

基于 Fast-OpenPose 的仰卧起坐姿态估计研究*

刘罡¹ 张旭² 侯恩翔²

(1. 无锡学院江苏省集成电路可靠性技术及检测系统工程研究中心 无锡 214105;

2. 南京信息工程大学电子与信息工程学院 南京 210044)

摘要:当前,许多学校体质测试项目中的仰卧起坐测试仍需通过手动计数,这不仅耗费人力,而且效率较低。为了促进体质健身的智能化发展,提出了一种基于人体姿态估计模型 Fast-OpenPose 和支持向量机(support vector machine, SVM)融合实现的仰卧起坐行为计数方法。通过 OpenPose 检测出仰卧起坐连续视频流中人体关键点的位置信息,再用 SVM 对获取到的每一帧人体关键点的坐标数据进行动作特征分类。鉴于原 OpenPose 网络复杂度高、模型参数量大、检测耗时长,用 FasterNet 对其主干特征提取部分进行轻量化改进,并在预测分支中优化更为高效的单分支网络结构和卷积类型,最后引入空间注意力(spatial group-wise enhance, SGE)来弥补精度损失。在 CoCo2017 数据集的基础上,额外扩充 1 000 张仰卧起坐场景的图片数据进行模型训练,实验结果表明,改进后的 Fast-OpenPose 在损失部分精度但不影响仰卧起坐姿态估计的情况下,模型参数量缩减近 80%,关键点检测速度提升 110%。与同系列其他改进模型相比,在保持相近平均精度均值(mAP)的同时,更具有轻量化与速度优势。

关键词: OpenPose; FasterNet; SGE 空间注意力; 姿态估计; 仰卧起坐

中图分类号: TP399 **文献标识码:** A **国家标准学科分类代码:** 520.20

Research on sit-up posture estimation based on Fast-OpenPose

Liu Gang¹ Zhang Xu² Hou Enxiang²

(1. Jiangsu Province Engineering Research Center of Integrated Circuit Reliability Technology and Testing System,

Wuxi University, Wuxi 214105, China; 2. School of Electronic and Information Engineering, Nanjing University
of Information Science and Technology, Nanjing 210044, China)

Abstract: At present, the sit-up test in many school physical fitness test projects still needs to be manually counted, which not only consumes manpower, but also is inefficient. In order to promote the intelligent development of physical fitness, this paper proposes a method for counting sit-up behavior based on a fusion implementation of the Fast-OpenPose human posture estimation model and support vector machine (SVM). The position information of the key points of the human body in the continuous video stream of sit-ups is detected by OpenPose, and the SVM is used to classify the motion features of the coordinate data of each frame of the key points of the human body. In view of the shortcomings of the original OpenPose network, such as high complexity, large number of model parameters and long detection time, this paper uses FasterNet to lightweightly improve its main feature extraction part, and optimizes the more efficient single branch network structure and convolution type in the prediction branch. Finally, spatial group-wise enhance (SGE) is introduced to make up for the loss of accuracy. Based on the CoCo2017 dataset, an additional 1 000 image data of sit-up scene are expanded for model training, and the experimental results show that the improved Fast-OpenPose reduces the number of model parameters by nearly 80% and improves the speed of keypoint detection by 110%, while losing part of the accuracy but not affecting the sit-up pose estimation. Compared with other improved models in the same series, it has more lightweight and speed advantages while maintaining similar mAP values.

Keywords: OpenPose; FasterNet; SGE spatial attention; posture estimation; sit-up

收稿日期:2024-02-06

* 基金项目:国家自然科学基金(62204172)、横向项目“体质监测设备研制”(苏技认字(2021)02050098)项目资助

0 引言

科技的发展推动了生活中运动健身方式以及质量的升级。近些年,为了重视学生的体质提升,教育部开始强调学生体质健康,同时,国家体育总局也增加了对健身场地和设施的资金支持。此外,一些先进的人工智能技术已经开始在体育运动中得到应用。体质测试作为大学生每年的必修科目,旧式的人工计数和传感器辅助测试的弊端日益凸显,其智能化的升级趋势越发迫切。从传统的人工计时和计数,到使用电子传感器辅助测量,再到基于计算机视觉技术的智能化测试方式,这些都预示着一个全新的趋势。

人体姿态估计又称人体关键点检测^[1],是计算机视觉领域的一个基础且具有挑战性的任务,旨在从图像或视频中准确地提取出人体的关键点位置,以还原人体的姿势。这些关键点包含了头部、肩膀、手肘、手腕、髋部、膝盖和脚踝等关键位置。人体关键点检测对于描述人体姿态、检测人体行为等至关重要。近年来,随着深度学习和卷积神经网络的不断发展,人体姿态估计在运动分析、智能交互、医疗康复、虚拟现实等领域有着广泛应用,但在体育测试和健身领域的应用案例却显得少之又少。本文将探索结合人体姿态估计技术和数据特征分类技术,共同实现体质测试中仰卧起坐项目的测试功能。

人体姿态估计不同于目标检测^[2],目标检测只能检测出人这个主体及其在图像中的大致位置,而姿态估计可以得到人体多个骨骼关节的具体空间信息,更有利于对人为动作的精确分类。姿态估计有传统方法和深度学习方法之分,深度学习方法克服了传统方法的局限性,在准确性和效率方面都取得了大幅提升;姿态估计也有二维和三维之分,三维对摄像头等传感器硬件有较高要求,二维虽不如三维那样具有丰富的空间深度信息,但它的优势在于计算量小、速度更快;姿态估计还有单人和多人之分,多人姿态估计能处理多人场景,应用更加广泛,但计算速度也相对较慢,更具挑战性。

由于多人姿态估计比单人姿态估计具备更强的泛化能力和通用性,因此本文采用的是 OpenPose 模型^[3],是目前最流行的二维多人姿态估计方法之一。使用的自底向上结构不同于 AlphaPose^[4]的自顶向下结构,自底向上结构先检测图像中人体部位,然后根据部位之间的连接关系来估计人体关键点的位置。二者虽准确性差距不大,但 OpenPose 的可定制性和实时性更好。近年来,不少学者

