

# 基于滑动窗口和卷积神经网络的可穿戴人体活动识别技术

何 坚<sup>\*①②</sup> 郭泽龙<sup>①</sup> 刘乐园<sup>①</sup> 苏予涵<sup>①</sup>

<sup>①</sup>(北京工业大学信息学部 北京 100124)

<sup>②</sup>(北京市物联网软件与系统工程技术研究中心 北京 100124)

**摘 要:** 由于缺少统一人体活动模型和相关规范,造成已有可穿戴人体活动识别技术采用的传感器类别、数量及部署位置不尽相同,并影响其推广应用。该文在分析人体活动骨架特征基础上结合人体活动力学特征,建立基于笛卡尔坐标的人体活动模型,并规范了模型中活动传感器部署位置及活动数据的归一化方法;其次,引入滑动窗口技术建立将人体活动数据转换为RGB位图的映射方法,并设计了人体活动识别卷积神经网络(HAR-CNN);最后,依据公开人体活动数据集Opportunity创建HAR-CNN实例并进行了实验测试。实验结果表明,HAR-CNN对周期性重复活动和离散性人体活动识别的F1值分别达到了90%和92%,同时算法具有良好的运行效率。

**关键词:** 人体活动识别;特征提取;卷积神经网络;滑动窗口;RGB位图

中图分类号: TN911.7; TP391

文献标识码: A

文章编号: 1009-5896(2022)01-0168-10

DOI: 10.11999/JEIT200942

## Human Activity Recognition Technology Based on Sliding Window and Convolutional Neural Network

HE Jian<sup>①②</sup> GUO Zelong<sup>①</sup> LIU Leyuan<sup>①</sup> SU Yuhan<sup>①</sup>

<sup>①</sup>(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

<sup>②</sup>(Beijing Engineering Research Center for IOT Software and Systems, Beijing 100124, China)

**Abstract:** Due to the lack of unified human activity model and related specifications, the existing wearable human activity recognition technology uses different types, numbers and deployment locations of sensors, and affects its promotion and application. Based on the analysis of human activity skeleton characteristics and human activity mechanics, a human activity model based on Cartesian coordinates is established and the normalization method of activity sensor deployment location and activity data in the model is standardized; Secondly, a sliding window technique is introduced to establish a mapping method to convert human activity data into RGB bitmap, and a Convolutional Neural Network is designed for Human Activity Recognition (HAR-CNN); Finally, a HAR-CNN instance is created and experimentally tested based on the public human activity dataset Opportunity. The experimental results show that HAR-CNN achieves the F1 values of 90% and 92% for periodic repetitive activity and discrete human activity recognition, respectively, while the algorithm has good operational efficiency.

**Key words:** Human Activity Recognition (HAR); Feature extraction; Convolutional Neural Network (CNN); Sliding window; RGB bitmap

### 1 引言

人体活动识别(Human Activity Recognition,

HAR)<sup>[1]</sup>广泛应用于人机交互<sup>[2]</sup>、医疗诊断<sup>[3]</sup>、智能家居控制<sup>[4]</sup>、异常行为检查等领域<sup>[5]</sup>,是普适计算和人机交互的重要研究领域。HAR主要依据由传感器采集的活动特征数据来实现活动识别。依据传感器的不同,HAR技术可以分为基于环境感知的HAR、基于视觉的HAR和基于可穿戴计算的HAR。其中,基于环境感知的HAR由嵌入人们工作生活环境的传感器(如压力传感器、被动式红外传感器等)采集人体活动数据,实现人体活动的分类识别;由于其采用的传感器需要部署在具体环境中,

收稿日期: 2020-11-04; 改回日期: 2021-06-02; 网络出版: 2021-08-24

\*通信作者: 何坚 Jianhee@bjut.edu.cn

基金项目: 国家重点研发计划(2020YFB2104400), 国家自然科学基金(61602016), 北京市科技计划(D171100004017003)

Foundation Items: The National Key R&D Program of China (2020YFB2104400), The National Natural Science Foundation of China (61602016), Beijing Science and Technology Plan (D171100004017003)

因此其通常仅适用于特定场景。基于视觉的HAR由部署于人们工作生活环境中的摄像机获得人体活动的视觉数据，并应用计算机视觉技术对人体活动进行分类识别<sup>[6]</sup>。基于视觉的HAR具有不影响用户日常活动、使用方便等优点，但其存在易受光线变化和遮挡干扰，并会涉及用户隐私等缺点<sup>[7]</sup>。基于可穿戴计算的HAR，将加速度计、陀螺仪等惯性传感器集成于可穿戴手表、背心等设备中，通过采集人体活动力学数据实现人体活动的分类识别<sup>[8]</sup>。可穿戴HAR具有成本低、易于使用且不影响用户日常活动等优点，成为HAR研究热点之一。

可穿戴HAR任务可分为人体活动数据感知、特征数据提取和活动分类识别3个阶段。随着微电子机械系统技术(Micro Electro Mechanical Systems, MEMS)的发展，3轴加速度计、陀螺仪、磁力计等惯性传感元器件精度越来越高、体积越来越小，研究人员开发了多种集成了惯性传感器的可穿戴设备来采集人体活动的3轴加速度、角速度等动力学数据，依据不同日常活动(如行走、上楼、下楼等)中3轴加速度、角速度数值及变化趋势的差异，提炼不同的人工特征(例如，瞬时加速度/角速度，合加速度/角速度，加速度/角速度的方差/均方差等)，并在此基础上研究不同的人体活动分类识别算法。

