

文章编号: 2096-1618(2022)05-0538-06

基于深度学习的体能测试计数算法改进与应用

杨磊, 李英祥, 张洪波

(成都信息工程大学通信工程学院, 四川 成都 610225)

摘要:为改变体能测试中人工计数的传统方式,提出一种基于深度神经网络的体能测试计数算法。该算法以运动视频作为输入,采用轻量级的姿态估计网络(lightweight-openpose)对人体关节点坐标进行检测,改进原模型的关节漏检和无效计算问题并新增了人物追踪模块;使用图像分类网络进行运动类型预测,采用批量正则化和迁移学习方式强化网络在自制4类常见体能测试数据集上的训练;制定4类体能测试运动计数标准,结合姿态估计和运动识别结果进行有效运动判断并计数。实验表明,姿态估计模型相较于原模型运行速度提升17.6%,人物追踪错误率低至3.2%;运动识别模型准确率提升5%~8%,达到94.84%;运动计数算法平均准确度在4类体能测试运动中达到95%。

关键词:体能测试;lightweight-openpose;模型改进;运动识别;运动计数

中图分类号:TP183

文献标志码:A

doi:10.16836/j.cnki.jcuit.2022.05.009

0 引言

体能测试是监测在校学生身体健康的重要手段,目前主要通过体育老师人工进行评估和计数,存在效率低下和标准难以准确执行等问题。因此,本文旨在设计基于深度学习的体能测试评估与计数算法。在本算法中,需要对体能测试视频中人物的关节点进行检测、运动类型进行识别,并针对不同的运动制定评价标准,判断有效运动并计数,以提高体能测试的检测效率与准确度。

当前,国内外学者对人体运动姿态识别与分析方法进行了一定程度的研究。满蔚仕等^[1]采用一种通过雷达回波模型和同步挤压小波变换(SSTAP)方法获得人体主要部位的特征,再通过同步挤压小波变换获得人体回波信号并解析为运动姿态信息。张永强^[2]设计了一种基于Hu不变矩阵特征优化的人体运动姿态识别算法,提取连续二维图像的7个Hu不变举证,并输入到SVM中训练最终完成对运动姿态识别。Gao L等^[3]设计了一种无线穿戴运动捕捉系统,通过姿态数据库中的姿态信号特征序列和后续三层识别算法精确地识别各种运动,但是硬件成本太高。Zhao L等^[4]采用惯性传感器技术对篮球运动中的运球、传球、接球和投篮4种姿势进行识别。Thar M C等^[5]基于Openpose对关节信息进行检测,并计算与标准动作之间的差距,从而完成对瑜伽的自学训练,但是该方法应用场景较为单一。综上所述,大部分研究基于传统方法和

深度学习模型,存在成本较高、准确率低和检测速度较慢等问题。

针对上述问题,本文设计了基于轻量级的深度学习网络的体测识别与计数算法,有效改善了体测运动识别与计数效率与准确率问题。该算法结构如图1所示,输入为视频中的每帧RGB图像,主要包括姿态估计、分类网络和计数3个主要模块。姿态估计模块能够对图像中的人物关节点进行检测和追踪,输出为视频中每个人物的关节点坐标时域信息;分类网络能够根据图像中人物的动作进行运动识别;计数模块结合上述模块的计算结果与运动评价标准完成对视频中体测运动的有效判断与计数。

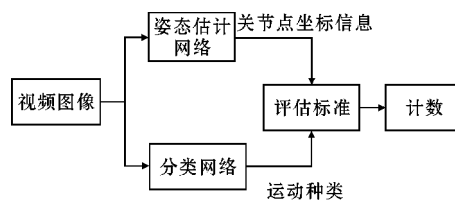


图1 算法结构图

1 姿态估计模块与改进

Openpose作为一种自底向上的姿态估计方法,在诸多领域得到了广泛的应用。本文采用Danil Osokin^[6]提出的一种对Openpose改进的姿态估计网络——lightweight-openpose作为姿态估计模块的基本框架。在该网络结构中,首先将轻量级的网络Mobilenet作为特征提取网络。在后续的网络中,Heatmap和PAFs预测采用公用的网络结构,在最后两层中,分别完成对Heatmap和PAF的预测。同时,将Openpose

