DOI:10.19651/j.cnki.emt.2212162

引入 GAN 与可变形注意力的多维人体运动分析*

孙文昊[1,2] 路光达[1,2] 秦转萍[1,2] 郭庭航[1,2] 赵壮壮[1,2]
(1.天津职业技术师范大学自动化与电气工程学院 天津 300222; 2.天津市信息传感与智能控制重点实验室 天津 300222)

摘 要: 研究了一种用于肢体状态评估和运动姿态校正的人体运动分析系统。首先,针对人体运动时易出现的遮挡等问题,通过引入可变形注意力和生成对抗网络优化人体关键点热图位置检测,在 Transformer 的基础上设计了一种人体关键点识别算法。其次,利用所提出的算法,结合人体姿态的肢体空间约束关系以及体态分析相关知识,设计了一套运动分析系统。最后,通过在公共数据集上和真实场景中的测试,从质化和量化两个角度对所提出的算法和系统的可行性进行了评估实验。实验结果证明,本文算法在公共数据集上的检测精度最高可达 93.7%;在实际场景的测试中,本文设计的算法和运动分析系统可以有效解决人体姿态识别中常见的遮挡等问题,并通过可视化系统展示了对人体运动姿态的多维度分析结果。

关键词: 人体运动分析;人体关键点检测;Transformer;生成对抗网络;可变形注意力
中图分类号: TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Multi-dimensional human motion analysis by introducing GAN and deformable attention

Sun Wenhao[1,2] Lu Guangda[1,2] Qin Zhuanping[1,2] Guo Tinghang[1,2] Zhao Zhuangzhuang[1,2]
(1. School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China;
2. Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education, Tianjin 300222, China)

Abstract: This paper studies a human motion analysis system for limb status assessment and motion posture correction. Firstly, to address the problems such as occlusion that are prone to occur during human motion, this paper introduces deformable attention and generative adversarial networks based on Transformer for optimal human key point location detection. Secondly, using the proposed algorithm, this paper designs a motion analysis system by combining the limb space constraint relationship of human posture and knowledge related to body posture analysis. Finally, through testing on public datasets and in real scenarios, this paper evaluates the feasibility of the proposed algorithm and system from both qualitative and quantitative perspectives in experiments. The experimental results prove that the detection accuracy of the algorithm in this paper can reach up to 93.7% on public datasets; in the tests on real scenes, the algorithm and the motion analysis system designed in this paper can effectively solve the common problems such as occlusion in human posture recognition, and show the multi-dimensional analysis results of human motion posture through the visualization system.

Keywords: human motion analysis; human key point detection; Transformer; generative adversarial networks; deformable attention

0 引 言

新冠病毒大流行更加突出了远程医疗的重要性。在针对脑卒中、帕金森等疾病引起的肢体运动障碍患者的康复治疗过程中,主要依靠的是专业医师的指导和辅助训练,以及对患者肢体运动进行的评估[1]。但是,该方法仅限于在医院等专业治疗场所下进行,场地限制和需要专业人士针对性指导等缺点导致患者的治疗成本较高。因此,基于深度学习和传感器技术的人体运动分析等辅助治疗方法[2-3]成为人体康复学等学科的重要研究方向,其目标是通过使

收稿日期:2022-11-22
* 基金项目:天津市科技支撑重点项目(20YFZCSY00600)、天津市教委科研计划项目(2021KJ011)、天津市科技计划项目(22ZYCGSN00480)、天津市教委科研计划项目(2022ZD009、2022ZD035、2022ZD036)、天津市研究生科研创新项目(2022SKY284)资助

用传感器捕捉人体相关点位坐标,从而获得人体运动数据。

人体姿态是人类在日常生活中出现的各种运动模式和状态,包括行走、跳跃等有规律的运动,以及意外跌倒等无规律的运动,姿态检测是人体运动学研究的重点。随着社会和科学技术的发展,人体姿态检测和人体运动分析的研究方法得到丰富和深化,使姿态检测有了更广泛的应用层面,进一步推动了更多领域的应用研究[4-6],包括开发自动化系统来检测人体运动并将检测数据与医疗等相关学科进行结合应用。

目前,人体姿态检测的研究主要分为基于可穿戴设备的人体姿势检测和基于视觉的人体姿势检测。基于可穿戴设备的人体姿态检测是通过将传感器设备佩戴在人体相关位置并读取传感器的数据,结合各传感器的数据进行运动姿态识别。如 Li 等[7]结合上肢表面肌电图信号(surface electromyogram, sEMG)和下肢三轴加速度以及足底压力信号的多模态信息,通过上下肢可穿戴传感器的不同组合,实现了动态检测人体运动姿势。可穿戴设备包含的传感器越多,其识别精度越准确,也能更精确地分析人体运动特性。但是这种设备的缺点也很明显,如携带不方便、传感器太多时影响人体运动状态等。随着研究的深入,许多研究人员开始尝试利用算法结合少量传感器进行实验[8-10],通过不断改进姿势检测方法,提高检测率。由于可穿戴识别设备始终受到环境和区域限制,因此,基于视觉的人体姿态识别仍然是研究人员最广泛关注的一种方法。

早些年间,研究人员通过在室内环境下架设多台经过专业校准的高清摄像机来实现人体动作捕捉[11]。这种识别方法不仅价格昂贵,而且对于非专业人士来说也很难实现。近年来,得益于深度学习技术的发展以及单目 RGB 相机和双目深度相机的性能提升,基于计算机视觉技术的人体动作捕捉和姿态检测方法被越来越广泛地使用[12-14]。从检测视角方面分析,基于视觉检测的人体运动捕捉方法主要分为两类,包括基于多视图的人体姿态重建和基于单视图的人体姿态识别方法。基于多视图的人体姿态识别已经有了很多成果。例如 Kwon 等[15]基于预先扫描的模板,将模板拟合到生成的视觉外壳,使用多视图轮廓跟踪主体的形状和姿势。Zhang 等[16]通过将语义分割与运动跟踪优化框架相结合,即使在严重遮挡下也能实现强大的多人运动捕获。尽管这些方法已经取得了很好的效果,但它们仍然需要为每一次的运动分析预先设置一个模板。此外,架设多台高清摄像机导致系统成本较高且操作复杂。因此,这种动作识别系统被限制在实验室环境下,并未被广泛使用,基于单视图的人体姿态识别方法仍然是现在最广泛研究的问题。

在单视图人体姿态检测方面,Long 等[17]从深度图估计 2D 关节位置或人体关键点特征图,这也是现在大部分人所用的方法。这种方法虽然非常有效,但发生严重遮挡时往往会产生不稳定的预测结果。并且双目相机的价格较单目相机要更加昂贵,成本偏高,采用单目相机结合深度学习算法的关键点识别是目前最流行的方法。Kim 等[18]使用卷积神经网络直接从 RGB 图像预测关节位置和关节位置热图,从而估计人体运动姿态。Jiang 等[19]训练了一个端到端的人体姿态识别网络,直接从 RGB 图像推断人体模型的形状和姿势参数。但是,当人体快速移动时会出现图像严重模糊和深度信息丢失的问题,从而产生识别错误或细节丢失等不稳定的结果。此外,大多数基于单目相机和深度学习的方法是按照视频流中的顺序独立处理每个帧,因此,可能会出现为姿势不连贯和人体重建错误等结果。

