DOI: 10.19650/j.cnki.cjsi.J2513902

基于三维姿态估计的智能康复运动检测系统应用研究*

张 堃[1],张鹏程[1],陈孝豪[1],张 彬[2],华 亮[1]
(1.南通大学电气与自动化学院 南通 226019; 2.苏州市体育专业运动队管理中心 苏州 215000)

摘 要:在康复运动场景中,运动输入通常是视频序列,基于主流的 2D 人体姿态估计方法和深度相机进行的伪 3D 方案无法对视频中的骨骼点测距,影响最终评估效果。为了解决这个问题,提出一种针对视频的序列到序列 3D 帧聚焦姿态识别方法用于康复评估。其目的是从最原始的二维噪声场景中直接提取更全面、更详细的三维坐标信息,并基于这些信息进行运动序列分析。该方法采用四支路流式变换器,能够捕获长序列时间与空间之间的交互关系,同时分别对原始 2D 输入进行时序与空间处理。这四支路信息通过可学习比例参数进行整合,并通过一个额外模块,结合空间编码器和增强型时间解码器获得最终输出。所提方法在 Human 3.6M 数据集上的表现优于最先进方法,平均关节位置误差仅为 14.4 mm,三维姿态坐标误差最低,证明了所提主干架构能够有效处理更复杂的康复运动视频序列任务,同时在实际康复视频序列的对比实验也验证了本方法的有效性。此外,基于先进的人体姿态估计方法,研发了一种新颖的多维度智能康复运动评估分析系统,能够对人体各个关节 120 个动作进行运动指标估计,已进入临床验证阶段,并完成 2 000 余例病人测试,平均准确率 93.2%。

关键词: 序列到序列;FFPose 算法;四支路流式变换器;可学习比例参数;无接触式
中图分类号: TH701 **文献标识码**: A **国家标准学科分类代码**: 520.40

Rehabilitation exercise detection method based on 3D human pose estimation

Zhang Kun[1], Zhang Pengcheng[1], Chen Xiaohao[1], Zhang Bin[2], Hua Liang[1]
(*1. School of Electrical Engineering and Automation, Nantong University, Nantong 226019, China;*
2. Suzhou Sports Professional Sports Team Management Center, Suzhou 215000, China)

Abstract: In rehabilitation exercise scenarios, motion input is typically in the form of video sequences. However, pseudo-3D solutions based on mainstream 2D human pose estimation methods and depth cameras are incapable of accurately measuring distances between skeletal points within videos, thereby affecting the final assessment performance. To address this issue, this paper proposes a sequence-to-sequence 3D frame-focused pose recognition method tailored for rehabilitation evaluation. The goal is to directly extract more comprehensive and detailed 3D coordinate information from the original noisy 2D scenarios and conduct motion sequence analysis based on this data. The proposed method adopts a four-branch streaming transformer architecture that captures the spatiotemporal interactions across long sequences by independently modeling the temporal and spatial aspects of the raw 2D input. These four branches are integrated through learnable proportional parameters, and an additional module combining a spatial encoder with an enhanced temporal decoder is employed to generate the final output. Our method outperforms state-of-the-art approaches on the Human 3.6M dataset, achieving a mean per-joint position error (MPJPE) of only 14.4 mm, the lowest 3D pose coordinate error reported to date. This demonstrates that the proposed backbone architecture is effective in handling more complex rehabilitation motion video sequence tasks. Moreover, comparative experiments on real-world rehabilitation video sequences further validate the effectiveness of our approach. Based on this advanced human pose estimation method, we have developed a novel multi-dimensional intelligent rehabilitation exercise evaluation and analysis system, capable of estimating motion metrics for 120 joint actions. The system has entered the clinical validation phase and has been tested on over 2 000 patients, achieving an average accuracy of 93.2%.

Keywords: sequence-to-sequence; FFPose algorithm; quadruple-stream Transformer; learnable scaling parameters; contactless

收稿日期:2025-04-09 Received Date: 2025-04-09
*基金项目:江苏省研究生实践创新项目(KYCX25_3747)资助

0 引 言

我国是全球人口第二多的国家,康复需求人数过亿,而康复师数量不足国外的 1/10。随着老龄化加剧,传统康复体系面临巨大压力,康复师亟需人工智能和机器人技术来提高治疗效率[1]。

在康复运动场景中,很多方法依赖于将医疗硬件设备固定在患者身上,通过传感器获取运动指标[2]。然而,这种方法常会给患者带来额外负担,尤其是对于重症患者,佩戴硬件设备不适合,最终导致无法依赖先进设备进行治疗。此外,一些基于智能视觉的无接触方法为解决这一问题提供了思路,本团队的先前工作[3]提出了一种基于二维人体姿态上肢无接触测量解决方案,在配合深度相机的情况下,可以较好地实时处理在 X 和 Z 位面的康复动作。然而,二维姿态检测的应用场景有限。在更复杂的场景中,例如 Y 位面肢体运动,要求患者调整姿势以提高准确度;同时,由于深度相机的限制,相近的骨骼点可能会被误认为是同一深度,从而导致误检。

相比之下,三维人体姿态估计不受这些限制,且已有许多研究将其应用于运动检测。三维姿态估计能够在推理过程中有效摆脱维度限制,通过二维人体坐标回归三维坐标信息[4]。这种方法广泛应用于动作识别和三维网格重建任务[5]。近年来,许多工作集中在首先从视频帧中提取二维人体骨架信息,然后通过主干网络进行处理并实现升维。然而,仅凭单帧内的骨架信息来学习三维坐标是一个极具挑战的任务。

为了解决这一问题,很多团队探索了优化二阶段升维操作的最佳案。传统的 Transformer 模块通过加入空间一致性分析只能捕捉帧间的空间信息。因此,文献[6]在 Transformer 基础上提出了步进编码器(stride Transformer, STE),并与原始编码器(vanilla Transformer, VTE)配合使用。尽管这一结构取得了良好的结果,但 VTE 结构的仅对单帧进行增强估计。文献[7]提出的混合时空编码器(mixed spatio-temporal encoder,MIXSTE)方法,结合了空间模块作为编码器和时间模块作为解码器,从而同时捕捉每帧的人体骨骼点空间信息和时间信息,并在测试中表现出较传统方法更高的指标。文献[8]在 MIXSTE 的基础上提出了双支路 Transformer 的结构:第 1 条支路仍然采用空间模块作为编码器、时间模块作为解码器,第 2 条支路则采用时间模块作为编码器、空间模块作为解码器,最后将两条支路的信息进行融合,以弥补单支路 Transformer 无法捕获的重要信息。虽然这些方法为二维坐标升维至三维坐标提供了卓越贡献,但无论是单流结构还是双流结构,都只考虑了空间和时序的交互过程,而忽略了时序模块和空间模块自身作用对模型的影响。导致这些方法在实现最佳性能的同时,仍然缺乏一些能够最小化误差的操作。

基于以上观点,提出了一种不依赖深度相机,能够通过帧间上下文信息准确获取三维骨骼点坐标的帧聚焦姿态识别方法(frame focused pose, FFPOSE)并运用于康复运动评估。FFPOSE 在网络结构上提出了一种创新的四支路流式变换器架构(quadruple-stream Transformer, QSFormer),该架构的 2 条支路用于捕捉不同帧间的时空交互信息,另外 2 条支路则用于捕捉时空模块自身对模型的影响作用。为增强模型性能,引入 Extra Part 增强模块,采用空间编码器和时间解码器对前一部分输出进一步提取融合信息,以获得最佳输出;同时,在该模块中并行化了一个改进的门控注意力单元(improved gated attention unit, IGAU),以优化原始 Transformer 模块,从而将 Transformer 信息和 IGAU 对长序列建模的能力结合,进一步提升模型的效果。基于该方法,设计了一种能够最大程度帮助康复师在复杂医疗场景下进行视觉康复评估的康复运动检测系统。该系统的硬件部分由智能计算模块和视觉采集模块构成;软件部分包括了一个具备强大扩展性的人机交互界面,以及多个功能模块以实现不同运动监测。系统可以对人体所有关节和肢体 120 项不同运动进行角度、速度、加速度、位移等指标估计。

