

DOI:10.19651/j.cnki.emt.2312680

基于多尺度通道注意力机制的行为识别方法^{*}

许晨阳¹ 范非易¹ 柯冠舟² 沈建飞^{1,3}(1. 中国科学院计算技术研究所 北京 100190; 2. 北京交通大学 数据科学与智能决策研究院 北京 100091;
3. 山东省智能计算技术研究院 济南 250102)

摘要: 针对可穿戴行为识别任务中小尺度的感受野难以提取长序列关联,大尺度感受野会导致特征压缩降低网络对信号特征的分辨率的问题。提出了一种基于多尺度通道注意力机制的行为识别方法。首先,从多个感受野提取时间特征和传感器通道特征,在保证信号具有低语义特征的同时提取信号的高语义特征;其次,在多尺度特征图之间建立跨通道关联,保证低语义特征和高语义特征之间的交互。多尺度通道注意力机制能够充分融合多尺度特征和多个特征图的关联信息,增强对微弱信号和剧烈信号的识别能力。在 UCIHAR、DSADS、PAMAP2 和 UniMib-SHAR 数据集上进行了对比实验,结果表明 MSCA-HAR 方法相比目前的主流方法在 4 个数据集上的分类准确率分别提升 0.43%,0.75%,2.90%和 0.83%。

关键词: 可穿戴设备;行为识别(HAR);多尺度通道注意力;深度学习

中图分类号: TP2 **文献标识码:** A **国家标准学科分类代码:** 530.2420

Human activity recognition method based on multi-scale channel attention mechanism

Xu Chenyang¹ Fan Feiyi¹ Ke Guanzhou² Shen Jianfei^{1,3}

(1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China; 2. Beijing Jiaotong University, Institute of Data Science and Intelligent Decision Support, Beijing 100091, China; 3. Shandong Academy of Intelligent Computing Technology, Jinan 250102, China)

Abstract: To address the problem that small-scale receptive fields for wearable activity recognition tasks make it difficult to extract long range associations and that large-scale receptive fields lead to feature compression reducing the network's resolution for signal features. In this paper, we propose a multi-scale channel attention mechanism based human activity recognition method. Firstly, temporal features and sensor channel features are extracted from multiple receptive fields, so that high semantic features and low semantic features are extracted at the same time to ensure high resolution of features. Secondly, cross channel association is established between multi-scale feature maps to obtain the interaction between low semantic features and high semantic features. multi-scale channel attention mechanism can fully integrate multi-scale features and correlation information of multiple feature maps, enhancing the recognition ability of weak signals and violent signals. The comparative experiments on the UCIHAR, DSADS, PAMAP2 and UniMib-SHAR datasets show that the classification accuracy of our method is improved by 0.43%, 0.75%, 2.90% and 0.83% respectively compared with the state of the art methods.

Keywords: wearable device; human activity recognition; multi-scale channel attention; deep learning

0 引 言

近年来,移动设备的广泛普及,使得基于可穿戴传感器

的人体行为识别(human activity recognition, HAR)^[1]成为人工智能和模式识别领域的一个重要的研究点。基于可穿戴设备的行为识别在环境监测、疾病预警、健身追踪、家

收稿日期:2023-01-28

^{*} 基金项目:国家重点研发计划(2020YFC2007104)、中国科学院战略性先导科技专项(A类)、黑土地保护与利用科技创新工程专项(XDA28040500)、北京市科技计划项目(Z221100002722009)、国家自然科学基金青年科学基金(62202455)、济南“高校20条”引进创新团队项目(2020GXRC030)资助

庭护理和异常监测领域有极大的应用前景。已有的 HAR 方法主要分为两类,一类是基于经典机器学习的 HAR 方法,另一类是基于深度学习的 HAR 方法。由于传统的机器学习算法严重依赖手工提取的特征,而这些特征通常受到个人领域知识的限制。因此本文主要研究基于深度学习的 HAR 方法。

伴随着深度学习方法的普及,基于卷积神经网络(convolutional neural networks, CNN)被研究人员应用到 HAR 领域。如 Zeng 等^[2]将 CNN 网络用于基于传感器数据的 HAR 领域,使用卷积网络提取局部的时空特征。在之后的研究中, Teng 等^[3]将局部损失方法嵌入 CNN 网络中,使网络在上游进行梯度的反向传播,降低上游网络的损失。另一种解决行为识别问题的网络是循环神经网络(recurrent neural network, RNN),有学者将 RNN 网络在时间维度和传感器通道维度分别提取长序列时间依赖和跨传感器通道关联^[4]。近年来,注意力机制常被嵌入深度学习网络,使网络对重要的特征加权。例如, Gao 等^[5]提出基于卷积注意模块(convolutional block attention module, CBAM)注意力机制的 HAR 方法,通过在特征通道维度和传感器时空维度提取长距离的依赖特征,进而提升网络对人体行为识别的准确率。

在上述方法中,基于 CNN 的方法由于感受野受限,会丢失时间关联性和传感器通道关联性。基于 RNN 方法对行为数据的不相关信号(噪声)敏感,会对信号噪声进行重复性编码,进而造成分类精度下降^[6]。基于 CBAM 注意力机制的 HAR 方法相比之前的基于 CNN 的方法有显著的性能改进,但在 CBAM 注意力机制中,空间注意力模块使用大尺度感受野会降低传感器特征数据的分辨率,对传感器通道数较少的数据集会造成细粒度特征损失,不利于特征的前向传播。

