软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2019,30(3):718−737 [doi: 10.13328/j.cnki.jos.005685]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

CNN 多位置穿戴式传感器人体活动识别*

邓诗卓[1], 王波涛[1], 杨传贵[1], 王国仁[2]

[1](东北大学 计算机科学与工程学院,辽宁 沈阳 110819)

[2](北京理工大学 计算机学院,北京 100081)

通讯作者: 王波涛, E-mail: wangbotao@cse.neu.edu.cn

摘 要: 随着人工智能的发展和可穿戴传感器设备的普及,基于传感器数据的人体活动识别(human activity recognition,简称 HAR)得到了广泛关注,且具有巨大的应用价值.抽取良好判别力的特征,是提高 HAR 准确率的关键因素.利用卷积神经网络(convolutional neural networks,简称 CNN)无需领域知识抽取原始数据良好特征的特点,针对现有基于传感器的 HAR 忽略三轴向传感器单一轴向多位置数据空间依赖性的不足,提出了两种动作图片构建方法 T-2D 和 M-2D,构建多位置单轴传感器动作图片和非三轴传感器动作图片;进而提出了卷积网络模型 T-2DCNN 和 M-2DCNN,抽取三组单一轴向动作图片的时空依赖性和非三轴传感器的时间依赖性,并将卷积得到的特征拼接为高层次特征用于分类;为了优化网络结构,减少卷积层训练参数数量,进一步提出了基于参数共享的卷积网络模型.在公开数据集上与现有的工作进行对比实验,默认参数情况下,该方法在公开数据集 OPPORTUNITY 和 SKODA 中 F_1 最大提升值分别为 6.68%和 1.09%;从传感器数量变化和单类识别准确性角度验证了模型的有效性;且基于共享参数模型,在保持识别效果的同时减少了训练参数.

关键词: 人体活动识别;卷积神经网络;穿戴式传感器;特征提取;动作图片

中图法分类号: TP18

中文引用格式: 邓诗卓,王波涛,杨传贵,王国仁.CNN 多位置穿戴式传感器人体活动识别.软件学报,2019,30(3):718−737. http://www.jos.org.cn/1000-9825/5685.htm

英文引用格式: Deng SZ, Wang BT, Yang CG, Wang GR. Convolutional neural networks for human activity recognition using multi-location wearable sensors. Ruan Jian Xue Bao/Journal of Software, 2019,30(3):718−737 (in Chinese). http://www.jos.org.cn/1000-9825/5685.htm

Convolutional Neural Networks for Human Activity Recognition Using Multi-location Wearable Sensors

DENG Shi-Zhuo[1], WANG Bo-Tao[1], YANG Chuan-Gui[1], WANG Guo-Ren[2]

[1](School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China)

[2](School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Wearable sensor-based human activity recognition (HAR) plays a significant role in the current smart applications with the development of the theory of artificial intelligence and popularity of the wearable sensors. Salient and discriminative features improve the performance of HAR. To capture the local dependence over time and space on the same axis from multi-location sensor data on convolutional neural networks (CNN), which is ignored by existing methods with 1D kernel and 2D kernel, this study proposes two methods, T-2D and M-2D. They construct three activity images from each axis of multi-location 3D accelerometers and one activity image from the other sensors. Then it implements the CNN networks named T-2DCNN and M-2DCNN based on T-2D and M-2D respectively,

* 基金项目: 国家自然科学基金(61872072, U1401256, 61173030, 61732003)
Foundation item: National Natural Science Foundation of China (61872072, U1401256, 61173030, 61732003)
本文由智能数据管理与分析技术专刊特约编辑樊文飞教授、王国仁教授、王朝坤副教授推荐.
收稿时间: 2018-07-18; 修改时间: 2018-09-20; 采用时间: 2018-11-01

which fuse the four activity image features for the classifier. To reduce the number of the CNN weight, the weight-shared CNN, TS-2DCNN and MS-2DCNN, are proposed. In the default experiment settings on public datasets, the proposed methods outperform the existing methods with the F_1-value increased by 6.68% and 1.09% at most in OPPORTUNITY and SKODA respectively. It concludes that both naïve and weight-shared model have better performance in most F_1-values with different number of sensors and F_1-value difference of each class.

Key words: human activity recognition; convolutional neural network; wearable sensor; feature extraction; activity image

近年来,随着智能生活概念的普及和可穿戴终端设备技术的快速发展,大量的时序传感器数据为普适计算、人工智能等领域带来了新的研究热点,因此,基于传感器数据的人体活动识别(human activity recognition,简称 HAR)[1,2]得到了广泛关注.基于传感器数据的 HAR 能够有效地避免基于视频数据动作识别[3,4]带来的隐私泄露等问题[5],且计算复杂度比视频数据低,使得基于传感器的 HAR 研究更加具有现实意义.目前,HAR 的研究应用包括健康监控[6−8]、智慧家庭[9−11]、工业环境[12]和运动员监测[13]等,例如:老人通过佩戴智能手环检测是否有走路姿态异常;工人通过佩戴传感器来记录和规范操作动作.因此,准确地识别和记录人体的活动姿态,能够为人们提供更为精确的服务.面对多形态多位置的复杂传感器数据,如何利用传感器数据特点提取具有良好判别力的特征以提高基于传感器数据的的 HAR 准确率,是具有研究价值和现实意义的研究问题.

基于传感器数据的 HAR 特征提取方法包括传统特征提取和深度学习特征表达两种方法[1,2].传统方法需人工对划分的数据抽取统计学意义特征向量,包括时域特征[14−16]、频域特征[17]以及其他特征向量[16,18].但是传统方法所抽取的特征都是浅层的,深度学习技术可以直接从原始数据自动地抽取更复杂的深层的特征[19,20],无需人工干预;同时,深度学习可以有效地解决类内差异和类间相似的问题.卷积神经网络(convolutional neural network,简称 CNN)是深度学习方法中研究较多的特征抽取方法[2,21].文献[19]的理论分析和实践证明了 HAR 中基于传感器的 CNN 可以更有效地抽取数据隐藏的特征.

本文基于多位置的传感器数据特点,利用深度学习中卷积神经网络技术,从原始数据中抽取有效的特征来提高 HAR 识别准确率.现有的基于传感器的 CNN 抽取特征方法根据卷积核维度分为一维卷积核方法 1DCNN[13,22−28]和二维卷积核方法 2DCNN[29−37].1DCNN 即提取基于时间依赖的特征,在时间轴上面进行卷积.虽然通过一维卷积核可以抽取具有良好分辨能力的特征,但是一维卷积核忽略了原始数据的不同位置传感器之间的空间依赖性.因此研究人员提出了基于传感器数据的二维卷积核,综合考虑了时间依赖性和空间依赖性.但是,这些 2D-CNN 方法仅仅是朴素地将不同类别传感器罗列拼接,或者利用一定算法构建传感器动作图片(activity image)作为二维卷积核的输入,并没有考虑不同位置三轴向传感器(三轴加速度、三轴陀螺仪等)的相同轴向数据之间隐藏的空间依赖性.

针对以上不足,为了进一步利用不同位置的三轴传感器之间的相同轴向多位置传感器原始数据的空间依赖性,本文提出了新颖的 CNN 二维卷积的输入动作图片构建方法,进而提出了卷积网络结构.为了优化网络结构,进一步提出了卷积层参数优化方法.本文的主要贡献如下:

(1) 提出了基于多位置三轴向传感器二维卷积核输入动作图片构建方法 T-2D(triaxial 2 dimension).该方法对来自不同位置传感器的原始数据按照不同轴向划分,生成 3 张动作图片.每一轴向动作图片构成二维卷积的输入;在 T-2D 基础之上,针对非三轴传感器提出 M-2D(mixed 2 dimension),构建非三轴传感器数据一维卷积核输入的动作图片,以抽取时间依赖特征,共生成 4 张动作图片;

(2) 提出了基于 T-2D 和 M-2D 多层卷积神经网络模型 T-2DCNN 和 M-2DCNN,将 M-2D 得到一维特征和 T-2D 得到的二维特征合并成高层次特征向量,实现了人体活动识别功能;为了减少卷积层训练参数数量,对卷积网络进行优化,进一步提出了基于参数共享的卷积网络 TS-2DCNN 和 MS-2DCNN;

(3) 在公开数据集上进行了对比实验,验证了基于本文所提出的特征抽取方法的 HAR 模型识别效果,同时验证了共享参数卷积网络在参数数量减少的情况下,可依然保持其识别效果;从传感器数量变化以及每个类别识别准确性的角度,分别验证了本文所提出的两种卷积输入动作图片构建方法的有效性和稳定性.

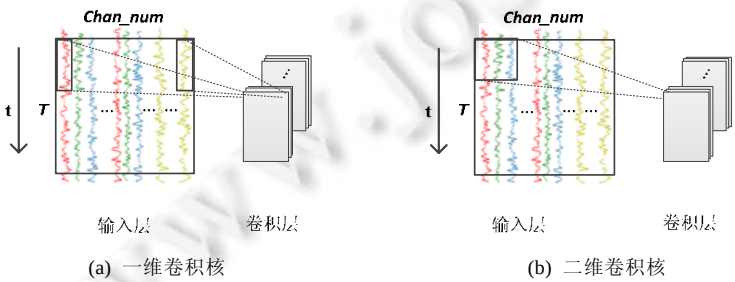
本文第 1 节介绍现有基于 CNN 的 HAR 中基于穿戴式传感器的两种典型特征抽取方法,并进行分析对比.第 2 节描述二维卷积核输入动作图片的两种构建方法和基于本文方法的 HAR 模型的卷积网络结构及其优化网络.第 3 节通过实验和对比分析,证明本文提出的二维卷积输入构建方法的良好性能.最后是总结与展望.

1 相关工作

卷积神经网络已经广泛并成功地应用于图像识别[38,39]和语音识别[40],在无需知道专业领域知识(domain knowledge)的前提下,自动地抽取具有高分辨能力的特征.近年来,CNN 在基于传感器数据的人体活动识别方面也取得了一定进展.与图像识别的原始数据不同,在 HAR 中的原始数据大多为时间序列数据、一维数据.在文献[19]中,从局部依赖(local dependency)和尺度不变性(scale invariance)这两个角度描述了在动作识别领域应用 CNN 的优势.本文主要从卷积核维度以及对应的输入构造方法对现有工作进行总结和分析.