网络中 7×7 的卷积核替换成两个 3×3 的卷积核,这样既大幅减少了网络的参数,同时也具有相同的感受野^[7],最后在 3×3 卷积块中加入了残差结构,以解决网络变深带来的影响。网络结构示意图如图 2 所示。

本文根据体能测试计数实际应用场景做了如下的改进:(1)通过计算相邻帧所有关节组的相似度,对视频中多个人物进行区分与追踪;(2)对由于遮挡导致关节检测不全的情况进行优化;(3)减少模型无效关节点和区域的计算。

1.1 人物追踪

在体能测试计数的场景下,常常会多人出现在一张图像中的情况。为便于后续流程对关节信息逐帧、逐人的分析,本文设计了人物追踪算法^[8]。如果一张图片中存在多个人,即多个姿态(关节点组)的时候,会遍历上

一帧中的所有姿态信息,对每一个姿态中的所有关节进行相似度计算。如果相似关节数大于阈值,即认为这相邻两帧的姿态属于同一个人。相邻帧中的两个姿态的相似度计算公式如下:

$$J_i = \begin{cases} 1 & \text{if } \exp\left(-\frac{d}{2(a+\varepsilon)}\right) \geq \theta \\ 0 & \text{if } \exp\left(-\frac{d}{2(a+\varepsilon)}\right) < \theta \end{cases} \quad (1)$$

$$N = \sum_{i=0}^{i=n} J \quad (2)$$

式中, d 为两个姿态所有关节点距离之和, a 为两个姿态框面积的最大值, n 为关节点个数, ε 为一个较小的实数,其中:

$$d = \sum_{i=0}^{i=n} (k_a^i - k_b^i)^2 \quad (3)$$

$$a = \max(\text{box}_a, \text{box}_b) \quad (4)$$

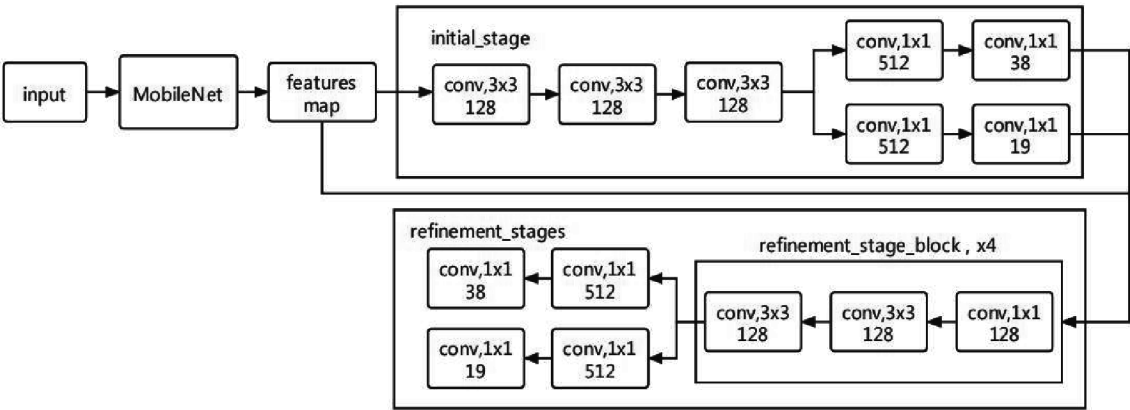


图 2 lightweight-openpose 网络结构示意图

1.2 关节点补全

在体能测试场景中,由于障碍物或者拍摄角度的问题,视频图像中的人物往往存在部分关节不全或者重叠的问题,导致模型无法对关节进行有效的预测,对后续运动分析模块造成较大的影响。对于上述问题,

文中采用人体测量学的方法对漏检的人体进行补全^[9]。具体方法为:对于身体对称部位,只需要获取其中一半的身体部位关节点即可通过以身体中轴线映射获取另一半关节的坐标信息;对于其他特殊部位,如鼻子,其坐标应该位于两肩中点的延长线上,长度可以通过其他可见的帧进行获取。改进效果如图 3 所示。