花费精力投入在 OpenPose 各个不同场景的应用研究以及模型本身的改进方法上。Sandler 等^[5]提出基于深度可分离卷积(depthwise separable convolution, DSC)的轻量级神经网络模型,在损失一定精度的情况下,大幅度压缩模型大小。肖文福等^[6]、伏娜娜等^[7]用深度可分离卷积结构优化 OpenPose 的主干网络,降低了卷积计算量和模型参数量,提升了模型在嵌入式设备和移动端上的运行速率。李一凡等^[8]用 ResNet18 网络替换底层特征提取部分,引入二阶项融合的残差网络提取特征图,使得改进后模型的检测精度和速度均有小幅提升。上述学者都较好的兼顾了模型的轻量化程度和检测精度,但模型的改进均基于特定的背景,应用场景单一,并不具备通用性。

1 方法概述

选取 OpenPose 基准模型属于 2D 多人姿态估计中自底向上方法的一种,分类如图 1 所示。本文将汲取文献[5-8]的改进思路,结合本领域最新的研究成果和自己的结构创新,对原模型的性能进行优化,并针对体育测试的应用背景,增加特定的数据集,提升模型的适应能力。

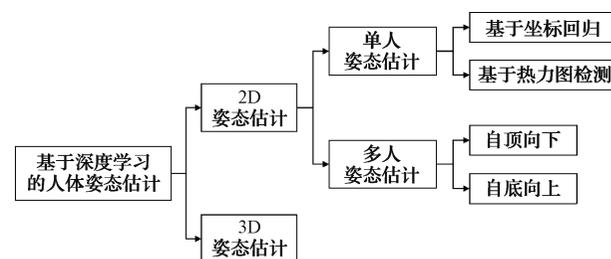


Fig. 1 Classification of human posture estimation

为了解决现实问题,本文顺应时代发展需求,研究了一种结合改进 OpenPose 人体姿态估计与支持向量机(support vector machine, SVM)^[9]多特征融合实现的仰卧起坐行为计数方法。该方案是先将视频转换成单帧图片,通过 OpenPose 模型一阶网络进行人体特征提取,二阶网络进行人体关键点的预测;得到关键点的位置信息后,利用 SVM 分类器做坐起(up)和平躺(down)两种姿态的二分类;接着,使用自定义的动作标准化检测模块判断有无违规动作,该模块可以通过骨骼关键点之间的坐标信息检测出学生有无双手抱头、有无双腿屈膝,并能在受测人员站起来时停止检测并输出运动个数。方案流程如图 2 所示。

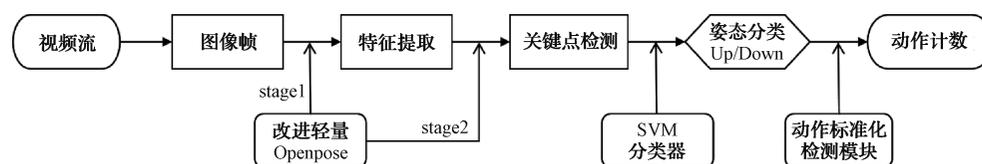


Fig. 2 Program flow

2 算法原理

2.1 原 OpenPose 基准模型

OpenPose 是由美国卡内基梅隆大学研发的高性能开源人体姿态估计检测器,在 COCO、MPI 等数据集上表现优秀,是人体姿态估计领域的里程碑之作。利用深度卷积神经网络(deep neural network, DNN)的自底向上架构,通过多个网络模块实现单图多人物姿态估计。模型可分解成主

干特征提取网络、姿态预测与部分关联分支两大结构。整体流程可概括为 3 大步骤:首先输入单张 RGB 图像或视频流,利用主干网络 VGG-19 的前 10 层提取底层特征信息;然后,得到特征图 F 将进入到关键点预测阶段,该阶段有两大分支机构,热力图预测分支预测关键点的二维坐标和置信度分数,部分亲和力分支预测身体不同部位的连接关系;最后,综合关键点的置信度分数和不同关键点的亲和力大小输出人体骨架信息。模型网络结构如图 3 所示。