1.1 HAR中一维卷积核1D-CNN应用

一维卷积核应用于一维时间序列数据,抽取具有时间依赖性的特征.尽管基于可穿戴设备传感器类型多种多样,包括加速度计、陀螺仪等,但是传感器数据或者传感器各通道数据都是一维时序数据,例如加速度计中每一个轴向的时序数据代表一个通道.图 1(a)描述了大部分一维卷积核的卷积方式,数据包含具备三轴属性的传感器数据,同时也有可能包含不具备三轴属性的传感器数据,例如心率等.输入矩阵是由滑动窗口(sliding window)大小 T 和传感器通道总数 *Chan_num* 构成的矩阵,一维卷积核方式在各通道数据上面进行卷积操作,因此在滑动窗口固定的情况下,能够自动抽取动作特征.但在具体相关文献中,传感器的选择以及 CNN 的结构略有不同.文献[19]中,作为早期的 CNN 在 HAR 中的应用,将三轴加速度计的 3 个轴向数据看成 3 个通道,分别在各轴向数据上沿时间轴进行卷积和池化,最后合并成该窗口的特征向量,并提出了局部共享权重的 CNN 卷积方式.该文献使用一对卷积层和池化层,因此抽取特征属于浅层特征.文献[22]同样利用单一加速度计的三轴向原始数据作为卷积网络的输入,但是通过改变卷积核的大小来适应三轴加速度信号特点.类似地,文献[13]提出了利用手腕传感器采集的三轴加速度识别排球动作.文献[23,24]利用加速度计三轴向数据和陀螺仪三轴向数据组成六通道,对输入使用一维卷积核,多卷积层和池化层抽取深层特征提高了识别准确率.文献[25]中除了使用了三轴加速计和多种传感器数据,将 3 个通道扩展成多通道,利用两层一维卷积核和池化抽取特征.文献[26]中提出了将 CNN 和长短期记忆循环网络(long short-term memory,简称 LSTM)结合的深度学习模型,利用多层 CNN 一维卷积核自动抽取特征,且没有池化层,使用 LSTM 进行分类识别,在通用数据集上取得了良好的效果.文献[27]采用一维卷积核抽取特征,结合 VGG 和 LSTM 组合模型.文献[28]利用三轴加速度性质,将三轴加速度分看成深度为 3 的一维输入,经过卷积之后生成 180 个特征映射,以此来捕捉数据之间的依赖性.



(a) 一维卷积核 (b) 二维卷积核

Fig.1 Architecture of convolutional neural network based on multiple channels

图 1 基于多通道时序数据的卷积结构

综上所述,一维卷积核的输入通常为一维的原始信号数据.一维卷积核在抽取特征过程中能够较好地挖掘一维信号的时间依赖性,但是往往忽略了不同类别传感器或者是同一传感器不同轴向数据之间的空间依赖性,影响人体动作的识别效果.

1.2 HAR中二维卷积核2D-CNN应用

针对一维卷积核 HAR 的不足,绝大多数 2D-CNN 对原始数据进行重新排列,形成类似图片一样的输入矩阵,即动作图片.在该矩阵上使用二维卷积核以及相应的池化操作,利用二维卷积核抽取时空依赖性,如图 1(b)所示,之后结合不同的学习模型.文献[29]将原始数据多通道输入看作一张图片,第 1 层使用二维卷积核,在整张图片上进行卷积形成一维特征映射,再次进行卷积时,应用一维卷积核.文献[30]利用不同位置的相同类型传感器按三轴顺序排序,并以零矩阵填充来分隔不同类型的传感器,同一张动作图片上面包含所有轴,在此输入结构基础上使用二维卷积核.该文献实现了时间依赖性和空间依赖性结合.在文献[30]二维卷积核的基础之上,文献[31]对 CNN 卷积核应用了局部权重共享和全局权重共享两种卷积方式.不同于图 1(b),文献[32]利用傅里叶变换将一维原始信号数据生成频谱,对三轴加速度计和三轴陀螺仪信号形成频谱,将频谱视为动作图片作为输入进行卷积操作.类似地,文献[33]提出一种算法,可将一维信号转换为二维图片信号,其目的在于让每一通道信号都可以与其他通道信号相连,即使是不同类型的通道信号,抽取每两个信号之间的依赖性.受文献[33]的启发,文献[34]修改了输入图片构造算法,同时调整了同一时刻不同类型三轴传感器之间(x,y,z)轴的数据排列顺序,确保该时刻所有不同类型的传感器数据两两相邻,以及同一类型传感器内的不同轴向数据两两相邻.文献[35]针对单一位置传感器,使用滑动窗口折叠法分别生成三轴动作图片,在每个动作图片上分别卷积.文献[36,37]同样针对手机上传感器中的三轴加速度计,将滑动窗口内三轴数据直接生成同一张方形动作图片,之后采用卷积网络.

相比于一维卷积核只考虑时间域上面的依赖性,二维卷积核应用的角度更多,可以更大限度地利用传感器数据的空间依赖性.基于二维卷积核的 HAR 依赖于输入数据的组织构造形式,但是现有的工作仅仅是不同传感器数据按照通道朴素排列并直接进行二维卷积或是对三轴向传感器的同一位置的三轴数据进行简单粗暴地排列组合.尽管相较于传统的基于统计学特征的 HAR 和基于一维卷积核的 HAR,现有的二维卷积方式可以提高准确率和精确率等评价指标,但是并没有充分利用不同位置的三轴向传感器数据的特点以及充分挖掘空间依赖性,如三轴加速度计.因此,本文根据三轴传感器的数据特点,充分挖掘相同轴向数据之间的空间依赖性和时间依赖性,设计新颖的输入组织形式构造多张动作图片来提高 HAR 的识别效果.

2 基于二维卷积核的多位置三轴向传感器 HAR 模型

本文提出了基于二维卷积核的多位置三轴向穿戴式设备 HAR 模型,该模型虽然沿用了 HAR 问题中的二维卷积应用的思想,但是增强了对时序数据的空间依赖性的挖掘.因此,结合三轴向传感器特点,本文提出了一种新颖的二维卷积核输入的动作图片构建方法及其改进方法,利用不同位置的三轴向传感器的相同轴向数据构成输入动作图片,将原始数据分别构建成 3 个或 4 个动作图片,基于输入动作图片构建卷积神经网络,分别抽取原始数据各轴向数据内部的时空依赖性,进而得到更加具有分辨能力的特征向量,提高模型识别准确率.为了简化模型参数,对卷积网络进行参数优化,提出了基于共享参数的 HAR 模型.

2.1 问题描述与定义

本文研究的根本问题是如何学习一个具有良好识别效果的人体活动识别模型,即模型参数;关键问题是如何从原始标签数据中抽取具有高分辨能力的特征;核心问题是如何构造用于二维卷积核输入的动作图片,原因在于不同的动作图片组织形式会导致二维卷积核抽取的特征具有不同的依赖性和表达能力.为了方便描述问题,形式化定义如下:

定义 1(动作图片). 相对于一维时序数据,动作图片是对一维信号通过一定规律进行升维操作所生成的二维信号数据.动作图片是卷积核的输入.

定义 2(滑动窗口). 滑动窗口是动作图片的时间窗口,是时序传感器数据的时间处理单元,令 ws 表示滑动窗口大小,如图 1 中所示,$ws=T$.

定义 3(模型 M). 给定时序数据长度 N,每时刻原始训练数据 $X=\{(l_1,x_1),\ldots,(l_N,x_N)\}$为带有标签数据,时刻数据标签 l_i,其中,$i\in[1,N]$.假设覆盖率为 0 且 $ws=T$,用于训练的实例数量为 $N'=N/T$,其中,$j\in[1,N']$.用于训练模型 M 的数据为 $D=\{(t_1,d_1),\ldots,d_1(t_{N'},d_{N'})\}$,则:

$$M_{(A,B)}=\mathrm{argmin}Loss(A,B|D) \tag{1}$$

其中,$\boldsymbol{A}$,$\boldsymbol{B}$ 分别为模型 M 的权重参数矩阵和偏置参数矩阵,$Loss$ 函数为损失函数.即:在输入数据 D 的条件下求取使得损失函数最小的模型参数 $\boldsymbol{A}$ 和 $\boldsymbol{B}$.

2.2 HAR模型二维动作图片构建

2.2.1 基于多位置三轴传感器二维动作图片构建原理

传感器形态多样化可分为传感器种类多样化和传感器位置多样化[30],本文中构造二维卷积输入的前提是传感器位置多样化.传感器种类多样化是基于多种类别传感器的动作识别,例如基于加速度计、陀螺仪、磁力计等传感器的识别.传感器位置多样化是基于多种位置的相同类型传感器或者不同类型传感器的动作识别.本文结合三轴向传感器特点,舍弃同一传感器三轴之间的空间依赖性,选择不同传感器的相同轴向数据的空间依赖性,并构建用于卷积网络输入的动作图片,从理论和数据分析两个角度阐释本文动作图片构建原理.

(1) 理论分析

以三轴加速度计为例,其三轴向传感器数值代表了 3 个相互垂直方向的加速度值,从空间角度上讲,同一传感器三轴向加速度在加速度方向上是相互独立的,相关性较弱,不同传感器的不同轴向之间相关性更弱.而不同传感器的相同轴向加速度值大致描绘的是同一个方向的整体加速度变化,类比于图片分类中的 RGB 三通道,每一个轴向具有相同的意义表达,描述的是一个动作在同一个轴向上所有传感器的加速度变化,基于同一个轴方向的加速度矩阵可以通过卷积操作提取更加具有分辨能力的特征,提高识别准确率.

(2) 数据分析

在理论分析的基础上,对原始数据同一传感器不同轴向数据和不同传感器同一轴向数据进行相关性分析.相关系数绝对值在 0.3 以下为极弱相关性,该相关系数矩阵极弱相关性比例远远高于多传感器同一轴向各个传感器加速度之间的比例.取本文实验中的测试数据集 SKODA 的 5 个三轴传感器计算相关性,如图 2 所示,在 105 对数据中,该矩阵极弱相关性系数比例达到 18.09%;而同一轴向所有传感器数据共有 10 对数据,3 个单轴向的所有传感器中则有 30 对数据,在 30 对数据中极弱相关性系数比例仅有 6.67%.这是因为该矩阵中存在大量的同一传感器的三轴之间和不同传感器的不同轴之间的相关性计算.

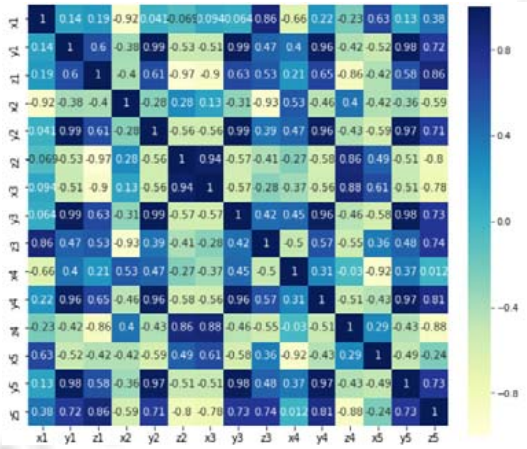


Fig.2 Correlation coefficient matrix of different axises of accelerometer sensors

图 2 不同轴向数据相关系数矩阵

2.2.2 基于多位置三轴传感器二维动作图片构建方法

根据多位置三轴传感器动作图片构建原理,本文提出了基于多位置三轴传感器输入构建方法 T-2D 和基于多位置混合传感器输入构造算法 M-2D.