在行为识别任务中,人体的动作具有较强的多样性(如摔倒等突发性动作和走路等周期性动作),在 CNN 网络中传统的单尺度感受野会使得网络丢失多传感器之间的关联特征,体现在小尺度的感受野仅可以提取局部时空关联特征,丢失长序列关联;大尺寸的感受野可以提取全局时间关联特征与传感器通道特征,但是会造成特征压缩,降低网络对特征的分辨率。因此,已有的固定尺度感受野的 CNN-HAR 方法无法适应传感器数据的多样性。受图像领域多尺度卷积的启发^[7],本文提出基于多尺度通道注意力机制(multi-scale channel attention, MSCA)的 HAR 方法, MSCA 机制通过从不同尺度的感受野提取不同分辨率的特征。本文提出的 MSCA 机制的主要实现步骤和本文的贡献总结如下:

1) 本文提出一种 MSCA 机制,从不同尺度感受野提取动作信号特征,在保留低语义信号特征的同时,并行地提取包含长序列依赖和跨通道关联的高语义特征,并在多个尺度的特征图之间建立跨特征通道关联。

2) 针对多个尺度的特征图,使用压缩激励特征通道提取模块对每个尺度的特征图建立跨通道关联;其次对不同尺度特征图的权重进行全局校准,从而增强网络对不同特征图的注意力,增强多个特征图的关联。

3) 本文通过在 4 个数据集上进行实验并对比了多个主流(state-of-the-art, SOTA)方法,实验结果表明所提出的基于 MSCA 机制的 HAR 方法在分类准确率上优于 SOTA 方法 0.43% (UCIHAR), 0.75% (DSADS), 2.90% (PAMAP2) 和 0.83% (UniMib-SHAR)。

本文将按以下方式展开。在第 1 章总结 HAR 领域的相关工作;在第 2 章介绍 MSCA 方法的细节;在第 3 章详细介绍所使用的 HAR 数据集和实验设计,从各个层面比较了实验结果并进行了讨论;在第 4 章进行了总结。

1 相关工作

1.1 基于机器学习的人体行为识别方法

传统的机器学习方法依赖于均值、方差、最大值等手工提取的特征。这些提取的特征被输入到支持向量机、随机森林、隐马尔可夫模型等经典机器学习方法中,进行行为识别。如 Awais 等^[8]使用支持向量机作为分类器来构建基于惯性传感器的老年人身体活动分类系统,精确的分类出日常生活的 4 种主要活动,包括坐、站、走和躺。在文献^[9]中, Ignatov 等首先使用主成分分析方法构建了行为动作的相位轨迹矩阵特征,其次使用 k 近邻方法对相位轨迹矩阵特征进行行为识别。随着极限学习机的发展,单隐层前馈神经网络具有运行速度快、泛化程度高等优点。它已普遍应用于 HAR 领域。极限学习机的问题在于不能自适应的调节算法权重,进而影响识别性能。同时算法随机生成的输入权重和隐藏层偏差使网络不稳定。Wu 等^[10]考虑到权重核的选择对极限学习机的性能有很大影响,提出了一种混合核的极限学习机方法,改善了极限学习机单一权重核的缺陷。在上述方法中,都需要进行手工制作数据特征,这使得实验人员必须具备该领域的先验知识,而且,模型的参数不能随着数据的变化而变化。

1.2 基于深度学习的人体行为识别方法

基于深度学习的 HAR 方法可以自动地从原始传感器信号中提取和学习特征,并有效地对行为进行识别。目前基于 HAR 的深度学习方法包括: CNN、RNN、基于注意力机制的 CNN 网络以及结合 CNN 和 RNN 的方法等。

在基于 CNN 的 HAR 方法中, Liu 等^[11]提出一种基于深度 CNN 的多变量时间序列数据架构,引入一种输入张量变换的方案,并使用卷积运算来捕获变量之间的局部关联。在文献^[2]中,提出一种基于 CNN 的 IMU 数据分类方法,其中 CNN 用于提取局部特征,并通过滑动卷积核达到提取全局特征信息的目的。在上述方法中, CNN 只局限于提取局部时空相关信息,并不能建立长序列关联。

在基于 RNN 的 HAR 方法中, Ullah 等^[12]提出一种基

于 RNN 的行为识别方法,首先对从陀螺仪和加速度计获得的数据进行归一化,然后将归一化的数据传递到堆叠的 RNN 网络,其输出结果经过 Softmax 层进行分类。在上述方法中,尽管 RNN 可以提取长序列信号,但是由于传感器存在不相关信号,会对 RNN 产生影响,体现在 RNN 网络对信号中噪声信号敏感,造成分类不准确^[6]。

在基于 CNN-RNN 结合的 HAR 方法中,Qian 等^[13]使用 CNN 提取局部时空关联性,并使用两个 RNN 提取长距离关联性和跨通道信息。在文献^[14]中,Dua 等提出多输入 CNN-RNN 架构,使用 3 个尺度的 CNN 从多个感受野提取局部相关性,RNN 提取长序列依赖关系。另一位学者提出基于注意力机制的人体行为识别方法,其将注意力机制与 CNN 和 RNN 相结合,在网络中增强对重要信号的权重,同时建立信号在传感器通道维度和时间维度的依赖关系^[12]。上述方法虽然可以对建立局部关联性和长序列依赖关系,但是忽略了特征通道维度关联性。