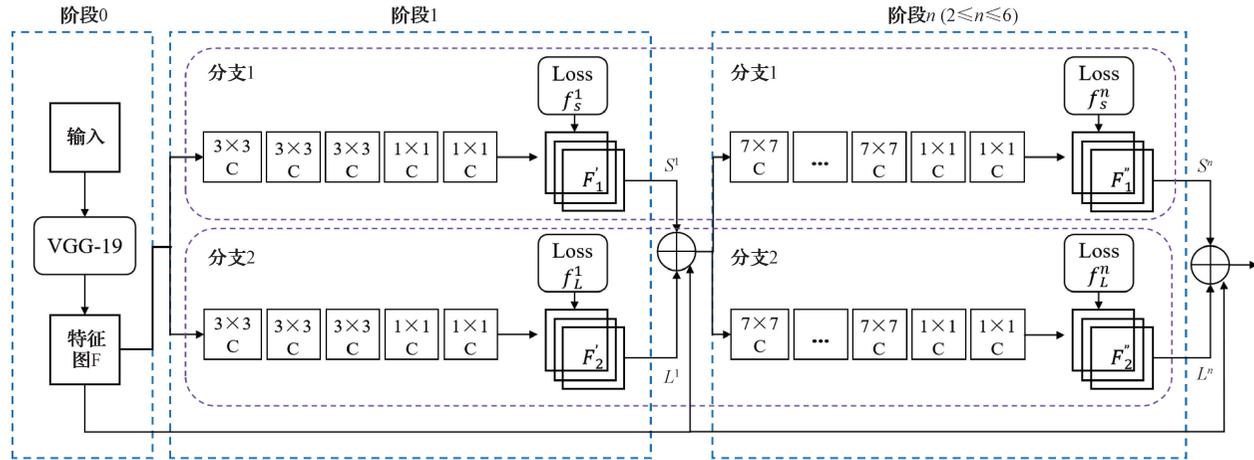


图 3 OpenPose 网络结构

Fig. 3 OpenPose network structure

由图 3 可知,特征图 F 不仅作为第 1 阶段分支 1 和分支 2 的输入,也和第 1 阶段的一组输出共同作为后续 5 个阶段的输入,这样的好处是多特征复用使得预测结果更加精确。关键点预测的多个阶段由上下两大分支结构组成,有置信度预测分支(confidence map for parts, CMPs)和亲和域关联分支(part affinity fields, PAFs)^[3]。上分支的输出 S^1 作为置信度映射,用于表征关键点的热力图位置;下分支的输出 L^1 为关联向量场,用于构造肢体向量场并将同一人体的高置信度关键点进行连接。不同阶段的关键点置信度与亲和力关系如下:

$$S^n = \rho^n(F, S^{n-1}, L^{n-1}), \forall n \geq 2 \quad (1)$$

$$L^n = \phi^n(F, S^{n-1}, L^{n-1}), \forall n \geq 2 \quad (2)$$

式中: ρ^n 和 ϕ^n 表示第 n 轮神经网络迭代的中间结果,与上一阶段的输出和特征图 F 相关连。

模型训练时,在每个阶段结束之后都要计算 CPMs 的损失函数 f_s 和 PAFs 的损失函数 f_l ,各阶段的结果将作为中间监督 Loss 储存,以防止梯度消失出现。最后,累加所有阶段的 Loss 作为总 Loss 计算梯度。随着阶段数目增加,网络训练加深,模型逐渐收敛,输出的预测图也愈发精确。损失函数 f_s 和 f_l 的计算公式如下:

$$f_s^n = \sum_{j=1}^J \sum_P W(P) \cdot \| S_j^n(P) - S_j^*(P) \|^2 \quad (3)$$

$$f_l^n = \sum_{c=1}^C \sum_P W(P) \cdot \| L_c^n(P) - L_c^*(P) \|^2 \quad (4)$$

式中: $W(p)$ 表示关键点在 p 区域的权重,当位置 p 被成功标注时 $W(p)=1$,否则为 0; S_j 和 L_c 表示关键点置信度和亲和力的预测值, S_j^* 和 L_c^* 表示真实值。

2.2 改进 OpenPose 模型

距 OpenPose 模型的发布已过去多时,近些年轻量神经网络(如 MobileNet、ShuffleNet 和 GhostNet 等)被学者广泛关注,并取得了丰富成果。科研人员和企业为了减少昂贵的计算设备开支,越来越倾向于低成本的轻量神经网络应用研究。而原 OpenPose 模型已经不能满足未来社会对它的实时性要求,考虑到仰卧起坐姿态估计有部署到算力有限的移动设备的可能性,针对原模型存在网络结构复杂、计算量大、检测费时的缺陷,结合最新的快速神经网络、卷积结构和注意力机制,提出了一种改进的轻量化 OpenPose 姿态估计的方法。改进后的网络结构如图 4 所示。

由图 3 和 4 可知,与原模型网络结构相比较,升级之处有 4 方面。1)优化主干网络,将过时的 VGG-19 卷积神经网络架构替换成 FasterNet 的新型网络架构;2)优化预测分支结构,将阶段 1~6 的双分支结构优化成前半部分单分支,后半部分双分支的结构形式,可使预测网络共享部分卷积运算成果而提升预测效率;3)替换阶段 2~6 公共分支的卷积类型,并重新设计卷积块内部的残差连接;4)在阶段 0 之后引入空间注意力机制(spatial group-wise enhance, SGE),赋予人体关键空间区域以更高权重,增

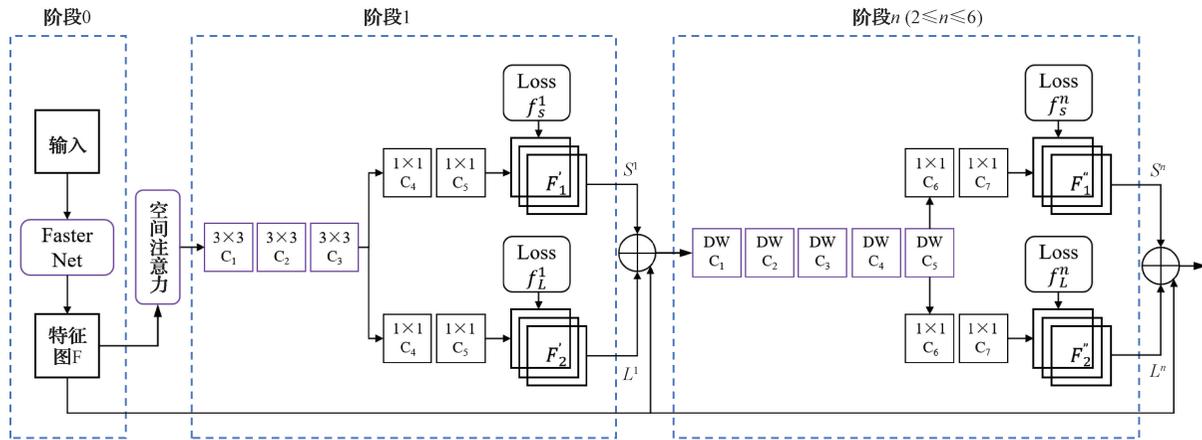


图4 改进 OpenPose 网络结构

Fig. 4 Improved OpenPose network structure

强模型对人体关键区域的特征提取与表示。

1) 优化主干网络

VGGNet^[10] 系列模型早在 2014 年就由牛津大学团队研发成功并开源, 本文 OpenPose 特征提取网络选用的是 VGG-19。它结构简单规整、特征表达能力强, 这使得它在当时的图像分类比赛中取得了很好的成绩。但 VGG-19 采用了过多的卷积层和全连接层堆叠, 网络层数过深导致网络参数量巨大、模型训练和推理成本过高, 与新网络相比已经失去了它的优势。将原 VGG-19 底层特征提取网络更替为 FasterNet^[11] 快速神经网络。

