引文格式:李雯静,刘鑫.顾及人体骨架区域特征的行为识别研究[J].武汉大学学报(信息科学版),2025,50(3):571-578.DOI:10.13203/j.whugis20220020

Citation: LI Wenjing, LIU Xin. Action Recognition Considering Skeleton Region Characteristics[J]. Geomatics and Information Science of Wuhan University, 2025, 50(3):571-578.DOI:10.13203/j.whugis20220020

顾及人体骨架区域特征的行为识别研究

李雯静[1] 刘 鑫[1]

1 武汉科技大学资源与环境工程学院,湖北 武汉,430000

摘 要:基于人体骨架数据的行为识别研究目前已经取得较好的进展,然而现有方法大多仅考虑关节点的空间位置信息,忽视了关节点的区域变化特征。提出一种顾及人体骨架区域特征的行为识别方法,使用人体骨架数据表征人体行为特征,按照人体运动规律对骨架图进行区域划分,在关节坐标数据的基础上考虑区域内关节的角度变化情况,并将两种数据分别作为时空图卷积网络的输入,对两种数据流的预测结果进行融合。实验结果表明,所提方法较单个数据流的检测结果提高了1.9%;与几种经典模型比较,其Top-1和Top-5准确率分别达到了32.4%和54.2%,相较其他模型有更好的检测结果。

关键词:行为识别;区域特征;骨架图;时空图卷积网络

中图分类号:P208 **文献标识码**:A **收稿日期**:2024-03-20

DOI:10.13203/j.whugis20220020 **文章编号**:1671-8860(2025)03-0571-08

Action Recognition Considering Skeleton Region Characteristics

LI Wenjing[1] *LIU Xin*[1]

1 School of Resources and Environmental Engineering, Wuhan University of Science and Technology, Wuhan 430000, China

Abstract: **Objectives**: The research on action recognition based on human skeleton data has made good progress. However, most of the existing methods only consider the spatial position information of joint points and ignore the regional change characteristics of joint points. In order to solve this problem, an action recognition method considering the regional characteristics of skeleton is proposed. **Methods**: First, based on the human body structure, the joints and bones are constructed into a spatiotemporal skeleton map which represents the human action characteristics. The skeleton map is divided into regions according to the law of human movement. Then, the coordinates of nodes in the skeleton graph and the angle change data of the nodes in the region are used as the inputs of the spatiotemporal graph convolution network. Finally, the prediction results of the two data streams are fused to realize human action recognition. **Results**: In order to prove the effectiveness of the proposed method, it is verified on Florence 3D dataset and dynamics action dataset, respectively. The results show that the accuracy of the proposed method reaches 91.1% on Florence 3D dataset, which is 1.9% higher than that of a single data stream. The accuracies of Top-1 and Top-5 on dynamics action dataset reach 32.4% and 54.2%, respectively. **Conclusions**: Compared with the existing methods, the proposed method is proved to have better recognition accuracy and higher effectiveness through multiple sets of experiments.

Key words: action recognition; regional characteristics; skeleton diagram; spatiotemporal graph convolution network

近年来,随着深度学习的快速发展,人体行为识别[1-3]成为新兴的热门研究方向,被应用于包括智能监控[4]、智能医疗[5]和智能家居[6]等多个领域。人体行为识别是对视频数据中发生的动作

基金项目:湖北省高等学校优秀中青年科技创新团队计划(T2020002);武汉市重点研发计划(2024050702030122)。

第一作者:李雯静,博士,教授,主要研究方向为时空数据挖掘。liwenjing@wust.edu.cn

进行分析,并判断行为的类别,其中的关键问题是如何表达人体行为特征。

传统的行为识别是基于图像[7-8]的研究,以RGB数据作为输入进行特征提取,通过图像分类方法对行为进行分类。但真实场景中的图像数据总是会受到背景混乱、光照复杂等因素的影响,这对特征信息提取带来一定的困难。基于人体骨架信息的行为识别在一定程度解决了上述问题,首先利用骨架估测算法提取图像中的人体关键部位,将关键部位通过骨骼连接构建人体骨架图,然后对骨架图进行分类,从而实现行为识别[9-12]。这种方法可以在不同复杂环境下的图像中提取出需要的人体关键信息,减少了图像中其他信息带来的干扰。

