



中国图形学报

2021
12
VOL.26

ISSN1006-8961
CN11-3758/TB





第26卷第12期（总第308期）
2021年12月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

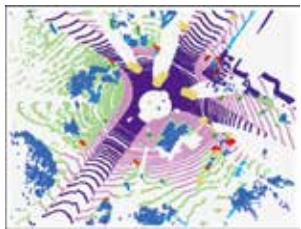
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

环视鱼眼图像处理深度学习
研究进展(第2778页)结合空间深度卷积和残差
的大尺度点云场景分割
(第2848页)曼哈顿世界环境下消失点非线
性优化估计算法(第2931页)

综述

深度对抗视觉生成综述

谭明奎, 许守恺, 张书海, 陈奇 2751

人机交互中的智能体情感计算研究

刘婷婷, 刘箴, 柴艳杰, 王瑾, 王媛怡 2767

环视鱼眼图像处理深度学习研究进展

包俊, 刘宏哲, 褚文博 2778

图像处理和编码

图像分解与色彩先验下的多曝光图像融合

李嫛源, 王琴, 朱智勤, 齐观秋 2800

双尺度分解和显著性分析相结合的红外与可见光图像融合

霍星, 邹韵, 陈影, 檀结庆 2813

轻量级注意力特征选择循环网络的超分辨率重建

徐雯捷, 宋慧慧, 袁晓彤, 刘青山 2826

区域级通道注意力融合高频损失的图像超分辨率重建

周波, 李成华, 陈伟 2836

图像分析和识别

结合空间深度卷积和残差的大尺度点云场景分割

刘盛, 黄圣跃, 程豪豪, 沈家瑜, 陈胜勇 2848

细粒度图像分类的互补注意力方法

赵勋, 王家宝, 李阳, 王亚鹏, 苗壮 2860

联合分块谱带优选和深度特征的高光谱人脸识别

谢志华, 李毅, 牛杰一 2870

时空双仿射微分不变量及骨架动作识别

李琪, 墨瀚林, 赵婧涵, 郝宏翔, 李华 2879

多尺度分割先验迭代重加权低秩恢复显著性检测

张荣国, 郑维佳, 赵建, 胡静, 刘小君 2892

融合环境特征与改进YOLOv4的安全帽佩戴检测

葛青青, 张智杰, 袁琰, 李秀梅, 孙军梅 2904

图像理解和计算机视觉

关节轴角先验对3维人体重建结果的影响

姚砾, 张幼安, 张梦雪, 万燕 2918

曼哈顿世界环境下消失点非线性优化估计算法

韩冰心, 卢惠民, 于清华, 张礼廉 2931

双目机器视觉及RetinaNet模型的路侧行人感知定位

连丽容, 罗文婷, 秦勇, 李林 2941

遥感图像处理

遥感图像中油罐目标精确定位与参数提取

江哈, 张月婷, 郭嘉逸, 赵鑫, 李芳芳, 黄丽佳, 胡玉新, 雷斌, 丁赤飏 2953

参数合成空间变换网络的遥感图像一致性配准

陈颖, 张祺, 李文举, 石艳娇, 陈磊 2964

改进Mask R-CNN模型的海洋锋检测

徐慧芳, 黄冬梅, 贺琪, 杜艳玲, 覃学标 2981

噪声鲁棒的轻量级深度遥感场景图像分类检索

王亚鹏, 李阳, 王家宝, 赵勋, 苗壮 2991

改进U-Net型网络的遥感图像道路提取

杨佳林, 郭学俊, 陈泽华 3005

弱监督深度语义分割网络的多源遥感影像水体检测

李鑫伟, 李彦胜, 张永军 3015

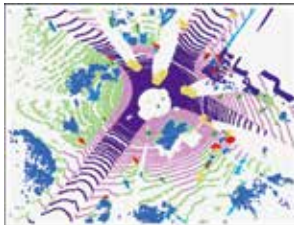
总目次 I

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Research progress of fisheye image processing based on deep learning(P2778)



A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation(P2848)



Vanishing point estimation based on non-linear optimization in manhattan world environments (P2931)

Review

A review on deep adversarial visual generation	
Tan Mingkui, Xu Shoukai, Zhang Shuhai, Chen Qi	2751
Agent affective computing in human-computer interaction	
Liu Tingting, Liu Zhen, Chai Yanjie, Wang Jin, Wang Yuanyi	2767
Research progress of fisheye image processing based on deep learning	
Bao Jun, Liu Hongzhe, Chu Wenbo	2778

Image Processing and Coding

Multi-exposure image fusion based on image decomposition and color prior	
Li Yuanyuan, Wang Qin, Zhu Zhiqin, Qi Guanqiu	2800
Dual-scale decomposition and saliency analysis based infrared and visible image fusion	
Huo Xing, Zou Yun, Chen Ying, Tan Jieqing	2813
Lightweight attention feature selection recursive network for super-resolution	
Xu Wenjie, Song Huihui, Yuan Xiaotong, Liu Qingshan	2826
Region-level channel attention for single image super-resolution combining high frequency loss	
Zhou Bo, Li Chenghua, Chen Wei	2836

Image Analysis and Recognition

A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation	
Liu Sheng, Huang Shengyue, Cheng Haohao, Shen Jiayu, Chen Shengyong	2848
Complemented attention method for fine-grained image classification	
Zhao Xun, Wang Jiabao, Li Yang, Wang Yapeng, Miao Zhuang	2860
Hyperspectral face recognition based on partition bands optimal selection and deep features	
Xie Zhihua, Li Yi, Niu Jieyi	2870
Spatio-temporal dual affine differential invariants for skeleton-based action recognition	
Li Qi, Mo Hanlin, Zhao Jinghan, Hao Hongxiang, Li Hua	2879
Iterative reweighted low-rank recovery salient object detection with multiscale segmentation prior	
Zhang Rongguo, Zheng Weijia, Zhao Jian, Hu Jing, Liu Xiaojun	2892
Safety helmet wearing detection method of fusing environmental features and improved YOLOv4	
Ge Qingqing, Zhang Zhijie, Yuan Long, Li Xiumei, Sun Junmei	2904

Image Understanding and Computer Vision

The impact of joint axis angle prior on the results of 3D human body reconstruction	
Yao Li, Zhang Youan, Zhang Mengxue, Wan Yan	2918
Vanishing point estimation based on non-linear optimization in manhattan world environments	
Han Bingxin, Lu Huimin, Yu Qinghua, Zhang Lilian	2931
Roadside pedestrian detection and location based on binocular machine vision and RetinaNet	
Lian Lirong, Luo Wenting, Qin Yong, Li Lin	2941

Remote Sensing Image Processing

Accurate localization and parameter extraction of oil tank in remote sensing images	
Jiang Han, Zhang Yueting, Guo Jiayi, Zhao Xin, Li Fangfang, Huang Lijia, Hu Yuxin, Lei Bin, Ding Chibiao	2953
Consistent registration of remote sensing images in parametric synthesized spatial transformation network	
Chen Ying, Zhang Qi, Li Wenju, Shi Yanjiao, Chen Lei	2964
Ocean front detection method based on improved Mask R-CNN	
Xu Huifang, Huang Dongmei, He Qi, Du Yanling, Qin Xuebiao	2981
A robust lightweight deep neural network for remote sensing scene image classification and retrieval under label noise	
Wang Yapeng, Li Yang, Wang Jiabao, Zhao Xun, Miao Zhuang	2991
Road extraction method from remote sensing images based on improved U-Net network	
Yang Jialin, Guo Xuejun, Chen Zehua	3005
Weakly supervised deep semantic segmentation network for water body extraction based on multi-source remote sensing imagery	
Li Xinwei, Li Yansheng, Zhang Yongjun	3015

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2021)12-2879-13

论文引用格式: Li Q, Mo H L, Zhao J H, Hao H X and Li H. 2021. Spatio-temporal dual affine differential invariants for skeleton-based action recognition. Journal of Image and Graphics, 26(12): 2879-2891 (李琪, 墨瀚林, 赵婧涵, 郝宏翔, 李华. 2021. 时空双仿射微分不变量及骨架动作识别. 中国图象图形学报, 26(12): 2879-2891) [DOI: 10.11834/jig.200453]

