

11-17-2020

Research and Application of a Lightweight Real-time Human Posture Detection Model

Hongkun Zhu

1. School of Electrical Engineering, Nantong University, Nantong 226000, China; ;

Jiawei Yin

1. School of Electrical Engineering, Nantong University, Nantong 226000, China; ;

Wenyu Feng

1. School of Electrical Engineering, Nantong University, Nantong 226000, China; ;

Hua Liang

1. School of Electrical Engineering, Nantong University, Nantong 226000, China; ;

See next page for additional authors

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Research and Application of a Lightweight Real-time Human Posture Detection Model

Abstract

Abstract: The traditional OpenPose model has good accuracy but slow speed in human posture detection. *In order to accelerate the detection speed and reduce the model on condition of the detection precision, based on the traditional OpenPose model, the residual network with second-order term fusion is used to extract the low-level features, the weights of the trained model are pruned by the L1 norm weight, and an improved OpenPose model is proposed.* Experiments show that when the detection accuracy is approximately equal to original model, the model size reduces to about 8%, the parameters reduces by nearly 83%, and the detection speed increases by 5 times. The improved OpenPose model is applied to the physical fitness test of sit-ups, and the results show that the detection accuracy of the model can reach 97%.

Keywords

human pose detection, OpenPose model, Residual network system, weight pruning

Authors

Hongkun Zhu, Jiawei Yin, Wenyu Feng, Hua Liang, Minrui Fei, and Kun Zhang

Recommended Citation

Zhu Hongkun, Yin Jiawei, Feng Wenyu, Hua Liang, Fei Minrui, Zhang Kun. Research and Application of a Lightweight Real-time Human Posture Detection Model[J]. Journal of System Simulation, 2020, 32(11): 2155-2165.

一种轻量化实时人体姿势检测模型研究与应用

朱洪堃¹, 殷佳炜¹, 冯文宇¹, 华亮¹, 费敏锐², 张堃^{1*}

(1. 南通大学 电气工程学院, 江苏 南通 226000; 2. 上海大学 机电工程与自动化学院, 上海市电站自动化技术重点实验室, 上海 210053)

摘要: 传统的 OpenPose 模型在人体姿势检测方面精度较好但速度较慢。为了在保证检测精度的前提下加快检测速度、缩小模型, 在传统 OpenPose 模型基础上, 使用添加二阶项融合的残差网络提取底层特征, 再通过 L1 范数权值对训练后的模型进行权值修剪, 提出了改进型 OpenPose 模型。实验结果表明, 在检测精度大致等同原模型情况下, 模型大小缩小至约 8%, 参数减少近 83%, 检测速度提升约 5 倍。将改进的 OpenPose 模型应用到仰卧起坐体育健康测试中, 结果表明该模型对仰卧起坐动作检测精度达到 97%。

关键词: 人体姿势检测; OpenPose 模型; 残差网络; 权值修剪

中图分类号: TP273; TH89 文献标识码: A 文章编号: 1004-731X (2020) 11-2155-11
DOI: 10.16182/j.issn1004731x.joss.20-FZ0308

Research and Application of a Lightweight Real-time Human Posture Detection Model

Zhu Hongkun¹, Yin Jiawei¹, Feng Wenyu¹, Hua Liang¹, Fei Minrui², Zhang Kun^{1*}

(1. School of Electrical Engineering, Nantong University, Nantong 226000, China; 2. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai Key Laboratory of Power Station Automation Technology, Shanghai 210053, China)

Abstract: The traditional OpenPose model has good accuracy but slow speed in human posture detection. In order to accelerate the detection speed and reduce the model on condition of the detection precision, based on the traditional OpenPose model, the residual network with second-order term fusion is used to extract the low-level features, the weights of the trained model are pruned by the L1 norm weight, and an improved OpenPose model is proposed. Experiments show that when the detection accuracy is approximately equal to original model, the model size reduces to about 8%, the parameters reduces by nearly 83%, and the detection speed increases by 5 times. The improved OpenPose model is applied to the physical fitness test of sit-ups, and the results show that the detection accuracy of the model can reach 97%.

Keywords: human pose detection; OpenPose model; Residual network system; weight pruning

引言

随着对人工智能研究的不断深入, 人工智能技术的应用方向也变得越来越广泛。对人体姿势的检

测是重要的应用方向。它在行为检测、视频捕捉和计算机图形学等领域都有着广阔的应用范围和丰富的应用价值。

传统的人工智能检测人体姿势方法, 如 Hao Fang 等^[1]提出的 RMPE 模型、Kaiming He 等^[2]提出的 Mask R-CNN 模型, 它们主要采用自顶向下的检测方法, 即对于一张输入的待检测图像, 首先判断该图像中是否有人, 若是则再判断图像中每个人体所占据的具体区域, 最后再在这些具体区域上检



收稿日期: 2020-06-05 修回日期: 2020-07-14;
基金项目: 江苏省六大人才高峰项目(XNY-039), 江苏省高等学校自然科学研究重大项目(19KJA350002), 国家级大学生创新创业训练计划(202010304065Z);
作者简介: 朱洪堃(1996-), 男, 河南, 硕士生, 研究方向为计算机视觉。

<http://www.china-simulation.com>

• 2155 •

测人体关键部分,还原人的整体姿势。目前,传统模型在 MPII 数据集的多人姿势数据集上的平均精度(mean Average Precision, mAP)到 76.7%。然而,这种传统方法在实际生产生活中并没有广泛应用,它主要具有如下缺点:(1) 模型太大,参数过多,检测速度慢;(2) 图像中人数越多,检测速度越慢;(3) 如果图像中人体部分区域超出图像范围或被遮挡较多,检测准确率就会明显下降。

