



计算机应用
Journal of Computer Applications
ISSN 1001-9081, CN 51-1307/TP

《计算机应用》网络首发论文

题目：基于骨骼时序散度特征的人体行为识别算法
作者：田志强，邓春华，张俊雯
收稿日期：2020-08-06
网络首发日期：2020-12-04
引用格式：田志强，邓春华，张俊雯. 基于骨骼时序散度特征的人体行为识别算法[J/OL]. 计算机应用.
<https://kns.cnki.net/kcms/detail/51.1307.tp.20201203.0859.008.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于骨骼时序散度特征的人体行为识别算法

田志强^{1,2,3}, 邓春华^{1,2,3*}, 张俊雯^{1,2,3}

(1. 武汉科技大学 计算机与技术学院, 武汉 430065; 2. 武汉科技大学 大数据科学与工程研究院, 武汉 430065; 3. 武汉科技大学 智能信息处理与实时工业系统湖北省重点实验室, 武汉 430065)
(* 通信作者电子邮箱 dchzx@wust.edu.cn)

摘要: 人体行为识别是智能监控、人机交互、机器人等领域的一项重要基础技术。图卷积神经网络(GCNs)在基于骨骼的人体行为识别方面取得了卓越的性能。利用 GCNs 在人体行为识别研究中存在以下难点问题: 1) 人体骨架骨骼点采用坐标表示, 缺乏骨骼点的运动细节信息; 2) 在某些视频中, 人体骨架的运动幅度太小, 关键骨骼点的表征信息不明显。针对上述问题, 首先提出骨骼点的时序散度模型, 描述骨骼点的运动状态, 放大了不同人体行为的类间方差; 并进一步提出了时序散度特征的注意力机制, 突显关键骨骼点, 进一步扩大类间方差。最后根据原始骨架的空间数据特征和时序散度特征的互补性构建了双流融合模型。该算法在权威的人体行为数据集 NTU-RGB+D 的两种划分策略下分别达到了 82.9 和 83.7 的准确率, 相比自适应图卷积网络(AGCN)提高了 1.3 和 0.5 个百分点。所提算法在数据集上精度的提升证明了本算法的有效性。

关键词: 骨骼; 行为识别; 图卷积; 时序散度; 注意力

中图分类号: TP18 **文献标志码:** A

Human behavior recognition algorithm based on skeletal temporal divergence feature

TIAN Zhiqiang^{1,2,3}, DENG Chunhua^{1,2,3*}, ZHANG Junwen^{1,2,3}

(1. School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan Hubei 430065, China;
2. Institute of Big Data Science and Engineering, Wuhan University of Science and Technology, Wuhan Hubei 430065, China;
3. Hubei Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan Hubei 430065, China)

Abstract: Human behavior recognition is an important basic technology in the fields of intelligent monitoring, human-computer interaction and robotics. Graph Convolutional Neural Networks (GCNs) have achieved excellent performance in skeleton-based human behavior recognition. The following problems exist in the research of human behavior recognition when using GCNs: 1) The human skeleton is represented by coordinates and lacks detailed information about the movement of the bone points. 2) In some videos, the motion amplitude of the human skeleton is too small, the characterization information of skeleton points is not obvious. Aiming at above problems, firstly, a temporal divergence model of skeleton-base was designed to describe the movement state of the skeleton, which amplifies the between-class variance of different human behaviors. In addition, the attention mechanism of temporal divergence features was designed to highlight pivotal points and further expand the between-class variances. Finally, a two-stream fusion model was constructed based on the spatial characteristics of the original skeleton and the temporal divergence characteristics. The algorithm achieved an accuracy rate of 82.9 and 83.7 respectively under two partitioning strategies of authoritative human behavior dataset NTU-RGB+D, which were 1.3 and 0.5 percentage points higher than that of Adaptive Graph Convolution Network (AGCN). The improvement of the accuracy of the proposed algorithm on the data set proves the effectiveness of the algorithm.

Key words: skeleton; behavior recognition; graph convolution; temporal divergence; attention

0 引言

视频序列中人体行为识别是计算机视觉中一个重要的任务,也是机器视觉、模式识别、人工智能等多个学科领域的交叉研究课题,在视频监控、人机交互、智能机器人、虚拟现实等领域被广泛应用^[1]。基于视频流的人体行为识别的方法很多,有些方法采用 RGB(Red Green Blue)图像序列^[2-3]、有些方

法采用深度图像序列^[4-5]、还有些方法采用双流融合的模式(例如,RGB+光流^[6]和人体骨架序列^[7]等)。人体骨架数据是人体关节和骨头的一种拓扑表示方式,在面对复杂背景以及人体尺度变化、视角变化和运动速度变化时具有先天优势,相比其它模式具有更小的运算消耗^[8]。此外,在深度传感器和人体姿态估计的技术不断成熟下,我们可以轻松获得准确的人体骨架数据。因此,许多研究人员和实际项目都使用人体

收稿日期:2020-08-06;修回日期:2020-11-15;录用日期:2020-11-17。

基金项目: 国家自然科学基金项目(61806150); 湖北省科技厅计划项目(2018CFB195); 湖北省教育厅科学技术研究计划青年人才项目(Q20181104); 智能信息处理与实时工业系统湖北省重点实验室开放基金项目(znxx2018QN09); 武汉科技大学国防预研基金项目(GF201814)。

作者简介: 田志强(1996—),男,湖北武汉人,硕士研究生,主要研究方向:计算机视觉、机器学习; 邓春华(1984—),男,湖南郴州人,副教授,博士,主要研究方向:计算机视觉、机器学习; 张俊雯(1997—),女,湖北荆门人,硕士研究生,主要研究方向:计算机视觉、机器学习。

