DOI: 10.19650/j.cnki.cjsi.J2514611

基于视觉的无标记运动学分析方法*

黄高华[1,2],李玉榕[1,2],姜海燕[1,2],陈建国[1,2]
(1. 福州大学电气工程与自动化学院 福州 350108; 2. 福建省医疗器械和医药技术重点实验室 福州 350108)

摘 要:针对基于标记的光学系统成本高、耗时长、专业性强的问题,提出了一种采用两个视角的视觉无标记运动学分析方法,旨在实现便捷、低成本的运动学评估。首先,集成 Swin Transformer 的全局上下文建模能力、坐标注意力的精准位置感知能力、双向特征金字塔网络的多尺度特征融合能力,构建二维特征提取架构,克服自遮挡、关键点小目标检测问题,有效提取二维特征。其次,提出以关键点位置合理性和肢体长度一致性为关节上下文约束的三角测量方法,并结合人体参数化模型对三维关键点进行重构,提高关键点估计精度。最后,构建关键点增强模型,获取解剖标记集并结合肌骨模型进行运动学分析。公开数据集上的运动学评估实验表明,所提方法的平均关节角度误差为 8.59°,平均关节位置误差为 42.02 mm,优于现有的高性能方法。同时,为验证方法在真实场景下的适用性,以商用动作捕捉系统 Xsens 作为评估标准,并与当前主流方法 OpenCap 进行比较,分别对肩关节和步态运动学展开分析。实验结果表明,在肩关节和步态运动学上,所提方法与 Xsens 的相关系数分别为 0.92 和 0.86,较 OpenCap 的相关系数分别提高 9.52% 和 7.40%;角度误差分别为 13.97°和 3.12°,较 OpenCap 的误差分别下降 27.01% 和 25.18%。综上所述,在公开数据集和真实场景下,所提方法可实现比当前主流方法更准确的运动学分析,对促进运动学分析相关应用的推广具有重要意义。

关键词: 视觉;无标记;三维关键点估计;关键点增强;肌骨模型;运动学分析
中图分类号: TP391.4 TH77 **文献标识码**: A **国家标准学科分类代码**: 510.40

Markerless vision-based kinematic analysis method

Huang Gaohua[1,2], Li Yurong[1,2], Jiang Haiyan[1,2], Chen Jianguo[1,2]
(*1. College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China*;
2. Fujian Provincial Key Laboratory of Medical Instrument and Pharmaceutical Technology, Fuzhou 350108, China)

Abstract: To address the high cost, time-consuming nature, and specialized expertise required by marker-based optical systems, this article proposes a visual kinematic analysis method utilizing two viewpoints to achieve convenient, low-cost kinematic evaluation. First, a two-dimensional feature extraction architecture is established by integrating the global context modelling capability of the Swin Transformer, the precise positional awareness of coordinate attention, and the multi-scale feature fusion capability of the bidirectional feature pyramid network. It overcomes challenges such as occlusion and small target detection for keypoints, enabling effective extraction of two-dimensional features. Secondly, a triangulation method is proposed, employing joint contextual constraints based on keypoints position plausibility and limb length consistency. This is combined with a parametric human model to reconstruct 3D keypoints, enhancing estimation accuracy. Finally, a keypoint augmentation model is formulated to obtain an anatomical label set, which is then integrated with a musculoskeletal model for kinematic analysis. Kinematic evaluation on public datasets demonstrates an average joint angular error of 8.59° and average joint positional error of 42.02 mm, outperforming existing high-performance methods. To validate real-world applicability, commercial motion capture system Xsens serves as the evaluation benchmark against the mainstream OpenCap method, with analyses conducted on shoulder joint and gait kinematics, respectively. Experimental results show that for shoulder joint and gait kinematics, the proposed method achieves correlation coefficients of 0.92 and 0.86, respectively, with Xsens, representing improvements of 9.52% and 7.40% over OpenCap. Angular errors are reduced to 13.97° and 3.12°, respectively, marking decreases

收稿日期:2025-10-22 Received Date: 2025-10-22
*基金项目:国家自然科学基金(62373108)、福建省技术创新重点攻关及产业化项目(校企联合类)(2024XQ001)资助

of 27.01% and 25.18% compared to OpenCap. In summary, the proposed method achieves more accurate kinematic analysis than current mainstream approaches on both public datasets and in real-world scenarios, holding significant implications for advancing applications related to kinematic analysis.

Keywords: vision; markerless; 3D keypoint estimation; keypoint enhancement; musculoskeletal model; kinematic analysis

0 引　　言

运动学分析是研究人体运动规律的重要学科,可辅助疾病的早期筛查,推动早期干预的实施,并优化患者康复方案的制定[1]。临床上依赖基于标记的光学系统来获取高精度的运动学结果。然而,基于标记的光学系统需要昂贵的高速红外相机阵列、专用的空间以及校准设备。同时还需要专业的操作人员进行耗时的红外标记点放置以及数据处理与分析。这些因素限制了人体运动学分析在临床上的应用和推广[2]。发展高效且低成本的人体运动学分析方法,有助于推动其相关应用的深化与普及。

随着计算机视觉和人工智能技术的快速发展,基于视觉的无标记人体运动学分析方法有望实现便捷、高效、无标记的运动学分析[3-4]。基于视觉的无标记人体运动学分析方法利用人体姿态估计算法提取人体三维关键点,并结合肌骨模型进行运动学分析[5]。文献[6-7]提出了新颖的架构,使用端到端的模式,直接利用图像数据得到生物力学骨骼模型参数,实现了便捷化的运动学分析。然而,由于图像和生物力学骨骼模型的配对数据缺乏多样性,端到端方法的泛化性能有待检验。并且,与进行肌骨模拟的方法相比,直接得到生物力学骨骼模型参数的方法不利于进一步分析动力学相关内容,比如关节力矩。而动力学分析在康复医学、运动医学和骨科学中至关重要,通过动力学分析可以揭示疾病的根本力学病因和康复的治疗机理[8]。在实际应用中,为了追求运动学分析的精度以及可靠性,大部分应用采用先获取人体三维关键点,再结合肌骨模型进行肌骨模拟的方法[9]。

人体姿态估计算法研究中,提取人体三维关键点主要有直接估计和先二维估计再三维估计两种范式[10]。直接估计的范式直接从图像中估计人体三维关键点,由于图像和三维关键点的配对数据缺乏多样性,并且深度模糊的影响,直接估计的范式对复杂场景的鲁棒性不高[11]。得益于二维关键点估计算法的成熟与普及,先估计二维关键点再从二维关键点提升到三维关键点的范式是目前研究的主流[12]。在二维关键点估计算法中,SimpleBaseline 方法[13]使用残差网络(residual network, ResNet)模型进行特征提取,并直接回归二维关节中心点的坐标。这种方法适用于轻量级网络和实时应用,但在面对复杂姿态时表现较差,鲁棒性不足。为了有效克服这些缺陷,Xu 等[14]使用基于关节中心点热图的方法,采用编码-解码架构,通过多尺度特征融合提高了二维关键点估计精度。Liang 等[15]提出高分辨率网络(high-resolution network, HRNet)架构,通过高分辨率特征和低分辨率特征的融合进一步提高了二维关键点估计精度。随着 Transformer 的发展,ViTPose (vision Transformer for human pose estimation)方法利用 Transformer 强大的全局上下文建模能力,以 Transformer 为主干,通过回归方式获取二维关键点[16]。在医学应用中,必须综合考虑二维关键点估计算法的精度、可靠性、鲁棒性[17]。

将二维关键点提升到三维关键点的方法主要包含三角测量方法、拟合人体模型方法、数据驱动的深度学习方法[18]。三角测量方法是医学应用中的常用方法,具有原理简单清晰,容易实现的优势。然而,基础的三角测量方法对二维关键点的检测误差高度敏感并且对遮挡和漏检的容错性较差。体素三角测量方法通过重投影和特征聚合的方式克服了这些缺陷[19]。然而,体素三角测量方法未考虑运动过程中关键点位置的合理性和肢体长度的一致性。同时,随着人体参数化模型的发展[20],使用人体参数化模型可以充分利用人体姿态先验,有利于提高三维关键点估计精度。