综上所述,基于计算机视觉技术的人体姿态检测虽然与可穿戴设备检测相比成本较低,且更加方便。但是,运动过程中的遮挡及快速运动产生的动态模糊问题仍是制约单视图检测精度的重要因素。目前较为流行的方法是通过优化相关算法实现更高精度的遮挡预测并解决模糊问题。基于上述分析,本文通过引入可变形注意力的 Transformer 进行关键点位置检测,结合生成对抗网络(generative adversarial networks, GAN)对人体关键点热图进行优化重构,从而解决人体肢体间遮挡和动态模糊等问题。通过人体关键点位置识别和关节约束关系,本文建立了一套仅依靠单目视觉检测实现对人体运动过程进行多维度分析的人体运动分析系统。

1 结构设计

本文设计的运动分析系统分为两大部分:识别部分和分析部分,结构如图 1 所示。首先,本文在 Transformer 的基础上,通过引入可变形注意力,在节约了计算成本的同时提高模型对全局位置关系的建模能力。为了增强识别算法对关键点识别的准确性以及增强出现遮挡或动态模糊时模型的抗干扰能力,本文在前期算法基础上结合生成对抗网络(generative adversarial network,GAN)设计了一个热图重构模块,用于生成关键点热图分布并优化网络对热图预测的能力。获取人体关键点位置后根据人体位置的约束关系以及运动过程中肢体变化过程实现对人体的体态和运动分析。分析的结果通过 PyQt 设计的数据可视化界面进行展示。

2 算法设计

人体动作时肢体间或肢体与物品间的遮挡以及运动模糊问题是影响人体关键点识别精度的重要因素。本文通过引入可变形注意力和生成对抗网络,在 Transformer 的基础上设计了一种注意力偏移、可通过对抗学习来重构预测部分的人体关键点识别网络。

2.1 关键点识别算法框架

人体关键点检测算法的框架如图 2 所示。初步提取特征采用 HRNet[20]中的 Bottleneck 和 Basicblock 模块,其参数量仅为原始 HRNet 网络模型的 1/4。其中的 Bottleneck 由 3 层组成,其结构如图 3(a)所示,依次为 1×1 卷积,3×3

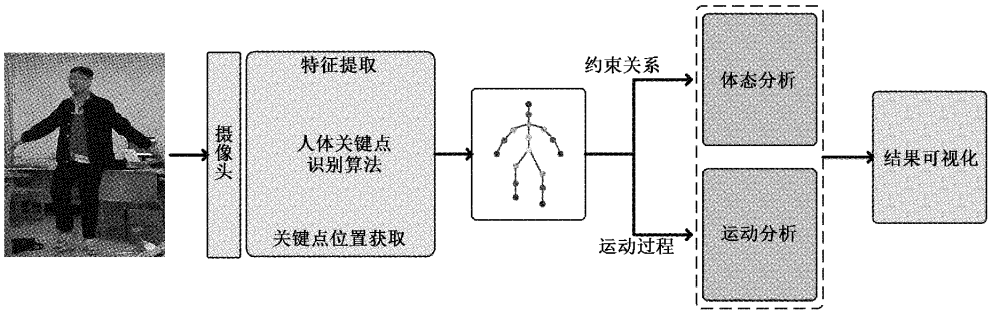


图 1　系统主要架构

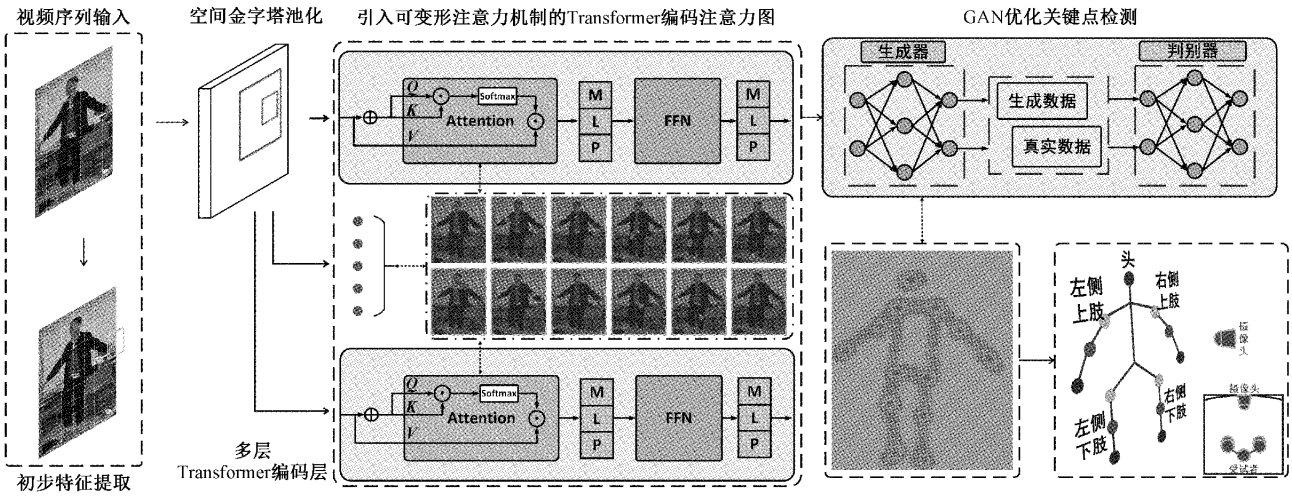


图 2　基于 Transformer 和 GAN 的人体关键点识别框架

卷积，1×1 卷积，其中 1×1 卷积负责缩小放大通道数 C。Basicblock 由两层组成，其结构如图 3(b)所示，均为 3×3 卷积，第 1 个 3×3 卷积的步长是可变的，可以调整特征图的尺寸；第 2 个 3×3 卷积的步长为 1，只用于改变通道数，不会改变特征图的大小。

在进行全局关系建模之前，为了获取更多层次的图像特征，将空间金字塔池化[21]嵌入 Transformer 中。经过三层池化后的特征图分别送入 Transformer 做自注意力计算。本文采用的空间金字塔池化与多头自注意力模块连接方式如图 4 所示。由于热图预测只需要对图像进行编码，因此特征提取阶段只使用 Transformer 中的编码器，将原始图像信息压缩成关键点的紧凑位置表示。通过 Transformer 获取的人体关键点注意力图，经过解码计算获得各人体关键点的二维坐标。将首次获得的关键点信息通过 GAN 生成一组新的注意力热图，用于表征前向计算后每个关键点位置的置信度分数。之后使用与生成模型相同架构的判别模型，将输入热图与 RGB 图像共同编码，再将其解码为一组新的热图，用以区分真假热图。生成对抗网络通过生成器和判别器之间的博弈逐渐优化人体关键点的检测精度，从而使模型输出的关键点坐标更加精确。

