

DOI:10.19651/j.cnki.emt.2210534

基于时空自适应图卷积网络的跌倒检测算法*

刘鹏飞 李伟彤

(广东工业大学信息工程学院 广州 510006)

摘要: 针对现有图卷积网络(GCN)需要预先定义人体骨架拓扑图和模型较大的问题,提出了基于时空自适应图卷积网络(ST-AGCN)的跌倒检测算法。该网络包括3个部分:利用HRNet姿态估计算法从视频中提取人体骨架点序列,并预处理成四维张量;引入归一化嵌入式高斯函数通过学习(无需人工预定义)得到人体拓扑图,利用空间自适应图卷积获取人体关联特征;利用多尺度卷积提取时间运动特征,提高模型获取动态信息的能力。在公开数据集和自建数据集上分别进行仿真,准确率分别达95.45%和99.55%。结果表明,该算法优于目前GCN方法,参数量只有后者的1/4甚至更少。本文算法还可以适用于不同的数据集。

关键词: 计算机视觉;跌倒检测;图卷积网络;人体姿态估计;嵌入式高斯函数

中图分类号: TP394.1 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Fall detection algorithm based on spatial-temporal adaptive graph convolution network

Liu Pengfei Li Weitong

(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: To solve the problem that existing graph convolution network (GCN) need to pre-define human skeleton topology and the model is large, a fall detection algorithm based on spatiotemporal adaptive graph convolutional network (ST-AGCN) is proposed. The network consists of three parts: firstly, HRNet, a human pose estimation algorithm, is used to extract human skeleton points from video and preprocess them into four-dimensional tensor. Secondly, the normalized embedded Gaussian function is introduced to obtain the human body topology by learning (without manual pre-definition), and the human body correlation features are obtained by spatial adaptive graph convolution. Thirdly, multi-scale convolution is used to extract temporal motion features to improve the model's ability to obtain dynamic information. Simulations are carried out on public and self-built dataset, and the accuracy rates are 95.45% and 99.55%, respectively. The results show that the proposed algorithm is better than the current GCN methods, and the number of parameters is only a quarter of the latter, or even less. Another advantage of our algorithm is that it can be applied to different datasets.

Keywords: computer vision; fall detection; graph convolution network; human pose estimation; embedded gaussian function

0 引言

近年来,我国人口增速放缓,人口老龄化程度日益严重。跌倒是造成老年人伤害或死亡的最大威胁,因此对老人进行跌倒检测是很有必要的。

当前主要跌倒检测方法有两种:基于传感器的方法和基于计算机视觉的方法。基于传感器的方法主要利用加速度传感器^[1]或多传感器联合^[2]进行跌倒检测,其缺点是使

用者需时时佩戴、存在侵入感和无法长时间使用。

基于计算机视觉的方法具有无侵入感,可长期使用等优点。Carrier等^[3]先通过光流法,对每两帧图像中提取光流图像,再将其送入到卷积神经网络(convolutional neural networks, CNN)中进行分类判断。提取光流图像受背景、光照和角度影响比较严重,特征提取难度大,数据处理复杂,同时存在隐私问题。金彦亮等^[4]使用帧间差分法提取相邻两帧图像间的人物轮廓信息,再从连续3帧中提取像

收稿日期:2022-07-01

* 基金项目:广东省科技计划项目(2017A010101016)资助

素点的变化值作为运动信息,最后送入 CNN 进行分类检测。此方法多依赖于人工设置合理阈值,对数据集适应性差。曹建荣等^[5]和马敬奇等^[6]首先从视频中提取人体骨架点,再根据人体骨架点进行人工特征计算,最后送入分类网络中判断。此方法虽然可以解决图像方法的一些缺点,具有对环境有强鲁棒性且数据量少等优点,但并没有充分利用人体的结构信息以及运动信息,单纯依靠人工设定特征,其网络效果难以令人满意。

骨架点属于非欧数据,无法用 CNN 等方法处理^[7],研究者提出图卷积网络(graph convolution network, GCN)^[8]用于处理这类数据,并在动作识别领域取得较好应用。时空图卷积网络(spatial-temporal graph convolution network, ST-GCN)^[9]分别在空间和时间上交替进行卷积,可以更好获取动态骨架信息,提高动作识别能力。但 ST-GCN 仍存在一些问题,比如需要手工定义人体拓扑图、建模复杂、权值固定和对数据集适用能力差等缺点。