传统基于人体骨架信息的行为识别方法是将每帧身体关节点坐标编码为特征向量进行分类[13-14],但是这些方法没有考虑关节点内部的联系,缺少对关节点关系及依赖的关注,而利用图卷积网络(graph convolutional network, GCN)[15]可以改善这一问题。文献[16]对此进行了研究并提出了时空图卷积网络(spatial temporal GCN, ST-GCN),同时在空间和时间两个维度提取特征,但仅考虑了一些骨骼直接连接的关节点特征,忽略了图结构中其他节点之间的联系。因此,文献[17]提出动作结构图卷积网络(actional-structural GCN, AS-GCN)识别骨架数据,引入动作链接以学习每个关节与其他任何关节存在的潜在关系,同时将更深层的骨架图关系表示为结构链接,实验结果证明其在精度上优于之前的网络,为采用GCN处理骨架数据提供了更好的思路。为了将空间注意机制和时间注意机制联系更加紧密,文献[18]提出一种新的嵌套时空注意力机制,通过将嵌套的时空注意模块嵌入到基本网络中进行行为识别。文献[19]将反馈机制引入到基于图卷积网络的行为识别当中,通过反馈图卷积块将高层语义信息特征和时间特征传递到浅层网络中,同时使用先验粗预测指导精确预测,从而完成行为识别。

上述方法在一定程度上提高了识别准确率,但是忽视了关节点坐标以外的属性信息。为此,本文提出一种顾及骨架区域特征的行为识别模型,以ST-GCN为基础模型,分别将关节位置信息和关节区域角度变化信息作为输入,将两个通道预测的得分相加得到最终预测结果。两组输入的数据流构成了双流网络,双流网络框架见图1。

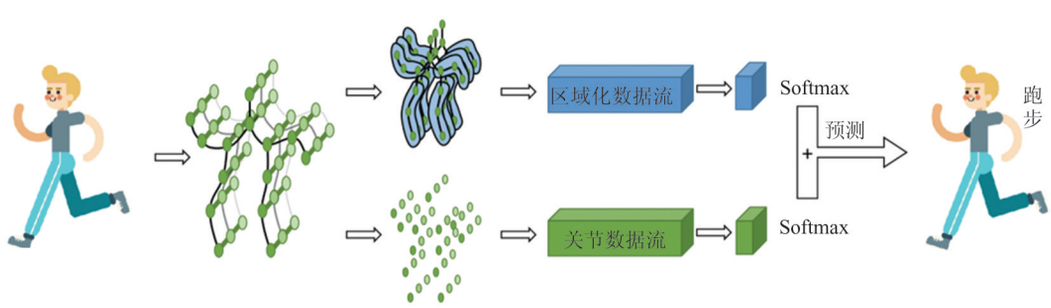


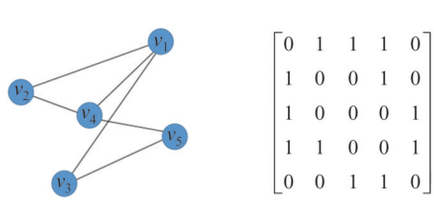
图1 双流网络框架

Fig. 1 Framework of Two-Stream Network

1 时空图卷积网络

1.1 时空骨架图结构

构建人体骨架图可以有效表达人体的运动信息。骨架图是以人体关节作为节点,骨骼作为边的一种图结构。图是一个具有广泛含义的对象,常被用来描述各类关系,在数学上是由一系列顶点和连接这些顶点的边构成的拓扑结构。将图表示为顶点和边的集合,记为$G=(V,E)$,其中$V=\{v_i\}$表示顶点集合,$i=1,2,\cdots,n$,$E=\{(v_i,v_j)|v_i,v_j\in V,i=1,2,\cdots,n,j=1,2,\cdots,n,i\neq j\}$,表示边集合。图的连接关系可以表示为邻接矩阵$\boldsymbol{A}$,$a_{ij}\in\boldsymbol{A}$表示$v_i$和$v_j$的连接关系,1表示连接,0表示不连接。图结构及其邻接矩阵如图2所示。



(a) 图结构 (b) 邻接矩阵

图2 图结构及其邻接矩阵

Fig. 2 Graph Structure and Adjacency Matrix

骨架图的建立可以有效表达视频每一帧的动作信息,包括节点的位置信息以及节点间的连接关系。但是建立单帧骨架图缺少了上下文时间信息的联系,因此需要构建时间与空间结合的时空骨架图,如图3所示。图3的节点集合为$V_t=\{v_{it}\}$,t表示帧数,$t=1,2,\cdots,m$。图3的结构由两部分组成,一部分为单帧内的所有关节点的连接,用$E_s=\{(v_{it},v_{jt})\}$表示;另一部分为相同关节点在不同帧下的连接,用$E_t=\{(v_{iT},v_{it})|T\neq t\}$表示。将顶点的坐标向量看作对应顶点的属性,记为$\boldsymbol{v}=(x,y)$或$(x,y,z)$。通过构建时空骨架图表达人体运动信息可以将关注点集中到人体结构本身,减少了图像噪音等因素的影响,提高了模型的适应能力。