骨架数据来做行为检测和识别^[9-13]。

传统的基于骨架的方法通常利用手工特征从特定的骨架序列中提取运动模式^[14]。然而,这些手工特征只在一些特定的数据集上表现良好,可能无法迁移到其它数据集,不具有普适性^[15]。随着深度学习方法在其它计算机视觉任务上的发展和优异表现,使用骨架数据的卷积神经网络(Convolutional Neural Networks, CNN)^[16-17]、循环神经网络(Recurrent Neural Network, RNN)^[18]、图卷积网络(Graph Convolutional Network, GCN)^[9]开始涌现。人体行为的骨架序列是关节点的自然时间序列,而RNN比较适合处理时间序列数据,因此基于RNN及其变种(例如,长短时记忆网络(Long Short-Term Memory, LSTM),门控循环单元(Gated Recurrent Unit, GRU))的骨架行为识别方法比较多。CNN网络模型通常将一个切片上的骨骼数据转换成伪图像的形式,然后再利用CNN在伪图像上提取相应的深度学习特征。当CNN处理骨架数据序列时,通常需要结合RNN模型。RNN的时序上下文信息和CNN丰富的空间信息相结合往往可以取得比单一的结构模型更好的效果。最近两年,很多学者开始将GCN应用于骨架的行为识别。GCN是图论与卷积神经网络的融合,其本质目的就是用来提取拓扑图的关系特征。人体骨架序列本身就是一个自然的拓扑图结构,GCN网络模型更适合描述骨骼关键点之间的空间和时序拓扑信息。因此,GCN处理骨架数据任务的时候,比RNN更具优势。

由于图论比较复杂,难以构造有效的深层GCN。文献[13]利用切比雪夫多项式逼近算法提出了一种近似的图卷积结构,使得图卷积操作变成了双线性模型,从而能够构造深层有效的GCN结构。ST-GCN(Spatial Temporal Graph Convolutional Networks)^[9]首次利用文献[13]的理论对人体骨架数据进行建模,性能明显优于同时期的RNN算法^[1]。该算法根据人体的拓扑结构构建邻接矩阵,用以描述人体骨架的空间结构,并提出了一种基于距离采样的函数作用于GCN的卷积层,使神经网络更好的学习到人体骨骼数据之间的空间和时序上下文信息。许多研究者在ST-GCN^[9]的基础上,通过改进,提出了一系列的基于人体骨架的行为识别算法^[10-12]。对于种类繁多的行为动作,固定的骨架图结构往往不是最佳的选择,2s-AGCN(Two-Stream Adaptive Graph Convolutional Networks)^[10]提出自适应图卷积层,根据不同的动作,生成相对有效的图模型;并利用双流模型,实现数据之间的信息互补,用以提升识别效果。AS-GCN(Actional-structural graph convolutional networks)^[11]引入一种解码-编码结构用于获取骨架点之间的依赖信息,并结合结构链接图,提出一种双流GCN的行为识别算法。RA-GCN(Richly activated graph convolutional network)^[12]在前人的基础上,直接发展成多流人体骨架数据,构建一种分支结构的GCN,最后再使用softmax分类器对多个分支就行融合。这些方法都在ST-GCN的基础上提出了特有的双流或者多流算法。

然而,目前的GCN算法在数据提取上仍然存在不足:输入的人体骨架数据切片仅仅包含人体骨骼点的空间信息,骨架的时序上下文信息需要通过GCN模型获取。人体骨架数据是由人体骨架的三维坐标数据组成,缺乏外观细节和运动细节信息,极大限制了骨架数据在行为识别上的性能。外观细节无法通过关节坐标点表征,然而运动细节信息却可以通过关节坐标点展现。虽然时序上下文信息可以通过大量样本

训练得到,但是基于数据驱动的特征具有更强的一般化运动信息而缺乏运动细节描述。针对上述问题,本文提出了一种基于骨骼点时序散度的特征模型,用来描述骨骼点在时间维度上的位移及运动信息。本研究发现同一种骨骼点在不同行为的时序骨架数据的分散度可能不同。在物理学中,表征空间点矢量场发散的强弱程度一般使用散度场理论来描述,其意义在于描述场的有源性。受此启发,本文将物理中的散度场理论引入到骨骼点的运动细节信息的描述,提出一种散度场的时序特征模型。考虑到骨架关键点的坐标误差对运动细节影响较大,本文提出了一种散度特征注意力机制,用以增强散度模型的有效性。时序散度特征和原始的骨骼数据有很强的互补性,分别从运动和静态的角度对骨骼数据进行了描述。在此基础之上,本文构建了一种新的双流GCN模型。

为了验证散度模型和本文的双流模型的有效性,我们在目前权威的人体行为数据集NTU-60^[19]和NTU-120^[20]上做了充分的验证,性能均超过了上述所提出的方法。本文主要工作有:

- (1)提出了一种骨骼数据的散度模型,描述骨骼点的运动细节信息;
- (2)提出了一种骨骼数据的注意力机制,增强散度模型的有效性;
- (3)根据原始骨架的空间数据特征和时序散度特征的互补性构建了双流融合模型。

1 图卷积神经网络

面对深度神经网络难以学习复杂数据之间依赖关系的问题,使用图卷积神经网络往往是不错的选择。目前图卷积的方法主要分为两种:空域卷积^[21-22]和频域卷积^[23-24],空域卷积是利用卷积核直接对图的顶点及其邻居节点进行卷积操作,并手动设计规则进行相应进行特征提取和归一化。频域卷积则利用图的邻接矩阵导出其频域上的拉普拉斯算子,利用傅里叶变换^[25],将图映射到频域上的欧式空间中进行卷积。本文采用的是空域卷积的方法。

图卷积神经网络中,需要通过数据间的依赖关系构建相应的图模型。为了描述人体骨骼点之间的依赖关系,本文将参照文献[9]中的方法,根据人体的拓扑结构构建图模型。文献[9]将邻接矩阵划分为3个子集。

①根节点本身;②向心集:比根节点更靠近骨架中心点的相邻节点;③离心集:比根节点更远离骨架中心点的相邻节点。表达式如式(1)所示。

$$A_{ij} = \begin{cases} 0 & \text{if } r_i = r_j \\ 1 & \text{if } r_i < r_j \\ 2 & \text{if } r_i > r_j \end{cases} \quad (1)$$

式(1)中, A_{ij} 表示*i*节点到*j*节点的相对距离, r_i 和 r_j 分别表示*i*,*j*两个骨骼点到人体重心点的相对距离,在NTU-RGB+D^[19-20]数据集中取人体胸口为重心点。