时空双仿射微分不变量及骨架动作识别

李琪*, 墨瀚林, 赵婧涵, 郝宏翔, 李华

1. 中国科学院计算技术研究所智能信息处理重点实验室, 北京 100190; 2. 中国科学院大学, 北京 100049

摘要: **目的** 人体骨架的动态变化对于动作识别具有重要意义。从关节轨迹的角度出发, 部分对动作类别判定具有价值的关节轨迹传达了最重要的信息。在同一动作的每次尝试中, 相应关节的轨迹一般具有相似的基本形状, 但其具体形式会受到一定的畸变影响。基于对畸变因素的分析, 将人体运动中关节轨迹的常见变换建模为时空双仿射变换。**方法** 首先用一个统一的表达式以内外变换的形式将时空双仿射变换进行描述。基于变换前后轨迹曲线的微分关系推导设计了双仿射微分不变量, 用于描述关节轨迹的局部属性。基于微分不变量和关节坐标在数据结构上的同构特点, 提出了一种通道增强方法, 使用微分不变量将输入数据沿通道维度扩展后, 输入神经网络进行训练与评估, 用于提高神经网络的泛化能力。**结果** 实验在两个大型动作识别数据集 NTU (Nanyang Technological University) RGB + D (NTU 60) 和 NTU RGB + D 120 (NTU 120) 上与若干最新方法及两种基线方法进行比较, 在两种实验设置 (跨参与者识别与跨视角识别) 中均取得了明显的改进结果。相比于使用原始数据的时空图神经卷积网络 (spatio-temporal graph convolutional networks, ST-GCN), 在 NTU 60 数据集中, 跨参与者与跨视角的识别准确率分别提高了 1.9% 和 3.0%; 在 NTU 120 数据集中, 跨参与者与跨环境的识别准确率分别提高了 5.6% 和 4.5%。同时对比于数据增强, 基于不变特征的通道增强方法在两种实验设置下都能有明显改善, 更为有效地提升了网络的泛化能力。**结论** 本文提出的不变特征与通道增强, 直观有效地综合了传统特征和深度学习的优点, 有效提高了骨架动作识别的准确性, 改善了神经网络的泛化能力。

关键词: 运动分析; 骨架动作识别; 时空双仿射变换; 微分不变量; 通道增强; 泛化能力

Spatio-temporal dual affine differential invariants for skeleton-based action recognition

Li Qi*, Mo Hanlin, Zhao Jinghan, Hao Hongxiang, Li Hua

1. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: **Objective** Skeleton-based action recognition has been concerned in recent years, as the dynamics of human skeletons has significant information for the task of action recognition. The action of human skeletons can be seen as time series of human poses, or the combination of human joint trajectories. The trajectory of important joints indicating the action class

收稿日期: 2020-08-21; 修回日期: 2020-11-20; 预印本日期: 2020-11-27

* 通信作者: 李琪 neilyoyoyo@163.com

基金项目: 国家重点研发计划资助 (2019YFF0301801, 2017YFB1002703); 国家重点基础研究规划项目 (2015CB554507); 国家自然科学基金项目 (61379082)

Supported by: National Key R&D Program of China (2019YFF0301801, 2017YFB1002703); National Key Basic Research Program of China (2015CB554507); National Natural Science Foundation of China (61379082)

has conveyed the most significant information among all the human joints. The trajectories of these joints have been subjected to some distortions when performing the same action under different attempts. In this case, two similar trajectories of corresponding joints should share a basic shape. However, these two trajectories have appeared in diverse kinds of distortions due to individual factors. These distortions have been caused by spatial and temporal factors. Spatial factors have included the change of viewpoints, different skeleton sizes and action amplitudes, while temporal factors indicate time scaling along the time series, denoting the order and speed of performing specific action. All the spatial factors can be modeled by the affine transformation in 3D space, whereas the uniform time scaling has been commonly discussed case, which can be seen as affine transformation in 1D space. These two kinds of distortions as the spatio-temporal dual affine transformation have been combined. A novel invariant feature under these distortions has been proposed and utilized for facilitating skeleton-based action recognition. A kind of feature invariant based on the spatio-temporal affine transformation has aided the identification of similar trajectories to be beneficial for action recognition. **Method** A general method for constructing spatio-temporal dual affine differential invariant (STDADI) has been proposed. The rational polynomial of derivatives of joint trajectories to obtain the invariants has been utilized in detail via eliminating the transformation parameters effectively. Robust, coordinate-system-independent feature has calculated directly from the 3D coordinates. Bounding the degree of polynomial and the order of derivatives, we generate 8 independent STDADIs and combine them as an invariant vector at each moment for each human joint. Moreover, an intuitive and effective method called channel augmentation has been proposed to extend input data with STDADI along the channel dimension for training and evaluation. Specifically, the coordinate vector and the STDADI vector at each joint for each frame have been concatenated. Channel augmentation has introduced invariant information into input data without changing the inner structure of neural networks. The spatio-temporal graph convolutional networks (ST-GCN) as the basic network have been used. The skeleton data modeling as a graph structure has involved spatial and temporal connections between human joints simultaneously. Particularly, it has exploited local pattern and correlation from human skeletons. In other words, the importance of joints along the action sequence has been expressed as the weights of human joints in the spatio-temporal graph. This is in line with our STDADI, because both of them focus on describing joint dynamics, and our features further provide an invariant expression which is not affected by the distortions.

Result The synthetic data has been examined to verify the effectiveness of STDADI as well as the large-scale action recognition dataset. First, 3D spiral line and selected joint trajectory based on NTU-RGB + D applied with random transformation parameters has shown that STDADI is invariant under the spatio-temporal affine transformations. Next, the effectiveness of the proposed feature and method has been validated on the large-scale action recognition dataset NTU (Nanyang Technological University) RGB + D (NTU 60) and its extended version NTU-RGB + D 120 (NTU 120), which is currently the largest dataset with 3D joint annotations captured in a constrained indoor environment, and perform some detailed study to examine the contributions of STDADI. A data augmentation technique as well as the original ST-GCN have been as the baseline methods. The data augmentation technique has involved rotation, scaling and shear transformations of 3D skeletons. The same training strategy and hyper-parameters as the original ST-GCN have been used. ST-GCN + channel augmentation has performed well. Compared with the ST-GCN using raw data, in NTU 60, the cross-subject and cross-view recognition accuracy has been increased by 1.9% and 3.0%, respectively; in NTU 120, the cross-subject and cross-setup recognition accuracy has increased by 5.6% and 4.5% respectively. As it is mainly consisted of 3D geometric transformations, the accuracy in cross-view recognition has been much improved but contributes little to the cross-subject setting for data augmentation. The spatio-temporal dual affine transformation assumption has been validated on both evaluation criteria. **Conclusion** A general method for constructing spatio-temporal dual affine differential invariant (STDADI) has been proposed. The effectiveness of this invariant feature using a channel augmentation technique has been proved on the large-scale action recognition dataset NTU-RGB + D and NTU-RGB + D 120. The combination of hand-crafted features and data-driven methods has improved the accuracy and generalization well.

Key words: motion analysis; skeleton-based action recognition; spatio-temporal dual affine transformation; differential invariant features; channel augmentation; generalization ability

0 引言

基于人体骨架的动作识别受到广泛关注与研究, 主要是因为人体骨架的动态变化已具备了判别多数动作类型的基本信息。相比于其他数据形式(如视频、深度图和光流图等), 骨架数据同时具有数据量少和信息密度高的优点(Wang 等, 2018)。骨架的动态变化既可以视做每帧人体姿态的时间序列, 又可以视为所有关节轨迹的组合。从后者的角度来看, 部分对动作类别判定具有价值的关节轨迹传达了最重要的信息。然而, 在不同的尝试下进行相同的动作时, 关节轨迹一般会受到一定的畸变影响。本文将这种畸变总结为时空双仿射变换, 提出了一种微分不变特征, 并通过通道增强的方法将该特征引入神经网络中以提高网络模型的泛化能力。

在同一动作的每次尝试中, 相应关节的关节轨迹一般具有相似的基本形状。然而, 由于各种因素影响, 同一人的关节轨迹一般也会产生一定变化, 从而对动作的正确识别产生影响。这些因素可概括为空间与时间两类。空间因素主要包括: 拍摄视角的变化导致描述动作的坐标系改变、不同参与者的身体形状大小导致骨架尺度变化以及不同参与者动作幅度变化导致轨迹尺度差异(Wu 和 Li, 2009; Wang 和 Wang, 2017)。时间因素则表示沿时间维度的尺度缩放, 即不同参与者执行同一动作, 或者同一人每次重复动作的快慢有所不同(Esling 和 Agon, 2012; He 等, 2014)。为处理这些差异与变化, 将空间因素建模为3维空间中的仿射变换。时间维度的尺度变换相对复杂, 一般是线性简化为均匀的时间尺度缩放, 将其建模为时间维度1维空间中的仿射变换。将这两种仿射变换组合, 即为时空双仿射变换。

如图1所示, 以一条已知曲线方程的3维螺旋线和选自NTU RGB + D(Shahroudy 等, 2016)的一条抛掷动作中的右肩轨迹为例, 演示时空双仿射下曲线的变化。从空间角度, 轨迹形状会受到旋转、平移和缩放等的作用。从时间角度, 轨迹上采样点的疏密发生变化, 体现为关节运动的快慢。在人体运动过程中, 关节轨迹在时空方面的畸变是广泛存在的, 而时空双仿射变换是一种有效的建模手段。基于以上分析, 提出一种时空双仿射变换下的微分不