早期可穿戴人体活动识别方法，以人体活动的人工特征数据为基础，分别从阈值、时域、频域方面进行特征提取和活动分类<sup>[9,10]</sup>。例如，Ravi等人<sup>[11]</sup>采用快速傅里叶变换(Fast Fourier Transform, FFT)提取人体活动的频域特征并进行活动分类识别。随着研究的深入，研究人员纷纷将传统的机器学习方法引入到可穿戴HAR中。例如，何坚等人<sup>[12]</sup>构造了可穿戴活动感知背心，引入K邻近(K-Nearest Neighbor, KNN)并结合卡尔曼滤波设计了可穿戴跌倒检测系统，跌倒检测的准确率达到98.9%。Tran等人<sup>[13]</sup>在对原始活动数据FFT变换后，在智能手机上采用支持向量机(Support Vector Machines, SVM)实现人体活动的快速准确识别。近年来随着深度学习在图像处理、语音识别等领域取得成功，研究人员纷纷将深度学习技术引入到HAR中。Hanai等人<sup>[14]</sup>将卷积神经网络(Convolutional Neural Network, CNN)引入到HAR中，实验表明由于CNN包含更多隐藏层，并通过逐层卷积和池化计算提取不同尺度的人体活动特征，因此获得了较高的人体活动识别准确率。针对人体活动的时序特征，文献<sup>[15]</sup>结合CNN和长短期记忆网络(Long Short Term Memory, LSTM)构造深度神经网络提取人

体活动的时空上下文特征，并通过实验证明该方法相比单纯CNN网络具有更高的活动识别准确率。

概括起来，由于缺少统一的协议和规范，不同可穿戴HAR系统采用活动传感器的类别、数量，以及传感器部署的位置不尽相同。例如，文献<sup>[16]</sup>将集成3轴加速度、陀螺仪的感知模块部署于可穿戴背心的胸背部；文献<sup>[17]</sup>将惯性传感器部署于人体腰部；文献<sup>[18]</sup>将惯性传感器集成于项链的吊坠上；文献<sup>[19]</sup>将惯性传感器部署于人体腰部和左右大腿部。由于可穿戴HAR传感器的类型、部署位置不尽相同，已有HAR采用的活动分类识别算法难以推广应用。此外，尽管深度学习算法取得了较高人体活动识别准确率，但其需要较高的运算和存储资源，需要GPU等专用处理器支持才能满足实时人体活动识别需要。

针对上述问题，本文借鉴人体骨架模型思想，建立通用人体活动力学模型；其次，建立3轴加速度、角速度与RGB像素之间的映射关系，通过引入滑动窗口技术构造基于人体活动力学模型的像素图像；最后，构造面向人体活动识别的卷积网络(Convolutional Neural Network is designed for Human Activity Recognition, HAR-CNN)提取人体活动的时空上下文信息，并采用公开的Opportunity数据进行算法验证测试。实验结果表明HAR-CNN具有较高人体活动识别率，并具有良好的运行效率。

## 2 人体活动力学模型

人体活动过程中，头、手、足、四肢、腰和胸部会呈现出不同的运动特征。对此，本文参考微软人体骨架模型<sup>[20]</sup>，并采用笛卡尔坐标系建立通用人体活动力学模型(如图1所示)，图1(a)–图1(c)为根据人体运动特征提取出的骨架关节点，图1(d)为关节点上传感器的加速度与角速度的方向示意图。

人体形态主要由四肢、躯干、手、足和头部通过关节铰链式连接构成。其中，在人体活动中肢体运动主要体现在胸部(如挺胸)、腰部(如弯腰)，因此胸部和腰部通常会作为人体活动力学数据的检测点。依据文献<sup>[21]</sup>的人体骨架模型，人体活动建模需要部署的动力学传感器节点集 $S = \{S_1, S_2, \dots, S_{13}, H_l, H_r, F_l, F_r\}$ 。如图1(c)所示， $S_1$ 为头部感知节点； $S_2, S_3, S_4, S_9, S_{10}, S_{11}$ 为左右臂感知节点； $H_l, H_r$ 为左右手感知节点； $S_7, S_8$ 为胸部、腰部脚感知节点； $S_5, S_6, S_{12}, S_{13}$ 为左右腿感知节点； $F_l, F_r$ 为左、右脚的感知节点。

为了规范表示人体活动力学特征，本文采用笛卡尔坐标系建立统一的人体活动力学模型。在图1(d)

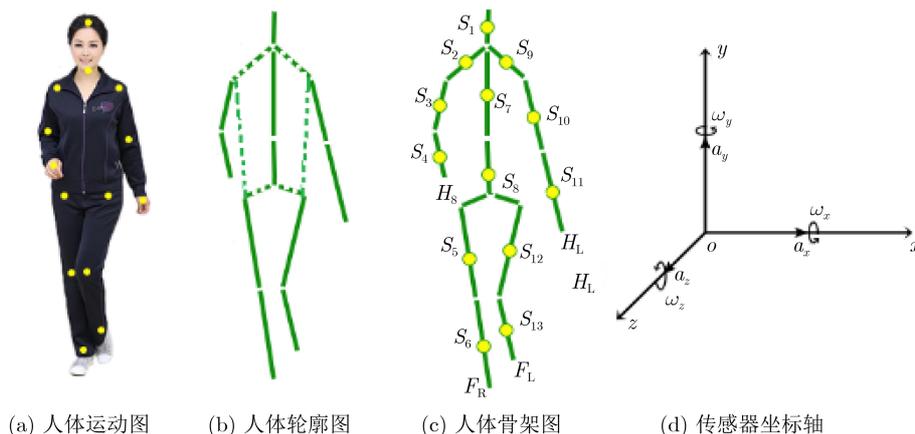


图1 基于笛卡儿坐标的人体活动动力学模型

中,  $a_x, a_y$  和  $a_z$  分别代表人体活动中3轴加速度计分别沿着坐标系中  $x$  轴、 $y$  轴和  $z$  轴采集得到的加速度分量数据;  $\omega_x, \omega_y$  和  $\omega_z$  分别代表人体活动时陀螺仪分别沿  $x$  轴、 $y$  轴和  $z$  轴感知到的人体倾转角加速度分量数据。设  $A$  表示人体活动,  $F_{\text{ext}}$  为特征提取函数,  $F_{\text{cls}}$  为人体活动分类函数。则

$$A = F_{\text{cls}}(F_{\text{ext}}(S_1, S_2, \dots, S_{13}, H_1, H_r, F_1, F_r)) \quad (1)$$