(a) 头部关节点漏检



(b) 头部关节点补全

图 3 关节点补全效果对比图

1.3 模型优化

原始模型共有 18 个关节点,但是在体能测试计数的场景中,部分头部关节点的坐标信息没有实际作用,反而加重了整个模型的运算负担^[10]。因此,将原始模型的关节点删减,部位包括左右耳朵、左右眼睛和脖子,以减轻模型的计算负担。

同时注意到在体能测试的场景中,可能会出现意

外入境的人物,原始模型往往会对其进行检测,但这却是无效的计算,因此需要抛弃。在改进的模型中,基于检测到的姿态对象,会生成一个检测框属性,该属性是包含所有关节点的最小面积的长方形框。为避免对无效区域的计算,对检测框区域面积所占比例小于整体图像的检测区域进行舍弃,从而进一步减少模型的计算量。模型改进效果对比如图 4 所示。

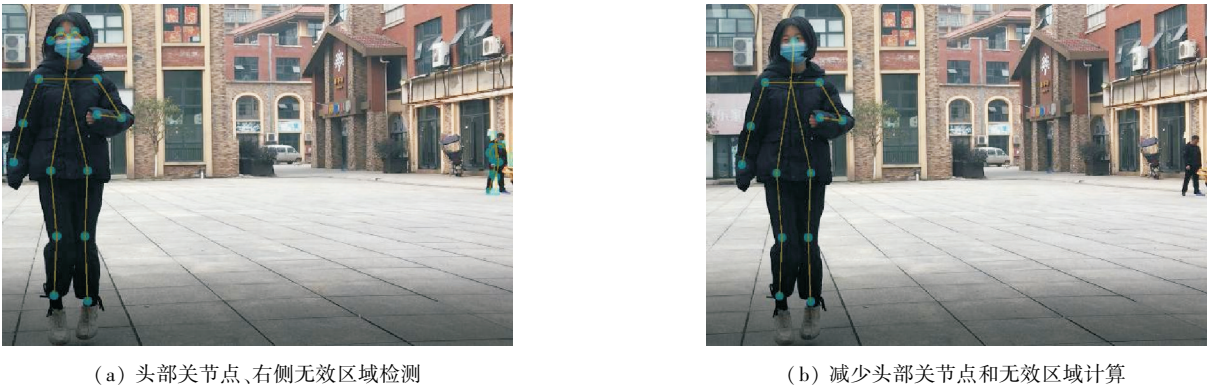


图 4 模型优化效果对比图

2 运动识别模块

2.1 数据集

针对常见的 4 种体能测试运动:引体向上、俯卧

撑、下蹲和仰卧起坐制作数据集。数据集图片的来源包括 bilibili、youtube、腾讯视频等视频网站;百度图片、必应图片、谷歌图片及搜狗图片的搜索引擎,通过爬虫和视频取帧等技术获取运动图片。该数据集共有 4 个分类、6160 张图片,随机分配测试集和训练集的比例为 3 : 7,部分数据集如图 5 所示。