FasterNet 是一种基于新卷积操作的高效神经网络, 包括部分卷积 (partial convolution, PConv) 和点卷积 (pointwise convolution, PWConv)。PConv 的设计思路是利用特征图中的冗余, 只在部分输入通道上进行常规卷积, 而不影响其他通道。用 PConv 和 PWConv 替代传统卷积设计出的 FasterNet 网络, 可以减少计算量和内存访问, 降低网络延迟。这使得它在图像分类和检测任务上超越 MobileNetV2、ShuffleNetV2 等轻量化主干网络, 达到更高的准确率和运行速度。结构如图 5 所示。

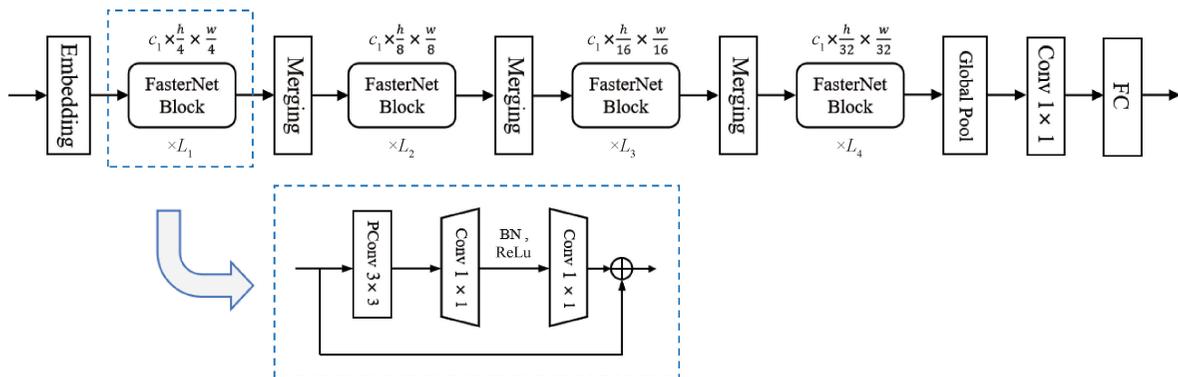


图5 FasterNet 模块结构

Fig. 5 FasterNet module structure

FasterNet 有 4 个阶段, 每个阶段前有一个嵌入层或合并层, 各阶段由若干个 FasterNet 块组成, 最后通过全局池化层和全连接层进行分类。块内部是 PConv + PWConv 的倒残差结构, 输入特征图分为两部分, 一部分经过一个 PConv 层和两个 1x1 卷积层, 另一部分不变。两部分特征图再经过一个元素级别的加法操作, 得到 PConv 块的输出。

2) 优化预测分支和卷积结构

为了生成人体关键点的置信度热力图和亲和域向量

场, 预测阶段从主干网络中获取特征, 并与前一阶段输出的关键点热图和亲和域估计连接起来。双分支预测的网络结构, 增加了模型的计算量, 极大的影响了模型推理速度。受周意乔等^[12]、马子健等^[13] 的启发, 决定在热力图和向量场两分支之间共享大部分计算, 在关键点估计和骨架关联时使用单个预测分支, 共享预测分支上除最后两层之外的所有卷积层。优化后的单分支结构, 可见图 4 中阶段 1~6。

考虑到预测阶段上有过多的 7x7 卷积, 而大卷积核

带来更大的感受野的情况下,也会导致更高的计算开销。所以,将 7×7 大小的卷积替换为具有相同感受野的卷积块,以捕获长程空间依赖性并降低网络复杂度。经过系列实验,将多个 DSC 层堆叠在一起,每个层都会增加感受野。通过多层串联与添加残差连接,网络的底层神经元就能够感知到上层信息,从而达到扩大感受野的效果。同时,使用深度可分离卷积代替传统卷积的另一个优势是它通过深度卷积(depthwise Conv, DWConv)和 PWConv 两步骤达到减少参数数量和计算复杂度的效果。优化后卷积结构如图 6 所示。

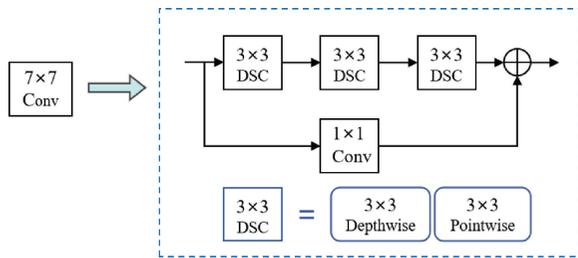


图 6 卷积块结构

Fig. 6 Convolutional block structure

3) 引入空间注意力模块

在网络模型的轻量化改进过程中,不可避免地会带来性能瓶颈受限、准确率下降、泛化能力降低等问题。因此,引入 SGE^[14] 提高模型对局部重要信息的感知能力以增强模型性能。经 FasterNet 主干网络生成的对象特征,通常以分组形式分布在每个层的特征向量中,用以表征不同语义实体。然而,像人体骨骼关键点这类子特征的提取会受到相似背景的噪声干扰,导致识别出错和定位偏差。SGE 注意力可以利用全局特征与局部特征间的相似性,建模人体关键点之间的依赖关系;还可以调整每个子特征的重要性,为每个特征组的不同空间位置生成注意力因子,提升关键区域的特征表达并抑制噪声。并且,该空间注意力增强模块几乎不增加额外参数和计算成本,与模型改进初衷不谋而合。