其中

$$S_i = (a_x^i, a_y^i, a_z^i, \omega_x^i, \omega_y^i, \omega_z^i) \quad (2)$$

$H_1, H_r$  和  $F_1, F_r$  构成与  $S_i$  相同。由于不同厂家传感器的量程和测量精度不尽相同, 为了避免这些因素对活动识别算法的影响, 可以采用式(3)对采样数据进行归一化处理, 使得加速度、角速度的取值在  $[0, 1]$ 。式(3)中,  $x$  代表传感器采集的原始数据,  $x_{\text{min}}$  代表传感器量程的最小值,  $\text{range}$  代表加速度计或陀螺仪的量程范围,  $x_{\text{scale}}$  代表归一化之后的数据。

$$x_{\text{scale}} = \frac{x - x_{\text{min}}}{\text{range}} \quad (3)$$

已有可穿戴活动感知技术中, 研究人员选取了图1所示的人体活动动力学模型中感知节点集的子集作为人体活动动力学数据的采样点, 设计了动力学数据的信号向量模、方差、均方差、熵等人工特征方法来提取人体活动特征, 并研究了不同分类算法以实现人体活动的分类识别。

### 3 面向HAR的卷积神经网络

深度学习可在无需事先知道检测目标所具有的人工特征情景下, 通过监督学习来提取目标特征并分类, 其在图像识别、自然语言处理等方面取得了巨大成功。对此, 本节引入滑动窗口技术构造人体活动的RGB位图, 进而构造基于卷积神经网络的深度学习算法提取人体活动特征, 并进行活动分类。

#### 3.1 人体活动动力学数据的RGB位图映射

由于人体活动具有连续性和长期性等特征, 运动动力学传感器实时采集的人体活动数据形成连续流数据。而传统针对固定静态数据的算法无法直接应用于流数据场景。对此, 本文通过滑动窗口技术缓存人体活动数据, 并为每一个感知数据设置对应时间戳。这样, 随着新数据的不断产生, 系统依据时间戳按照先进先出的原则, 不断更新窗口内的老数据。

图2所示为人体活动数据滑动窗口示例。其中, 新感知数据元素来自窗口右测, 窗口左侧元素为按照先进先出原则移出窗口的数据。滑动窗口包含  $T_s \times n$  时间段的感知数据。其中,  $T_s$  为采样周期,  $n$  为采样次数。窗口中每个元素  $e_j = (S_1^j, S_2^j, \dots, S_{13}^j, H_1^j, H_r^j, F_1^j, F_r^j)$ 。根据式(2)可知每个  $S_i^j$  可以映射为两个像素。则滑动窗口  $W$  中包含的总像素个数为  $17 \times 2 \times n$ 。

对人体活动模型中的3轴活动动力学数据与图像的RGB通道数据建立映射关系(例如:  $x$  轴分量值对应红色通道子像素值,  $y$  轴分量值对应绿色通道子像素值,  $z$  轴分量值对应蓝色通道子像素值), 则可以将人体活动的3轴加速度、角速度数据分别转化为对应的RGB像素(如图2所示)。从图2可知, 将人体活动动力学数据转换为RGB像素图, 不仅保存了同一时刻中人体活动3轴加速度、角速度之间的关联关系, 同时也保存了滑动窗口中不同时刻3轴加速度、角速度之间的依赖关系, 这为设计基于卷积神经网络的深度学习算法提供了基础。

由于加速度计和陀螺仪传感器的量程与精度不尽相同, 为此本文采用式(4)对加速度计和陀螺仪传感器感知的数据进行规范化处理, 通过四舍五入将结果转化为整型数据。其中, RGB图像中子通道通常为1 Byte, 即取值为  $0 \sim 255$  的整型数据。

$$\text{result} = 255 \times \frac{x - x_{\min}}{\text{range}} \quad (4)$$

通过式(4)将3轴加速度计和陀螺仪采集的 $x$ 、 $y$ 和 $z$ 轴数据均规范到0~255范围内。这样就可将每次3轴加速度计和陀螺仪采集到的数据转换成一个像素点。

### 3.2 面向HAR的卷积网络架构

通过对LeNet-5<sup>[22]</sup>、AlexNet<sup>[23]</sup>、VGGNet<sup>[24]</sup>等经典的卷积神经网络进行归纳总结，本文提出了面向人体活动识别的卷积网络架构，即HAR-CNN。HAR-CNN由输入层、卷积层、池化层、全连接层、softmax层构成，如图3所示。其中，第1层为输入层，输入 $A \times A \times 3$ ，其中 $A$ 为RGB位图长/宽度。

中间层为 $m \times n$ 个卷积单元。每个卷积单元 $U_{m,n}(m,n=1, 2, 3, \dots)$ 包含1或多个卷积层，0或1个池化层。每一列卷积单元串联成一个卷积通道，多个卷积通道可以并联成多通道的卷积网络。

每个卷积单元 $U_{m,n}$ 中的 $C_{m,n,i}(i=1, 2, 3, \dots)$ 为卷积层。在每个卷积单元中有 $i$ 个卷积层，每个卷积单元中的所有卷积层采用相同大小的卷积核( $K_1, S_1, C_1, O_1$ )，其中 $K_1$ 为卷积核大小， $S_1$ 为卷积核的步长， $C_1$ 为特征图输入通道数， $O_1$ 为特征图输出