本研究主要贡献为:

1)提出了四支路流式变换器 QSFormer,该网络架构综合考虑了时空交互作用,以及时空模块对模型的影响。

2)主干网络后半部分引入了 Extra Part 增强模块,其中包含改进的 IGAU,与原始的 Transformer 协同工作,以增强信息提取能力。

3)主干网络有效模拟了长序列任务,表现出强大的适应性,并最大限度地减少了运动误差,在多个数据集上取得了最先进的性能。

4)将三维人体姿势检测应用于康复训练,实现了对康复运动视频序列的检测,并进行了速度、加速度、位移、角度等指标估计。基于提出的核心技术,建立了基于姿态捕捉的智能康复系统,已在相关医院进行临床验证,完成了 2 000 余例病人测试,平均检测准确度为 93.2%。

1 理论分析

1.1 方法概述

如前文所述的事实和实际应用案例[9],本方法基于 Transformer 架构,充分考虑长序列任务中时空交互的关键作用,本研究提出了一种创新的三维姿态估计主干网络,旨在同时捕获时空交互信息,并有效学习时间和空间

模块的特征。具体而言,如图 1(a)所示,本方法首先利用二维人体姿态检测器(SR-POSE)提取骨架信息,随后采用序列到序列方法将二维骨架数据映射至三维空间,并基于所得三维骨架数据实现康复指标检测。

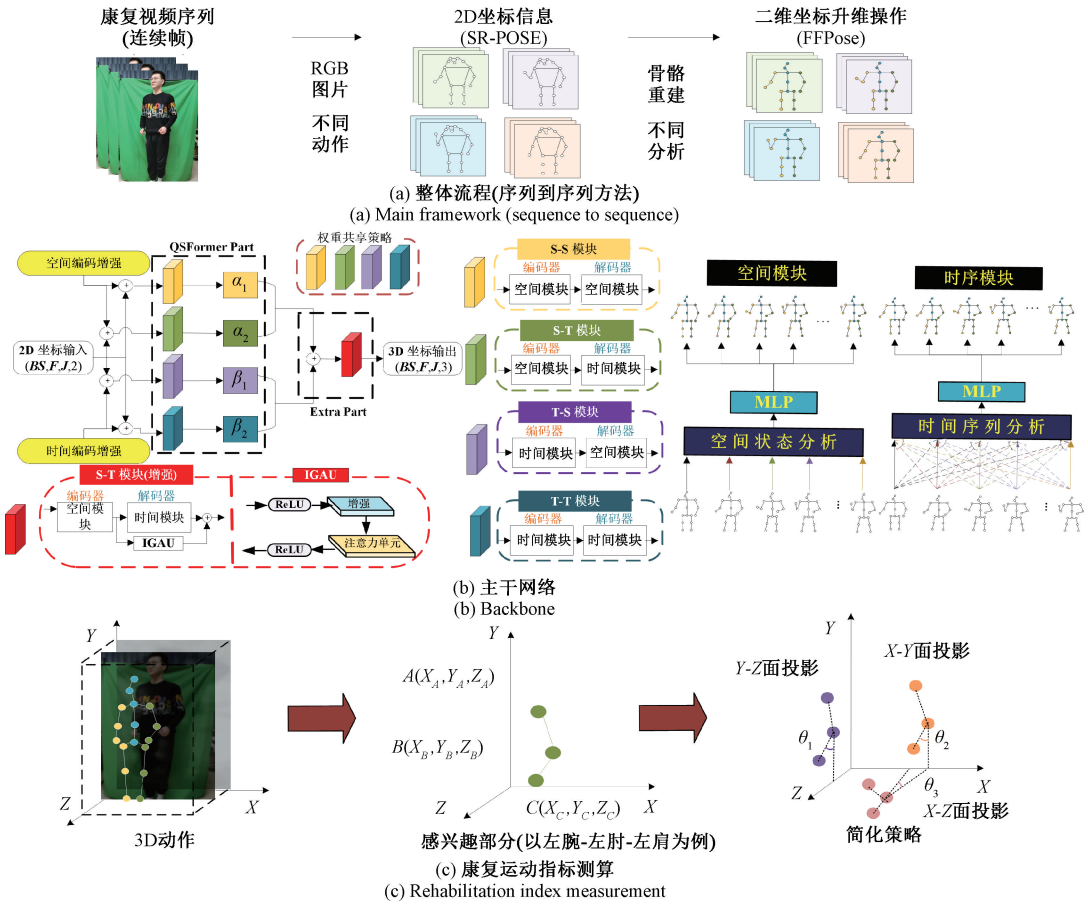


图 1　FFPose 姿态检测算法模型和总体结构

Fig. 1　FFPose attitude detection algorithm model and overall structure are explained in detail

1.2　主干网络

如图 1(b)所示,本方法的主干网络由 2 部分组成:四支路模块(QSFormer)和增强特征提取模块(Extra Part)。在处理视频序列时,每个序列的帧数表示为 F,批次大小表示为 BS,并由硬件计算能力决定。对于输入的康复运动视频序列,网络首先按帧大小对视频进行分组,并根据硬件条件确定适当的批次大小。人体骨骼关键点 J 的数量固定为 17 个。

模型的输入数据为二通道坐标,经过网络处理后,其通道维度被扩展为 512。在四支路结构中,不同编码器负责不同类型的特征增强:前 2 条空间编码器支路用于空间编码增强,后 2 条时间编码器支路用于时序编码增强。每条支路均由 4 个可学习比例参数进行动态调节,以优化信息表达。前半部分各支路的输出可表示如式(1)所示。

$$\begin{cases} Z^1_{out} = \alpha_1(S - S(pos_s(x))) \\ Z^2_{out} = \alpha_2(S - T(pos_s(x))) \\ Z^3_{out} = \beta_1(T - S(pos_t(x))) \\ Z^4_{out} = \beta_2(T - T(pos_t(x))) \end{cases} \tag{1}$$

其中,Z^1_{out}、Z^2_{out}、Z^3_{out}、Z^4_{out} 分别表示四支路结构(从上至下)的输出;α_1、α_2、β_1、β_2 为对应的可学习参数;S-S、S-T、T-S、T-T 表示 4 个模块在网络中的不同功能作用;$pos_s(x)$、$pos_t(x)$ 分别表示空间编码增强操作和时序编码增强操作。因此,总输出如式(2)所示。

$$Z^{half}_{out} = \frac{1}{2}\sum_{k=1}^{4} Z^i_{out} \tag{2}$$

其中,Z^{half}_{out} 为 QSFormer 的综合输出,该输出将作为 Extra Part 模块的输入,经过增强特征提取后,生成最终的三维人体姿态信息。Extra Part 由空间编码器和增强型时序解码器组成,因此,三维人体姿态最终输出可表示如式(3)所示。

$$Z_{out}^{final} = S' - T'(Z_{out}^{half}) \tag{3}$$

其中，Z_{out}^{final} 为最终的人体姿态输出，形如 $x_{out} \in \mathrm{R}^{BS\times F\times J\times 3}$；$S' - T'$ 为增强型空间与时序模块的作用。

1.3 空间模块

对于长视频序列，本研究采用空间模块对输入数据进行建模[10]。为提升信息提取能力，引入了注意力机制，以增强模型的学习能力。通常，注意力机制需要确定 3 个关键变量：查询（$\boldsymbol{query}_s$，$\boldsymbol{q}_s$）、键（$\boldsymbol{key}_s$，$\boldsymbol{k}_s$）、值（$\boldsymbol{value}_s$，$\boldsymbol{v}_s$）。在本研究中，重点分析了多头自注意力机制（multi-head self-attention，MHSA）的影响。对于经过位置编码的 512 通道向量，考虑到时空模块不涉及时序建模，其 3 个关键变量可表示如式（4）所示。

$$\boldsymbol{q}_s = \boldsymbol{k}_s = \boldsymbol{v}_s = \boldsymbol{x}_s \in \mathrm{R}^{BS\cdot F\times J\times 512} \tag{4}$$