变量(spatio-temporal dual affine differential invariant, STDADI)。对于每个时刻的每个关节点, 均有一个8维的不变特征向量用于描述关节轨迹的局部属性, 并在双仿射变换下保持不变。

对基于人体骨架的动作识别算法, 研究者倾向于探索数据驱动方法在其中的潜力, 典型的即为各种神经网络算法。当考虑提高不同变换下神经网络的泛化能力时, 常见的做法是数据增强。但是, 额外的数据预处理会生成更多样本, 并在训练阶段花费更多时间。特别地, 针对人体骨架数据的数据增强(Wang 和 Wang, 2017)一般仅考虑空间方面因素, 而针对时间变换, 则需要对骨架数据进行插值处理, 这种操作不仅复杂, 而且无法保证生成数据的可靠性。对此, 本文提出一种直观有效的方法, 使用STDADI将输入数据沿通道维度扩展后, 输入神经网络进行训练与评估, 并将这种做法称为通道增强。实验表明, 基于STDADI的通道增强方法可以实现更好的泛化能力。

1 相关工作

1.1 基于人体骨架的动作识别

在深度学习兴起之前, 部分研究工作使用手工特征解决基于人体骨架的动作识别。Wang 等人(2012)提出使用关节之间的相对位置作为运动特征。Hussein 等人(2013)利用关节轨迹的协方差矩阵进行运动描述。Vemulapalli 等人(2014)利用关节位置的旋转和平移关系来描述骨架的动态变化。然而, 这类基于手工特征方法的性能是有限的, 主要是因为手工设计的特征一般不能覆盖影响识别的各项主要因素, 从而使识别的结果产生一定的偏差。

得益于深度学习的成功, 数据驱动方法取得了性能上的优势。这些方法可以根据神经网络的类型划分为基于递归神经网络、卷积神经网络和图神经网络的方法。基于递归神经网络的方法将人体关节坐标的向量序列作为网络输入并按时序递归预测动作标签(Veeriah 等, 2015; Du 等, 2015; Liu 等, 2016; Shahroudy 等, 2016; Zhang 等, 2017; Liu 等, 2017b, c; Wang 和 Wang, 2017)。基于卷积神经网络的方法将骨架数据表示为伪图像输入常规的2维卷积神经网络(Wang 等, 2016; Liu 等, 2017a; Liu 等, 2017d; Li 等, 2017a; Li 等, 2017b; Liu 和 Yuan, 2018; Cao 等,

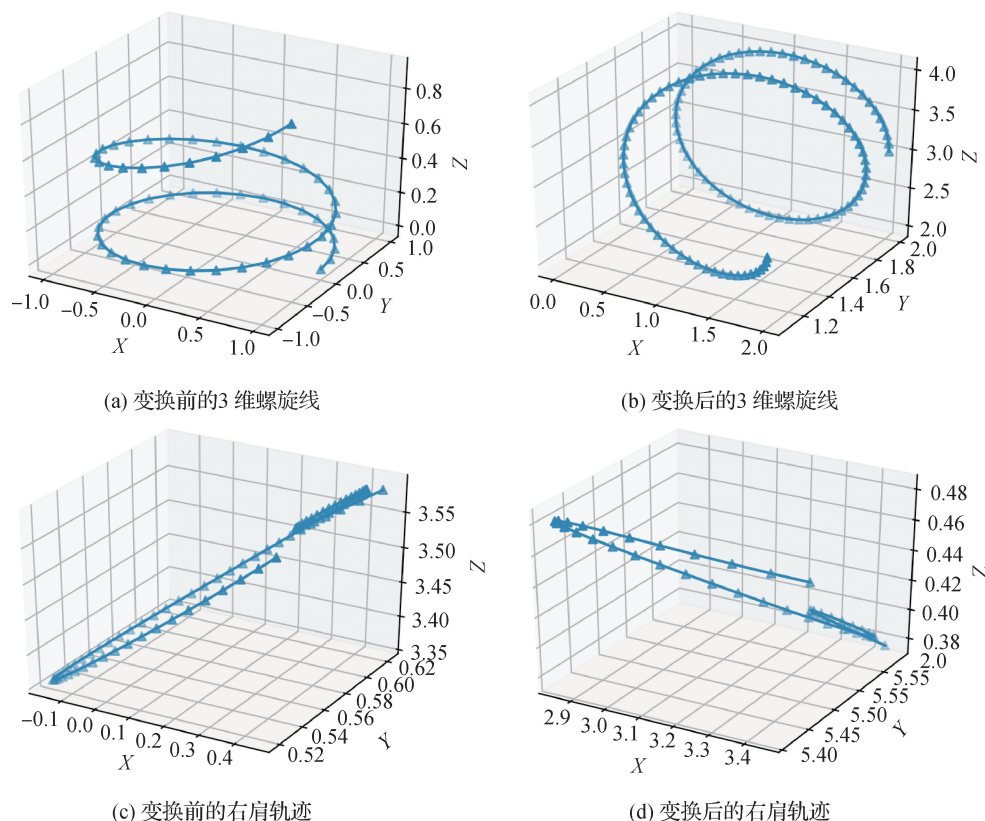


图1 时空双仿射变换前后的3维螺旋线和抛掷动作右肩轨迹

Fig. 1 The 3D spiral line and trajectory of right shoulder of a throw action within spatio-temporal dual affine transformation
((a) 3D spiral line before transformation; (b) 3D spiral line after transformation; (c) right shoulder trajectory before transformation; (d) right shoulder trajectory after transformation)

2019;Pham 等,2019)或坐标向量序列输入时序卷积神经网络(Ke 等,2017,2018;Kim 和 Reiter,2017)。与前两种方法相比,基于图卷积网络的方法是依据人体关节之间的自然连接关系建模的,在表述上更为自然。Yan 等人(2018)首先提出时空图卷积网络(spatio-temporal graph convolutional networks, STGCN)用于解决动作识别问题,后续工作进一步在图的连接方式上进行调整,在识别准确率上得到了有效的改进(Li 等,2019a;Li 等,2019b;Shi 等,2019a,b;Si 等,2019)。

1.2 骨架数据的不变特征

从变换角度理解动作识别问题时,研究者通常讨论的是空间变换,主要考虑拍摄视角的变化。Müller 等人(2005)使用1组与4个关节相关的布尔特征来描述它们的相对位置,这一特征相对于骨架的位置、方向与尺寸是不变的。Vemulapalli 等人(2014)利用关节之间的刚体变换描述骨架,这种变换具有几何不变性。Shao 和 Li(2015)使用积分不

变特征作为关节轨迹的局部描述。Boulahia 等人(2016)整合了一组受手写识别启发的特征,在其中使用了相似变换下的矩不变量(Sadjadi 和 Hall,1980)。

时间尺度缩放作为沿时间维度的一种变换,以往的基于人体骨架的动作识别工作较少涉及。部分工作(Wu 和 Li,2009;Ke 等,2018;Anirudh 等,2017;Kacem 等,2020)使用动态时间规整(dynamic time warping, DTW)技术进行时间对准或轨迹匹配。王扬扬等人(2013)使用图像序列的时空超兴趣点描述人体动作。郑潇等人(2018)利用2维姿态时空特征描述运动。丁重阳等人(2020)提出时空权重姿态运动特征。钟秋波等人(2020)在图卷积网络基础上提出时空域融合网络。以上方法一般仅利用时间维度信息,而未从变换角度考虑。Esling 和 Agon(2016)明确定义了时间的尺度缩放,并将其分类为均匀的或动态的。本文遵循均匀尺度缩放的定义,并将其建模为时间仿射变换。

在深度学习领域,数据增强是一种用于提高泛

化能力的常用方法。但是,这种方法在训练过程中相对耗时。Wang 和 Wang (2017) 提出了一种适用于骨架数据 3 维坐标的数据增强方法,包括空间的旋转、缩放与错切变换,并在其提出的双流递归神经网络中证明了有效性。针对时间尺度缩放的数据增强是很困难的,一般需要对离散的姿态序列进行插值与采样,而插值方法则与动作类型和关节种类密切相关,即使完成也很难保证生成数据的可靠性。实验证明,基于 STDADI 的通道增强方法可以有效提高神经网络的泛化能力,这一提升兼顾了时空两方面的变换。