通道数。每个卷积单元中的所有卷积层采用相同的Padding(方式为VALID或SAME)，不同的Padding方式将导致卷积层输出的特征图大小不同。

$S_{m,n}$ 为卷积单元中的池化层。每个卷积单元最后的卷积层后可以添加0或1个池化层。池化通常取整个池化区域的最大值或平均值。

$F_k(k=1, 2, \dots, K)$ 是全连接层。在卷积单元后添加全连接层将卷积单元输出的特征结合起来，全连接层可以为1或多层。

HAR-CNN的最后一层为softmax层。通过softmax层输出预测结果，将概率最大的类别作为预测结果输出。

## 4 HAR-CNN实现与实验分析

本文采用公开的Opportunity<sup>[25]</sup>人体活动数据集，参考其对应的实验环境和传感器部署情况(如图4所示)设计构造了HAR-CNN实例，并进行相应网络的训练和测试。

### 4.1 Opportunity数据集

为了开展人体活动识别技术研究，欧盟在第7框架项目中搭建了包含躺椅、餐桌、冰箱、橱柜等设施的日常生活仿真环境，按照图4设计了包含

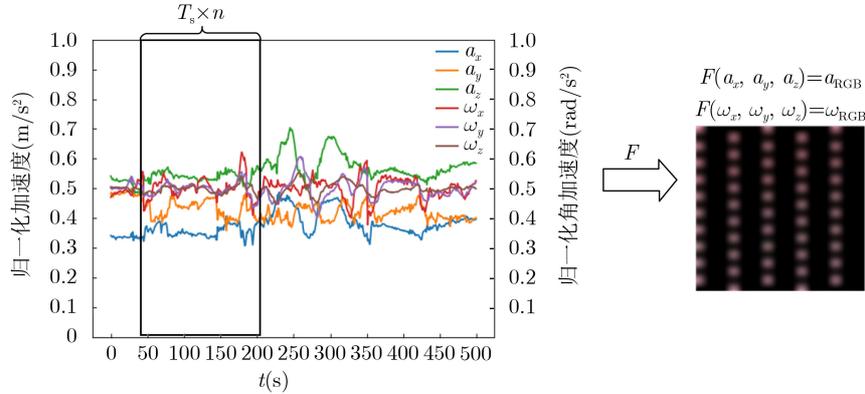


图2 人体活动力学数据转位图示意

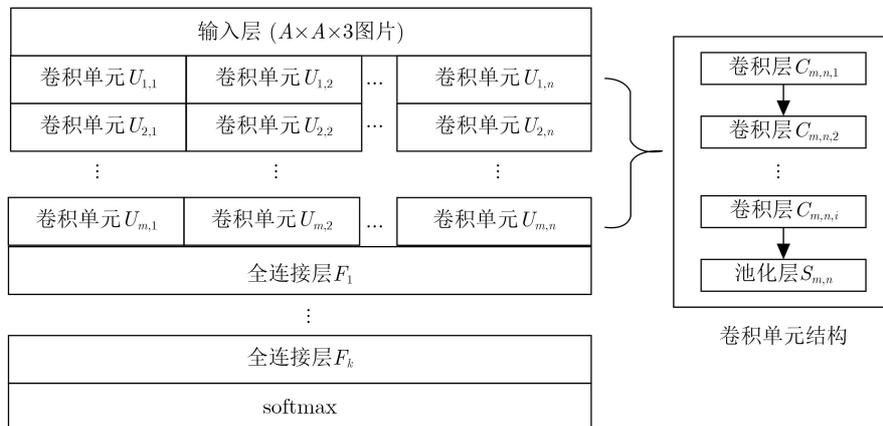


图3 HAR-CNN架构

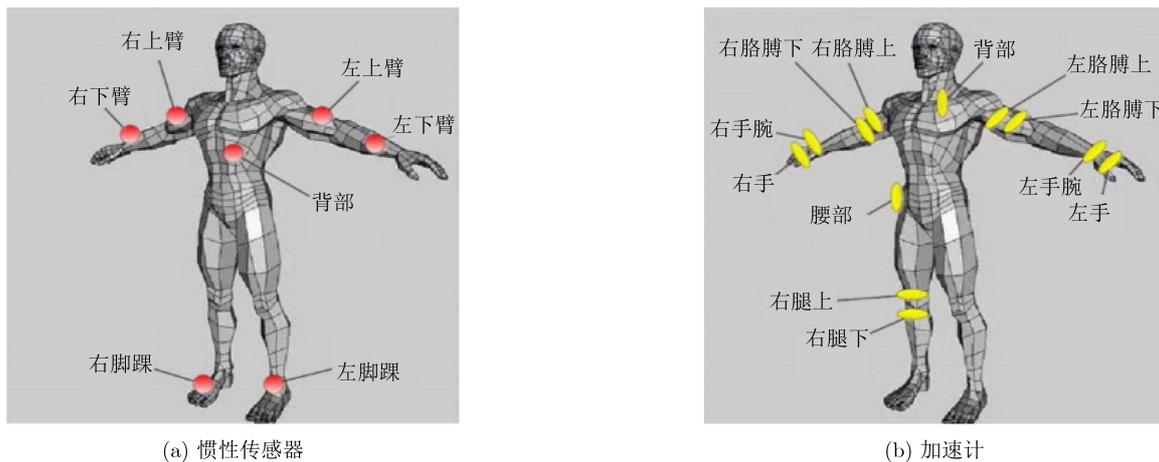


图4 Opportunity数据集传感器分布图

7个惯性传感器和12个3轴加速计的可穿戴设备以采集人体活动数据。4位被试按照设定的场景描述各自独立完成5组日常活动和1组训练活动。其中,日常活动由被试从躺椅上起身、活动准备、室外散步、准备咖啡、喝咖啡、准备三明治、吃三明治、清理餐桌、躺躺椅上休息9个顺序活动构成。在1组训练活动中被试重复20次如下9组活动:打开/关闭冰箱、打开/关闭洗碗机、打开/关闭不同高度的3个抽屉、开/关门1、开/关门2、打开/关闭灯、清理餐桌、站立喝饮料、坐着喝饮料。