在基于注意力机制的行为识别方法中,Gao 等^[15]提出一种选择内核机制和自适应的选择感受野方法,增强行为识别任务的分类准确率。Tang 等人^[16]提出了 Triple Cross-Domain 注意力机制,通过在时间、特征通道、传感器通道维度建立两两的关联,在维度间建立了跨维度交互,实现更精准的 HAR 分类。Gao 等^[5]提出基于 CBAM 注意力机制的 HAR 方法,该方法的空间注意力模块仅使用大尺度的感受野提取操作长距离跨传感器通道关联的能力,会造成特征压缩,降低信号特征的分辨率,不利于特征的前向传播。

2 基于 MSCA 机制的 HAR 方法设计

本文使用卷积操作处理原始信号图像进而增加信号特征维度,然后使用 MSCA 机制获得多尺度特征并在多个特征图之间建立关联,进而增强网络对信号特征的识别能力。

2.1 数据预处理

为了使用可穿戴设备进行人体行为识别,一般将惯性测量单元(inertial measurement unit, IMU)放置在人体的不同部位,以同步地采集对应位置的加速度、陀螺仪和磁力计信号。这里的 IMU 单元可以是独立的传感器,也可以是搭载在智能手表或者智能手机上的传感器。在数据处理过程中,数据需要填充空缺值,防止模型在矩阵运算中出现错误;其次需要对陀螺仪、加速度计、地磁计的数据重采样进而达到统一的频率,便于后期的同维度拼接;最后,将多个通道的传感器数据组合。

数据集只需要根据需识别的行为动作时长,使用滑动窗口操作将多通道的信号分割为多个信号图像作为网络的输入,无需进行特征提取操作。由于本文需要使用 MSCA 机制提取多尺度特征,因此窗口尺寸需要尽可能覆盖更多行为时长以获得更全面的信息。滑动窗口操作如图 1 所示。

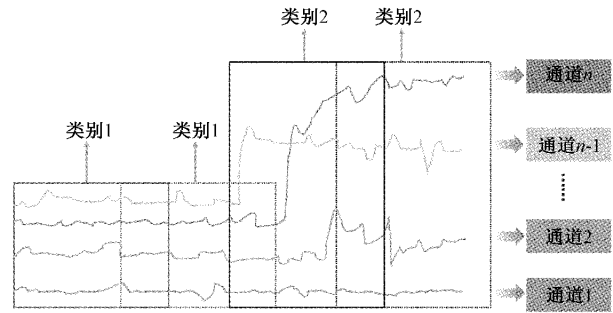


图 1 滑动窗口操作

本文将单帧信号图像定义为 $S_i = \{s_1, s_2, s_3, \dots, s_n\}$, $S_i \in \mathbf{R}^{m \times n}$ 。其中, S_i 代表输入网络的信号图像(信号图像由多个传感器通道数据经过一个滑动窗口组成), i 代表第 i 帧信号图像, m 代表时间维度, n 代表传感器通道维度。针对给定时间段的信号图像,本文将训练行为识别模型,识别该时间段内的活动(如坐、站和走等)。

2.2 多尺度通道注意力(MSCA)机制的设计

为解决感受野过小或过大导致关联性丢失和特征压缩问题,本文提出基于 MSCA 机制的 HAR 方法。MSCA 机制主要由以下几个步骤组成:1)通过多尺度卷积块从不同的感受野提取的多尺度特征(在 1 部分详细阐述);2)利用 Squeeze and Excitation (SE)^[17]通道特征提取模块,沿着多尺度的特征图的特征通道维度提取跨通道关联信息,得到特征通道维度的注意力向量(将在 2 部分详细阐述);3)使用 Softmax 函数重新校准通道维度注意力向量,获得特征通道维度的校准权重,并对重新校准的权重和对应的多尺度特征图进行逐个元素乘积运算;4)将校准的多尺度特征图通过 1×1 的卷积进行特征融合。5)为了防止梯度消失,本文将输入特征与多尺度特征经过残差连接(3)~5)步骤将在 3 部分详细阐述)。

1)多尺度特征提取

令 $X \in \mathbf{R}^{C \times H \times W}$ 作为输入网络的特征图,其中 C 代表特征通道数、 H 代表高度(时间维度)、 W 代表宽度(传感器通道维度)。如图 2 所示,本文对 C 通道的信号特征使用 1×1 的卷积操作和 ReLu 函数进行非线性组合,保证送入多尺度卷积块的特征图具有更强的表达能力。其次,将所有 C 个通道信号特征图送入多尺度卷积块,从不同的感受野提取时间关联特征与传感器通道关联特征。以下为多尺度特征图的生成函数:

$$F_i = Conv(k_i \times k_i)(X), i = 0, 1, 2, 3 \quad (1)$$

其中, k_i 代表第 i 个尺度的卷积核尺寸,例如 k_1 代表第 1 个尺度的感受野。

2)多尺度时空特征通道关联

上述方法仅能得到时间关联与传感器通道关联的特征图,由于不同尺度的特征图有多个通道,所以本文引入了 Squeeze and Excitation (SE)^[17]特征通道提取模块,在某一