2 时空双仿射微分不变量与通道增强

2.1 时空双仿射变换

一般地,3 维人体关节轨迹可以表示为空间参数曲线 $f(t)$,以时间 t 作为参数,即

$$f(t) = (f_x(t), f_y(t), f_z(t))^T \quad (1)$$

式中, x, y, z 分别表示 3 维坐标的 x, y, z 分量。

若以 $f(t)$ 和 $g(u)$ 分别表示变换前后的两条曲线,则可以将时空双仿射变换的定义表示为

$$g(u) = Af(t) + B, \quad u = \frac{1}{c}(t - d) \quad (2)$$

式中,前式为外变换,以空间仿射变换表示,后式为内变换,以时间仿射变换表示。矩阵 A 和向量 B 为空间仿射变换的参数。矩阵 A 控制坐标的旋转、缩放与错切,是一个满秩矩阵;向量 B 控制平移。标量 c 和 d 为时间仿射变换的参数。 c 表示尺度缩放,代表动作的快慢; d 表示相位平移,代表不同的开始时间。通过内外变换的表达方式,可以将时间与空间的仿射变换结合在一起,用一个统一的关系式进行表达。

2.2 双仿射微分不变量

基于变换前后关节轨迹的微分关系推导不变特征,具体地,基于式(2),可以推导出变换前后的关节轨迹的 1 阶微分之间的关系。即

$$\frac{dg}{du} = \frac{dg}{df} \times \frac{df}{dt} \times \frac{dt}{du} = A \times \frac{df}{dt} \times c = cA \frac{df}{dt} \quad (3)$$

类似地,可以基于链式法则获取任意阶导数之间的关系,即

$$g^i = c^i A f^i \quad (4)$$

式中, i 表示微分的阶数。

特别地,当阶数 i 设为 0 时,式(4)等价于没有平移向量 B 的外变换。可以通过对变换前后的轨迹减去均值以消除平移向量的影响,即

$$\hat{g} = g - \bar{g} = A(f - \bar{f}) = c^0 A \hat{f} \quad (5)$$

式中, $\hat{g}(\hat{f})$ 表示关节坐标减去均值 $\bar{g}(\bar{f})$ 后的坐标向量。通过这种处理,在式(4)中,上标 i 可以扩展为非负整数。

基于式(4)中的关系,可以以曲线的不同阶导数作为列向量构造矩阵,并得到

$$M_g^{ijk} = (g^i, g^j, g^k) = (c^i A f^i, c^j A f^j, c^k A f^k) = A(c^i f^i, c^j f^j, c^k f^k) \quad (6)$$

式中,矩阵 M 的上标 i, j, k 为互不相同的非负整数,表示构造过程中所用到的导数阶数,下标代表基于对应曲线进行构造。

矩阵 M 的行列式即为双仿射变换下的相对不变量,对变换前后的轨迹曲线计算该值时,仅差一个与变换参数 c 和 A 有关的常数,具体为

$$\|M_g^{ijk}\| = c^{i+j+k} \|A\| \| (f^i, f^j, f^k) \| = c^{i+j+k} \|A\| \|M_f^{ijk}\| \quad (7)$$

进一步地,可以通过构造矩阵 M 行列式的有理式消去变换参数的影响,具体为

$$\frac{\prod_{\sigma=1}^N \|M_g^{i_{\sigma} j_{\sigma} k_{\sigma}}\|}{\prod_{\sigma=1}^N \|M_g^{l_{\sigma} m_{\sigma} n_{\sigma}}\|} = \frac{\prod_{\sigma=1}^N \|M_f^{i_{\sigma} j_{\sigma} k_{\sigma}}\|}{\prod_{\sigma=1}^N \|M_f^{l_{\sigma} m_{\sigma} n_{\sigma}}\|} \quad (8)$$

由于等式两边形式相同的构造式对变换前后的曲线保持不变,因此可以去掉下标,构造完成微分不变特征(STDADI),具体为

$$\frac{\prod_{\sigma=1}^N \|M^{i_{\sigma} j_{\sigma} k_{\sigma}}\|}{\prod_{\sigma=1}^N \|M^{l_{\sigma} m_{\sigma} n_{\sigma}}\|} \quad (9)$$

式中,以曲线不同导数阶数组合而成的矩阵 M 的行列式为基本单元,分子和分母分别表示基本单元的连乘式,分子分母的连乘次数 N 应该相等,以保证消去矩阵 A 的行列式值。这里将 N 称为不变特征的次数。为了保证有效消去参数 c ,需要对阶数组合进行约束,具体为

$$\sum_{\sigma=1}^N (i_{\sigma} + j_{\sigma} + k_{\sigma}) = \sum_{\sigma=1}^N (l_{\sigma} + m_{\sigma} + n_{\sigma}) \quad (10)$$

为了保证每一个行列式的值不为 0,还需进行一些约束,具体为

$$i_{\sigma} \neq j_{\sigma} \neq k_{\sigma}, l_{\sigma} \neq m_{\sigma} \neq n_{\sigma}, \forall \sigma \quad (11)$$

在构造不变量时,使用导数的最高阶数称为不变量的阶数。设置不变特征的最高次数和阶数分别为 2 和 4,这样即可基于构造方法组合获取共 55 个不变量,从中基于函数独立的分析方法 (Brown, 1935) 筛选出 8 个不变量,它们之间的相关性较低,因此可以作为更有效的不变量组合。具体为

$$\begin{aligned} & \frac{\|M^{023}\|}{\|M^{014}\|}, \frac{\|M^{123}\|}{\|M^{024}\|}, \frac{\|M^{034}\|}{\|M^{124}\|}, \frac{\|M^{012}\| \|M^{023}\|}{\|M^{013}\|^2}, \\ & \frac{\|M^{013}\| \|M^{123}\|}{\|M^{014}\|^2}, \frac{\|M^{023}\| \|M^{124}\|}{\|M^{123}\|^2}, \\ & \frac{\|M^{123}\| \|M^{134}\|}{\|M^{124}\|^2}, \frac{\|M^{124}\| \|M^{234}\|}{\|M^{134}\|^2} \end{aligned} \quad (12)$$

在计算时,由于需要估计关节轨迹的 4 阶导数,首先使用 5 阶 B 样条曲线拟合轨迹,并在此基础上进行估算。在获取关节轨迹的各阶导数之后,基于式(9)的定义和式(12)的上标组合计算 8 个微分不变特征,并将其组合为向量形式。在每个时刻的每个关节点,都有一个 8 维的不变特征向量。

2.3 通道增强

由于在每个时刻的每个关节点均有一个 8 维的不变特征向量,与关节点坐标在张量维度上具有同构的特点,因此,沿通道方向将坐标向量与不变特征向量串联,并作为神经网络的输入,即为通道增强方法,如图 2 所示。在串联前,对不变特征向量施加一个双曲正切函数将不变特征的数值范围限制在 $[-1, 1]$ 内。这样既可以保证其数值范围与坐标向量一致,又可以抑制某些异常值的影响。通道增强方法通过在输入数据中引入不变信息,而无需对神经网络的结构进行修改。

在实验中,选择使用时空图卷积网络(ST-GCN)

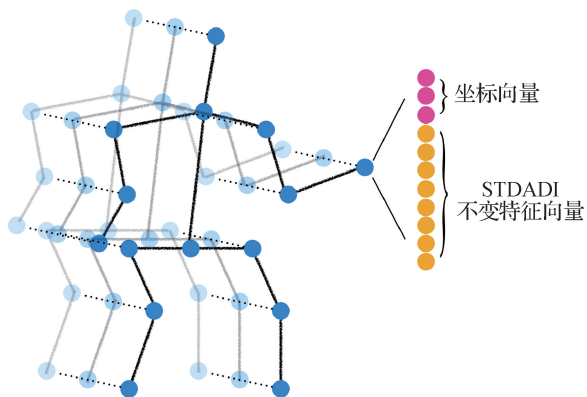


图2 人体骨架数据在时空图中的通道增强

Fig. 2 Illustration of channel augmentation for the spatio-temporal graph of human skeletons

作为基础网络模型,将人体骨架的动作序列建模为时空图,表述骨架数据更为自然,可以帮助发现对分类动作有重要意义的关节轨迹,并以时空图中的权重进行显示。

3 实验结果与分析

为验证时空双仿射微分不变量的有效性,进行了以下两种实验。首先进行仿真实验,通过对两种曲线轨迹施加时空双仿射变换后计算微分不变量以检验特征的不变性与区分性。在两个大型的动作识别数据集 NTU RGB + D (NTU 60) (Shahroudy 等, 2016) 和 NTU RGB + D 120 (NTU 120) (Liu 等, 2020) 上验证不变特征和通道增强的有效性。除了基础的 ST-GCN,还与 Wang 和 Wang (2017) 中提出的人体骨架数据增强方法作为基线进行对比。所有实验均使用与 Yan 等人 (2018) 方法一致的训练策略和超参数设置。