(1) 基于多位置三轴传感器输入构建方法算法 T-2D.

图 3 描述了基于多位置三轴传感器动作图片构建方法 T-2D.原始时序数据来自于 n 个位置三轴传感器,将所有位置传感器的数据根据(x,y,z)三轴方向进行抽取,分别得到全部 x 轴时序数据、全部 y 轴时序数据和全部 z 轴时序数据,并且分别对这 3 组数据进行归一化处理.根据定义 2,分别对 3 组数据以滑动窗口 T 和预设覆盖率进行滑动窗口划分,获取时序传感器数据的处理单元,该时间单元的传感器数据为动作图片.

3 个动作图片均作为二维卷积神经网络的输入,并用于抽取不同位置传感器的相同轴向数据的时间依赖性和空间依赖性.目前,三轴加速度计是在人体活动识别过程中使用较为普遍的传感器,大多数设备中都包含三轴加速度计,如手环、手机等,因此,本文三轴向传感器使用三轴加速度计.

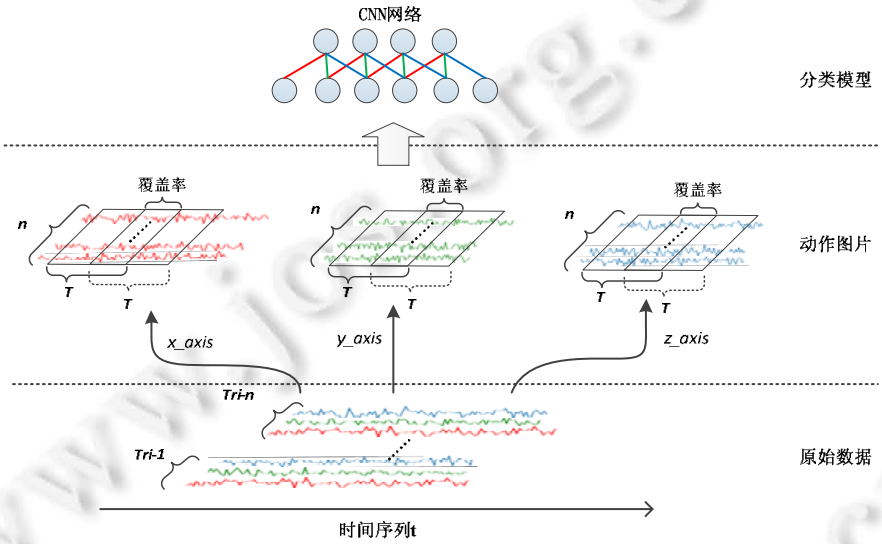


Fig.3 T-2D based on multi-location 3D accelerometer sensors

图 3 基于多位置三轴传感器输入构建方法 T-2D

(2) 基于多位置混合传感器二维输入构建方法 M-2D.

大部分设备内置三轴加速度计的同时,也配有其他的传感器,这些传感器一部分与三轴加速度计一样具有三轴特性的,例如陀螺仪、磁力计等;而有一些就是简单的时序数据,例如心率计.因此,针对这种更为实际且广泛存在的情况,在 T-2D 的三轴加速度计输入构造算法的基础上,本文扩充了一种混合传感器数据输入构造算法 M-2D,在生成三轴加速度的 3 个动作图片的同时,对其他传感器数据按序排列,构成第 4 个动作图片.由于该动作图片在卷积过程中采用一维卷积核而非二维卷积核,仅抽取这些时序数据的时间依赖性.

2.2.3 算法实现

算法 1 描述了本文提出的基于多位置混合传感器二维输入动作图片构建方法 M-2DCNN,由于 M- 2DCNN 是 T-2DCNN 的扩充方法,其不同之处在于增加了算法 1 中的第 4 行和第 7 行.算法 1 输入为 $X=\{(l_1,x_1),\ldots,(l_N,x_N)\}$,其中,每个时刻数据 $x_i=(x_{i1},y_{i1},z_{i1},\ldots,x_{in},y_{in},z_{in},r_{i1},\ldots,r_{im})$,$n$ 为三轴传感器数量,x_{in},y_{in},z_{in} 分别表示三轴数据,r_{im} 表示非三轴传感器.该算法返回的结果是对输入数据按照滑动窗口大小进行分割之后的动作图片以及相应的标签.

算法 1. 基于二维输入动作图片构造方法 M-2D.

输入:$X=\{(l_1,x_1),\ldots,(l_N,x_N)\}$,$ws$;

输出:所有分割好的动作图片以及该组动作图片对应标签 $List((X_{imag},Y_{imag},Z_{imag},R_{imag}),t)$.

//其中,二维卷积核输入动作图片 $X_{imag},Y_{imag},Z_{imag}$ 一维卷积核输入动作图片 R_{imag}.

1. $X_{in}=Load(X,X_{column})$; //抽取 x 轴所有列数据
2. $Y_{in}=Load(X,Y_{column})$; //抽取 y 轴所有列数据

```
3.  Z_in=Load(X,Z_column);                    //抽取 z 轴所有列数据
4.  R_in=Load(X,R_column);                    //抽取剩余所有列数据
5.  for each in (X_in,Y_in,Z_in)              //对每个轴向输入做矩阵归一化
6.      NormMatrix(each);
7.  NormColumn(R_in);                         //对剩余所有列数据以列为单位归一化
8.  for each in (X_in,Y_in,Z_in,R_in)         //对每个卷积输入按照滑动窗口大小进行划分
9.          (X_imag,Y_imag,Z_imag,R_imag)=Segment(each,ws);
10. t=getLabel(X,ws);                         //获取窗口标签
11. Return List((X_imag,Y_imag,Z_imag,R_imag),t);
```

2.3 基于二维卷积核的HAR模型

在基于传感器时序数据的活动识别领域,与传统基于统计学特征不同,CNN 能够在无需专业领域知识的前提下自动地从原始数据中提取具有高分辨能力的特征,因此,CNN 在该领域得到了广泛应用.本文在构建动作图片的基础上,使用卷积神经网络提取特征并分类.

2.3.1 二维卷积网络结构

由于基于多位置混合传感器的二维卷积结构是基于多位置加速度计传感器二维卷积结构的改进方法,图 4 既可以描述 T-2DCNN 又可以描述 M-2DCNN 卷积模型结构,虚线框内为 T-2DCNN 的二维卷积操作.

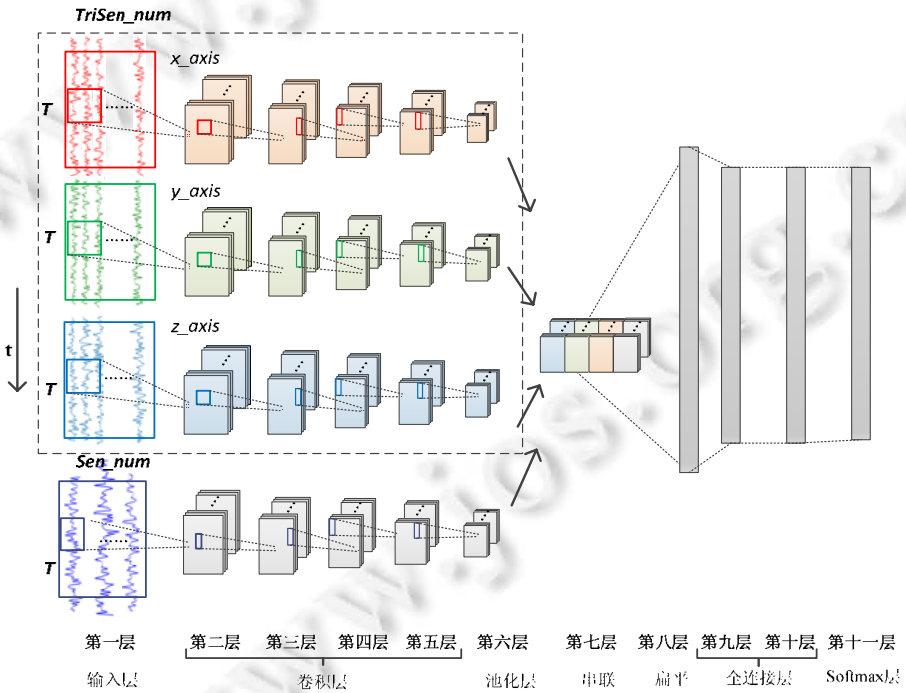


Fig.4 M-2DCNN architecture of 2D CNN based on multi-location mixed sensors

图 4 基于多位置混合传感器的二维卷积模型网络结构 M-2DCNN

该网络结构包括输入层在内共 11 层,其中,卷积层 4 层,池化层 1 层,数据转化层 2 层,全连接层 2 层,连接分类器 1 层.本文中,卷积层和全连接层的 feature map 的数量同文献[26],为 C(64)-C(64)-C(64)-C(64)-D(128)-D(128).多层卷积可以提取出更加具有分辨能力的特征,二维卷积核是为了得到原始数据每个轴向的时间和空间的依赖性,再使用一维卷积核降低参数数量.同时增设最大值池化层用来减少特征数量和训练参数,抽象数据

轮廓特征.如图 4 所示,

- 第 1 层为输入层,从上到下依次为 x 轴、y 轴和 z 轴数据构成的 3 个动作图片,其中,T 为滑动窗口大小,*TriSen_num* 为三轴加速度计的数量;
- 第 2 层~第 5 层均为卷积层,其中,前两层使用二维卷积核,卷积核(过滤器)大小为 $K_h \times K_w$,针对每一个动作图片分别进行卷积操作时,提取原始数据时间和空间的依赖性;同时,为了减少参数数量,第 4 层和第 5 层采用一维卷积核,卷积核大小为 $K_h \times 1$;
- 卷积之后进行池化(第 6 层),本文采用最大值池化可降低由于卷积层参数误差引起的估计均值的偏移;
- 第 7 层和第 8 层均为对池化后的矩阵进行重新整理和融合特征,第 7 层将池化后的矩阵数据沿时间轴方向将特征映射进行拼接,池化层每个轴向数据尺寸为 $64@h_l \times w_l$,则第 7 层数据尺寸为 $64@\ h_l \times (w_l \times 3)$;第 8 层将数据平铺变成一维向量,作为全连接层的输入(第 9 层),两层全连接层抽取更为抽象的特征,实现特征非线性组合,方便连接分类器 Softmax(第 11 层).