尺度的特征图的多个特征通道间建立关联性,使网络有选择性地权衡每个通道的重要性。

SE 特征通道提取模块由压缩和激励两部分组成,用于提取特征通道全局信息。在压缩部分,使用全局平均池化操作生成全局特征通道信息,操作公式如下:

$$gp = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_i \quad (2)$$

其中, gp 代表全局池化操作, F_i 代表第 i 个尺度的特

征图。激励部分需要基于以上的全局特征获得注意力权重,其计算公示如下:

$$Z_i = \sigma(T_1(\delta(T_0(gp(F_i))))), i = 0, 1, 2, 3 \quad (3)$$

其中, δ 代表 ReLU 激活函数, T_0 和 T_1 代表全连接层, σ 代表 Sigmoid 激活函数, gp 代表沿特征通道维度对所有信号特征图池化操作, Z_i 代表某一尺度的通道注意力权重向量。在激励部分,使用全连接层在特征通道维度组合内部的线性关系,使高维和低维通道信息的关联。

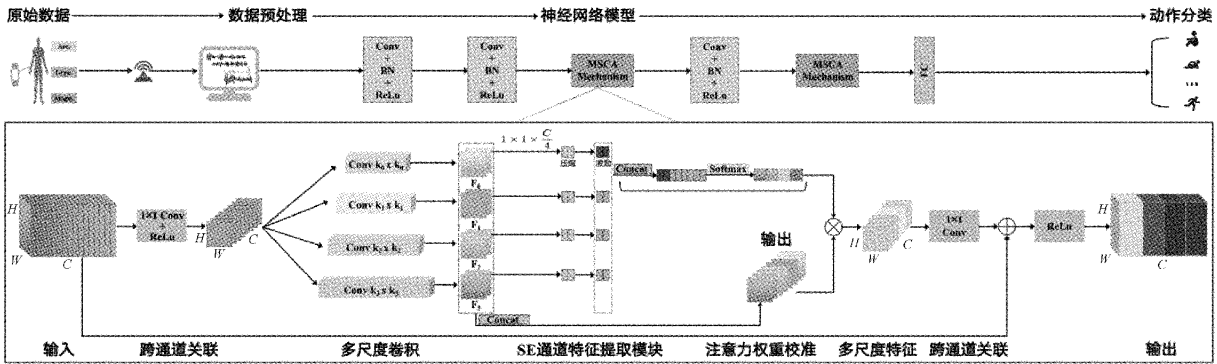


图2 基于 MSCA 机制的人体行为识别方法框架

3)多尺度通道注意力生成

本文使用 Softmax 函数在特征通道维度建立多个尺度的特征图之间的关联,最终每个尺度特征图的通道权重如下:

$$att_i = Softmax(Z_i) = \frac{\exp(Z_i)}{\sum_{i=0}^3 \exp(Z_i)}, i = 0, 1, 2, 3 \quad (4)$$

通过将不同尺度通道注意力的权重 att_i 与对应尺度的特征图 F_i 相乘得到加权后的特征图,如下式所示:

$$Y_i = F_i \odot att_i, i = 0, 1, 2, 3 \quad (5)$$

其中, \odot 代表对应元素相乘操作, Y_i 代表经过通道权重校准的多尺度特征图。

为了增强多尺度特征图的信息表达能力,本文将经过校准的多尺度特征图送入 1×1 的卷积操作和 ReLU 函数,非线性地在特征通道维度组合多尺度特征图。最后,针对网络层数加深造成的梯度消失问题, MSCA 机制将输入特征和输出特征进行残差连接。

通过以上的分析, MSCA 机制可以利用多尺度的时空特征,并将多尺度时空关联信息和跨通道信息整合到一起,为每个尺度时空特征图建立全局通道关联。

2.3 基线网络设计

因为 IMU 数据的数据量较小,不适合使用深度神经网络进行行为识别,过于深层的网络反而会造成过拟合,因此,本文使用浅层网络作为基线网络^[15]。本文使用的两个基线网络分别是 3 层 CNN 和 3 层 ResNet。在 CNN 网络中,将 MSCA 机制添加到第 2 层和第 3 层网络之后。在 ResNet 网络中,将 MSCA 机制添加到第 1 层和第 2 层网

络之后,如图 3 所示。

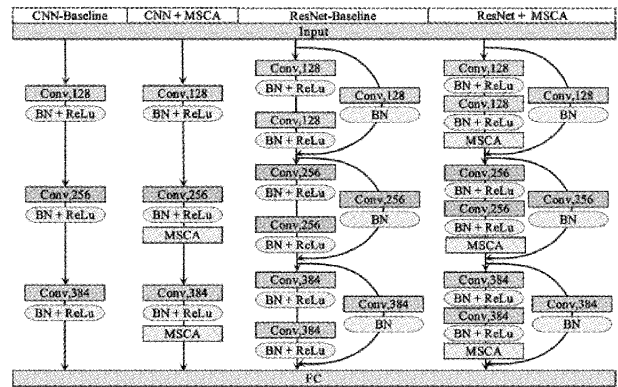


图3 网络结构图

3 实验

3.1 数据集

本文在 4 个公有数据集 (UCIHAR、DSADS、PAMAP2、UniMib-SHAR) 上进行实验,数据集划分细节如表 1 所示,数据集活动类别如表 2 所示。

表1 数据集划分细节

数据集 操作	UCIHAR	DSADS	PAMAP2	UniMib- SHAR
类别	6	9	12	17
训练集比例	70%	80%	70%	70%
测试集比例	30%	20%	30%	30%
滑窗尺寸	128	无	512	151
滑窗步长	64	无	256	75