3.1 仿真实验

以图 1 所示的两种曲线轨迹(3 维螺旋线和抛掷动作中的右肩轨迹)为基准,施加 5 次随机参数的时空双仿射变换,控制变换矩阵 A 为满秩矩阵,变换参数 c 的范围为 $[0.5, 1.5]$ 。计算每个时刻的微分不变量,并连接成曲线图,如图 3 和图 4 所示。8 个不变特征按照式(12)赋予序号。对于每个不变特征,首先绘制其在时间对齐之前的曲线图,如图 3 (a) 和图 4 (a) 所示。随后,为了进一步展现出它对应时间点的不变性,基于变换参数进行时间对齐,如图 3 (b) 和图 4 (b) 所示。在对齐前,不变特征的曲线在形状上是相似的,仅在时间维度上有所伸缩,而在对应时刻点的不变量数值保持不变。对齐后的实验结果验证了这一点。

进一步地,验证不变特征对不同形状曲线的区分性。仍以图 1 中的两种曲线为基准,分别进行 50 次随机参数的时空双仿射变换,计算微分不变特征后,使用 DTW 技术计算样本距离。作为对比,基于坐标向量,同样使用 DTW 计算样本距离矩阵。实验结果如图 5 所示。可以看出,不变特征可以很好地区分两种曲线轨迹,且在其特征空间中,无论进行怎样的时空双仿射变换,两种轨迹的特征距离保持稳定。而仅从坐标向量的角度观察时,无法有效区分两种曲线轨迹。

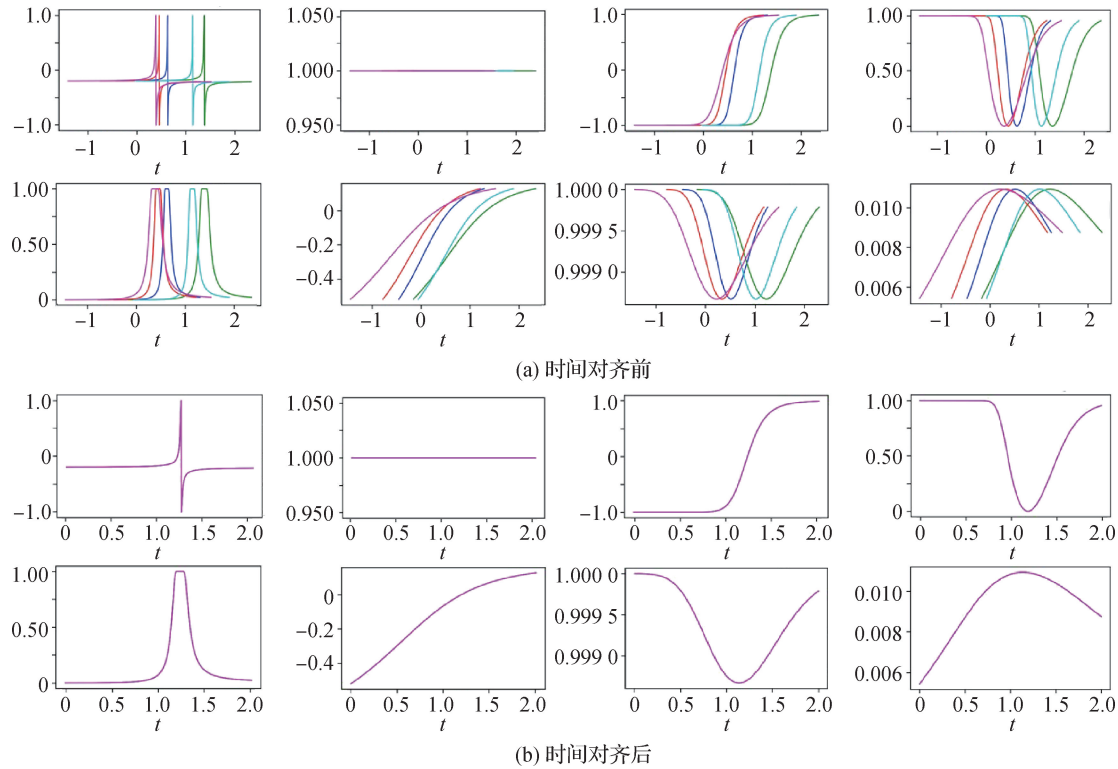


图3 3维螺旋线的8个不变特征变化曲线图

Fig.3 The 8 STDADI curves for the 3D spiral line((a) before time alignment;(b) after time alignment)

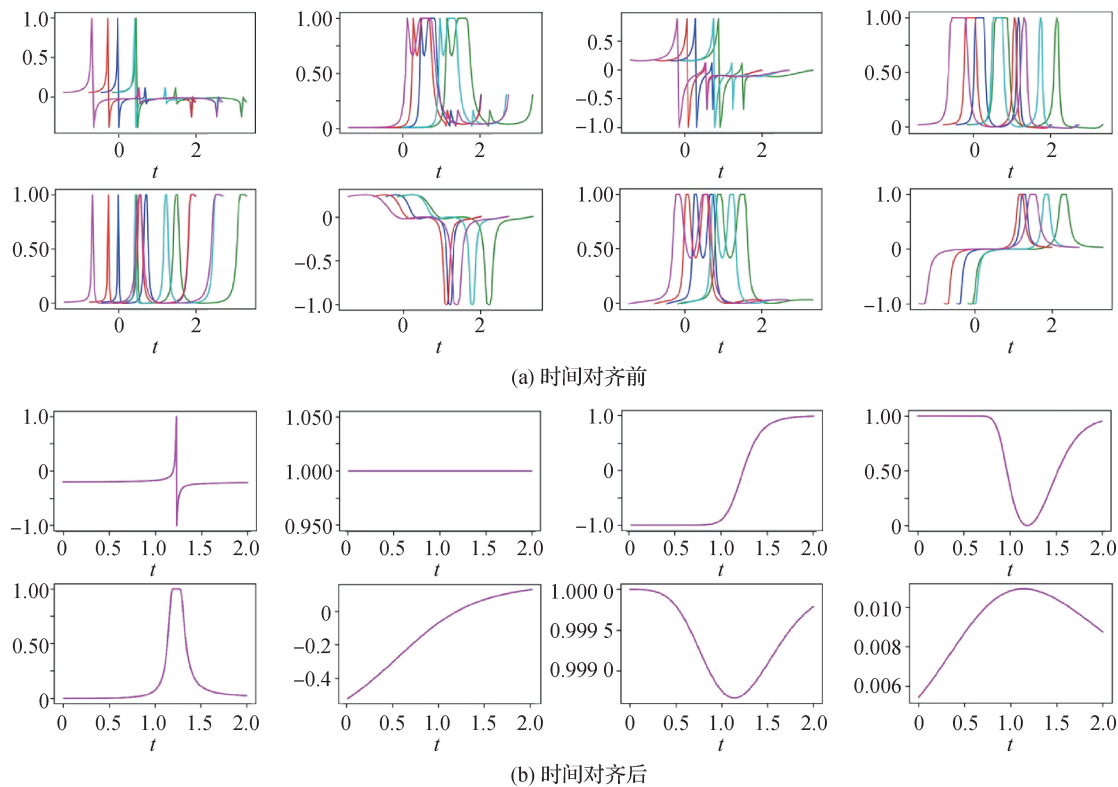


图4 抛掷动作中的右肩轨迹的8个不变特征变化曲线图

Fig.4 The 8 STDADI curves for trajectory of right shoulder of a throw action((a) before time alignment;(b) after time alignment)

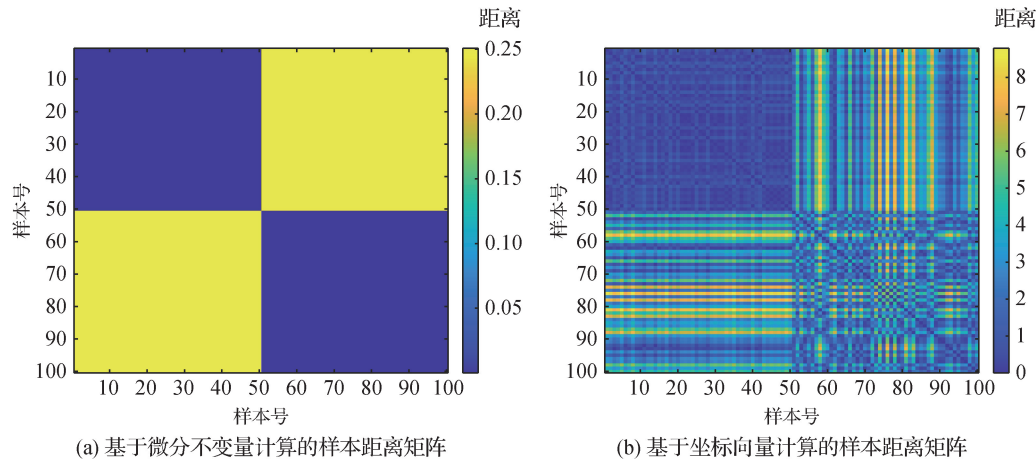


图 5 两种曲线轨迹时空双仿射变换后生成的样本距离矩阵

Fig. 5 Distance matrix of samples generated by two curve trajectories by spatio-temporal dual affine transformation
((a) sample distance matrix calculated by differential invariants; (b) sample distance matrix calculated by coordinates)

