Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2021, 10(3), 248-256
Published Online August 2021 in Hans. http://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2021.103025

Hans汉斯

基于改进时空图卷积网络的乒乓球击球动作识别

刘明方[1,2]，汪语哲[1,3]，尹真杰[1,2]，张皓天[1,3]，段晓东[1,2]

[1]大连民族大学大数据应用技术国家民委重点实验室，辽宁 大连
[2]大连民族大学计算机科学与工程学院，辽宁 大连
[3]大连民族大学机电工程学院，辽宁 大连
Email: 124478369@qq.com

收稿日期：2021年7月1日；录用日期：2021年7月15日；发布日期：2021年8月4日

摘 要

本文研究了计算机视觉辅助开展乒乓球训练中的乒乓球击球动作识别问题。基于骨骼关键点方式的动作识别算法，只对人体骨骼点的时空信息进行学习，可以去除环境、光线等干扰因素。通过摄像机采集了正手击球、反手击球、正手拉球、反手拉球和非击球动作5类动作在内的体育运动视频，使用OpenPose提取18个人体骨骼关键点，构建了乒乓球击球骨骼点数据集。根据乒乓球击球核心力量区域对ST-GCN网络的卷积核进行调整，最终训练模型的击球动作精准度可以达到98%；并在文章创建数据集之外的乒乓球击球动作视频上进行了泛化测试，对比ST-GCN网络的泛化效果，结果文章调整后的时空图卷积网络方法效果更好，具有较高的实用价值。

关键词

乒乓球击球，动作识别，骨骼关键点，图卷积网络

Recognition of Table-Tennis Action Based on Improved Spatio-Temporal Graph Convolutional Network

Mingfang Liu[1,2], YuzheWang[1,3], Zhenjie Yin[1,2], Haotian Zhang[1,3], Xiaodong Duan[1,2]

[1]SEAC Key Laboratory of Big Data Applied Technology, Dalian Minzu University, Dalian Liaoning
[2]College of Computer Science and Engineering, Dalian Minzu University, Dalian Liaoning
[3]College of Mechanical and Electronic Engineering, Dalian Minzu University, Dalian Liaoning
Email: 124478369@qq.com

文章引用: 刘明方, 汪语哲, 尹真杰, 张皓天, 段晓东. 基于改进时空图卷积网络的乒乓球击球动作识别[J]. 人工智能与机器人研究, 2021, 10(3): 248-256. DOI: 10.12677/airr.2021.103025

Received: Jul. 1st, 2021; accepted: Jul. 15th, 2021; published: Aug. 4th, 2021

Abstract

The problem of table tennis training with assistance of computer video was studied in this paper. Action recognition algorithm based on method of key point of the bone only learns the spatio-temporal information of the human bone points, and can remove interference factors such as environment and light. Video of sports activities including forehand, backhand, forehand, backhand, and non-hit action was collected through the camera, and 18 key points of human bones were extracted using OpenPose to construct a dataset of bones of players playing table tennis. Convolution kernel of the ST-GCN network was adjusted according to the core strength area of table tennis striking, and accuracy of the final training model's striking action can reach 98%. Generalization test was performed on video of table tennis striking beyond data set proposed in this paper, and generalization effect showed that the proposed spatio-temporal graph convolutional network method showed better results and thus had higher practical value than the proposed ST-GCN network.

Keywords

Striking of Table Tennis, Human Action Recognition, Skeleton Key Points, Graph Convolutional Network

1. 引言

随着体育科技的发展，计算机技术已成为辅助开展体育训练、提升竞技水平的重要手段。通过计算机视觉对体育运动者的动作行为进行识别，从而辅助运动者提升个人技战术能力，对体育科技的发展有着很重要的影响。乒乓球被称为我们国家的“国球”，同时乒乓球运动也是一项老少皆宜的动作，但是对于大多数乒乓球业余爱好来说，由于没有受过专业培训，个人技战术水平如何提升是一项较为困难的事情。通过乒乓球发球机结合计算机手段是辅助训练者提升个人技战术水平的一项重要选择。

