

# 基于端云协同的商用设备运动纠正系统的设计与实现<sup>①</sup>

徐渊<sup>②\*\*</sup> 刘志刚<sup>\*\*</sup> 王卅<sup>\*</sup> 包云岗<sup>\*</sup>

(\*中国科学院计算技术研究所, 北京 100190)

(\*\*中国科学院大学 北京 100049)

**摘要** 动作的准确性对于提高锻炼质量和防止各种身体损伤至关重要。现有的解决方案一般依赖于昂贵的光学传感器(例如, Kinetic), 或者复杂的运动跟踪传感器(例如, 惯性测量单元(Inertial Measurement Units, IMU), 测力板)。这些传感器不具备便携性, 无法让普通用户随时随地地使用。本文介绍了一种运动纠正系统原型, GymBeats, 并首次展示了用户如何独自通过商用设备(例如, 手机, 笔记本)校正运动中身体姿势的错误。具体而言, 我们提出了一种新颖的上下文感知运动纠正方法, 包括状态识别和错误识别。我们还开发基于端云协同的系统, 并通过一种典型的全范围练习运动(杠铃深蹲)来描述我们系统的可行性。我们面向 15 位用户对 GymBeats 的两种反馈形式进行评估: 1) 实时语音反馈, 2) 可视化报告反馈。我们的结果表明, GymBeats 可以以低成本和便捷的方式显著提高身体姿势的正确性。

**关键词:** 运动纠正, 商用设备传感器, 骨架提取, 上下文感知识别, 实时语音反馈, 可视化报告反馈

## 0 引言

反馈在人们运动锻炼过程中十分重要, 因为它可以监控运动中身体的姿势, 并纠正错误。不幸的是, 在多数情况下, 人们几乎无法获取系统性的反馈来帮助他们纠正身体姿势。他们获取反馈的唯一信息来源可能是一面无法提供任何专业指导的镜子。更糟糕的是, 他们需要分散注意力来观察镜中的自己, 这进一步地降低了运动质量。

尽管近几年有很多运动纠正相关的工作, 来帮助经验不足的用户提高运动质量, 但是它们提供的方法既不便捷, 也不精确。一方面, 一些现有的工作通过基于深度信息的光学传感器获取骨骼信息, 并根据这些信息推断不同部位的关节角度<sup>[1,2]</sup>。这种方法通常只能提取人体骨骼正视图, 并测量运动涉及到的关节, 粒度较粗, 因此它们并不够精确。另一方面, 一些基于传统光学跟踪系统<sup>[3,4]</sup>和基于 IMU 的跟踪系统<sup>[5,6]</sup>通过在重要的解剖位置附加光学标记或者 IMU 传感器, 从而提供精确的运动信息。但是, 这些方法需要使用的传感器价格昂贵, 很难在普通用户中普及。

本文探索如下一个问题: 我们是否可以利用诸如

智能手机或者笔记本电脑这样的商用设备上的传感器来纠正运动时的错误姿势? 如果可以, 这将为普通用户提供一套既廉价又便捷的运动纠正解决方案。设想一下, 每次锻炼时, 您都可以通过便携式智能手机和一个支架获得实时的语音反馈, 并在运动后得到一份可视化的分析报告。这种方法可以提升您的动作标准度, 防止身体的损伤, 更好刺激肌肉, 甚至燃烧更多的卡路里。

实现这一目标的挑战在于运动纠正应用的计算需求与商用移动设备的硬件算力之间的不匹配。具体而言, 诸如智能手机和平板电脑之类移动设备依靠单目相机只能生成不带有深度数据的 2D 图像。而从这些 2D 图像中提取骨架信息需要应用深度神经网络(Deep Neural Networks, DNN)。这导致在精确性和响应性的约束下, 骨架信息提取过程对硬件资源的需求很高。不幸的是, 移动设备为了满足功耗和成本的约束, 其算力往往非常有限, 只能进行轻量级的运算。因此, 研究人员对在商用移动设备中运行此类人工智能应用程序的性能提出了担忧<sup>[7,8,9,10]</sup>。

