.

近年来,基于CNN[12-13],循环神经网络(RNN)[14-15]和Transformer[16-18]等神经网络的深度学习算法已经在行人轨迹预测领域取得可观的进展.然而,由于感受野的限制,CNN难以对长期依赖性的信息进行建模;RNN在提取序列中局部特征的能力上往往存在着缺点,这可能导致有时包含预测未来轨迹的关键线索丢失;Transformer虽然具有高效的融合多模态信息的能力,但会花费更大的计算资源和运行时间.因此,许多学者基于深度学习技术提出了不同的预测算法,以提高行人轨迹预测的性能.Styles等[13]使用FlowNet2-CSS获得场景图片中的光流特征,然后基于ResNet来计算过去场景信息的光流特征并与历史轨迹融合和输出预测结果.Rasouli等[10]将预测任务分为三个部分,首先使用多个卷积层(Conv)和长短时记忆网络(LSTM)模块预测未来的行人意图和车辆速度,然后使用上述未来特征与历史轨迹来预测未来的轨迹.Yin等[18]为了弥补先前方法中多模态特征融合能力的不足,同时实现模型轻量化,仅使用Transformer架构实现轨迹预测,通过Self-Attention和Cross-Attention来建立特征与自身和与其他特征在不同位置上的上下文关系.Yang等[19]在GCN中引入了一个空间变换器,用于建模动态和静态的空间相关性,然后使用Temporal Conv和Transformer分别处理长期时间依赖性和融合时空特征.

尽管上述方法已经实现优越的预测性能,但是如何在计算资源有限的移动车辆上准确地实现行人轨迹预测仍然是一个备受关注的问题.近年来图卷积网络被引入到轨迹预测领域,并取得了显著的成果[20-21].轨迹预测任务中不同行人之间的空间关系对于预测任务非常重要,GCN首次出现并被用于捕捉物体之间的空间布局关系,同时通过在图中传递信息,还可以用于捕捉不同物体之间的互动关系.另一方面,轨迹预测领域中一个重要的研究问题是轨迹的不确定性,通过使用GCN可以对不确定性进行建模,使模型输出预测实现概率分布,而不仅仅是单一的预测结果.这有助于更好地理解预测结果的可信度,并且GCN还可以用于高效融合不同模态的数据,从而提供更全面的信息来进行预测.GCN方法通过将图节点信息与其相邻节点信息进行加权聚合,大大优化了参数尺寸、推理速度和信息融合效率.Mohamed等[20]引入用于学习时空信息的时空GCN,通过对社交图中的节点进行卷积操作,可以从历史轨迹中提取特征,从而捕捉人类行为的动态变化和时空关系.Cadena等[21]提出基于图卷积网络的行人过马路预测模型,通过使用GCN架构从行人的历史行为和交通环境提取特征,可以捕捉到行人的行为模式和交通环境的动态变化,从而快速、准确地完成预测任务.

针对上述内容,本文提出了一个轻量化的行人轨迹预测网络模型.模型基于CNN模块设计了多尺度特征处理模块,通过设定不同的输入输出尺寸将时间序列数据在时间和空间尺度上进行

融合和压缩,以获得不同尺度上的轨迹特征和感受野下的场景特征,同时使用全局特征以避免因CNN邻接运算导致的关键信息缺失;然后基于GCN设计了多尺度特征融合模块,将不同时间尺度和空间尺度上的轨迹特征与场景特征一一对应并融合以得到不同感受野的预测表示;最后将多个预测表示进行二次融合得到最终的预测结果.本文模型在特征融合模块中使用两个GCN模块,第一个GCN融合不同尺度的轨迹特征和场景特征,以建立轨迹与场景之间的信息交互关系,第二个GCN模块用于增强模型对瞬态变化信息中时序变化的理解同时调整输出格式,并减少非必要参数的使用,避免了CNN,RNN和Transformer架构多模态特征融合存在的不足.

1　基于GCN和CNN的行人轨迹预测算法

本文提出的行人轨迹预测网络架构包含多尺度轨迹特征提取模块和多尺度场景特征提取模块和多尺度特征融合模块.算法整体架构如图1所示.