为了解决上述问题,同时提高跌倒检测能力,本文提出了基于时空自适应图卷积网络(spatial-temporal adaptive graph convolution network, ST-AGCN)的跌倒检测算法。本文从视频中提取骨架信息,解决常规图像前景提取的缺点;使用自适应图卷积网络,分别获取人体空间信息和时间动态信息。该网络的空间卷积单元使用归一化高斯函数计算骨架点之间的相似性,并以此创建拓扑图,从而解决了人工干预和拓扑图需繁琐定义等问题。时间卷积单元利用多尺度卷积,提取不同感受野的运动信息,进而提高识别效果。本文具体步骤是:首先提取输入视频的人体骨架点序列,然后将其处理成多维向量,最后送入时空自适应图卷积网络中,判断是否跌倒。

1 骨架点提取及数据预处理

1.1 HRNet 人体姿态估计算法

人体姿态估计算法主要分为自上而下和自下而上两类, RMPE^[10]、Mask-RCNN^[11]和 HRNet^[12]属于前者, OpenPose^[13]属于后者。

表1给出上述4种算法的仿真结果, HRNet 的各项指标都是最优^[11]。这是因为 HRNet 采用了多分辨率和并行融合处理,能够更准确预测骨架点位置;同时 HRNet 配置比较容易、处理速度快,因此本文选取该网络进行骨架点提取。

表1 主要人体姿态估计算法对比 %

算法	AP(0.50)	AP(0.75)	AP	AR
OpenPose ^[13]	84.9	67.5	61.8	66.5
Mask-RCNN ^[11]	87.3	68.7	63.1	—
RMPE ^[10]	89.2	79.1	72.3	—
HRNet-W48 ^[12]	92.5	83.3	75.5	80.5

1.2 数据预处理

使用 HRNet 对初始视频进行骨架点提取,如图1所示,视频帧数为 T (不超过 300),每帧包含 V 个 C 维骨架点坐标。图1(b)为 $V=17$ 时所对应的骨架点序列,将该序列转换为三维张量 $C \times T \times V$,如图1(c)所示。帧内骨架点顺序对最终预测结果没有影响,因此只需保持帧间骨架点顺序一致。类似地,可以将 N 个视频的骨架点序列转换为四维张量 $N \times C \times T \times V$ 。

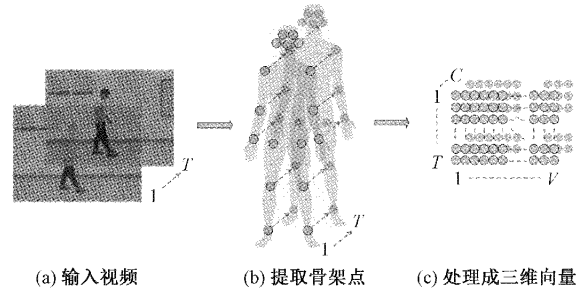


图1 数据预处理

2 时空图卷积网络(ST-GCN)

在 ST-GCN 中,数据用时空图 $G=(M, E)$ 表示,其中点集 $M=\{n_{ti} | t=1, \dots, T; i=1, \dots, V\}$ 是人体所有骨架点坐标的集合,可以反映人体静止时的状态, T 表示帧数, V 表示每帧内的骨架点数;边集 $E=\{E_s, E_f\}$ 可以反映人体运动变化,其中 $E_s=\{v_{ti}, v_{tj}\}$ 为帧内连接边, $E_f=\{v_{ti}, v_{(t+1)i}\}$ 为帧间连接边。

每帧内的骨架连接(即 E_s)用邻接矩阵 $\mathbf{A}=\{A_{ij}\}_{N \times N}$ 和单位矩阵 \mathbf{I} 表示,若骨架点 v_i, v_j 相连,则 $A_{ij}=1$, 否则 $A_{ij}=0$; 单位矩阵 \mathbf{I} 用来表示骨架点自身连接。采用多层图卷积网络^[14]进行迭代,如式(1)所示。

$$f_{\text{out}} = f_{\text{in}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{W} \quad (1)$$

其中, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, $\mathbf{D} = \{\mathbf{D}_{ij}\}_{N \times N}$ 为度矩阵, $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{A}_{ij}$ 为与骨架点 v_i 相连的骨架点数。对所有邻接点采用划分策略,根据邻接点的重要程度分区并赋予不同权值,建立完整骨架拓扑图,如式(2)所示。

$$f_{\text{out}} = \sum_j f_{\text{in}} \tilde{\mathbf{D}}_j^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}_j^{-\frac{1}{2}} \mathbf{W}_j \quad (2)$$