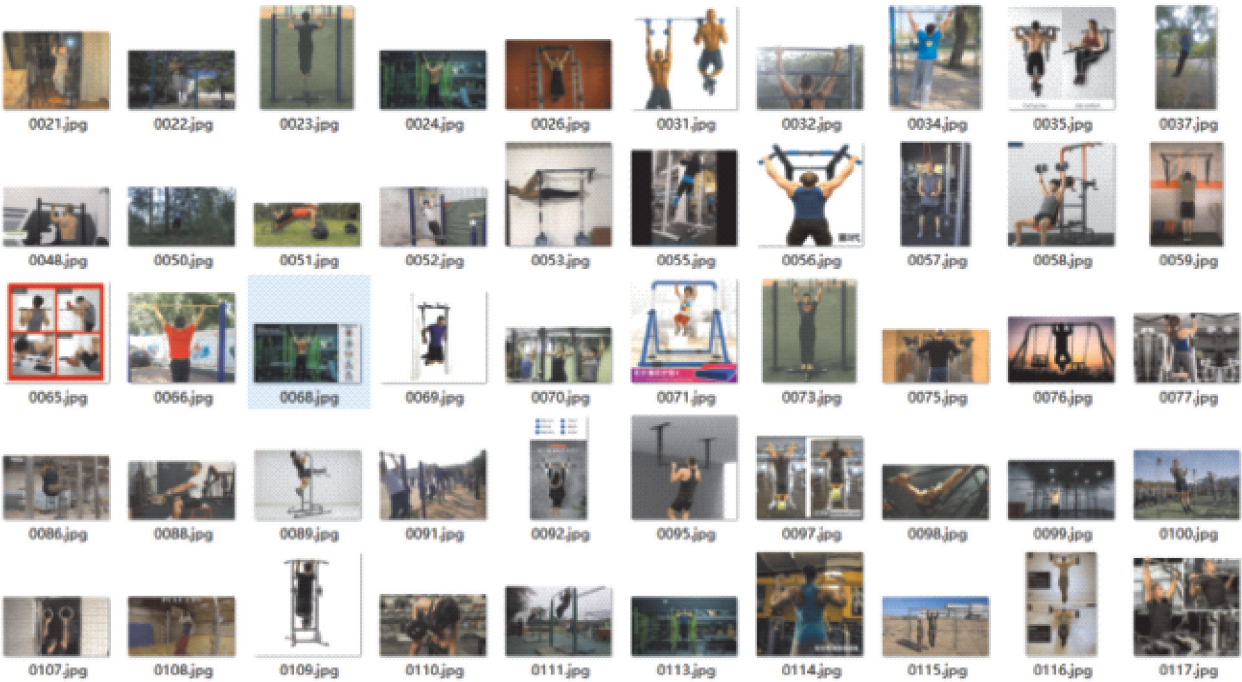


图 5 部分数据集

2.2 运动识别

由于不同的体测运动具有不同的评判和计数标

准,因此运动视频或者图片需要输入到分类网络,对运动种类进行预测^[11]。考虑到实际场景中,改变运动类型的时间间隔普遍较长,因此采取抽样的策略进行预

测。根据上述功能需求,选择了常见的分类网络,包括 ResNet、MobileNet^[12]等。首先,对各个原始网络在数据集上进行训练,但是预测准确率表现不佳。因此,对训练过程做如下优化:(1)对数据进行批量正则化(batch normalization)处理,目的是使一批输入图像的 Feature Map 满足均值为 0、方差为 1 的分布规律,加速网络的收敛并提升准确率^[13]; (2)在训练过程中使用迁移学习^[14]的方法,通过官方提供的 ImageNet 数据集上的预训练模型,使在较少的数据集上得到更好的准确率。通过训练与测试,经过上述对训练过程的改进,各个网络的预测精度均提升了 5%~8%。

3 体能测试计数模块

在体能测试和日常训练中,如果关键关节或肢体伸展不够,不仅不能达到预期的锻炼效果,而且在体能测试过程中不能算作有效运动。因此,体能测试与计数模块需要对视频中人物的运动有效性进行判断,并记录有效的运动次数。具体方法为:根据姿态估计网络获得运动中各个关节坐标的时域信息,并以此计算各个关键肢体弯曲角度的时域信息,从而判断关键部位的伸展情况,计算有效的运动次数,并通过图表的方式进行可视化。本文涉及的 4 种运动及其评价方式为:俯卧撑需要关注手臂和臀部的弯曲角度;引体向上通过手臂拉伸角度判断是否有效完成运动;仰卧起坐通过整个身体在臀部的弯曲程度判断是否有效运动;下蹲运动需要膝盖处角度判断是否下蹲到位。计算公式如表 1 所示,其中 $\angle \text{elb}$ 、 $\angle \text{hip}$ 和 $\angle \text{kne}$ 分别代表手肘、臀部和膝盖的角度, θ 和 α 代表允许的误差范围。

表 1 各运动评价标准	
运动名称	评价标准
俯卧撑	$\frac{\pi}{2} \pm \theta < \angle \text{elb} < \frac{5\pi}{6} \pm \theta$
	$\pi + \theta < \angle \text{hip} < \pi - \theta$
引体向上	$\frac{\pi}{3} \pm \theta < \angle \text{elb} < \frac{5\pi}{6} \pm \theta$
仰卧起坐	$\frac{\pi}{4} \pm \theta < \angle \text{hip} < \frac{2\pi}{3} \pm \theta$
下蹲运动	$\frac{\pi}{2} \pm \theta < \angle \text{kne} < \pi \pm \theta$