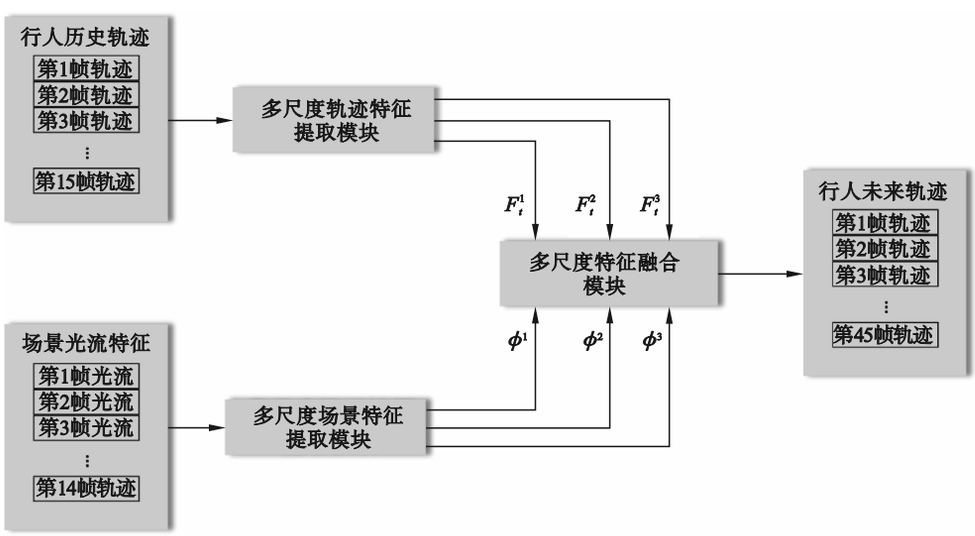

图1　算法整体架构

Fig. 1　Overall architecture of the algorithm

1)多尺度轨迹特征提取模块包含两个卷积模块,每个卷积模块的输出作为每个感受野下的局部特征,与轨迹编码层输出的全局特征组成多尺度轨迹特征.

2)场景特征提取模块包含3个卷积模块,每个卷积模块会对场景编码特征下采样,获得不同感受野下的全局和局部场景特征.

3)多尺度特征融合模块使用嵌套GCN模块融合不同感受野下的轨迹与场景的全局特征和局部特征并输出预测结果.

1.1　多尺度轨迹特征提取模块

为了优化模型的尺寸和特征提取能力,本文基于卷积模块设计了多尺度轨迹特征提取模块,提升了模型推理效率.通过设置不同卷积参数,帮助模型获取轨迹信息中不同时间尺度和空间尺度上有效的全局特征和局部特征.

如图2所示,轨迹信息$T_{\mathrm{h}} \in \mathbf{R}^{t_{\mathrm{h}} \times 4}$首先经过编码层中的MLP层,与一个和编码尺寸相关的常量系数δ相乘,并经过Relu激活函数来剔除无效信息后,能够得到全局特征$F_t^1 \in \mathbf{R}^{t_{\mathrm{h}} \times c}$;然后全局特征$F_t^1$分别经过两个不同设定参数的卷积模块对其进行下采样操作,以获得不同时空尺度上的局部特征$F_t^2 \in \mathbf{R}^{t_{\mathrm{h}} \times c}$和$F_t^3 \in \mathbf{R}^{i_{\mathrm{h}} \times c}$;与编码层一样,为了剔除冗余信息和在特征融合时保证特征矩阵尺寸一致,本文在每个卷积模块后均使用MLP和Relu层.多尺度轨迹特征提取模块定义为

$$F_t^1 = \mathrm{Relu}\left(\delta \mathrm{MLP}\left(T_{\mathrm{h}}\right)\right), \tag{1}$$

$$F_t^2 = \mathrm{Relu}(\mathrm{MLP}(\mathrm{Conv}(F_t^1))), \tag{2}$$

$$F_t^3 = \mathrm{Relu}(\mathrm{MLP}(\mathrm{Conv}(F_t^1))). \tag{3}$$

其中:t_{h}是历史轨迹长度;c是编码尺寸;$\delta = \sqrt{c}$;第一层卷积的输入尺寸和输出尺寸与历史观测长度t_{h}一致为15;第二层卷积的输入尺寸为15,输出尺寸i_{h}为7.第一个卷积模块的卷积核为2,第二个卷积模块的卷积核为4,其余卷积模块参数设定相同,步长为2,填充层为0,均使用可学习的偏置矩阵.