Shi 等^[15]提出双流自适应图卷积网络(two-stream adaptive graph convolution network, 2S-AGCN),增加可变参数用来学习数据特征,但仍保留需要定义的人体拓扑图,如式(3)所示。

$$f_{\text{out}} = \sum_j f_{\text{in}} (\mathbf{A}_j + \mathbf{B}_j + \mathbf{C}_j) \mathbf{W}_j \quad (3)$$

其中, \mathbf{A}_j 是需要预定义的邻接矩阵, $\mathbf{B}_j, \mathbf{C}_j$ 为与 \mathbf{A}_j 维度相同的可变参数。

ST-GCN 需要首先根据先验知识预定义人体骨架拓扑图、划分子集,然后根据子集重要性或距离质心的远近,赋

予骨架点不同的权重。

不同数据集提取的骨架点数可能不同,造成相应的邻接矩阵维数也可能不同。将 ST-GCN 从一个数据集移植到另一个数据集时,若提取的骨架点数不同,需要重新定义输入特征和邻接矩阵,因此 ST-GCN 无法适应所有的数据集。另外,目前还没有一种拓扑图可以最佳地表示骨架点间的关系,进而更有效提取关联特征,因此直接预定义拓扑图可能会丢失部分信息。

基于以上考虑,本文对 ST-GCN 进行改进,提出了 ST-AGCN,利用数据学习人体骨架拓扑图,并将其与网络一起进行优化。该网络的优点是无需手动预先定义固定的拓扑图,因此增加了模型的灵活性和对不同数据的适用性。

3 时空自适应图卷积网络(ST-AGCN)

3.1 ST-AGCN 框架

本文整体框架如图 2 所示,将预处理后的数据送入网络,经过 3 个时空自适应图卷积块(ST-AGCN Block),模块下面的数字分别表示输入通道数、输出通道数和卷积步幅,如 64,128,2 代表输入通道数是 64,输出通道数是 128,卷积步幅为 2。最后进行全局平均池化,送入 SoftMax 层进行分类判断。

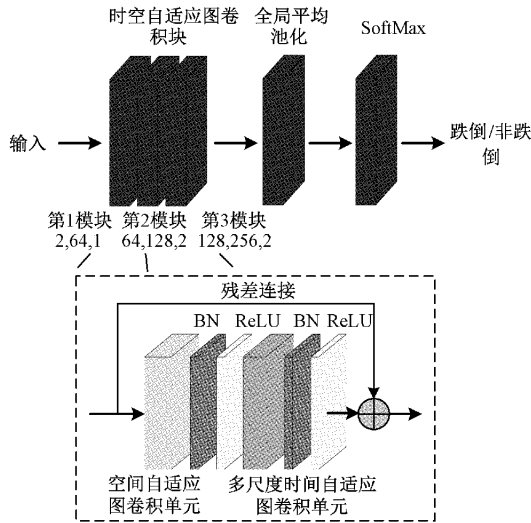


图 2 ST-AGCN 模型结构

ST-AGCN 块采用先空间后时间的串行输入方式。首先,利用空间自适应图卷积(spatial-AGCN, S-AGCN)建立人体骨架点间的关系,并赋予不同权重;然后,利用多尺度时间自适应图卷积(multi-scale temporal-AGCN, MST-AGCN)提取运动信息,并进行跌倒判决。在空间和时间卷积之后分别加入批归一化(batch normalization, BN)层和 ReLU 激活层,数据经过 BN 层后能够保持同分布,可以加快模型收敛速度。另外,为了保持模型稳定,在块与块之间添加了残差连接。ST-AGCN 块输出与输入之间的关系为:

$$f_{out} = MST - AGCN[S - AGCN(f_{in})] \quad (4)$$

3.2 空间自适应图卷积(S-AGCN)

本文的拓扑图通过数据学习方式建立:利用归一化嵌入式高斯函数^[16]计算两个骨架点间的相似性,确定拓扑图的连接程度。同一帧中两个骨架点 v_i 与 v_j 的相似度计算如式(5)所示。

$$f(n_i, n_j) = \frac{\exp[\theta(v_i)^T \phi(v_j)]}{\sum_{j=1}^V \exp[\theta(v_i)^T \phi(v_j)]} \quad (5)$$