其中，BS 为批次大小，F 为截取的视频帧数，$BS \cdot F$ 表示将批次大小和截取帧数的维度合并，J 代表骨骼关键点的数量。

随后，这 3 个变量将输入多头注意力机制。当注意力头数为 n 时，每个注意力头继承原变量的一部分通道数，以实现分批建模，从而提取输入骨骼关键点的空间特征。该过程可表示如式（5）所示。

$$\begin{cases} out_{sn} = softmax\left(\dfrac{\boldsymbol{q}_{sn}\boldsymbol{k}_{sn}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{v}_{sn} \\ d = \dfrac{512}{n} \end{cases} \tag{5}$$

其中，d 为单头注意力机制输入量的维度，其值等于总通道数与注意力头数的比值；out_{sn} 表示第 n 个注意力头的输出；$\boldsymbol{q}_{sn}$、$\boldsymbol{k}_{sn}^{\mathrm{T}}$、$\boldsymbol{v}_{sn}$ 分别表示第 n 个头的查询、键的转置和值；$Softmax$ 为激活操作，用于将注意力分数转换为概率分布，从而衡量不同关键点的重要性。

因此，空间多头注意力机制可表示如式（6）所示。

$$\boldsymbol{SA} = [out_{s0}\cdots out_{sn}]\boldsymbol{W}_s + \boldsymbol{x}_s = \left[softmax\left(\frac{\boldsymbol{q}_{s0}\boldsymbol{k}_{s0}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{v}_{s0}\cdots softmax\left(\frac{\boldsymbol{q}_{sn}\boldsymbol{k}_{sn}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{v}_{sn}\right]\boldsymbol{W}_s + \boldsymbol{x}_s \tag{6}$$

其中，$\boldsymbol{SA}$ 表示空间注意力机制的总输出，$\boldsymbol{W}_s$ 为参数投影矩阵，$\boldsymbol{x}_s$ 为原始输出。

为了获得空间多头注意力机制的最终输出，需对其进行层归一化操作 $Layernorm$，并通过多层感知机 MLP 进一步处理。该过程可表示如式（7）所示。

$$\boldsymbol{SA}_f = \boldsymbol{SA} + MLP(Layernorm(\boldsymbol{SA})) \tag{7}$$

其中，$\boldsymbol{SA}_f$ 为整个空间模块的输出。为防止主路径学习到错误信息，引入残差连接，以辅助矫正并增强多头注意力机制的输出稳定性。

1.4 时间模块

在时序模块计算过程中，对于编码增强后的输入，不再执行类似于空间模块的维度压缩操作，而是保持其原始形态，直接输入到多头注意力机制。此时，查询、键、值可以表示如式（8）所示。

$$\boldsymbol{q}_t = \boldsymbol{k}_t = \boldsymbol{v}_t = \boldsymbol{x} \in \mathrm{R}^{BS\times F\times J\times 512} \tag{8}$$

其中，BS 为批次大小，F 为截取的视频帧数，J 为骨骼关键点的个数。在时序建模过程中，时间维度被单独作为一个独立维度，模型在执行多头注意力计算时，能够隐式学习视频中的时序信息。随后，3 个输入将按照空间模块中类似的多头注意力机制进行处理，最终生成时序模块的输出。

1.5 Extra Part 与改进型门控注意力单元

IGAU 已被证明在长序列建模中具有优越的性能，在本研究的增强模块可有效发挥作用，同时避免引入大量额外参数，该方法已广泛应用于基于卷积神经网络（convolutional neural networks，CNN）的视觉任务[11]。

受此启发，采用空间模块作为编码器，时间模块作为解码器，并在解码器两端并联改进型 IGAU，构成增强型解码器，用于对来自 QSFormer 的输出特征进行增强与再提取。

在 IGAU 内，设定第 2 维的关键点个数为 17，以确保特征提取的针对性。为降低计算复杂度，对输入编码层的 512 通道向量进行全连接操作 $Linear$，将通道数降至 256，可表示如式（9）所示。

$$\boldsymbol{x}_{gin} = relu(Linear(\boldsymbol{x})) \tag{9}$$

随后，对输入向量执行激活操作，并通过全连接操作进一步扩展通道数。其中，选取 128 个通道的数据作为查询和键的基础向量，512 个通道的数据作为值向量。然后，对基础向量进行增强，引入可学习参数 ∂ 和 b，得到增强后向量 $\boldsymbol{x}_{emb}$ 如式（10）所示。

$$\boldsymbol{x}_{emb} = \boldsymbol{x}_{gin} \times \partial + b \tag{10}$$

将查询和键设置为增强向量，则注意力机制可表示如式（11）所示。

$$\boldsymbol{x}_{gout} = Dropout\left(relu\left(\frac{\boldsymbol{q}_g\boldsymbol{k}_g^{\mathrm{T}}}{\sqrt{128}}\right)\right)\boldsymbol{v}_g \tag{11}$$

其中，$\boldsymbol{x}_{gout}$ 为 IGAU 的输出向量，$Dropout$ 为随机丢弃操作，丢弃率为 0.5，$relu$ 为激活操作，$\boldsymbol{q}_g$ 为 IGAU 的查询，$\boldsymbol{k}_g^{\mathrm{T}}$ 为 IGAU 的键的转置，$\boldsymbol{v}_g$ 为值向量。

因此 Extra Part 解码器部分的最终输出 x_{final} 可表示如式（12）所示。

$$\boldsymbol{x}_{final} = (\boldsymbol{x}_{decoder} + \boldsymbol{x}_{gout}) \times 0.5 \tag{12}$$

其中，$\boldsymbol{x}_{decoder}$ 为空间解码器的基础输出。对两路输出取平均值，以综合考虑模块间的相互作用。最后，经过 Extra Part 模块后的向量需通过最终全连接层，将通道数降至 3，得到回归后的三维坐标数据。

1.6 损失函数

在预训练阶段，采用平均关节位置误差作为损失函

数之一,如式(13)所示。

$$l_{mpjpe} = \frac{1}{N}\sum_{n=0}^{N} \| x_{3d}^{n} - x_{r3d}^{n} \|_2 \tag{13}$$

其中,x_{3d}^{n} 为第 n 个关键点的三维预测值,x_{r3d}^{n} 为对应的三维真值,N 为关键点的总数。

为评估预测值和真值之间的一致性,对序列数据进行差分处理,即从第 2 个时间步开始,逐一减去前一时间步的值,以计算预测值与真值的时序差异,并定义一致性损失 l_{mpjve},如式(14)所示。

$$l_{mpjve} = \frac{1}{N-1}\sum_{n=0}^{N-1} \| x_{3d}^{n} - \tilde{x}_{3d}^{n} \|_1 \tag{14}$$

随后,将三维坐标重投影到二维坐标系,计算二维投影误差损失 l_{2d},如式(15)所示。

$$l_{2d} = \frac{1}{N}\sum_{n=0}^{N} \| \theta_{2d}^{n} - \theta_{r2d}^{n} \|_2 \tag{15}$$

其中,θ_{2d}^{n} 为第 n 个关键点的二维投影预测值,θ_{r2d}^{n} 为其二维真值。

因此,训练阶段的总损失函数 l_{pre} 如式(16)所示。

$$l_{pre} = \gamma_1(l_{mpjpe} + l_{mpjve}) + (1-\gamma_1)l_{2d} \tag{16}$$

其中,γ_1 为三维影响系数,本文设置为 0.8。

1.7　康复指标估计算法

受二维康复运动检测的启发,本文所提出的康复评估算法同样基于美国国家体能协会(National Strength and Conditioning Association,NSCA)指南进行测试和评估[12]。该指南明确规定了肩外展等康复运动的关键关节角度,包括内收、伸展或屈曲、肘关节伸展、屈曲、髋关节伸展、屈曲和外展等动作的理论参考角度。