4 实验过程与分析

4.1 姿态估计算法

在人物追踪算法实验过程中,关键点在于对参数 θ 、 N 的选择。其中, θ 为某相邻帧、同一个关节的相似

程度,而 N 代表相似关节的个数。实验选择 lightweight-openpose 官方提供的预训练目标检测模型,对视频中图像进行关节点预测与人物追踪,并对保存的帧图像进行验证。实验发现,错误的参数设置可能导致将单个人检测成多人的情况。 θ 、 N 的选择与人物追踪错误率如表 2 所示,由表可见当 $\theta=0.5$ 、 $N=3$ 时算法有最佳的表现。

表 2 θ 、 N 选择对于人物追踪的正确率影响		
θ	N	错误率/%
0.3	3	6.4
0.4	3	5.1
0.5	3	3.2
0.5	2	6.2
0.5	4	8.1

基于上述改进的模型,继续对关节点数量和对无效检测区域进行优化,对头部无效计算的耳朵、眼睛和脖子等关节点进行删减。无效检测区域的比例设置为 10%,实验图片数量为 1000 张,平台为 NVIDIA GTX 1080 GPU,模型单帧平均耗时如表 3 所示,改进模型在运行速度上提升了 17.6%,达到约 36 帧/秒。

表 3 姿态估计模型改进前后耗时		单位:s
原始方法	优化方法	
0.034	0.028	

4.2 运动识别

运动识别模块使用 Pytorch 框架搭建分类网络,包括 ResNet 和 MobileNet 家族常用网络,数据集采用自制数据集,共 6160 张图片,随机分配验证集和训练集,比例为 3:7。训练过程中,对数据的预处理包括:(1)随机剪裁至大小为 224×224 ; (2)以概率为 0.5 随机水平旋转^[14]; (3)归一化处理。训练平台为 Nvidia GTX 1080 GPU,迁移学习预训练模型采用 Pytorch 官网提供的模型, batch_size 设置为 16,优化器选择 Adam 优化器,损失函数选择交叉熵损失函数,模型的平均运行速度如表 4 所示,各个网络在该分类任务上的精度如表 5 所示。

表 4 各个模型运行速度		单位:s
模型名称	单帧平均运行时间	
ResNet34	0.01613	
ResNet50	0.01845	
MobileNet_v2	0.01591	
MobileNet_v3_large	0.01761	

表 5 各个模型在验证集上的准确率

模型	平均	俯卧撑	引体向上	仰卧起坐	下蹲运动
resnet34	0.8987	0.9239	0.9191	0.8900	0.8621
resnet34 *	0.9386	0.9528	0.9526	0.9327	0.9163
resnet50	0.8817	0.9239	0.8930	0.8900	0.8201
resnet50 *	0.93735	0.9528	0.9461	0.9381	0.9121
mobilenet_v2	0.8690	0.9291	0.8956	0.8591	0.7992
mobilenet_v2 *	0.9365	0.9475	0.9530	0.9210	0.9247
mobilenet_v3_large	0.8610	0.8924	0.9008	0.8351	0.8159
mobilenet_v3_large *	0.9484	0.9633	0.9661	0.9447	0.9196

注：* 代表对训练过程进行优化后的结果

从表 5 可以看出,由于数据集的限制,各个模型在验证集上的准确率均不高,但是通过迁移学习和正则化预处理的方法提高准确率约 5% ~ 8%。其中 MobileNet_v3_large 的准确率最高,达到了94.49%。但是结合表 4 中各个模型的平均运行速度,在实际应用中选择 MobileNet_v2 作为运动识别预测模型。

4.3 运动计数

采用各个视频网站收集和现场录制的体能测试运动视频片段作为输入,依次通过姿态估计网络和运动识别网络,结合有效计数标准,通过图表分析关键肢体角度的时域信息,并计算峰值滤除无效运动最终完成计数,部分实验结果如图 6 所示。实验过程中,误差范围 θ 设置为 10° ,并进行重复实验,实验结果与人工计数结果进行对