表 2 数据集活动类别

数据集	活动类别
UCIHAR	走路, 上楼, 下楼, 坐, 站, 躺
DSADS	下楼, 静止站在电梯, 在电梯里四处走动, 在停车场散步, 以 4 公里/小时的速度在跑步机上行走 (平坡), 以 4 公里/小时的速度在跑步机上行走 (15 度爬坡), 在台阶上锻炼, 在交叉训练器上锻炼, 打篮球
UniMib-SHAR	从坐到站, 从躺到站, 走路, 跑步, 上楼, 跳, 下楼, 从站到躺, 坐, 向前摔倒并用手抑制, 向右倒, 向后倒, 撞击障碍物, 有防护策略的摔倒, 坐在椅子上向后摔倒, 晕厥, 向左倒
PAMAP2	躺下, 坐, 站, 走, 跑, 骑自行车, 越野行走, 上楼, 下楼, 用吸尘器清扫, 熨衣服, 跳绳

UCIHAR 数据集^[18]: 该数据集包含 30 名志愿者的数据。志愿者的年龄分布在 19~30 岁。该数据由三星 GalaxyS2 手机内置的陀螺仪和加速度计作为传感器采集。数据采集频率设定为 50 Hz, 经过预处理后的训练样本集为 7 352 个信号图像, 测试样本集为 2 947 个信号图像。该数据集包括 6 种类型的活动: 上楼、走路、坐、下楼、站和躺。

DSADS 数据集^[19]: 该数据集包含了 8 名志愿者 (4 位女士, 4 位男士) 活动数据, 共计 19 种活动, 每个活动采集 5 min。工作人员将传感器分别放置在志愿者的胸口、右手臂、左手臂、右腿和左腿。该数据集以 25 Hz 采样频率采集数据, 经过预处理后的训练样本集为 3 024 个信号图像, 测试样本集为 1 296 个信号图像。本文使用了该数据集中常用的 9 种活动。

PAMAP2 数据集^[20]: 该数据集采集了 9 名志愿者的数据。志愿者被要求进行 18 项体育活动, 包括 12 项特定活动 (步行、骑自行车、跳绳等) 和一些可选活动 (看电视、踢足球、汽车驾驶等)。每名志愿者佩戴 3 个 IMU 传感器, 分别位于胸部、手部和脚踝。数据以 100 Hz 的采样率采集, 经过预处理后的训练样本集为 1 769 个信号图像, 测试样本集为 759 个信号图像。

UniMib-SHAR 数据集^[21]: 该数据集由 11 771 个样本组成, 这些样本来自 30 名年龄在 18~60 岁之间的志愿者。该数据使用三星 NexusI9250 手机内置的 BMA220 加速度计采集。每名志愿者进行 8 种日常生活活动 (如行走、站立和坐着) 以及 9 种跌倒 (如向左跌倒、向右跌倒、向后跌倒和晕厥)。该数据集以 50 Hz 的频率采集, 经过预处理后的训练样本集为 7 734 个信号图像, 测试样本集为 4 037 个信号图像。

3.2 实验设置

1) 实验平台

本文所有实验均在一台 Linux 系统的服务器完成。服务器的 GPU 配置为四张 NVIDIA RTX A5000, 每张 GPU 显存 24 GB, CPU 配置为 Intel Xeon Sliver 4214R 2.4 GHz, 内存为 32 GB。本文所有实验基于 PyTorch 库实现。

2) 超参数设置

在前文的图 3 中描述了本文使用的网络结构。本文使用的 Batch Size 在 UCIHAR、DSADS、PAMAP2 和 UniMib-SHAR 4 个数据集上分别为 64、64、8、64。在所有

数据集上初始学习率为 0.001。

3) 对比实验

本文通过与多个 SOTA 方法的对比实验对基于 MSCA 机制的行为识别模型性能进行了评估, 结果如表 3 所示。通过复现 5 位学者的方法与引用其他文献的公开结果, 表明本文提出方法的有效性。本文复现方法描述如下:

选择卷积方法^[22] (2021): 基于选择卷积的网络结构, 其使用 ECDM 矩阵衡量每个通道的贡献性, 并用高贡献通道替换低贡献通道。

浅卷积网络^[23] (2021): 基于图卷积的网络结构。其使用图卷积提取跨通道信息, 并将图卷积模块嵌入 3 层卷积神经网络中, 进而实现对 HAR 的分类任务。

分布嵌入神经网络^[13] (2019): 该方法是一种基于 RNN 与 CNN 结合的方法。该方法首先将数据映射到希尔伯特空间提取高维度特征, 然后使用两个 RNN 网络分别提取时空信息, 并结合 CNN 提取的局部信息。

局部损失方法^[3] (2020): 基于局部损失方法与 CNN 结合的方法。该方法首先使用相似度函数在第 1 层卷积之后衡量损失, 然后在网络下游执行分类损失。

基于 CBAM 注意力机制的 HAR 方法^[5] (2021): 基于 CBAM^[24] 注意力机制的 CNN 网络。该方法分别在时间和空间维度提取相关性信息。

本文将基于 MSCA 的 HAR 方法与上述方法对比, 如表 3 所示, 在 UCIHAR、DSADS、PAMAP2 和 UniMib-SHAR 四个数据集上相比 SOTA 方法在准确率上分别提升 0.43%、0.75%、2.90% 和 0.83%。实验结果证明, 基于 MSCA 机制的 HAR 方法优于目前许多方法。