如图 1(c)所示,单目相机采集的康复视频流首先经过二维姿态检测提取人体骨架信息,然后通过三维姿态估计算法将二维坐标升维至三维坐标,并转换至世界坐标系。以左侧肩点 $A(X_A,Y_A,Z_A)$、左侧肘关节 $B(X_B,Y_B,Z_B)$ 和左侧腕关节 $C(X_C,Y_C,Z_C)$ 为例,探讨肘关节的旋转角度。在空间坐标系中,肢体的旋转角度通常具有多维特性,即使某一维度被刻意固定,该维度的旋转角度仍可能出现细微变化。

不同于点透视问题(perspective-n-point, PnP)等综合处理三维运动的方法,提出了一种简化策略,即将肢体在空间中相对于特定坐标平面 $(X-Y,Y-Z,X-Z)$ 进行投影,以获得投影坐标。以 $X-Y$ 平面的投影为例,其过程如式(17)所示。

$$\begin{bmatrix} X_p \\ Y_p \\ Z_p \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{17}$$

其中,(X_p,Y_p,Z_p) 表示 $X-Y$ 位面投影后的坐标,在该投影过程中,Z_p 轴坐标将被设置为 0。因此,左侧肘关节在 $X-Y$ 平面的投影坐标为 $B(X_B,Y_B,0)$,左侧腕关节的投影坐标为 $C(X_C,Y_C,0)$。由此,三维角度问题可简化为二维角度求解问题。进一步地,将左侧肘关节的坐标再次投影至 x 轴上,坐标为 $C(X_C,0,0)$,最终肢体在此位面上的运动角度如式(18)所示。

$$\theta_2 = \arctan\left(\frac{Y_B - Y_C}{X_B - X_C}\right) - \arctan\left(\frac{0 - Y_C}{X_C - X_C}\right) = \arctan\left(\frac{Y_B - Y_C}{X_B - X_C}\right) + \frac{\pi}{2} \tag{18}$$

其中,θ_2 代表在该位面 $X-Y$ 投影后的旋转角度。同理,其他位面的角度 θ_1 与 θ_3 也可以通过相同方法进行计算。但在计算过程中,需根据反正切函数的性质及实际需求对符号转换进行相应调整。

运动速度[13]是康复训练中关键评估指标,用于衡量患者对康复训练的适应性。基于上述角度计算,提出了一种多帧融合的康复运动速度计算方法。由于视频通常包含高帧率信息,短时间内的帧间变化肉眼难以察觉,但这些细微变化在运动分析中同样重要,因此需要对其进行量化计算。

对于一个包含 30 帧的视频序列 $t \in [1,2,\cdots,T]$,本文将每秒帧数划分为两部分,即每 15 帧作为一个子区间,并提取其角度信息进行曲线拟合,拟合曲线 $f(x)$ 如式(19)所示。

$$f(x) = \sum_{i=1}^{n} a_i \times p_i(x) \tag{19}$$

其中,n 为系数的维度,a_i 为第 i 个系数的值,$p_i(x)$ 代表第 i 个多项式基函数。如式(20)所示,函数 $p_i(x)$ 可进一步表示为:

$$p_i(x) = x^i \tag{20}$$

对该函数求一阶导数,得到 $p_i'(x)$ 并代入式(19)可得最终速度表达式 $f'(x)$,如式(21)所示。

$$f'(x) = \sum_{i=1}^{n} a_i \times p_i'(x) = \sum_{i=1}^{n} a_i x^{i-1} i \tag{21}$$

因此,15 帧窗口的平均速度如式(22)所示。

$$v_{mean} = \frac{\sum_{j=0}^{15} f'(j)}{15} \tag{22}$$

以此类推,最终的康复运动速度可通过多个小区间的速度计算得到,从而全面反映患者在康复训练过程中的运动表现。

2　实验验证

2.1　实验平台的建立

本研究由两张 Nvidia RTX 3090Ti 显卡构建的深度学习实验平台进行模型训练与性能评估。在数据集选择方面,采用经典的三维人体姿态估计数据集 Human

3.6 M[14]，该数据集包含 360 万张图片及其对应的姿态标注信息，主要用于室内环境下的人体姿态估计。数据涵盖了 11 名演员的多个动作样本，其中仅有 7 位演员提供了三维标注数据[15]。根据前人研究经验[16]，选择演员 S1、S5、S6、S7、S8 作为训练集，S9 和 S11 作为测试集。

2.2 训练参数设置

在训练参数设置方面，采用 Adam 优化器，初始学习率设定为 0.001。对于常规模块，最大通道数设置为 512；而在 Extra Part 增强模块中的 IGAU 部分，通道数初始设置为 256，并在内部处理后最终升至 512 通道，以增强特征表达。在 Transformer 模块中，使用多头注意力机制，注意力头设定为设置为 8，整体网络架构重复 4 次。在训练过程中，时序模块所截取的帧长度设为变量 $\bar{T}$，参考前人研究经验[17]，分别选取了 $\bar{T}=27$ 与 $\bar{T}=243$ 作为实验变量。实验结果表明，当时序效果在 $\bar{T}=243$ 时，时序建模效果最优，且随着 $\bar{T}$ 的减少，预测误差呈非线性增大。

此外，Extra Part 增强模块基于前文提出的空间-时间模块进行设计。

实验表明，该配置可实现最佳性能匹配。最终，将模型训练设定为 120 轮，如图 2 所示，在 40 轮后，模型逐步收敛，误差水平呈现一定波动。

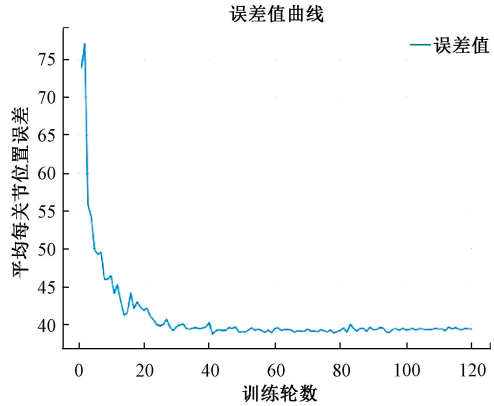


图 2 训练曲线

Fig. 2 Training curve

2.3 三维人体姿态估计表现

如表 1 所示，在 Human 3.6M 数据集的测试集中对所提出的方法进行了评估，并以毫米为单位记录误差指标。主要采用的衡量标准为平均每关节位置误差（mean per joint position error，MPJPE），该指标用于量化预测的三维坐标与真实值之间的欧式距离，数值越小表明预测精度越高。为全面评估本文方法的优势，本文选取了多个研究进行对比。表格采用了 3 种误差计算标准：

表 1 三维人体姿态检测方法和一些目前最先进技术在 Human 3.6M 数据集上的指标比较

Table 1 Comparison of 3D human pose detection method and some SOTA technologies on the Human 3.6M dataset