①国家重点研发计划(2016YFB1000201), 国家自然科学基金(Grant No. 61702480, 61672499)和中国科学院青年创新促进会(2013073, 2020105)资助项目。

②男, 1991 年生, 博士; 研究方向: 计算机系统结构, 人工智能, 性能优化; E-mail: [xuyuan@ict.ac.cn](mailto:xuyuan@ict.ac.cn)

(收稿日期: 2020-xx-xx)

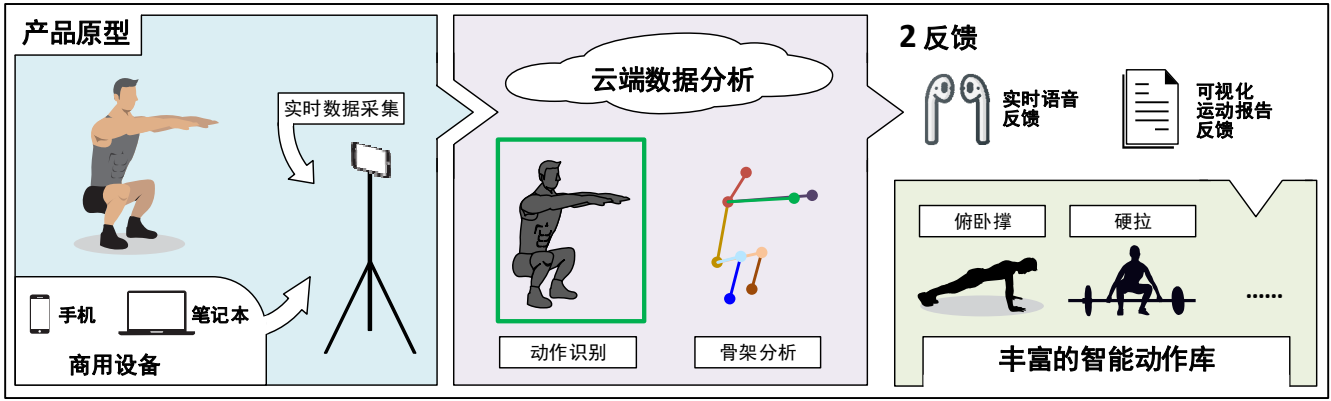


图 1 使用智能手机的相机进行动作纠正应用程序实例。智能手机可以通过将图像发送到云端进行动作分析。GymBeats 系统首先识别运动类型，从 2D 图像中提取骨骼信息。接着，它在丰富的运动动作库中搜索该运动类型的规则信息，并与实时骨骼信息进行匹配。最后，用户可以通过系统中两个反馈机制（实时语音反馈和可视化运动报告反馈）进行身体姿势的纠正

为了应对这一挑战，我们证明了可以通过构建物端-云端协同计算环境来满足上述约束。具体而言，基于云服务商提供的强大的服务器，我们可以将计算密集型的骨架提取任务卸载到云端服务器中，从而获得实时分析结果。我们的实验进一步表明，尽管本地设备和远程服务器之间的数据传输会引入网络延迟，但是这些额外的开销是可以接受的，并不会打破端到端实时性的约束。

基于上述观察，我们设计了一个基于端云协同的运动纠正系统——GymBeats（图 1），可以为没有经验的用户提供专业的运动纠正反馈。具体而言，我们首先将缩放后的 2D 图像从智能手机传输到云服务器，然后从每个图像中提取身体关键点。为了验证动作的正确性，我们设计了一种新颖的上下文感知运动纠正方法，将从身体关键点获取到的特征与我们动作库中每个动作的状态进行匹配。我们的方法包括两个阶段：状态识别阶段和错误识别阶段。状态识别阶段用于估计练习中每个动作的状态，包括站立，下蹲，底部，站起。错误识别阶段用于检查用户动作与标准规则是否匹配。