获取人体三维关键点后,结合肌骨模型进行肌骨模拟后可得到运动学结果。通过肌骨模拟得到运动学结果一般需要经过模型缩放、配准、逆运动学等步骤[21]。进行肌骨模拟时,关键点数量过于稀疏会导致无法对动作进行充分约束,引起运动学分析精度下降[22]。为解决关键点过于稀疏问题,SynthPose (synthetic pose)方法[23]通过人体网格来合成解剖标记数据标签,进而利用合成的数据来微调关键点估计模型以获得更为密集的解剖标记集。文献[24]则通过设计标记增强模型来解决这个问题。标记增强模型以稀疏的关键点为输入得到更为全面的解剖标记集。

总的来说,基于双目立体视觉的方法[25-26],本文提出一种采用两个视角的视觉无标记运动学分析方法。采用双目相机,避免了基于标记的光学系统的高成本硬件要求,并且通过视觉的方式实现无标记的运动学分析,避免了基于标记的光学系统方法中复杂的反光标记放置和数据处理与分析过程。为解决视觉无标记运动学分析中的自遮挡、关键点估计精度不足、关键点数量稀疏问题,本文方法的主要创新点包括:

1)为克服自遮挡以及关键点小目标检测问题,集成 Swin Transformer 的全局上下文建模能力、坐标注意力的

精准位置感知能力以及双向特征金字塔网络的多尺度特征融合能力设计二维关键点特征提取架构,实现更加有效的二维关键点特征提取。

2)为提高关键点估计精度,在体素三角测量方法中引入具有关节位置合理性和肢体长度一致性的关节上下文约束,并利用人体参数化模型对三维关键点进行重构。

3)为解决关键点过于稀疏问题,设计手臂和身体关键点增强模型,对三维关键点进行增强获取对应的解剖标记集,并通过 Opensim 平台进行肌骨模拟来获取运动学结果。

最后,为验证方法的适用性,本文在真实场景下,分别对肩关节运动学和步态运动学展开分析。

1 整体方法概述

本文方法流程如图 1 所示,流程包含视频数据获取、三维解剖标记点估计、Opensim 运动学分析 3 个步骤。首先,使用校准并同步的双目相机进行数据采集。其次,对采集到的视频数据进行二维特征提取、三角测量,得到初始的三维关节位置。并结合人体参数化模型对三维关节位置进行重构。进一步,使用关键点增强模型对重构的三维关节位置进行关键点数据增强,得到三维解剖标记点。最后,使用三维解剖标记点结合肌骨模型在 Opensim 平台上进行运动学分析,得到运动学参数。

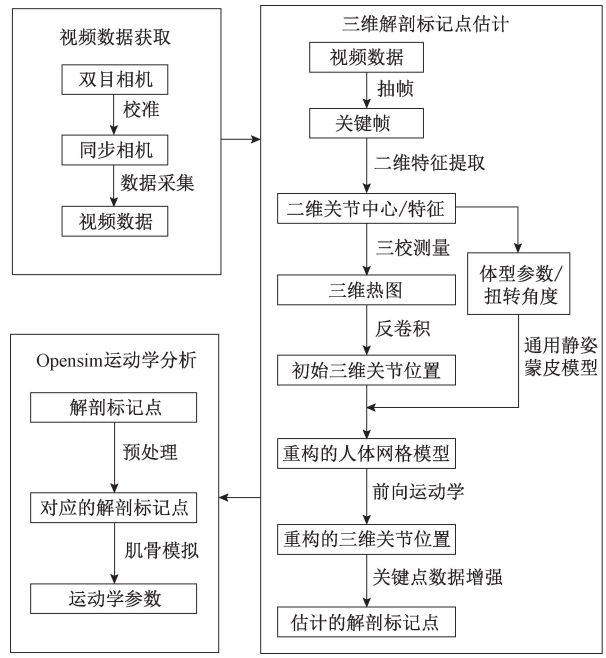


图 1 整体流程

Fig. 1 Overall flowchart

2 运动学分析方法

本章阐述运动学分析所涉及的方法与模型。

2.1 二维特征提取

YOLO-Pose (you only look once for pose estimation)[27] 是人体关节中心点检测的先进方法之一。在 YOLO-Pose 的骨干网络中,通过 Focus 的切片操作实现无信息丢失的下采样,能够保留更多细粒度特征。并且利用 C3 模块的跨阶段融合优势构建层级结构,在提取多分辨率特征的同时保持较高的计算效率。在 YOLO-Pose 的瓶颈层中,利用多分辨率特征构建多尺度特征金字塔,并通过自顶向下和自底向上的路径方式增强浅层特征和深层特征的交互,使提取的特征能够包含强语义信息和精准的空间细节。因此,本文以 YOLO-Pose 为基线。

然而,由于卷积操作的局部性,YOLO-Pose 模型提取并利用全局上下文信息的能力有限。并且连续的下采样操作会导致空间精度的丢失。同时,由于特征融合机制导致处于浅层的小目标特征丢失。

针对这些缺陷,首先,在 YOLO-Pose 骨干网络的 C3 层中嵌入 Swin Transformer 编码器[28],利用 Swin Transformer 的全局自注意力机制,提高模型提取并利用全局上下文信息的能力,缓解姿态估计中的自遮挡问题。C3STR 模块的结构如图 2 所示,其公式可表示为式(1),即:

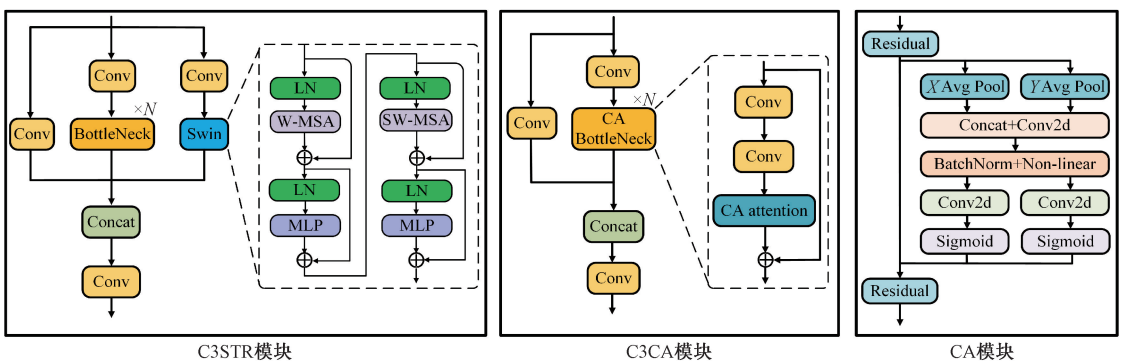


图 2 C3 模块的改进结构

Fig. 2 The improved structure of the C3 module

$$\boldsymbol{F}_{out} = \mathrm{Conv}(\mathrm{Concat}(\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3)) \tag{1}$$

其中,$\boldsymbol{F}_1$ 为通过基础卷积层提取的特征,$\boldsymbol{F}_2$ 为通过 N 个 Bottleneck 残差模块提取的特征,$\boldsymbol{F}_3$ 为通过 Swin Transformer 编码器捕获的全局特征。

其次,CA 模块、C3CA 模块结构如图 2 所示,在 YOLO-Pose 的连接头部的每个 C3 层中嵌入坐标注意力(coordinate attention, CA)机制[29],增强算法对不同尺度下关节中心点位置信息的敏感性,缓解由于下采样操作导致的空间精度丢失问题,以提高关节中心回归精度。CA 模块先进行坐标信息嵌入,如式(2)所示。

$$\begin{cases} \boldsymbol{z}_c^h = \dfrac{1}{W}\displaystyle\sum_{0\leqslant j<W} \boldsymbol{F}_{in}(:, j) \\ \boldsymbol{z}_c^w = \dfrac{1}{H}\displaystyle\sum_{0\leqslant i<H} \boldsymbol{F}_{in}(i, :) \end{cases} \tag{2}$$