2.2　基于 Transformer 的人体关键点识别

使用 Transformer[22]架构建立人体姿态识别模型的优势是可以显式地捕获关键点特征之间的空间依赖性。由于卷积算子的感受野较小，因此卷积神经网络在提取局部特征方面具有优势，但捕获全局特征的能力较差。仅通过叠加卷积深度来扩大感受野也并不能有效地捕获全局依赖关系。Transformer 架构在建模高阶特征方面具有很大的优势，自注意力层的计算使模型能够获得任意位置之间的交叉关系，并作为一个瞬时记忆去存储这些依赖关系。多头自注意力模块是在 n 个分支内进行自注意力操作，能在有限的网络层中对全局内容进行计算。注意力机制计算式为：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{1}$$

图 5 为 Transformer 中的多头自注意力机制的结构。为了使网络能更好获取图像中人体关键点之间的关联性，将带有位置信息的特征图输入到带有 3 个权重参数矩阵 W_q、W_k、W_v 的多头自注意力模块中，$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别表示对应的查询量(*query*)、键(*key*)和值(*value*)。位置编码方式与文献[22]中的编码方式相同，本文不再赘述。自注意力机制通过将值矩阵 $\boldsymbol{Q}$ 中的每个值与 W 中的相应权重进行

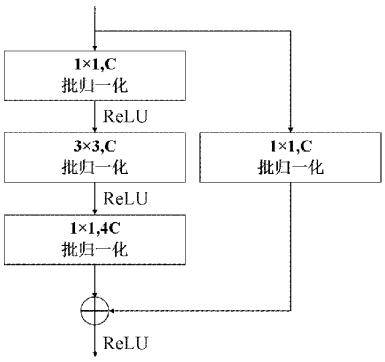


图 3 Bottleneck 和 Basicblock 结构

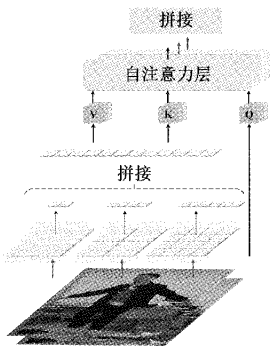


图 4 加入空间金字塔池化的多头自注意力模块

线性组合的方式作为动态权重实现了对特征向量的更新。通过将得到的注意力图与低维特征图进行融合,可以实现不同尺度特征信息的交换,以达到获取更优的关键点预测热图的作用。

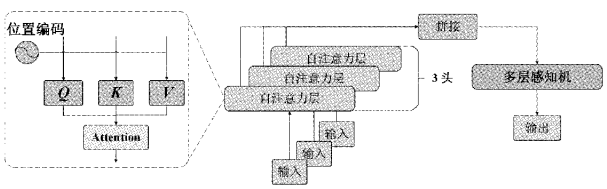


图 5 Transformer 中的多头自注意力机制

关键点热图输出的方式为在图像的每个位置预测一个分数,该分数用以表征该位置属于关键点的置信度,这种方法更好地保留了空间位置信息。训练人体姿态估计网络时,以热图为标签进行训练,将原图的关键点像素坐标转换为降采样后分辨率下的关键点坐标,并利用高斯模糊转换成热图。之后,通过将相应的分辨率下的热图转换回原始坐标空间,获得关键点坐标在原始图像坐标空间中的位置。

2.3 可变形注意力

Transformer 在各种视觉任务上表现出卓越的性能,标准的 Transformer 具有全局关注、特征远距离依赖等优点,强大的全局特征捕获能力赋予了 Transformer 模型比卷积神经网络模型更强的全局建模能力。然而,简单地扩充注意力计算也会引发一些问题。一方面,使用密集注意力计算会导致过多的内存和计算成本,并且特征可能会受到超出感兴趣区域的无关部分的影响。另一方面,采用稀疏注意力机制计算时,有很大几率会涉及到无关或不感兴趣区域数据的计算,从而限制模型对远程关系建模的能力。与可变形卷积类似,可变形卷积的每一个部分都进行了一定的偏移,可变形注意力(deformable attention)[23]则是 patch 的大小和整体位置的改变。同时,patch 的位置和大小一般是由一个线性层直接获取,并且位置的相关预测基本上和整体任务是独立的,因此,在全局关系计算中,可变形注意力在算力上与普通注意力的区别并不大,但是特征计算的能力却有很大提升,用于密集预测型任务也有更大的优势。图 6 为本文中可变形注意力的运算结构图。

2.4 GAN

GAN 是由 Goodfellow 等[24]提出,通过两个卷积神经网络之间的对抗学习来训练两个网络模型的能力。图 7 为应用于优化关键点检测的 GAN 的简化模型。GAN 主要由生成网络(generator network)和判别网络(discriminator network)组成。判别网络的作用是判断输入的图像数据是否为真,生成网络的作用是根据真实图片的数据分布生成一个类似真实图片的图片,初始状态下生成网络和判别网络都是没有经过训练的。经过多次对抗训练,即生成网络创造一张图片去欺骗判别网络,判别网络去判断这张创造出的图片是真是假。在不断对抗训练和学习的过程中,两个网络的能力越来越强,最终达到均衡状态。

在本文中,生成对抗网络被用于解析人体关键点的分布规律,并以此对产生严重遮挡的部分和动态模糊部分进行预测。类似于传统的 GAN 模型,生成网络生成一组热图,用于表征经 Transformer 计算后每个关键点位置的置信度。判别网络使用与生成网络相同的架构,将输入热图与 RGB 图像一起编码,并将其解码为一组新的热图,以区分真假热图。通过生成网络和判别网络之间的对抗逐渐优化人体关节识别的准确度。

GAN 训练过程中总的损失函数为:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(\boldsymbol{x})] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(\boldsymbol{z})))] \quad (2)$$