比,得到体能测试计数算法的准确率如表 6 所示。

表 6 4 类体能测试计数准确率

运动	测试量	正确量	准确率
引体向上	30	29	0.9667
俯卧撑	30	30	1.0
仰卧起坐	30	26	0.8667
下蹲运动	30	29	0.9667
平均	120	114	0.950

由图 6 可知,本算法能对 4 种常见的体能测试运动进行有效性区分并准确计数。在图 6(d)中,由于下蹲程度不够,导致部分运动未进行计数。从表 6 可以看到,体能测试算法平均准确率达到了 95%。但是,由于仰卧起坐运动中未对手臂和腿等关节进行计算,导致计数错误率较高。

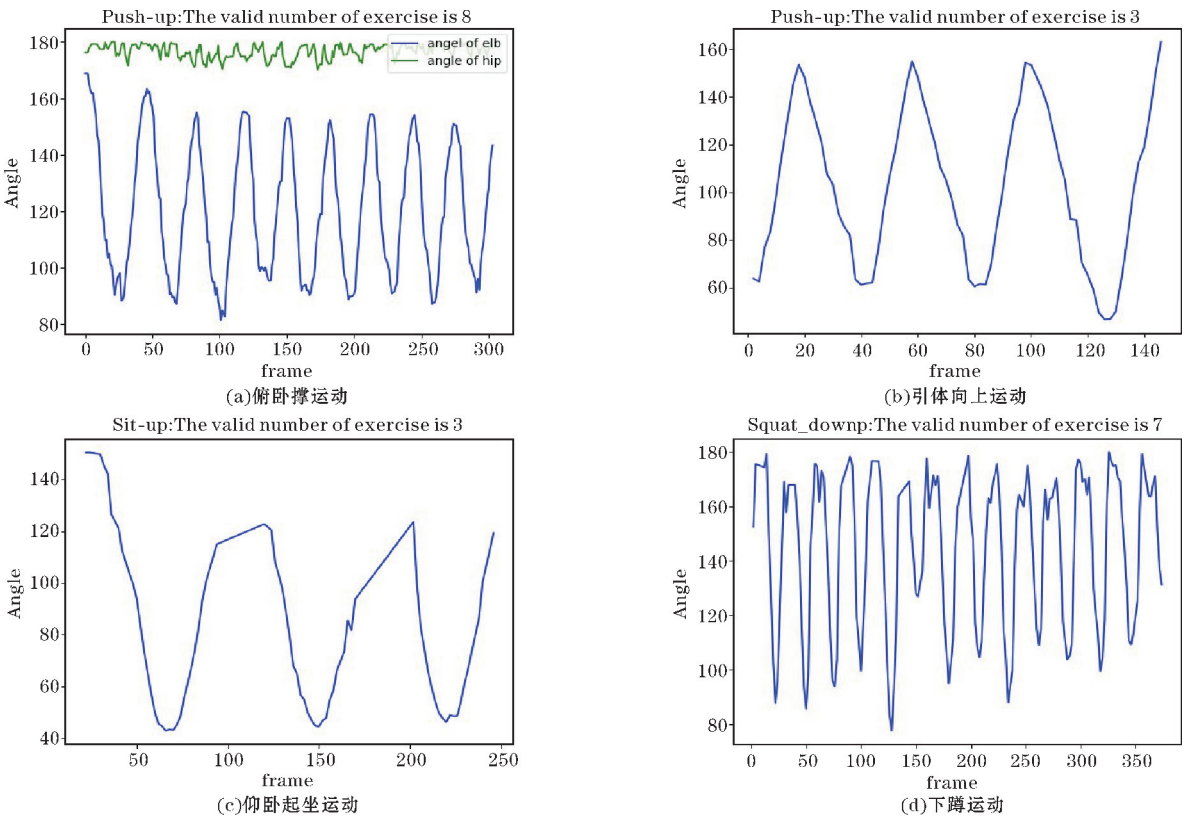


图 6 运动计数实验结果

5 结束语

本文提出的体能测试计数方法解决了人工计数标准难以执行和效率低的问题。本算法改进了姿态估计算法,并制作数据集训练分类网络,结合计数标准对有效运动进行计数。实验表明,姿态估计模块人物追踪错误率最低为3.2%,速率提升了17%;动作识别网络通过改进模型训练过程使得准确率达到94.84%;有效运动计数模块平均准确率达到95%。但是该方法也存在一定的缺陷:(1)运动识别数据集较小,且场景比较单一,导致分类网络预测精度较低;(2)体能测试运动有效性仅对部分关键部位的简单分析,并且实验数据量偏小。在后续研究中将针对上述问题,扩容数据集和实验数据,考虑更多的肢体评价指标,提升计数准确率。