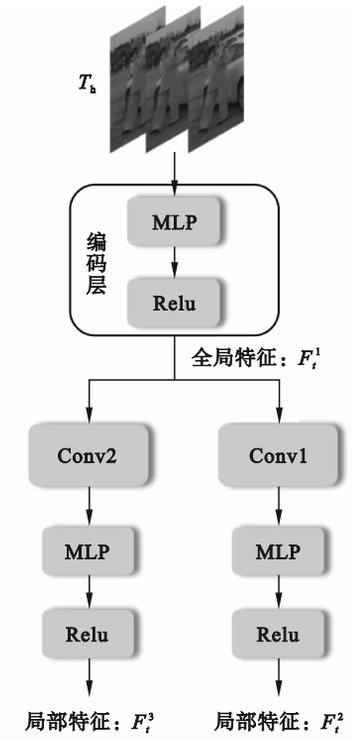

图2 多尺度轨迹特征提取模块

Fig. 2 Multi-scale trajectory feature extraction module

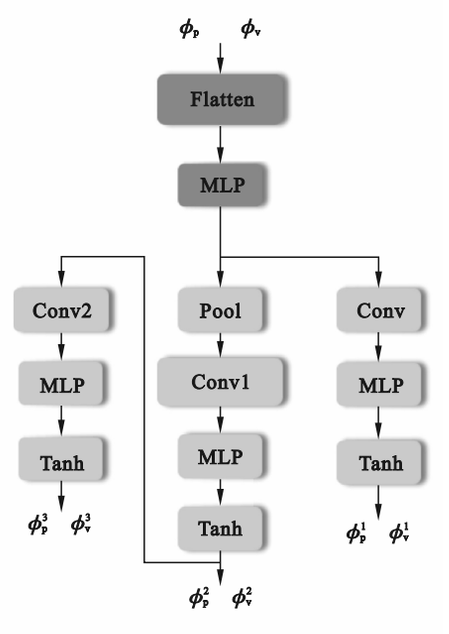

图3 多尺度场景特征提取模块

Fig. 3 Multi-scale scene feature extraction module

1.2 多尺度场景特征提取模块

与传统方法中将每一帧的场景图片作为输入信息不同，本文采用的移动车辆视角中包含有车辆、行人高度动态变化信息的光流数据.光流数据的中心区域和行人边界框区域分别表示移动车辆和行人的运动变化情况，因此引入行人光流和车辆光流能够补偿实际动态场景中因车运动使得观察视角快速变化而丢失的瞬态信息.

为捕获场景光流信息中有效的全局特征和局部特征，本文采用多个卷积模块构造多尺度场景特征提取模块，并且行人光流信息和车辆光流信息均使用相同的特征提取模块.如图3所示，为了保证输入进模块的光流信息的尺寸相匹配，本文首先对行人光流 $\boldsymbol{\phi}_{\mathrm{p}} \in \mathbf{R}^{t_\phi \times h_{\mathrm{p}} \times c_{\mathrm{p}}}$ 和车辆光流 $\boldsymbol{\phi}_{\mathrm{v}} \in \mathbf{R}^{t_\phi \times h_{\mathrm{v}} \times c_{\mathrm{v}}}$ 作了展平操作，然后经过MLP层将行人与车辆光流特征的尺寸调整为 $\dot{\boldsymbol{\phi}}_{\mathrm{v}}, \dot{\boldsymbol{\phi}}_{\mathrm{p}} \in \mathbf{R}^{t_\phi^1 \times c_\phi}$.以行人光流为例，接下来的输出特征 $\dot{\boldsymbol{\phi}}_{\mathrm{p}}$ 会依次经过Conv，MLP和激活函数Tanh得到行人光流特征 $\boldsymbol{\phi}_{\mathrm{p}}^1 \in \mathbf{R}^{t_\phi^1 \times c}$，其中Conv在不改变时间尺度的前提下，融合不同空间位置的信息以获得全局场景特征 $\boldsymbol{\phi}_{\mathrm{p}}^1$；与轨迹特征提取局部特征不同，光流局部特征提取采取了嵌套卷积的方式，$\dot{\boldsymbol{\phi}}_{\mathrm{p}}$ 首先经过Max-pooling层，在不改变时间尺度的情况下保留关键的局部空间信息和降低计算复杂度，然后依次经过Conv1，MLP和Tanh作下采样操作，并获得行人光流特征 $\boldsymbol{\phi}_{\mathrm{p}}^2 \in \mathbf{R}^{t_\phi^2 \times c}$，其中第2层卷积模块用于融合局部特征的不同空间位置信息并获得第1层行人局部光流特征 $\boldsymbol{\phi}_{\mathrm{p}}^2$；然后，$\boldsymbol{\phi}_{\mathrm{p}}^2$ 依次经过Conv2，MLP和Tanh作第二次下采样操作以获得第2个行人光流特征 $\boldsymbol{\phi}_{\mathrm{p}}^3 \in \mathbf{R}^{t_\phi^3 \times c}$，其中Conv2会挤压时间尺度，在时空尺度上提取场景特征，与Conv和Conv1结合能够最大限度地保留光流信息中有效的全局和局部特征.为了与不同感受野下的轨迹特征中的矩阵尺寸相匹配，同时保留每个节点信息，在每个卷积模块后都加入了MLP来调整特征的尺寸.多尺度场景特征提取模块定义为