从时序上可将Opportunity活动数据分为周期性重复活动(如坐、站、躺、走等)、非重复性周期活动(打开/关闭门、打开/关闭冰箱、打开/关闭洗碗机、移动杯子、清理餐桌等)。表1所示为Opportunity数据集中周期性、非周期性活动数据集描述。其中,数据的采样频率为30 Hz。由于原始Opportunity数据集存在部分数据缺失,对此本文采用线性插值进行数据补齐。

本文按照式(3)对Opportunity数据进行归一化处理,并采用3.1节的滑动窗口方法生成Opportunity数据对应的RGB位图。本文将4位被试活动数据的80%作为训练样本,其余20%作为测试样本。

#### 4.2 HAR-CNN结构

HAR-CNN网络架构如图5所示。其中,对于周期性活动,HAR-CNN的输入为26像素 $\times$ 48像素的3通道RGB位图。由于Opportunity数据集中针对非周期性活动还包含了环境传感器(如门、抽屉开关电磁阀等)采集的数据,因此对于非周期性活动的HAR-CNN输入数据扩充为57像素 $\times$ 48像素的3通道RGB位图。HAR-CNN共有6个网络层(不包含输入层)。图5中,网络的卷积层标注为C,池化层标注为S $x$ ,全连接层标注为F $x$ 。 $x$ 为网络的层号。

C1为第1层卷积层。在卷积的过程中如果卷积

表1 Opportunity数据集描述

	活动	次数
周期性	站	960
	坐	169
	走路	1711
	躺	40
非周期性	开门1	125
	关门1	122
	开门2	119
	关门2	120
	开冰箱	209
	关冰箱	213
	喝饮料	136
	清理餐桌	134
	开/关灯	129
	开洗碗机	128
	关洗碗机	124
	开抽屉1	123
关抽屉1	125	
开抽屉2	116	
关抽屉2	112	
开抽屉3	210	
关抽屉3	216	

核步长不能被输入图片尺寸整除,会发生丢弃数据的情况。为了防止输入的数据丢失,本文首先对数据的边缘进行扩充,在输入图片的四周补0,将图像扩充为28 $\times$ 50 $\times$ 3尺寸来进行卷积。卷积层1有32个卷积核,卷积核大小为2 $\times$ 2 $\times$ 3,卷积的步长为1,每个卷积核同时在图像的3个通道中卷积,输出1维的卷积图像。每个卷积核有2 $\times$ 2 $\times$ 3以及1个偏置共13个连接参数,该层共416个参数,最后输出32张26 $\times$ 48的特征图。

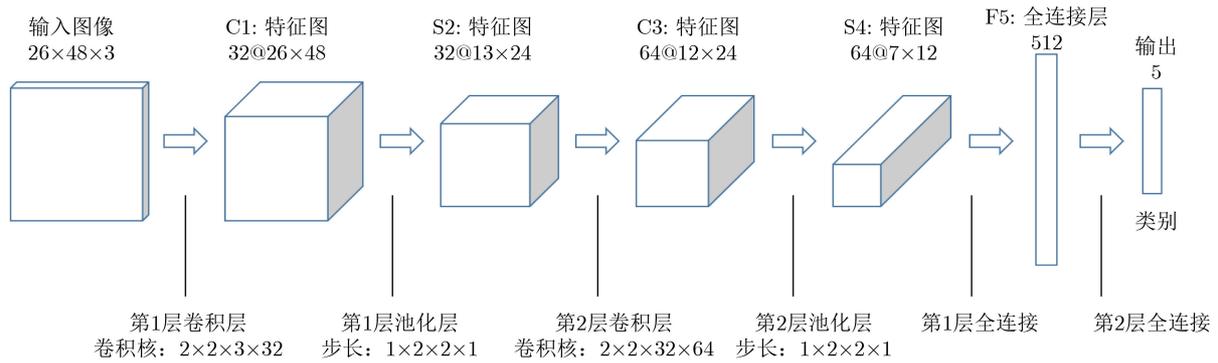


图 5 HAR-CNN网络架构

S2为第1层池化层。第1层池化层的输入为第1层卷积层后输出的 $26 \times 48 \times 32$ 的特征图，在池化操作前对输入特征图进行边缘扩充。池化的窗口为 $2 \times 2$ ，步长为2，每次仅提取窗口中的最大值作为输出，该层最后输出32张 $13 \times 24$ 的特征图。

C3为第2层卷积层。第2层卷积层的输入为第1层池化层输出的特征图，在卷积前对输入特征图进行边缘补充操作。卷积层2的卷积核数量为64，大小为 $2 \times 2 \times 3$ ，卷积步长为1。该层最后输出64张 $13 \times 24$ 的特征图。

S4为第2层池化层。第2层池化层的输入为第2层卷积层后输出的特征图。第2层池化层的窗口大小为 $2 \times 2$ ，步长为2，选择窗口中最大的数据进行保留，该层最后输出 $7 \times 12 \times 64$ 特征图。

F5为网络的全连接层。全连接层内有512个神经元，将池化层2的结果展平作为全连接层的输入，为了防止网络过拟合，添加了dropout层，随机地丢掉神经元，dropout的参数设置为0.8。

Outputs为网络最后的输出层，输出将全连接层的输出与5个周期性活动标签类进行全连接操作，通过softmax函数计算出每个活动类别的概率，其中概率最大的为预测的活动标签。

本文选择了一台搭载Ubuntu操作系统的服务器(Xeon E5-2620 v4 CPU, 128 GB内存, TESLA M40图形加速卡)，采用Tensorflow和Python设计实现HAR-CNN实例。使用Opportunity公开数据对HAR-CNN进行模型训练和测试。其中，网络训练的学习率0.0001，batch size设置为750，训练迭代1000次后识别准确率不再提高，课题组停止模型训练。此外，课题组使用了Adam方法对网络进行优化。