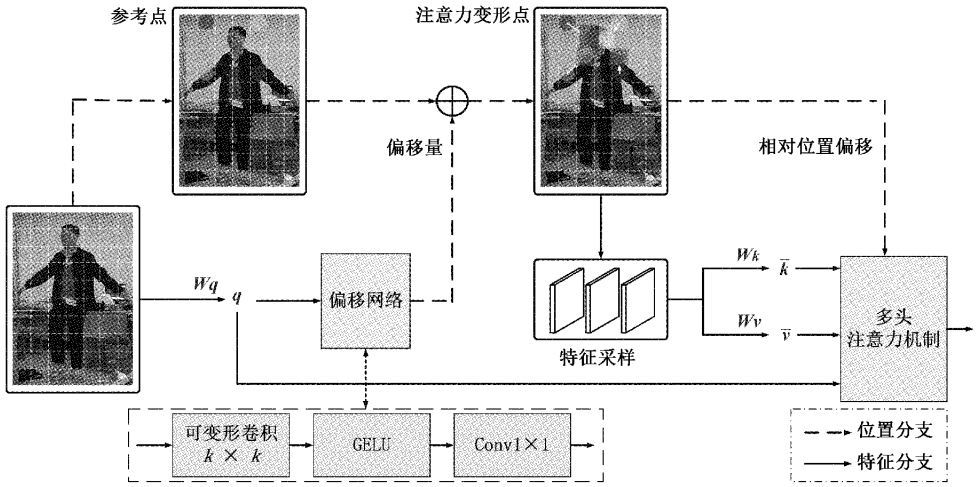


图 6　可变形注意力运算结构图

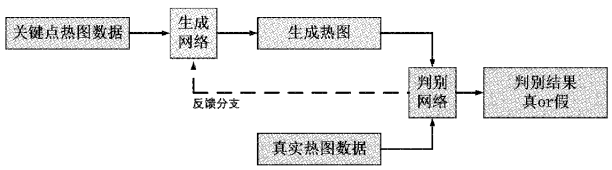


图 7　用于优化关键点检测的 GAN

式中：$D(x)$为判别网络的函数，$G(z)$为生成网络的函数；$P_z(z)$为输入噪声数据分布；$V(D,G)$为表征真实样本与生成样本之间差异程度的价值函数；max 为固定生成网络 G 时判别网络 D 最大化地判别给定数据为真或假，min 为在固定判别网络 D 的条件下使生成网络最小化真实样本与生成样本的差异。

训练过程分为两个阶段，首先训练用于判断是真样本还是假样本的判别网络。训练判别网络时，从式(2)中分离出 D 的函数，使用梯度下降法进行优化，损失函数为：

$$\mathcal{L}_D = -\mathbb{E}_{x\sim p_{gt}(x)}[\log D(\boldsymbol{x})] - \mathbb{E}_{z\sim N(z|0,I)}[\log(1-D(G(z))] \tag{3}$$

在训练生成网络时，从式(2)中分离出 G 的函数，同样使用梯度下降法进行优化，损失函数为：

$$\mathcal{L}_G = \mathbb{E}_{z\sim N(z|0,I)}[\log(1-D(G(z))] \tag{4}$$

在实验过程中，由于判别网络的梯度下降过快，在与生成网络对抗时出现了判别网络过于强势的情况，即生成网络无法产生使判别网络发生错误的热图数据。因此，在 GAN 中添加反馈分支将判别网络提取的热图信息经归一化后反馈至生成网络中，可以有效增强生成网络对人体关键节点预测的准确度。

3　系统环境与实验分析

3.1　系统设计环境

本文的模型和系统基于 Ubuntu18.04，在 Python3.7、PyTorch1.7.0 深度学习框架下进行编译。模型训练和验证基于 NVIDIA RTX3090-24G GPU。在真实场景下的实验中使用了奥比中光 Astra Pro 以 30 fps 的采样率采集视频数据。系统使用 PyQt5 进行界面设计，运动分析系统界面如图 8 所示。

图 8　运动分析系统界面

3.2　公共数据集实验

本文采用 COCO test-dev2017[25] 和 MPII[26] 两个数据集分别进行了训练和测试。COCO test-dev2017 是微软公司发布的关键点检测数据集，包含超过 20×10^4 张图像和 25×10^4 个标有关键点的人实例，每人包含 17 个可能的关键点，例如眼、臀部、脚踝等；MPII 单人人体姿态数据集由大约 2.5×10^4 张带有 16 个关键点标注的图像组成，涵盖了 410 种不同的人类活动，其中 1.5×10^4 张是训练样本，3×10^3 张图像是验证样本，7×10^3 张图像是测试样本。训练时，本文使用了单块 RTX3090 GPU 进行训练，初始学习率为 0.000 1，批量大小为 100，迭代次数为 300，优化器选择了带有正则化的 AdamW，优化器参数设置为 $\beta_1=0.000\,1$，$\beta_2=0.05$。

表 1 和 2 为本文算法与一些经典关键点检测算法在上述两个数据集上的效果对比。由于本文算法综合了多种算法模块,因此在测试中,不仅与专门用于关键点检测的算法进行了比较,同时也使用 HRNet 和基于 Tansformer 的人体姿态检测模型 TransPose 进行了测试。为了更好地比较算法之间的性能差异,本文采用了各算法原文献中共同采用的性能指标进行比较。

表 1　在 COCO test-dev2017 数据集上的测试结果　%

模型	AP	AP^M	AP^L	AP^{50}	AP^{75}
ResNet-101[27]	64.8	60.4	71.5	87.8	71.1
Mask R-CNN[28]	63.1	57.8	71.4	87.3	68.7
CPN[29]	72.1	68.7	77.2	91.4	80.0
HRNet-W48[20]	75.5	71.9	81.5	92.5	83.3
TransPose[30]	75.0	71.3	81.1	92.2	82.3
本文算法	72.8	70.2	78.3	91.6	81.1

表 2　在 MPII 数据集上的测试结果　%

模型	OpenPose	HRNet-W32	TransPose	本文算法
PCKh	88.8	92.3	93.5	92.4

表 1 中 AP(average precision)为平均精确度,APL 和 APM 分别表示在识别较大目标(area$>96^2$)和中等目标($32^2<$area$<96^2$)时的平均精确度,AP50 和 AP 75 分别表示性能指标 IoU 的阈值为 0.5 和 0.75 时的平均精确度。由表 1 可知,在 COCO test-dev2017 数据集上,本文算法在各评价指标上并非在处于最优位置。但是用于关键点位置检测中时,本文算法在牺牲了纯 Transformer 和深度卷积神经网络较大的运算量后,仍然在各评价指标中获得较好的结果。例如在 IoU 的阈值为 0.5 和 0.75 时以及对较小目标和较大目标的检测效果的平均精确度均高于经典人体姿态检测算法,但略低于 TransPose 和 HRNet-W48。

表 2 为在 MPII 数据集上的测试结果。与 COCO test-dev2017 数据集不同的是,MPII 数据集中并没有对人体部位进行掩码操作,且数据量偏少,因此其检测难度较低。本文在 MPII 数据集上测试时采用的评价指标为 PCK,即检测的关键点与其对应的正确标注间的归一化距离小于设定阈值的比例。MPII 中是以头部距离作为归一化参考,即 PCKh。表 2 中展示的结果同样说明了本文算法在人体关键点识别中的有效性。这种结果进一步证明了本文的算法是有效的。将算法部署至上位机系统中,在真实场景中测试时,特别是引入可变形注意力以及 GAN 之后,目标遮挡的预测已得到极大改善,然而表 1 和 2 中的评价指标无法将遮挡部分的预测效果进行表征。因此,本文在 3.3 节从质化评估和量化评估两个角度对算法在真实场景中检测人体关键点的能力进行了评价。

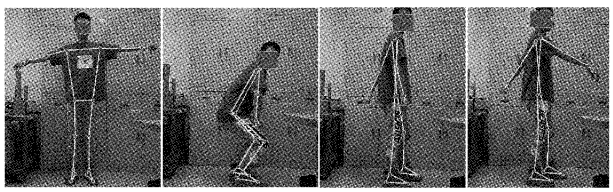
表 3 为上述实验中采用的模型的特征提取主干的参数量与本文算法模型中特征提取主干参数量的对比。从表 3 中可以看出,在 COCO test-dev2017 数据集和 MPII 数据集上表现较为优越的 HRNet-W48 和 TransPose 模型,其浮点运算量要比本文算法更高。本文算法的优越之处在于采用了部分 HRNet 进行局部建模与 Transformer 中全局信息编码的优势,以较小的计算代价获得了相近甚至超越大模型的特征提取能力。在结合可变形注意力以及生成对抗网络后,模型在小目标以及密集目标检测方面取得了更好的效果。