经过坐标信息嵌入后,进一步计算注意力权重并进行特征加权输出,表示为如式(3)和(4)所示。

$$\boldsymbol{a} = \mathrm{Sigmoid}(\mathrm{Conv}(\mathrm{Concat}(\boldsymbol{z}_c^h, \boldsymbol{z}_c^w))) \tag{3}$$

$$\boldsymbol{F}_{out} = \boldsymbol{F}_{in} \otimes \boldsymbol{a} \tag{4}$$

其中,$\boldsymbol{z}_c^h$ 和 $\boldsymbol{z}_c^w$ 分别是高度和宽度方向的全局池化特征,$\otimes$ 表示逐元素相乘。

最后,引入了基于小目标检测层的加权双向特征金字塔网络(bidirectional feature pyramid network, BiFPN)[30],通过加权的方式进行特征融合,缓解特征融合过程中浅层的关键点小目标特征丢失问题。BiFPN 的结构如图 3 所示。其结构可以表示为如式(5)所示。

$$\boldsymbol{P}_l^{out} = \mathrm{Conv}\left(\frac{\sum_i \boldsymbol{w}_i \cdot \boldsymbol{P}_l^i}{\sum_i \boldsymbol{w}_i + \theta}\right) \tag{5}$$

其中,$\boldsymbol{P}_l^i$ 表示参与融合的各层输入特征,$\boldsymbol{w}_i$ 表示可学习权重,θ 为数值稳定项,其值为 1×10^{-4}。

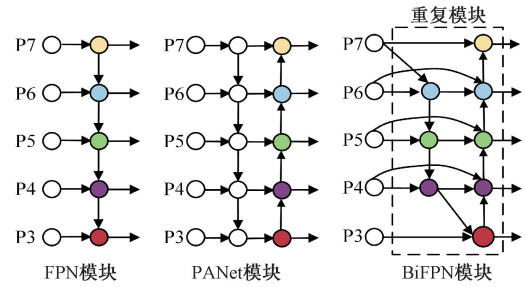


图 3　FPN、PANet 和 BiFPN 架构比较

Fig. 3　Comparison of FPN, PANet and BiFPN architectures

本文将改进后的模型称为 YOLO-Rpose,其结构如图 4 所示,并将其作为提取二维特征的主干。

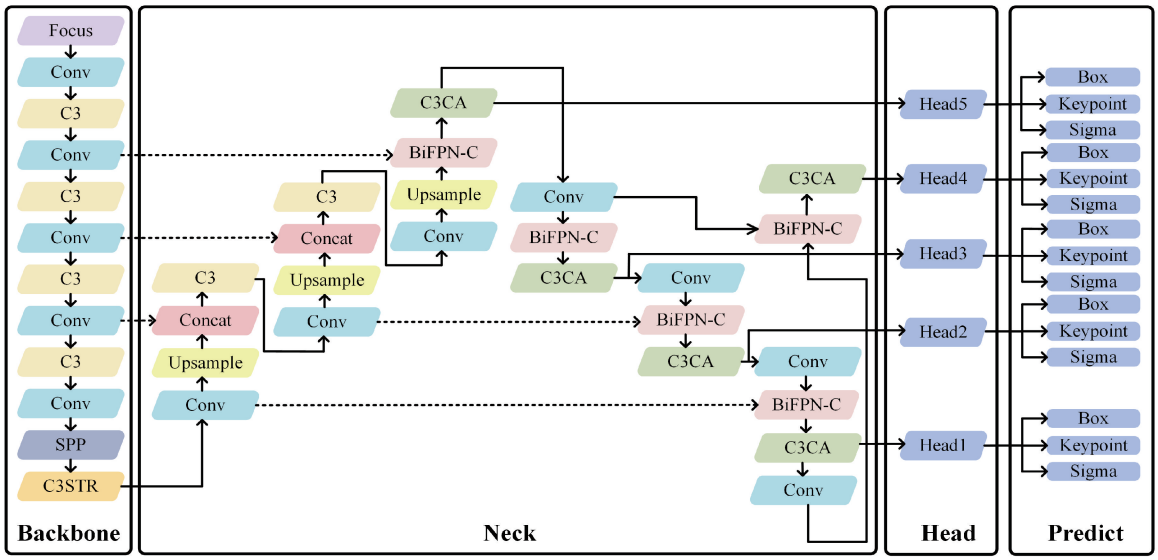


图 4　YOLO-Rpose 模型结构

Fig. 4　YOLO-Rpose model architecture

2.2　三角测量与关键点重构

体素三角测量方法通过重投影和特征聚合的方式实现端到端优化,可以有效整合多视角特征,提高模型的鲁棒性。该方法的核心思想是将二维关节特征运用三角测量的方法投影到三维体积中,形成三维体素网格。三维姿态估计网络如图 5 所示,本文使用体素三角测量的重投影和特征聚合的方式来处理多视角特征,并引入全局注意力机制(global attention, GA)与成对注意力机制(pairwise attention, PA)[31]。通过 GA 和 PA 对运动过程中的关节位置和肢体长度进行约束,形成具有关节上下文约束的三维关节热图估计。

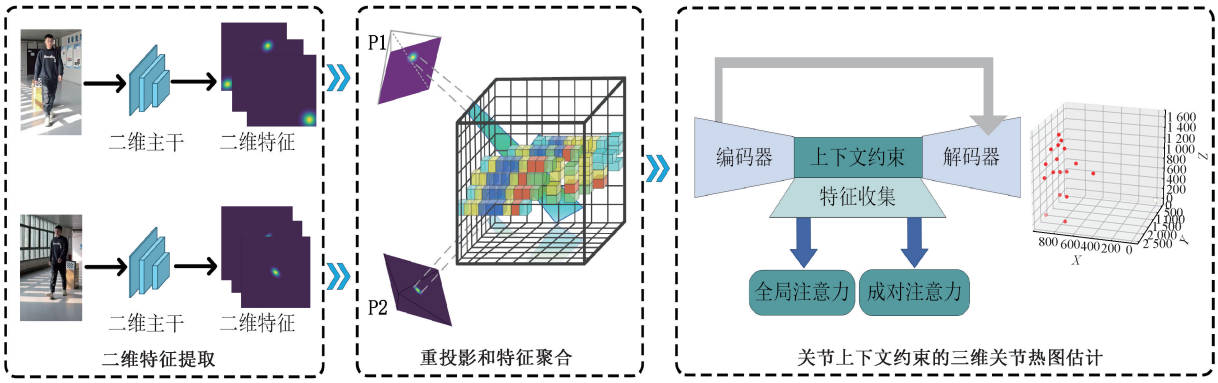


图 5 三维姿态估计网络

Fig. 5 Three-dimensional pose estimation network

由于简化骨架模型缺少轴向旋转自由度,会降低关节位置估计的准确性。本文借鉴文献[32]中的思路,将原始旋转分解成扭转和摆动,以提高三维关节中心点的估计精度。具体来说,姿态重构网络如图 6 所示,以回归得到的人体三维姿态为初始人体三维姿态。进一步利用全连接网络,从多视角聚合的关键点特征学习人体的关节扭转角度参数以及人体形状参数,所使用的全连接网络为 3 层结构,前两层为全连接层,每层包含 1 024 个神经元,最后一层为输出层,包含 56 个神经元,其中 10 维是形状参数,46 维是扭转角度参数。