CAO Zhe 等^[3]提出的 OpenPose 模型,有别于传统检测方式,采用了自底向上的检测方法,即对于一张输入的待检测图像,先检测图像中人体部位的关键部位,再使用局部区域亲和算法(Part Affinity Fields, PAFs)^[4]将这些关键部位连接起来,从而弥补了传统检测方式的不足之处。它具有如下优点:(1) 检测速度相较于传统模型有明显的提升;(2) 检测精度鲁棒性强,对一些多人且姿势复杂的图像也有较好的检测精度;(3) 图像中的人数并不会显著影响检测速度。到 2018 年为止,该模型检测能力的各项性能指标大部分都是最好的,但若想将其应用于实际场景中,它仍存在着检测实时性差、模型参数过多、模型太大等不足之处。

国内也有部分学者对 OpenPose 模型做了进一步的研究。在模型改进方面,周意乔等^[5]将 OpenPose 模型与双向长短时记忆网络(long short-term memory, LSTM)相结合,有效提高了复杂环境下人体姿势识别的准确率;周德良等^[6]先使用 OpenPose 模型获取人体关键部分坐标,再用 BP 神经网络进一步迭代优化该坐标,实现了检测精度的整体提高。在模型应用方面,郭园等^[7]运用 OpenPose 模型及非接触式 Kinect 运动捕捉系统分析中小学生的学习坐姿随时间的变化,以此指导学习桌的设计;彭秋萍等^[8]将 OpenPose 模型应用于自动扶梯摔倒监测上,在模型识别到乘客摔倒行为后,通过无线通讯方式启动扶梯急停装置,可及时有效的保护乘客安全。

本文在 OpenPose 模型的基础上加以改进,提出了改进型 OpenPose 模型,旨在保持检测精度与原模型大致相等的情况下,加快检测速度,降低参

数数量,缩小模型大小,以便于实际应用。本文对原模型的主要修改如下:(1) 将原模型提取底层特征所用的 VGG-19 网络^[9]替换为添加二阶项融合的残差网络^[10],以提升其检测速度和检测精度;(2) 对训练后的模型进行权值修剪,以压缩模型和提高检测速度。此外,本文还以所提出的改进型 OpenOpse 模型为核心,设计并测试了一套仰卧起坐测试系统。该系统硬件结构简单,人体姿势检测速度快,检测精度较高,具有较强的实用性。

1 OpenPose 模型介绍

1.1 OpenPose 模型处理流程介绍

OpenPose 模型本质上是一种双并行卷积网络模型。它同时使用 2 个卷积网络,一个卷积网络从图像中定位人体关键部位的位置,另一个卷积网络则用于连接候选关键部位组成肢体,再将这 2 个卷积网络的结果汇总进行姿势组装,完成对图像中人体姿势的检测。显然,这种双并行卷积网络的构造会极大的消耗计算资源,因此在 OpenPose 模型的实现中,是先使用一个卷积网络初步提取图像特征,再将该特征分别输入至双并行卷积网络中进行后续处理,等同于将双并行卷积网络较底层的部分合并为一个卷积网络,以节约计算资源。

图 1 展示了 OpenPose 模型的检测流程,图 2 是该流程的可视化。首先,使用 VGG-19 网络^[9]提取输入图像的底层特征,如图 2(a)所示是其中部分卷积层的输出;然后,底层特征分别输入到两个并行卷积网络中,其中一个网络使用非极大值抑制算法(Non-maximum suppression)^[11]生成置信度图,用于人体关键部位的定位,图 2(b)是该过程中膝盖和脚踝部位的置信图;另一个网络使用局部区域亲和算法,提供关键部位之间的连接方式以组成肢体,图 2(c)是该过程中左小腿的检测效果;之后,汇总这 2 个卷积网络的处理结果,采用匈牙利算法(Hungarian algorithm)^[12],进行姿势组装,最终输出待测图像中的人体姿势,图 2(d)、(e)分别是姿势组装示意图和最终检测结果。