表 3　(主干)参数量对比实验结果

模型	HRNet-W48	TransPose	本文主干
参数量/M	63.6	17.5	21.1
浮点运算量/G	32.9	21.8	16.3

3.3　人体运动分析结果

本文算法在真实场景中的测试如图 9 所示,图 9(a)为正面站立情况下的关键点识别可视化结果,图 9(b)为侧面半蹲时的效果图,图 9(c)和图 9(d)分别为侧面手臂微抬以及手臂高抬时的效果图。一般的性能评价指标无法对遮挡部分的预测效果进行评估,因此,本文根据对受试人员在实际运动中的测量数据,从质化评估和量化评估两个角度对本文算法重建人体关键点的能力进行了评估。



(a) 正面站立图　(b) 侧面半蹲　(c) 侧面微摆臂　(d) 侧面高抬臂

图 9　关键点识别效果

1) 质化评估

为了更加清晰地表现出本文算法对关键点检测的贡献,特别是预测遮挡部分关键点的能力,本节通过将算法捕获的人体关键点位置重建为受试者的真实骨骼结构,并与使用物理测量方法获得的骨骼结构数据进行比对,进一步展示捕获数据与真实数据之间的差距。在测试过程中,本文以功能性动作筛查(functional movement screen, FMS)[31]的方法为例,规定并采集了跨栏步、深蹲、直线弓步蹲、肩部灵活性、直腿主动抬高、躯干稳定俯卧撑和旋转稳定七种动作,部分原始图和重建图的对比结果如图 10 所示。

图 10 中,圆形点位置分布为通过本文算法获得的受试者在动作时的骨骼点位图,三角形点位置分布为通过实际测量得到的经归一化获得的数据。其中,图 10(a)和(b)分别为从正面和侧面获得的跨栏步动作数据,图 10(c)和(d)分别为从正面和侧面获得的深蹲动作数据,图 10(e)为直

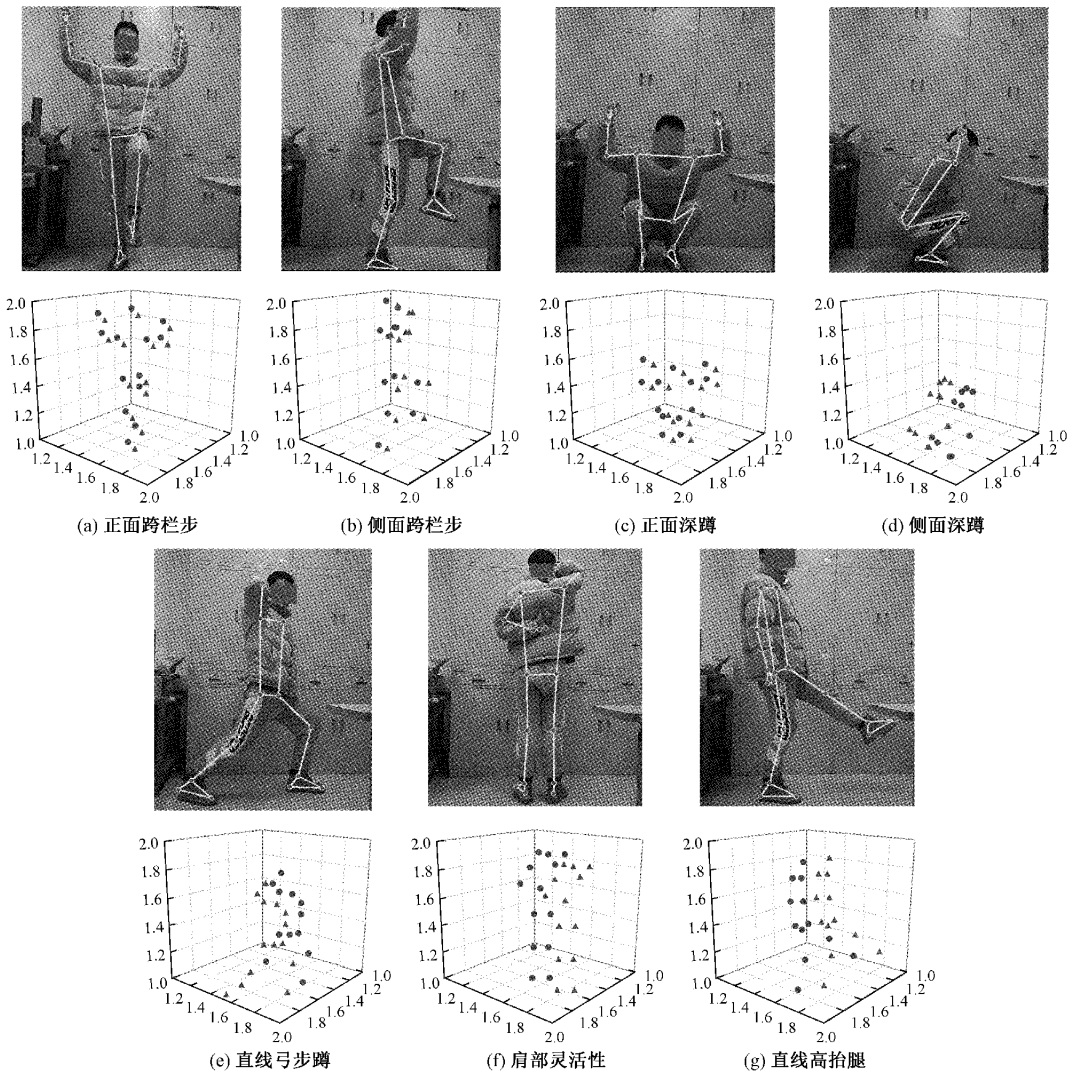


图 10　部分 FMS 动作识别与关键点重建对比

线弓步蹲的动作数据，图 10(f)为肩部灵活性的测量数据，图 10(g)为直腿主动抬高动作的数据。从图 10 中可以看出，在正视时的点位深度估计和侧视时的手臂遮挡预测都取得了较好的效果，在自然伸展的情况下所有骨骼点位置均可以被显式标出。在红蓝点位置的分布中也可以展现出本文算法捕获的数据与真实数据分布的相关性很高，但是在肢体交错情况复杂及发生严重遮挡时仍会出现问题。如图 10(d)中侧视受试者的深蹲动作，其骨骼点经还原后的膝部与髋部的位置与真实位置有轻微偏差。