动作行为跟踪识别在计算视觉领域是近来的研究热点，其目标是通过视频或者摄像头识别人体行为、人与环境交互、人与人交互等活动，一般分为基于骨骼关键点和时空特征分析两个主要部分，基于骨骼关键点主要算法是图卷积网络(GCN) [1]，时空图卷积网络(ST-GCN)是通过图卷积研究动作识别的开山之作[2]，通过骨骼提取算法提取人体骨骼关键点信息，时空图卷积对骨骼关键点的时空特征信息进行学习以完成行为识别的目的。基于时空特征分析的主要算法有三种：双流法(Two-Stream) [3]，使用两种分类器同时训练并融合的方式达到识别人体行为的目的，一种是 RGB 图像的分类器，另一种是光流图像分类器；3D 卷积神经网络(3D CNN) [4]方式直接使用 3D 卷积核学习视频帧序列的时空特征；卷积神经网络与长短时记忆网络结合(CNN + LSTM) [5]方式，使用卷积神经网络学习视频帧序列的空间特征，用长短时记忆网络学习视频序列的时间特征。对于时空特征分析的动作识别算法，其数据集中为正常视频数据集，视频图像中干扰识别的环境因素较多，利于用在人与环境交互的场景。而基于骨骼关键点的动作识别算法是对人体骨骼点的时空信息进行分析，不受环境的干扰，利于在个人动作行为识别场合使用。

2. 算法流程设计

文章构建数据集并训练乒乓球击球模型，以辅助乒乓球爱好者自己通过乒乓球发球机、乒乓球动作训练器等硬件设备训练时矫正发力动作。通过与具备专业乒乓球技能的老师和学生合作，采用 60 fps 相机采集其击球动作，利用 OpenPose 提取人体骨骼关键点，最后根据乒乓球击球动作对 ST-GCN 网络的卷积核更改并对数据集进行训练，得到的训练模型可以完成对乒乓球击球动作的识别，并且泛化性能和鲁棒性较强。乒乓球击球动作识别流程图如图 1 所示。

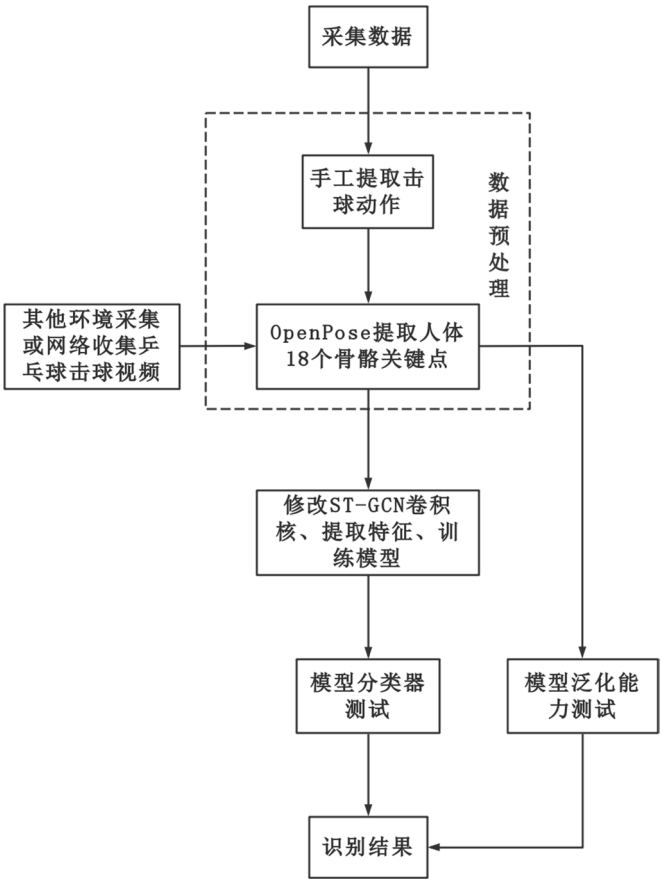


Figure 1. Flowchart of recognition of table tennis striking action
图 1. 乒乓球击球动作识别流程图