（mm）

对比方案	文献	$\bar{T}$	Dir	Disc	Eat	Greet	Phone	Photo	Pose	Pur	Sit	SitD	Smoke	Wait	WalkD.	Walk	WalkT	平均
Protocol#1 在 MPJPE 指标下本文方法与其他先进方法的表现	[16]	1	51.8	56.2	58.1	59.0	69.5	78.4	55.2	58.1	74.0	94.6	62.3	59.1	65.1	49.5	52.4	62.9
	[17]	1	45.2	49.9	47.5	50.9	54.9	66.1	48.5	46.3	59.7	71.5	51.4	48.6	53.9	39.9	44.1	51.9
	[18]	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	50.2
	[19]	7	44.6	47.4	45.6	48.8	50.8	59.0	47.2	43.9	57.9	61.9	49.7	46.6	51.3	37.1	39.4	48.8
	[20]	243	45.2	46.7	43.3	45.6	48.1	55.1	44.6	44.3	57.3	65.8	47.1	44.0	49.0	32.8	33.9	46.8
	[21]	243	44.8	46.1	43.3	46.4	49.0	55.2	44.6	44.0	58.3	62.7	47.1	43.9	48.6	32.7	33.3	46.7
	[22]	243	41.8	44.8	41.1	44.9	47.4	54.1	43.4	42.2	56.2	63.6	45.3	43.5	45.3	31.3	32.2	45.1
	[5]	81	41.5	44.8	39.8	42.5	46.5	51.6	42.1	42.0	53.3	60.7	45.5	43.3	46.1	31.8	32.2	47.3
	[23]	200	38.5	42.5	39.9	41.7	46.5	51.6	39.9	40.8	49.5	56.8	45.3	46.4	46.8	37.8	40.4	44.3
	[9]	351	39.2	43.1	40.1	40.9	44.9	51.2	40.6	41.3	53.5	60.3	43.7	41.1	43.8	29.8	30.6	43.0
	[6]	243	40.6	43.3	40.2	42.3	45.6	52.3	41.8	40.5	55.9	60.6	44.2	43.0	44.2	30.1	30.1	43.7
	本文	**243**	**35.6**	**37.4**	**37.3**	**32.3**	**41.1**	49.3	**36.2**	**34.1**	50.5	**53.8**	41.2	**37.2**	**37.0**	25.9	**26.3**	**38.3**
Protocol#2 在使用二维坐标真值情况下的 MPJPE 指标对比	[8]	243	21.6	22.0	20.4	21.0	20.8	24.3	24.7	21.9	26.9	24.9	21.2	21.5	20.8	14.7	15.6	21.6
	[7]	243	16.7	19.9	17.1	16.5	17.4	18.8	19.3	20.5	24.0	22.1	18.6	16.8	16.7	10.8	11.5	17.8
	本文	**243**	**13.9**	**15.0**	**14.3**	**12.9**	**14.5**	15.4	**15.6**	**15.5**	21.7	**18.4**	**15.2**	**13.3**	**8.36**	**12.7**	**9.36**	**14.4**

Protocol#1 和 Protocol#2 为平均每关节位置误差(MPJPE)。表 1 中,下划线表示了列表中的第 2 好的工作结果,加粗表示了列表中的最好工作结果。根据表 1(Protocol#1)的结果,本方法在完整三维坐标的预测上展现出明显优势。此外,与当前的最先进方法的对比表明,本方法在三维运动信息预测方面具有更优表现。在表 1(Protocol#2)中,为了评估单一距离的误差,实验采用真实值替代预测坐标中的前两维,仅对第三维(深度信息)进行误差分析。结果表明,本方法在单一距离估计方面具有较高的精度。综合表 1 的结果可见,本方法在三维姿态估计任务中表现优异,并进一步验证了基于姿态估计的康复运动系统的可行性。

2.4　消融研究实验

重点分析模型的四支路结构、Extra Part 模块以及用于增强 Extra Part 输出的 IGAU 模块的独立作用效果。这些分析有助于深入理解各组件在整体网络中的贡献及其对模型性能的影响。

如表 2 所示,展示了本模型各组件在 Human3.6M 测试数据集的作用效果。可以观察到,QSFormer 结构已达到较优性能,其在测试集上的 MPJPE 指标为 39.0 mm;进一步加入 Extra Part 模块和 IGAU 模块后,模型性能分别提升了 0.4 和 0.3 mm。从消融研究的角度来看,本文所提出的各模块均能有效优化三维人体姿态估计,验证了其对整体模型性能的积极贡献。

表 2　消融研究

Table 2　Ablation study

QSFormer	Extra Part	IGAU	MPJPE/mm
√			39.0
√	√		38.6
√	√	√	38.3

如表 3 所示,展示了当 Extra Part 选择不同编码器和解码器时的实验结果。

表 3　Extra Part 的选择

Table 3　Choice of Extra Part

S-T	T-S	S-S	T-T	MPJPE/mm
				39.0
√				38.3
	√			38.7
		√		38.9
			√	38.5

第 1 行代表不采用该模块的结果。实验结果表明,Extra Part 选择空间编码器和时间解码器(对应表 3 中 S-T)作为主干结构时,能够最小化误差;同时相较于不采用该模块,能够降低 1.7% 的误差,从而进一步优化模型的预测精度。

2.5　康复运动指标估计研究

本方法在误差指标和鲁棒性方面均展现出优异的性能。在本节中,将进一步探讨该方法在康复运动视频序列中的实际应用表现。根据前述流程,视频帧提取由 SR-Pose 二维检测器完成,以提取二维骨架信息,随后,该信息被输入至三维检测器进行升维处理,最终计算相关运动指标。

基于南通市三家医院康复科的数据,本方法已在 2 000 余例神经、脑和骨损伤患者中进行临床测试。为验证系统的有效性,本研究随机选取了 100 例患者,评估其身高、运动速度等生物运动特征信息。本研究已获得相关医院伦理审查委员会批准,并在实验前告知并征得患者同意。

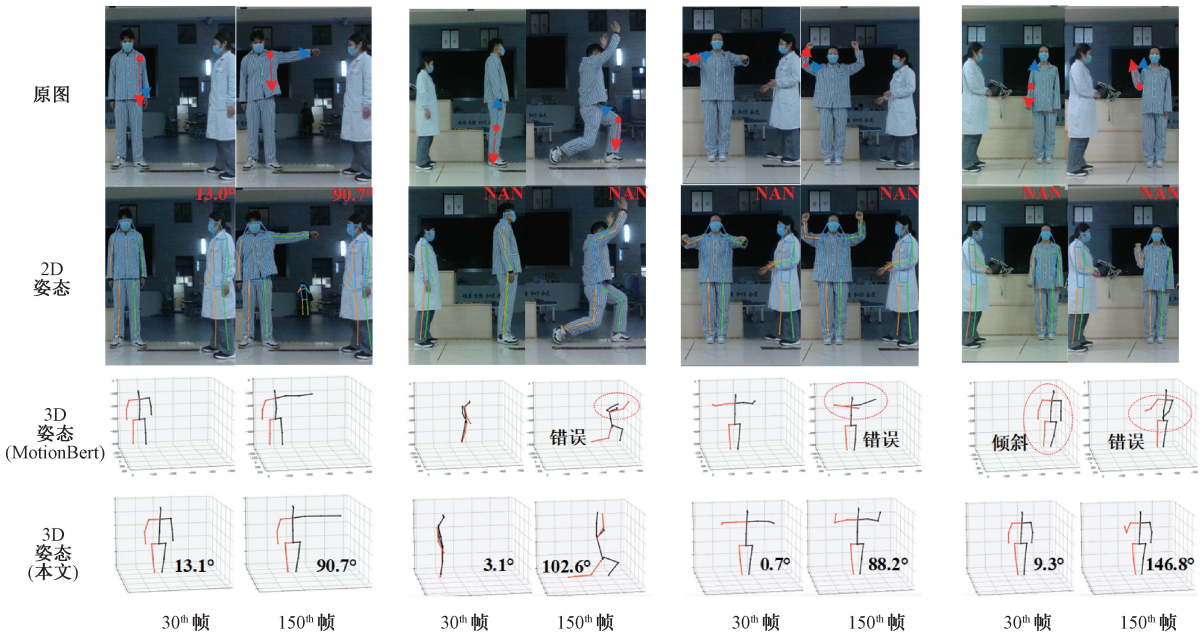
如图 3 所示,实验过程中,患者分别执行左肩外展、膝盖屈曲、小臂外展和右肘屈曲等康复动作。针对这些动作,分别采用二维和三维方法进行对比分析。值得注意的是,在左肩外展这一平面动作中,两种方法均取得了良好效果,因为该动作不依赖骨骼点的深度信息。然而,当动作模式变得更加复杂,并超出单一平面时,二维方法因缺乏深度信息,无法为关键点提供有效的深度估计,最终导致所有运动指标变为非数字(not a number,NaN)值。

相比之下,基于三维人体姿态估计的方法有效克服了上述限制,能够顺利计算运动角度。与当前先进的三维姿态估计方法进行比较后,尽管本文方法在误差指标上未出现显著下降,但在实际视频序列建模效果上,其表现优于当前最先进的方法(如图 3(b)~(d)所示)。综上所述,本文方法在实际应用中能够有效实现预期的康复评估目标,为康复训练提供更精准、稳定的姿态分析能力。