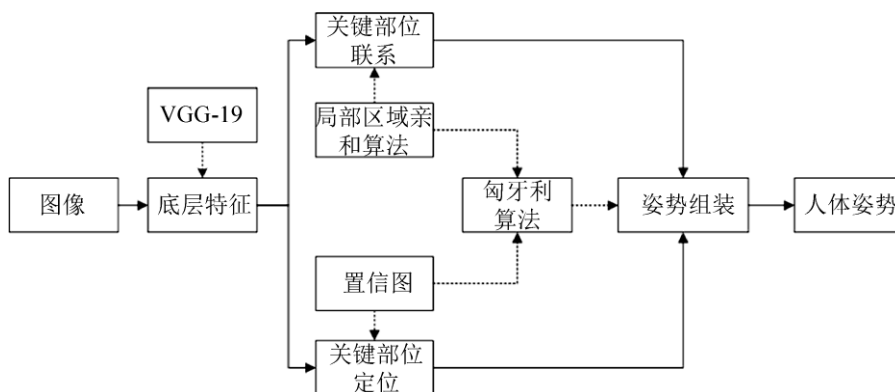


图 1 OpenPose 模型处理流程
Fig. 1 OpenPose model processing flow

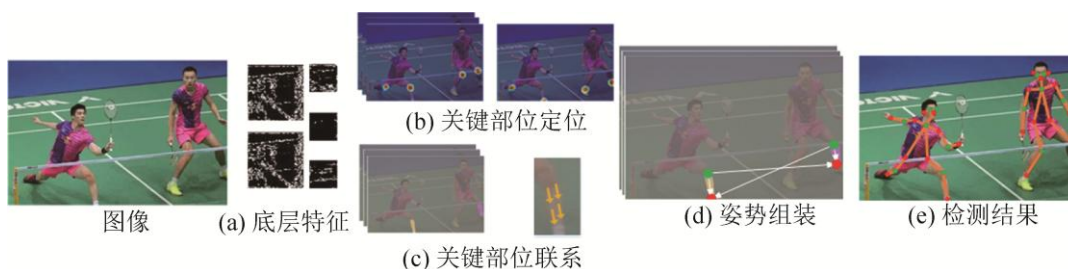


图 2 OpenPose 模型处理可视化展示
Fig. 2 OpenPose model handles visual presentation

1.2 OpenPose 模型主要算法介绍

1.2.1 局部区域亲和算法

局部区域亲和算法的作用是计算候选关键部位相互连接组成候选肢体的置信度。

如图 3 所示, 设 j_1, j_2 是 2 个不同的关键部位, \mathbf{v} 为从 j_1 指向 j_2 的单位向量, \mathbf{v}_T 为向量 \mathbf{v} 的垂直向量, 肢体 c 的长度和宽度分别设为 $l_c, \sigma_c, X_{j_1}, X_{j_2}$ 分别表示候选关键部位 j_1, j_2 的位置坐标。

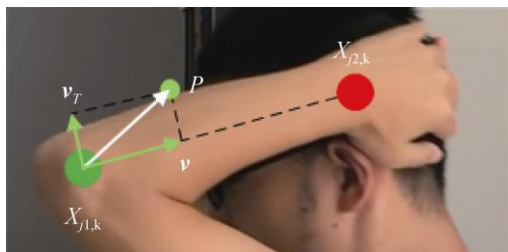


图 3 局部区域亲和算法示例
Fig. 3 Example of PAFs

首先, 如公式(1)所示, 计算图像上某一点 p 在肢体 c 上的置信向量 $A_c(p)$ 。如果点 p 在肢体 c

上, 则 $A_c(p)$ 就等于肢体 c 的单位方向向量 \mathbf{v} , 否则为零向量。公式(2)~(3)是判断点 p 是否在肢体 c 上的公式, 若均成立则点 p 在肢体 c 上, 否则不在。

$$A_c(p) = \begin{cases} \mathbf{v}, & p \text{ 在肢体上} \\ 0, & p \text{ 不在肢体上} \end{cases} \quad (1)$$

$$0 \leq \mathbf{v} \cdot (p - X_{j_1}) \leq l_c \quad (2)$$

$$|\mathbf{v}_T \cdot (p - X_{j_1})| \leq \sigma_c \quad (3)$$

之后, 就可用公式(4)得出肢体 c 的置信度 E_c , 它是计算关键部位 j_1, j_2 之间连线上所有点的置信向量的定积分得来的。

$$E_c = \int_0^1 A_c(p(u)) \frac{X_{j_2} - X_{j_1}}{\|X_{j_2} - X_{j_1}\|_2} du \quad (4)$$

1.2.2 匈牙利算法

在姿势组装过程中, 由于输入图像中人数未知, 则不能简单的采用遍历全部姿势组装方案, 找出候选肢体置信度总和最大的方案, 因为这是一个 NP 问题, 会占用很多计算资源。因此, OpenPose 模型采用最大二分图匹配法找出可能的姿势组装

方案,再寻找其中候选肢体置信度总和最大的方案。

匈牙利算法^[12]就是 OpenPose 模型用于完成最大二分图匹配的算法。假设现有 3 种关键部位,它们的可能连接方式如图 4 所示。

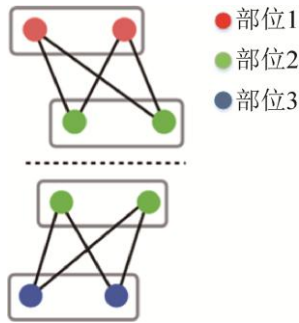


图 4 匈牙利算法示意

Fig. 4 Diagram of Hungarian algorithm

匈牙利算法计算过程为:

首先,由于最大二分图匹配法不允许多边共用一个节点的情况(即一个手臂不可能连接 2 个手腕),所以有 2 个约束条件,如公式(5)、(6)所示:

$$\forall m \in D_1, \sum_{n \in D_2} Z_{1,2}^{m,n} \leq 1 \quad (5)$$

设使用 $Z_{1,2}^{m,n} = 1$ 表示部位 1 类第 m 个关键部位与部位 2 类第 n 个关键部位具有连接的可能性,则对于任意的部位 1 类第 m 个关键部位,所有与此关键部位相连的部位 2 类关键部位的置信度之和必须不超过 1,否则代表与此关键部位相连的部位 2 类关键部位超过了 1 个,这种连接方式是违规的。

$$\forall n \in D_2, \sum_{m \in D_1} Z_{1,2}^{m,n} \leq 1 \quad (6)$$

公式(6)含义与公式(5)相同,是对部位 2 类第 n 个关键部位的约束。

$$\max_{Z_c} E_c = \max_{Z_c} \sum_{m \in D_1} \sum_{n \in D_2} E_{mn} Z_{1,2}^{m,n} \quad (7)$$

随后,将所有满足约束条件的可能的连接方式,使用公式(7)找出最大的积分和,则其所对应的连接方式就是最可能的连接方式,于是就找到了某一条肢体两端所对应的关键部位,对其他肢体重复以上步骤,最后将共用的同一个关键部位的肢体组装在一起,就实现了多人姿势的检测。