3. 数据集采集

3.1. 数据集采集硬件设备与乒乓球运动员选择

乒乓球辅助训练的硬件设备为乒乓球发球机，能发出上旋、下旋转和侧旋球，同时具备调节发球速度、角度、频率等功能，因此可满足基本技术动作训练。

视频采集设备为摄像机，为保证击球动作速度较快时图片不会出现运动模糊现象，摄像机参数调整为：分辨率 720 像素、帧率 60 fps。

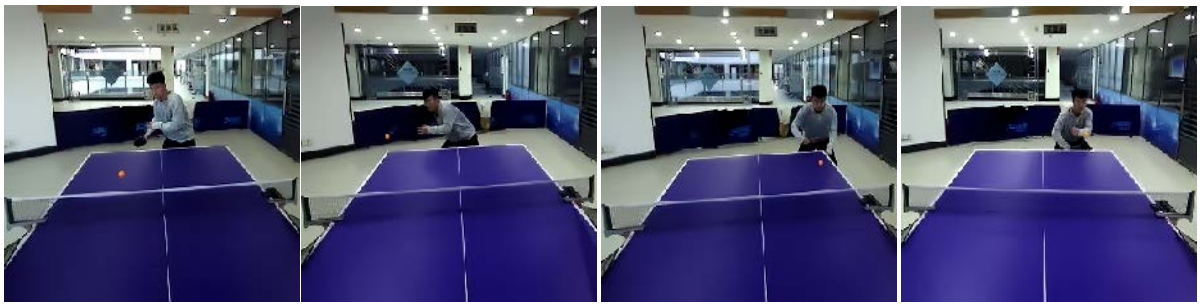
目前，乒乓球击球动作没有一个准确的动作集标准，然而项目组邀请的专业教练员和运动员的测试结果表明，他们的击球时肢体动作大致一致，因此选择与专业乒乓球运动员和教练进行合作，以他们的击球动作作为标准数据集，辅助乒乓球击球爱好者通过硬件设备训练时矫正自身动作。

3.2. 数据集采集

香港中文大学研究者于 2020 年构建了一个大规模体操运动人体动作数据集：FineGym [6]，数据集通过层级化标注对动作的细粒度进行了区分，最细划分到一个具体的体操动作，例如平衡木体操下马动作中的团身前空翻动作。文章根据此数据所划分的细粒度动作，构建了乒乓球击球动作数据集。

乒乓球击球动握拍方式大致可以分为横拍和直拍两种，通过分析视频中的连续帧发现，在图 2 中所示，直拍中有些击球动作轨迹相似，例如正手击球和正手拉球在即将击球的瞬间的连续 2~3 帧动作相似，反手击球和反手拉球在准备接球和击球的瞬间都有 1~3 帧的动作相似，这也是击球动作中区分的难点。因此，在采集数据和做动作区分时，选择了对正手击球、正手拉球、反手击球、反手拉球的动作进行采集。采集的数据集中，击球动作之外的动作将其称为其他动作，对 4 类击球动作和 1 类其他动作进行识别，以便于模型更好区分击球动作。

(a1) 正手击球　(b1) 正手拉球　(c1) 反手击球　(d1) 反手拉球



(a2) 正手击球　(b2) 正手拉球　(c2) 反手击球　(d2) 反手拉球

Figure 2. Similar actions in table-tennis striking action: (a1) and (b1) are a certain frame with similar actions in the process of forehand and forehand; (a2) and (b2) are a certain frame with difference between forehand and forehand;(c1) and (d1) are a certain frame of similar actions in the process of backhand and backhand

图 2. 击球动作中的相似动作：(a1)和(b1)为正手击球和正手拉球过程中相似动作的某一帧；(a2)和(b2)为正手击球和正手拉球过程中动作区别的某一帧；(c1)和(d1)为反手击球和反手拉球过程中的相似动作的某一帧