受 GymSoles<sup>[11]</sup>和 FEMO 启发<sup>[6]</sup>，我们提供了两种反馈机制：基于语音的实时反馈，和基于可视化报告的运动后反馈。我们访问了大学校园内的一群健身爱好者来确定用户需求，并通过咨询一些专业教练来设计不同运动的匹配规则。

概括而言，我们做出了如下贡献：

- 我们对比分析了当前最先进运动纠正技术的优缺点（第 1 节），并提出了一种新颖的基于上下文感知的运动纠正方法，包括状态识别和错误识别。我们以杠铃深蹲这种典型的全方位运动为例，来描述我们的方法（第 2 节）。
- 我们首次展示了一种基于廉价商用设备的运动

纠正解决方案。它采用了端云协同的系统设计，可以生成实时语音反馈并提供可视化反馈报告（第 3 节）。

- 基于我们的实践经验，我们为运动反馈系统的研究人员和从业人员提供了我们自己的见解（第 4 节）。

## 1 研究动机

我们建议使用商用设备来进行动作纠正，主要是因为当前的动作纠正方案并不能完全满足用户的需求。本节，我们首先通过调研确定了用户对于运动反馈系统的真实需求。然后逐个分析了现有解决方案的问题。最后我们介绍了本文的方案——GymBeats。

### 2.1 用户需求

为了了解用户在运动时的实际需求，我们在大学校园内对一群健身爱好者进行了半结构化访谈，让他们说出理想的运动纠正设备应当具备的功能。我们总结了三个最主要的功能，如下所示：

- **便携的廉价设备（Portable Cheap Device (PCD)）。**在我们收到的所有回复中，最关心的问题就是设备的便携性和价格。首先就是设备价格不能太贵，假如设备太贵的话，受访人通常情愿花钱请私教上以获取更人性化的指导。此外，他们希望设备具有便携性，以便他们可以在任何地方使用设备，例如健身房或家里。
- **丰富的运动库（Rich Exercises Benchmark (REB)）。**考虑到单一锻炼对肌肉刺激的局限性以及锻炼种类的多样性，受访人希望随着他们的需求变化，系统可以随之不断丰富他们的锻炼动作类型。因此该软件应当可以识别各种运动类型并提供纠正建议，从而指导用户。

- **专业的反馈 (Professional Feedback (PF))**。我们的目标用户是非专业的初学者，他们往往需要专业的反馈意见，以帮助他们提高动作质量。因此，该软件应当准确识别每个错误并提供实时反馈。

## 1.2 相关工作

先前关于运动纠正的方法主要包括使用光学跟踪相机，基于 IMU 的传感器和压敏设备。如表 1 所示，本小节根据上述三个要求进一步讨论当前技术存在的问题。

表 1 我们的方法与运动纠正其他相关工作比较

		PCD	REB	PF
光学运动跟踪	基于光学标记		✓	✓
	基于光学深度		✓	
	GymBeats	✓	✓	✓
基于 IMU 的运动跟踪		✓	✓	
压力数据跟踪				✓

### ● 光学运动跟踪

如今最先进的光学运动跟踪系统根据传感源不同，可以分为两类。第一类是基于光学标记的系统，例如 Vicon<sup>[3]</sup>和 OptiTrack<sup>[4]</sup>。这些商业系统通过将光学标记安装在用户的重要解剖位置（例如关节）来提供高精度的跟踪结果。但是，这种高精度的反馈需要一套昂贵的配件<sup>[12]</sup>，包括镜头，光学标记及其配套软件。普通用户通常负担不起这么高昂的价格。

第二种是基于光学深度信息的系统。这些系统使用基于深度信息的设备（例如 Microsoft Kinect<sup>[2]</sup>和 ASUS Xtion<sup>[1]</sup>）从获取的深度图像中检测骨骼位置，并推断每个像素的身体部位的分布<sup>[13]</sup>。基于骨骼信息，这些系统可以提取与关节角度有关的信息，以匹配标准运动姿势。开源和无需使用光学标记的特性使此类系统广泛用于各种训练的纠正<sup>[14,15,16]</sup>。然而基于光学深度的系统具有三个问题。首先，无论是 Kinetic 还是 Xtion 都尚未广泛普及，并且由于用户需要将其连接到笔记本电脑上使用，所以并不便携。其次，仅凭借正面测量出的角度信息不足以识别一些运动错误，例如后仰。第三，这种系统提供的是视觉的实时反馈，这会在用户运动时分散注意力。