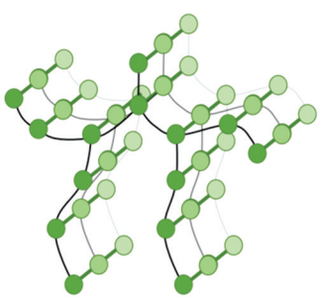


图3　时空骨架图

Fig. 3　Spatial Temporal Graph

1.2　时空图卷积

上述时空骨架图需要使用多层卷积网络提取特征。参照图像的卷积,定义一个$k\times k$的卷积核以及通道数为c的输入特征映射f_{in},则在点p处卷积后的输出f_{out}为:

$$f_{\text{out}}(x)=\sum_{h=1}^{k}\sum_{w=1}^{k}f_{\text{in}}(P(x,h,w))\cdot W(h,w) \quad (1)$$

式中,h和w分别表示输入的长和宽;P为采样函数,表示点p及其邻域,这里取邻域半径为1;W为权重函数,它提供了一个c维空间的权向量,且该权重与输入位置无关,所以在输入图上的任何位置都是共享权重的。

图像上的采样函数是在点p的相邻像素上定义的,推广到图上,可以先定义一个顶点v_i的1-邻域集$B(v_i)=\{v_j|d(v_i,v_j)\leqslant 1\}$,其中$d(v_i,v_j)$是点$v_i$与$v_j$间的距离,该距离用最小路径来表示。

图卷积的权重函数和图像的卷积不同,中心位置的邻域点没有固定的空间顺序,并且采样区域中顶点的数目是变化的,而权重向量的维数是固定的,所以引入函数l作为顶点和权重向量的唯一映射。该映射并没有对每个相邻节点设定一个唯一标签,而是通过将联合节点的节点邻域集划分为固定数量的n个子集来简化这个过程,并对每一个子集设定一个标签,实现标签映射是实现时空卷积的重点。

人体骨架是基于空间定位的,根据人体骨架的空间结构将节点的邻域集分为3个部分:(1)根节点集;(2)向心节点集,该子集元素为相较于根节点更接近骨骼重心的相邻节点;(3)离心节点集,该子集元素为相较于根节点更远离骨骼重心的相邻节点。骨骼重心为所有关节的平均坐标。分区策略的计算式为:

$$l_t(v_{ij})=\begin{cases}0, r_j=r_i\\ 1, r_j<r_i\\ 2, r_j>r_i\end{cases} \quad (2)$$

式中,$l_t(v_{ij})$表示t帧中点v_i到点v_j的映射;r_i是所有帧的骨架图中骨骼重心与v_i的平均距离;r_j是骨骼重心与v_j的距离。3种节点邻域子集的示意图如图4所示,其中黑色圆点为骨骼重心,黄色节点为根节点,蓝色节点为向心点,灰色节点属于离心点,节点上数字表示其与根节点的距离,蓝色区域为根节点的1邻域。

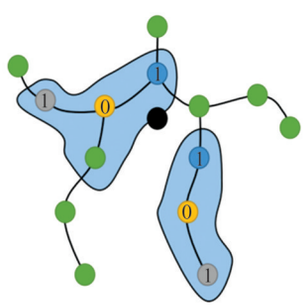


图4　节点邻域的分区示意图

Fig. 4　Partition Strategy Diagram

根据上述改进采样函数以及引入空间区域划分的映射函数,在空间维度下,节点v_i的卷积公式可改进为:

$$f_{\text{out}}(v_i)=\sum_{v_i\in B_i}\frac{1}{Z_{ij}}f_{\text{in}}(v_j)\cdot W(l_i(v_j)) \quad (3)$$

式中,B_i为v_i的采样区域,定义为与v_i距离为1的点集合;Z_{ij}为点v_i所在B_i子集的基数;l_i为引入的基于空间区域划分的映射函数。

2　顾及关节区域特征的行为识别方法

2.1　时空图卷积模型

在单帧情况下,时空图卷积的计算式为:

$$f_{\text{out}}=\boldsymbol{\Lambda}^{-\frac{1}{2}}(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{\Lambda}^{-\frac{1}{2}}f_{\text{in}}\boldsymbol{W} \quad (4)$$

式中,$\boldsymbol{\Lambda}$为度矩阵,用于归一化;$\boldsymbol{A}$为邻接矩阵;$\boldsymbol{I}$为单位矩阵;$\boldsymbol{W}$为权重矩阵。从图2(b)中可以看出,由于只考虑各边之间连接关系而忽略了节点自身的特征,所以邻接矩阵的对角线元素都是0。为了避免这个问题,在邻接矩阵$\boldsymbol{A}$的基础上加上单位矩阵$\boldsymbol{I}$。