参考文献:

- [1] 满蔚仕,朱宗耀,张志禹,等.采用同步挤压小波变换的人体运动姿态分析[J].西安交通大学学报,2017,51(12):8-13.
- [2] 张永强.基于Hu不变矩特征优化的人体运动姿态识别算法[J].计算机科学,2014,41(3):306-309.
- [3] Gao L,Zhang G,Yu B,et al. Wearable human motion posture capture and medical health monitoring based on wireless sensor networks[J]. Measurement,2020,166(4):108252-108263.
- [4] Zhao L,Chen W. Detection and recognition of human body posture in motion based on sensor technology[J]. IEEJ Transactions on Electrical and Electronic Engineering,2020,15(5):766-770.
- [5] Thar M C,Winn K Z N,Funabiki N. A Proposal of Yoga Pose Assessment Method Using Pose Detection for Self-Learning[C]. 2019 International Conference on Advanced Information Technologies(IC-AIT). IEEE,2019:137-142.
- [6] Osokin D. Real-time 2D Multi-Person Pose Estimation on CPU: Lightweight OpenPose[J/OL]. https://arxiv.org/abs/1811.12004,2018,11,29.
- [7] Wu Z,Shen C,Hengel A. Wider or Deeper: Revisiting the ResNet Model for Visual Recognition[J]. Pattern Recognition,2016,90(2019):119-133.
- [8] Fang H S,Xie S,Tai Y W,et al. RMPE: Regional Multi-person Pose Estimation[C]. 2017 IEEE International Conference on Computer Vision(ICCV). IEEE,2017:2334-2343.
- [9] Zhe C,Simon T,Wei S E,et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE,2017:1302-1310.
- [10] 闫航,陈刚,佟瑶,等.基于姿态估计与GRU网络的人体康复动作识别[J].计算机工程,2021,47(1):12-20.
- [11] 苏超,王国中.基于改进OpenPose的学生行为识别研究[J].计算机应用研究,2021,38(10):3183-3188.
- [12] Sandler M,Howard A,Zhu M,et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE,2018:4510-4520.
- [13] Ioffe S,Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International conference on machine learning. PMLR,2015:448-456.
- [14] Priya D T,Udayan J D. Transfer learning techniques for emotion classification on visual features of images in the deep learning network[J]. International Journal of Speech Technology,2020,23(8):361-372.
- [15] Krizhevsky A,Sutskever I,Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Advances in neural information processing systems. 2012,25(4):1097-1105.

Improvement and Application of Physical Fitness Test based on Deep Learning

YANG Lei, LI Yingxiang, ZHANG Hongbo

(College of Communication Engineering, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: In order to change the traditional way of manual counting in physical fitness test, a physical fitness test counting algorithm based on deep neural network is proposed. The algorithm takes the motion video as the input, uses the lightweight-openpose network to detect the coordinates of human joints, improves the joint missing detection and invalid calculation of the original model, and adds a human tracking module. The image classification network is used to predict the types of sports, and the batch normalization and transfer learning are used to strengthen the training of the network on the self-made four kinds of common physical fitness test data sets. Four kinds of sports counting standards for physical fitness test are formulated, and valid test judgment and counting are carried out combined with the results of pose estimation and motion recognition. Experiments show that the running speed of the pose estimation model is improved by 17.6% compared with the original model, and the error rate of human tracking is as low as 3.2%; The accuracy of motion recognition model is improved by 5%-8%, reaching 94.84%; The average accuracy of the motion counting algorithm reached 95% in the four types of physical fitness test.

Keywords: physical fitness test; lightweight-openpose; model improvement; motion recognition; motion count