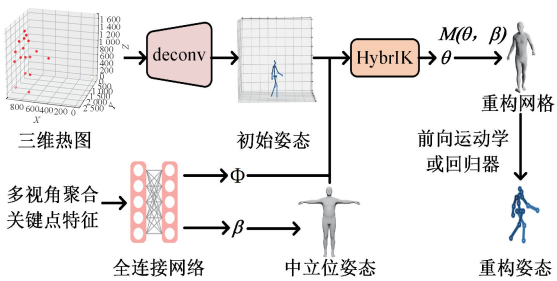


图 6 姿态重构网络

Fig. 6 Posture reconstruction network

最后,以初始人体三维姿态、扭转角度参数、以形状参数初始化的中立位姿态为输入,利用混合逆运动学模块(hybrid inverse kinematics, HybrIK),得到重构的人体网格。并使用回归器或前向运动学从重构的人体网格获取重构的人体三维姿态。姿态重构过程见算法 1。

算法 1:关键点重构算法

输入:初始三维姿态 $\boldsymbol{J}$,多视角聚合关键点特征 $\boldsymbol{F}$

输出:重构网格 $\boldsymbol{M}$,重构姿态 $\boldsymbol{J\_refined}$

1. #参数回归
2. $\boldsymbol{\beta}$, $\boldsymbol{\theta\_twist}$ ← FCN($\boldsymbol{J}$, $\boldsymbol{F}$) #形状参数和扭转角
3. #中立位姿态获取
4. $\boldsymbol{T\_pose}$ ← SMPL_Template($\boldsymbol{\beta}$)
5. #摆动旋转计算
6. **for** 每个关节 j **do**
7. #当前骨骼向量 $\boldsymbol{v\_target}$ ← $\boldsymbol{J}[j]-\boldsymbol{J}[\boldsymbol{parent}(j)]$
8. #标准姿态骨骼向量 $\boldsymbol{v\_rest}$ ← $\boldsymbol{T\_pose}[j]-\boldsymbol{T\_pose}[\boldsymbol{parent}(j)]$
9. #摆动旋转 $\boldsymbol{R\_swing}$ ← CalculateRotation($\boldsymbol{v\_rest}$, $\boldsymbol{v\_target}$)
10. **end for**
11. #完整旋转构建
12. $\boldsymbol{R\_local}$ ← $\boldsymbol{R\_swing}\cdot\boldsymbol{R\_twist}(\boldsymbol{\theta\_twist})$
13. #重构网格
14. $\boldsymbol{M}$ ← $\boldsymbol{T\_pose}$, $\boldsymbol{R\_local}$
15. #重构姿态
16. $\boldsymbol{J\_refined}$ ← ForwardKinematics($\boldsymbol{T\_pose}$, $\boldsymbol{R\_local}$)
17. **OR** $\boldsymbol{J\_refined}$ ← MeshRegressor($\boldsymbol{M}$)

总的来说,本文以 YOLO-Rpose 为二维主干网络,提取二维特征。进一步,通过重投影和特征聚合的方法处理多视角二维特征,并引入 GA 和 PA 模块保证关节位置的合理性和肢体长度的一致性。最后,通过 HybrIK 模块对人体三维姿态进行重构,进一步提高人体三维姿态估计的精度。

2.3 关键点增强

为缓解人体三维姿态过于稀疏而造成的运动学分析精度下降问题,本文基于文献[33]中的思路,利用长短期记忆(long short-term memory, LSTM)网络设计关键点增强模型。在上肢部分,设计了手臂模型,以手臂和肩部的 7 个关节中心点为输入预测 8 个解剖标记点,用于定义肩关节和肘关节的运动。在身体部分,设计了身体模型,以下肢和躯干部位的 15 个关键点为输入预测 35 个解剖标记点,用于定义下肢和躯干部位的运动。通过标记点增强模型将原来重构的人体三维姿态增强到具有

43 个解剖标记点的解剖标记集。关键点增强模型示意如图 7 所示。

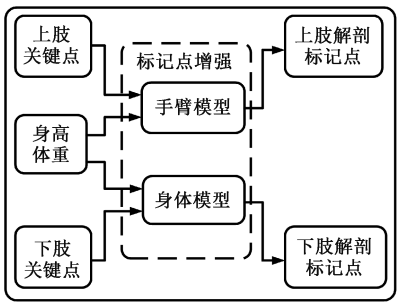


图 7　关键点增强模型示意图

Fig. 7　Diagram of the human keypoint enhancement model

2.4　Opensim 运动学分析

肌骨模型能够对人体的骨骼和肌肉的物理行为以及相互作用进行精确建模。结合肌骨模型进行运动学分析可以缓解关键点遮挡的影响和不合理姿势的产生。在 Opensim 平台,利用获得的解剖标记集结合肌骨模型进行运动学分析。运动学分析主要需要经过模型缩放、配准、逆运动学等步骤。首先,通过模型个性化缩放使模型适配不同个体。其次,通过配准对模型的姿态进行初始化。最后,进行逆运动学,通过优化的方法最小化肌骨模型上的虚拟标记点和解剖标记点位置之间的误差,从而得到运动学结果。

3　公开数据集下的实验结果与分析

本章依次对各模块模型的训练过程以及实验设置进行详细阐述,并通过相关实验验证所提改进方法的有效性。实验所采用的硬件与软件环境如表 1 所示。

表 1　实验设备及软件环境

Table 1　Experimental equipment and software environment

开发环境	参数
CPU 处理器	13th Gen Intel(R)Core(TM)i7-13700KF 3.40 GHz
GPU	NVIDIA GeForce RTX 4090
内存	32 G
操作系统	Ubuntu 20.04.6
深度学习计算库	Pytorch 2.3.1+cu118
编程语言	Python 3.9.21
虚拟环境管理	Anaconda

3.1　二维特征提取主干训练与评估

本文使用 COCO 全身数据集[34] 和 Haple 全身数据集[35] 对二维特征提取主干进行训练。COCO 全身数据集具有超过 20 万张的图像数据,并且包含 133 个全身关键点标注。Haple 全身数据集具有超过 25 万张的图像数据,并且包含 136 个全身关键点标注。两个数据集充分覆盖了日常与特定任务下的各类动作,包含丰富的场景与姿态多样性。

YOLO-Rpose 是在 YOLO-pose 的基础之上进行改进的,采用与 YOLO-pose 模型相同的损失函数进行训练,损失函数主要由热图回归损失、中心度损失、关键点坐标回归损失 3 部分组成。同时,利用 YOLO-pose 的部分预训练模型加速训练过程。

训练数据图像使用真实的人体边界框进行裁剪,并采用随机缩放(±30%)、随机旋转(±40°)以及图片翻转数据增强技术。训练时输入图像的大小设置为 256×192,采用 Adam 优化器进行训练,训练周期设置为 300,批次大小设置为 32,初始学习率设置为 1×10^{-3},并采用学习率阶梯式衰减的策略,分别在第 200 个周期和第 250 个周期进行衰减,衰减率为 0.1。

为全面评估算法的性能,使用平均精确度(average precision, AP)指标评估算法精度,包括 AP、AP50、AP75。使用平均召回率(average recall, AR)指标评估模型在不同人体关节或部位上能有效识别和预测目标的能力。在 COCO 全身数据集上,以 YOLO-pose 为基线模型,对 3 个改进点展开实验。二维特征提取实验结果如表 2 所示(加粗表示最优结果),本文算法的平均精确度为 66.13%,较基线模型提高了 5.14%,平均召回率为 72.83%,较基线模型提高了 3.40%。与基线模型相比,本文的方法能更加充分地提取图像数据中的二维特征,有利于后续运动学分析的展开。

表 2　二维特征提取结果

Table 2　Two-dimensional feature extraction results

(%)

C3STR	C3CA	BiFPN	*AP*	*AP*50	*AP*75	*AR*
			62.90	87.70	69.40	70.30
√			62.66	85.31	69.60	71.37
√	√		64.02	87.13	70.77	70.96
√	√	√	**66.13**	**88.92**	**72.30**	**72.83**