### ● 基于 IMU 的运动跟踪

基于 IMU 的运动跟踪系统提供了基于商用可穿戴设备（例如智能手表和手环）的便携式廉价解决方案。通过分析多轴加速度计的测量数据，这些系统能执行活动识别<sup>[17,18,19,20]</sup>，步态分析<sup>[21,22,23,24]</sup>和健康监测<sup>[25,26,27,28]</sup>等任务。此外，一些工作着重于使用 IMU 数据来检测和协助运动<sup>[29,30,31]</sup>。例如，FEMO 使用附着在哑铃上的无源 RFID 标签的反射后向散射信号的

多普勒频移曲线来识别现场的自由重量训练，评估动作质量，并向用户提供有用的反馈。与光学运动跟踪系统不同，基于 IMU 的运动跟踪系统要实现高精度，就需要在身体上部署多个 IMU 传感器，并进行一些额外的工作，包括广泛的校准<sup>[32]</sup>以及精确地将每个传感器安放到身体的特定部位<sup>[33]</sup>。通常，只从智能手表获取到的 IMU 的数据不足以准确跟踪身体姿势<sup>[34]</sup>。

### ● 压力数据跟踪

压力数据跟踪系统通过测量地面反作用力和压力中心（Center of Pressure, CoP）的位移来分析平衡性和脚部动力学信息。这些数据从多是压敏设备（如测力板和支撑智能鞋垫）中采集得到。由于成本昂贵和缺乏便携性，压力数据跟踪系统主要用于实验室的康复研究<sup>[35,36]</sup>和步态分析<sup>[37,38,39]</sup>。近期的一项工作提出了一种基于智能鞋垫的运动纠正系统 GymSoles<sup>[11]</sup>，该系统基于脚部 CoP 提供震动和视觉反馈，从而纠正下蹲和举重时的身体姿势。尽管压力数据跟踪系统可以实现准确的 CoP 跟踪数据，但仍有一些不足。具体而言，此跟踪系统只能用于某些与压力有关的特定运动，例如下蹲和硬拉。然而，大部分其他运动都无法通过 CoP 进行分析，例如哑铃前举，哑铃侧举和杠铃卧推。

## 1.3 GymBeats

随着计算机视觉技术的飞速发展，越来越多的研究工作开始关注如何利用卷积神经网络（Convolution Neural Network, CNN）算法从没有深度信息的 2D 图片中检测出姿势信息<sup>[40,41,42]</sup>。2D 图片数据可以通过智能手机和笔记本电脑等商用设备轻松获取。基于这些工作，我们提出了 GymBeats，这是一种使用商用设备的运动纠正系统，可以为用户提供专业的反馈。在接下来的几节中，我们将详细描述运动纠正方法（第 2 节），系统架构和反馈（第 3 节）。此外，我们还将展示志愿者对 GymBeats 的体验结果（第 4 节）。

## 2 方法论

本小节首先简要描述一种典型运动的动作姿势规范（第 2.1 节），然后提出一种新颖的上下文感知运动纠正方法，以识别不满足规范的错误动作（第 2.2 节）

### 2.1 动作要求规范

我们以杠铃深蹲为例来描述我们的上下文感知纠正方法。选择该运动的原因有两个。首先，杠铃深蹲是一种全方位的力量运动，可以直接训练核心和背部肌肉群，包括脊柱直肌，绳肌，臀大肌，内收肌，股四头肌和斜方肌<sup>[43]</sup>。由于杠铃深蹲非常复杂，我们