其中, V 是骨架点个数, $\theta(v_i) = W_\theta v_i, \phi(v_j) = W_\phi v_j$ 。

模型实现步骤为:1)选取 1×1 卷积作为嵌入函数,将维度为 $C_{in} \times T \times V$ 的数据特征分别送入两个不同嵌入函数 θ 和 ϕ 中,得到两个 $C_e \times T \times V$ 的特征张量;2)将两个嵌入特征分别展平为 $V \times C_e T$ 和 $C_e T \times V$,通过矩阵相乘,得到相似度矩阵 $S = \{S_{ij}\}_{V \times V}$,其中 S_{ij} 为骨架点 v_i 与 v_j 的相似度;3)通过 SoftMax 将矩阵值进行归一化得到拓扑图。 S 可用下式计算:

$$S = \text{SoftMax}(f_{in}^T W_\theta^T W_\phi f_{in}) \quad (6)$$

其中, f_{in} 为输入, W_θ 和 W_ϕ 分别是嵌入函数 θ 和 ϕ 的权重矩阵。将两个嵌入函数中的参数 C_e 设置为输出通道数的 $1/4$,能够减少计算量。

由式(1)可知,图卷积的传播主要依赖于拓扑图,将式(1)变为下列形式:

$$f_{out} = f_{in} S W_w \quad (7)$$

式中: W_w 对应式(1)中的加权函数 W 。

S-AGCN 单元如图 3 所示, \oplus 表示元素加法运算, \otimes 表示矩阵乘法运算。残差分支中的虚线框表示:若输入、输出通道数不等,则使用 1×1 卷积匹配通道数。S-AGCN 单元计算公式为:

$$f_{out} = f_{in} S W_w + \text{Res}(f_{in}) \quad (8)$$

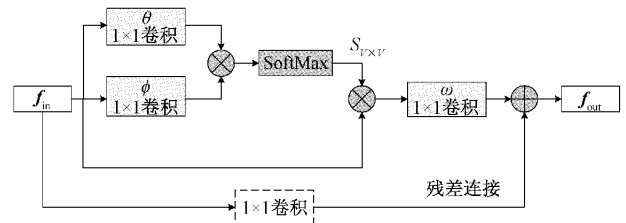


图 3 空间自适应图卷积单元

3.3 多尺度时间自适应图卷积(MST-AGCN)

跌倒是与时间信息有关的持续性动作,通过人体运动变化可以更加准确地进行跌倒检测。若只依靠单一静止状态,容易出现误判,如躺着和跌倒不动两者非常相似,难以检测。

文献[17]首次利用逐点卷积(point-wise convolution, 1×1 卷积),提出实现跨通道信息交互的池化结构。其优点是能够进行通道数降维,减少网络参数,进行更深层的训

练,提高检测效果。文献[18]利用多层空洞卷积,聚合多尺度的信息提高模型预测能力,能够扩大卷积感受野,同时不增大计算成本。

本文采用多个尺度不同的卷积核来提取时间信息,构造多尺度时间自适应图卷积单元,如图4所示。具体实现为:1)整个单元包括4个分支,每个分支都分别利用 1×1 卷积进行跨通道信息聚合,同时对输入通道数降维,每个卷积的输出通道数为整个单元输出通道数除以分支数;2)在其中的两个分支后面各加一个空洞卷积,其卷积核大小相同(9×1)、扩张系数(dilation)不同(分别设为1、2)。目的是对不同感受野进行特征提取,同时可以降低卷积操作所产生的计算成本;3)在另一个分支中使用最大池化层获取主要特征并减少参数;4)对多个卷积输出结果进行拼接处理,聚合各个分支的特征。5)增加残差连接,目的是保持模型稳定,加快收敛速度。公式如下:

$$\begin{aligned} f_{out} &= \text{Concat}(t_1, t_2, t_3, t_4) + \text{Res}(f_{in}) \\ t_1 &= \text{Conv}_{2D(9 \times 1, 1)} [\text{Conv}_{2D(1 \times 1, 1)}(f_{in})] \\ t_2 &= \text{Conv}_{2D(9 \times 1, 2)} [\text{Conv}_{2D(1 \times 1, 1)}(f_{in})] \\ t_3 &= \text{MaxPool}_{2D(3 \times 1, 1)} [\text{Conv}_{2D(1 \times 1, 1)}(f_{in})] \\ t_4 &= \text{Conv}_{2D(1 \times 1, 1)}(f_{in}) \end{aligned} \quad (9)$$