3.2 NTU 60 和 NTU 120 数据集

NTU RGB + D (NTU 60) 及其扩展版本 NTU RGB + D120(NTU 120) 是目前最大的基于 3 维人体骨架序列的动作识别数据集, 每个样本为 Microsoft Kinect V2 相机在受限的室内环境中获取的动作序列, 每个时刻包含相机坐标系下 25 个主要人体关节的 3 维坐标。NTU 60 数据集包含 40 个参与者执行的 60 个动作类别中的 56 880 个样本, NTU 120 数据集通过添加 57 600 个样本来扩展原始样本, 将动作类别和参与者数量分别扩展到 120 和 106。在两个数据集上均进行跨参与者识别(cross-subject) 与跨视角识别(cross-view) 实验。其中跨参与者识别实验以 50%

参与者样本为训练集, 其余 50% 参与者样本为测试集; 跨视角识别实验以其中两个视角样本为训练集, 另外一个视角样本为测试集。NTU 120 引入了更多影响视角的因素, 包括相机到动作参与者的高度距离, 将跨视角识别扩展为跨环境识别(cross-setup)。两项实验从不同角度考察算法模型的学习能力。实验选择测试集的 top-1 识别准确率作为主要评估指标。

3.3 实验结果与分析

首先与部分已有方法以及两种基线方法进行对比, 其中多数已有方法的结果取自 Liu 等人(2020) 的研究, 对比结果如表 1 所示。可以看出, 基于不变特征的通道增强方法在两个数据集上均取得了最优

表 1 不同方法在 NTU 60 和 NTU 120 数据集上的识别准确率
Table 1 The recognition accuracy of different methods on NTU 60 and NTU 120 datasets

方法	NTU 60 数据集		NTU 120 数据集	
	跨参与者识别	跨视角识别	跨参与者识别	跨环境识别
part-aware LSTM(Shahroudy 等, 2016)	62.9	70.3	25.5	26.3
spatio-temporal LSTM(Liu 等, 2016)	69.2	77.7	55.7	57.9
GCA-LSTM(Liu 等, 2017b)	74.4	82.8	58.3	59.2
two-stream attention LSTM(Liu 等, 2017a)	76.1	84.0	61.2	63.3
skeleton visualization(Liu 等, 2017c)	80.0	87.2	60.3	63.2
body pose evolution map(Liu 等, 2018)	82.4	86.7	64.6	66.9
multi-task learning network(Ke 等, 2017)	79.6	84.8	58.4	57.9
multi-task CNN with RotClips(Ke 等, 2018)	81.1	87.4	62.2	61.8
ST-GCN(Yan 等, 2018)	81.5	88.3	71.7	74.3
ST-GCN + 数据增强	80.6	90.5	72.2	79.0
ST-GCN + 通道增强	83.4	91.3	77.3	78.8

注: 加粗字体表示每列最优结果。

或者接近最优的结果。特别地,在与两种基线方法对比时,本文方法在两项实验中的识别准确率均取得了明显改进。图6为不同实验设置下在测试集上绘制的ROC(receiver operating characteristic)曲线图,与识别准确率的结果保持一致。考虑到数据增强方法对样本施加旋转、缩放与错切变换,其主要是从空间变换角度使训练样本具有多样性,考虑因素以视角方向与远近变化为主,因此在跨视角识别中取得了不错的性能提升。而在跨参与者识别中,不同参与者的动作差异往往体现在动作幅度与动作节奏上,动作幅度的变化可依据关节轨迹的尺度变换

建模,而动作节奏则是时间维度的变换。因此,基于空间变换的数据增强方法在跨参与者识别时无法获得明显的性能改进,甚至在NTU 60上出现了性能的倒退。基于可视化样本,可以推断这可能是由于部分空间变换的幅度超出了线性变换所能描述的正常人体姿态的范畴,从而在训练数据中引入了额外的噪声。作为对比,本文方法综合考虑了时间空间变换,在跨视角的实验设置上取得了与数据增强方法相近的性能,在跨参与者的实验设置上取得了较为明显的性能提升。因此,本文方法对时空两方面的变换均可以提升网络的鲁棒性。

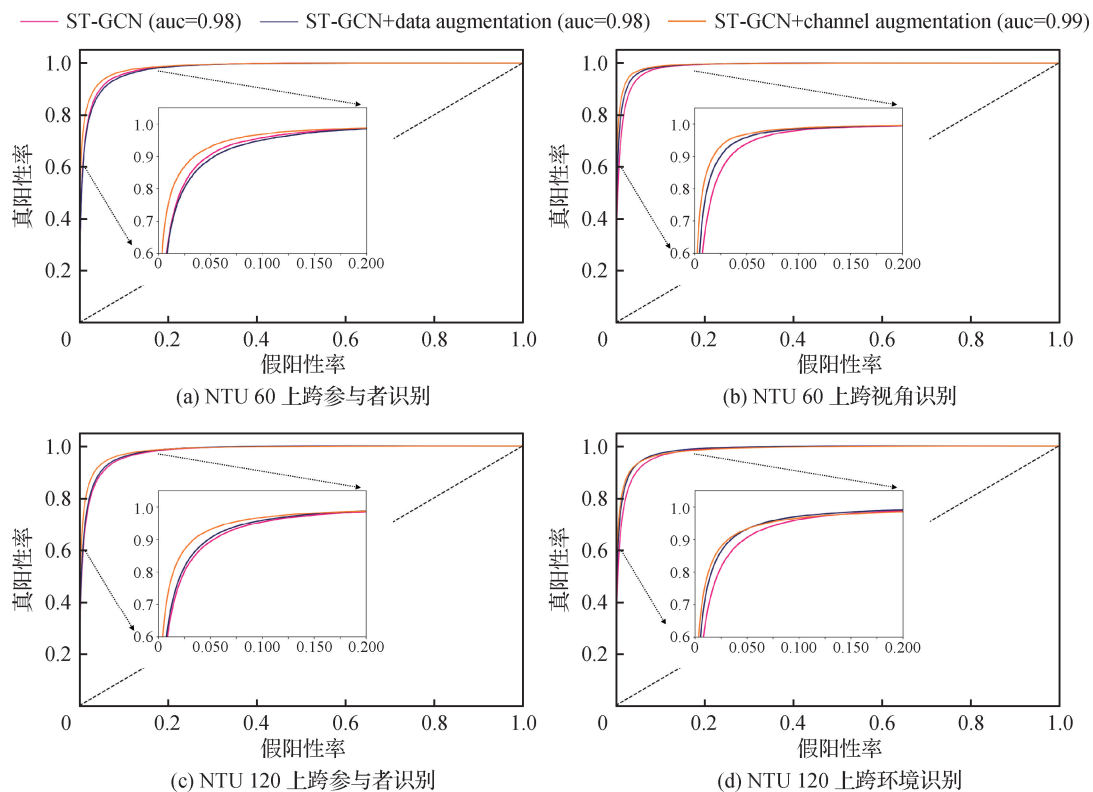


图6 不同实验设置下的ROC曲线图

Fig. 6 ROC curves under different experimental setting((a) NTU 60 cross-subject setting;
(b) NTU 60 cross-view setting; (c) NTU 120 cross-subject setting; (d) NTU 120 cross-setup setting)

为了定性分析不变特征和通道增强的方法相对原神经网络的变化幅度,对逐类别的识别准确率变化进行可视化。图7为NTU 60的跨参与者识别结果,列出了准确率变化超过 $\pm 2\%$ 的动作类别,并以条形图绘制准确率变化幅度。可以看出,指向某物和敬礼获得了最大程度的性能提升。准确率提升较

大的动作一般具有特定的关节轨迹运动模式。当执行指向某物或敬礼等动作时,不同运动者的肩、肘、腕关节的轨迹在形状上是相似的,但是会受到时空畸变的影响。相应地,洗头发的动作在准确率上有明显降低,而其关节轨迹(如腕部)往往是杂乱无章的。然而,对于多数动作类型而言,关节轨迹的运动