如图 4 所示,当处理非三轴加速度计数据时(虚线外),基于 M-2D 共生成 4 个动作图片.在卷积层,针对由非三轴全部使用一维卷积核,其工作同文献[25]卷积操作类似.当传感器中也包含三轴向数据时,将该数据看做 3 个通道,依次排列.由于各个卷积层使用的卷积核在时间维度上的尺寸是相同的,因此在第 7 层串联时,可以与三轴加速度计产生的 3 个特征映射在时间轴上面进行拼接,此时,第 7 层特征维度为 $64@h_l \times (w_l \times 3 + Sen_num)$,其中,*Sen_num* 为第 4 张动作图片的传感器数量,之后的操作则与多位置加速度计传感器二维卷积结构的后四层操作相同.

2.3.2 基于参数共享二维卷积网络结构

尽管卷积网络与传统全连接网络相比极大地降低了参数数量,但 M-2DCNN 和 T-2DCNN 在训练过程中,卷积操作依然产生大量的参数,原因在于每一张动作图片都需要单独的卷积操作.

每一个动作图片的前向传播的卷积过程如图 5 所示,其中,卷积核大小为 $K_h^l \times K_w^l$,第 l 层共有 K_n^l 个 feature maps,第 l–1 层共有 K_n^{l-1} 个 feature maps.

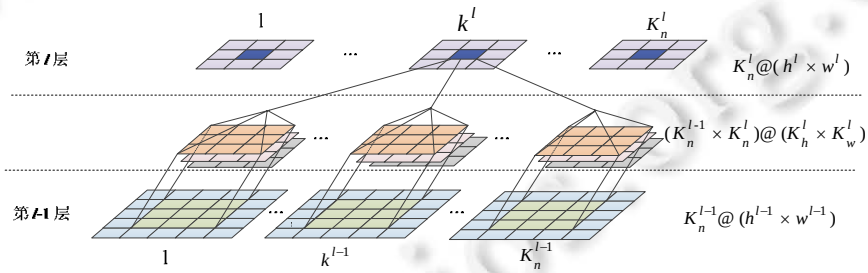


Fig.5 Computation of 2D CNN kernals

图 5 二维卷积计算过程

因此,每一个动作图片卷积过程中的第 l–1 层和第 l 层之间需要训练的参数见公式(2):

$$P_{num} = K_n^{l-1} \times K_n^l \times K_h^l \times K_w^l + K_n^l \tag{2}$$

根据图 4 所示的卷积网络训练模型,每个动作图片的卷积操作过程中模型参数都是独立的,那么 3 个动作图片的每层卷积操作中共需要 $3 \times P_{num}$ 数量的参数,且参数数量是随着卷积核大小以及 feature map 数量的增加而增加的.

因此,本文在 T-2DCNN 和 M-2DCNN 的基础上提出了基于共享卷积层参数的卷积网络模型 TS-2DCNN 和 MS-2DCNN.结合本文中动作图片单一轴向数据的特点以及模型的特点,将利用共享同一组参数对 3 组动作图片进行特征提取,即:在图 4 之中的 *x_axis*,*y_axis* 和 *z_axis* 这 3 个方向使用同一组模型参数取代之前的 3 组模型参数.可以共用同一组参数的原因在于该卷积操作中训练 3 组不同参数与同一组参数都可以达到在各个方向上提取数据特征的目的,因此可在保证识别效果的同时减少训练参数数量.采用表 4 中的卷积核参数,利用公式

(2),对图 4 的网络结构中卷积层之间的参数数量进行计算,见表 1,其中,*Input_X* 表示输入层中 *X* 轴的动作图片,*Cov*2_*X* 表示通过 *X* 轴动作图片获取的第 2 层卷积层,则第 2 层中 *Cov*2_*X* 与上一层 *Input_X* 之间需要训练的参数数量为 3×3×64+64=640,以此类推.TS-2DCNN 和 MS-2DCNN 在卷积层能够将参数数量缩减至原来的三分之一.

Table 1 Comparison of the number of the CNN parameters

表 1 卷积层参数数量对比

卷积层	上一层	方法	
		T-2DCNN/M-2DCNN	TS-2DCNN/MS-2DCNN
Cov2_X	*Input_X*	3×3×64+64=640	640
Cov2_Y	*Input_Y*	640	
Cov2_Z	*Input_Z*	640	
Cov3_X	*Cov2_X*	3×3×64×64+64=36928	36 928
Cov3_Y	*Cov2_Y*	36928	
Cov3_Z	*Cov2_Z*	36928	
Cov4_X	*Cov3_X*	5×1×64×64+64=20544	20 544
Cov4_Y	*Cov3_Y*	20 544	
Cov4_Z	*Cov3_Z*	20 544	
Cov5_X	*Cov4_X*	20 544	20 544
Cov5_Y	*Cov4_Y*	20 544	
Cov5_Z	*Cov4_Z*	20 544	
总计		235 968	78 656

2.3.3 模型训练与实现

- 前向传播过程

本文中使用了二维卷积核和一维卷积.结合图 5,二维卷积在前向传播过程中的计算见公式(3):

$$z_{i,j}^{k^l}=\sum_{k^{l-1}=1}^{K_n^{l-1}}\sum_{u=1}^{K_h^l}\sum_{v=1}^{K_w^l}a_{u,v}^{(k^{l-1},k^1)}x_{(i-1)\cdot h_s^l+u,(j-1)\cdot w_s^l+v}^{k^{l-1}}+b_{k^l} \tag{3}$$

其中, $z_{i,j}^{k^l}$ 表示第 l 层的第 k^l 个特征映射坐标为(i,j)的输入值, $k^l\in[1,K_n^l]$, $a_{u,v}^{(k^{l-1},k^1)}$ 表示 l–1 层的第 k^{l-1} 个特征映射与 l 层的第 k^{l-1} 个特征映射之间的卷积核中,坐标为(u,v)的权重值. $x_{(i-1)\cdot h_s^l+u,(j-1)\cdot w_s^l+v}^{k^{l-1}}$ 为第 k^{l-1} 个特征映射坐标为 $((i-1)h_s^l+u,(j-1)w_s^l+v)$ 的输出值,其中, h_s^l 和 w_s^l 分别为卷积过程中的移动步长.

第 k^l 个特征映射的输出为

$$x_{i,j}^{k^l}=\sigma(z_{i,j}^{k^l}) \tag{4}$$

其中,$\sigma(x)$为激活函数.为了加快求导数速度并防止梯度弥散,本文中使用修正线性单元 ReLU(rectified linear unit)作为激活函数,在池化层,即子采样层,选取卷积层子区域内所有神经元的最大值构成池化层神经元.

- 反向传播过程

本文中的神经网络结构中包含两种情况:卷积层连接卷积层和卷积层连接池化层.当 l 层为卷积层时,第 l–1 层第 k^{l-1} 特征映射的残差项来自于 l 层的所有神经元单元,定义⊗为卷积操作,那么第 k^{l-1} 特征映射中坐标为(i,j)的残差项为

$$\delta_{i,j}^{k^{l-1}}=\frac{\partial L}{\partial z^{k^{l-1}}}=\sum_{k^l=1}^{K_n^l}\frac{\partial L}{\partial z^{k^l}}\cdot\frac{\partial z^{k^l}}{\partial z^{k^{l-1}}}=\sum_{k^l=1}^{K_n^l}\delta^{k^l}\cdot\frac{\partial z^{k^l}}{\partial z^{k^{l-1}}}=\sum_{k^l=1}^{K_n^l}\delta^{k^l}\otimes rot180(a^{(k^{l-1},k^l)})\cdot\sigma'(z^{k^{l-1}}) \tag{5}$$

在得到残差项计算公式之后,第 l–1 层的第 k^{l-1} 个特征映射神经元卷积核 $a_{u,v}^{(k^{l-1},k^1)}$ 的梯度公式为

$$\frac{\partial L}{\partial a_{u,v}^{(k^{l-1},k^l)}}=\sum_{i=1}^{h^l}\sum_{j=1}^{w^l}x_{(i-1)\cdot h_s^l+u,(j-1)\cdot w_s^l+v}^{k^{l-1}}\cdot\delta_{i,j}^{k^l} \tag{6}$$

- 模型实现

算法 2 描述了基于二维卷积动作图片的 HAR 模型训练过程:分别对每个动作图片进行 4 次卷积和一次池化操作(第 4 行、第 5 行),并将特征映射进行串联拼接和扁平处理(第 6 行).第 7 行、第 8 行对应全连接层和输

出层的操作,全连接层的输入为第 6 行处理之后的一维特征向量.第 9 行包含对所有层反向传播计算,当参数变化小于阈值或者训练完全部数据时,可以得到该模型的参数.值得注意的是:在实际运算过程中,是将训练数据划分成批次,以矩阵形式进行计算.

算法 2. 基于二维卷积输入动作图片的 HAR 模型训练.

输入:$List((X_{imag},Y_{imag},Z_{imag},R_{imag}),t)$,$Para=\{\theta\}$; //其中,$\theta$为权重迭代阈值;

输出:M. //得到模型 M.

1. *Initialize*(***A***,***B***); //初始化权重和偏置
2. **for** i in $List((X_{imag},Y_{imag},Z_{imag},R_{imag}),t)$
3. **for** *each* in $(X_{imag},Y_{imag},Z_{imag},R_{imag})$
4. ***FP***(l=*CNN*); //四次卷积
5. ***FP***(l=Maxpool);
6. ***Flatten***(X_f,Y_f,Z_f,R_f); //将卷积池化后特征映射进行串联扁平
7. ***FP***(l=*FC*); //全连接层网络
8. ***FP***(l=*OUT*); //输出层
9. ***BP***(L); //所有层进行反向传播
10. **if** $\Delta A<\theta$
11. **break**;
12. M=***GetModel***(***A***,***B***,*Para*); //得到训练模型

算法 3 描述 HAR 模型预测过程.对于输入无标签数据,使用 M-2DCNN(MS-2DCNN)或者 T-2DCNN(TS-2DCNN)方法对输入进行处理生成相应动作图片,利用训练好的参数预测每个时间窗口数据的标签类别 p_i.

算法 3. HAR 模型预测过程.

输入:$X_{UL}=\{(x_1),\ldots,(x_{N'})\}$,$ws$,模型 M; //其中,X_{UL} 为无标签数据

输出:p_i.