在确定图模型之后,需要构建图卷积神经网络来对数据提取相应特征。基于图的卷积网络实现起来比2D或3D卷积复杂许多。本文按照文献[9]中的方法构建图卷积网络模型,其定义式如下:

$$f_{\text{out}} = \Lambda^{1/2} (A_k + I) \Lambda^{1/2} f_{\text{in}} M, \quad (2)$$

其中 f_{in} 和 f_{out} 分别表示数据的输入和输出, M 为权重矩阵。 I 表示单位矩阵, $\Lambda^{1/2} (A_k + I) \Lambda^{1/2}$ 表示第*k*个样本邻接矩阵和关

联矩阵的拉普拉斯归一化。输入特征 f_{in} 为 $C \times T \times N$, C 表示特征维度, T 表示帧数, N 表示单个骨骼的骨骼点个数。

2 时序散度特征模型及其双流网络

本节主要介绍本文的方法原理, 流程如图 1 所示。首先探究人体骨骼点在不同行为中的运动差异性, 并引出骨骼点时序散度的基本概念, 由此构建骨骼点时序散度特征模型; 并为骨骼点散度特征模型添加注意力机制, 进一步增大类间方差; 接着为了探究不同数据流之间的互补关系, 将骨骼点散度特征流与空间特征流相融合, 训练分类器预测结果; 最后给出详细的实施步骤以及伪代码。

2.1 构建骨骼点的时序散度特征流

对于行为识别任务, 人们往往能通过某些骨骼点的运动差异性来分辨某些人体行为。例如, 当人们主观去辨别“挥手”和“踢腿”这两个行为时, 可以通过手臂和腿部骨骼的运动差异性来对这两种行为做出辨别。在此, 本文做了一个简单的实验来论证上述观点。

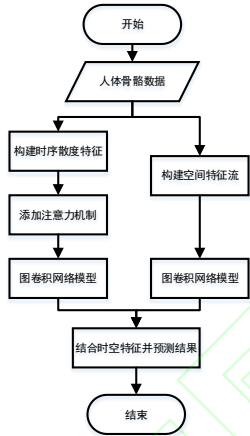


图 1 本文方法流程

Fig. 1 Flow chart of the proposed method

本文提取 NTU-RGB+D60^[17] 数据集中类别为“挥手”和“踢腿”的两个视频, 计算视频序列中所有时刻手腕和脚踝在 x, y, z 三个方向上的平均位移矢量 $\Delta x, \Delta y, \Delta z$, 并将其标记在三维坐标系上。

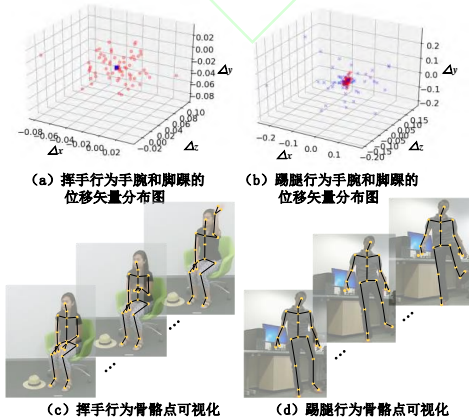


图 2 不同人体行为间的运动差异

Fig. 2 Movement differences between different human behaviors

图 2 中, (a) 和 (b) 分别表示“挥手”和“踢腿”两个人体行为类别中人物骨骼点位移矢量的分布图。(c) 和 (d) 分别表示

“挥手”行为及“踢腿”行为的骨骼点的可视化。(a) 和 (b) 中的红点与蓝“x”分别表示视频中所有时刻手腕和脚踝在 x, y, z 三个方向上的位移矢量。从图中可以明显的看出, 对于“挥手”这个类别, 其手腕的位移矢量分布的离散程度要大于脚踝。而对于“踢腿”这一类, 脚踝的位移矢量分布的离散程度则要大于手腕。依靠骨骼点位移矢量分布的离散程度, 人们就能很容易的区分“鼓掌”和“踢腿”这两种行为。

为了更好的描述上述试验中提出的骨骼点位移矢量的分布状态, 本文给出骨骼点时序散度的概念: 骨骼点的时序散度可以用来描述人体骨骼点在时间维度的运动状态。例如, 在某一时刻, 骨骼点时序散度的大小可以表示骨骼点在该时刻的运动幅度; 而在宏观上, 在某一时间段内, 每个时刻散度分布的离散程度可以反应骨骼点的运动幅度。后文会围绕上述概念对骨骼点的时序散度给出公式及定义, 在此之前, 本文将率先介绍笛卡尔坐标系中向量场的散度定义。

在三维笛卡尔坐标系中, 一个连续可微向量场的散度场可以用来描述该向量场的强弱, 其表达示如下:

$$\text{div} \mathbf{F} = \nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} \mathbf{i} + \frac{\partial F_y}{\partial y} \mathbf{j} + \frac{\partial F_z}{\partial z} \mathbf{k}, \quad (3)$$

其中, $\frac{\partial F_x}{\partial x}, \frac{\partial F_y}{\partial y}, \frac{\partial F_z}{\partial z}$ 分别表示向量场中任意一点在 x, y, z 方向上的偏导数, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ 分别表示 x, y, z 轴方向上的单位向量。

人体骨骼点在某一时间段内的运动轨迹恰好构成了向量场。则对于任意时间段 $[t, t + \Delta t]$, 骨骼点的轨迹向量场可表示如下:

$$\mathbf{V}^t = \mathbf{C}^{t+\Delta t} - \mathbf{C}^t, \quad (4)$$

其中 \mathbf{C} 表示骨骼点的点源坐标集合, \mathbf{C}^t 表示视频 t 时刻骨骼点的坐标特征。

轨迹向量场可以放大不同行为之间的类间方差。式(3)中首描述的是物理学中向量场在空间坐标系下的变化状态。而本文所研究的是骨骼点轨迹向量场在时间维度上的变化状态。参照式(3), 可以衍生出骨骼点时序散度的定义:

在任意时间段 $[t, t + \Delta t]$ (Δt 趋近于 0), 人体骨骼点的位移矢量与时间变化量的比值即为骨骼点在时刻 t 时刻的时序散度。在空间坐标系中, 骨骼点的位移矢量可分解为 x, y, z 三个方向上位移分量, 骨骼点的时序散度的定义式如下:

$$\text{div} \mathbf{V}^t = \nabla \mathbf{V}^t = \frac{V_x^t}{dt} \mathbf{i} + \frac{V_y^t}{dt} \mathbf{j} + \frac{V_z^t}{dt} \mathbf{k}, \quad (5)$$

其中 $\mathbf{i}, \mathbf{j}, \mathbf{k}$ 分别表示 x, y, z 轴方向上的单位向量。如式(5)所示, 对于任意时刻, 每一个体的骨骼点都可以提取 3 维时序散度。其物理意义是描述骨骼点在时间维度上的运动幅度。

在某一完整的人体行为视频序列中, 融合每一帧所有骨骼点的时序散度则可得到该视频的时序散度特征:

$$\mathbf{T}_s = \{ \text{div} \mathbf{V}_k^t | t \in T, k \in K \}, \quad (6)$$

其中 K 表示骨骼点总个数, 在 NTU-RGB+D 数据集中, 人体的骨骼点个数为 25。时序散度特征即为骨骼点每一时刻时序散度的集合。宏观上可以看作是人体骨骼连续的运动轨迹及瞬时的运动幅度。

2.2 骨骼点时序散度特征的注意力机制

在行为识别任务中, 人们往往只需要通过“手”, “足”这些具有代表特征的骨骼点就能判别某些行为。这些骨骼点都有一个共同特点: 他们距离人体重心点的距离较远, 运动半径更

大,灵活性更强。相比于人体上灵活性相对较弱的骨骼点(例如肩膀,跨部),**手腕与脚踝这些骨骼点往往对人体行为识别任务起到更为关键的作用。**

为了突显这些骨骼点的特征,本文将为这些骨骼点添加注意力机制。具体算法如下文。

选取人体重心点为基准,用重心点到其他各个骨骼点的连线 l 表示各个骨骼点到人体重心的相对距离。则在相邻时刻 $[t, t + dt]$,骨骼点 j_k 的到重心点的平均距离 L 可表示如下:

$$\bar{l}_k = (|l_k^t| + |l_k^{t+dt}|)/2, \quad (7)$$

其中 $\bar{l} \in [0, 1]$,其数值过小,无法起到表征的作用。在此,本文使用 δ 激活函数,将 \bar{l} 归一化到新的区间 $[a, b]$ 。 δ 激活函数的定义式如下:

$$\delta(x) = \ln(x + p) + q, \quad (8)$$

令 $x = L$,将原始区间 $[0, 1]$ 以及欲生成的新区间 $[a, b]$ 代入式(8),可解 $p = 1/(e^{b-a} - 1)$, $q = b - \ln(e^{b-a}/(e^{b-a} - 1))$ 。

在实验过程中, a, b 的默认取值分别为0.8和1。对于任意时刻 t ,骨骼点 j_k 重心散度特征的定义式如式(9)所示:

$$\text{div } \bar{V}_k^t = \text{div } V_k^t \delta(\bar{l}_k^t), \quad (9)$$

其中 $\text{div } V_k^t$ 表示式(5)中的骨骼点的时序散度,其描述的是骨骼点的运动方向和运动幅度。 \bar{l}_k^t 为式(7)中时刻骨骼点 j_k 到重心点的平均距离。它所代表的物理意义是骨骼点距离人体重心点的相对位置关系,其值越大,则该骨骼点距离重心点越远;其值越小,则该骨骼点距离重心点越近。最终得到添加注意力机制的时序散度特征的表达式如下:

$$Ts = \{\text{div } \bar{V}_k^t | t \in T, k \in K\}. \quad (10)$$

注意力机制可以增强距离人体重心点较远以及运动幅度较大的骨骼点的表征能力,进一步增加不同人体行为的类间方差,使神经网络更容易辨别到不同人体行为。

2.3 时序散度特征与空间特征相融合

在ST_GCN^[9]以及AGCN^[10]中,研究者都是将人体骨架以点源坐标的形式输入到神经网络中,**点源坐标属于人体骨骼点的空间特征**,而本文所提出时序散度特征流为骨骼点的时序特征。本文认为空间特征与时间特征有良好的互补关系。为了验证上述观点,本文将构建双流网络,将时序散度特征分别与点源流特征相结合,训练分类器并预测最终结果。其表达式如下:

$$res = \sigma(Ts(x_i) \oplus Js(x_i)), \quad (11)$$

其中, $\sigma(\cdot)$ 表示激活函数,在本文试验中 $\sigma(x) = \text{softmax}(x)$, res 表示预期结果。 $Js(\cdot)$ 和 $Ts(\cdot)$ 分别表示骨骼点的时序散度特征流以及点源流特征的图卷积网络模型。 \oplus 表示矩阵拼接。

2.4 算法的实现

上文中介绍了本文中所提出的方法及原理。本文将在本节描述整体流程:本文首先获取视频中骨骼点的坐标数据矩阵 $Data$ 。该矩阵维度为 $[N, T, C, V]$,这些参数分别表示所有视频中的总人数、视频帧数、特征维度以及骨骼点个数。在NTU-RGB+D^[17-18]数据集中, $C = 3, V = 25$ 。接着利用式(4)求出轨迹向量 t 。确认归一化区间 $[a, b]$,并根据2.2中所示方法解出式(8)中的实参 p 和 q 。并根据式(7)求得骨骼点到人体重心点的平均距离 L ,将 $x = L$ 代入式(8)中将其归一化到区间 $[a, b]$ 之间。利用式(9)可求得 x_t, x_i 表示公式时序散度特征。

然后分别构建时序图卷积神经网络 $Ts(\cdot)$ 以及空间图卷积神经网络 $Js(\cdot)$ 。用它们分别提取时序散度和点源的抽象特征,接着采用矩阵合并的方式将上述抽象特征融合,最后使用 softmax 得到分类结果 res 。其伪代码如下。