4) 消融实验

本文通过消融实验对 MSCA 机制的有效性进行了评估。如表 3 所示, 本文提出的基于 CNN 网络结构的 MSCA 模型在 UCIHAR、DSADS、PAMAP2 和 UniMib-SHAR 相比 CNN 基线网络分类准确率分别提升 0.77%、1.95%、2.68%、3.85%。基于 ResNet 网络结构的 MSCA 模型在上述 4 个数据集相比 ResNet 基线网络分类准确率分别提升 1.03%、0.35%、2.67% 和 1.96%。可见本文提出的基于 MSCA 的 HAR 方法能够有效提高 HAR 分类的准确性。

表3 基于MSCA机制的行为识别模型在不同数据集的平均准确率

数据集 方法	UCIHAR	DSADS	PAMAP2	UniMib-SHAR
ResNet+MSCA	97.20	94.68	97.49	79.39
ResNet	96.46	94.33	94.51	77.43
CNN+MSCA	96.37	95.57	97.13	78.07
CNN	95.60	93.42	94.45	74.22
文献[22]	96.40	94.44	94.16	76.16
文献[23]	96.48	94.52	94.86	75.42
文献[13]	84.13	82.25	84.55	73.21
文献[3]	95.23	94.29	94.59	76.16
文献[5]	96.65	94.82	93.79	77.29
文献[2]*	96.37	—	94.27	—
文献[16]*	96.77	—	93.20	78.56
文献[25]*	—	94.28	—	—
文献[26]*	—	93.42	—	—

注:*代表从文献中获取的结果

5) 参数量与时间对比实验

我们对 DSADS 数据集的测试部分进行了执行时间比较和参数量比较,其中包括 1 296 个测试数据,结果如表 4 所示,虽然所提出的方法不是最快的,但它没有明显的缺点。本文方法在时间执行上并不优越的主要原因是我们的 MSCA 机制,它增加了多个并行的卷积操作,以及特征通道上的跨通道关联操作,导致在可接受的执行时间内,能达到更好的性能。在局部损失方法中,使用余弦相似度来衡量网络上游的局部损失,虽然执行速度快,但是分类准确率并不是很高。在模型参数量上,由于使用多个并行卷积操作,因此导致了参数量增加,但是相比分布嵌入神经网络,本文提出方法的参数量并不是最大的,而且本文提出的方法在分类准确率上相比其他方法有明显的优势。

表4 模型的时间消耗和参数量对比(DSADS数据集)

方法	CNN +MSCA	文献 [22]	文献 [23]	文献 [13]	文献 [3]	文献 [5]
参数(M)	5.05	1.22	3.68	7.11	0.42	3.69
时间(S)	1.33	0.06	2.88	0.44	0.11	0.14

6) MSCA 机制位置对模型效果影响

以 PAMAP2 数据集为例,本文通过实验对比 MSCA 机制在网络的不同位置对模型的影响,实验结果如图 4 所示。

在基于 CNN 的网络中,在第 2 层第 3 层嵌入 MSCA 机制得到的效果最优,这是由于上游网络提取了信号的低语义信息, MSCA 机制在低语义特征的基础上并行地提取高语义特征,在保证具备高分辨率特征的同时,提取具备

长序列时间依赖和跨传感器通道关联的高语义特征。在 ResNet 网络中,由于每层包括两个卷积操作,因此将 MSCA 机制嵌入第 1 层、第 2 层和第 3 层达到了最优的分类效果。

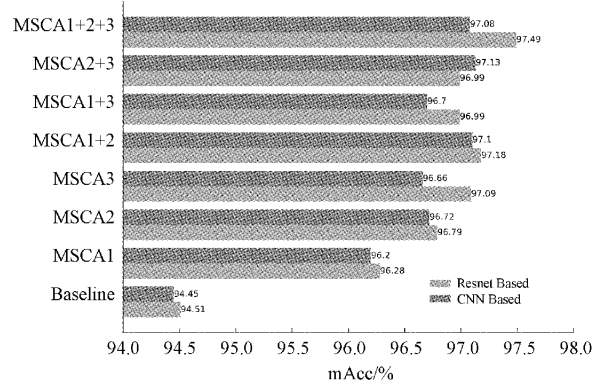


图4 MSCA 机制在网络不同位置对模型的影响 (上侧柱形图表示基于 CNN 的方法,下侧的表示基于 ResNet 的方法)

7) 讨论

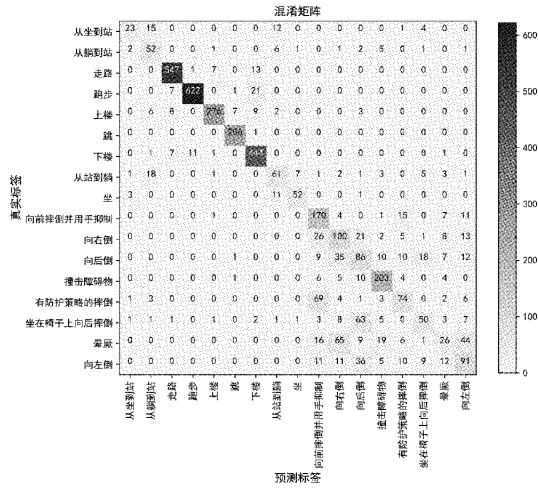
由表 3 可知,与 Gao 等^[3]方法相比,本文方法在准确率具有明显优势。Gao 等方法在 UCIHAR 数据集上取得了较好的效果,精度已经达到 96.65%,超越该方法有较高难度。但是,本文方法在 UCIHAR 数据集上取得了更好的结果,精度达到 97.20%,超出 Gao 等方法 0.55%。此外,本文方法在 DSADS 和 UniMib-SHAR 数据集上分类准确率相比 SOTA 方法分别提升 1.09%、0.59%。在 PAMAP2 数据集上超过 SOTA 方法 2.63%。因为数据集的采集方式不同, PAMAP2 数据集动作完成度高, DSADS, UniMib-SHAR 和 UCIHAR 数据集在采集过程中动作完成难度较大,需要网络提取多个传感器通道的特征,所以有较大的识别难度。通过使用多个感受野提取特征,增强多个传感器通道之间的关联,同时使用小尺度感受野保证特征的分辨率,进而提升网络对动作的分类精度。