1. //预测过程
2. $List=M\text{-}2D(X_{UL},ws)$;
3. **for** i in $List$
4. $p_i=M(i)$;

3 实验与分析

为了验证本文提出的两种二维输入构造算法 T-2DCNN(TS-2DCNN)和 M-2DCNN(MS-2DCNN)的特征提取的有效性,本文在公开数据集上对比了现有典型且特征提取效果较好的两种算法的姿态识别效果,即传统的一维卷积输入 1DCNN[26]和现有的二维卷积输入 2DCNN[30].文献[26,30]中所述的两种卷积方式是目前在相同公开数据集上表现较好且利用不同数据组织形式和卷积方式提取特征的方法,具有一定代表性.此外,本文重心不是测试现有 CNN 网络结构以及 CNN 和不同分类模型融合的混合模型,而是侧重于对特征的提取效果的验证,因此选取 1DCNN 和 2DCNN 两种对比方法.同时,分类模型全部使用 softmax 进行分类.

3.1 实验数据集描述

在基于传感器信号的人体活动识别领域,根据传感器位置,用于测试的公开数据集通常分为两类:一种是仅用测试者身上的传感器收集的数据;另一种是包含周边环境和物体传感器的数据,例如在抽屉、茶杯之类的物品上面放置传感器,结合身上的传感器一起识别动作.本文中使用的公开数据集属于第 1 种类型的数据——OPPORTUNITY[41]数据集和 SKODA 数据集[42].

- OPPORTUNITY 数据集(https://archive.ics.uci.edu/ml/datasets/opportunity+activity+recognition)

OPPORTUNITY 数据集收集场景为日常生活,主要活动范围在厨房.该数据集数共有 4 名测试者,3 种类型

传感器分别为:测试者身上传感器(body-worn sensors)、物体传感器(object sensors)和周围环境传感器(ambient sensors).本文利用身上传感器识别测试者的动作,包括 7 个 IMU(inertial measurement unit)、12 个三轴加速度传感器,其分布见表 2.对于每一个测试者记录了 6 个文件,其中 5 个 ADL 和 1 个 Drill.ADL 是测试者在自然状态下的活动记录,包括在椅子上躺着、在房间走动以及准备咖啡、三明治等一些列活动;Drill 是对一系列动作的 20 次重复,包括开关冰箱、开关抽屉等活动.对于活动类别,该数据进行了等级划分,分为运动模式识别(modes of locomotion)和姿态识别(gesture recognition),姿态识别分为高层次活动[43]、中层次活动[25,26]和低层次活动[44].高层次活动为抽象活动,例如咖啡时间、三明治时间等;中层次活动为高层次活动细化,例如咖啡时间中包含开关抽屉等活动;低层次活动为中层次活动的活动单元.本文选择中层次动作作为识别的标签.而运动模式识别中描述运动的状态,如站立、躺、坐等.为了描述实验方便,用 GR 和 LM 来代表姿态识中层次标签识别任务和运动模式标签识别任务,标签描述见表 3.在原始数据中存在很多缺失值,因此利用线性插值对缺失数据进行填充.数据采样频率为 30HZ.

Table 2 Location of the sensors in OPPORTUNITY dataset

表 2 OPPORTUNITY 数据集传感器位置分布

位置	IMU	Accelerometer (ACC)
左臂	LUA,LLA	LUA^,LUA_,LWR,LH
右臂	RUA,RLA	RUA_,RUA^,RWR
左腿	LSHOE	
右腿	RSHOE	RKN^,RKN_
主躯干	BACK-I	BACK-A,HIP

Table 3 Details of OPPORTUNITY and SKODA

表 3 OPPORTUNITY 和 SKODA 数据集描述

OPPORTUNITY						SKODA		
GR			LM					
标签	数量	符号	标签	数量	符号	标签	数量	符号
close dishwasher	716	G1	Stand	22 388	L1	write on notepad	1 386	S1
close drawer 3	624	G2	Walk	13 183	L2	open hood	1 643	S2
close drawer 2	444	G3	Sit	9 427	L3	close hood	1 540	S3
close door 1	871	G4	Lie	1 674	L4	check gaps (front door)	1 141	S4
close door 2	927	G5	Null	9 693	L0	open left front door	675	S5
close drawer 1	453	G6				close left front door	639	S6
close fridge	1 010	G7				close both left door	1 188	S7
toggle switch	740	G8				check trunk gaps	1 315	S8
open dishwasher	765	G9				open and close trunk	1 573	S9
open drawer 3	634	G10				check steering wheel	884	S10
open drawer 2	506	G11						
open door 1	923	G12						
open door 2	982	G13						
open drawer 1	525	G14						
open fridge	1 157	G15						
drink from cup	3 559	G16						
clean table	1 001	G17						
Null	40 528	G0						

- SKODA 数据集(http://www.ife.ee.ethz.ch/research/activity-recognition-datasets.html)

SKODA 数据集描述汽车工人对汽车进行操作的场景,该数据集总共包含 19 个 USB 传感器,每个传感器包含三轴加速度传感器,其中,在右臂和左臂上分别放置10个和9个传感器,其具体位置见文献[42].在本文中,我们仅适用右臂中的 10 个传感器进行测试,传感器编号从手腕到大臂依次为{2,27,16,29,1,18,14,24,22, 21},该数据集包含 10 类标签,具体动作见表 3.

3.2 实验设置

- 参数设置

本文实验环境为 Google Colaboratory 提供的 GPU.编程语言为 Python 2.0,框架为 Keras.本文中实验过程所

需参数见表 4.为了验证本文所提算法在不同滑动窗口下的性能,因此设计了滑动窗口参数范围,在没有明确指出滑动窗口参数时缺省值为 24,即 24 个时刻数据作为一个数据单元进行训练和预测,该窗口最后一个时刻数据作为该数据单元的标签,窗口覆盖率为 50%.由于 SKODA 采样频率较大,为了实现实验参数共享并且方便与 OPPORTUNITY 数据集对比,对 SKODA 数据进行降采样,降采样之后频率为 32HZ,与 OPPORTUNITY 数据集采样频率相当.当滑动窗口为 24 时,OPPORTUNITY 数据集和 SKODA 数据集各类训练样本数量见表 3,并对各类标签进行了符号化表示.结合数据传感器类型特点,本文利用 OPPORTUNITY 数据集验证 M-2DCNN (MS-2DCNN)识别效果,利用 SKODA 验证 T-2DCNN(TS-2DCNN)识别效果.

Table 4 Experiment parameters

表 4 实验参数

	参数	缺省值
滑动窗口	24,48	24
OPPORTUNITY 传感器类别	8 类: ACC(右臂)和 IMU(左臂,右臂,主躯干) 10 类: ACC(右臂,右腿)和 IMU(左臂,右臂,主躯干) 14 类: ACC(右臂,右腿,左臂)和 IMU(左臂,右臂,主躯干) 16 类: ACC(右臂,右腿,左臂,主躯干)和 IMU(左臂,右臂,主躯干)	16 类
SKODA 传感器类别	3 类:2,27,16(小臂) 8 类:2,27,16,18,14,24,22,21(小臂,大臂) 10 类:2,27,16,18,14,24,22,21,29,1(小臂,大臂,手肘)	10 类
卷积核	第 2 层:3×3&3×1 第 3 层:3×3&3×1 第 4 层:5×1 第 5 层:5×1 第 6 层:3×1	同参数
学习速率	0.01	同参数
卷积步长	1	同参数

同时,本文从传感器数量角度对提出的卷积网络模型以及共享参数模型的特征抽取效果进行了验证,对于 OPPORTUNITY 数据集,添加部分 IMU 传感器并按照传感器在人体分布情况逐次添加传感器,总共有 4 组传感器,默认缺省值为最后一组包含 16 类传感器.类似地,对于 SKODA 数据集,按照传感器分布位置区域进行添加,共有 3 组:第 1 组包含小臂传感器 3 类,第 2 组包含大小臂传感器 8 类,最后一组包含 10 类,默认为最后一组.

在使用 OPPORTUNITY 和 SKODA 数据集的文献中,由于各种参数设置不同以及衡量标准不同,无法统一衡量,因此本文结合参考文献中的方法,统一了输入数据参数进行对比实验.结合第 2.3 节中所提出的卷积框架,卷积层和池化层的卷积核参数见表 4.对于第 2 层和第 3 层卷积层,该结构在梯度下降过程中学习速率为 0.01.本文中用于对比实验的一维卷积 1DCNN 输入构造和网络结构来自文献[26],为了使实验对比结果更加令人信服,将原文献中前两个卷积层的卷积核定义 3×1,网络架构与原文相同.二维卷积 2DCNN 输入构造来自文献[30],还原了文献中的二维卷积输入构造和卷积网络结构.

- 评价指标

本文使用 F_1 值作为识别效果的评价指标,F_1 同时受到精确率和召回率的影响,计算公式为

$$F_1=\sum_{i=1}^{C}F_i=\sum_{i=1}^{C}w_i\cdot\frac{2P_i\cdot R_i}{P_i+R_i} \tag{7}$$

其中,P_i 和 R_i 分别为第 i 类的精确率和召回率,w_i 为该类样本占所有样本比例,目的是解决类间样本数量不平衡问题.

- 训练集和测试集选取

对于 SKODA 数据集,训练集和测试集比例为 4:1.对于 OPPORTUNITY 数据集,结合文献中的实验源码,训练数据包括测试者 1 的所有 ADL 和 Drill、测试者 2 和测试者 3 的 ADL1,ADL2,ADL3 和 Drill,测试数据为测试者 2 和测试者 3 的 ADL4 和 ADL5.所有数据均没有进行数据类间均衡.在 OPPORTUNITY 数据集中,尽管 IMU 传感器数据中包含多维度数据,我们仅仅抽取其中的三轴加速度与 ACC 的三轴加速度结合,按照三轴方向构造

3 组二维输入,IMU 传感器剩余数据包括三轴陀螺仪和磁力器均作为一维卷积输入.

3.3 实验结果与分析

- F_1 对比分析

表 5 和表 6 分别描述了本文提出的两种输入构造法以及经过优化参数的网络在 OPPORTUNITY 数据集和 SKODA 数据集上面的表现.对于 OPPORTUNITY 数据中 GR 标签和 LM 标签,均分为两种情况:一种是包含 Null class 类,一种是不包含 Null class 类.