2 改进方式

本文所提出改进型 OpenPose 模型在检测流程上继承 OpenPose 模型思想,而在底层特征提取方式和模型压缩方面进行改进。

2.1 添加二阶项融合的残差网络提取底层特征

OpenPose 模型提取底层特征所用的 VGG-19 网络虽然是一种应用较为广泛的网络,但随着卷积层数的增加,该网络存在梯度消失、性能退化等缺陷。人工智能技术飞速发展,较早提出的 VGG 网络在性能上已经落后于新提出的网络。因此,本文将底层特征提取网络更换为 Res-18 网络^[13],以提升改进后模型的检测精度和检测速度。该网络参数如表 1 所示。

表 1 Res-18 网络参数
Tab. 1 Res-18 network parameters

层名	输出尺寸	卷积层参数
卷积 1	736×736	7×7×64 卷积
卷积 2	368×368	3×3 最大池化
卷积 3	184×184	2×[3×3×64]卷积
卷积 4	92×92	2×[3×3×128]卷积
卷积 5	46×46	2×[3×3×256]卷积

本文参考了 Wang 等^[14]所提出的 SORT (Second Order Response Transform)方法,该方法适用于一般分支结构和残差结构,仅增加 5% 的检测时间,就能有效提高网络性能。考虑到底层特征对整个模型预测精度的决定性作用,本文为 Res-18 网络中的残差块添加了一个二阶项 F'' 及一个较小的偏置量,以加强该网络的非线性拟合能力。

图 5 为添加了二阶项的残差块和一般残差块的对比,其输出 f 计算方式如公式(8)~(9)所示:

$$F'' = \sqrt{F(X \odot X') + b} \quad (8)$$

$$f = X + X' + F'' \quad (9)$$

式中: F'' 为二阶项; b 为其偏置量; F 为激活函数; \odot 为矩阵点乘。

此外,函数 $F \odot$ 必须可微分,这样才能对其进行梯度下降操作以优化参数,参考 SORT 方法中实

验所得结论, 将该函数设置为 $F(\bullet) = \sqrt{\bullet}$ 。

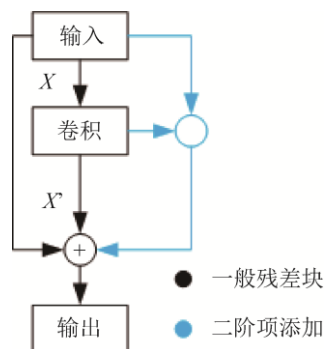


图 5 添加二阶项融合的残差块
Fig. 5 Add residual block of second-order term fusion

通常来说非线性越强, 模型的泛化能力就越好, 因此为了进一步加强模型的非线性, 本文还对 X, X' 使用激活函数进行处理。在目前常用的激活函数中, Leaky-ReLU 函数的非线性最强^[15], 因此本文使用 Leaky-ReLU 函数进行激活。如图 6 所示, 是传统残差块和本文所述添加二阶项残差块以 Leaky-ReLU 函数作为激活函数时的激活值对比。

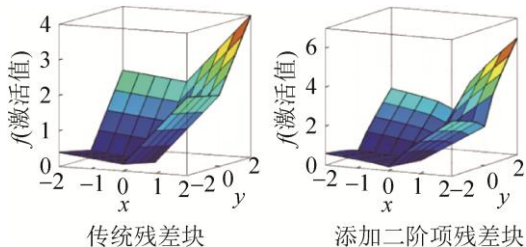


图 6 2 种残差块激活值对比
Fig. 6 Comparison of activation values of 2 residual blocks

2.2 权值修剪

由于本文所提的改进型 OpenPose 模型是一种密集网络, 将前层的所有特征作为后层的输入。文献[16]指出, 后层并非都需要将所有前层的输入作为输出, 有时过低层的特征并不起什么作用。因此, 本文使用权值修剪^[17]去除那些权值低于阈值的连接, 这样就能在保证检测精度的前提下, 大幅度缩减模型大小, 加快检测速度。

如图 7 所示, 是修剪前后的神经元模型对比示意图。本文权值修剪的范围为提取底层特征后的部分, 权值修剪过程如下:

- (1) 先用一半的训练集训练模型, 找出权值低于阈值的连接;
- (2) 用 0 权值填充这些连接;
- (3) 再用另一半训练集训练模型, 再次找出并填充权值低于阈值的连接;
- (4) 保存修剪后模型。

在训练过程中使用的权值计算方式见公式(10):

$$L(w) = \frac{1}{N} \sum_{i=1}^N (f(x_i; w) - y_i)^2 + \lambda \|w\|_1 \quad (10)$$

式中: $L(w)$ 为损失值; N 为样本数; $f(x_i; w)$ 为第 i 个样本的预测值; y_i 为第 i 个样本真实值; $\|w\|_1$ 为权重的 $L1$ 范数^[18]; λ 为该范数的系数。

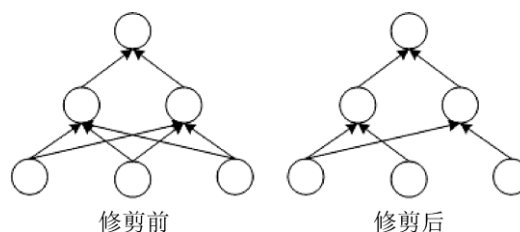


图 7 权值修剪对比
Fig. 7 Weight trim comparison

通过在损失函数后加上参数的 $L1$ 范数对参数 w 进行约束, 使其变成一个稀疏矩阵(即矩阵中大部分元素为 0), 从而完成对权重值的筛选。此外, 带有 $L1$ 范数的权值计算方式还可以在在一定程度上防止过拟合现象^[19]。