此外,还对本方法的结果与物理传感器测量结果进行了角度对比实验。为此,选取了惯性测量传感器(inertial measurement unit, IMU)[24]作为物理传感器进行验证。IMU 作为目前广泛使用的接触式检测方法之一[25],其测量角度的均方根误差通常低于 2°,具有较高测量精度[26]。

本研究对患者进行了 5 次测试,每次测试中,除了计算所有患者的平均角度指标外,还通过计算本方法获取的角度区间值与物理传感器测量的角度区间值的比值,得出了动作完成度,如式(23)所示。

$$c_{com}=\frac{\| x_{omax}-x_{omin} \|_1}{\| x_{smax}-x_{smin} \|_1} \tag{23}$$



(a) 左肩外展
(a) Total abduction

(b) 膝盖屈曲(视频)
(b) Knee flexion (video)

(c) 小臂外展(视频)
(c) Elbow rotation (video)

(d) 右肘屈曲(视频)
(d) Elbow flexion (video)

图 3 不同的康复评估效果(角度)

Fig. 3 Different rehabilitation evaluation results (angle)

其中，c_{com} 为完成度，x_{omax} 和 x_{omin} 为本方法获得平均指标最大值和平均最小值，x_{smax} 和 x_{smin} 为传感器获得平均指标最大值和平均最小值，当比值>1 时取倒数获得结果。该指标用于衡量本方法与标准传感器方法在动作评估上的一致性，从而验证其能否有效地模拟传统传感器[27]的测量结果，实验结果汇总于表 4。

表 4 方法在实际角度测试中结果

Table 4 Results in actual testing environments

实验次数	物理传感器 (平均-平均)	本方法 (平均-平均)	完成度 (平均)
1	0.6°-91.8°	0.9°-90.9°	0.986
2	0.1°-89.8°	0.1°-89.0°	0.991
3	0.2°-90.5°	0.5°-90.4°	0.984
4	0.1°-92.1°	0.3°-92.2°	0.998
5	0.5°-88.7°	0.1°-90.5°	0.975

如表 4 所示，在 5 次测试的平均值背景下，本方法所测得的角度指标与传统传感器测量结果基本一致。特别是在第 4 位患者的动作评估中，本文方法表现出最高的准确性，表明其在某些特定患者的动作检测任务上具有更高的可靠性和精准度。本方法已进入临床验证阶段，并完成 2 000 余例病人测试，平均准确率 93.2%，进一步提升了本方法的应用价值。

3 基于姿态捕捉的智能康复系统设计

3.1 硬件选型

如图 4 所示，结合先前研究与新方法，设计了一套智能康复运动评估系统。该系统的硬件架构由智算模块和视觉采集模块组成，其中视觉采集模块采用 Intel RealSense D435i 相机。该相机具备 RGB 图像采集功能，支持高清分辨率，帧率可达 30 fps，并具有广视野范围，能够满足系统对高精度捕捉的需求。

图 4 系统总体结构

Fig. 4 Structure of system

3.2 系统流程

系统首先由智算模块发送指令，控制视觉采集模块捕获患者的彩色图像(如图 5 所示)。随后，智算模块接

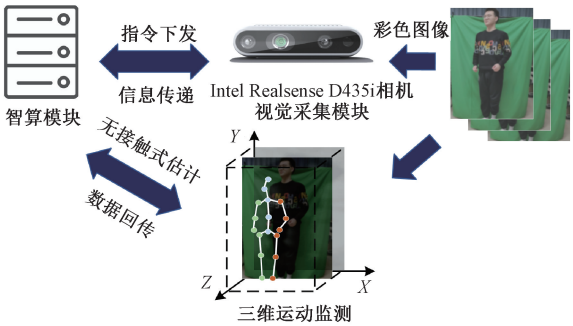


图 5　系统硬件检测流程

Fig. 5　Process of system

收图像并应用提出的三维方法进行无接触姿态估计,从而提取运动指标。

如图 6 所示,本系统支持体姿评估、平衡估计、关节活动度分析等十大核心动作,并可进一步细分至 120 项子动作,实现精细化康复评估。

软件部分提供了便捷的人机交互界面,通过消息队列遥测传输协议(message queuing telemetry transport, MQTT)协议实时传输硬件计算结果,并将数据动态显示在交互界面上,如图 7 所示。

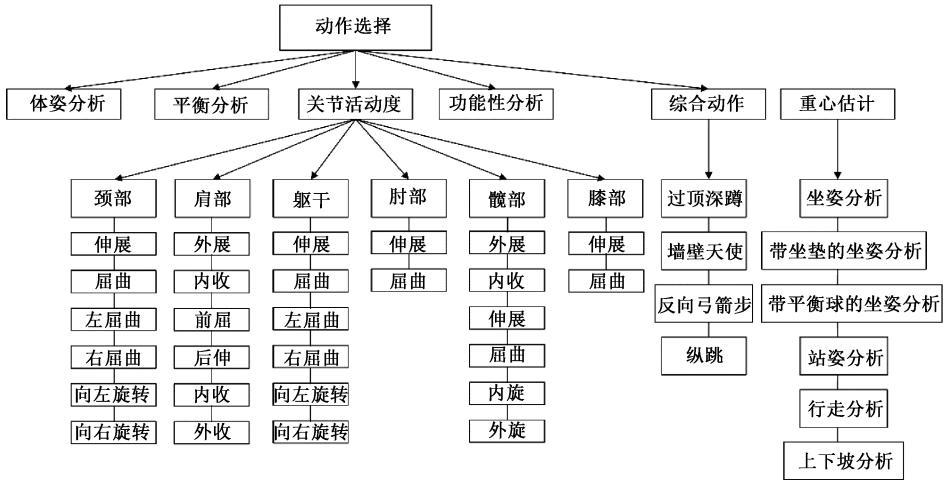


图 6　系统支持的模块

Fig. 6　Supported modules of the system

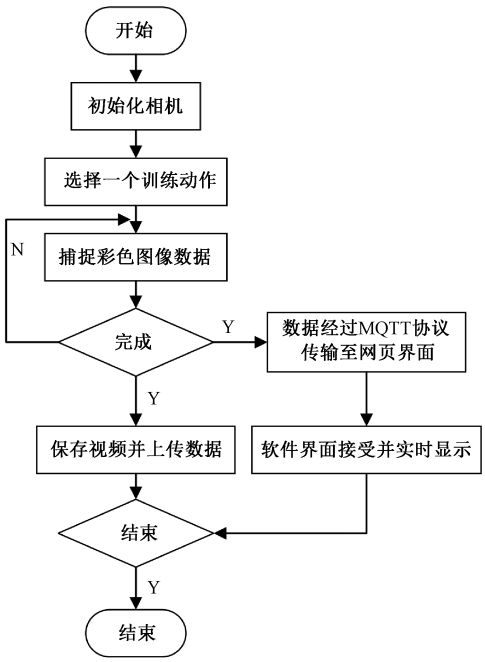


图 7　软件部分流程

Fig. 7　Flowchart of software part

3.3　系统功能展示

以体姿分析为例,如图 8 所示,在患者活动过程中,系统可实时追踪人体关节连线的水平夹角和高度差,从而快速检测高低肩、高低髋等异常情况,并加速运动纠正方案的执行。在系统界面左下角,系统可实时估算人体关键点之间的距离,并以中心点为基准捕捉前后微小变化。从而及时发现不自然的代偿现象,帮助改进运动中难以察觉的问题。界面右侧展示运动过程中人体骨骼与关节夹角信息,下方设有对齐圈,引导运动者快速调整姿势,以优化训练效果。

图 8　体姿分析动作

Fig. 8　Balance analysis

如图 9 所示,系统可对患者特定关节的活动度进行无接触式估计,并提供合理的参考值,便于直观对比。选取躯干作为分析对象,并以躯干向左屈曲为例,其中实际测量指标和参考指标分别用外环圈和内环圈表示,以清晰展现偏差情况。