Table 5 F_1-value on OPPORTUNITY dataset

表 5 OPPORTUNITY 数据集 F_1 值

数据集参数		方法			
ws	标签类型	1DCNN	M-2DCNN	MS-2DCNN	2DCNN
24	GR	0.794 6	0.813 0	**0.819 1**	0.752 3
	GR (null)	0.895 4	**0.899 2**	0.899 1	0.886 1
	LM	0.903 9	0.919 5	**0.920 7**	0.908 3
	LM (null)	0.856 9	0.868 5	**0.875 0**	0.817 1
48	GR	0.743 9	0.772 2	**0.801 5**	0.759 8
	GR (null)	0.878 4	0.891 2	**0.893 0**	0.875 6
	LM	0.906 4	0.913 7	**0.921 1**	0.898 0
	LM (null)	0.865 0	0.862 0	0.854 9	**0.870 6**

Table 6 F_1-value on SKODA dataset

表 6 SKODA 数据集 F_1 值

数据集参数	方法			
ws	1DCNN	T-2DCNN	TS-2DCNN	2DCNN
24	0.949 1	**0.951 8**	0.951 7	0.940 9
48	0. 921 2	**0.963 7**	0.949 3	0.937 5

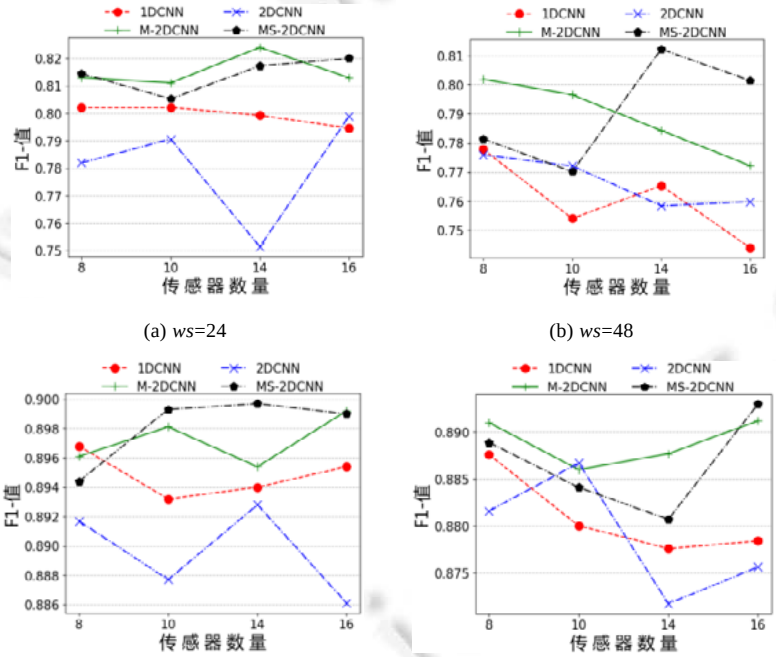
在表 5 中,加粗数字表示同等实验参数和环境下最高 F_1 值.通过表 5 亦可知:滑动窗口取 48 时,GR 和 GR (null)的 F_1 值与窗口为 24 时相比均有降低.无论窗口取 24 还是 48,基于本文提出的输入构造方法的卷积神经网络 M-2DCNN 和 MS-2DCNN 与现有方法 1DCNN 和 2DCNN 的 F_1 值相比,在绝大多数情况下表现效果良好.这是因为 OPPORTUNITY 传感器遍布全身且动作具有连续性,本文所述动作图片构建方法能够有效地卷积出单一轴方向上的数据特征,其时空特征表现良好.在默认参数情况下,针对 4 种分类任务,利用 2DCNN 实现 GR 分类时,MS-2DCNN 的 F_1 提升值最大为 6.68%,M-2DCNN 次之为 6.07%;在利用 1DCNN 实现 GR(null)时,F_1 提升值最小,为 0.38%.而在窗口取 48 时,LM(null)分类效果中,本文方法略占劣势,与最高值相比经降低了 1.57%.其原因在于 1DCNN 和 2DCNN 更依赖于时序数据的时间特征,而本文方法需要较多的训练样本.在这 4 种方法中,LM 识别 F_1 值总体上比 GR 的高.这是因为 LM 识别种类少,包括 Null class 在内共 5 种类别,而 GR 包含 Null class 在内共有 18 种类别,相比于 LM,GR 类间差异会相对不明显,因此整体 F_1 值低于 LM.在 GR 中,由于 Null class 数量相对于其他类别数据数量明显增多,占 GR 样本总数的 71.9%,因此包含 Null class 的 GR 的识别效果要好于不包含 Null class 的 GR.在 LM(null)识别任务中,Null class 仅占总样本数 17.2%,Null class 的加入会造成整体 F_1 值降低.

表 6 描述了本文算法在 SKODA 数据集上的表现,由表可知:T-2DCNN 和 TS-2DCNN 表现良好,在默认参数情况下,T-2DCNN 的 F_1 高达 95.18%,与相同条件下使用朴素 2DCNN0 相比提高了 1.09%.该数量级 F_1 值提升非常有意义,说明本文提出的二维卷积输入构建方法可以更高层次地抽取传感器数据之间的空间依赖性;同样与 1DCNN 相比,比单纯的抽取时间依赖性提高了 0.27%.当超过采样频率之后即窗口取 48 时,在默认传感器数量情况下,整体 F_1 值有所下降.其原因在于训练样本随着窗口增加而减少,训练样本不足导致.同时, T-2DCNN 和 TS-2DCNN 依然保持优势.这是因为样本不足时,1DCNN 时间特征不足,2DCNN 空间特征表达没有 T-2DCNN 和 TS-2DCNN 明显.

- 传感器数量与 F_1 关系分析

为了更好地说明本文提出的输入组织形式可以有效地抽取数据的空间依赖特征和时间依赖特征,并在不同传感器数量以及输入数据窗口的情况下保持整体的平均识别效果,本节从传感器数量角度出发,采用表 4 中传感器数量参数,对基于 4 种卷积网络的识别效果进行验证.OPPORTUNITY 数据集和 SKODA 数据集传感器添加方式均是将不同区域的传感器按照逐一区域添加.

图 6 描述了 OPPORTUNITY 中 GR 和 GR(null)F_1 传感器数量之间的关系.随着传感器数量增加,不同滑动窗口大小的 F_1 趋势不同,甚至出现反复现象.由此可以说明传感器数量越多,其计算效果不一定越好.原因在于识别效果和传感器空间分布以及滑动窗口的变化有关,同时,传感器较多容易引起过拟合,如 ws=48 时 GR 中 M-2DCNN,1DCNN 和 2DCNN.从整体上讲,GR 中的 M-2DCNN 和 MS-2DCNN 可以有效地兼顾数据时间和空间的特征,在传感器数量较少时,能够有从单一轴向动作图片中挖掘时间依赖性;在传感器数量较多时,本文描述的动作图片可以挖掘同一轴向上的不同位置传感器的空间依赖性.相对于 1DCNN 和 2DCNN 的 F_1 提升较大.但在 GR(null)中,虽然 F_1 值也有提升,但提升没有 GR 中明显.其原因在于 null 类在训练过程中所占比例较高.随着传感器数量增多,F_1 值先下降后上升,这是因为新加入的传感器属性能够更好地表达特征,降低了之前加入传感器数造成的数据干扰.



(a) ws=24 (b) ws=48

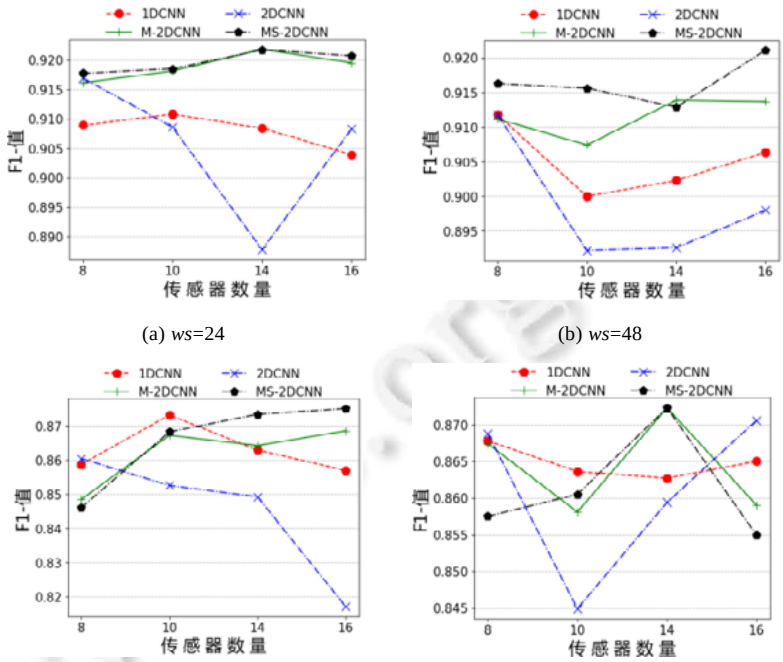
(c) ws=24(null) (d) ws=48(null)

Fig.6 F_1-values of GR with different number of sensors on OPPORTUNITY

图 6 OPPORTUNITY 数据集不同数量传感器的 GR 任务的 F_1 值

图 7 描述了 OPPORTUNITY 中 LM 和 LM(null)中的动作识别效果随着传感器数量变化的的趋势.在 LM 中,M-2DCNN 和 MS-2DCNN 在窗口为 24 时的 F_1 值整体上随着传感器数量增加而提高,在窗口为 48 时小范围波动,而其他两种方法在窗口为 24 时整体低于 M-2DCNN 和 MS-2DCNN 并呈现下降趋势.这说明 M-2DCNN 和 MS-2DCNN 在小窗口时可以有效地从少量多个位置、多种传感器中抽取空间特征.当 LM(null)分类任务窗口为 48 时,1DCNN 识别效果较为稳定,这说明 1DCNN 能够在大窗口传感器数据中得到时间特征.而 2DCNN 的 F_1 值范围波动较大,M-2DCNN 和 MS-2DCNN 也有小范围波动,说明在该窗口下数据空间特征受传感器分布影响较大,不能够在逐一区域添加传感器过程中得到稳定的时空特征.当加入 null class 之后,4 种方法对于特征学

习能力下降,这是因为在 LM 标签中 null class 学习样本较少,增大了误差.



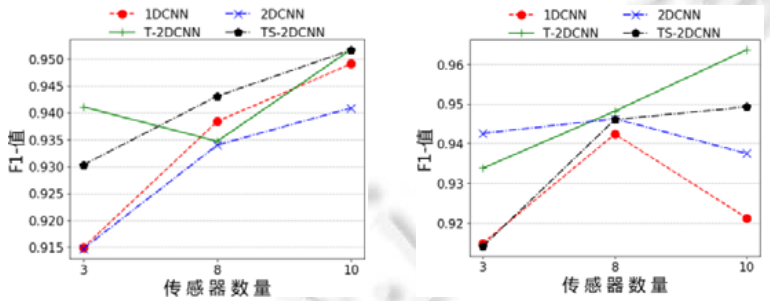
(a) *ws*=24 (b) *ws*=48

(c) *ws*=24(null) (d) *ws*=48(null)

Fig.7 F_1-values of LM with different number of sensors on OPPORTUNITY

图 7 OPPORTUNITY 数据集不同数量传感器的 LM 任务的 F_1 值

图 8 描述了 SKODA 数据集上不同滑动窗口情况下 F_1 随着传感器数量变化.