2) 量化评估

人体关节之间的位置及角度变化可以明显表征出一个人的体态与运动过程。如图 11 所示，本文中设定的关节分析部分包括膝关节角度、肘关节角度、肩颈夹角、肩-躯干夹角、髋-躯干夹角、肩-大臂夹角、髋-大腿夹角、颈-躯干夹角、肩水平夹角、髋水平夹角、躯干竖直夹角、肘关节与锁骨距离、腕关节与锁骨距离、髋关节与锁骨距离、膝关节与锁骨距离、踝关节与锁骨距离、腕关节相对左右肩距离、踝关节相对左右髋距离、步频分析、角度特征汇总、X 坐标差汇总、Y 坐标差汇总、腿型分析图等。

式(5)为人体关键点间距离的计算式：

$$d_{x,y}=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \tag{5}$$

其中，$(x_i,\ x_j)$和$(y_i,\ y_j)$为图像二维像素坐标系中人体关键点的坐标。

由式(6)可计算出两个关节之间的向量角：

$$\theta=\arccos\frac{d_1^{\ 2}+d_2^{\ 2}+d_3^{\ 2}}{2d_1d_2} \tag{6}$$

基于上述知识，根据 3.3 节 1)中的功能性动作筛查(FMS)方法以及人体姿态的约束关系，本文将算法捕获的人体运动过程进行了多维度分析。分析的内容包括将获取的各关节位置进行相关计算，对关节之间的角度、相对位置关系等进行了多维度的分析。

以髋关节的相关分析为例，分析的可视化结果如图 12～14 所示。

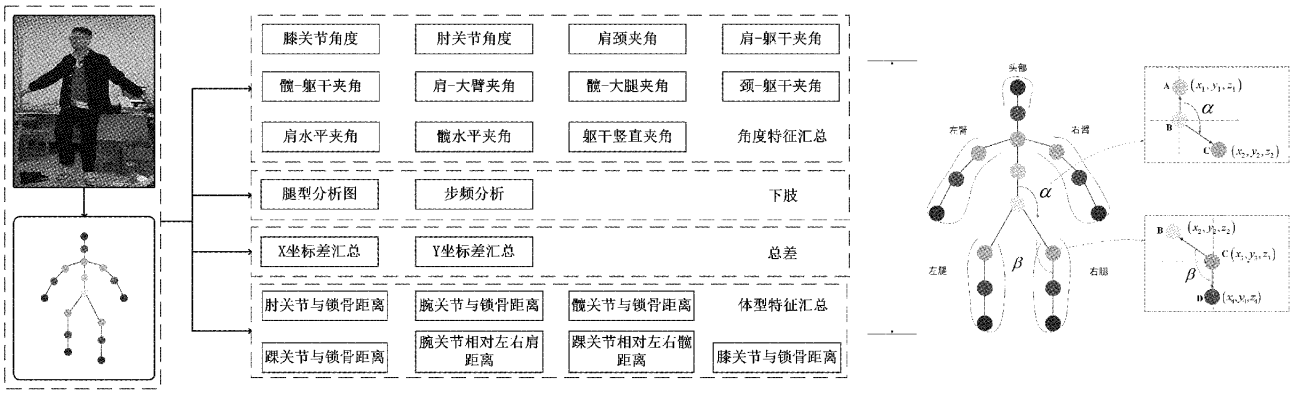
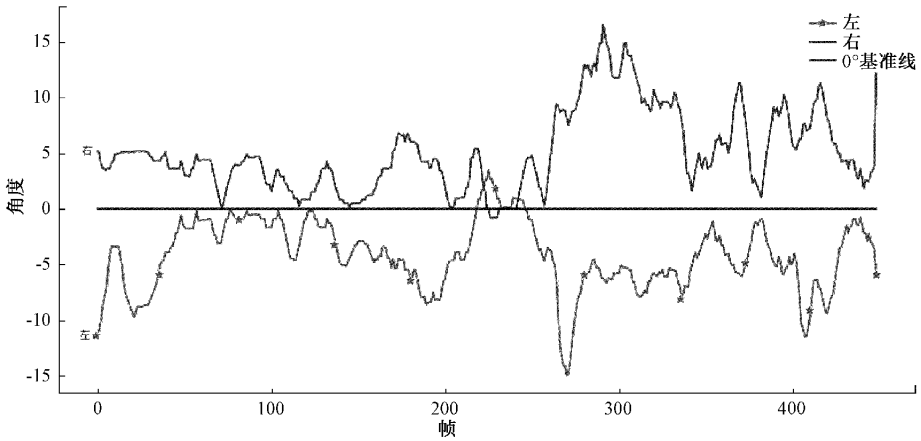
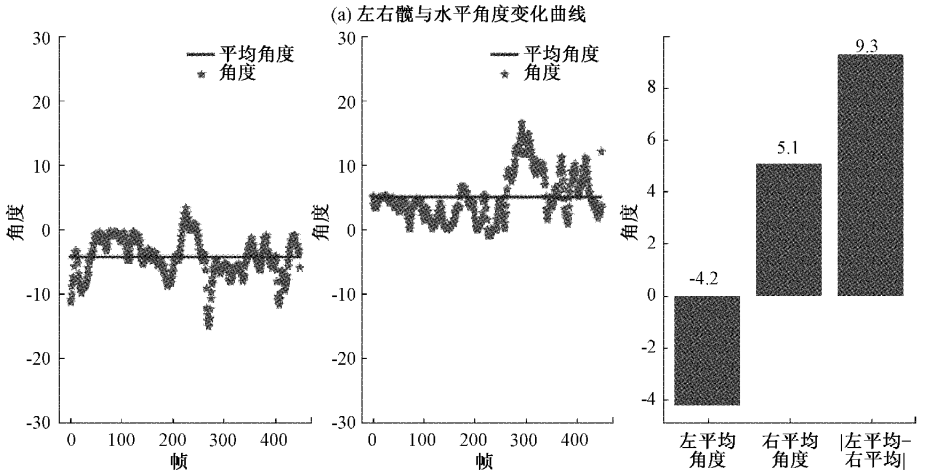