输入:视频中的骨骼数据 $S = [N, T, C, V]$,间隔时间 c ,归一化区间 $[a, b]$,绝对值函数 $\text{abs}(\cdot)$,点源特征 x_j ,时序图卷积神经网络 $Ts(\cdot)$,空间图卷积神经网络 $Js(\cdot)$,矩阵拼接函数 $\text{concat}(\cdot)$,分类器 $\text{softmax}(\cdot)$ 。

输出:预测结果 res

```

1) BEGIN
2)  $V \leftarrow (S[:, c: T: c, :, :] - S[:, 0: (T - c): c, :, :]) / c$ 
3)  $p \leftarrow 1 / (e^{(b - a)} - 1)$ 
4)  $q \leftarrow b - \ln(e^{(b - a)} / (e^{(b - a)} - 1))$ 
5)  $L \leftarrow V - V[:, :, 20: 21, :]$ 
6)  $L \leftarrow \text{abs}(L[:, 0: T: 1, :, :] + L[:, 0: (T - 1): 1, :, :]) / 2$ 
7)  $x_t \leftarrow \ln(L + p) + q$ 
8)  $J \leftarrow Js(x_j)$ 
9)  $T \leftarrow Ts(x_t)$ 
10)  $res \leftarrow \text{softmax}(\text{concat}(J, T))$ 
11) END

```

3 实验与分析

3.1 数据集

NTU-RGB+D 60^[19]: NTU-RGB+D 60是公开的3D人体行为数据集。该数据集包含60个动作类别,总共包含57880个视频样本。该数据集的拍摄和剪辑工作由40个年龄段在10至35岁之间的志愿者所执行。每个动作由3台摄像机拍摄,这3台摄像机拍摄时选择的高度相同,但水平角度不同,分别为: $-45^\circ, 0^\circ, 45^\circ$ 。该数据集利用 $kinect$ 深度传感器检测到每一帧人体的3D骨骼点序列,骨骼序列中每个人有25个骨骼点,骨骼点的分布如图3所示,且所有视频中的总人数不超过2人。论文[19]中提供了2个基准来对数据集进行划分数据集: $X\text{-Sub}$ 和 $X\text{-View}$ 。其中 $X\text{-Sub}$ 按拍摄视频中参与者的不同来将数据集划分为训练集(40320个视频)和验证集(16560个视频); $X\text{-View}$ 将1号和3号摄像机拍摄到的37920个视频作为训练集,将2号摄像机拍摄到18960个视频作为验证集。本文将遵循上述两种划分方式来评估本文所提出的模型。

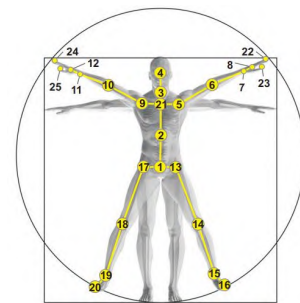


图3 NTU-RGB+D 骨架示意图^[17]

Fig. 3 NTU-RGB+D skeleton diagram

NTU-RGB+D 120^[20]: 和NTU-RGB+D 60^[19]一样,该数据集是3D行为识别的公开数据集。文献[20]中提供了2个基准来对数据集进行划分: $X\text{-Sub}$ 和 $X\text{-View}$ 。其中 $X\text{-Sub}$ 按拍摄视频中参与者的不同来将数据集划分为训练集(63026个

视频)和验证集(50 919个视频)。X-View 根据视频编号来对数据集进行划分,将编号为偶数的视频作为训练集(75 823个视频),编号为奇数的视频作为测试集(38 122个视频)。本文将遵循上述两种划分方式来评估本文所提出的模型。

3.2 训练详情

本文的所有实验均使用 Pytorch^[39]作为深度学习框架。训练方式参考文献^[10]:在 NTU-RGB+D^[19-20]数据集上,单次训练所选取的样本数为 64($baseline=64$)。选择交叉熵作为反向传播梯度的损失函数,权重衰减率设置为 0.0001。对于 NTU-RGB+D^[19-20]数据集,本文选取的样本人数为 2。若视频

中总人数大于 2,则取其中 2 人;若总人数小于 2,则取目标视频中的总人数,并在数据矩阵中用 0 补全空缺位置。视频总共取 300 帧,如果所选视频帧数少于 300,则用 0 补全空缺位置;总的迭代次数设置为 60,初始的学习率设为 0.1,并在第 30 次迭代时和第 40 次迭代时的时候分别除以 10。用上述方法完成两个单流网络的训练后,本文将分别提取网络输出特征。并将得到的特征融合,接入链式神经网络继续训练,得到预测的最终结果。训练链式神经网络时,单次训练所选取的样本数设置为 256,初始学习率为 0.1,迭代次数设置为

20,损失函数同上文所述相同。

3.3 融合特征的实验结果分析

为了探求骨骼空间特征与骨骼点时序散度特征的互补关系,本文将构建双流网络,并在 NTU-RGB+D^[19-20]上做相应的测试。实验结果如表 1 与表 2 所示。其中 JT-STGCN (Joint and Temporal- Spatial Temporal Graph Convolutional Networks) 与 JT-AGCN 均表示特征融合双流网络,所融合的特征数据分别表示点源流 J_s 以及时序散度流 T_s ,所选用的基准模型分别为 ST-GCN^[9]与 AGCN^[10]。

如表 1 和表 2 所示,在 NTU-RGB+D60^[19]上,2s-AGCN^[10]在 X-Sub 和 X-View 两种划分方法下的准确率分别为 88.5% 和 95.1%;本文结合时序散度特征流以及点源特征流,在 sub 和 view 两种划分方法下的准确率分别为 89.3% 和 95.5%。在 NTU-RGB+D120^[20]上,2s-AGCN^[10]在 X-Sub 和 X-View 两种划分方法下的准确率分别为 81.6% 和 93.2%;本文结合时序散度特征流以及点源特征流,在 X-Sub 和 X-View 两种划分方法下的准确率分别为 82.9% 和 83.7%。可以看出时序特征与空间特征的互补能力要强于原版 2s-AGCN^[2]中的双流网络。