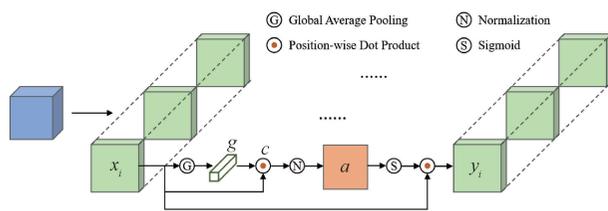


图 7 SGE 注意力模块

Fig. 7 SGE attention module

SGE 注意力模块的创新性在于空间分组增强。它的核心思想是将输入特征图分为多个组,每一组视为表示一个语义特征;对每个组内每个空间位置,根据其特征与组统计特征的相似度,生成注意力权重,按权重增强语义特

征的空间分布,并抑制不相关区域响应。SGE 原理如图 7 所示,具体步骤如下:1)输入特征图,按通道数均分为多组,每一组并行进行操作;2)对每组求全局平均池化,得到统计向量 g ;3)向量 g 与组内特征图 x_i 的每个位置元素做点积,得到相似度 c ;4)对相似度 c 进行标准化处理及 $[0,1]$ 的权重映射,得到规范化权重 a ;5)权重 a 经过 Sigmoid 激活后,再与特征图 x_i 做点积得到增强后的输出特征图 y_i 。统计向量的公式为:

$$g = F_{gp}(x) = \frac{1}{m} \sum_{i=1}^m x_i \quad (5)$$

式中: $F_{gp}(x)$ 为空间平均函数; x_i 为组空间局部特征向量; m 为像素的数目。

3 实验与分析

3.1 实验准备

1) 实验数据

用于训练人体姿态估计模型的数据集有多种,包括 LSP、FLIC 等单人数据集和 COCO、MPII 等多人数据集。不同数据集来自不同的复杂场景,包含的样本数量和关键点个数也各有不同。选取最为主流的 COCO2017 作为本文算法的人体关键点训练数据集。原本用于微软的姿态估计比赛,拥有超 10 万张的标注数据,包含身体的 17 个骨架关键点,OpenPose 又增添一个颈部关键点后共计 18 个。关键点分布如图 8 所示。这里将 COCO 中无关的复杂场景剔除后,并额外增加了 2 000 张自制的仰卧起坐人体姿态数据集,以提升特定应用场景中模型的适应能力和泛化性能。最后,统一修改输入尺寸为 368×368 ,并将融合后的数据划分出训练集 64 115 张、验证集 5 288 张、测试集 21 600 张。

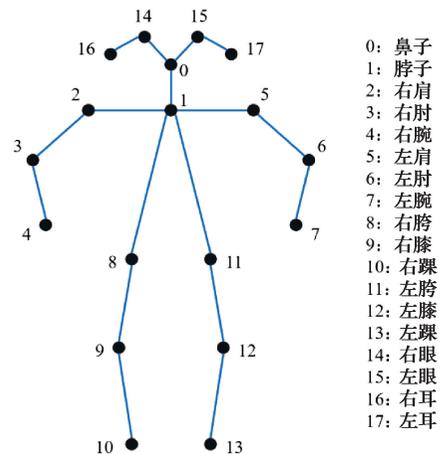


图 8 人体关键点分布

Fig. 8 Distribution of human body key points

2) 实验环境

CPU 为 InterR Core i7-13790F,显卡为 NVIDIA GeForce RTX™4070TI,显存 12 GB,运行内存 24 GB,操作系

统为 Win10, 编译器为 Pycharm, 编程语言为 Python3.9, 软件环境为 CUDA11.0 和 Cudnn8.0, 还包括 OpenCV4.5.3 视觉库和 TensorFlow2.4.0 深度学习加速框架。

3) 参数设置与模型训练

模型训练时优化器采用随机梯度下降 (stochastic gradient descent, SGD) 来迭代调整模型参数, 帮助模型快速收敛, 它的初始学习率为 0.01, 动量为 0.9, 权重衰减率为 5×10^{-5} , 批大小 (batch size) 设为 16, 轮数 (epoch) 设为 300。

训练过程中模型的状态如图 9 所示。可以看出, 随着迭代次数的增加, 训练集和验证集的总损失函数值曲线逐步收敛。在达到 250 轮之后训练损失值与验证损失值无限接近, 且趋于平缓地稳定在了 0.1 以下, 说明了模型训练效果较好。

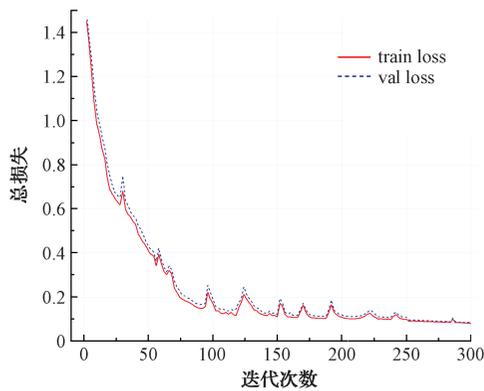


图 9 训练与验证损失曲线

Fig. 9 Training and verification loss curve

3.2 改进算法性能评估

人体关键点检测涉及将预测的关键点与真实值进行一一对应。然而, 由于姿态的多样性, 确保准确的对应关系是较为困难的。其次, 算法无法追溯预测关键点与真实值之间的先前对应关系。这使得难以确定先前某个预测结果是否存在误检或漏检, 从而影响算法的全面评估。因此需要采用一种更适合且复杂的人体关键点的度量指标, 即人体关键点相似度 (object keypoint similarity, OKS)^[15]。是一种用于衡量预测的关键点与真实关键点之间的准确性的度量标准, 取值在 0~1 之间, 越接近 1 表示预测的关键点位置越准确。计算方式通常涉及计算加权欧氏距离, 以考虑关键点之间的空间关系, 表达式如下:

$$OKS = \frac{\sum_i \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right)\delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \quad (6)$$

式中: i 为关键点号码; d 为预测关键点位置与真实位置的欧氏距离; s 为人体在整张图像的尺度因子; k 为衰减常数; δ 为冲激函数; $v_i > 0$ 表示第 i 个关键点可见。