3 仰卧起坐测试系统的设计

为了验证所设计模型的实用性, 本文还设计了一套仰卧起坐测试系统作为改进型 OpenPose 模型的应用实例。

3.1 仰卧起坐测试系统的结构

该测试系统的硬件结构简单, 只需笔记本电脑, 摄像头及相应连接线。将摄像头放置在受测人员左侧面中间位置, 尽量使受测人员位于画面中间即可。

测试系统的软件结构如图 8 所示, 其中核心部分是根据改进型 OpenPose 模型的检测结果进行仰卧起坐计数。

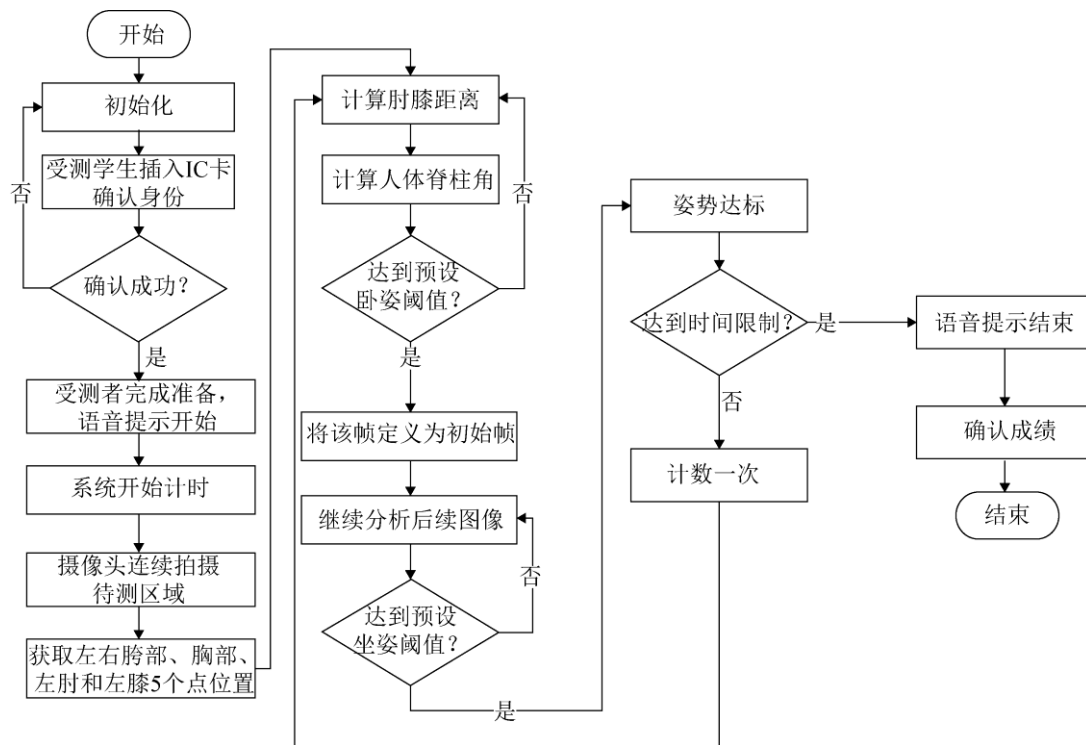


图 8 软件结构

Fig. 8 Software structure

表 2 是本文所提改进型 OpenPose 模型检测的 19 个人体关键部位, 关键部位以序号的形式存储, 分为可见、不可见和不在图中 3 种状态。但是, 在仰卧起坐测试系统的后续处理中并不需要使用所有关键部位的检测结果, 本文选择这 19 个关键部位的主要目的是: (1) 覆盖几乎全部人体关键部位, 在实验中能有效体现模型的检测效果; (2) 保留模型通用性, 方便应用于其他领域。

表 2 人体关键部位对应序号
Tab. 2 Serial number of key body parts

序号	名称	序号	名称	序号	名称
1	左耳	8	右肩	15	右髌
2	右耳	9	左肘	16	左膝
3	左眼	10	右肘	17	右膝
4	右眼	11	左腕	18	左踝
5	鼻子	12	右腕	19	右踝
6	胸部	13	肩部		
7	左肩	14	左髌		

该系统判断是否完成仰卧起坐方式如下: 在检测结果中挑选出左右肩部和胸部的 3 个关键点, 找

出胸部点到左右肩部两点连线的垂线, 将该线与水平线夹角定义为人体脊柱角; 挑选左肘和左膝 2 个关键点并连接起来, 将该连线长度定义为肘膝距离。通过对人体脊柱角和肘膝距离的变化分析, 即可得出受测人员属于准备阶段、起身阶段、坐姿阶段还是下躺阶段, 该阶段动作是否达标等等, 从而完成对仰卧起坐的自动计数。

如图 9(a)是系统在准备阶段中检测到受测人员姿势符合标准, 将在倒计时 5 s 后开始测试; 图 9(b)是开始测试后受测人员处于下躺阶段, 系统检测到该动作不符合标准, 则提示未达标并不计数。



(a) 准备阶段



(b) 下躺阶段

图 9 检测效果
Fig. 9 Test results

3.2 针对不同运动的优化

通常情况下,改进型 OpenPose 模型所检测的人体姿势均为人的全身姿势,而在体育测试系统的实际运行中,不少运动并不需要检测人的全身姿势。例如在引体向上测试中,动作的重点在于上半身。因此,在体育测试系统测试引体向上时,可以仅检测上半身部分关键点,即可显著减少运算量,进一步加快检测速度,使得检测实时性更高。