式(4)只考虑了根节点的邻接矩阵,而根据§1.2所述节点邻域的分区策略,式(4)中邻接矩阵$\boldsymbol{A}$被分解为多个矩阵,用$\boldsymbol{A}+\boldsymbol{I}=\sum_j \boldsymbol{A}_j$表示,当$j=0$时,$\boldsymbol{A}_j=\boldsymbol{I}$;当$j=1$时,$\boldsymbol{A}_j=\boldsymbol{A}$。因此式(4)进一步转化为:

$$f_{\text{out}}=\sum_j \boldsymbol{\Lambda}_j^{-\frac{1}{2}}(\boldsymbol{A}_j\otimes\boldsymbol{M})\boldsymbol{\Lambda}_j^{-\frac{1}{2}}f_{\text{in}}W_j \tag{5}$$

式中,按照节点邻域的分区策略,$\boldsymbol{A}_j$分为$j=0$时根节点矩阵$\boldsymbol{A}_0$、$j=1$时向心节点矩阵$\boldsymbol{A}_1$和$j=2$时根节点矩阵$\boldsymbol{A}_2$;$\boldsymbol{M}$是一个可以学习的权重矩阵;$\otimes$表示两个矩阵之间的元素乘积;W_j表示v_j的权重函数。

时间维度下的卷积是用一个$k\times1$的卷积核对$c\times t\times n$的张量进行卷积,t为时间长度,n为顶点的数量。一个时空卷积基本模块主要由一个时间卷积层和一个空间卷积层组成,每个卷积层后面都有一个批量标准化层和一个激活函数(rectified linear unit,ReLU)层,基本模块如图5所示。完整的网络结构是由9层基本模块组合而成,数据在输入模型后首先通过归一化层对数据进行标准化,在模型末端执行全局平均汇集层,用于将不同样本的特征汇集到相同大小,最终输出到Softmax分类器以获得预测结果。

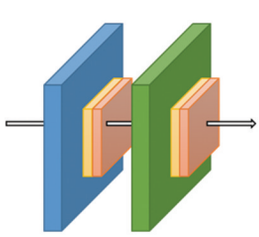
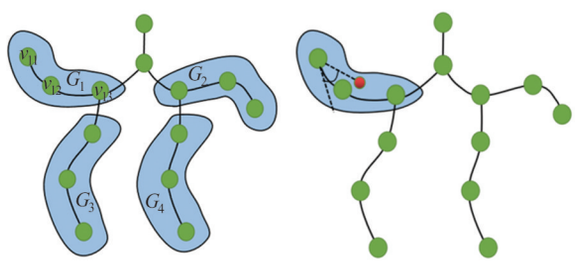


图5 基本模块

Fig. 5 Basic Module

2.2 关节的区域化信息

考虑到动作进行过程中关节和骨骼有按区域运动的规律,因此将成组的关节及骨骼变化信息作为输入是必要的。为了增强模型的适应能力,需要将关节坐标信息和区域角度变化信息分别作为模型的两个输入流。与节点邻域子集的划分不同,关节的区域划分是固定的,首先将骨骼及关节分成4个区域,两侧的手、肘和肩分别形成两个区域G_1、G_2,同样的,两侧的脚、膝和髋也分别形成两个区域G_3、G_4;每个区域都包括3个关节,将每个区域的关节点平均坐标作为该区域的重心,即$g_i=(x_i,y_i,z_i)$,骨架图的区域和角度划分如图6所示。



(a) 区域划分 (b) 角度划分

图6 骨架图的区域划分和角度划分

Fig. 6 Area Division and Angle Division in Skeleton Map

关节在区域内的变化情况是相对于区域重心的角度变化,具体步骤如下:

1)将每个区域的边缘关节(仅连接一个骨骼的关节)记为v_{i1},如图6(a)中的G_1区域,其边缘关节为手,将相邻关节记作v_{i2},第3个关节记作v_{i3}。

2)计算关节相对于区域重心的角度,计算式为:

$$\theta_{ij}=\arccos\frac{\overrightarrow{d_{ij}g_i}\cdot\overrightarrow{d_{ij}d_{i(j+1)}}}{\left|\overrightarrow{d_{ij}g_i}\right|\cdot\left|\overrightarrow{d_{ij}d_{i(j+1)}}\right|} \tag{6}$$

式中,θ_{ij}表示该区域重心与v_j的角度;g_i为区域的重心;d_{ij}为区域i的第j个关节。图6中$i\in[1,4]$,$j\in[1,3]$。

考虑到运动过程中区域外关节与骨骼的变化相对较小,所以为这些关节分配一个0°角,则按照上述步骤,每个关节都被分配了一个角度,这种区域化的关节数据就以角度的形式与每个关节唯一对应,和关节坐标数据一样作为图结构的节点特征。将两种数据流分别作为ST-GCN的输入,最后将两个通道预测的Softmax得分相加,作为最终预测结果。