4. 数据集处理

4.1. 手工处理数据

对乒乓球击球视频的每一帧进行提取，经过研究对比发现，每种动作在 32 帧内完成击球，将包含一个击球动作的连续帧进行提取，形成一个击球动作的时空特征数据，而其非击球动作一般在 48 个连续帧之内，提取非击球动作的连续帧，形成一个非击球动作的时空特征数据。经过手工提取数据后的数据集包括了 824 个动作片段，共 4 类击球动作，1 类非击球动作。表 1 显示了动作类别和动作数量。

Table 1. System resulting data of standard experiment
表 1. 数据集动作类别和动作数量

击球动作	动作样本数量
正手击球	179
正手拉球	122
反手击球	206
反手拉球	158
其他动作	159

4.2. OpenPose 提取骨骼关键点

OpenPose [7]是美国卡耐基梅隆大学的研究者在人体姿态识别项目中提出的一个模型[8]，此模型可以实时跟踪识别 15、18 或者 25 个身体关键点，单只手上的 21 个手部关键点，70 个面部关键点[7]。此模型是通过自底向上的方法实时检测出图像中多人的人体、面部和手部的关键点，是基于深度学习的实时多人姿态估计应用的开山之作[8]。乒乓球横拍击球是腰部和手臂动作的配合，使用 OpenPose 提取人体中的 18 个关键点，所提取获取的是每个骨骼关键点的索引(index)、像素坐标(x,y)和置信度，如图 3 所示。

通过研究将手工区分数据集中 16 个连续视频帧视为一个击球动作，对于大于 16 帧的动作片段，根据式 1 进行跳帧提取，首先保存连续动作的第一帧，设置所选取的动作帧f初始值为 1，C_{fram} 为手工处理后一个动作连续帧的总数量，m 为 C_{fram} 除以 16 并向下取整后的正整数，随后 15 帧从f开始每隔 m 帧取一帧。

$$\begin{cases} f = 1 \\ m = \left\lfloor \dfrac{C_{fram}}{16} \right\rfloor \\ f += m \end{cases} \tag{1}$$

通过 OpenPose 提取出每 16 连续帧中人体骨骼关键点，由于数据集中没有其他人员的干扰，于是只设置提取一个人的骨骼关键点即可，最终用一个 json 文件存放一个动作的连续骨骼点，文件存储格式如表 2。键 date 和其值存储骨骼点的数据。键 frame_index 和其值确定哪一帧的数据。键 skeleton 和其值存储一帧骨骼点的坐标和精准度。键 pose 对应的值是 36 个归一化到 1.0 以内的像素坐标值，存储形式每两个数值为一个骨骼点的在图像坐标轴上的坐标，共 18 个骨骼点。键 score 和其值存储骨骼点的精准度。键 label 和其值存储动作的名称。键 label_index 和其值存储动作的标签。

Table 2. Storage of continuous bone key point using json format
表 2. 连续骨骼关键点存储 json 格式

```
{“data”:[{
“frame_index”: 0
“ skeleton” :[{
“pose”: [ 0.0 - 1.0  之间的骨骼点(x,y)像素坐标值,共 36 个值],
“score”:[0.0 - 1.0  之间的骨骼点的置信度]
                        }]
        },
        {
            后 15 帧格式同 frame_index 0
        }
      ],
"label": "Backhand", "label_index": 0}
```