$$\boldsymbol{\phi}_{\mathrm{p}}^1 = \mathrm{Tanh}\left(\mathrm{MLP}(\mathrm{Conv}(\boldsymbol{\phi}_{\mathrm{p}}))\right), \tag{4}$$

$$\boldsymbol{\phi}_{\mathrm{p}}^2 = \mathrm{Tanh}\left(\mathrm{MLP}\left(\mathrm{Conv}\left(\mathrm{Pool}\left(\boldsymbol{\phi}_{\mathrm{p}}\right)\right)\right)\right), \tag{5}$$

$$\boldsymbol{\phi}_{\mathrm{p}}^3 = \mathrm{Tanh}\left(\mathrm{MLP}\left(\mathrm{Conv}\left(\boldsymbol{\phi}_{\mathrm{p}}^2\right)\right)\right). \tag{6}$$

其中：历史光流信息长度 t_ϕ^1 为14；行人光流的尺寸 h_{p} 和 c_{p} 分别为9，2；c_ϕ 为128；车辆光流的尺寸 h_{v} 和 c_{v} 分别为64，2；Conv的输入尺寸和输出尺寸为14；Conv1的输入尺寸为14，输出尺寸 t_ϕ^2 为7；Conv2输入尺寸为7，输出尺寸 t_ϕ^3 为4. 3个卷积层的卷积核和步长均为2，填充层为0，均使用可学习的偏置矩阵.

本文通过使用Max-pooling层、多个卷积模块逐层对光流信息在时空尺度上作下采样操作,每层下采样能够获得有效的局部光流特征;同时为了避免在下采样过程中有效信息的丢失,在局部特征提取前加入了全局特征提取模块.在提取完行人光流和车辆光流的全局、局部特征后,将行人光流特征$\boldsymbol{\phi}_{\mathrm{p}}^{1},\boldsymbol{\phi}_{\mathrm{p}}^{2},\boldsymbol{\phi}_{\mathrm{p}}^{3}$与车辆光流特征$\boldsymbol{\phi}_{\mathrm{v}}^{1},\boldsymbol{\phi}_{\mathrm{v}}^{2},\boldsymbol{\phi}_{\mathrm{v}}^{3}$在第二维度上拼接便得到了3个感受野下的混合光流特征$\boldsymbol{\phi}^{1}\in\mathbf{R}^{(2t_{\varphi}^{1})\times c},\boldsymbol{\phi}^{2}\in\mathbf{R}^{(2t_{\varphi}^{2})\times c},\boldsymbol{\phi}^{3}\in\mathbf{R}^{(2t_{\varphi}^{3})\times c}$.

1.3　多尺度特征融合模块

GCN可用于拓扑图矩阵中的空间特征,基于GCN捕获每个场景节点与轨迹特征中其他节点的关系,从而快速融合不同时空尺度上的特征并建立轨迹与场景的上下文连接,从而大大降低模型的计算复杂度和提高模型的推理效率.因此本文基于GCN模块设计了多尺度特征融合模块.

如图4所示,轨迹特征F_t^n在转置操作后经MLP将特征尺寸调整至与预测步长t_{p}一致,并获得特征$\dot{F}_t^n$.然后场景光流特征$\boldsymbol{\phi}^n$与$\dot{F}_t^n$经过第一层GCN,能够捕获场景光流信息与行人轨迹在相同时间尺度上不同空间位置间的相互关系并得到混合表示$P_{\mathrm{f}}^n\in\mathbf{R}^{(2t_{\varphi}^n)\times y_{\mathrm{p}}}$;为进一步增强模型理解高速运动状态信息的能力,将混合表示P_{f}^n作转置操作后与场景光流特征$\boldsymbol{\phi}^n$输入进第二层GCN模相融合后获得不同时间尺度上的行人预测表示$P_{\mathrm{p}}^n\in\mathbf{R}^{t_{\mathrm{p}}c}$;最后将多尺度预测表示在第三维度拼接后,依次经过Relu,MLP以计算每个尺度预测表示与行人未来轨迹坐标之间的概率,并得到预测结果$T_{\mathrm{p}}\in\mathbf{R}^{4t_{\mathrm{p}}}$.每层GCN中输出的特征矩阵需要分别与可学习的偏置矩阵$\boldsymbol{M}\in\mathbf{R}^{t_{\mathrm{p}}}$和$\boldsymbol{N}\in\mathbf{R}^{c}$相加.多尺度特征融合模块定义:

$$P_{\mathrm{f}}^n=\mathrm{Relu}(\mathrm{GCN}(\boldsymbol{\phi}^n,\mathrm{MLP}(\dot{F}_n))+\boldsymbol{M}),\tag{7}$$

$$P_{\mathrm{p}}^n=\mathrm{Relu}(\mathrm{GCN}(\mathrm{Tr}(P_{\mathrm{f}}^n),\boldsymbol{\phi}^n)+\boldsymbol{N}),\tag{8}$$

$$T_{\mathrm{p}}=\mathrm{MLP}(\mathrm{Relu}(\mathrm{Concat}(P_{\mathrm{p}}^n))).\tag{9}$$

其中:$n\in(1,2,3)$;Concat表示矩阵拼接操作;GCN表示图卷积操作;Tr表示转置操作;预测步长t_{p}为45.

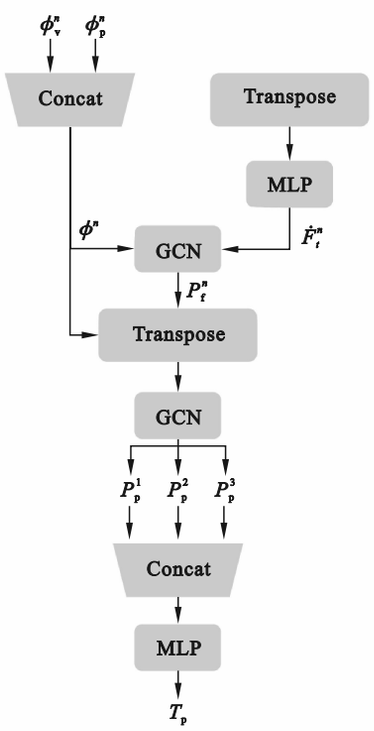

图4　多尺度特征融合模块

Fig. 4　Multi-scale feature integration module

本文使用两层GCN架构的融合方法能够仅依赖于基础的矩阵计算操作便实现快速融合轨迹信息和场景光流信息,同时建立二者在不同时间和空间尺度上的关系.由于轨迹信息自身性质使得特征尺寸很小,经过处理后的行人光流和车辆光流的尺寸也得到缩减,为了让模型将注意力均匀地分配到各个节点特征上,本文在使用GCN进行逐层特征融合时将边定义为全1矩阵,保证消息在逐层传递时模型不会降低对第1层信息的关注度.嵌套GCN可视化如图5所示,预测结果的任一节点关系如图6所示.与传统的RNN和Transformer不同,该方法既能够快速建立特征间的相互关系,又能大大降低模型的计算复杂度.

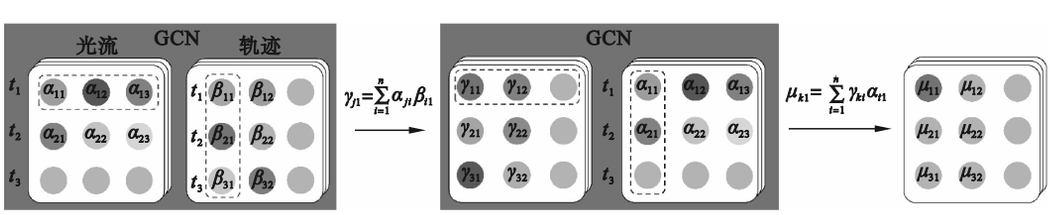

图5　多尺度特征融合可视化

Fig. 5　Visualization of multi-scale feature integration

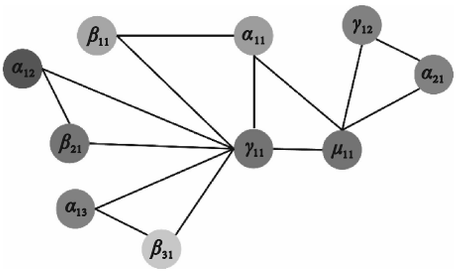

图6 预测节点关系图

Fig. 6 Forecast node relationship graph

2 实 验

2.1 数据集介绍

本文使用两个被广泛使用的第一人称行人轨迹预测数据集,分别是PIE和JAAD.它们以每秒30帧(frame per second,FPS)的采样速度录制以车辆自我中心为视角的车辆移动视频.PIE数据集是一个大型的第一人称视角驾驶数据集,从高清相机在加拿大的白天拍摄中获得,其中包括37个视频、1 842条人行道和909 480帧场景图片.JAAD数据集中包含2 856条人行道和346个视频片段,每个视频片段由200~400帧组成,共包含82 032帧场景图片和2 800个带注释的行人.