3.2　三角测量与关键点重构训练与评估

对于体素三角测量部分,使用 3.1 节中得到的模型作为二维特征提取主干的预训练模型。进一步,使用 H3WB 数据集[36] 和 Human3.6 m 数据集[37] 进行二维主干网络与三维体素网络的联合训练与微调。H3WB 数据集具有超过 10 万张的图像数据,并且包含 67 个三维全身关键点标注。Human3.6 m 数据集具有超过 360 万帧视频数据,并且包含 32 个三维身体关键点标注。两个数

据集涵盖多样的姿态数据。训练过程采用 4 个摄像头的 RGB 图像数据。在评估过程中可使用 2 个及以上的摄像头的图像数据进行评估。采用和 3.1 节中一样的数据增强技术。训练时输入图像的大小设置为 256×192,采用 Adam 优化器进行训练,训练周期设置为 50,批次大小设置为 32,二维主干网络的学习率设置为 1×10^{-4},三维体积网络的学习率设置为 5×10^{-4}。

对于关键点重构部分,以体素三角测量得到的热图和中间过程得到的三维关键点特征为输入,结合文献[32]中提供的预训练权重。设置输入图像的大小为 256×192,采用 Adam 优化器进行训练,训练周期设置为 50,批次大小设置为 32,初始学习率设置为 1×10^{-4},在第 40 个周期时下降为 1×10^{-5},之后保持不变。在整个训练以及微调过程中,体素三角测量部分的损失函数主要使用三维位置 L1 损失和弱热图正则化项。关键点重构部分的损失函数主要使用形状参数 L2 损失和姿态参数 L2 损失。

采用平均每个关节位置误差(mean per joint position error, MPJPE)指标评估模型三维关键点估计的性能。此外,采用平均每个肢体长度误差(mean per limb length error, MPLLE)和平均每个肢体角度误差(mean per limb angle error, MPLAE)指标评估模型对肢体长度约束的效果。在 Human3.6 m 数据集的测试集部分展开评估,三角测量与关键点重构结果如表 3 所示(加粗表示最优结果),本文结果在 MPJPE、MPLLE、MPLAE 这 3 个评价指标上均有一定程度的下降,表明在体素三角测量方法的基础之上引入全局注意力和成对注意力模块之后,可以有效地保证关节位置的合理性和肢体长度的一致性。此外,结合人体参数化模型对关键点进行重构有利于关键点估计性能的提升。

表 3 三角测量与关键点重构结果

Table 3 Triangulation and key point reconstruction results

GA	PA	HyBrIK	S9			S11		
			MPJPE/mm	*MPLLE*/mm	*MPLAE*/(°)	*MPJPE*/mm	*MPLLE*/mm	*MPLAE*/(°)
	√		52.00	14.58	0.151 7	35.16	9.78	0.124 0
√			52.46	14.16	0.152 4	34.98	9.67	0.122 4
√	√		50.24	14.13	0.150 9	34.10	9.50	0.121 7
√	√	√	**49.13**	**14.05**	**0.133 2**	**33.80**	**9.00**	**0.116 8**

3.3 关键点增强训练与评估

本文使用大型专家标注数据集[21]对手臂模型和身体模型进行训练,该数据集涵盖 273 名参与者、包含超过 2 400 万帧具有光学标记位置标注的数据。为了增加姿态的多样性,以垂直于横断面为旋转轴,围绕根关节进行 6 次旋转。模型输入的三维位置采用相对于根关节的表示方法,并使用受试者的身高体重进行标准化。两个模型均采用 LSTM 模型结构,对于身体模型,为 4 层结构,每层包含 128 个单元。对于手臂模型为 5 层结构,每层包含 128 个单元。从受试者层面按照训练集 80%、验证集 20% 的比例划分数据集。在训练过程中,使用加权均方根(mean squared error, MSE)作为损失函数,采用 Adam 优化器,学习率设置为 6×10^{-5},批次大小设置为 64。每个模型均独立训练两次,最终依据在验证集上的性能择优选取模型。

利用提出 OpenCap 方法的文献[33]提供的 LabValidation 数据集评估关键点增强模型在不同动作下的性能表现,该数据集包含 11 名受试者进行深蹲、跳跃、步行等动作的视频数据,并且包含光学标记位置标注以及运动学数据标注。原始关键点经过关键点增强后得到解剖标记点,以解剖标记点的位置误差作为评价关键点增强模型的性能指标。不同动作下关键点增强实验结果如图 8 所示,在不同动作下,本文的关键点增强模型性能明显优于其他两种方法,并具有更高的稳定性。

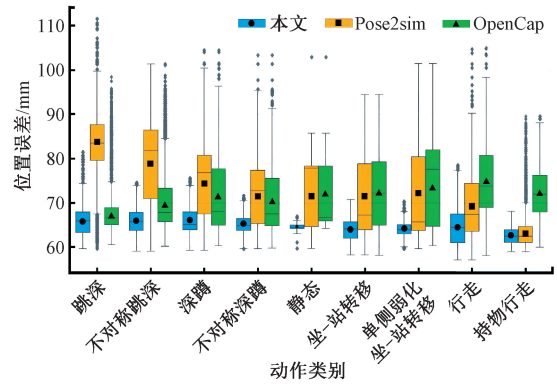


图 8 关键点增强得到的解剖标记位置误差

Fig. 8 The positional error of anatomical markers obtained through key point enhancement

同时,为评估关键点增强对运动学评估的影响,本文在关键点增强前后对运动学进行评估。关键点增强前后的运动学实验结果如图 9 所示。

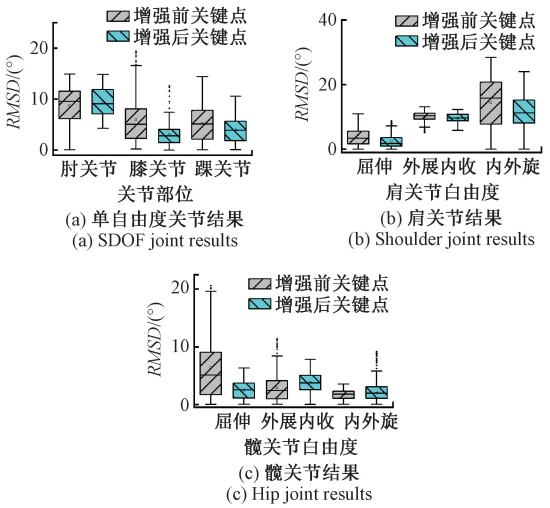


图 9　关键点增强前后的运动学误差对比

Fig. 9　Comparison of kinematic errors before and after key point enhancement

图 9 中英文缩写表示为单自由度(single degree of freedom, SDOF)。相较于使用增强前的原始关键点进行运动学分析。使用增强后的解剖标记点进行运动学分析可以获得更加准确、更加稳定的结果。

3.4　整体运动学分析实验

针对整个流程展开运动学分析实验,分别利用 LabValidation 数据集[33]、BMLmovi 数据集[38]以及 RRIS40 数据集[36]展开实验。3 个数据集中主要包含深蹲、跳跃、步行、坐姿、搬运、抬腿等动作。为评估运动学分析的精度,采用平均每个关节角度误差(mean per joint angle error, MPJAE)评估关节角度的准确性,采用平均每个关节位置误差 MPJPE 评估关节中心位置估计的准确性。

在 3 个数据集上,分别将本文的方法与 OpenCap 方法[33]、RRIS40 方法[39]以及 Pose2sim 方法[40]进行比较,整体运动学评估结果如表 4 所示(加粗表示最优结果),本文方法在上述 3 个数据集表现良好,在 RRIS40 数据集上本文方法的性能与基线方法大致相当,在其余两个数据集上均优于基线方法。