(a) *ws*=24 (b) *ws*=48

Fig.8 F_1-values of recognition with different number of sensors on SKODA

图 8 SKODA 数据集不同数量传感器的活动识别 F_1 值

随着传感器数量的增加,T-2DCNN,TS-2DCNN 呈现上升趋势.这是因为动作图片构建需要一定数量的传感器,在不受噪音传感器干预的情况下,传感器数量越多,则在每一个轴向的动作图片中所提取动作的空间依赖性越强.在窗口为 48 时,1DCNN 和 2DCNN 呈现了先上升后下降的趋势,其原因是在该参数情况下,训练数据较少,且传感器数量较多,产生了过拟合现象.同时可观测到,TS-2DCNN 和 T-2DCNN 在传感器数量为 3 时 F_1 值很低.这是由于该数量传感器构成的 3 个动作图片的空间特征分辨力不足,但随着传感器数量增多,F_1 值逐渐提高.

为了更加清晰地描述基于这 4 种方法的在各种传感器分布情况下的卷积网络效果,我们对图 6~图 8 获取

最高 F_1 值的方法出现的次数进行了统计,见表 7 和表 8.该统计的目的在于说明这 4 种方法在不具体指出传感器位置分布的情况下的总体识别效果,即该算法的通用性,而不针对于某一种情况下作比较.在 OPPORTUNITY 上,M-2DCNN 和 MS-2DCNN 获取最高 F_1 值的频次较高;SKODA 数据集上,本文所述方法频次最高.因此说明我们提出的方法对实验参数和数据集的变化具有较好的的适应能力,可以有效地从多位置传感器组合中抽取有效的特征.同时,通过图 6~图 8 也可以看出:基于共享参数的卷积网络 MS-2DCNN 和 TS-2DCNN 与 M-2DCNN 和 T-2DCNN 达到同等效果,即使 F_1 在某些情况下略有降低,但整体好于 2DCNN 和 1DCNN.

Table 7 Frequency of the largest F_1-value on OPPORTUNITY

表 7 OPPORTUNITY 数据集最高 F_1 值次数

方法	标签类型				
	GR	GR (null)	LM	LM (null)	总计
1DCNN	0	1	0	2	3
M/MS-2DCNN	8	6	8	3	25
2DCNN	0	1	0	3	4

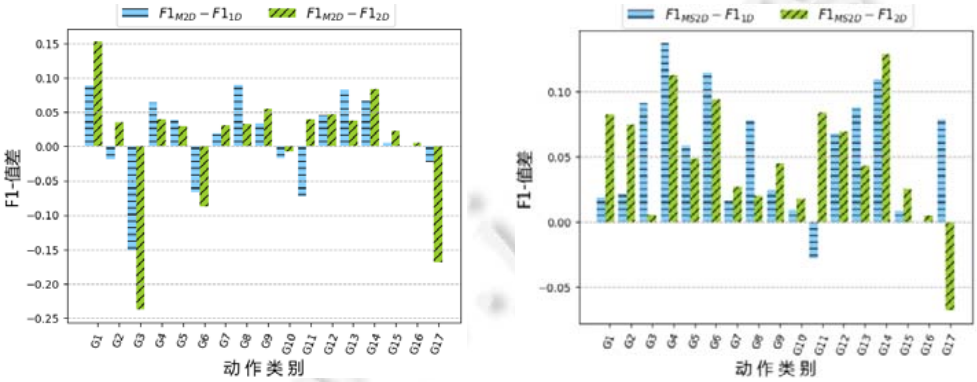
Table 8 Frequency of the largest F_1-value on SKODA

表 8 SKODA 数据集最高 F_1 值次数

方法	滑动窗口		
	24	48	总计
1DCNN	0	0	0
T/TS-2DCNN	3	2	5
2DCNN	0	1	1

- 相同类别 F_1 值对比分析

本文未采用混淆矩阵对每一类的识别效果进行分析,而是纵向对比了每一类采用不同方法的 F_1 差值.可以更清晰和直观地观测到不同的输入构造和卷积对每一类识别效果的影响.图 9 描述了 OPPORTUNITY 中 M-2DCNN 和 MS-2DCNN 每个类别 F_1 值与待对比方法法 F_1 值差异,横坐标为类别标签符号,其符号与类别之间对应关系见表 3,纵坐标为不同方法 F_1 值之间的差值.



(a) M-2DCNN 差异 (b) MS-2DCNN 差异

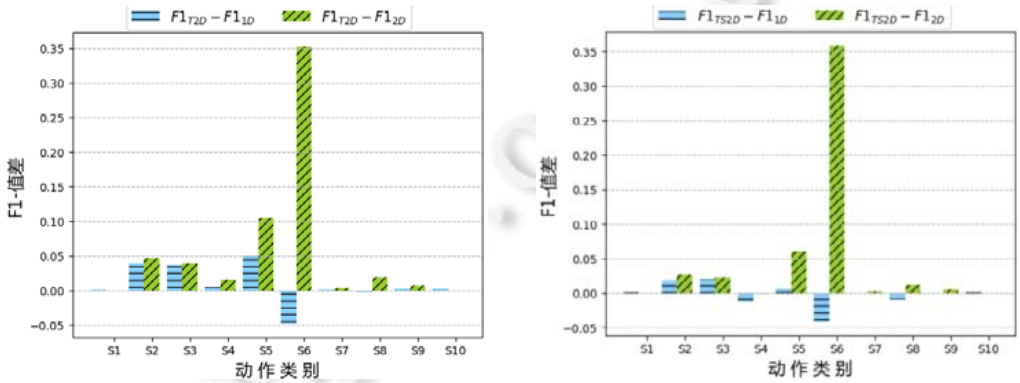
Fig.9 F_1-values' difference of GR in different CNN architectures on OPPORTUNITY (ws=24)

图 9 OPPORTUNITY 数据集 GR 各类别不同输入构造方法 F_1 值差异(ws=24)

在 GR 的 17 类动作中,M-2DCNN 在 11 个动作中的识别效果均优于其他两种方法,完全低于其他两种方法的有 4 类动作(G3,G6,G10 和 G17).在 MS-2DCNN 中,不存在完全低于其他两种方法的动作标签,对于类间差异较小的动作,如开关不同的抽屉和门,本文方法能够有效地进行区分,同时说明了 OPPORTUNITY 数据集中传感器分布使得同一轴向的动作图片数据特征表达较好,即使共享卷积参数,依然在各类别的识别过程中表现良好.

但对于 G17(clean table)的动作识别效果不如 2DCNN,这是因为 G17 不同于其他动作,该动作持续时间长,因此在同一轴向小窗口数据上的空间依赖性表达不如 2DCNN 明显.

图 10 描述了 SKODA 数据集中,默认情况下使用 T-2DCNN 和 TS-2DCNN 与待对比方法 F_1 值的差异. T-2DCNN 在 9 种动作中的识别效果均优于其他两种方法;TS-2DCNN 可以较好地识别出 8 种动作;而 1DCNN 在第 6 类动作识别效果略好,这是因为 S6 动作样本数量较少,且该动作完全由左臂完成,文中所使用 SKODA 传感器均佩戴在右臂,在完成动作时右臂动作幅度不大,使用一维卷积核时能够完整地抽取传感器各个轴向的时间特征,该动作空间特征不明显,出现了本文方法识别效果不如 1DCNN,但却优于 2DCNN 的现象.



(a) 2DCNN 差异 (b) TS-2DCNN 差异

Fig.10 F_1-Values' difference in different CNN architectures on SKODA (ws=24)

图 10 SKODA 数据集各类别不同输入构造方法 F_1 值差异(ws=24)

4 结 论

本文针对现有二维卷积输入构建方法中对多位置三轴向传感器相同轴向数据之间的空间依赖性挖掘不足的现象进行研究并提出解决方案.本文提出了 T-2DCNN 和 M-2DCNN 两种网络结构,其核心是利用 T-2D 和 M-2D 构建方法将多个位置的三轴加速度计的数据划分成 3 张独立的动作图片用于二维卷积核的输入,同时,结合非三轴传感器构建一维卷积核输入的动作图片,融合多张动作图片所得的特征映射来抽取高层次特征,实现了以本文所提方案为特征抽取方法的 HAR 模型.同时,提出了基于共享参数 TS-2DCNN 和 MS-2DCNN 的网络结构,在保证识别准确率的同时,减少了卷积层训练参数数量,并通过设置不同实验参数验证了本文所述方法的识别效果.在未来研究工作中,将结合项目研究组采集的数据对真实环境建模和动作识别.

References:

[1] Wang JD, Chen YQ, Hao SJ, Peng XH, Hu LS. Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 2018,3(33):1–9. [doi: 10.1016/j.patrec.2018.02.010]

[2] Nweke HF, Teh YW, Al-garadi MA, Alo UR. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. Expert Systems with Applications, 2018,105:233–261. [doi: 10.1016/j.eswa.2018.03.056]

[3] Poppe R. A survey on vision-based human action recognition. Image and Vision Computing, 2010,28(6):976–990. [doi: 10.1016/j.imavis.2009.11.014]

[4] Asadi-Aghbolaghi M, Clapes A, Bellantonio M, Escalante HJ, Ponce-López V, Baró X, Guyon I, Kasaei S, Sergio E. A survey on deep learning based approaches for action and gesture recognition in image sequences. In: Proc. of the 12th IEEE Int'l Conf. on Automatic Face & Gesture Recognition. IEEE, 2017. 476–483. [doi: 10.1109/FG.2017.150]

[5] Lara OD, Labrador MA. A survey on human activity recognition using wearable sensors. IEEE Communications Surveys and Tutorials, 2013,15(3):1192−1209. [doi: 10.1109/SURV.2012.110112.00192]

[6] Kasteren TL, Englebienne G, Kröse BJA. An activity monitoring system for elderly care using generative and discriminative models. Personal and Ubiquitous Computing, 2010,14(6):489−498. [doi: 10.1007/s00779-009-0277-9]

[7] Avci A, Bosch S, Marin-Perianu M, Marin-Perianu R, Paul H. Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In: Proc. of the 23rd Int'l Conf. on Architecture of Computing Systems (ARCS). 2010. 167−176.