3 实验与分析

3.1 实验配置

本文实验是基于Windows 10操作系统,中央处理器为i7-10750H,显卡为RTX3060。所有实验都是在PyTorch深度学习框架下进行。采用动量为0.9的随机梯度下降作为优化策略,训练批量大小设为32,训练轮次为65,初始学习率设为

0.1,在第45轮和55轮时为0.01。

3.2　数据集

Kinetics人体动作数据集包含从YouTube检索到的大约300 000个视频片段,包含了多达400多种人类动作类别,从日常活动、运动场景到复杂的互动动作都有涉及,其中各类动作的每个片段持续约10 s。该数据集仅提供没有骨架数据的原始视频剪辑,所以本文对每一帧的图像使用openpose算法提取人体的15个关节的2D坐标(x,y),同时每个坐标包含一个置信度。Kinetics人体动作数据集的数据按照帧数、坐标和置信度存储为json格式。

Florence 3D动作数据集由佛罗伦萨大学使用Kinect相机拍摄所得,它包括9项活动:挥手、喝水、接电话、鼓掌、系领带、坐下、站起来、看手表、鞠躬。该数据集解析格式为15个部位的3D坐标,按照样本序号、演员序号、动作类别、坐标的形式存储。

3.3　实验结果分析

3.3.1　人体行为可视化分析

本文的主要工作是引进了区域化的关节变化信息,为了说明人体运动过程中人体的关键部位具有区域化运动的特点,图7为挥手、坐下、鼓掌3种行为在不同帧下的骨架图可视化图。

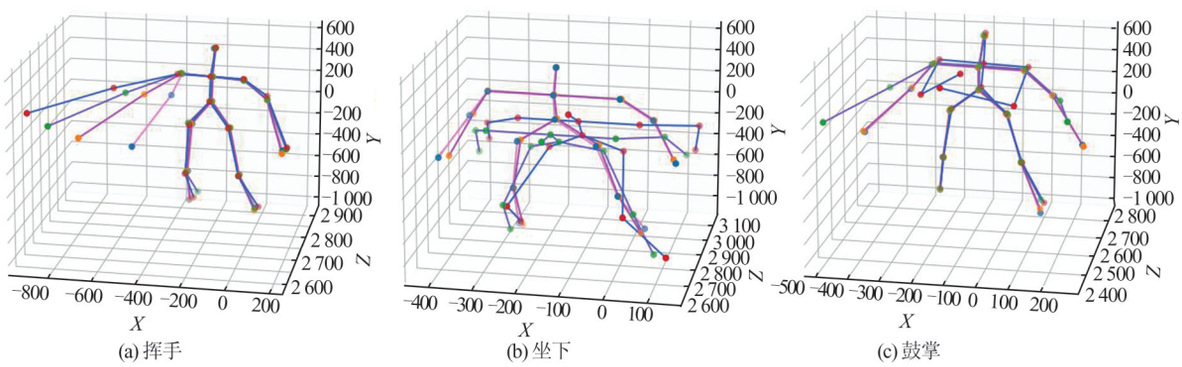


图7　不同行为的骨架图可视化图

Fig. 7　Skeleton Diagram Visualization of Different Behaviors

从图7可以看出,不同行为下人体的骨架运动状态不同,但是不同的人进行相同行为时,其部分区域的几个相邻关节运动具有相同规律。如图7(a)两个挥手动作,按照§2.2对关节的区域划分,不同的人在挥手时,其手、肘、肩区域运动的规律相同。

图8展示了挥手、坐下、鼓掌3个动作进行过程中,每个区域最外部的两个部位(8个部位依次为左手、左肘、右手、右肘、左脚、左膝、右脚、右膝)的变化关系强度的热力图,其中每个网格表示这两个部位在运动过程中相对该区域重心的角度变化率的相对变化强度,变化率越接近,其关系越强。部位与区域重心所构成的角度按照§2.2方法计算,每个动作的计算取随机3个片段的数据平均值。

从图8中可以看出,每个动作中相同区域的部位其角度变化率更接近,且上(下)半身的两个区域间不同部位的角度变化率比其与下(上)半身的两个区域的部位更接近。究其原因是大多数动作在进行过程中某两个区域也呈现同时成组运动的趋势,例如人在坐下时双腿的运动情况是一样的。