图 11　关节位置相关分析内容

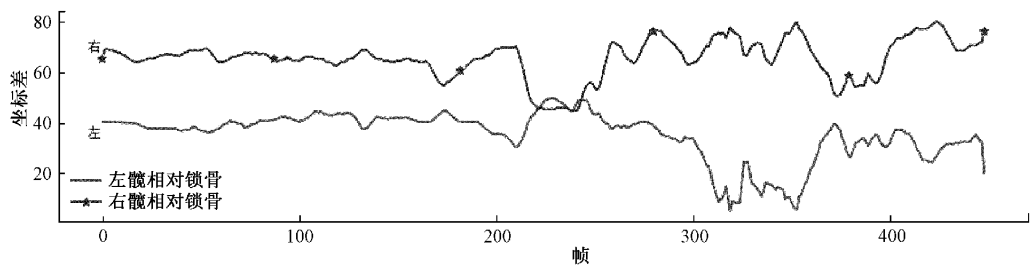


(a) 左右髋与水平角度变化曲线

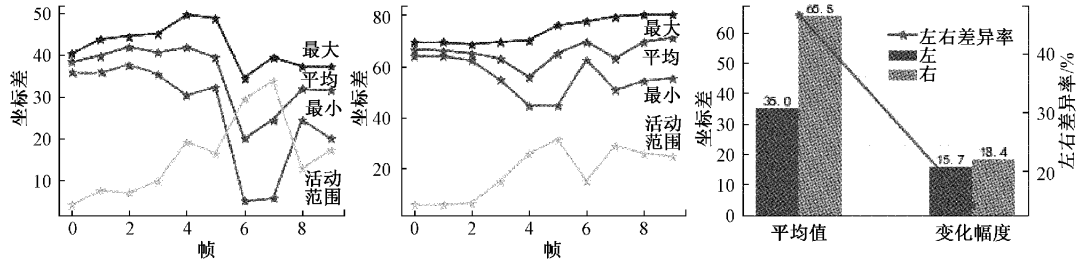


(b) 左髋与水平角度散点图　(c) 右髋与水平角度散点图　(d) 左右髋与水平角度平均值对比

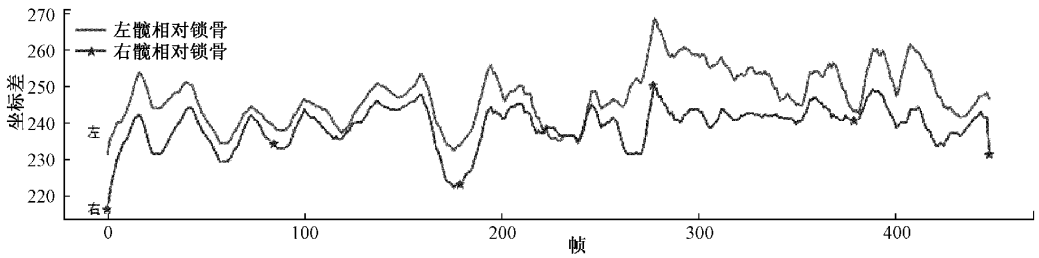
图 12　左右髋关节与水平角度分析



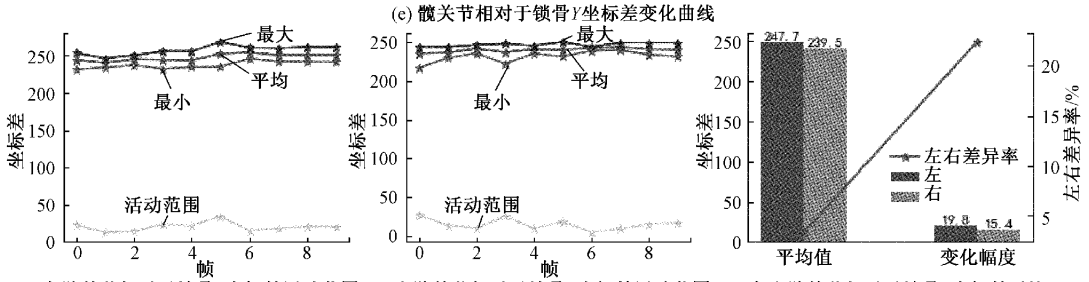
(a) 髋关节相对于锁骨X坐标差变化曲线



(b) 左髋关节相对于锁骨X坐标差活动范围 (c) 右髋关节相对于锁骨X坐标差活动范围 (d) 左右髋关节相对于锁骨X坐标差对比

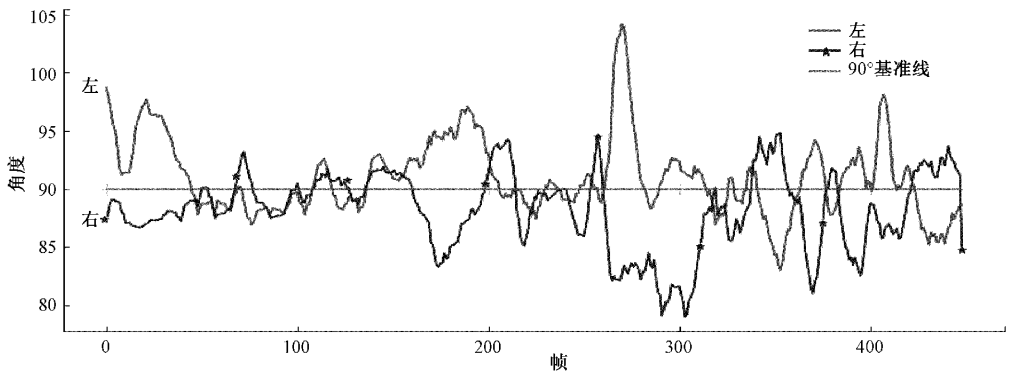


(e) 髋关节相对于锁骨Y坐标差变化曲线

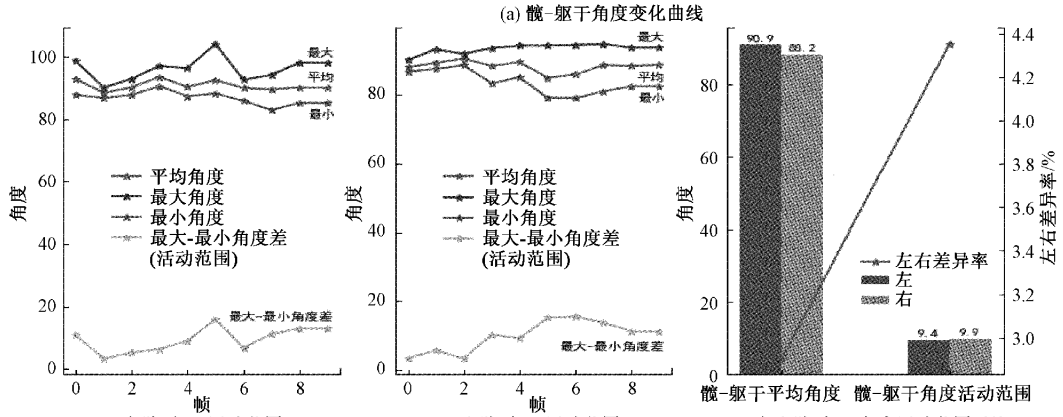


(f) 左髋关节相对于锁骨Y坐标差活动范围 (g) 右髋关节相对于锁骨Y坐标差活动范围 (h) 左右髋关节相对于锁骨Y坐标差对比

图 13　左右髋关节相对锁骨坐标变化



(a) 髋-躯干角度变化曲线



(b) 左髋-躯干活动范围　(c) 右髋-躯干活动范围　(d) 左右髋-躯干角度活动范围对比

图 14　髋关节与躯干角度变化

4 结 论

本文根据 Transformer 在全局关系建模上具有较强能力的特点,在模型结构和训练策略等方面进行了优化设计,结合人体姿态识别的下游任务,构建了一种能有效捕获人体关键点特征的算法,并在此算法基础上设计了一套人体运动分析系统。首先采用部分 HRNet 网络以及空间特征金字塔池化,从不同尺度对输入图像进行特征提取,以增强对图像中较小目标的检测能力;其次,通过在 Transformer 中引入可变形注意力,在控制注意力模块计算量的基础上对特征密集区域进行更加有效的处理,使算法的特征注意力点尽可能地保持在需要识别的位置,以减少计算冗余;最后使用生成对抗网络对 Transformer 编码生成的热图进行博弈训练,不断提高网络对关键点检测的能力。经过在 COCO test-dev2017 和 MPII 两个数据集上进行测试,其对应的性能指标 AP 和 PCKh 分别达到了 72.8%和 92.4%。在本文算法的基础上,根据功能性动作筛查的 7 种动作,以及人体姿态的约束条件,建立了一套包含人体各个关节运动分析的系统。从质化评估和量化评估两个角度进行分析后,本文建立的运动分析系统的可行性和准确性已得到验证。经过不断测试和调整,将所设计的系统在 Ubuntu 设备上进行部署,实现了在真实场景下的可视化运行,将来可应用于临床诊断,为住院医师和康复工程学科等提供一定的辅助和指导作用。