可视化分析:以 UniMib-SHAR 数据集为例,本文使用混淆矩阵展示出基于 MSCA 的 HAR 模型和基线网络模型在 UniMib-SHAR 数据集上的分类效果,如图 5 所示。例如,“Falling Right”与“Falling Back”两个动作识别分类中,因为它们动作相似,只是摔倒方向不同。因此,网络需要同时考虑加速度信号和陀螺仪信号,如果仅使用固定尺度卷积核,很难提取到加速度信号和陀螺仪信号的关联特征。本文提出的 MSCA 机制通过使用小尺度的感受野提取高分辨率的特征信息,并使用大尺度的感受野提取多个传感器信号之间的关联,进而达到对相似动作更精准的分类。

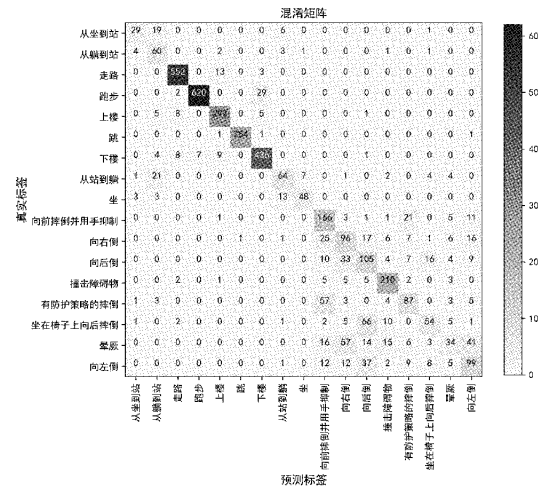
多尺度感受野尺寸设置:由于数据集传感器通道数量多样,采样频率不同,以及动作持续时间与信号变化多样。

大尺度卷积核可以通过增大感受野进而提取时间维度长序列依赖和传感器通道维度的跨通道关联。但是,其会造成信号特征尺寸缩小,不利于下游网络提取特征。为了解

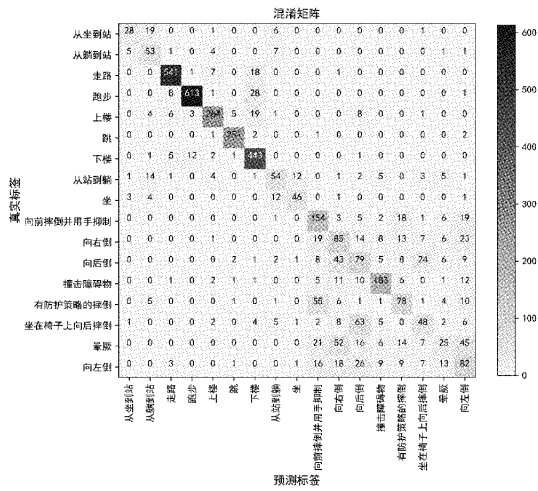
决以上问题,本文将 4 个不同尺度的感受野尺寸设置为 1、3、5 和 7。如果将大尺度的感受野设置为 9,会对特征造成更大压缩进而丢失低语义信息。



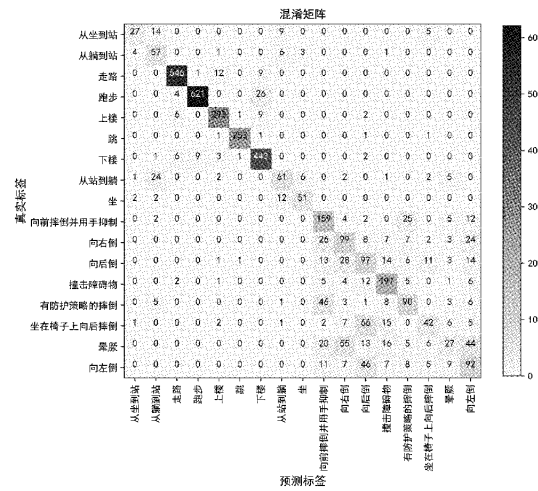
(a) ResNet网络



(b) ResNet+MSCA网络



(c) CNN网络



(d) CNN+MSCA网络

图 5 不同网络结构在 UniMib-SHAR 数据集上的混淆矩阵

因此本文将多个感受野尺寸设置为 1、3、5 和 7,这样既不会造成信息丢失,又可以提取多尺度特征信息。

4 结 论

本文提出了一种基于 MSCA 机制的人体行为识方法。MSCA 机制可以通过提取多个尺度的特征,并在特征通道维度建立多通道非线性关联,进一步提升了模型的识别能力。在 UCIHAR、DSADS、PAMAP2 和 UniMib-SHAR 四个数据集上的实验,结果表明加入 MSCA 机制对识别人体动作的准确率有显著的提升,对走路和跌倒等动作能够更有效的区分。上述实验均证明了 MSCA 机制对提取突变信号的动作以及周期信号的动作的特征是有效的。在未来的工作中,我们将进一步研究自适应尺度的行为识别方

法,针对不同信号特征,采取自适应卷积尺寸提取特征。

参考文献

[1] PEI L, GUINNESS R, CHEN R, et al. Human behavior cognition using smartphone sensors [J]. Sensors, 2013, 13(2): 1402-1424.