表 1 在 NTU-RGB+D 120 数据集上的对比实验结果

Tab. 1 Comparative experiment results on the NTU-RGB+D 120 dataset

网络模型	准确率/%	
	X-Sub	X-View
Part-Aware LSTM ^[19]	25.5	26.3
Soft RNN ^[26]	36.3	44.9
Dynamic Skeleton ^[27]	50.8	54.7
Spatio-Temporal LSTM ^[28]	55.7	57.9
Internal Feature Fusion ^[29]	58.2	60.9
GCA-LSTM(Global Context-aware attention LSTM networks) ^[30]	58.3	59.2
Multi-Task Learning Network ^[31]	58.4	57.9
FSNet(Fixed Scale Network) ^[32]	59.9	62.4
Skeleton Visualization (Single Stream) ^[33]	60.3	63.2
Two-Stream Attention LSTM ^[34]	61.2	63.3
Multi-Task CNN with RotClips ^[35]	62.2	61.8
Body Pose Evolution Map ^[36]	64.6	66.9
Magnitude-Oriented (TSA) ^[37] + Yang et al. ^[38]	67.7	66.9
ST-GCN ^[9]	71.9	72.5
2s-AGCN ^[10]	81.6	83.2
JT-AGCN(2s)(Joint and Temporal-Adaptive Graph Convolutional Networks)	82.9	83.7

表 2 在 NTU-RGB+D 60 数据集上对比实验结果

Tab. 2 Comparative experiment results on the NTU-RGB+D 60 dataset

网络模型	准确率/%	
	X-sub	X-View
Lie Group ^[40]	50.1	82.8
ARRN-LSTM(Attentional Recurrent Relational Network-Long-Short Term Memory) ^[41]	80.7	88.8
Ind-RNN(Independently-Recurrent Neural Network) ^[42]	81.8	88.0
CNN+Motion+Trans ^[43]	83.2	89.3
3scale ResNet152 ^[44]	85.0	92.3
ST-GCN ^[1]	81.5	88.3
DPRL+GCNN(Deep Progressive Reinforcement Learning + Graph Convolutional Neural Networks) ^[45]	83.5	89.8
AS-GCN ^[11]	86.8	94.2
2s-AGCN ^[10]	88.5	95.1
JT-STGCN(2s)	85.8	93.8
JT-AGCN(2s)	89.3	95.5

为了探求时序特征与和点源特征的优劣性,本文以 ST-GCN^[1]为基准模型,在 X-View 的划分方法下,单独计算了

NTU-RGB+D60^[17]数据集中单类的准确率。时序散度流(T_s)和点源特征流(J_s)在类别“穿鞋”的上准确率分别为 93% 和

70%,在类别“检查时间”上的准确率分别为“69%”和“94%”。由此可以看出,时序散度特征与点源特征在不同的类别上各有优劣。如图4(a)所示,“穿鞋”这一行为,包含弯腰,伸手,提鞋子这些动作分支,人体动作比较复杂。时序散度特征对复杂动作有极强的表征能力,故在“穿鞋”这一行为上的识别率要远高于点源流特征;如图4(b)所示,“检查时间”这个动作仅仅是抬起手臂,动作较为单一。时序散度特征对单一的动作没有很强的表征能力,故对于“检查时间”这一行为,点源特征的识别率要高于时序散度特征。在结合上述两流的特征后,“穿鞋”和“检查时间”的识别率分别达到了95%和93%。可以看出,结合后的双流网络继承了时序散度流以及点源流的优势,弥补了各自的不足,整体上有着更好的性能。

通过上述实验,可以得出以下结论:时序散度特征对于人体行为中复杂的运动状态有着极强的表征能力,点源特征则更适用于运动状态单一的人体行为。两者各有优劣,同时也具有极强的互补关系。这也论证了上文2.3节中所提出的观点:时序特征与空间特征有着极强的互补关系。

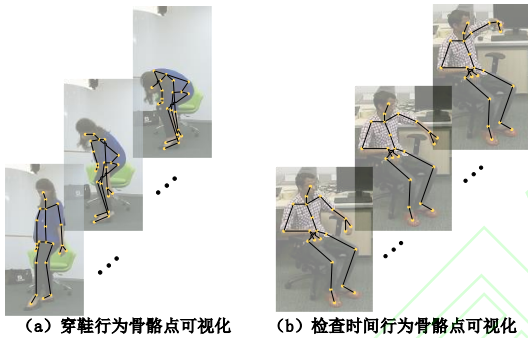


图4 不同人体行为的骨骼点可视化

Fig. 4 Visualization of skeletal points of different human behaviors

3.4 消融实验结果分析

为了验证散度特征注意力机制的有效性,本文做了相应的消融实验。如表3所示,Ts-no-A表示无注意力机制的时序散度特征。可以看出,在加入注意力机制后,实验性能都有显著的提升。由此验证了2.2节中所阐述的观点:在人体行为识别任务中,相比于人体上灵活性相对较弱的骨骼点(例如肩膀,跨部),手腕与脚踝这些骨骼点往往对人体行为识别任务起到更为关键的作用。

表3 关于注意力机制的消融实验结果

Tab. 3 Results of ablation experiments on the attention mechanism

网络模型	数据模型		准确率(%)	
	Ts-no-A	Ts	sub	view
ST-GCN ^[9]	√		82.1	90.5
		√	83.1	91.7
AGCN ^[11]	√		84.9	92.9
		√	86.2	93.8

为了对比Js和Ts的性能,本文选用ST-GCN以及AGCN为基准模型,特征流分别选用点源特征流Js和时序散度特征流Ts,在NTU-RGB+D^[19-20]数据集上采用单流网络做了消融实验,实验结果如表4-5所示。

从表4-5中可以得知,对于不同的基准模型,时序散度特征的性能也有所不同。在ST-GCN^[9]的基础上,相对于原始的点源数据而言,采用时序散度特征的建模方法可以达到更加

显著的效果;但是在AGCN^[10]上,采用轨迹向量建模,效果却不是那么明显。本文认为,由于AGCN^[10]中自适应卷积层的作用是通过初始的数据信息预测出有效的特征图。因此AGCN^[2]模型数据有一定的要求:初始数据必须对人体的空间特征有一定的表达能力。然而轨迹向量的建模方法会使数据丢失一部分的空间表达能力。故在AGCN^[2]的单流网络中,散度特征的性能没有很好的展现出来。