参考文献

[1] SEVCENKO K, LINDGREN I. The effects of virtual reality training in stroke and Parkinson′s disease rehabilitation: A systematic review and a perspective on usability [J]. European Review of Aging and Physical Activity, 2022, 19(1):4-4.

[2] 竺明月,刘志鹏,张可,等.基于磁定位的下肢康复步态采集系统[J].电子测量技术,2020,43(15):84-88.

[3] KIM J K, CHOO Y J, CHANG M C. Prediction of motor function in stroke patients using machine learning algorithm: Development of practical models[J]. Journal of Stroke and Cerebrovascular Diseases, 2021, 30(8):105856.

[4] 张宇,温光照,米思娅,等.基于深度学习的二维人体姿态估计综述[J].软件学报,2022,33(11):4173-4191.

[5] ROCHA I E, OROS FLORES M I, ALMANZA-OJEDA D L, et al. Kinect validation of ergonomics in human pick and place activities through lateral automatic posture detection[J]. IEEE Access, 2021, 9:109067-109079.

[6] 段俊臣,梁美祥,王瑞.基于人体骨骼点检测与多层感知机的人体姿态识别[J].电子测量技术,2020,43(12):168-172.

[7] LI X, ZHOU Z, WU J, et al. Human posture detection method based on wearable devices [J]. Journal of Healthcare Engineering, 2021(2):8879061.

[8] HAN J, SONG W, GOZHO A, et al. LoRa-based smart IoT application for smart city: An example of human posture detection[J]. Wireless Communications and Mobile Computing, 2020, 2020(2):1-15.

[9] GOCHOO M, TAN T H, BATJARGAL T, et al. Device-free non-privacy invasive indoor human posture recognition using low-resolution infrared sensor-based wireless sensor networks and DCNN[C]. 2018 IEEE International Conference on Systems, Man, and Cybernetics. Miyazaki: IEEE, 2018:2311-2316.

[10] JILIANG M U, XIAN S, JUNBIN Y U, et al. Flexible and wearable BaTiO3/polyacrylonitrile-based piezoelectric sensor for human posture monitoring[J]. Science China Technological Sciences, 2022, 65(4):858-869.

[11] WU C, AGHAJAN H, KLEIHORST R. Real-time human posture reconstruction in wireless smart camera networks[C]. International Conference on Information Processing in Sensor Networks. USA: IEEE Computer Society, 2008:321-331.

[12] 于乃功,柏德国.基于姿态估计的实时跌倒检测算法[J].控制与决策,2020,35(11):2761-2766.

[13] 刘今越,刘彦开,贾晓辉,等.基于模型约束的人体姿态视觉识别算法研究[J].仪器仪表学报,2020,41(4):208-217.

[14] PATRUNO C, MARANI R, CICIRELLI G, et al. People re-identification using skeleton standard posture and color descriptors from RGB-D data[J]. Pattern Recognition, 2019, 89:77-90.

[15] KWON B, KIM J, LEE K, et al. Implementation of a virtual training simulator based on 360° multi-view human action recognition[J]. IEEE Access, 2017, 5:12496-12511.

[16] ZHANG Y, LI Z, AN L, et al. Lightweight multi-person total motion capture using sparse multi-view cameras [C]. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, DOI: 10.1109/ICCV48922.2021.00551.

[17] LONG Y, YU H, LIU B. Depth completion towards different sensor configurations via relative depth map estimation and scale recovery[J]. Journal of Visual Communication and Image Representation, 2021, 80:103272.

[18] KIM Y, KIM D. A CNN-based 3D human pose estimation based on projection of depth and ridge data[J].

Pattern Recognition, 2020, 106:107462.

[19] JIANG M, YU Z, ZHANG Y, et al. Reweighted sparse representation with residual compensation for 3D human pose estimation from a single RGB image [J]. Neurocomputing, 2019, 358(SEP. 17):332-343.

[20] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. DOI: 10.1109/CVPR.2019.00584.

[21] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9):1904-1916.

[22] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need [C]. In Proceedings of the31st International Conference on Neural Information Processing Systems, 2017, DOI: 10.48550/arXiv.1706.03762.

[23] XIA Z, PAN X, SONG S, et al. Vision transformer with deformable attention[C]. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, DOI: 10.48550/arXiv.2201.00520.

[24] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks [J]. Communications of the ACM, 2020, 63(11): 139-144.

[25] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]. European Conference on Computer Vision. Springer International Publishing, 2014, DOI: 10.1007/978-3-319-10602-1_48.

[26] ANDRILUKA M, PISHCHULIN L, GEHLER P, et al. Human pose estimation: New benchmark and state of the art analysis[C]. Computer Vision and Pattern Recognition, 2014, DOI: 10.1109/CVPR.2014.471.

[27] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, DOI: 10.1109/CVPR.2016.90.

[28] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017, DOI: 10.1109/ICCV.2017.322.

[29] CHEN Y, WANG Z, PENG Y, et al. Cascaded pyramid network for multi-person pose estimation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2018, DOI: 10.48550/arXiv.1711.07319.

[30] YANG S, QUAN Z, NIE M, et al. Transpose: Keypoint localization via Transformer [C]. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, DOI: 10.48550/arXiv.2012.14214.

[31] HARVEY A, GRAHAM H K, MORRIS M E, et al. The functional mobility scale: Ability to detect change following single event multilevel surgery [J]. Developmental Medicine & Child Neurology, 2007, 49(8): 603-607.

作者简介

孙文昊，硕士研究生，主要研究方向为深度学习与机器人技术。

E-mail:whsun@tute.edu.cn

路光达（通信作者），博士，教授，主要研究方向为机器视觉、机械设计及理论、康复机器人技术等。

E-mail:lugd1229@163.com

秦转萍，博士，讲师，主要研究方向为康复机器人、医学图像处理等。

E-mail:qzp2013@tute.edu.cn

郭庭航，博士，讲师，主要研究方向为先进测量技术、检测技术与仪器等。

E-mail:guotinghang@tute.edu.cn

赵壮壮，硕士研究生，主要研究方向为康复机器人设计。

E-mail:Z1069593108@163.com