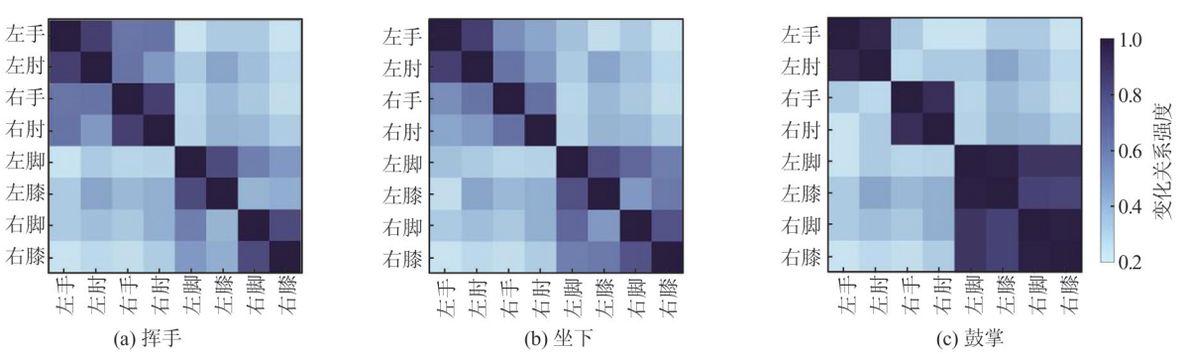


图8　不同动作重要部位相对于区域重心的角度变化率相对强度关系

Fig. 8　Angle Change Rate and Relative Intensity of Each Part of Different Actions Relative to the Regional Center

图9展示了3个不同的测试者在挥手、坐下、鼓掌时,每个区域最外部的两个部位相对该区域重心的角度变化情况。从图9可以看出,不同人进行相同的动作时,相同区域内的部位相对于区域重心的角度变化情况是相同的。同时从图9中还可以看到不同人做同一动作时由于其动作幅度不同、视频的获取方位不同等原因导致关节角度本身的差别,所以仅通过关节的坐标信息作为网络输入并不能完整表达人体的行为特征。

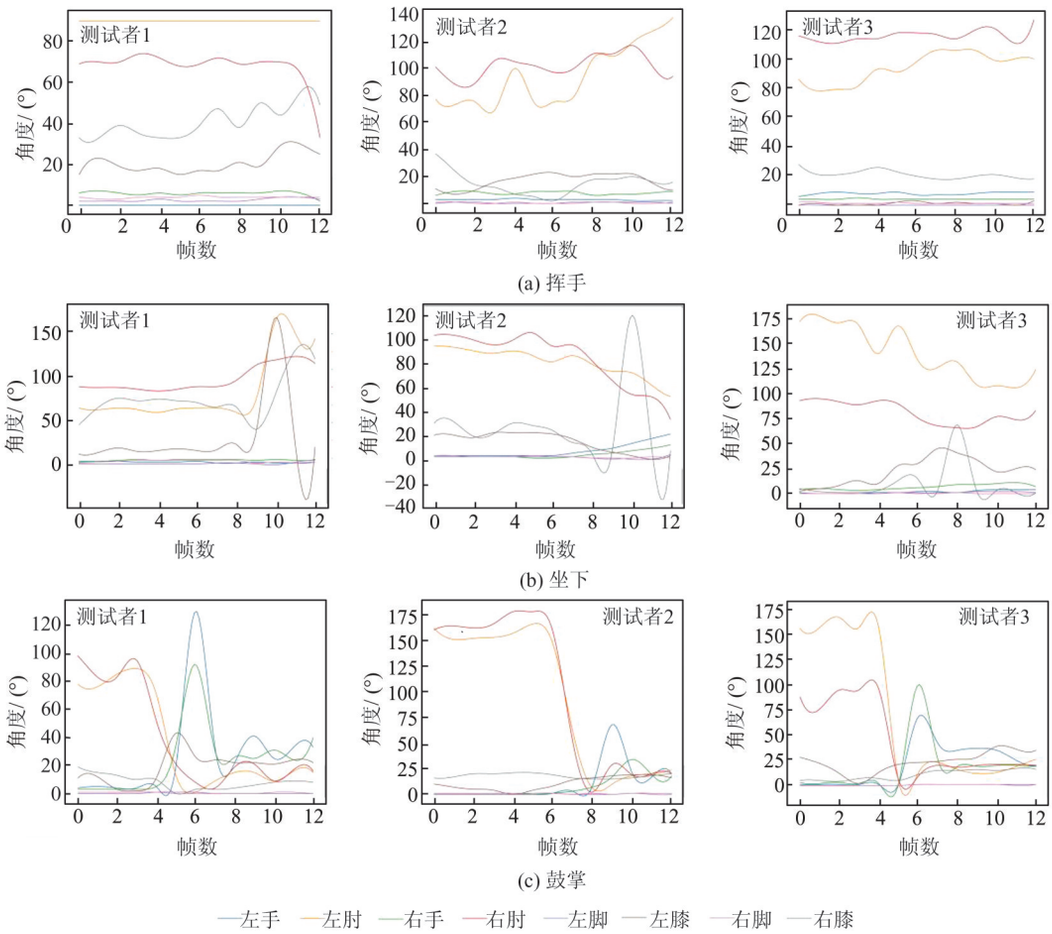


图9 不同动作的重要部位相对于区域重心的角度变化

Fig. 9 Angle Changes of Important Parts of Different Actions Relative to the Regional Center