2.2 损失函数和评估指标

在训练过程中,本文使用均方误差(mean square error,MSE)来计算每个时间步长上预测的行人边界框与实际边界框之间的误差.MSE定义为

$$\mathrm{MSE}=\frac{1}{L}\sum_{t=1}^{L}\left\|\check{l}_t-l_t\right\|^2. \tag{10}$$

其中:$\check{l}_t$和l_t分别为t时刻的预测边界框和真实边界框;L表示预测长度.在测试过程中,本文还使用C_{MSE}和$\mathrm{CF}_{\mathrm{MSE}}$评估算法在空间位置误差计算和长期预测结果上的性能.C_{MSE}定义为

$$C_{\mathrm{MSE}}=\frac{1}{L}\sum_{t=1}^{L}\left\|\check{C}_t-C_t\right\|^2. \tag{11}$$

其中:$\check{C}_t$和C_t分别是t时刻的行人边界框的中心位置和实际边界框的中心位置.$\mathrm{CF}_{\mathrm{MSE}}$是最后预测步长上的中心位置预测误差.所有轨迹预测结果均以像素为单位.

2.3 实现细节

本文使用MTN研究作为数据集分割和实验结果对比的基准.所有实验都在RTX 2080上进行,使用Adam优化器来训练算法,学习率初始化为0.000 1,批量大小为128,编码尺寸c设定为64,c_ϕ尺寸为128.PIE和JAAD数据集中历史数据观测序列长度被设定为15帧(0.5 s),未来的预测长度被设定为45帧(1.5 s).本文所使用的场景光流信息[18]为研究中预处理后的特征,由于光流信息需要从两帧图像中捕获瞬态变化的特征,所以历史观测长度t_ϕ为14帧.将PIE数据集分割后能得到44 383个训练样本、10 214个验证样本和36 208个测试样本;JAAD数据集分割后能够得到14 155个训练样本、2 879个验证样本和5 124个测试样本.然后与DTP-MOF[13],PIEfull[9],STED[22],MTN[18]和SGnet[23]方法进行比较.

2.4 定量评估

为了验证模型的有效性,将本文方法行人轨迹预测结果与现有的研究方法相比.表1展示了本文模型和其他方法的对比结果,能够发现本文方法在PIE数据集上的MSE,C_{MSE}和$\mathrm{CF}_{\mathrm{MSE}}$上取得具有竞争力的性能;在JAAD数据集上,该模型在空间定位能力和长期预测能力中都展示出了比现有方法更优越的性能.在模型参数方面,能够发现该模型比参数最少的方法MTN减少73%,比预测性能最好的方法SGnet减少99.5%,这表明该方法具有最轻量化的模型尺寸.在模型推理速度方面,使用相同的环境配置对PIE数据集的测试样本进行测试,为了更加明显地体现推理速度的差异,将测试批次设定为6 000,同时为保证测试的准确性,本文采用截尾均值来对比不同算法的推理耗时.

2.5 定性分析

图7中的4个场景分别表示不同算法在PIE和JAAD数据集上行人轨迹预测结果的可视化图,其中白色边框表示行人初始位置,Groundtruth边框表示行人最终边界框,其他边框是不同算法的预测结果,不同点表示连续时间内行人轨迹中心的变化情况.结果表明本文方法能够预测出不同场景下的行人未来的轨迹坐标.这证明所提方法具有很好的预测性能,并且在两个大型的行人轨迹预测数据集上具有一定的竞争力.图8的实验结果表明,本文算法在每个批次上的推理速度远低于MTN.图9和图10更加具体地展示了不同方法在PIE和JAAD数据集上每个时间步长上C_{MSE}的对比结果,能够发现本文方法在PIE数据集上45个时间步上具有一定的竞争力,而在JAAD数据集上展示出了比MTN方法更好的长期预测能力.