表 4　整体运动学评估结果

Table 4　Overall kinematic assessment results

方法	LabValidation 数据集		BMLmovi 数据集		RRIS40 数据集	
	MPJAE/(°)	*MPJPE*/mm	*MPJAE*/(°)	*MPJPE*/mm	*MPJAE*/(°)	*MPJPE*/mm
OpenCap	7.43	39.86	13.35	53.48	15.22	43.72
Pose2sim	9.70	45.53	11.57	55.72	13.61	42.18
RRIS40	8.35	45.53	12.42	44.37	**8.55**	42.52
本文	**7.36**	**40.37**	**9.64**	**43.54**	8.76	**42.15**

4　真实场景下的运动学分析

为验证方法在真实场景下的适用性,本文在真实场景下进行肩关节和步态运动学分析。商用动作捕捉系统 Xsens 方法是一种基于惯性测量单元(inertial measurement unit, IMU)的运动捕捉方法,具有高精度和高可靠性的优势,被广泛认可作为可靠的、便携式的评估标准。因此,本文以 Xsens 方法作为评估标准,并与同样采用两个视角的高性能方法 OpenCap 进行比较。3 个方法间通过肩关节或髋关节角度最高点进行同步比较。

步态运动学分析的实验场景设置如图 10(a)所示,肩关节运动学分析的实验场景设置如图 10(b)所示,Xsens 方法的 IMU 放置如图 10(c)所示。本文采用 Stereo Calibration 方法实现两个视角相机校准[41],并通过最大化二维关节中心点的垂直速度相关性确定最佳的时间偏移,实现两个相机的时间同步。

4.1　肩关节运动学分析

本文针对肩关节活动的 3 个自由度展开实验评估,具体包括肩关节的外展内收、前屈后伸、外旋内旋。实验

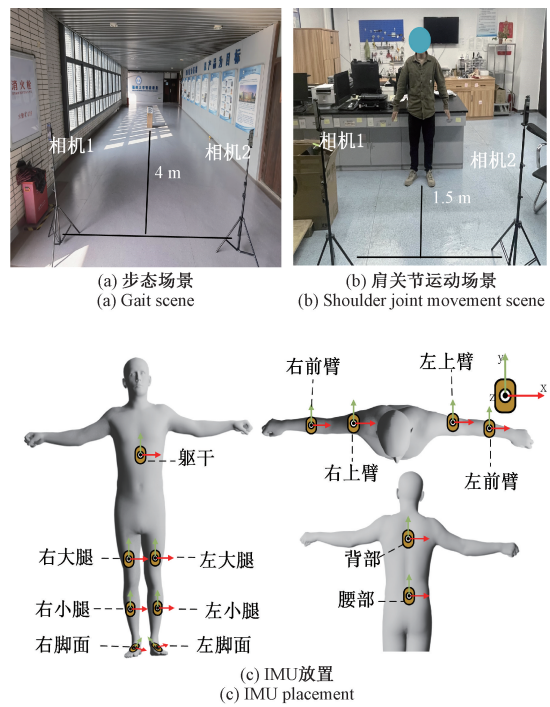


图 10　实验数据采集

Fig. 10　Experimental data acquisition

过程中,受试者在距离相机约 1.5 m 处进行相应动作,每一动作均重复 3 次。肩关节动作示意图如图 11 所示。

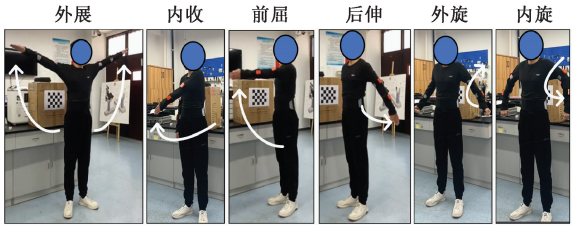


图 11 肩关节动作示意图

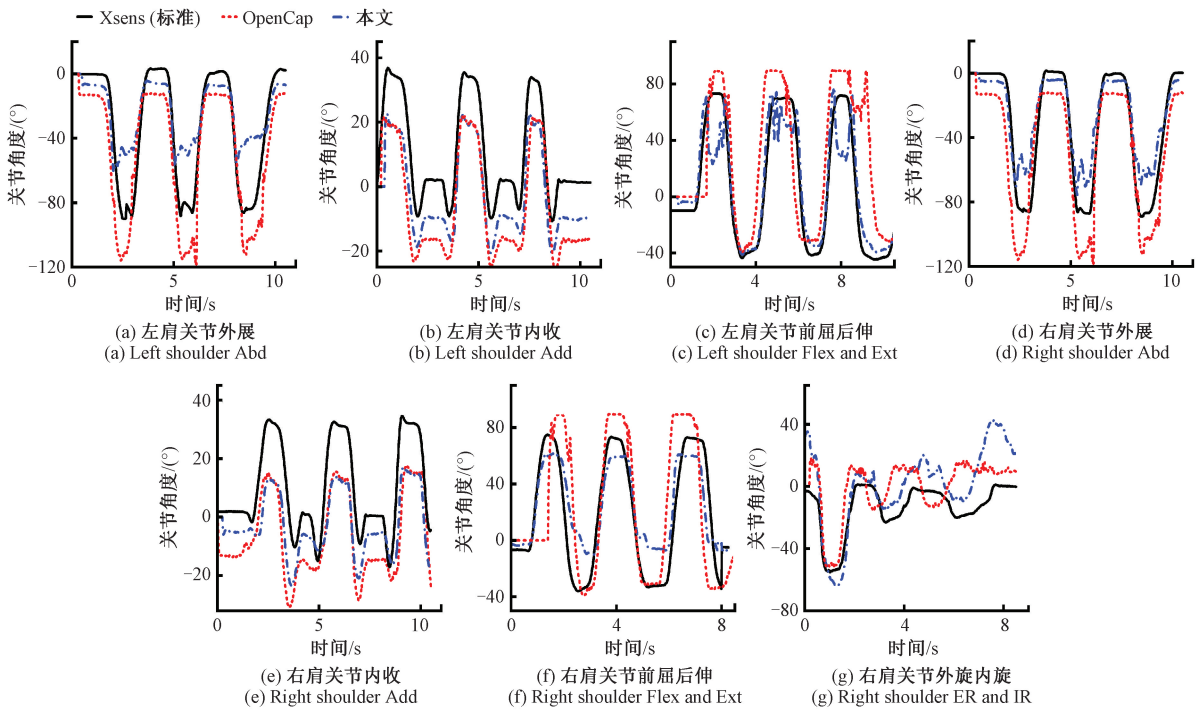
Fig. 11 Diagram of shoulder joint movement

肩关节运动学分析的定量性能比较结果如表 5 和图 12 所示。图 12 中英文缩写表示为外展(abduction, Abd)、内收(adduction, Add)、前屈(flexion, Flex)、后伸(extension, Ext)、外旋(external rotation, ER)、内旋(internal rotation, IR)。

表 5 肩关节运动学分析结果

Table 5 Results of shoulder joint kinematics analysis

自由度名称	OpenCap 方法		本文方法	
	MAE/(°)	*r*	*MAE*/(°)	*r*
右肩关节外展	16.84	0.94	10.89	0.97
左肩关节外展	17.81	0.95	15.73	0.92
右肩关节内收	25.46	0.72	16.83	0.90
左肩关节内收	26.19	0.78	13.35	0.92
右肩关节前屈后伸	16.26	0.89	12.56	0.90
左肩关节前屈后伸	15.82	0.94	12.28	0.98
右肩关节外旋内旋	15.65	0.65	16.16	0.83
平均	19.15	0.84	13.97	0.92



(a) 左肩关节外展 (a) Left shoulder Abd
(b) 左肩关节内收 (b) Left shoulder Add
(c) 左肩关节前屈后伸 (c) Left shoulder Flex and Ext
(d) 右肩关节外展 (d) Right shoulder Abd
(e) 右肩关节内收 (e) Right shoulder Add
(f) 右肩关节前屈后伸 (f) Right shoulder Flex and Ext
(g) 右肩关节外旋内旋 (g) Right shoulder ER and IR