可以借此对我们的方法进行全面的描述。其次,杠铃深蹲是一种危险的力量运动,动作错误和杠铃重量过大都会对锻炼者造成严重损伤。因此,杠铃深蹲是最近的工作中主要关注的锻炼之一<sup>[11,16]</sup>。

**基本概念。**尽管有许多书籍介绍杠铃深蹲的动作规范,但是本文遵循最具权威的杠铃训练书籍,即《Starting Strength Basic Barbell Training》<sup>[44]</sup>。在描述需求规范之前,我们首先对负重下的身体运动系统做一个必要介绍。当我们把举重者和杠铃视为同一个系统时,系统的重心(Center of Mass, COM)将从髋骨的中部位置移动到背部负重杠铃中部。当举重者本身负载更重的杠铃时,举重者-杠铃系统的 COM 可以近似为杠铃自身的中点。因此,我们要对杠铃进行很大范围的移动时,考虑杠铃本身的平衡性就非常重要。

脚中点(Mid-foot Point)位于足弓下方,是身体保持平衡的关键位置。图 2 显示了在举重者-杠铃系统中背部杠铃和脚中点之间的关系。显然,当脚中点位于 COM 的垂直线上时,系统将达到平衡。此外,任何杠铃运动的本质都是抵抗重力的向下做功。因此,最高效的杠铃移动姿势是保持 COM 与脚中点形成一条直线,仅垂直移动杠铃。