其中,下标 $2D(k \times 1, d)$ 中的 $2D$ 是二维卷积, $k \times 1$ 为核尺寸, d 为扩张系数。

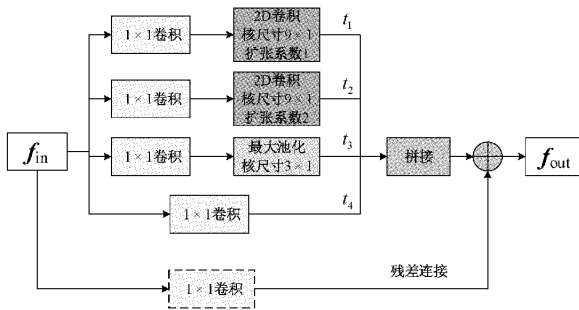


图4 多尺度时间自适应图卷积单元

4 数据和实验结果分析

4.1 实验测试环境与数据集

实验所使用的硬件平台为 Intel i7-6700HQ、NVIDIA GEFORCE GTX965M,内存为 16 G,软件使用 Windows10 操作系统、PyTorch 框架。

本文使用公开数据集 URFD^[19]和自建数据集。其中 URFD 数据集包含 40 个日常动作、30 个跌倒动作共 70 组视频数据,每个视频都保存为 PNG 图像序列格式。由于数据量较少,不能充分验证模型的优劣,故本文自行建立数据集进行补充。

自建数据集包括 10 个类别,其中 6 个类别为正常行为(行走、蹲下、弯腰、坐、躺下、趴),4 个类别为跌倒行为(随机倾倒、绊倒、重心不稳跌倒、起身昏倒)。采用 3 台 $1\ 280 \times 720$ 分辨率摄像机,分别固定在相同高度的不同角

度(0° 、 45° 、 90°)进行数据采集,共有 9 位实验人员(6 名男性、3 名女性)参与,每位实验人员重复每个动作 3 次,共采集 810 个视频数据,每个视频中均只有一人出现,视频长度不超过 300 帧。将视频随机按 7:3 比例划分为训练集和测试集,对视频的预处理提取 17 个人体骨架点,自建数据集部分示例如图 5 所示。

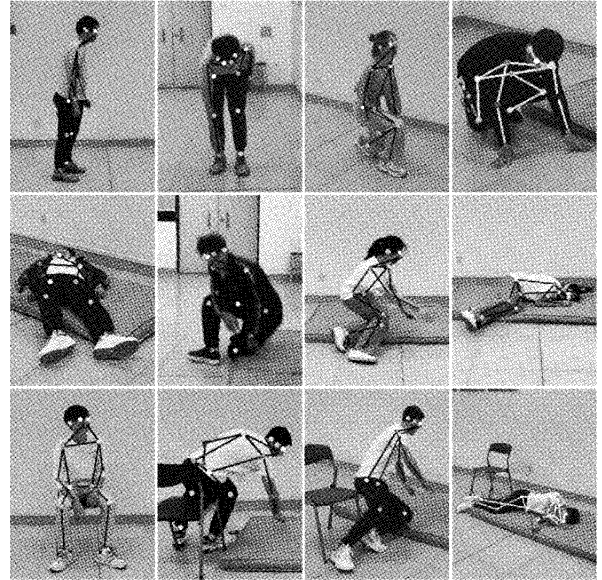


图5 自建数据集示例

4.2 评价指标

为了评价各个模型的优劣,采用广泛认同的衡量指标——准确率(accuracy, Acc)^[3]、精确率(precision, Pr)^[5]、灵敏度(sensitivity, Se)^[5]和 F1 分数,另外添加模型参数量(parameters, Pa)指标,各个指标的含义如下:

Acc:预测正确的样本数与总样本数之比,表示预测正确的概率;

Pr:预测为跌倒中的真实跌倒数与预测跌倒数之比,体现预测跌倒的准确性;

Se:实际跌倒中检测出的跌倒数与实际跌倒数之比,体现跌倒预测的全面性;

F1:精确率 Pr 与灵敏度 Se 的调和平均值,是衡量模型综合能力的指标;

Pa:模型参数数量,用来衡量模型的复杂度。

4.3 网络结构改进对性能影响

在自建数据集上,通过控制变量,分别对 S-AGCN 中的学习拓扑图、MST-AGCN 提取多尺度时间信息和 ST-AGCN 模块数对网络性能的影响,进行详细实验,从而确定最佳模型。

1) S-AGCN 学习拓扑图

本文使用 S-AGCN 单元中归一化嵌入式高斯函数学习人体拓扑图,替换 ST-GCN^[9]中的手工定义拓扑图,通过消融实验对比改进前后网络性能的变化,结果如表 2 所示。

表 2 S-AGCN 优化模型 %

空间方法	Acc	Pr	Se	F1
S-GCN	94.80	95.93	89.90	92.74
S-AGCN	95.43	94.76	92.81	93.77

表 2 中第一行为基准数据,为 ST-GCN 使用普通 S-GCN 的结果,灵敏度较低,仅为 89.90%,容易出现漏检现象。将空间卷积变成 S-AGCN,其灵敏度提升 2.91%,精确率稍微下降 1.17%,准确率和 F1 分数分别提升了 0.63%和 1.03%,可以看出经过学习的拓扑图能够更加准确表达人体骨架点间的关系,训练后的拓扑图如图 6 所示。