由于身体有多个关键点, 很难将每一个关键点相似度单独衡量统计。为了更直观展示整体的检测精度, 借鉴 YOLO-pose 的性能评估方式, 用平均精度 (average precision, AP) 体现人体关键点的匹配效果, 包括了平均精度均值 (mAP)、AP@0.5 和 AP@0.75。其中, mAP 代表所有关键点的平均匹配精度, AP@0.5 代表 OKS 阈值为 0.5 时的匹配精度, AP@0.75 代表 OKS 阈值为 0.75 时的匹配精度。又为了展现本文改进模型的轻量化程度, 增加参数量 (parameters) 和每秒帧数作为模型复杂度和运行速度的评价标准。为验证本文算法的改进效果, 将其与其他优化算法进行比较实验, 结果如表 1 所示。

表 1 不同模型的性能比较

Table 1 Performance comparison of different models

模型	Network	mAP/%	AP@0.5/%	AP@0.75/%	Parameters/($\times 10^6$)	帧率/fps
OpenPose	VGG-19	64.3	85.9	71.5	143	15
PersonLab	ResNet-152	66.1	87.0	72.6	68.7	14
HRNet	HRNet-W48	65.4	86.8	70.4	63.6	17
Lightweight OpenPose	MobileNetv1	61.2	82.8	67.9	47	26
Fast-OpenPose(本文)	FasterNet	61.7	83.3	68.5	28	32

表 1 对比了 OpenPose 的原模型、PersonLab 和 HRNet 主流模型、OpenPose 的轻量化代表性模型、以及本文的 Fast-OpenPose 模型在 mAP、AP@0.5 和 AP@0.75 的精度指标和 Parameters 与 FPS 的轻量化指标上的性能差异。分析可知, 前三者原模型在 mAP、AP@0.5 和 AP@0.75 的精度值上均高于后两者的轻量化模型, 但模型参数量却远高于后两者, 并且检测帧率也远不及后者。这是因为原 OpenPose 模型和主流的姿态估计模型追求更精确的关键点检测效果, 采用了网络层数较深且卷积

量更大的特征提取网络 (如 VGGNet、ResNet 等)。后两者轻量化模型的精度虽不如前三者原模型高, 但不可否认它带来了更低的参数量和更快的检测速度。如最具影响力的 OpenPose 改进模型 Lightweight OpenPose, 它替换了基于深度可分离卷积的主干网络 MobileNetv1, 并在优化了预测阶段的网络分支结构, 使得它在损失 3.1% 的平均精度的情况下, 模型参数量降低至原来的 33%, 模型检测速度提升 70%。而本文的方法, 在借鉴了前人最优方案 (分支结构优化、卷积核改进) 基础上, 又引入了

快速特征提取网络 FasterNet 以及能提升关键点定位精度的 SGD 空间注意力。最终结果表明,本文提出的 Fast-OpenPose 模型,关键点平均定位精度 (mAP) 下降了 2.6%,模型参数量下降至原来的 19.6%,模型运行速度提升 113%。虽然在检测精度上低于原 OpenPose 和它

的竞争模型,但模型的轻量化程度和实时性能却远超它们。并且与 OpenPose 的代表性改进模型 Lightweight OpenPose 相比较,在 mAP 仍提升 0.5% 的优势前提下,参数量又降低了 19×10^6 、帧率也提升了 6 fps,实时性能取得了更高的飞跃。

表 2 不同改进模块的消融实验数据

Table 2 Experimental data of ablation with different modified modules

模型	FasterNet	单分支+DSconv	SGE 注意力	mAP/%	Parameters/($\times 10^6$)	帧率/fps
原 OpenPose				64.3	143	15
改进主干网络	✓			60.8	68	26
改进主干网络和预测分支	✓	✓		59.2	27	33
改进主干网络并加入注意力	✓		✓	62.5	69	25
本文	✓	✓	✓	61.7	28	32

通过逐步剔除模型的次要组成部分并改变模型的特定结构,以便深入理解模型的工作原理、关键组件对模型性能的影响,以及模型在不同条件下的性能表现。这里将创新模块做消融实验去验证本文改进思路的正确性,结果如表 2 所示。可知,原模型的 VGG 主干网络具有大量参数,使用 FasterNet 替换后能降低计算代价;再优化单分支和卷积类型后,模型参数量进一步降低,推理速度达到最快,但精度损失过高;再通过轻量化的 SGE 空间注意力给予人体关键区域更高权重后,人体关键点的检测精度得到了小幅补偿,且保持了较快的检测速度。

3.3 仰卧起坐场景应用

1) 分类器的选取

在 Fast-OpenPose 提取出人体关键点数据后,需对其进行进一步的分类处理,以判断不同的行为特征。当前,传统机器学习方法的 SVM、决策树 (decision tree, DT)、DNN 及 K 近邻 (K nearest neighbors, KNN) 等算法在人体动作分类任务上被广泛使用。根据冯文字等^[16]、李聪林等^[17]在人体行为分类领域的研究成果可知,KNN 主要通过已知数据点的距离进行分类,适合简单的分类任务;DNN 是一种多层神经网络结构,在大规模数据集上表现出色,是三者里性能最好、最耗费计算资源的;SVM 在高维空间中具有良好的泛化能力,对小样本数据更加友善,且训练耗时最低,准确率仅略低于 DNN。

为了验证引用文献结果在本应用场景下的适用性,这里又对 KNN、SVM 及 DNN 在仰卧起坐的 Up 和 Down 两种姿态坐标进行分类实验。自制仰卧起坐图片数据 2 000 张,其中 Up 和 Down 的姿态各占总量的 1/2,并以 8:2 的划分训练集和验证集。由于仰卧起坐运动的复杂性,存在侧身遮挡、视角偏差等现实因素,会导致远离相机那一侧的关键点定位错误和漏检等情况。因此,为了排除定位偏差关键点的错误数据干扰,降低计算复杂度,提升分类精度,故选取靠近相机一侧的关键点坐标作为数据样本。图 8 中 0~4、8~10、14~16 号点位。