3.3.2 对比实验

为了直观地说明本文方法有更好的效果,这里将两类数据输入单独进行实验以比较性能。为了方便记录,将不同帧的关节坐标数据记为一阶骨架信息;关节在区域内的角度变化数据记为二阶骨架信息。在Florence 3D动作数据集上分别检测了两种输入方法以及本文方法,其中一阶骨架信息输入的预测准确率为89.2%,二阶骨架信息输入的预测准确率为88.9%,而本文将两种方法融合后的预测准确率为91.1%。从结果可以看出,两种骨架数据单独输入时准确率相差不大,但是当融合结果后,其预测准确率有明显的提升。

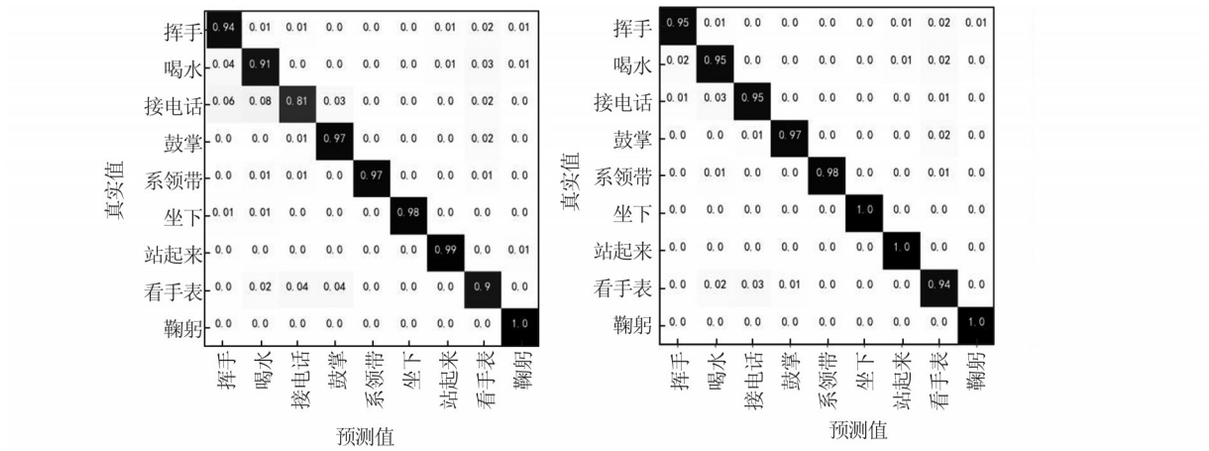
将仅输入一阶骨架信息的方法和本文方法分别在Florence 3D动作数据集上进行验证,得到标准化的预测混淆矩阵如图10所示。从图10中可以看出,本文方法在各类别动作的预测中都展示了较好的准确率,其中在接电话、喝水等局部动作中与没有增加关节区域变化信息相比具有更好的识别效果。

在Kinetics人体动作数据集上,通过Top-1和Top-5准确率两个指标将本文方法与几种经典方法的识别效果进行比较,其计算式分别为:

$$P_{\text{Top-1}}=\frac{S_1}{S} \tag{7}$$

$$P_{\text{Top-5}}=\frac{S_5}{S} \tag{8}$$

式中,$P_{\text{Top-1}}$、$P_{\text{Top-5}}$分别为Top-1和Top-5的准确率;S_1、S_5分别为预测结果最好的1个和5个类别与实际结果相符的样本个数;S为样本总数。



(a) 输入一阶骨架信息的方法　　(b) 本文方法

图 10　两种方法在 Florence 3D 动作数据集上的混淆矩阵

Fig. 10　Confusion Matrix of Two Methods on Florence 3D Action Dataset

表 1 是本文方法与其他几种经典网络模型预测结果的比较。从表 1 可以看出,与传统的几种网络模型相比,本文方法的识别准确率有所提高,这说明将区域化的关节角度变化信息作为输入可以提升模型的识别准确率。

表 1　双流网络与其他网络的精度对比/%

Table 1　Comparison of Two-Stream Network with Other Typical Networks/%

模型	Top-1 准确率	Top-5 准确率
Feature Enc[14]	14.9	25.8
Deep-LSTM[20]	16.4	35.3
TCN[21]	20.3	40.0
ST-GCN[16]	30.7	52.8
一阶骨架信息[16]	30.7	52.8
二阶骨架信息	28.8	50.5
本文方法	32.4	54.2

4　结　语

本文以人体骨架数据的物理连接作为关节邻接矩阵的基础,提出了一种顾及骨架区域特征的行为识别方法。分析人体运动过程中关节和骨骼的区域性运动规律,将关节点坐标数据与关节区域角度变化数据分别作为网络的输入。经实验验证,本文方法提高了基于人体骨架数据的行为识别准确率,与经典的行为识别方法相比有更好的表现。同时,证明了骨架关节坐标数据并不是基于骨架数据行为识别方法的唯一选择,顾及骨架区域化特征的行为识别方法为提高行为识别准确率提供了新的思路。

参　考　文　献

[1] 樊恒, 徐俊, 邓勇, 等. 基于深度学习的人体行为识别[J]. 武汉大学学报(信息科学版), 2016, 41(4): 492-497.
FAN Heng, XU Jun, DENG Yong, et al. Behavior Recognition of Human Based on Deep Learning[J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(4): 492-497.