### 4.3 实验结果分析

本文采用同时兼顾精确率和召回率的 $F_1$ 值(如式(5)所示)作为模型的评估指标。

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

为了验证本文提出的网络模型的有效性，本文选取了MS-2DCNN<sup>[7]</sup>和DeepConvLSTM<sup>[15]</sup>以及DRNN<sup>[26]</sup> 3种有监督机器学习算法以及Weka中的Bayes Network, Random Forest, Naïve Bayes, Random Tree 4种分类算法进行活动识别性能对比。其中，MS-2DCNN将不同轴向的传感器数据分别输入神经网络，然后对卷积得到的不同轴向的特征进行拼接分类，该网络共有11层，包括输入输出层、4层卷积层、1层池化层、1层数据转化层、2层全连接层以及1层连接分类器；DeepConvLSTM采用卷积神经网络和长短时记忆网络分别提取人体活动的时空特征并分类。DeepConvLSTM除输入层外，包括4个卷积层、2个LSTM层及一个softmax输出层。每个卷积层的卷积核长度为 $5 \times 1$ ，卷积核个数为64个，每个LSTM层包括128个LSTM单元。DRNN网络为一个包含3层隐藏层的循环神经网络。

Opportunity数据集中周期性重复活动称为运动模式(Modes of Locomotion, ML)，非重复性离散活动称为姿态模式(Gesture, GR)，表2所示为两种对比算法以及Weka中的4种分类算法与本文算法的 $F_1$ 值对比。其中，Gesture(NULL)和ML(NULL)分别表示包含NULL的姿态和运动模式，NULL类为间隔时间不含任何特征活动。

通过表2可以看出，Weka上4种算法的人体活动识别 $F_1$ 值普遍低于HAR-CNN，DRNN，MS-2DCNN和DeepConvLSTM算法，这可能是其提取的人体活动特征少于MS-2DCNN，DRNN，HAR-CNN和DeepConvLSTM方法的原因。DeepConvLSTM算法在每一类活动分类上都取得了较高的 $F_1$ 值，其在周期性重复活动ML高于本文的HAR-CNN算法；但由于其采用卷积神经网络和LSTM相结合的网络结构，网络参数较多，导致其训练和识别的时间相比HAR-CNN和MS-2DCNN较长。DRNN网络每一项都取得了较高的得分但是其使用

的循环神经网络在训练时易遇到梯度爆炸与消失等问题,并且算法的运行时间较长。MS-2DCNN算法在每一项分类中的 $F_1$ 值均与本文算法相近,但是其网络层数较多导致运行时间较长。本文HAR-CNN网络在离散类活动Gesture(NULL)和Gesture以及周期类活动ML(NULL)中取得了最高 $F_1$ 值,说明其对周期性、离散性活动均有较好的识别效果,同时运行时间也小于DeepConvLSTM,MS-2DCNN和DRNN算法。

本文还对数据分割使用的滑动窗口长度和步长进行了实验分析(如图6所示),图6(a)、图6(b)为随着窗口长度增加两种活动识别的 $F_1$ 值变化,图6(c)、图6(d)为随着滑动步长变化两种活动识别的 $F_1$ 值变化。从图6(a)、图6(b)可以看出ML和Gesture活动识别结果随着窗口从40 ms增加到1500 ms时,人体活动识别 $F_1$ 值总体上逐步提高(去除局部波动)。其中,针对周期性重复活动在窗口宽度为

1600 ms时, $F_1$ 值达到峰值,然后随着窗口增大 $F_1$ 值逐渐减小;离散活动在窗口宽度为1300 ms时, $F_1$ 值达到峰值,然后随着窗口增大 $F_1$ 值逐渐减小。这说明不是滑动窗口越长,人体活动识别 $F_1$ 越好,而是滑动窗口大小有一个峰值,在该状态下人体活动识别准确率最高。这是因为当窗口过小时,活动特征无法被完全提取会导致 $F_1$ 值较低。而当窗口太大时窗口中可能包含两类以上的活动,导致网络无法学习有效特征导致 $F_1$ 值下降。

本文通过实验分析每类活动的持续时间最终选定1.6 s作为滑动窗口大小。确定滑动窗口大小后,本文继续实验分析每次滑动窗口移动步长对活动识别 $F_1$ 值的影响。窗口滑动步长通常小于窗口长度,这使得相邻两个窗口间有部分重叠活动数据。从图6(c)、图6(d)可以看出ML和Gesture活动识别结果随着窗口滑动步长从40 ms增加到800 ms时,人体活动识别 $F_1$ 值总体上逐步提高,并在步长

表2 不同算法 $F_1$ 值对比

算法	Gesture	Gesture (NULL)	ML	ML (NULL)	运行时间(s)
Bayes Network	0.79	0.81	0.82	0.74	32.9
Random Forest	0.63	0.72	0.73	0.69	21.2
Naïve Bayes	0.54	0.66	0.75	0.74	8.1
Random Tree	0.75	0.88	0.87	0.85	7.0
DeepConvLSTM	0.86	0.91	0.93	0.89	6.6
MS-2DCNN	0.81	0.89	0.92	0.85	5.8
DRNN	0.84	0.92	0.91	0.88	7.3
HAR-CNN	0.90	0.92	0.92	0.90	3.7