4 实验过程与数据分析

4.1 实验环境与评价指标

实验所用的训练平台使用 Google Colab 免费深度学习计算平台,该平台设备为 Tesla k80 24GB 显卡;测试平台使用 GTX-1080Ti 11 GB 显卡,其他型号数据可能有所偏差。

实验中模型所检测的人体关键部位共 19 个,其序号和类别如表 2 所示。实验所用的精度评价标准为平均精度(mean Average Precision, mAP),分别使用 50%, 75% 和 90% 作为每类关键点的检测精度阈值,最后计算其平均值 mAP 作为模型的平均检测精度;所用的速度评价指标是 FPS(Frame Per Second),表示平均每秒钟模型能处理的图像数量。

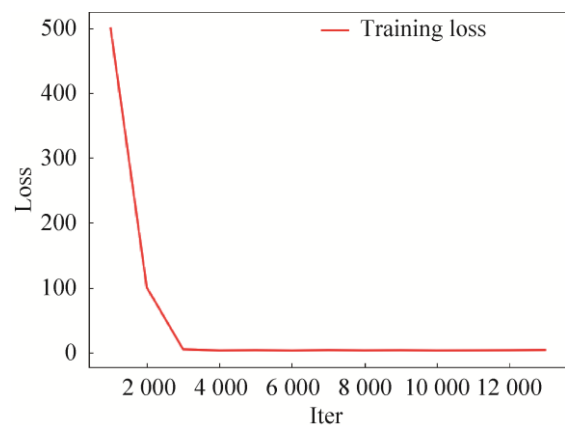
4.2 模型参数选择实验

4.2.1 训练数据集的选择

OpenPose 模型采用 COCO 数据集作为训练数据,文献[20]指出,该数据集具有以下 4 个问题:

(1) 未标注在图片区域外的关键部位;(2) 未标注遮挡情况下的关键部位;(3) 有些可见的关键部位未被标注;(4) 部分可忽略区域未被标注。此外,文献[20]还对 COCO 数据集进行了修正,并证明采用修正过后的数据集进行训练可加速模型收敛,提高模型检测精度。因此,本文采用该文献开源的修正后 COCO 数据集进行训练。

如图 10 所示,是本文所提改进型 OpenPose 模型在训练过程中损失曲线,模型共训练 13 000 轮,每训练 1 000 轮记录一次该 1 000 轮训练中的平均损失值。可见,训练 13 000 轮后模型已经达到收敛。

图 10 损失曲线
Fig. 10 Loss curve

4.2.2 底层特征提取方式对比

本文在 2.1 节中将 OpenPose 模型提取底层特征所用的 VGG 网络改进为二阶项残差网络。为了验证这项改进方式是否起到预期作用,本节实验中分别测试使用 VGG 网络、传统残差网络和二阶项残差网络提取底层特征时模型的检测性能。表 3 是这 3 种底层特征提取方式下模型性能对比。

表 3 3 种提取方式下模型性能对比
Tab. 3 Comparison of model performance under 3 extraction methods

提取方式	mAP/%	FPS/(f/s)
VGG 网络	77.2	11.8
残差网络	77.6	12.6
二阶项残差网络	77.9	12.1

由表 3 可得出, 采用残差网络代替 VGG 网络提取底层特征增加了模型的检测精度与速度; 对残差网络添加二阶项改进后则更进一步提升了模型的检测精度, 虽然其检测速度有所降低, 但仍比使用 VGG 网络的模型有所提高。因此, 本文采用添加二阶项的残差网络代替原模型所用的 VGG 网络提取底层特征。

如图 11 所示, 是改进前后模型对输入图像所提取的部分底层特征及对图中左肩部位检测置信图的对比。可见随着卷积层的加深, 改进后模型所提取的底层特征更加集中于关键部位分布密集的区域, 而原模型所提取特征分布则较为分散; 对比两者的左肩置信图可知, 相较于原模型, 改进后模型的置信度不仅分布较为集中, 而且消除了对称部位上的错误置信度分布。

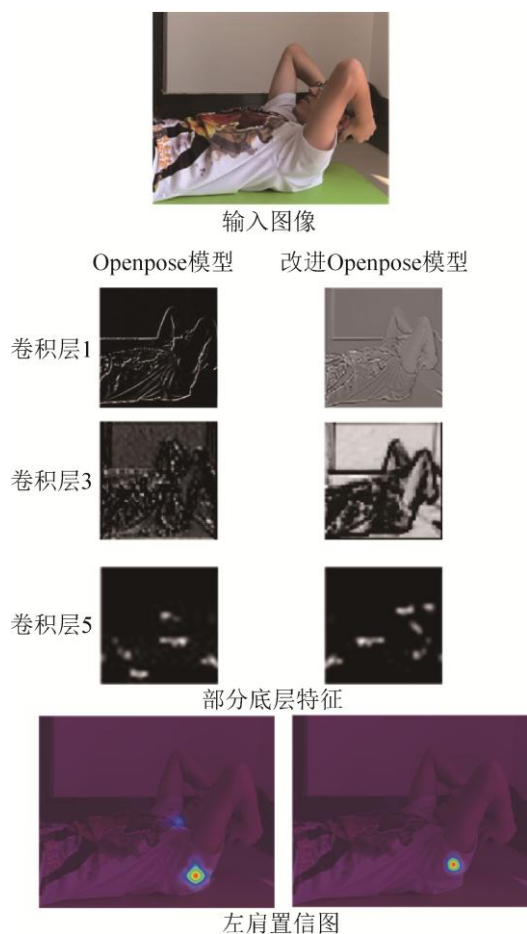


图 11 模型底层特征及置信图对比
Fig. 11 Comparison of low-level feature and confidence diagram

4.2.3 修剪程度的选择

本文 2.2 节中通过权值修剪^[17]方式对 OpenPose 模型进行压缩, 修剪范围为提取底层特征后的部分, 修剪程度为原始模型参数数量的百分比。本节实验中分别将修剪程度设置为 4.5%, 6%, 9% 和 15%, 并分别在测试它们与原始模型检测精度的差距, 实验结果如表 4 所示。