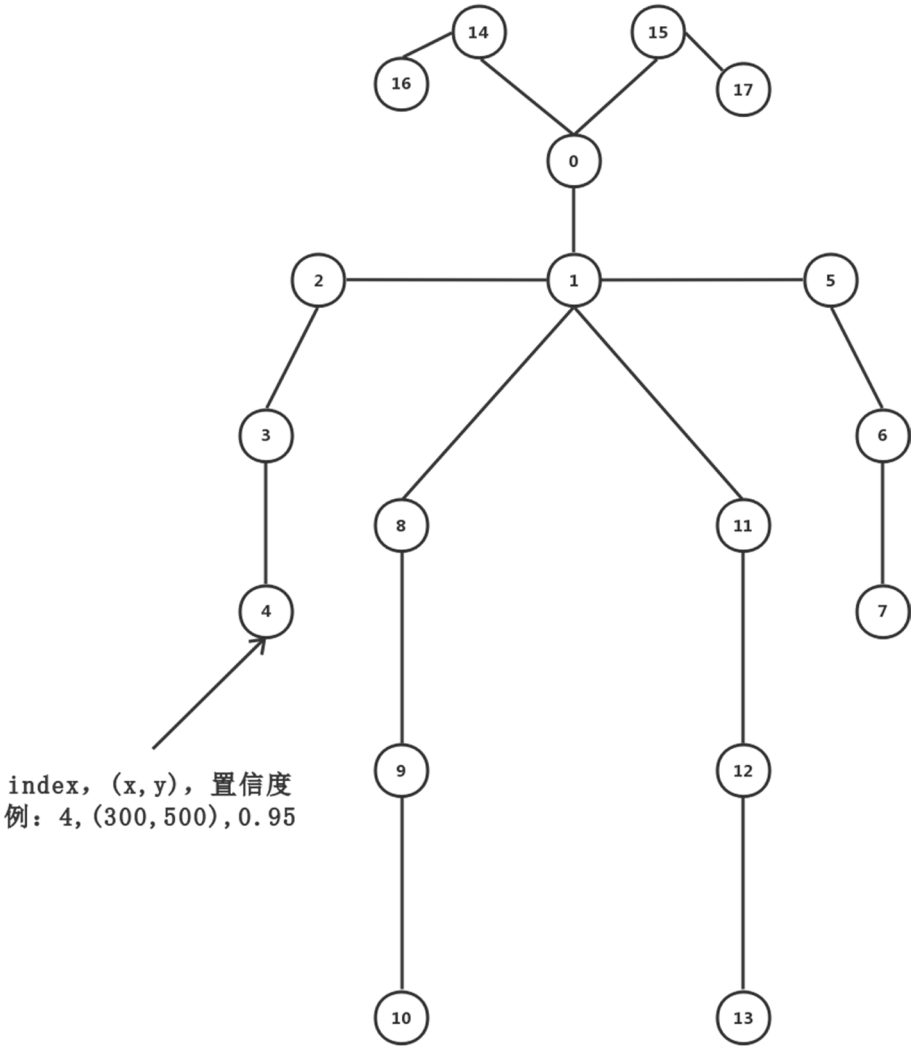


Figure 3. 18 bone points of human body extracted by OpenPose
图 3. OpenPose 提取的人体 18 个骨骼点