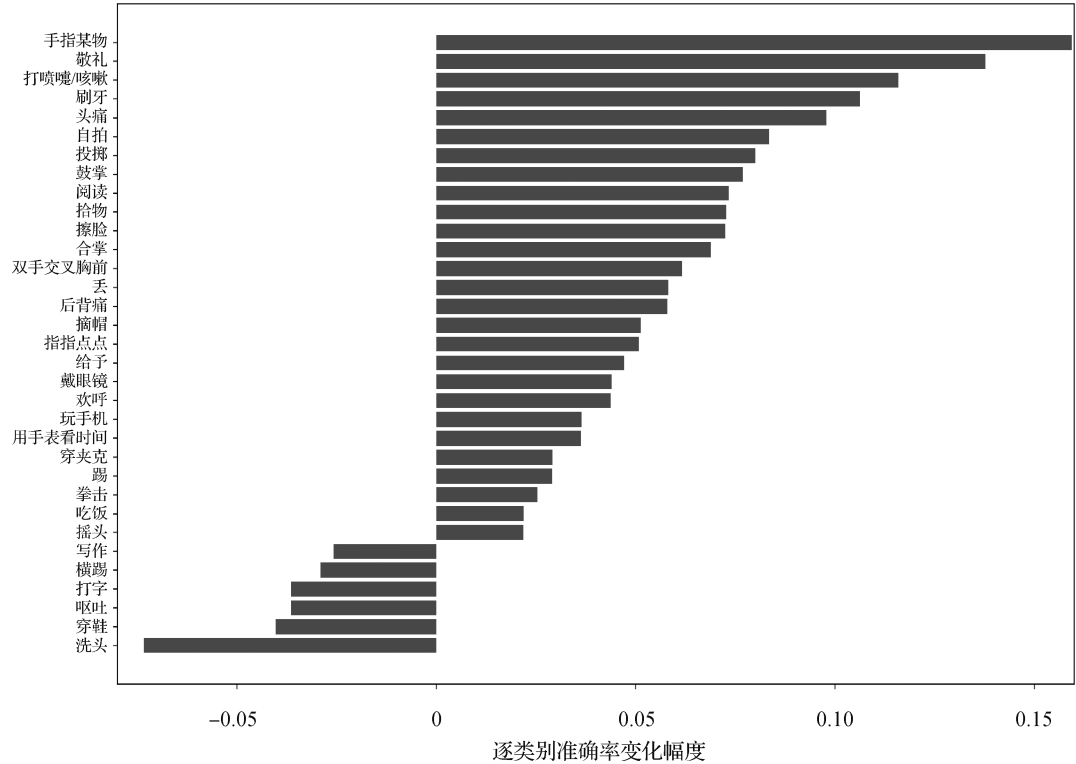


图 7 在 NTU 60 数据集上通道增强后的跨参与者识别结果逐类别的准确率变化幅度

Fig. 7 Accuracy change of each category after channel augmentation for cross-subject on NTU 60 dataset

模式是一致的,因此本文方法可以实现识别准确率的提升。

为了进一步考察不变特征的有效性,在通道增强框架下尝试使用各阶导数的输入设置。导数向量包含关节轨迹的1,2,3阶导数,共9维,在维度数量上与不变特征相近。不同输入设置在 NTU 60 数据集上的识别准确率如表 2 所示,每行分别表示仅使用微分不变量(only STDADI)、通道增强使用导数(+ derivatives)、仅使用坐标(original)、通道增强使用微分不变量(+ STDADI)作为输入的识别准确率。可以看出,使用导数向量导致两种实验设置下

表 2 不同输入设置在 NTU 60 数据集上的识别准确率

Table 2 Recognition accuracy of different input settings on NTU 60 dataset		
方法	跨参与者识别	跨视角识别
仅使用 STDADI 作为输入	52.8	56.9
通道增强方法添加导数特征	80.4	87.6
仅使用原始坐标作为输入	81.5	88.3
通道增强方法添加 STDADI 特征	83.4	91.3

注:加粗字体表示每列最优结果。

的识别准确率均有降低,这从侧面证明了准确率的提高源于不变特征在畸变下的不变性。然而,如果不将不变特征通过通道增强的办法与坐标向量结合,而仅将样本的不变特征作为神经网络的输入时,可以发现识别准确率有明显降低。

训练过程的损失函数曲线如图 8 所示。可以看出,有些情况下网络难以收敛。对于网络难以收敛的原因,推断可能是由于不变特征主要基于曲线的低阶微分进行计算,在保持不变性的同时失去了一

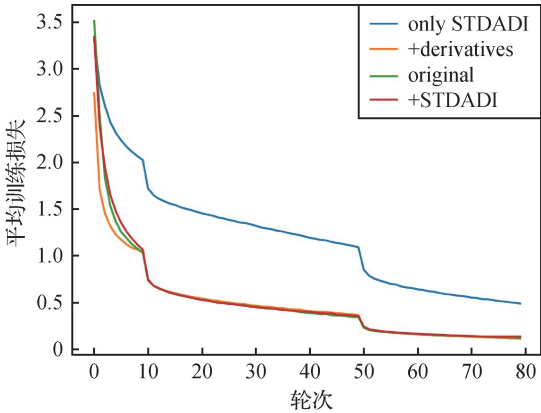


图 8 不同输入设置的训练损失函数曲线

Fig. 8 Training loss history with different input settings

定的样本信息,因此样本之间的区分性随之变弱。然而,将不变特征与关节坐标通过通道增强结合后,可以有效弥补这一缺陷。

4 结 论

从时间与空间变换的角度考察人体运动,用时空双仿射变换描述运动过程中人体关节轨迹的畸变,设计微分不变特征,使用通道增强方法,在骨架动作识别任务中提升了学习模型的泛化能力。值得注意的是,不同于数据增强,本文方法无需生成更多的样本,因此提升了模型训练的效率。实验结果证明,本文方法优于仅考虑了空间变换的数据增强,在跨参与者识别与跨视角识别中均取得了识别准确率的有效提升。然而,微分不变量在保持不变性的同时失去了一定的原始样本信息,与原始数据结合后才能取得较好效果。未来希望进一步探索时空双仿射变换在人体运动建模中的可行性,并研究获取更为有效的不变量,以期对人体运动的理解与分析不受各种形式变换的影响。

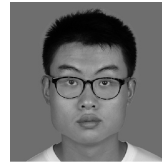
参考文献 (References)

- Anirudh R, Turaga P, Su J Y and Srivastava A. 2017. Elastic functional coding of riemannian trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(5): 922-936 [DOI: 10.1109/TPAMI.2016.2564409]
- Boulahia S Y, Anquetil E, Kulpa R and Multon F. 2016. HIF3D: hand-writing-inspired features for 3D skeleton-based action recognition// *Proceedings of the 23rd International Conference on Pattern Recognition*. Cancun, Mexico; IEEE: 985-990 [DOI: 10.1109/ICPR.2016.7899764]
- Brown A B. 1935. Functional dependence. *Transactions of the American Mathematical Society*, 38(2): 379-379 [DOI: 10.1090/S0002-9947-1935-1501816-5]
- Cao C Q, Lan C L, Zhang Y F, Zeng W J, Lu H Q and Zhang Y N. 2019. Skeleton-based action recognition with gated convolutional neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(11): 3247-3257 [DOI: 10.1109/TCSVT.2018.2879913]
- Ding C Y, Liu K, Li G, Yan L, Chen B Y and Zhong Y M. 2020. Spatio-temporal weighted posture motion features for human skeleton action recognition research. *Chinese Journal of Computers*, 43(1): 29-40 (丁重阳, 刘凯, 李光, 闫林, 陈博洋, 钟育民. 2020. 基于时空权重姿态运动特征的人体骨架行为识别研究. *计算机学报*, 43(1): 29-40) [DOI: 10.11897/SP. J. 1016. 2020. 00029]
- Du Y, Wang W and Wang L. 2015. Hierarchical recurrent neural network for skeleton based action recognition// *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA; IEEE: 1110-1118 [DOI: 10.1109/CVPR.2015.7298714]
- Esling P and Agon C. 2012. Time-series data mining. *ACM Computing Surveys*, 45(1): #12 [DOI: 10.1145/2379776.2379788]
- He X X, Shao C X and Xiong Y. 2014. A new similarity measure based on shape information for invariant with multiple distortions. *Neurocomputing*, 129: 556-569 [DOI: 10.1016/j.neucom.2013.09.003]
- Hussein M E, Torki M, Gawayyed M A and El-Saban M. 2013. Human action recognition using a temporal hierarchy of covariance descriptors on 3D joint locations// *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*. Beijing, China; AAAI Press: 2466-2472
- Kacem A, Daoudi M, Amor B B, Berretti S and Alvarez-Paiva J C. 2020. A novel geometric framework on gram matrix trajectories for human behavior understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1): 1-14 [DOI: 10.1109/TPAMI.2018.2872564]
- Ke Q H, Bennamoun M, An S J, Sohel F and Boussaid F. 2017. A new representation of skeleton sequences for 3D action recognition// *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA; IEEE: 4570-4579 [DOI: 10.1109/CVPR.2017.486]
- Ke Q H, Bennamoun M, An S J, Sohel F and Boussaid F. 2018. Learning clip representations for skeleton-based 3d action recognition. *IEEE Transactions on Image Processing*, 27(6): 2842-2855 [DOI: 10.1109/TIP.2018.2812099]
- Kim T S and Reiter A. 2017. Interpretable 3D human action analysis with temporal convolutional networks// *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Honolulu, USA; IEEE: 1623-1631 [DOI: 10.1109/CVPRW.2017.207]
- Li B, Dai Y C, Cheng X L, Chen H H, Lin Y and He M Y. 2017a. Skeleton based action recognition using translation-scale invariant image mapping and multi-scale deep CNN// *Proceedings of 2017 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. Hong Kong, China; IEEE: 601-604 [DOI: 10.1109/ICMEW.2017.8026282]
- Li B, Li X, Zhang Z F and Wu F. 2019a. Spatio-temporal graph routing for skeleton-based action recognition. *Proceedings of 2019 AAAI Conference on Artificial Intelligence*, 33(1): 8561-8568 [DOI: 10.1609/aaai.v33i01.33018561]
- Li C, Zhong Q Y, Xie D and Pu S L. 2017b. Skeleton-based action recognition with convolutional neural networks// *Proceedings of 2017 IEEE International Conference on Multimedia and Expo Workshops*