表4 消融实验在NTU-RGB+D 60数据集上的结果

Tab. 4 Results of ablation experiments on the NTU-RGB+D 60 dataset

网络模型	数据模型		准确率/%	
	Js	Ts(ours)	sub	view
ST-GCN ^[9]	√		81.5	88.3
		√	83.1	91.7
AGCN ^[11]	√		86.8	93.7
		√	86.2	93.8

表5 消融实验在NTU-RGB+D 120数据集上的结果

Tab. 5 Results of ablation experiments on the NTU-RGB+D 120 dataset

网络模型	数据模型		准确率/%	
	Js	Ts(ours)	sub	view
ST-GCN ^[9]	√		71.9	72.5
		√	74.7	76.5
AGCN ^[11]	√		79.5	81.2
		√	77.5	79.8

4 结语

在基于骨骼点的行为识别算法中,点源数据缺乏时序信息。为了解决上述问题,本文提出了基于骨骼点的时序散度特征,用来描述人体骨骼点在时间维度上的运动状态。接着为时序散度特征添加注意力机制,增加不同行为之间的类间方差,使神经网络更容易辨别不同的人体行为。最后构建时空双流网络模型,融合时序散度特征以及点源空间特征,并在NTU-RGB+D^[19-20]数据集上做了大量实验,双流模型的实验效果均超过所采用的基准模型。

从实验结果可以看出,时序散度特征可以弥补点源特征缺乏时间维度上运动表征的不足,能有效的与点源空间特征相结合,两者具有很强的互补关系。这也证实了本文方法的有效性。但由于人体骨架缺乏场景信息,该研究方向依然有很大的提升空间。可以考虑将场景信息与骨骼点将结合,并使用自适应卷积层过滤掉场景信息中的噪声,保留关键信息。本文在未来的行为识别算法研究中,将会在上述两个方面深入探索。

参考文献 (References)

[1] 朱煜, 赵江坤, 王逸宁, 等. 基于深度学习的人体行为识别算法综述[J]. 自动化学报, 2016, 42(6): 848-857. (ZHU Y, ZHAO J K, WANG Y N, et al. A review of human action recognition based on deep learning[J], Acta Automatica Sinica, 2016, 42(6):848-857.)

[2] FEICHTENHOFER C, FAN H, HE K, et al. SlowFast Networks for Video Recognition [C]// Proceedings of the 2019 IEEE International Conference on Computer Vision, Piscataway, NJ: IEEE, 2019:6202-6211.

[3] DU T, WANG H, et al. A closer look at spatiotemporal convolutions for action recognition [C]// Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition,

- Piscataway, NJ: IEEE, 2018:6450-6459.
- [4] 陆中秋, 侯振杰, 陈宸, 等. 基于深度图像与骨骼数据的行为识别[J]. 计算机应用, 2016(11):2979-2984. (LU Z Q, HOU Z J, CHEN C, et al. Action recognition based on depth images and skeleton data[J]. Journal of Computer Applications, 2016(11): 2979-2984.)
 - [5] 许艳, 侯振杰, 梁久祯, 等. 深度图像与骨骼数据的多特征融合人体行为识别[J]. 小型微型计算机系统, 2018, 39(8):1865-1870. (XU Y, HOU Z J, LIANG J Z, et al. Human action recognition with multi-feature fusion by depth image and skeleton data[J]. Journal of Chinese Mini-Micro Computer Systems, 2018, 39(8):1865-1870.)
 - [6] KAREN S, ANDREW Z. Two-stream convolutional networks for action recognition in videos[C]// Proceedings of the 2014 Advances in neural information processing systems. Piscataway, NJ: IEEE, 2014: 568-576.
 - [7] JOHANSSON G. Visual perception of biological motion and a model for its analysis[J]. Perception & psychophysics, 1973, 14(2): 201-211.
 - [8] REN B, LIU M Y, DING R W, et al. A survey on 3D skeleton-based action recognition using learning methods [EB/OL]. [2020-02-14]. <https://arxiv.org/pdf/2002.05907.pdf>.
 - [9] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition [EB/OL]. [2018-1-25]. <https://arxiv.org/pdf/1801.07455.pdf>.
 - [10] SHI L, ZHANG Y, CHENG J, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition[C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 12026-12035.
 - [11] LI M, CHEN S, CHEN X, et al. Actional-structural graph convolutional networks for skeleton-based action recognition[C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 3595-3603.
 - [12] SONG Y, ZHANG Z, WANG L. Richly activated graph convolutional network for action recognition with incomplete skeletons [C]// Proceedings of the 2019 IEEE International Conference on Image Processing. Piscataway, NJ: IEEE, 2019: 1-5.
 - [13] KIPF THOMAS N, WELING MAX. Semi-supervised classification with graph convolutional networks [EB/OL]. [2017-2-22]. <https://arxiv.org/pdf/1609.02907.pdf>.
 - [14] HU J, ZHENG W, LAI J, et al. Jointly learning heterogeneous features for RGB-D activity recognition [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE, 2015:5344-5352.
 - [15] WANG L, HUYNH DU Q, KONIUSZ PIOTR. A comparative review of recent kinect-based action recognition algorithms [J]. IEEE Transactions on Image Processing, 2019, 29:15-28.
 - [16] 盖赞, 荆国栋. 多尺度方法结合卷积神经网络的行为识别[J]. 计算机工程与应用, 2019, 55(2):100-103. (GE Y, JING G D. Human action recognition based on convolution neural network combined with multi-scale method[J]. Computer Engineering and Applications, 2019, 55(2):100-103.)
 - [17] 管珊珊, 张益农. 基于残差时空图卷积网络的3D人体行为识别[J]. 计算机应用与软件, 2020, 37(3):198-201. (Guan S S, Zhang Y N. 3D human behavior recognition based on residual spatio-temporal graph convolution network [J]. Computer Applications and Software, 2020, 37(3):198-201.)
 - [18] 万晓依. 基于时空结构关系的3D人体行为识别研究[D]. 苏州: 苏州大学, 2018:1-3. (WAN X Y. Research on 3D human action recognition base on spatio-temporal structure relationship [D]. Suzhou: Soochow University, 2018:1-3)
 - [19] SHAHROUDY AMIR, LIU JUN, NG TIAN-TSONG, et al. NTU RGB+ D: a large scale dataset for 3d human activity analysis[C]// Proceedings of the 2016 IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2016: 1010-1019.
 - [20] LIU J, AMIR S, MAURICIO P, et al. NTU RGB+D 120: a large-scale benchmark for 3d human activity understanding [C]// Proceedings of the 2019 IEEE Transactions on Pattern Analysis and Machine Intelligence, Piscataway, NJ: IEEE, 2019 42(10): 2684-2707.
 - [21] DUVENAUD DAVID K, DOUGAL M, JORGE I, et al. Convolutional networks on graphs for learning molecular fingerprints [C]// Proceedings of the 2015 Advances in neural information processing systems. Piscataway, NJ: IEEE, 2015: 2224-2232.
 - [22] MATHIAS N, MOHAMED A, KONSTANTIN K. Learning convolutional neural networks for graphs [C]// Proceedings of the International conference on machine learning. Piscataway, NJ: IEEE, 2016: 2014-2023.
 - [23] JOAN B, WOJCIECH Z, ARTHUR S, et al. Spectral networks and locally connected networks on graphs [EB/OL]. [2014-5-21]. <https://arxiv.org/pdf/1312.6203.pdf>.
 - [24] KE Q, MOHAMMED B, AN S, et al. A new representation of skeleton sequences for 3d action recognition [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE, 2017: 3288-3297.
 - [25] SHUMAN DAVID I, NARANG SUNIL K., FROSSARD PASCAL, et al. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains [J]. IEEE Signal Processing Magazine, 2013, 30(3):83-98.
 - [26] HU J, ZHENG W, MA LG, et al. Early action prediction by soft regression [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41:2568-2583.
 - [27] HU J, ZHENG W, LAI J, et al. Jointly learning heterogeneous features for RGB-D activity recognition [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015:5344-5352.
 - [28] LIU J, AMIR S, XU D, et al. Spatio-temporal lstm with trust gates for 3D human action recognition [C]// Proceedings of the 2016 European Conference on Computer Vision. Cham: Springer, 2016: 816-833.
 - [29] LIU J, AMIR S, XU D, et al. Skeleton-based action recognition using spatio-temporal lstm network with trust gates [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(12):3007-3021.
 - [30] LIU J, WANG G, HU P, et al. Global Context-aware attention LSTM networks for 3D action recognition [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 1647-1656.
 - [31] KE Q, MOHAMMED B, AN S, et al. A new representation of skeleton sequences for 3d action recognition [C]// Proceedings of