表 4 不同修剪程度下模型性能变化
Tab. 4 Model performance changes with different degree of pruning

参数数量/%	检测精度变化/%	检测速度变化/(f/s)
4.5	-6.1	+44
6	-2.3	+42
9	-0.9	+39
15	-0.0	+27

由表 4 可得, 模型检测精度随着模型参数数量的缩小而不断下降, 且随着参数数量一步步减少, 即使参数数量减少幅度降低, 但检测精度下降的幅度反而增大。

因此, 本文所提的改进模型选择将参数数量修剪至 9%。表中参数数量修剪至 15% 时所对应的检测精度变化值为 -0.0% 并非表示检测精度未发生变化, 而是其绝对值小于 0.1% 忽略不计。

4.3 模型间对比实验

本节实验内容为不同模型间性能对比, 所用的模型共 3 种, 分别为本文所提出的改进型 OpenPose 模型、OpenPose 模型和 Mask R-CNN 模型。实验分别从检测精度与检测速度对比和模型大小与参数数量对比两方面进行。在检测精度与检测速度对比试验中, 为了更好体现不同模型在不同情况下的对比, 本文将测试数据集分为单人数据集和多人数据集并分别进行测试。

表 5 是 3 个模型的检测精度和检测速度对比。一方面, OpenPose 模型相较于 Mask R-CNN 模型, 精度大约降低不到 1 个百分点, 但其检测速度对图像中人数不敏感(即对于多人图像的检测速度并不显著低于单人图像), 检测速度和泛用性都优于

Mask R-CNN 模型; 另一方面, 改进型 OpenPose 模型在单人数据集上检测精度仅比 OpenPose 模型低了 0.5%, 在多人数据集上则低了 2.5%, 但改进型 OpenPose 模型 FPS 相较于其他模型有明显增加, 约为另外 2 个模型的 5 倍, 而且同样具有检测速度对图像中人数不敏感的特性。可见, 改进型 OpenPose 模型在保持检测精度大致相同的情况下, 检测速度有了大幅度的提高。

表 5 不同模型检测精度与检测速度对比

Tab. 5 Comparison of detection accuracy and speed of different models

数据集	模型	mAP/%	FPS/(f/s)
单人	Mask R-CNN	78.5	11.5
	OpenPose	77.9	12.2
	Quick-OpenPose	77.4	60.8
多人	Mask R-CNN	76.7	8.6
	OpenPose	76.4	11.3
	Quick-OpenPose	73.9	56.9

如图 12 所示, 是以仰卧起坐姿势为例, 3 种模型检测效果对比。



Mask R-CNN



OpenPose



改进型 OpenPose

图 12 3 种模型检测效果对比

Fig. 12 Comparison of detection effects of the three models

由图 12 可见, Mask R-CNN 模型相较于其他模型多检测到双耳部位; OpenPose 模型和改进型 OpenPose 模型检测效果大致相同, 仅在关键部位的位置上有所区别, 但都属于位置正确范围内。而且, 图 11 中 3 种模型的检测精度均表现得比表 5 中对应的检测精度高, 这是由于 COCO 数据集包含一些视角较远、分辨率较低和姿势复杂的图像, 是人体姿势数据集中检测难度较高的数据集。

如表 6 所示, 是 3 个模型的模型大小与提取底层特征后的参数数量对比。不难看出, 改进型 OpenPose 模型在这方面具有明显优势, 其模型大小只有 OpenPose 模型^[3]的约 8%, 参数数量减少近 83%, 产生这种现象的原因主要是改进型 OpenPose 模型进行了权值修剪操作。

表 6 不同模型大小对比

Tab. 6 Comparison of different model sizes

模型	大小/MB	参数大小
Mask R-CNN ^[2]	218	27 432
OpenPose	199	29 124
改进型 OpenPose	17.8	4 918

综上所述, 改进型 OpenPose 模型检测精度与 OpenPose 模型和 Mask R-CNN 模型相差无几, 而检测速度有了大幅提升, 模型大小和参数数量显著减少, 实时性和泛用性更强。

5 结论

本文提出的改进型 OpenPose 模型, 可对人体关键部位进行检测, 并用连线的方式展现人的整体姿势。它是在对 OpenPose 模型^[3]进行充分研究, 总结其优缺点后进行修改的。本文主要解决了以下问题: (1) 大幅提高检测速度; (2) 显著缩小模型大小; (3) 以改进型 OpenPose 模型为核心, 设计并测试一套仰卧起坐计数系统; (4) 对比改进型 OpenPose 模型不同改进操作参数选择的影响。

但同时, 该系统也具有不足之处, 有些情况没有较好的解决: (1) 本文所提出模型在检测精确度上仍稍逊于 OpenPose 模型^[3], 尤其是多人姿势检

测上差距较大；(2) 本文所提出模型在一些复杂情况下错误率较大；(3) 本文所提出模型需要大量优质数据集进行训练，不适用于无大量优质训练集的其他任务。

针对以上的 3 个缺陷，可从下述方向进行改进：(1) 该问题可能是由于权值修剪所产生的，可尝试使用更优秀的修剪方式；(2) 可尝试优化肢体匹配原理；(3) 可尝试在训练时引入半监督学习、无监督学习研究领域内容。