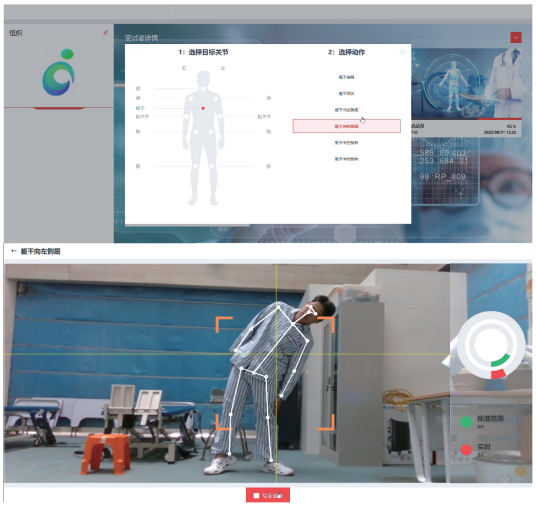


图 9 关节活动度

Fig. 9 Joint activity

以平衡运动模块为例,在患者单脚站立过程中,系统首先通过相机录制实时运动视频,随后将视频序列输入本文提出的模型,提取三维关节信息并进行角度指标估计。在分析过程中,系统将实时估算肩关节、髋关节、膝关节的旋转角度,以及头部的前后倾斜角度和旋转角度。为更直观地展示运动变化,系统提供了如图 10 所示的曲线图,用于动态呈现关键运动参数的变化趋势。

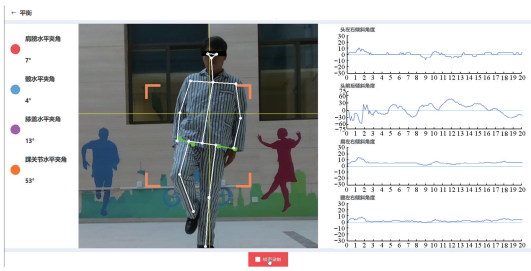


图 10 平衡运动检测

Fig. 10 Evaluation of balance

弓箭步测试要求测试者双手举过头顶,并完成前倾单腿下蹲动作(如图 11 所示)。系统重点评估以下关键运动参数:上肢肩-肘-腕对齐性、肩连线的水平夹角及轴旋转角度,下肢膝外翻程度及膝盖是否超过脚尖。综合上述指标,系统可自动判断动作是否标准,并通过“未完成/完成”标识显示在界面上直观显示,以便测试者快速获取关键运动指标,避免主观分析,提高评估的准确性。

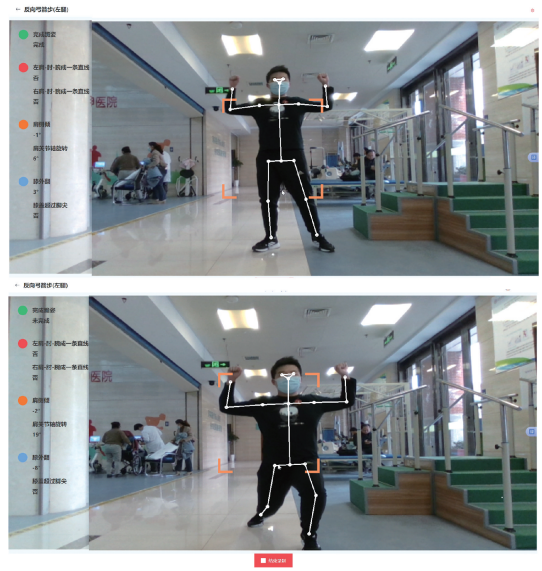


图 11 弓箭步

Fig. 11 Lunges

如图 12 所示,系统支持过顶深蹲动作评估。测试者需高举双手至头顶并完成深蹲。在此过程中,系统可实时检测以下运动参数:大腿角度、大腿与膝盖的距离、大腿相对水平面平行程度、肩侧倾斜角度、肩部轴旋转角度、肩-肘-腕对齐情况及膝盖外翻程度。通过上述指标,系统能够迅速识别 O 型腿等异常姿态,并为后续姿态矫正提供科学依据。

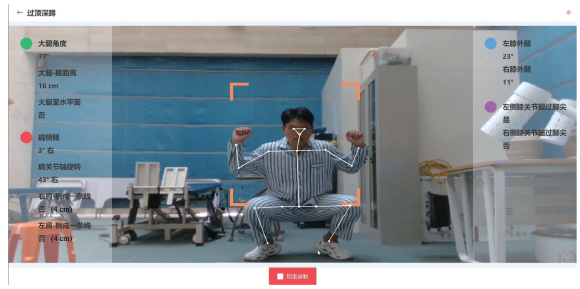


图 12 过顶深蹲

Fig. 12 Lung squat

系统最终生成分析报告,以直观方式呈现康复评估结果。第 1 页报告(如图 13 所示)记录患者信息及正面图像,并提供头部与脊柱角度、角速度及角加速度的分析图。

第 2 页报告(如图 14 所示)重点评估脊柱偏转方向及重心偏移距离,以进一步分析患者的姿态稳定性。通过康复评估报告,康复师可依据数据对比结果动态调整康复方案。例如,在本例中,患者的重心控制与左右平衡恢复良好,因此可继续沿用现有康复治疗方案。

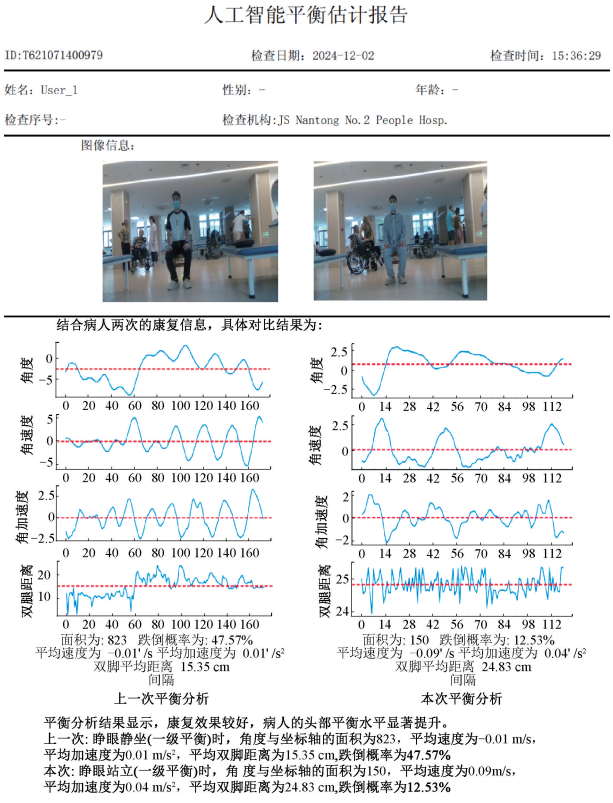


图 13　报告(第 1 页)

Fig. 13　Report(first page)

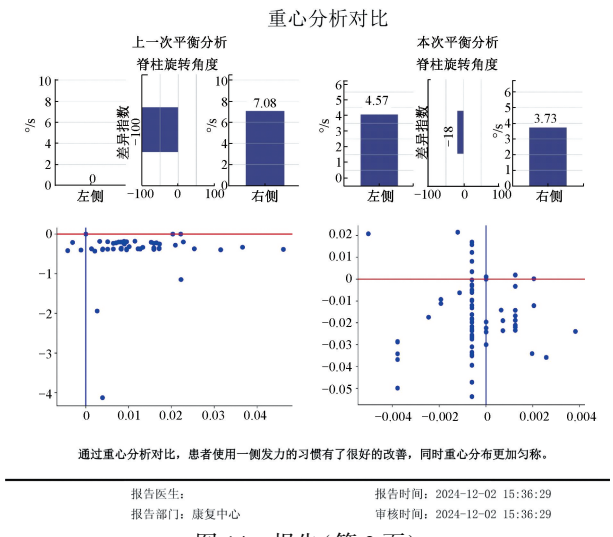


图 14　报告(第 2 页)