从 Python 的 sklearn 库中调用各分类算法,并设置各分类算法的性能参数为最优 (KNN 中的 $K=6$;SVM 中的核函数为 RBF,惩罚系数 $C=5$;DNN 的隐藏层数目为 3,各层神经元的数量为 100)。在相同实验环境下各分类器的性能对比如表 3 所示。

表 3 不同分类器的性能对比

Table 3 Performance comparison of different classifiers

名称	训练准确率/%	验证准确率/%	总共耗时/s
KNN	96.2	94.5	2 456
SVM	97.7	96.4	2 310
DNN	99.1	97.9	2 641

由表 3 可知,SVM 分类器相较于其他分类器在准确率和分类速率上有着较为均衡的成绩。在训练准确率和验证准确率上低于 DNN 深度神经网络,但高于 KNN 近邻算法;在分类时耗上均优于 KNN 和 DNN。综合考虑不同分类器的准确率和耗时,最终选用 SVM 作为仰卧起坐项目的行为特征分类器。

2) 测试功能实现

SVM 是一种监督学习算法,在人体姿态这种二分类任务中应用颇多。通过构建超平面,最大化不同类别样本的间隔距离来实现分类。目标是找到一个最优决策边界,即最大间隔超平面,使不同类别样本分布在该超平面的两侧。主要参数设置如下:核函数类型为 RBF 核、误差惩罚参数 $C=1$ 、收敛的误差容限 $tol=0.005$ 。

测试时,USB 摄像头选在距离人员 2 m、距地面 1.5 m 处固定。将视频流输入到算法模型,先检测到人体关键点的空间信息,接着通过训练好的 SVM 分类文件比对当前图像帧中的关键点坐标特征,并计算出动作类别的置信度。当检测到 down 或 up 的姿态置信度高于 0.9 时,运动个数 (number) 各加 0.5,并将 number 设置整型变量 (int) 后,输出最终取整结果。截取两种类别最高置信度的视频帧汇总如图 10 所示。

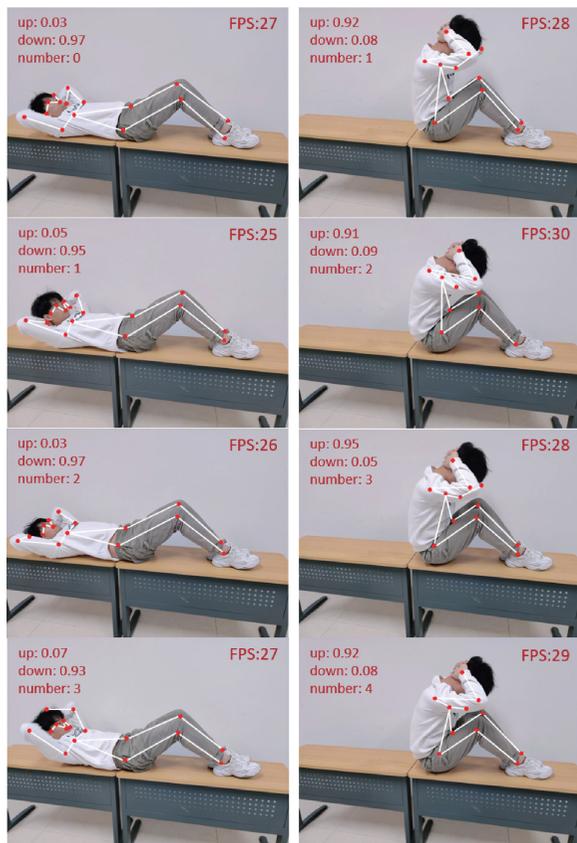


图 10 仰卧起坐测试效果

Fig. 10 Graph of sit-up test results

由测试结果可见,虽然仰卧起坐姿态复杂且存在角度偏差和躯干遮挡,导致了内侧手肘、手腕和部分头部关键点定位错误,但改进后的算法依旧能检测出绝大多数的人体关键点。并且,由于SVM分类器选取的特征点数据皆为外侧显著区域,因此行为特征分类和运动动作计数的正确性几乎不受影响。因在用OpenPose检测出人体关键点之后,又加入SVM去做动作分类,额外消耗了计算机的算力,所以这里的视频流帧率较Fast-OpenPose的关键点检测帧率有所下降,但实时性能依旧得到了满足。

动作标准化检测模块是通过计算关键点之前的欧氏距离和关键点连线之间的夹角来判断测试人员是否双手抱头和弯曲膝盖。例如,捕捉一个连贯的躺和起动作内的视频帧中是否存在3号肘部关键点和14号耳部关键点的像素距离小于50,以及腿部8、9、10号关键点之间的夹角是否在 $100^{\circ}\sim 150^{\circ}$ 之间。若二者皆存在,则该动作有效,计数加1;若有一不存在,则该动作无效,计数不变。当测试人员走离视频框,未检测到关键点信息时,算法结束并输出最终结果。因该模块的实现原理较易,本文不再赘述,效果如图11所示。

由图11可知,动作标准化检测模块实时计算目标关

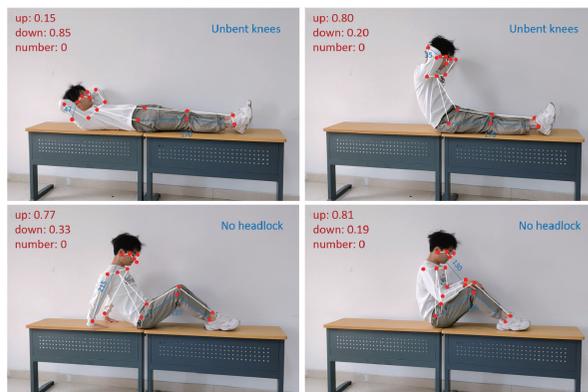


图 11 动作标准检测结果

Fig. 11 Action standard test results

节点的欧氏距离以及关节点连线间的夹角大小。蓝色的线条为计算辅助线,手部的蓝色数字为像素距离,腿部的蓝色字体为角度大小,它们皆通过调用OpenCV视觉库在检测结果图上画出。当检测到目标关节点间的连线夹角或欧氏距离不在定义区间内时,会有“未屈膝”和“未抱头”提示,且这段时间内完成的动作无效,这进一步保证了视觉测试方案的规范性。