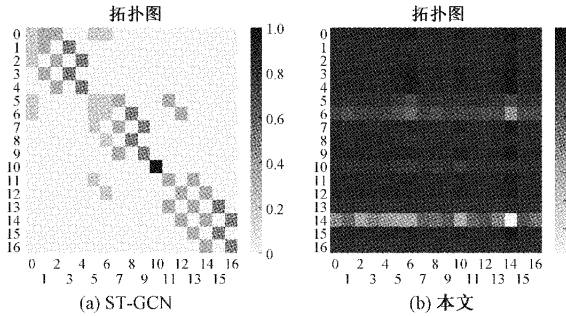


图 6 拓扑图

图 6(a)是 ST-GCN 的人工预定义拓扑图,在训练过程中固定不变,当前骨架点只与邻近骨架点有关。图 6(b)是本文从数据中学习的拓扑图,图中颜色深浅代表骨架点权重的差异,可以看出对 6(右肩)、10(右髋)和 14(右膝)骨架点赋予较大权值,与其他骨架点相比,在跌倒检测中所起的作用更大。在日常动作中,这 3 个骨架点一般处于不同高度,而在跌倒中经常处于同一水平面。

2) MST-AGCN 提取多尺度时间信息

时间图卷积单元利用 MST-AGCN 提取运动信息,通过消融实验,进一步验证各个分支对模型的影响,结果如表 3 所示。

表 3 各分支对模型的影响 %

方法	Acc	Pr	Se	F1
1	95.43	94.76	92.81	93.77
2	93.31	91.82	89.97	90.88
3	93.33	95.64	85.92	90.51
4	97.73	98.77	95.06	96.88
5	97.21	99.72	92.74	96.10
6	98.42	97.76	97.99	97.88
7	99.55	99.79	99.01	99.39

方法 1 是基准数据,没有使用任何分支,准确率和灵敏度较低,容易出现漏检现象。方法 2 使用第 3 个分支,性能有所下降;方法 3 同时使用第 3、4 分支,虽然精确率有所提

升,但是灵敏度下降较多,说明单独使用最大池化层还不能提升模型性能,反而可能因为只保留主要特征而丢失其他特征,造成性能下降。方法 4 只使用第 1 分支,各方面指标均有所提升;方法 5 同时使用第 1、3 分支,精确率提升了 0.95%,但是灵敏度下降了 2.32%,说明同时使用卷积和最大池化可以提高精确率,但是会降低灵敏度。方法 6 同时使用第 1、2 分支,同时使用两个空洞卷积,使得提升灵敏度了 2.93%。方法 7 使用全部分支获得了最优结果,相比不使用 MST-AGCN,准确率、精确率、灵敏度和 F1 分数分别提升了 4.12%、5.03%、6.20%和 5.62%,体现了 MST-AGCN 具有更强的时间特征分析能力,能够更加准确地检测跌倒,同时减少漏检。

进一步验证 1×1 卷积对 MST-AGCN 的影响,分别进行添加、去除操作,性能对比结果如表 4 所示。

表 4 逐点卷积对网络的影响

1×1 卷积	Pa/M	Acc/%	Pr/%	Se/%	F1/%
未使用	1.13	98.24	97.50	97.75	97.62
已使用	0.63	99.55	99.79	99.01	99.39

表 4 中第 2 行是使用卷积得到的结果,与不使用卷积比较,参数量几乎减少 1/2,准确率、精确率、灵敏度和 F1 分数分别提升了 1.31%、2.29%、1.26%和 1.77%。这可以说明卷积不仅可以减少参数量,还能实现跨通道间的信息交互,提高模型性能。

3) ST-AGCN 模块数

为了研究 ST-AGCN 模块数对网络性能的影响,分别利用 1~4 个模块进行消融实验,同时设置并确定不同模块的输出通道数,给出最优模型。从表 5 可以看出,使用 1 个模块(输出通道数为 64)时参数量最低,仅为 14.63 k,同时还可以达到 96.81%的准确率,这种适用于轻量网络。增加 1 个模块(通道数为 128),参数量增加约 10 倍,各个指标都相应提升了 2%~3%。再增加 1 个模块(通道数为 256),参数量提升至 627.05 k,仍然在可接受范围内;此时模型的各个指标都较好,均达到 99%以上。若再增加 1 个模块(通道数为 64),此时网络参数量增加了 31.97 k,灵敏度达到最高的 100.00%,但是精确率较之前降低了 1.96%,而且其他指标还稍微有所降低。综合考虑,使用 3 个输出通道(通道数分别为 64,128,256)的 ST-AGCN 模块。