图 12 肩关节运动学分析结果

Fig. 12 Results of shoulder joint kinematics analysis

本文方法与评估标准 Xsens 的平均皮尔逊相关系数为 0.92,高于 OpenCap 方法的 0.84。上肢运动学平均角度误差为 13.97°,低于 OpenCap 方法的 19.15°。实验结果表明,在肩关节运动学分析中,本文方法比 OpenCap 方法具有更高的准确性。特别是前屈后伸动作,虽然存在自遮挡问题,但本文方法仍能保持稳定的性能。

4.2 步态运动学分析

本文针对步态运动展开实验评估,评估步态过程中髋关节、膝关节、踝关节的运动角度,受试者在离相机约 4 m 的位置开始步行,约步行 2 个步态周期。

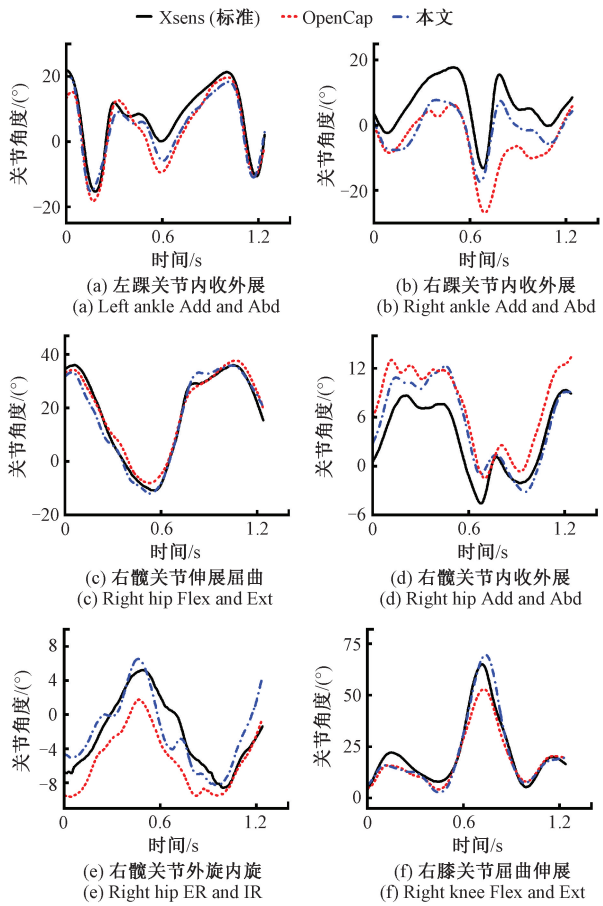
步态运动学分析的定量性能比较结果如表 6 和图 13 所示。图 13 中出现的英文缩写意义与图 12 中的相同。

表 6　步态运动学分析结果

Table 6　Results of gait kinematics analysis

自由度名称	OpenCap 方法		本文方法	
	MAE/(°)	*r*	*MAE*/(°)	*r*
右髋关节伸展屈曲	2. 04	0. 99	2. 14	0. 99
左髋关节伸展屈曲	3. 94	0. 96	1. 51	0. 99
右髋关节内收外展	2. 35	0. 91	3. 87	0. 95
左髋关节内收外展	2. 16	0. 91	2. 54	0. 88
右髋关节内旋外旋	2. 98	0. 91	1. 75	0. 87
左髋关节内旋外旋	4. 25	0. 00	3. 50	0. 04
右膝关节伸展屈曲	3. 69	0. 98	3. 04	0. 98
左膝关节伸展屈曲	5. 01	0. 94	2. 08	0. 99
右踝关节内收外展	11. 41	0. 74	7. 54	0. 92
左踝关节内收外展	3. 85	0. 96	3. 24	0. 97
平均	4. 17	0. 83	3. 12	0. 86

本文方法与评估标准 Xsens 的平均皮尔逊相关系数为 0. 86,高于 OpenCap 方法的 0. 83。下肢运动学角度误差为 3. 12°,低于 OpenCap 的 4. 17°。实验结果表明,在步态分析中,本文方法比 OpenCap 方法具有更高的准确性。然而,由于步行过程中盆骨抖动,造成髋关节外旋内旋角度相对误差较大,与评估标准 Xsens 的相关性较低。



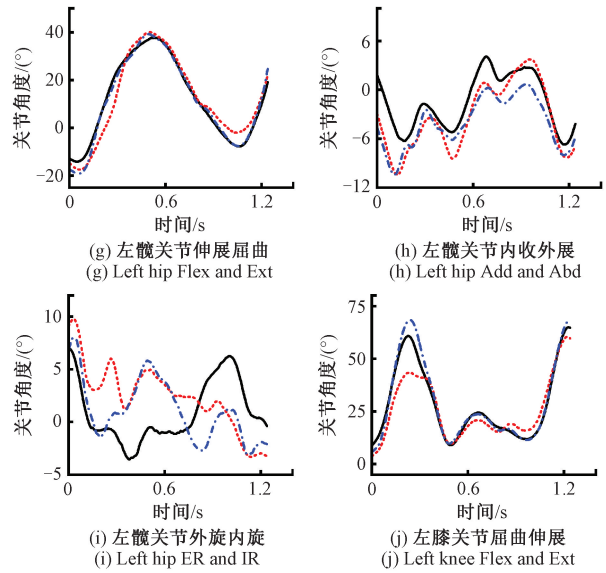


图 13　步态运动学分析结果

Fig. 13　Results of gait kinematics analysis

5　结　　论

本文提出一种采用两个视角的视觉无标记运动学分析方法。在 3 个常用公开数据集的整体运动学评估实验中,本文方法优于现有的高性能方法。在真实场景的肩关节和步态运动学评估中,与同样采用两个视角的主流方法 OpenCap 相比,本文方法更加准确,并且与评估标准 Xsens 的相关性更高。总体而言,本文方法在真实场景下具有适用性,相较于基于标记的光学系统,本文方法仅需两个相机,并且不需要复杂的反光标记放置和数据处理与分析过程,可以实现低成本、便捷的运动学分析,有利于推动运动学分析在临床上的应用。但是,由于模型大小和肌骨模拟的影响,本文方法尚无法实现实时性分析,在要求实时反馈的应用场景下受到限制。

参考文献

[1] 张堃,张鹏程,陈孝豪,等. 基于三维姿态估计的智能康复运动检测系统应用研究[J]. 仪器仪表学报, 2025,46(6):181-193.
ZHANG K, ZHANG P CH, CHEN X H, et al. Rehabilitation exercise detection method based on 3D human pose estimation[J]. Chinese Journal of Scientific Instrument, 2025, 46(6): 181-193.

[2] CERIOLA L, TABORRI J, DONATI M, et al. Comparative analysis of markerless motion capture systems for measuring human kinematics [J]. IEEE Sensors Journal, 2024, 24(17): 28135-28144.

[3] GU CH Y, LIN W C, HE X Y, et al. IMU-based motion capture system for rehabilitation applications: A systematic review [J]. Biomimetic Intelligence and Robotics, 2023, 3(2): 100097.

[4] 杨傲雷,任海燕,费敏锐,等. 基于多元回归的人体动态视觉定位方法[J]. 仪器仪表学报,2020,41(7): 252-260.
YANG AO L, REN H Y, FEI M R, et al. Dynamic body vision localization approach based on multiple regression[J]. Chinese Journal of Scientific Instrument, 2020, 41(7): 252-260.

[5] DE BORBA E F, STORNIOLO J L, CERFOGLIO S, et al. Effect of walking speed on the reliability of a smartphone-based markerless gait analysis system [J]. Sensors, 2025, 25(20): 6474.

[6] KOLEINI F, SALEEM M U, WANG P, et al. BioPose: Biomechanically accurate 3D pose estimation from monocular videos [C]. 2025 IEEE/CVF Winter Conference on Applications of Computer Vision, 2025: 6330-6339.