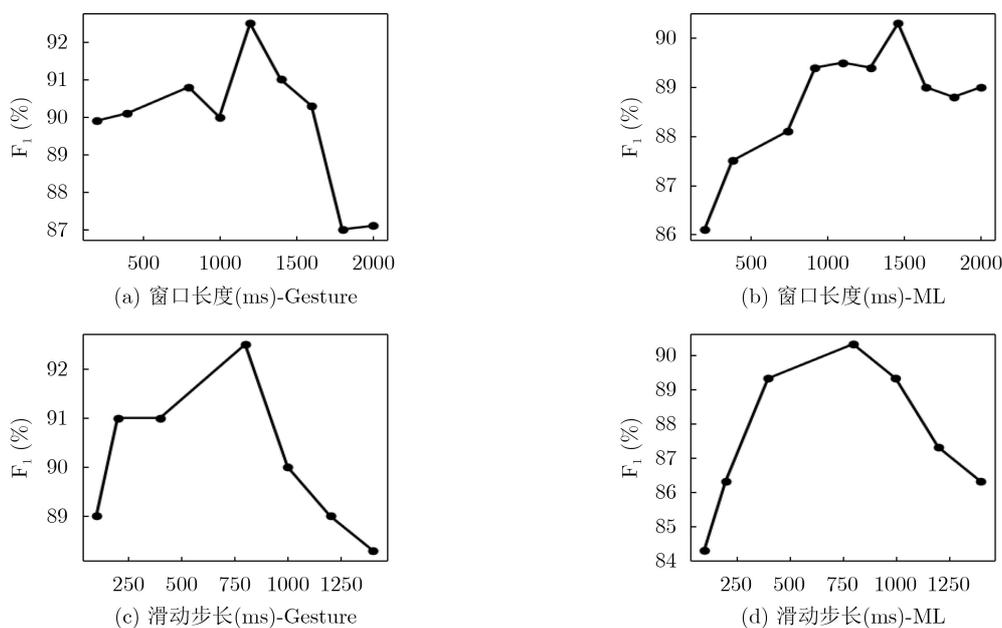


图6 不同滑动窗口长度和步长的 $F_1$ 值

为800 ms时 $F_1$ 值达到峰值，然后随着步长增大 $F_1$ 值逐渐减小。本文通过实验得到的最优滑动步长为滑动窗口大小的一半时，人体活动识别准确率最高。

如表3和表4列出了包含NULL类时本文算法和DRNN算法针对ML和Gesture两类活动的识别准确率、召回率的比较。表3、表4中本文算法的平均准确率都达到了90%优于DRNN算法，证明本文的算法有良好的识别效果。在ML标签活动识别中，两种算法的站立活动识别准确率都较高，而走路识别准确率却最低。分析Opportunity数据集发现站立数据占数据集的40%，而走路数据仅占23%。站立的训练数据较多使得站立活动的识别准确率较高。本文算法在GR标签活动识别中，因数据集中NULL

类占比高达58%，故大多数的错误识别均是将活动类别识别为NULL类。

## 5 结束语

本文建立基于人体骨架的通用人体活动力学模型，并构造了面向人体活动识别的HAR-CNN网络。通过对比另外3种深度学习算法与4种传统机器学习算法在Opportunity数据集的活动识别结果，证明了本文设计的HAR-CNN算法有较高的准确率和运算效率，是一种有效的活动识别算法。未来课题组将继续优化网络结构，包括网络的卷积核大小、连接方法等。此外，会进一步探索解决人体活动数据不平衡性(诸如NULL过多等问题)导致识别准确率降低等问题。

表 3 ML活动识别的准确率和召回率(%)

活动类别	DRNN准确率	DRNN召回率	HAR-CNN 准确率	HAR-CNN 召回率
NULL	90	91	92	92
站立	92	91	96	92
走路	79	82	81	94
坐下	92	85	93	73
躺	90	85	90	85
平均值	89	87	90	87
标准差	5	4	5	8

表 4 GR活动识别的准确率和召回率(%)

活动类别	DRNN准确率	DRNN召回率	HAR-CNN 准确率	HAR-CNN 召回率
NULL	95	97	94	95
开门1	87	92	92	87
开门2	93	93	96	93
关门1	90	92	96	80
关门2	95	92	96	91
开冰箱	85	65	90	80
关冰箱	83	86	87	74
开洗碗机	88	74	83	83
关洗碗机	80	81	85	83
开抽屉1	79	77	88	86
关抽屉1	80	77	85	75
开抽屉2	77	83	97	82
关抽屉2	79	88	83	83
开抽屉3	86	85	96	90
关抽屉3	86	84	91	85
清理餐桌	93	82	100	100
喝饮料	89	88	78	95
开/关灯	94	69	91	95
平均值	87	84	90	87
标准差	6	8	6	7