5. 基于图卷积的动作识别算法

5.1. ST-GCN 网络

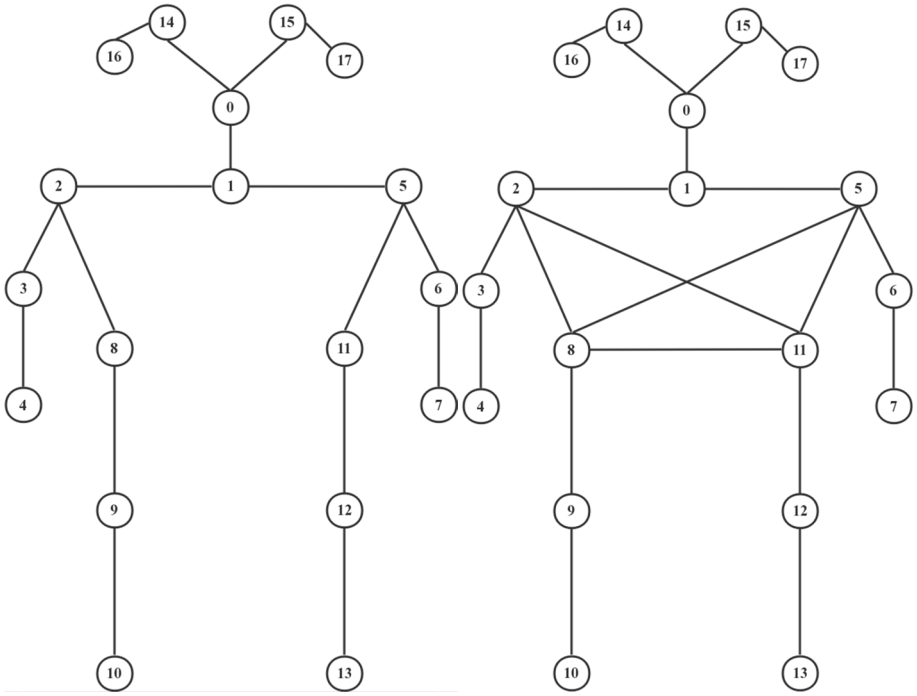
图卷积网络与图像的卷积网络不同，最大的区别在于图这种非欧式空间中邻居节点数目并不确定，不像图像那种邻居节点数目固定的空间结构，这也就导致了图卷积网络和图像卷积网络的不同。图卷积的卷积核可由拉普拉斯矩阵变换表示[9]，矩阵聚合方式有多种如归一化、随机游走归一化、对称归一化等[10]。在 ST-GCN 网络的实际代码中卷积核使用的是归一化拉普拉斯算子，公式如式(2)，其中 D 和 A 分别为拉普拉斯矩阵的度矩阵和邻接矩阵[11]。

$$x = D^{-1}AX \tag{2}$$

5.2. 骨骼点连通方式

在乒乓球击球过程中，击球的核心力量区位置是肩关节以下、髋关节以上的区域，击球核心肌肉群处于人体上半身身体躯干周围，核心力量区域内的核心肌肉群对乒乓球运动员击球身体姿势有着稳定和

支撑的作用[12]。



(a) ST-GCN 使用的连接方式　　(b) 文章提出的连接方式

Figure 4. Connection of skeleton point: (a) connection method used in the ST-GCN project; (b) connection method using improved skeletal point proposed in the article based on core strength area of table tennis striking
图 4. 骨骼点连接：(a)为 ST-GCN 项目中使用的连通方式；(b)是文章根据乒乓球击球核心力量区域改进的骨骼点连通方式

乒乓球的核心力量区域决定着在提取数据连接骨骼点时要考虑到肩关节和髋关节的连通，考虑整个击球过程是由核心力量在支撑运动员的击球动作，因此基于 OpenPose 提取的骨骼点中两个肩部骨骼点和两个臀部的骨骼点要完全连通，如图 4 所示，(a)为 ST-GCN 原有的骨骼连通方式，根据乒乓球击球核心力量区域的要求，其连通方式并不适合对乒乓球击球动作进行分类识别。

骨骼点连通方式改变后，在空间和时间上都可以更有效的提取到乒乓球击球核心力量区的特征。根据图 4(a)的骨骼连接方式，骨骼点 8 学习到与骨骼点 11 的空间特征关系需要多次卷才能提取，其中学习到的特征可能会有其它骨骼点特征的干扰，经过图 4(b)连接后，骨骼点 8 与骨骼点 11 的特征上有直接的联系，提取到的特征更符合乒乓球击球核心力量区域的动作表现。

6. 实验结果与分析

文章中基于自建数据集训练乒乓球击球动作识别模型实验硬件环境为：处理器 Intel(R) Core(TM) i7-7800X CPU@3.50GHz 3.50 GHz 和 NVIDIA GeForce RTX 2080Ti GPU；软件环境为：python 3.6 语言，Windows10 系统，Pycharm 编译器。

6.1. 模型训练实验结果分析

训练模型的训练集和验证集均来自于自建乒乓球击球骨骼数据集，对每个击球动作划分出 80%为训练集和 20%为验证集，则最后总数据集训练集和验证集的划分也分别为数据集的 80%和 20%。分别测试了文章改进的 ST-GCN 和原网络在文章自建的骨骼数据集上的表现，以及在开源数据集 kinetics-400 骨骼