- the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE, 2017:3288-3297.
- [32] LIU J, AMIR S, WANG G, et al. Skeleton-based online action prediction using scale selection network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,42:1453-1467.
- [33] LIU M, LIU H, CHEN C. Enhanced skeleton visualization for view invariant human action recognition[J]. Pattern Recognition, 2017, 68:346-362.
- [34] LIU J, WANG G, Duan L, et al. Skeleton-based human action recognition with global context-aware attention LSTM networks [J]. IEEE Transactions on Image Processing, 2018, 27(4):1586-1599.
- [35] KE Q, MOHAMMED B, AN S, et al. Learning clip representations for skeleton-based 3d action recognition [J]. IEEE Transactions on Image Processing, 2018, 27(6):2842-2855.
- [36] LIU M, YUAN J. Recognizing human actions as the evolution of pose estimation maps[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018:1159-1168.
- [37] CEATANO C, JESSICA S, BRÉMOND F, et al. SkeleMotion: a new representation of skeleton joint sequences based on motion information for 3D action recognition[C]// Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway, NJ: IEEE, 2019:1-8.
- [38] YANG Z, LI Y, YANG J, et al. Action recognition with spatio-temporal visual attention on skeleton image sequences [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(8):2405-2415.
- [39] ADAM P, SAM G, SOUMITH C, et al. Automatic differentiation in PyTorch [EB/OL]. [2017-10-29] <https://openreview.net/pdf?id=BJJsrmfCZ>.
- [40] RAVITEJA V, FELIPE A, RAMA C. Human action recognition by representing 3d skeletons as points in a lie group [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 588-595.
- [41] ZHENG W, LI L, ZHANG Z, et al. Relational network for skeleton-based action recognition [C]// Proceedings of the 2018 IEEE International Conference on Multimedia and Expo, Piscataway, NJ: IEEE, 2018:826-831.
- [42] Li S, Li W, Cook C, et al. Independently recurrent neural network (indrnn) : Building a longer and deeper rnn [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 5457-5466.
- [43] LI C, ZHONG A, XIE D, et al. Skeleton-based action recognition with convolutional neural networks [C]//Proceedings of the 2017 IEEE International Conference on Multimedia and Expo Workshops, Piscataway, NJ: IEEE, 2017:597-600.
- [44] LI B, DAI Y, CHENG X, et al. Skeleton based action recognition using translation-scale invariant image mapping and multi-scale deep CNN [C]// Proceedings of the 2017 IEEE International Conference on Multimedia and Expo Workshops, Piscataway, NJ: IEEE, 2017:601 - 604
- [45] TANG Y, TIAN Y, LU J, et al. Deep progressive reinforcement learning for skeleton-based action recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018:2323-5332.
- This work is partially supported by the National Natural Science Foundation of China (61806150), Hubei Provincial Department of Science and Technology (2018CFB195), Hubei Provincial Department of Education Science and Technology Research Program Young Talent Project (Q20181104), Foundation of Hubei key laboratory of intelligent information processing and real-time industrial system (znxx2018QN09), National Defense Advanced Research Foundation of Wuhan University of Science and Technology (GF201814).
- TIAN ZHIQIANG**, born in 1996, M. S. candidate. His research interests include Computer Vision, Machine Learning.
- DENG CHUNHAU**, born in 1984, Ph. D., associate professor. His research interests include Computer Vision, Machine Learning.
- ZHANG JUNWEN**, born in 1997, M. S. candidate. Her research interests include Computer Vision, Machine Learning.