- (ICMEW). Hong Kong, China; IEEE: 597-600 [DOI: 10.1109/ICMEW.2017.8026285]
- Li M S, Chen S, Chen X, Zhang Y, Wang Y F and Tian Q. 2019b. Actional-structural graph convolutional networks for skeleton-based action recognition//Proceedings of 2019 IEEE/CVF Conference on Computer vision and Pattern Recognition. Long Beach, USA; IEEE: 3590-3598 [DOI: 10.1109/CVPR.2019.00371]
- Liu H, Tu J H and Liu M Y. 2017 a. Two-stream 3D convolutional neural network for skeleton-based action recognition [EB/OL]. [2020-08-06]. <https://arxiv.org/pdf/1705.08106.pdf>
- Liu J, Shahroudy A, Perez M, Wang G, Duan L Y and Kot A C. 2020. NTU RGB + D 120: a large-scale benchmark for 3D human activity understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(10): 2684-2701 [DOI: 10.1109/TPAMI.2019.2916873]
- Liu J, Shahroudy A, Xu D and Wang G. 2016. Spatio-temporal LSTM with trust gates for 3D human action recognition//Proceedings of 2016 European Conference on Computer Vision. Amsterdam, the Netherlands; Springer: 816-833 [DOI: 10.1007/978-3-319-46487-9_50]
- Liu J, Wang G, Duan L Y, Abdiyeva K and Kot A C. 2017b. Skeleton-based human action recognition with global context-aware attention LSTM networks. IEEE Transactions on Image Processing, 27(4): 1586-1599 [DOI: 10.1109/TIP.2017.2785279]
- Liu J, Wang G, Hu P, Duan L Y and Kot A C. 2017c. Global context-aware attention LSTM networks for 3D action recognition//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 3671-3680 [DOI: 10.1109/CVPR.2017.391]
- Liu M Y, Liu H and Chen C. 2017d. Enhanced skeleton visualization for view invariant human action recognition. Pattern Recognition, 68: 346-362 [DOI: 10.1016/j.patcog.2017.02.030]
- Liu M Y and Yuan J S. 2018. Recognizing human actions as the evolution of pose estimation maps//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 1159-1168 [DOI: 10.1109/CVPR.2018.00127]
- Müller M, Röder T and Clausen M. 2005. Efficient content-based retrieval of motion capture data. ACM Transactions on Graphics, 24(3): 677-685 [DOI: 10.1145/1073204.1073247]
- Pham H H, Salmane H, Khoudour L, Crouzil A, Zegers P and Velastin S A. 2019. Spatio-temporal image representation of 3D skeletal movements for view-invariant action recognition with deep convolutional neural networks. Sensors, 19(8): #1932 [DOI: 10.3390/s19081932]
- Sadjadi F A and Hall E L. 1980. Three-dimensional moment invariants. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2(2): 127-136 [DOI: 10.1109/TPAMI.1980.4766990]
- Shahroudy A, Liu J, Ng T T and Wang G. 2016. NTU RGB + D: a large scale dataset for 3D human activity analysis//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 1010-1019 [DOI: 10.1109/CVPR.2016.115]
- Shao Z P and Li Y F. 2015. Integral invariants for space motion trajectory matching and recognition. Pattern Recognition, 48(8): 2418-2432 [DOI: 10.1016/j.patcog.2015.02.029]
- Shi L, Zhang Y F, Cheng J and Lu H Q. 2019a. Skeleton-based action recognition with directed graph neural networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 7904-7913 [DOI: 10.1109/CVPR.2019.00810]
- Shi L, Zhang Y F, Cheng J and Lu H Q. 2019b. Two-stream adaptive graph convolutional networks for skeleton-based action recognition//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 12018-12027 [DOI: 10.1109/CVPR.2019.01230]
- Si C Y, Chen W T, Wang W, Wang L and Tan T N. 2019. An attention enhanced graph convolutional LSTM network for skeleton-based action recognition//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 1227-1236 [DOI: 10.1109/CVPR.2019.00132]
- Veeriah V, Zhuang N F and Qi G J. 2015. Differential recurrent neural networks for action recognition//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE: 4041-4049 [DOI: 10.1109/ICCV.2015.460]
- Vemulapalli R, Arrate F and Chellappa R. 2014. Human action recognition by representing 3D skeletons as points in a lie group//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA; IEEE: 588-595 [DOI: 10.1109/CVPR.2014.82]
- Wang H S and Wang L. 2017. Modeling temporal dynamics and spatial configurations of actions using two-stream recurrent neural networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 3633-3642 [DOI: 10.1109/CVPR.2017.387]
- Wang J, Liu Z C, Wu Y and Yuan J S. 2012. Mining actionlet ensemble for action recognition with depth cameras//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA; IEEE: 1290-1297 [DOI: 10.1109/CVPR.2012.6247813]
- Wang P C, Li Z Y, Hou Y H and Li W Q. 2016. Action recognition based on joint trajectory maps using convolutional neural networks//Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, the Netherlands; ACM: 102-106 [DOI: 10.1145/2964284.2967191]
- Wang P C, Li W Q, Ogunbona P, Wan J and Escalera S. 2018. RGB-D-based human motion recognition with deep learning: a survey. Computer Vision and Image Understanding, 171: 118-139 [DOI: 10.1016/j.cviu.2018.04.007]

- Wang Y Y, Li Y B and Ji X F. 2013. Human action recognition based on super-interest points features. *Journal of Image and Graphics*, 18(7): 805-812 (王扬扬, 李一波, 姬晓飞. 2013. 人体动作的超兴趣点特征表述及识别. *中国图象图形学报*, 18(7): 805-812) [DOI: 10.11834/jig.20130710]
- Wu S D and Li Y F. 2009. Flexible signature descriptions for adaptive motion trajectory representation, perception and recognition. *Pattern Recognition*, 42(1): 194-214 [DOI: 10.1016/j.patcog.2008.06.023]
- Yan S J, Xiong Y J and Lin D H. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition [EB/OL]. [2020-08-06]. <https://arxiv.org/pdf/1801.07455.pdf>
- Zhang P E, Lan C L, Xing J L, Zeng W J, Xue J R and Zheng N N. 2017. View adaptive recurrent neural networks for high performance human action recognition from skeleton data//*Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE*: 2136-2145 [DOI: 10.1109/ICCV.2017.233]
- Zheng X, Peng X D and Wang J X. 2018. Human action recognition based on pose spatio-temporal features. *Journal of Computer-Aided Design and Computer Graphics*, 30(9): 1615-1624 (郑潇, 彭晓东, 王嘉璇. 2018. 基于姿态时空特征的人体行为识别方法. *计算机辅助设计与图形学学报*, 30(9): 1615-1624) [DOI: 10.3724/SP.J.1089.2018.16848]
- Zhong Q B, Zheng C M and Piao S H. 2020. Research on skeleton-based action recognition with spatiotemporal fusion and human-robot interaction. *CAAI Transactions on Intelligent Systems*, 15(3): 601-608 (钟秋波, 郑彩明, 朴松昊. 2020. 时空域融合的骨架动作识别与交互研究. *智能系统学报*, 15(3): 601-608) [DOI: 10.11992/tis.202006029]

作者简介



李琪, 1992年生, 男, 博士研究生, 主要研究方向为人体运动分析。

E-mail: neilyoyoyo@163.com

墨瀚林, 男, 博士研究生, 主要研究方向为形状分析。

E-mail: mohanlin@ict.ac.cn

赵婧涵, 女, 硕士研究生, 主要研究方向为时间序列分析。

E-mail: zhaojinghan17s@ict.ac.cn

郝宏翔, 男, 硕士研究生, 主要研究方向为图像模糊识别。

E-mail: haohongxiang18s@ict.ac.cn

李华, 男, 教授, 主要研究方向为计算机图形学、不变特征分析和形状分析。E-mail: lihua@ict.ac.cn