[7] XIA Y, ZHOU X W, VOUGA E, et al. Reconstructing humans with a biomechanically accurate skeleton [C]. 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025: 5355-5365.

[8] TAN T, WANG D X, SHULL P B, et al. IMU and smartphone camera fusion for knee adduction and knee flexion moment estimation during walking [J]. IEEE Transactions on Industrial Informatics, 2022, 19(2): 1445-1455.

[9] GÓRRIZ J M, ÁLVAREZ ILLÁN I, ÁLVAREZ MARQUINA A, et al. Computational approaches to explainable artificial intelligence: Advances in theory, applications and trends[J]. Information Fusion, 2023, 100: 101945.

[10] NIU Z H, LU K, XUE J, et al. From method to application: A review of deep 3D human motion capture[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 11340-11359.

[11] WADE L, NEEDHAM L, MCGUIGAN P, et al. Applications and limitations of current markerless motion capture methods for clinical gait biomechanics [J]. PeerJ, 2022, 10: e12995.

[12] DUBEY S, DIXIT M. A comprehensive survey on human pose estimation approaches [J]. Multimedia Systems, 2023, 29(1): 167-195.

[13] XIAO B, WU H P, WEI Y CH. Simple baselines for human pose estimation and tracking[C]. 15th European Conference on Computer Vision, 2018: 472-487.

[14] XU T H, TAKANO W. Graph stacked hourglass networks for 3D human pose estimation[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 16100-16109.

[15] LIANG SH, SUN X, WEI Y CH. Compositional human pose regression [J]. Computer Vision and Image Understanding, 2018, 176/177: 1-8.

[16] XU Y F, ZHANG J, ZHANG Q M, et al. ViTpose++: Vision Transformer for generic body pose estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(2): 1212-1230.

[17] WASHABAUGH E P, SHANMUGAM T A, RANGANATHAN R, et al. Comparing the accuracy of open-source pose estimation methods for measuring gait kinematics[J]. Gait & Posture, 2022, 97: 188-195.

[18] 陈慧娴,吴一全,张耀. 基于深度学习的三维点云分析方法研究进展[J]. 仪器仪表学报, 2023, 44(11): 130-158.
CHEN H X, WU Y Q, ZHANG Y. Research progress of 3D point cloud analysis methods based on deep learning[J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 130-158.

[19] ISKAKOV K, BURKOV E, LEMPITSKY V, et al. Learnable triangulation of human pose[C]. 2019 IEEE/CVF International Conference on Computer Vision, 2019: 7717-7726.

[20] LOPER M, MAHMOOD N, ROMERO J, et al. SMPL: A skinned multi-person linear model [J]. ACM Transactions on Graphics, 2015, 34(6): 1-16.

[21] WERLING K, BIANCO N A, RAITOR M, et al. AddBiomechanics: Automating model scaling, inverse kinematics, and inverse dynamics from human motion data through sequential optimization [J]. PLoS One, 2023, 18(11): e0295152.

[22] RUESCAS-NICOLAU A V, MEDINA-RIPOLL E, DE ROSARIO H, et al. A deep learning model for markerless pose estimation based on keypoint augmentation: What factors influence errors in biomechanical applications? [J]. Sensors, 2024, 24(6): 1923.

[23] GOZLAN Y, FALISSE A, UHLRICH S, et al. OpenCapBench: A benchmark to bridge pose estimation and biomechanics [C]. 2025 IEEE/CVF Winter Conference on Applications of Computer Vision, 2025: 4056-4065.

[24] FALISSE A, UHLRICH S D, CHAUDHARI A S, et al. Marker data enhancement for markerless motion capture[J]. IEEE Transactions on Biomedical Engineering, 2025, 72(6): 2013-2022.

[25] 王新, 谷亚东. 基于双目立体视觉的盲人避障技术研究[J]. 电子测量技术, 2025, 48(7): 98-106.
WANG X, GU Y D. Research on obstacle avoidance technology for the blind based on binocular stereo vision[J]. Electronic Measurement Technology, 2025, 48(7): 98-106.

[26] 赖东阳, 邱志斌, 杨泽鼎, 等. 基于 YOLOv9-SOEP 与双目立体视觉的输电线路山火火焰高度测量[J]. 电子测量与仪器学报, 2025, 39(8): 258-268.
LAI D Y, QIU ZH B, YANG Z D, et al. Measurement of wildfire flame height for transmission lines based on YOLOv9-SOEP and binocular stereo vision[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(8): 258-268.

[27] MAJI D, NAGORI S, MATHEW M, et al. YOLO-Pose: Enhancing YOLO for multi person pose estimation using object Keypoint similarity loss [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2022: 2636-2645.

[28] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical vision Transformer using shifted windows[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021: 9992-10002.

[29] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13708-13717.

[30] CHEN J, MAI H SH, LUO L B, et al. Effective feature fusion network in BIFPN for small object detection[C]. 2021 IEEE International Conference on Image Processing, 2021: 699-703.

[31] MA X X, SU J J, WANG CH Y, et al. Context modeling in 3D human pose estimation: A unified perspective[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 6234-6243.

[32] LI J F, XU CH, CHEN ZH C, et al. HybriK: A hybrid analytical-neural inverse kinematics solution for 3D human pose and shape estimation[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 3382-3392.

[33] UHLRICH S D, FALISSE A, KIDZIŃSKI Ł, et al. OpenCap: Human movement dynamics from smartphone videos [J]. PLoS Computational Biology, 2023, 19(10): e1011462.

[34] XU L M, JIN SH, LIU W T, et al. ZoomNAS: Searching for whole-body human pose estimation in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(4): 5296-5313.

[35] FANG H SH, LI J F, TANG H Y, et al. AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(6): 7157-7173.

[36] ZHU Y, SAMET N, PICARD D. H3WB: Human3.6M 3D wholebody dataset and benchmark[C]. 2023 IEEE/CVF International Conference on Computer Vision, 2023: 20109-20120.

[37] IONESCU C, PAPAVA D, OLARU V, et al. Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7): 1325-1339.

[38] GHORBANI S, MAHDAVIANI K, THALER A, et al. MoVi: A large multi-purpose human motion and video dataset[J]. PLoS One, 2021, 16(6): e0253157.

[39] JATESIKTAT P, LIM G M, LIM W S, et al. Anatomical-marker-driven 3D markerless human motion capture [J]. IEEE Journal of Biomedical and Health Informatics, 2025, 29(9): 6186-6199.

[40] PAGNON D, DOMALAIN M, REVERET L. Pose2Sim: An end-to-end workflow for 3D markerless sports kinematics—Part 1: Robustness [J]. Sensors, 2021,

21(19): 6530.

[41] MA X, ZHU P CH, LI X, et al. A minimal set of parameters-based depth-dependent distortion model and its calibration method for stereo vision systems[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-11.

作者简介

黄高华,2023 年于集美大学获得学士学位,现为福州大学硕士研究生,主要研究方向为视觉无标记运动学分析。

E-mail: 2382752230@qq.com

Huang Gaohua received his B. Sc. degree from Jimei University in 2023. He is currently a master student at Fuzhou University. His main research interest includes visual markerless kinematic analysis.

李玉榕(通信作者),1994 年于福州大学获得学士学位,1997 年于浙江大学获得硕士学位,2001 年于浙江大学获得博士学位,现为福州大学电气工程与自动化学院教授,主要研究方向为智能评估与康复技术的研究与应用。

E-mail: liyurong@fzu.edu.cn

Li Yurong (Corresponding author) received her B. Sc. degree from Fuzhou University in 1994, and her M. Sc. and Ph. D. degrees both from Zhejiang University in 1997 and 2001, respectively. She is currently a professor with the College of Electrical Engineering and Automation at Fuzhou University. Her main research interests include the research and application of intelligent assessment and rehabilitation technology.