[2] 毛琳, 陈思宇, 杨大伟. 引导式的卷积神经网络视频行人动作分类改进方法[J]. 武汉大学学报(信息科学版), 2021, 46(8): 1241-1246.
MAO Lin, CHEN Siyu, YANG Dawei. A Guided Method for Improving the Video Human Action Classification in Convolutional Neural Networks[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(8): 1241-1246.

[3] 周于涛, 吴华意, 成洪权, 等. 结合自注意力机制和结伴行为特征的行人轨迹预测模型[J]. 武汉大学学报(信息科学版), 2020, 45(12): 1989-1996.
ZHOU Yutao, WU Huayi, CHENG Hongquan, et al. Pedestrian Trajectory Prediction Model Based on Self-Attention Mechanism and Group Behavior Characteristics[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(12): 1989-1996.

[4] 徐敬海, 杜东升, 李枝军, 等. 一种应用传感器网和实景三维模型的复杂建筑物实时动态监测方法[J]. 武汉大学学报(信息科学版), 2021, 46(5): 630-639.
XU Jinghai, DU Dongsheng, LI Zhijun, et al. A Real-Time Dynamic Monitoring Method for Complex Building Applying Sensor Network and Reality 3D Model[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(5): 630-639.

[5] BAI L M. Intelligent Body Behavior Feature Extraction Based on Convolution Neural Network in Patients with Craniocerebral Injury[J]. *Mathematical Biosciences and Engineering*, 2021, 18(4): 3781-3789.

[6] 刘勇,谢若莹,丰阳,等. 智能家居中的居民日常行为识别综述[J]. 计算机工程与应用, 2021, 57(4): 35-42.
LIU Yong, XIE Ruoying, FENG Yang, et al. Survey on Resident's Daily Activity Recognition in Smart Homes[J]. *Computer Engineering and Applications*, 2021, 57(4): 35-42.

[7] XIE S N, SUN C, HUANG J, et al. Rethinking Spatiotemporal Feature Learning: Speed-Accuracy Trade-Offs in Video Classification[C]//The 15th European Conference on Computer Vision, Munich, Germany, 2018.

[8] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local Neural Networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.

[9] HAN J G, SHAO L, XU D, et al. Enhanced Computer Vision with Microsoft Kinect Sensor: A Review[J]. *IEEE Transactions on Cybernetics*, 2013, 43(5): 1318-1334.

[10] SAINATH T N, VINYALS O, SENIOR A, et al. Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks[C]//IEEE International Conference on Acoustics, Speech and Signal Processing, South Brisbane, Australia, 2015.

[11] KE Q H, BENNAMOUN M, AN S J, et al. A New Representation of Skeleton Sequences for 3D Action Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017.

[12] LI S, LI W Q, COOK C, et al. Independently Recurrent Neural Network (IndRNN): Building a Longer and Deeper RNN[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.

[13] DU Y, WANG W, WANG L. Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015.

[14] FERNANDO B, GAVVES E, JOSÉ ORAMAS M, et al. Modeling Video Evolution for Action Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015.

[15] DONG X W, THANOU D, RABBAT M, et al. Learning Graphs from Data: A Signal Representation Perspective[J]. *IEEE Signal Processing Magazine*, 2019, 36(3): 44-63.

[16] YAN S J, XIONG Y J, LIN D H. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition[C]//The 32nd AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, 2018.

[17] LI M S, CHEN S H, CHEN X, et al. Actional-Structural Graph Convolutional Networks for Skeleton-Based Action Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019.

[18] LI J P, WEI P, ZHENG N N. Nesting Spatiotemporal Attention Networks for Action Recognition[J]. *Neurocomputing*, 2021, 459(1): 338-348.

[19] YANG H, YAN D, ZHANG L, et al. Feedback Graph Convolutional Network for Skeleton-Based Action Recognition[J]. *IEEE Transactions on Image Processing*, 2021, 31(1): 164-175.

[20] SHAHROUDY A, LIU J, NG T T, et al. NTU RGB D: A Large Scale Dataset for 3D Human Activity Analysis[C]//IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016.

[21] KIM T S, REITER A. Interpretable 3D Human Action Analysis with Temporal Convolutional Networks[C]//IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017.