## 参考文献

- [1] SINGH R and SRIVASTAVA R. Some contemporary approaches for human activity recognition: A survey[C]. 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control, Mathura, India, 2020: 544–548.
- [2] GOTA D I, PUSCASIU A, FANCA A, *et al.* Human-Computer Interaction using hand gestures[C]. 2020 IEEE International Conference on Automation, Quality and Testing, Robotics, Cluj-Napoca, Romania, 2020: 1–5.
- [3] HU Ning, SU Shen, TANG Chang, *et al.* Wearable-sensors based activity recognition for smart human healthcare using internet of things[C]. 2020 International Wireless Communications and Mobile Computing, Limassol, Cyprus, 2020: 1909–1915. doi: [10.1109/IWCMC48107.2020.9148197](https://doi.org/10.1109/IWCMC48107.2020.9148197).
- [4] NATANI A, SHARMA A, PERUMA T, *et al.* Deep learning for multi-resident activity recognition in ambient sensing smart homes[C]. The IEEE 8th Global Conference on Consumer Electronics, Osaka, Japan, 2019: 340–341.
- [5] RODRIGUES R, BHARGAVA N, VELMURUGAN R, *et al.* Multi-timescale trajectory prediction for abnormal human activity detection[C]. 2020 IEEE Winter Conference on Applications of Computer Vision, Snowmass, USA, 2020: 2615–2623.
- [6] DEEP S and ZHENG Xi. Leveraging CNN and transfer learning for vision-based human activity recognition[C]. The 29th International Telecommunication Networks and Applications Conference, Auckland, New Zealand, 2019: 1–4.
- [7] 邓诗卓, 王波涛, 杨传贵, 等. CNN多位置穿戴式传感器人体活动识别[J]. 软件学报, 2019, 30(3): 718–737. doi: [10.13328/j.cnki.jos.005685](https://doi.org/10.13328/j.cnki.jos.005685).  
DENG Shizhuo, WANG Botao, YANG Chuangui, *et al.* Convolutional neural networks for human activity recognition using multi-location wearable sensors[J]. *Journal of Software*, 2019, 30(3): 718–737. doi: [10.13328/j.cnki.jos.005685](https://doi.org/10.13328/j.cnki.jos.005685).
- [8] WANG Yan, CANG Shuang, and YU Hongnian. A survey on wearable sensor modality centred human activity recognition in health care[J]. *Expert Systems with Applications*, 2019, 137: 167–190. doi: [10.1016/j.eswa.2019.04.057](https://doi.org/10.1016/j.eswa.2019.04.057).
- [9] BAO Ling and INTILLE S S. Activity recognition from user-annotated acceleration data[C]. The 2nd International Conference on Pervasive Computing, Vienna, Austria, 2004: 1–17.
- [10] NURMI P, FLORÉEN P, PRZYBILSKI M, *et al.* A framework for distributed activity recognition in ubiquitous systems[C]. International Conference on Artificial Intelligence, Las Vegas, USA, 2005: 650–655.
- [11] RAVI N, DANDEKAR N, MYSORE P, *et al.* Activity recognition from accelerometer data[C]. The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, Pittsburgh, USA, 2005: 1541–1546.
- [12] 何坚, 周明我, 王晓懿. 基于卡尔曼滤波与k-NN算法的可穿戴跌倒检测技术研究[J]. 电子与信息学报, 2017, 39(11): 2627–2634. doi: [10.11999/JEIT170173](https://doi.org/10.11999/JEIT170173).  
HE Jian, ZHOU Mingwo, and WANG Xiaoyi. Wearable method for fall detection based on Kalman filter and k-NN algorithm[J]. *Journal of Electronics & Information Technology*, 2017, 39(11): 2627–2634. doi: [10.11999/JEIT170173](https://doi.org/10.11999/JEIT170173).
- [13] TRAN D N and PHAN D D. Human activities recognition in android smartphone using support vector machine[C]. The 7th International Conference on Intelligent Systems, Modelling and Simulation, Bangkok, Thailand, 2016: 64–68.
- [14] HANAI Y, NISHIMURA J, and KURODA T. Haar-like filtering for human activity recognition using 3D accelerometer[C]. The 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop. Marco Island, USA, 2009: 675–678.
- [15] ORDÓÑEZ F J and ROGGEN D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition[J]. *Sensors*, 2016, 16(1): 115. doi: [10.3390/s16010115](https://doi.org/10.3390/s16010115).
- [16] 何坚, 张子浩, 王伟东. 基于ZigBee和CNN的低功耗跌倒检测技术[J]. 天津大学学报: 自然科学与工程技术版, 2019, 52(10): 1045–1054. doi: [10.11784/tdxbz201808059](https://doi.org/10.11784/tdxbz201808059).  
HE Jian, ZHANG Zihao, and WANG Weidong. Low-power fall detection technology based on ZigBee and CNN algorithm[J]. *Journal of Tianjin University: Science and Technology*, 2019, 52(10): 1045–1054. doi: [10.11784/tdxbz201808059](https://doi.org/10.11784/tdxbz201808059).
- [17] GJORESKI H, LUSTREK M, and GAMS M. Accelerometer placement for posture recognition and fall detection[C]. 2011 Seventh International Conference on Intelligent Environments, Nottingham, UK, 2011: 47–54.
- [18] WANG Changhong, LU Wei, NARAYANAN M R, *et al.* Low-power fall detector using triaxial accelerometry and barometric pressure sensing[J]. *IEEE Transactions on Industrial Informatics*, 2016, 12(6): 2302–2311. doi: [10.1109/TII.2016.2587761](https://doi.org/10.1109/TII.2016.2587761).
- [19] GJORESKI H, KOZINA S, GAMS M, *et al.* RAReFall - Real-time activity recognition and fall detection system[C]. 2014 IEEE International Conference on Pervasive Computing and Communication Workshops, Budapest, Hungary, 2014: 145–147.

- [20] SHOTTON J, FITZGIBBON A, COOK M, *et al.* Real-time human pose recognition in parts from single depth images[C]. The CVPR 2011, Colorado, USA, 2011: 1297–1304. doi: [10.1109/CVPR.2011.5995316](https://doi.org/10.1109/CVPR.2011.5995316).
- [21] KONIUSZ P, CHERIAN A, and PORIKLI F. Tensor representations via kernel linearization for action recognition from 3D skeletons[C]. The 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2106: 37–53.
- [22] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324. doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
- [23] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90. doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [24] SIMONYAN K and ZISSERMAN A. Very Deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2015.
- [25] Opportunity dataset[EB/OL]. <https://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition>, 2015.
- [26] MURAD A and PYUN J Y. Deep recurrent neural networks for human activity recognition[J]. *Sensors*, 2017, 17(11): 2556. doi: [10.3390/s17112556](https://doi.org/10.3390/s17112556).
- 何 坚：男，1969年生，副教授，主要研究方向为智能人机交互、普适计算和物联网。
- 郭泽龙：男，1996年生，硕士生，研究方向为智能人机交互和模式识别。
- 刘乐园：男，1990年生，博士生，主要研究方向为物联网、机器学习、软件工程。
- 苏予涵：男，1997年生，硕士生，研究方向为智能人机交互和模式识别。

责任编辑：马秀强