Fig. 14　Report(second page)

4　结　　论

本研究克服了二维姿态检测技术在康复视觉检测中的局限性,并提出了一种适用于康复运动场景的三维姿态检测方法。具体而言,设计了 QSFormer 进行粗特征提取,并结合空间-时间建模技术,精准捕捉运动过程中的时空特征。此外,通过 Extra Part 模块对现有模型进行精细化分析,从连续康复视频中提取骨骼信息,进一步增强检测精度。实验结果表明,提出的三维姿态检测技术在 4 项康复运动实验中均表现优异,其医学参数测量的有效性优于传统基于机器视觉的距离测量方法。

未来的工作将聚焦模型参数优化,提升推理速度,以满足实时应用需求;改进系统硬件形态,采用 AI 芯片代替传统计算机系统,以提升便携性;引入过程性评价指标,帮助康复师更精细地监控运动指标变化,优化康复方案。

参考文献

[1] 蒋青, 张雨. 新形势下运动损伤特点及细胞生物治疗的应用前景和挑战[J]. 北京大学学报(医学版), 2021, 53(5): 828-831.
JIANG Q, ZHANG Y. Characteristics of sports injuries under the new situation and the application prospects and challenges of cell biological therapy [J]. Journal of Peking University (Health Sciences), 2021, 53(5): 828-831.

[2] 姚玉峰, 裴硕, 郭军龙,等. 上肢康复机器人研究综述[J]. 机械工程学报, 2024, 60(11): 115-134.
YAO Y F, PEI SH, GUO J L, et al. A review of research on upper limb rehabilitation robots[J]. Journal of Mechanical Engineering, 2024, 60(11): 115-134.

[3] ZHANG K, ZHANG P CH, TU X T, et al. SR-POSE: A novel non-contact real-time rehabilitation evaluation method using lightweight technology[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 4179-4188.

[4] ZHANG C, YANG T Y, WENG J W, et al. Unsupervised pre-training for temporal action localization tasks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 14031-14041.

[5] ZHAO Q T, ZHENG C, LIU M Y, et al. Poseformerv2: Exploring frequency domain for efficient and robust 3D human pose estimation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 8877-8886.

[6] LI W H, LIU H, DING R W, et al. Exploiting temporal contexts with strided Transformer for 3D human pose estimation[J]. IEEE Transactions on Multimedia, 2022,

25: 1282-1293.

[7] ZHU W T, MA X X, LIU ZH Y, et al. Motionbert: A unified perspective on learning human motion representations [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 15085-15099.

[8] ZHANG J L, TU ZH G, YANG J Y, et al. Mixste: Seq2seq mixed spatio-temporal encoder for 3D human pose estimation in video[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 13232-13242.

[9] LI W H, LIU H, TANG H, et al. Mhformer: Multi-hypothesis Transformer for 3D human pose estimation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 13147-13156.

[10] YANG ZH D, ZENG AI L, YUAN CH, et al. Effective whole-body pose estimation with two-stages distillation[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 4210-4220.

[11] HUA W ZH, DAI Z H, LIU H X, et al. Transformer quality in linear time[C]. International Conference on Machine Learning. PMLR, 2022: 9099-9117.

[12] Miller T A. NSCA's guide to tests and assessments[M]. Human Kinetics, 2012.

[13] 罗智杰, 王泽宇, 岑飘,等. 基于改进 YOLOv8pose 的校园体测运动姿势识别研究[J]. 电子测量技术, 2024,47(19):24-33.
LUO ZH J, WANG Z Y, CEN P, et al. Research on human motion pose recognition algorithm based on improved YOLOv8 pose [J]. Electronic Measurement Technology, 2024,47(19):24-33.

[14] 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, 2013, 36(7): 1325-1339.

[15] SIGAL L, BALAN A O, BLACK M J. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion[J]. International Journal of Computer Vision, 2010, 87(1): 4-27.

[16] MARTINEZ J, HOSSAIN R, ROMERO J, et al. A simple yet effective baseline for 3D human pose estimation[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2640-2649.

[17] XU T H, TAKANO W. Graph stacked hourglass networks for 3D human pose estimation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 16105-16114.

[18] GONG K H, ZHANG J F, FENG J SH. Poseaug: A differentiable pose augmentation framework for 3D human pose estimation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8575-8584.

[19] CHEN T L, FANG CH, SHEN X H, et al. Anatomy-aware 3D human pose estimation with bone-based pose decomposition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(1): 198-209.

[20] PAVLLO D, FEICHTENHOFER C, GRANGIER D, et al. 3D human pose estimation in video with temporal convolutions and semi-supervised training [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7753-7762.

[21] YEH R A, HU Y T, SCHWING A G. Chirality nets for human pose regression [J]. Advances in Neural Information Processing Systems, 2019, 32: 1-10.

[22] LIU R X, SHEN J, WANG H, et al. Attention mechanism exploits temporal contexts: Real-time 3D human pose reconstruction [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5064-5073.

[23] WEHRBEIN T, RUDOLPH M, ROSENHAHN B, et al. Probabilistic monocular 3D human pose estimation with normalizing flows [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 11199-11208.

[24] 李昊楠, 毛剑琳, 张凯翔, 等. 一种基于安全区间的多机器人路径 k 鲁棒规划算法[J]. 仪器仪表学报, 2023, 44(10): 274-282.
LI H N, MAO J L, ZHANG K X, et al. Multi-robot path k robust planning algorithm based on safe interval[J]. Chinese Journal of Scientific Instrument, 2023, 44(10): 274-282.

[25] XU J W, YU ZH B, NI B B, et al. Deep kinematics analysis for monocular 3D human pose estimation[C]. Proceedings of the IEEE/CVF Conference on computer

Vision and Pattern Recognition, 2020: 899-908.

[26] ZHAO Y, WANG T H, LI W J, et al. Structural design of high-precision positioning system in weak signal environment based on UWB and IMU fusion [J]. Instrumentation, 2023, 10(2): 30-39.

[27] 虎勇, 吕辉岩, 李绍荣, 等. MEMS 姿态传感器在边坡表面位移监测的应用研究[J]. 电子测量与仪器学报, 2023, 37(7): 53-61.
HU Y, LYU H Y, LI SH R, et al. Research on the application of MEMS attitude sensor in slope surface displacement monitoring[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(7): 53-61.

作者简介

张堃,2005 年于江苏大学获得学士学位,2010 年于浙江大学获得硕士学位,2016 年于上海大学获得博士学位,现为南通大学电气与自动化学院教授,主要研究方向为医工结合人工智能、工业机器视觉、智能控制研究。

E-mail:zhangkun_nt@ 163. com

Zhang Kun received his B. Sc. degree from Jiangsu University in 2005, M. Sc. degree from Zhejiang University in 2010, and Ph. D. degree from Shanghai University in 2016. He is currently a professor at the School of Electrical and Automation Engineering, Nantong University. His main research interests include the integration of artificial intelligence in medicine and engineering, industrial machine vision, and intelligent control.

张彬,2009 年于北京体育大学获得硕士学位,现工作于苏州市体育专业运动队管理中心,主要研究方向为运动损伤防治。

E-mail:zhbnn333@ 126. com

Zhang Bin received his M. Sc. degree from Beijing Sport University in 2009 and currently working at the Suzhou Sports Professional Team Management Center. His primary research focus on the prevention and treatment of sports injuries.

华亮(通信作者),2001 年于南通工学院获得学士学位,2007 年和 2014 年于浙江工业大学分别获得硕士学位和博士学位,现为南通大学副校长、教授,主要研究方向为机器人及控制、图像处理与模式识别。

E-mail:hualiang@ ntu. edu. cn

Hua Liang(Corresponding author) received the B. Sc. degree from Nantong Instituted Technology in 2001, received M. Sc. and Ph. D. degrees are both from Zhejiang University of Technology in 2007 and 2014, respectively. Now he is a professor at Nantong University. His research interests include robot control, machine vision, pattern recognition.