参考文献：

- [1] Fang H, Xie S, Tai Y W, et al. RMPE: Regional Multi-person Pose Estimation [C]. The 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy: IEEE Press, 2017: 2353-2362.
- [2] He K, Gkioxari G, Dollár P, et al. Mask R-CNN [C]. The 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE Press, 2017: 2980-2988.
- [3] Cao Z, Simon T, Wei S, et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[C]. Computer vision and pattern recognition(CVPR). Puerto Rico, USA: IEEE Press, 2017: 1302-1310.
- [4] Wei S, Ramakrishna V, Kanade T, et al. Convolutional Pose Machines[C]. Computer Vision and Pattern Recognition(CVPR). Las Vegas, USA: IEEE Press, 2016: 4724-4732.
- [5] 周意乔, 徐昱琳. 基于双向 LSTM 的复杂环境下实时人体姿势识别[J]. 仪器仪表学报, 2020, 41(3): 192-201. Zhou Yiqiao, Xu Yulin. Real-time Human Posture Recognition in Complex Environment Based on Bidirectional LSTM[J]. Chinese Journal of Scientific Instrument, 2020, 41(3): 192-201.
- [6] 周德良. OpenPose 结合 BP 网络的行为检测方法[J]. 贵州大学学报(自然科学版), 2020, 37(3): 87-92. Zhou Deliang. Behavior Detection Method of OpenPose Combined With BP Network[J]. Journal of Guizhou University (Natural Science Edition), 2020, 37(3): 87-92.
- [7] 郭园, 郭晨旭, 时新, 等. 基于 OpenPose 学习坐姿分析的桌椅人机适应性研究[J]. 林业工程学报, 2020, 5(2): 179-185. Guo Yuan, Guo Chenxu, Shi Xin, et al. Study on Human-machine Adaptability of Tables and Chairs Based on OpenPose Learning Sitting Posture Analysis[J]. Journal of Forestry Engineering, 2020, 5(2): 179-185.
- [8] 彭秋萍, 万莉莉, 孙梦圆, 等. 基于机器视觉的扶梯自动急停系统[J]. 计算机系统应用, 2020, 29(3): 87-92. Peng Qiuping, Wan Lili, Sun Mengyuan, et al. Automatic Emergency Stop System of Escalator Based on Machine Vision[J]. Application of Computer System, 2020, 29(3): 87-92.
- [9] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. The IEEE Conference on Computer Vision and Pattern Recognition (ICCV). Boston, USA: IEEE Press, 2015: 1-9.
- [10] Costillareyes O, Verarodriguez R, Scully P J, et al. Analysis of Spatio-Temporal Representations for Robust Footstep Recognition with Deep Residual Neural Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (S0162-8828), 2019, 41(2): 285-296.
- [11] 刘望, 孙金玉, 马世伟. 基于时空特征金字塔网络的动作时序检测方法[J]. 系统仿真学报, 2019, 31(11): 2382-2387. Liu Wang, Sun Jinyu, Ma Shiwei. A Temporal Action Detection Algorithm Based on Spatio-Temporal Feature Pyramid Network[J]. Journal of System Simulation, 2019, 31(11): 2382-2387.
- [12] Liao Z, Wang J, Zhang S, et al. Minimizing Movement for Target Coverage and Network Connectivity in Mobile Sensor Networks[J]. IEEE Transactions on Parallel and Distributed Systems (S1045-9219), 2015, 26(7): 1971-1983.
- [13] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE Press, 2016: 770-778.
- [14] Wang Y, Xie L, Liu C, et al. SORT: Second-Order Response Transform for Visual Recognition[C]. International Conference on Computer Vision (ICCV). Venice, Italy: IEEE Press, 2017: 1368-1377.
- [15] 石敏, 姚瀚钦, 李淳芑, 等. 基于深度 Alignment 网络的足部测量[J]. 系统仿真学报, 2020, 32(7): 1267-1278. Shi Min, Yao Hanqin, Li Chunqiang, et al. Foot Measurement Based on Deep Alignment Network[J]. Journal of System Simulation, 2020, 32(7): 1267-1278.
- [16] Huang G, Liu S, Der Maaten L V, et al. CondenseNet: An Efficient DenseNet Using Learned Group Convolutions [C]. Computer vision and pattern recognition (CVPR). Salt Lake City, USA: IEEE Press, 2018: 2752-2761.
- [17] Zhou Z, Zhou W, Li H, et al. Online Filter Clustering and Pruning for Efficient Convnets[C]. International conference on image processing (ICIP). Athens, Greece:

- IEEE Press, 2018: 11-15.
- [18] 刘小雍, 熊中刚, 阎昌国. 逼近误差的范数最小化的区间模糊模型建模[J]. 系统仿真学报, 2018, 30(3): 1203-1209.
- Liu Xiaoyong, Xiong Zhonggang, Yan Changguo. Interval Fuzzy Modeling Based on Minimizing-norm on Approximation Error[J]. Journal of Forestry Engineering, 2018, 30(3): 1203-1209.
- [19] 徐嘉荟. 基于模型剪枝的神经网络压缩技术研究[J]. 信息通信, 2019(12): 165-167.
- Xu Jiahui. Research on Neural Network Compression Technology Based on Model Pruning [J]. Information & Communications, 2019(12): 165-167.
- [20] Duan P, Wang T, Cui M, et al. Multi-person Pose Estimation Based on a Deep Convolutional Neural Network[J]. Journal of Visual Communication and Image Representation (S1047-3203), 2019, 61: 245-252.