4 结论

当前,人体姿态估计技术仍在飞速地发展,基于改进OpenPose模型实现的人体行为检测应用在诸多领域。研发人员想方设法地对该模型的主干特征提取网络和卷积结构进行优化,以平衡关键点的检测精度和速度。本文从实际应用场景出发,扩充特定背景下的姿态数据,升级过时的主干特征网络,引入空间注意力机制,设计高效的共享分支结构,并优化内部卷积结构,推出了Fast-OpenPose快速姿态估计算法。将人体外侧置信度较高的关键点坐标喂入SVM分类器用于仰卧起坐姿态的分类任务,再与动作标准化检测模块相结合,有效提升仰卧起坐运动计数的准确性与可靠性。基于Fast-OpenPose的仰卧起坐姿态估计丰富了人体行为识别的案例,也可以应用于其他规则性的体育健身项目,促进体育健身的智能化发展。

参考文献

- [1] SAMKARI E, ARIF M, ALGHAMDI M, et al. Human pose estimation using deep learning: A systematic literature review [J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1612-1659.
- [2] KAUR R, SINGH S. A comprehensive review of object detection with deep learning [J]. Digital Signal Processing, 2023, 132: 103812.
- [3] CAO Z, SIMON T, WEI S E, et al. Realtime multi-

- person 2d pose estimation using part affinity fields[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7291-7299.
- [4] MUNDT M, BORN Z, GOLDACRE M, et al. Estimating ground reaction forces from two-dimensional pose data: A biomechanics-based comparison of AlphaPose, BlazePose, and OpenPose [J]. Sensors, 2022, 23(1): 78.
- [5] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2018: 4510-4520.
- [6] 肖文福, 孙焱, 陈恩庆. 一种基于深度可分离卷积的实时多人姿态估计[J]. 计算机应用研究, 2020, 37(S2): 348-350.
- XIAO W F, SUN Y, CHEN E Q. Real-time multi-person pose estimation based on depthwise separable convolution[J]. Application Research of Computers, 2020, 37(S2): 348-350.
- [7] 伏娜娜, 刘大铭, 程晓婷, 等. 基于轻量级 OpenPose 模型的跌倒检测算法[J]. 传感器与微统, 2021, 40(11): 131-134, 138.
- FU N N, LIU D M, CHENG X T, et al. Fall detection algorithm based on the lightweight OpenPose model [J]. Sensors and Microsystems, 2021, 40(11): 131-134, 138.
- [8] 李一凡, 袁龙健, 王瑞. 基于 OpenPose 改进的轻量化人体动作识别模型[J]. 电子测量技术, 2022, 45(1): 89-95.
- LI Y F, YUAN L J, WANG R. An improved lightweight human action recognition model based on OpenPose[J]. Electronic measurement technology, 2022, 45(1): 89-95.
- [9] CHANDRA M A, BEDI S S. Survey on SVM and their application in image classification[J]. International Journal of Information Technology, 2021, 13: 1-11.
- [10] SHAFAPOURTEHRANY M, REZAIIE F, JUN C, et al. Mapping post-earthquake landslide susceptibility using U-Net, VGG-16, VGG-19, and metaheuristic algorithms[J]. Remote Sensing, 2023, 15(18): 4501.
- [11] QING Y, HUIJUAN M, YUCHEN G, et al. A real-time object detection method for underwater complex environments based on FasterNet-YOLOv7[J]. Journal of Real-Time Image Processing, 2023, 21(1): 8.
- [12] 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 Y Q, XU Y L. Real-time human pose recognition in complex environments based on bidirectional LSTM [J]. Instruments Journal, 2020, 41(3): 192-201.
- [13] 马子健, 林雨衡, 王志强, 等. 封闭环境下人体姿态识别及打架行为监测[J]. 计算机应用, 2021, 41(S2): 214-220.
- MA Z J, LIN Y H, WANG ZH Q, et al. Human pose recognition and fighting behavior monitoring in closed environment[J]. Computer Application, 2021, 41(S2): 214-220.
- [14] LI Y, LI X, YANG J. Spatial group-wise enhance: Enhancing semantic feature learning in CNN [C]. Proceedings of the Asian Conference on Computer Vision, 2022: 687-702.
- [15] 党选举, 李启煌. 混合坐标注意力与改进空间金字塔池化融合的物体位姿估计[J]. 国外电子测量技术, 2023, 42(1): 178-186.
- DANG X J, LI Q H. Hybrid coordinate attention and improved spatial pyramid pooling fusion object pose estimation [J]. Foreign Electronic Measurement Technology, 2023, 42(1): 178-186.
- [16] 冯文字, 张宇豪, 张堃, 等. 复杂环境下课堂多人状态检测算法研究[J]. 电子测量与仪器学报, 2021, 35(6): 53-62.
- FENG W Y, ZHANG Y H, ZHANG K, et al. Research on multi-person state detection algorithm in complex environment[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(6): 53-62.
- [17] 李聪林, 王琪冰, 陆佳炜, 等. 基于数字孪生的电梯乘客异常行为建模与识别方法[J]. 计算机工程与应用, 2023, 59(19): 274-284.
- LI C L, WANG Q B, LU J W, et al. Elevator passenger abnormal behavior modeling and recognition method based on digital twin[J]. Computer Engineering and Application, 2023, 59(19): 274-284.

作者简介

刘罡(通信作者),硕士生导师,高级工程师,主要研究方向为计算机视觉与人工智能。

E-mail:mrliu@163.com

张旭,硕士研究生,主要研究方向为计算机视觉与人

体姿态估计。

E-mail:1793697124@qq.com

侯恩翔,硕士研究生,主要研究方向为深度学习、目标检测。

E-mail:575072816@qq.com