[8] Mazilu S, Blanke U, Hardegger M, Röster G, Gazit E, Hausdorff JM. GaitAssist: A daily-life support and training system for parkinson's disease patients with freezing of gait. In: Proc. of the 32nd Annual ACM Conf. on Human Factors in Computing Systems. ACM Press, 2014. 2531−2540. [doi: 10.1145/2556288.2557278]

[9] Rashidi P, Cook DJ. Keeping the resident in the loop: Adapting the smart home to the user. IEEE Trans. on Systems, Man, and Cybernetics—Part A: Systems and Humans, 2009,39(5):949−959. [doi: 10.1109/TSMCA.2009.2025137]

[10] Tolstikov A, Hong X, Biswas J, Nugent C, Chen L, Parente G. Comparison of fusion methods based on dst and DBN in human activity recognition. Journal of Control Theory and Applications, 2011,9(1):18−27. [doi: 10.1007/s11768-011-0260-7]

[11] Hong J, Ohtsuki T. A state classification method based on space-time signal processing using SVM for wireless monitoring systems. In: Proc. of the 2011 IEEE 22nd Int'l Symp. on Personal Indoor and Mobile Radio Communications (PIMRC). IEEE, 2011. 2229−2233. [doi: 10.1109/PIMRC.2011.6139913]

[12] Stiefmeier T, Roggen D, Ogris G, Lukowicz P, Tröster G. Wearable activity tracking in car manufacturing. IEEE Pervasive Computing, 2008,7(2):42−50. [doi: 10.1109/MPRV.2008.40]

[13] Kautz T, Groh BH, Hannink J, Jensen U, Strubberg H, M.Eskofier B. Activity recognition in beach volleyball using a deep convolutional neural network. Data Mining and Knowledge Discovery, 2017,31(6):1678−1705. [doi: 10.1007/s10618-017-0495-0]

[14] Parkka J, Ermes M, Korpipaa P, eMantyjarvi J, Peltola J, Korhonen I. Activity classification using realistic data from wearable sensors. IEEE Trans. on Information Technology in Biomedicine, 2006,10(1):119−128. [doi: 10.1109/TITB.2005.856863]

[15] Shoaib M, Bosch S, Incel OD, Scholten H, Havinga P. Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors, 2016,16(4):426. [doi: 10.3390/s16040426]

[16] Chen YP, Yang JY, Liou SN, Lee GY, Wang JS. Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. Applied Mathematics and Computation, 2008,205(2):849−860. [doi: 10.1016/j.amc.2008.05.099]

[17] Figo D, Diniz PC, Ferreira DR, Cardoso J. Preprocessing techniques for context recognition from accelerometer data. Personal and Ubiquitous Computing, 2010,14(7):645−662. [doi: 10.1007/s00779-010-0293-9]

[18] He Z, Jin L. Activity recognition from acceleration data based on discrete consine transform and SVM. In: Proc. of the IEEE Int'l Conf. on Systems, Man and Cybernetics. IEEE, 2009. 5041−5044. [doi: 10.1109/ICSMC.2009.5346042]

[19] Zeng M, Nguyen LT, Yu B, *et al*. Convolutional neural networks for human activity recognition using mobile sensors. In: Proc. of the 6th Int'l Conf. on Mobile Computing, Applications and Services (MobiCASE). IEEE, 2014. 197−205. [doi: 10.4108/icst.mobicase.2014.257786]

[20] Kuang XH, He J, Hu ZH, Zhou Y. Comparison of deep feature learning methods for human activity recognition. Application Reaearch of Computers, 2017,35(9) (in Chinese with English abstract). http://www.arocmag.com/article/02-2018-09-005.html [doi: 10.3969 /j.issn.1001-3695.2018.09.059]

[21] Hammerla NY, Halloran S, Ploetz T. Deep, convolutional, and recurrent models for human activity recognition using wearables. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence. New York, 2016. 1533−1540.

[22] Chen Y, Xue Y. A deep learning approach to human activity recognition based on single accelerometer. In: Proc. of the 2015 IEEE Int'l Conf. on Systems, Man, and Cybernetics (SMC). Hong Kong: IEEE, 2015. 1488−1492. [doi: 10.1109/SMC.2015.263]

[23] Ronao CA, Cho SB. Deep convolutional neural networks for human activity recognition with smartphone sensors. In: Proc. of the Int'l Conf. on Neural Information Processing. Cham: Springer-Verlag, 2015. 46−53. [doi: 10.1007/978-3-319-26561-2_6]

[24] Ronao CA, Cho SB. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 2016,59:235−244. [doi: 10.1016/j.eswa.2016.04.032]

[25] Yang JB, Nguyen MN, San PP, Li XL, Krishnaswamy S. Deep convolutional neural networks on multichannel time series for human activity recognition. In: Proc. of the 24th Int'l Conf. on Artificial Intelligence. Buenos Aires. 2015. 3995−4001.

[26] Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors, 2016,16(1):115. [doi: 10.3390/s16010115]

[27] Teng QL, Kelishomi AE, Cai ZM. A deep learning model for human activity recognition using motion sensor data. Journal of Xi'an Jiaotong University, 2018,52(8):1−8 (in Chinese with English abstract). [doi: 10.7652/xjtuxb201808010]

[28] Chen B, Yu QT, Chen TM. On deep-learning-model-based sensor activity recognition. Journal of Zhejiang University of Technology, 2018,46(4):375−381 (in Chinese with English abstract).

[29] Babu GS, Zhao P, Li XL. Deep convolutional neural network based regression approach for estimation of remaining useful life. In: Proc. of the Int'l Conf. on Database Systems for Advanced Applications. Cham: Springer-Verlag, 2016. 214−228. [doi: 10.1007/978-3-319-32025-014]

[30] Ha S, Yun JM, Choi S. Multi-modal convolutional neural networks for activity recognition. In: Proc. of the IEEE Int'l Conf. on Systems, Man, and Cybernetics (SMC). IEEE, 2015. 3017−3022. [doi: 10.1109/SMC.2015.525]

[31] Ha S, Choi S. Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. In: Proc. of the Int'l Joint Conf. on Neural Networks (IJCNN). IEEE, 2016. 381−388. [doi: 10.1109/IJCNN.2016.7727224]

[32] Ravi D, Wong C, Lo B, Yang GZ. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. IEEE Journal of Biomedical and Health Informatics, 2017,21(1):56−64. [doi: 10.1109/JBHI.2016.2633287]

[33] Jiang W, Yin Z. Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proc. of the 23rd ACM Int'l Conf. on Multimedia. ACM, 2015. 1307−1310. [doi: 10.1145/2733373.2806333]

[34] Chen K, Yao L, Gu T, Yu Z, Wang XZ, Zhang DL. Fullie and wiselie: A dual-stream recurrent convolutional attention mmmodel for Activity Recognition. arXiv preprint arXiv:1711.07661, 2017.

[35] Wu J, Xiao KC. Human activity recognition based on deep convolutional neural networks. Journal of Huazhong University of Science and Technology (Natural Science), 2016,44:190−194 (in Chinese with English abstract). [doi: 10.13245/j.hust.16S139]

[36] Wang ZM, Cao HJ, Fan L. Human activity recognition, deep learning, convolutional neural networks. Computer Science, 2016, 43(Z11):56−58 (in Chinese with English abstract). [doi: 10.11896/j.issn.1002-137X.2016.11A.012]

[37] Wang ZM, Zhang C, Heng X. Research on new human behavior recognition method basd on CNN and decision tree. Application Research of Computers, 2017,34(12):3569−3572 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-3695.2017.12.011]

[38] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proc. of the Advances in Neural Information Processing Systems. 2012. 1097−1105.

[39] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Li FF. Large-scale video classification with convolutional neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 1725−1732.

[40] Abdel-Hamid O, Mohamed A, Jiang H, Deng L, Penn G, Yu D. Convolutional neural networks for speech recognition. ACM Trans. on Audio, Speech, and Language Processing, 2014,22(10):1533−1545. [doi: 10.1109/TASLP.2014.2339736]

[41] Roggen D, Calatroni A, Rossi M, *et al*. Collecting complex activity datasets in highly rich networked sensor environments. In: Proc. of the 7th Int'l Conf. on Networked Sensing Systems (INSS). IEEE, 2010. 233−240. [doi: 10.1109/INSS.2010.5573462]

[42] Zappi P, Lombriser C, Stiefmeier T, *et al*. Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection. In: Proc. of the Wireless Sensor Networks. Heidelberg Berlin: Springer-Verlag, 2008. 17−33. [doi: 10.1007/978-3-540-77690-1_2]

[43] Edel M, Köppe E. Binarized-blstm-rnn based human activity recognition. In: Proc. of the Int'l Conf. on Indoor Positioning and Indoor Navigation (IPIN). IEEE, 2016. 1−7. [doi: 10.1109/IPIN.2016.7743581]

[44] Xu Y, Shen Z, Zhang X, Gao Y, Deng S, Wang Y, Fan Y, Chang E. Learning multi-level features for sensor-based human action recognition. Pervasive and Mobile Computing, 2017,40:324−338. [doi: /10.1016/j.pmcj.2017.07.001]

附中文参考文献:

[20] 匡晓华,何军,胡昭华,周媛.面向人体行为识别的深度特征学习方法比较.计算机应用研究,2017,35(9). http://www.arocmag.com/article/02-2018-09-005.html [doi: 10.3969/j.issn.1001-3695.2018.09.059]

[27] 滕千礼,Kelishomi AE,蔡忠闵.采用运动传感器的人体运动识别深度模型.西安交通大学学报,2018,52(8):1−8. [doi: 10.7652/xjtuxb201808010]

[28] 陈波,余秋婷,陈铁明.基于传感器人体行为识别深度学习模型的研究.浙江工业大学学报,2018,46(4):375−381.
[35] 吴军,肖克聪.基于深度卷积神经网络的人体活动识别.华中科技大学学报,2016,44:190−194. [doi: 1013245/j.hust.16S139]
[36] 王忠民,曹洪江,范琳.一种基于卷积神经网络深度学习的人体行为识别方法.计算机科学,2016,43(Z11):56−58. [doi: 10.11896/j.issn.1002-137X.2016.11A.012]
[37] 王忠民,张琮,衡霞.CNN 与决策树结合的新型人体行为识别方法研究.计算机应用研究,2017,34(12):3569−3572. [doi: 10.3969/j.issn.1001-3695.2017.12.011]

邓诗卓(1990－),女,辽宁凌源人,博士生,主要研究领域为大数据分析,机器学习,时序数据分析.

杨传贵(1995－),男,硕士生,主要研究领域为机器学习,时序数据分析.

王波涛(1968－),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为云计算,大数据,位置服务,隐私保护.

王国仁(1966－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为不确定数据管理,数据密集型计算,可视媒体数据管理与分析,非结构化数据管理,分布式查询处理与优化技术.

www.jos.org.cn
www.jos.org.cn