数据集上的表现。通过实验，网络模型性能对比结果如表 3 所示。

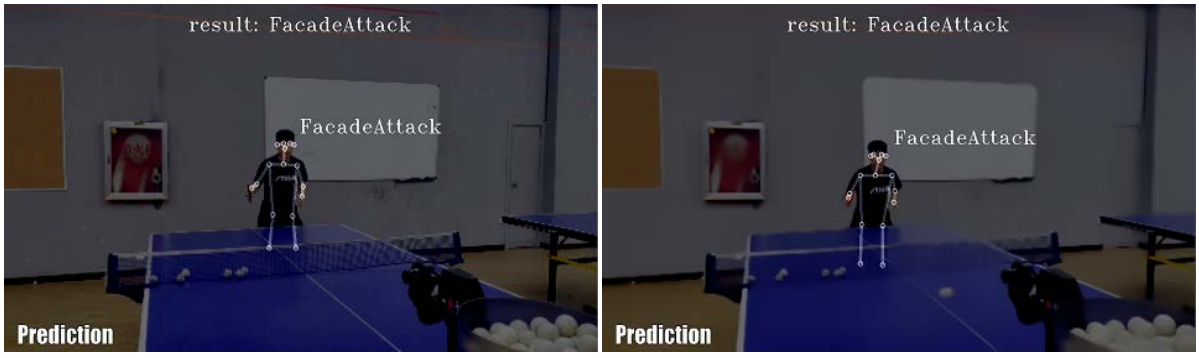
Table 3. Comparison of performance among network models
表 3. 网络模型性能对比

	Kinetics-400 Top1	Kinetics-400 Top5	文章数据集 Top1
ST-GCN	30.70%	52.80%	90.18%
文章改进模型	31.78%	54.73%	98.16%

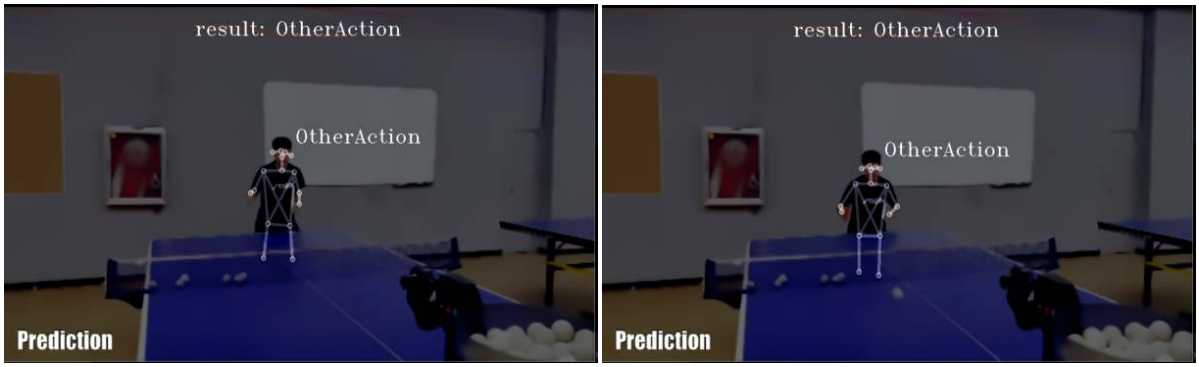
文章自建乒乓球击球动作骨骼数据集研究的是较为细粒度的人体动作，输入时空图卷积的数据时间序列长度为 16，时间序列上干扰因素较少，精准度较高；Kinetics-400 骨骼数据集是属于行为识别的范畴，输入网络的数据时间序列长度为 150，时间序列上的干扰因素较多。通过模型训练结果对比，乒乓球击球核心力量区域的骨骼点连通后能提高动作的识别精度。同时在 Kinetics-400 骨骼数据集上训练精度有所提升，可以表明乒乓球击球核心力量区域对其他人体行为也存在影响。