表1　与其他方法在PIE和JAAD上的结果对比

Table 1　Comparison with other methods on PIE and JAAD

算法	参数×10^{-6}	PIE			JAAD		
		MSE	C_{MSE}	CF_{MSE}	MSE	C_{MSE}	CF_{MSE}
DTP-MOF	11.3	665	566	2 373	1 158	1 014	4 143
PIE_{full}	3.07	551	512	2 232	—	—	—
STED	13.94	461	415	1 871	1 044	960	4 031
MTN	0.134	444	414	1 627	1 005	951	4 011
SGnet	7.622	442	413	1 761	1 049	996	4 076
本文	**0.036**	458	425	1 805	**994**	**943**	**3 821**

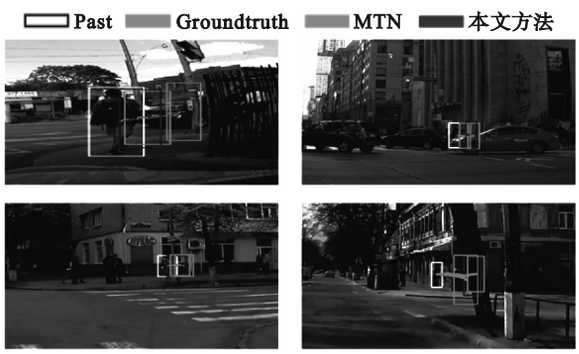

图7　行人轨迹预测成功的可视化图

Fig. 7　Visualization of successful pedestrian trajectory prediction

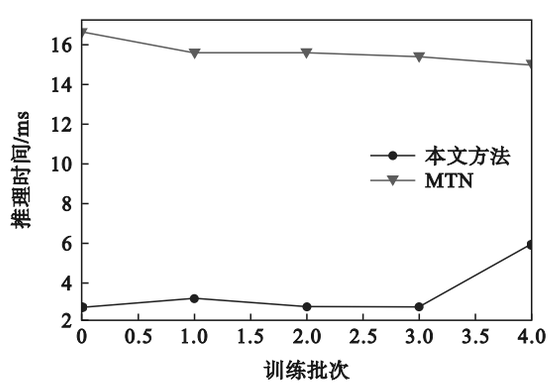

图8　推理时间在PIE的对比

Fig. 8　Comparison of inference time in PIE

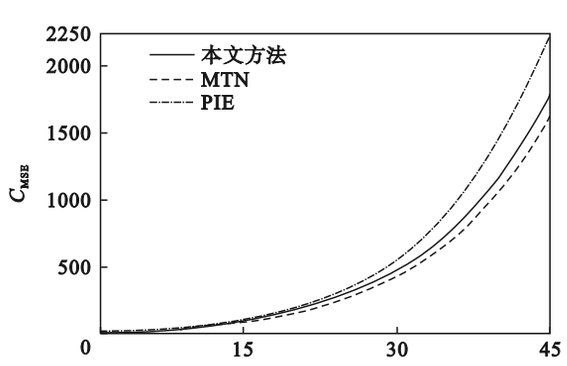

图9　不同时间步长在PIE的对比

Fig. 9　Comparison of different time steps in PIE

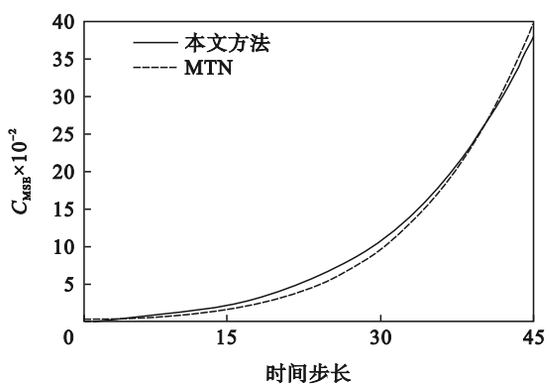

图10　不同时间步长在JAAD的对比

Fig. 10　Comparison of different time steps in JAAD

2.6　消融实验

为进一步分析本文所提方法各部分的有效性,接下来在PIE数据集上对模型各个部分设计了消融实验来对算法框架中的各个部分进行对比分析.

如表2所示,为验证不同感受野下轨迹、场景的全局特征和局部特征的作用.首先保持场景光流特征不变,将融合的3个轨迹特征依次替换为F_t^1,F_t^2,F_t^3,能够发现预测性能会逐渐提高,当使用全部轨迹特征时,预测性能达到最佳的458 MSE;然后保持轨迹特征不变,将融合使用的3个光流特征依次替换为$\boldsymbol{\phi}^1$,$\boldsymbol{\phi}^2$,$\boldsymbol{\phi}^3$,能够发现仅使用第2层光流特征时,预测指标能达到509 MSE,但比使用全部场景特征时的预测性能高51 MSE.可以看出,当轨迹特征和场景特征均使用不同感受野下的特征相融合时,本文方法会因有效获取每个感受野下的关键特征而取得优越的预测性能.