表 5 ST-AGCN 模块数对网络性能影响

模块数	Pa/k	Acc/%	Pr/%	Se/%	F1/%
1	14.63	96.81	95.59	95.82	95.70
2	138.21	98.89	99.29	97.69	98.48
3	627.05	99.55	99.79	99.01	99.39
4	659.02	99.18	97.83	100.00	98.80

4.4 网络性能横向比较

为了验证本文算法的准确性及有效性,在两个数据集上进行测试。对数据集都采用同样预处理方式,在模型的各种超参数和运行设备一致的前提下,与目前其他经典的图卷积方法(ST-GCN^[9], 2S-AGCN^[15], MS-AAGCN^[20], G3D^[21])进行对比。表6为在URFD数据集上的对比结果。

表6 跌倒检测算法结果对比(URFD数据集)

算法	Pa/M	Acc/%	Pr/%	Se/%	F1/%
ST-GCN ^[9]	3.07	95.45	90.91	100.00	95.24
2S-AGCN ^[15]	3.44	90.91	90.00	90.00	90.00
MS-AAGCN ^[20]	3.74	95.45	90.91	100.00	95.24
G3D ^[21]	2.88	90.91	83.33	100.00	90.91
本文	0.63	95.45	90.91	100.00	95.24

本文算法具有100%的灵敏度,同时在准确率、精确率方面取得较好结果,分别为95.45%、90.91%。文献[21]算法虽然可以达到相同的灵敏度,但是其83.33%的精确率却难以令人满意。另外两个算法(文献[9]、[20])与本文在该数据集测试中获得同样效果,但是文献[9]算法需要预先定义人体拓扑图,本文算法根据数据学习得到拓扑图,不需要预先定义,具有更强的适用性;另外本文模型的参数量只有文献[20]算法的1/6。

各模型的详细识别情况混淆矩阵如图7所示。图7(a)是文献[15]算法的结果,将跌倒和未跌倒数据集中的各一个视频误判。从图7(b)可以看出,文献[21]将两个未跌倒判定为跌倒,由此造成精确率较低。另外两个算法(文献[9]、[20])和本文识别结果如图7(c)所示,都是将一个未跌倒视频误判为跌倒,在URFD数据集中呈现相同的结果。

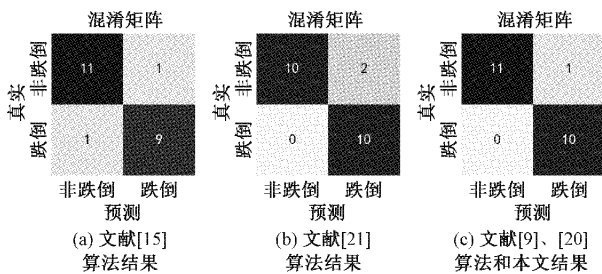


图7 混淆矩阵

由于URFD数据集数据量较小,容易出现多个指标相同的情况。本文继续在自建数据库中进行测试比较,结果如表7所示。

本文算法保持了99.01%的高灵敏度,准确率、精确率和F1分数也取得较优结果,分别为99.55%、99.79%和99.39%,灵敏度相比其他算法分别提高了9.11%、7.32%、2.30%和2.24%,其他指标也有所提高。文献[9, 15, 21]

表7 跌倒检测算法结果对比(自建数据集)

算法	Pa/M	Acc/%	Pr/%	Se/%	F1/%
ST-GCN ^[9]	3.07	94.80	95.93	89.90	92.74
2S-AGCN ^[15]	3.44	96.11	97.66	91.69	94.57
MS-AAGCN ^[20]	3.74	97.94	97.72	96.71	97.20
G3D ^[21]	2.88	96.80	95.24	96.77	95.88
本文	0.63	99.55	99.79	99.01	99.39

都需要人工定义骨架点关系,并且所有指标与本文算法都有一些差距。文献[20]算法虽然也能够达到比较接近的识别效果,其他指标都相差在2%左右,但灵敏度相差较大,为2.30%。对于跌倒检测任务来说,灵敏度尤为重要,漏检所造成的后果比较严重。

图8给出了各个算法的收敛过程,文献[9]算法在训练过程中有波动,且收敛较慢,稳定在较低的准确率。文献[21]算法前期收敛较快,后期缓慢收敛。本文算法在100轮左右已经接近收敛,后期进行微调达到最佳。