6.2. 模型泛化效果测试

文章对模型的识别效果和泛化效果进行测试，发现乒乓球的击球核心力量区域骨骼的连通性对动作识别的效果有很大的影响。例如图 5(a)中 ST-GCN 原网络实际结果应是其他动作这一类，结果识别成了正手击球，而文章改进的模型识别结果符合实际，如图 5(b)所示，图片为连续动作 16 帧中截随机提取两帧动作。



(a) ST-GCN 网络模型泛化效果



(b) 文章改进后网络模型泛化效果测试

Figure 5. Comparison of generalization effect between ST-GCN network and network improved in this paper
图 5. ST-GCN 网络与文章改进后网络泛化效果对比

7. 结语

文章自建乒乓球击球动作骨骼数据集，通过研究得知乒乓球击球动作的核心力量区是肩关节以下、髋关节以上，从而在数据预处理部分对乒乓球击球核心区域骨骼关键点的连接进行调整，骨骼关键点连通调整也就改变了图卷积的卷积核，致使原 ST-GCN 卷积核的卷积视野扩大，能更好地学习到乒乓球核心力量区域骨骼点关系的特征。最终，改进卷积核视野的时空图卷积网络在 Kinetics-400 骨骼动作数据集上训练后，Top1 精度和 Top5 精度分别提升了 1.07%和 1.93%，在自建的乒乓球击球骨骼数据中精准度为 98.16%，比原时空图卷积效果提升了约 8%的精准度。通过对模型实际应用效果进行测试，调整后骨骼连接的模型识别乒乓球动作更符合自建数据集的动作要求。下一步将在此数据上扩增其他乒乓球击球动作，进一步提高乒乓球击球动作识别模型的实际应用可能性。

基金项目

大连市科技创新基金项目“面向足球青训的技战术分析算法及配套可穿戴设备研发”(项目编号：2020JJ26GX038)；大连民族大学学科团队项目“基于机器学习的乒乓球接发球动作识别与水平评估算法研究”。

参考文献

[1] Kipf, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks. arXiv:1609.02907 [cs.LG]

[2] Yan, S., Xiong, Y. and Lin, D. (2018) Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. arXiv:1801.07455 [cs.CV]

[3] Simonyan, K. and Zisserman, A. (2014) Two-Stream Convolutional Networks for Action Recognition in Video. arXiv:1406.2199 [cs.CV]

[4] Tran, D., Bourdev, L., Fergus, R., *et al.* (2015) Learning Spatiotemporal Features with 3d Convolutional Networks. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 4489-4497. https://doi.org/10.1109/ICCV.2015.510

[5] Donahue, J., Anne Hendricks, L., Guadarrama, S., *et al.* (2017) Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 677-691. https://doi.org/10.1109/TPAMI.2016.2599174

[6] Shao, D., Zhao, Y., Dai, B., *et al.* (2020) FineGym: A Hierarchical Video Dataset for Fine-Grained Action Understanding. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 13-19 June 2020, 2616-2625. https://doi.org/10.1109/CVPR42600.2020.00269

[7] Zhe, C., Simon, T., Wei, S.E., *et al.* (2017) Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, HI, 21-26 July 2017, 1302-1310.

[8] 人工智能小技巧. Github 开源人体姿态识别项目 OpenPose 中文文档[Z/OL]. https://www.jianshu.com/p/3aa810b35a5d, 2018-11-11.

[9] Kip, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks. arXiv:1609.02907 [cs.LG]

[10] 日知. 如何理解 Graph Convolutional Network (GCN) [Z/OL]. https://www.zhihu.com/question/54504471/answer/611222866, 2019.

[11] Yan, S., Xiong, Y. and Lin, D. (2018) Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. arXiv:1801.07455 [cs.CV]

[12] 李亚娇. 核心力量训练在乒乓球教学中的应用[J]. 宿州教育学院学报, 2019, 22(3): 114-116.