在消融实验中,本文在算法中进行了细小的改动以匹配模型各个模块的输入尺寸.当轨迹特征分别仅使用F_t^1和F_t^2特征时,将其与$\boldsymbol{\phi}^3$融合所使用的特征融合模块中MLP的输入尺寸修改为

15,输出尺寸保持为c;当仅使用F_t^3特征时,将其与ϕ^1和ϕ^2融合所使用的特征融合模块中MLP的输入尺寸修改为7,输出尺寸不变.

表2 PIE上消融实验的对比结果

Table 2 Comparison of ablation experiment results on PIE

轨迹特征	场景特征	MSE	C_{MSE}	$\mathrm{CF}_{\mathrm{MSE}}$
F_t^1,F_t^1,F_t^1	ϕ^1,ϕ^2,ϕ^3	793	751	2 872
F_t^2,F_t^2,F_t^2	ϕ^1,ϕ^2,ϕ^3	509	476	1 964
F_t^3,F_t^3,F_t^3	ϕ^1,ϕ^2,ϕ^3	499	467	1 944
F_t^1,F_t^2,F_t^3	ϕ^1,ϕ^2,ϕ^3	520	486	2 006
F_t^1,F_t^2,F_t^3	ϕ^1,ϕ^2,ϕ^3	509	475	1 997
F_t^1,F_t^2,F_t^3	ϕ^1,ϕ^2,ϕ^3	579	546	2 151
F_t^1,F_t^2,F_t^3	ϕ^1,ϕ^2,ϕ^3	458	425	1 805

2.7 预测失败案例

尽管本文的方法达到了具有竞争力的预测性能,但在高度动态场景下的预测结果仍然存在较大的偏差.图11展示了本文方法在PIE和JAAD数据集上的预测失败的4个场景,其中边框和边框分别表示在最后时间步长上本文算法的预测轨迹和真实的行人轨迹.在第1行PIE数据集的2个场景中,车辆均在保持直行,行人在历史时刻分别保持静止状态和向车辆移动的状态,此时的行人轨迹预测结果与实际轨迹偏差较大.在第2行JAAD数据集的2个场景中,车辆分别保持向左转弯和直行的运动状态,行人在历史时刻也在持续的运动,此时行人轨迹预测结果与实际轨迹之间存在较大的偏差.因此,本文算法在车辆保持运动的情况下,会因行人轨迹快速的运动状态使得模型预测结果偏差较大,但是仍然能够大致预测出行人未来的运动方向.

图11 预测失败的可视化

Fig. 11 Visualization of prediction failure

3 结 语

针对以移动车辆为中心视角下的行人轨迹预测中现有方法计算复杂度高、运行时间长等问题,本文提出一种基于CNN和GCN的轻量化的行人轨迹预测算法.基于CNN构造了多尺度特征提取模块,使用多个CNN模块来捕获时间序列在不同时间和空间尺度上的特征.基于GCN构造了多尺度特征融合模块,使用两层GCN模块来建立不同时空尺度上轨迹特征和轨迹场景特征的上下文关系并得到每个时空尺度的预测表示,最后融合多个预测表示得到最终的预测结果.实验结果表明,本文方法以最轻量化的模型尺寸在PIE和JAAD数据集上取得了有竞争力和较优的预测性能.

参考文献:

[1] Xie G T, Gao H B, Qian L J, et al. Vehicle trajectoryprediction by integrating physics-and maneuver-based approaches using interactive multiple models[J]. *IEEE Transactions on Industrial Electronics*, 2017, 65(7): 5999-6008.

[2] Gao H B, Cheng B, Wang J Q, et al. Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment[J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(9): 4224-4231.

[3] Gao H B, Lyu C, Zhang T, et al. A structure constraint matrix factorization framework for human behavior segmentation[J]. *IEEE Transactions on Cybernetics*, 2021, 52(12): 12978-12988.

[4] 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). Long Beach, 2019: 5718-5727.

[5] Kosaraju V, Sadeghian A, Martin R, et al. Social-BiGAT: multimodal trajectory forecasting using bicycle-GAN and graph attention networks [J]. *ArXiv e-Prints*, 2019: 1907.03395.

[6] 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). Long Beach, 2019: 1349-1358.

[7] 姚冲,周晖.基于时空图的行人多模态轨迹预测方法[J].计算机工程与设计,2022,43(10):2918-2925.
(Yao Chong, Zhou Hui. Pedestrian multi-modal traectory prediction method based on spatial-temporal graph [J]. *Computer Engineering and Design*, 2022, 43(10): 2918-2925.)

[8] Rasouli A, Kotseruba I, Tsotsos J K. Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior [C]// Proceedings of the IEEE International Conference on Computer Vision Workshops. Venice, 2017: 206-213.

(下转第1594页)