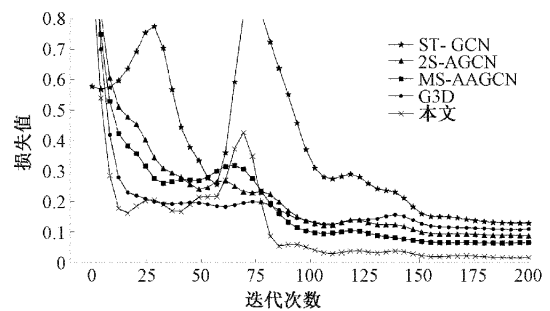


图8 各个算法的损失值比较

5 结论

本文提出了一种时空自适应图卷积跌倒检测算法,将视频处理为骨架点序列,减少周围环境的影响;通过归一化嵌入式高斯函数寻找人体骨架点间的关系,而无需利用先验知识预先定义人体拓扑图;使用多尺度卷积单元提取时间运动特征,提升检测效果。实验数据显示本文算法在公开和自建数据集上分别达到95.45%和99.55%的准确率,能够胜任跌倒检测任务。相比于其他图卷积算法,本文算法摆脱了人体建模的手工操作过程,增加了数据集的适用性;能够有效提高识别性能;参数量仅只有其他算法的1/4,甚至更少,因此可以有效降低运算成本,更适用于低成本计算设备。由于本文算法基于人体骨架数据,其提取骨架点的准确与否对算法的识别准确率影响较大。若存在其他物品对人体产生遮挡,会严重影响骨架点提取,造成识别不准确。本文ST-AGCN部分可以每秒处理35帧数据,但是由于骨架点提取部分需要大量运算,从而使得整个系统无法满足实时性要求,系统的实际应用还依赖于骨架点提取算法的改进。

参考文献

- [1] SHAHZAD A, KIM K. FallDroid: An automated smart-phone-based fall detection system using multiple kernel learning[J]. IEEE Transactions on Industrial Informatics, 2018, 15(1): 35-44.
- [2] 彭亚平, 贺乾格, 柯希焱, 等. 一种基于加速度传感器的摔倒检测腰带[J]. 电子测量技术, 2018, 41(11): 117-120.
- [3] CARLIER A, PEYRAMAURE P, FAVRE K, et al. Fall detector adapted to nursing home needs through an optical-flow based CNN [C]. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020: 5741-5744.
- [4] 金彦亮, 陈刚. 基于双流卷积的实时跌倒行为检测[J]. 计算机工程与设计, 2021, 42(9): 2621-2626.
- [5] 曹建荣, 朱亚琴, 张玉婷, 等. 基于关节点特征的跌倒检测算法[J]. 计算机应用, 2022, 42(2): 622-630.
- [6] 马敬奇, 雷欢, 陈敏翼. 基于 AlphaPose 优化模型的老人跌倒行为检测算法[J]. 计算机应用, 2022, 42(1): 294-301.
- [7] ZHOU J, CUI G, HU S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.
- [8] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[C]. International Conference on Learning Representations, 2014.
- [9] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. Thirty-second AAAI Conference on Artificial Intelligence, 2018: 7444-7452.
- [10] FANG H S, XIE S, TAI Y W, et al. Rmpe: Regional multi-person pose estimation[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2334-2343.
- [11] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [12] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5693-5703.
- [13] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7291-7299.
- [14] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[C]. International Conference on Learning Representations, 2017.
- [15] SHI L, ZHANG Y, CHENG J, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12026-12035.
- [16] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7794-7803.
- [17] LIN M, CHEN Q, YAN S. Network in network [C]. International Conference on Learning Representations, 2014.
- [18] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [C]. International Conference on Learning Representations, 2016.
- [19] KWOLEK B, KEPSKI M. Human fall detection on embedded platform using depth maps and wireless accelerometer [J]. Computer Methods and Programs in Biomedicine, 2014, 117(3): 489-501.
- [20] SHI L, ZHANG Y, CHENG J, et al. Skeleton-based action recognition with multi-stream adaptive graph convolutional networks [J]. IEEE Transactions on Image Processing, 2020, 29: 9532-9545.
- [21] LIU Z, ZHANG H, CHEN Z, et al. Disentangling and unifying graph convolutions for skeleton-based action recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 143-152.

作者简介

刘鹏飞, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 2112003187@mail2.gdut.edu.cn

李伟彤 (通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为图像融合、图像质量评估等。

E-mail: liweitong@gdut.edu.cn