[2] ZENG M, NGUYEN L T, YU B, et al. Convolutional neural networks for human activity recognition using mobile sensors[C]. 6th international conference on mobile computing, applications and services. IEEE, 2014: 197-205.

[3] TENG Q, WANG K, ZHANG L, et al. The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition[J].

- IEEE Sensors Journal, 2020, 20(13): 7265-7274.
- [4] 杨万鹏, 李擎, 雷明. 人体动作识别的特征级融合 LSTM-CNN 方法研究[J]. 电子测量技术, 2021, 44(17): 173-180.
- [5] GAO W, ZHANG L, TENG Q, et al. DanHAR: Dual attention network for multimodal human activity recognition using wearable sensors[J]. Applied Soft Computing, 2021, 111: 107728.
- [6] ZENG M, GAO H, YU T, et al. Understanding and improving recurrent networks for human activity recognition by continuous attention[C]. Proceedings of the 2018 ACM international symposium on wearable computers, 2018: 56-63.
- [7] ZHANG H, ZU K, LU J, et al. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network [J]. ArXiv Preprint, 2021, ArXiv: 2105.14447.
- [8] AWAIS M, CHIARI L, IHLEN E A F, et al. Physical activity classification for elderly people in free-living conditions[J]. IEEE journal of biomedical and health informatics, 2018, 23(1): 197-207.
- [9] IGNATOV A D, STRIJOV V V. Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer[J]. Multimedia tools and applications, 2016, 75: 7257-7270.
- [10] WU D, WANG Z, CHEN Y, et al. Mixed-kernel based weighted extreme learning machine for inertial sensor based human activity recognition with imbalanced dataset[J]. Neurocomputing, 2016, 190: 35-49.
- [11] LIU C L, HSAIO W H, TU Y C. Time series classification with multivariate convolutional neural network [J]. IEEE Transactions on industrial electronics, 2018, 66(6): 4788-4797.
- [12] MA H, LI W, ZHANG X, et al. AttnSense: Multi-level attention mechanism for multimodal human activity recognition[C]. IJCAI, 2019: 3109-3115.
- [13] QIAN H, PAN S J, DA B, et al. A Novel Distribution-Embedded Neural Network for Sensor-Based Activity Recognition[C]. IJCAI: 2019, 2019: 5614-5620.
- [14] DUA N, SINGH S N, SEMWAL V B. Multi-input CNN-GRU based human activity recognition using wearable sensors [J]. Computing, 2021, 103: 1461-1478.
- [15] GAO W, ZHANG L, HUANG W, et al. Deep neural networks for sensor-based human activity recognition using selective kernel convolution [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-13.
- [16] TANG Y, ZHANG L, TENG Q, et al. Triple Cross-Domain Attention on Human Activity Recognition Using Wearable Sensors [J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022, 6(5): 1167-1176.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 7132-7141.
- [18] ANGUITA D, GHIO A, ONETO L, et al. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine [C]. Ambient Assisted Living and Home Care: 4th International Workshop, 2012: 216-223.
- [19] ALTUN K, BARSHAN B, TUNÇEL O. Comparative study on classifying human activities with miniature inertial and magnetic sensors [J]. Pattern Recognition, 2010, 43(10): 3605-3620.
- [20] REISS A, STRICKER D. Introducing a new benchmarked dataset for activity monitoring[C]. 2012 16th international symposium on wearable computers. IEEE, 2012: 108-109.
- [21] MICUCCI D, MOBILIO M, NAPOLETANO P. UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones[J]. Applied Sciences, 2017, 7(10): 1101.
- [22] HUANG W, ZHANG L, TENG Q, et al. The convolutional neural networks training with channel-selectivity for human activity recognition based on sensors[J]. IEEE Journal of Biomedical and Health Informatics, 2021, 25(10): 3834-3843.
- [23] HUANG W, ZHANG L, GAO W, et al. Shallow convolutional neural networks for human activity recognition using wearable sensors [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-11.
- [24] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision (ECCV), 2018: 3-19.
- [25] LIU S, YAO S, LI J, et al. Giobalfusion: A global attentional deep learning framework for multisensor information fusion[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2020, 4(1): 1-27.
- [26] ZHANG Y, WANG J, CHEN Y, et al. Adaptive

memory networks with self-supervised learning for unsupervised anomaly detection [J]. IEEE Transactions on Knowledge and Data Engineering, 2022,3:1164-1171.

作者简介

许晨炀, 硕士研究生, 主要研究方向为人工智能、行为

识别。

范非易, 工程师, 主要研究方向为生理信号处理等。

柯冠舟, 博士研究生, 主要研究方向为深度学习, 多模态学习。

沈建飞(通信作者), 工程师, 主要研究方向为普适计算领域可穿戴感知、无线信号感知和非接触生理监测等。