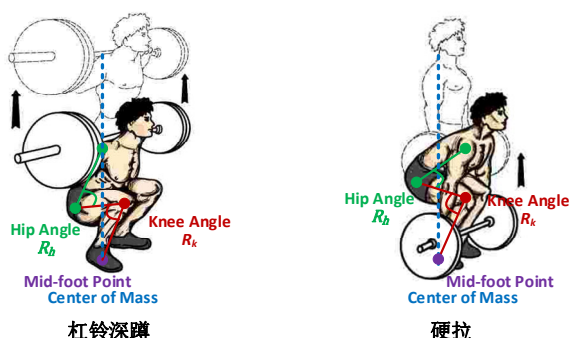


图 2 杠铃系统的诊断信息,包括髋骨角度,膝盖角度,脚中点和重心(蓝色虚线)。我们用  $R_h$  和  $R_k$  分别表示运动员的髋骨角度和膝盖角度

**杠铃深蹲。**我们在图 3 中总结了 4 个杠铃深蹲的基本要求规范。

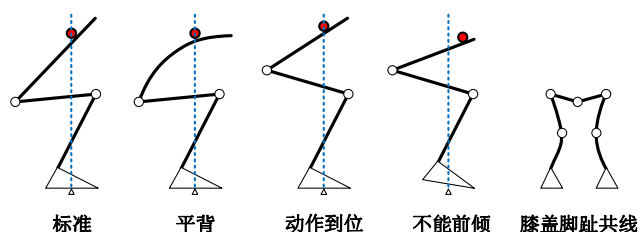


图 3 杠铃深蹲的基本要求规范。蓝色虚线表示 COM, 红色实心圆圈表示杠铃

1. **平背。**深蹲最大危险就是腰部受伤。这些损伤是由于椎骨在负重作用下运动不当,向前或者向后挤压椎间盘造成的。举重者应当保证背部平坦,

因为平坦的背部可以防止他的腰椎在负重过程中过度伸展或弯曲,并确保力在椎间盘上正确分布。

2. **动作到位。**深蹲的深度是影响关节稳定性与安全性的重要因素。正确的下蹲动作要求臀部下降到膝盖骨顶端以下的位置,以使得股四头肌产生向前力与腘绳肌产生的向后力保持平衡。在此深度以上的深蹲称为部分深蹲。该动作不会给臀大肌,内收肌或腘绳肌施加压力,而是给膝盖和四头肌施加压力。此外,由于运动范围短,部分深蹲会让举重者负载更大的重量,从而增大脊柱过渡负重导致的背部损伤风险。
3. **不能前倾。**在下蹲过程中,举重者的膝盖会向前弯曲,目光前视。因此,当身体失去平衡时,举重者会本能地前倾。失衡会使 COM (杠铃) 和脚中点之间的水平偏移产生力矩(旋转力)。举重者必须提供一定的力来抵消旋转力的影响并重新获得平衡性。因此,在整个负载运动中,举重者必须确保 COM 始终与脚中点在一条垂直线上。
4. **膝盖脚趾共线。**杠铃深蹲过程中和膝盖相关的错误类型通常有两种:膝盖内扣和前伸。本文着重于第一类错误类型。因为第二类错误类型取决于举重者的股骨,胫骨和躯干尺寸。对于一些股骨短而胫骨长的人,膝盖是可以在脚趾的前面。而对于第一类错误类型,当膝盖彼此向内移动时,股骨内侧和外侧的肌肉功能会减弱,这会导致举重者失去平衡并磨损膝盖软骨。因此,正确的动作应是将膝盖保持与脚趾在一条直线上。

## 2.2 上下文感知运动纠正

尽管此前有相关工作<sup>[11,45,46]</sup>通过考虑髋骨角度,膝盖角度以及这两个角度之间的比率来分析运动姿势错误。但是我们认为这些粗粒度方法无法准确识别出以上所有四个错误。例如,对动作到位中的错误识别需要检查举重者是否已经蹲到底部,而对于膝盖脚趾共线的错误识别则需要考虑膝盖之间的距离而不是角度。

为了解决这些问题,我们设计了一种新颖的上下文感知运动纠正方法(参见图 4)。如表 2 所示,除了传统的膝盖角度  $R_k$  和髋骨角度  $R_h$  外,我们还提出了一些新的识别指标,包括膝盖角速度  $V_k$ , 深度角  $R_d$ , 杠铃距离  $D_b$  和膝盖距离  $D_k$ 。我们指标的功能大体可以分为两类:状态识别和错误识别。

表 2 上下文感知运动纠正方法的度量值

度量值	功能	描述
$R_k$	状态识别	膝盖角指胫骨和股骨的平面夹角
$V_k$		膝盖角速度定义为膝盖角的变化率
$R_h$	平背	髌骨角指躯干和股骨的平面夹角
$R_d$	动作到位	深度角指躯干的平面和膝盖的水平线夹角
$D_b$	不能前倾	杠铃距离定义为杠铃与标准 COM 的偏移量
$I_b$		杠铃距离指数定义为杠铃距离与人的身高比
$D_k$	膝盖脚趾共线	膝盖距离定义为膝盖关节到脚趾垂直线的偏移量
$I_k$		膝盖距离指数定义为 $D_k$ 与人的身高比

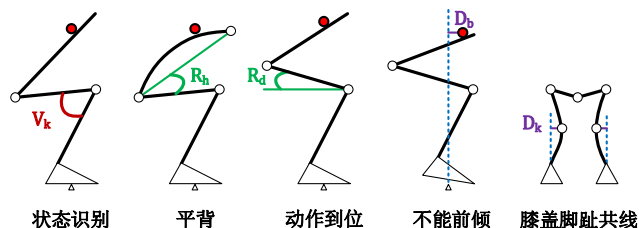


图 4 针对不同错误的上下文感知运动纠正方法

**状态识别。**通常,下蹲时,举重者可能处于一下四种状态之一:站立,下蹲,底部,站起。我们同时使用膝盖角度和膝盖角速度来识别负重过程中的每个状态。具体而言, $V_k$ 为正值表示举重者处于站起状态,为负值表示处于下蹲状态。当举重者开始/结束下蹲并处于站立状态时, $V_k$ 和 $R_k$ 分别接近  $0 \text{ rad/s}$  和  $180^\circ$ 。如果在上述条件下 $R_k$ 远小于  $180^\circ$ ,则意味着举重者处于底部状态。

**错误识别。**对于每一种错误类型,我们采用以下纠正方法:

1. **平背。**通过追踪底部状态下髌骨角  $R_h$  识别。如果髌骨角接近  $40^\circ$  或更大,则认为该动作符合平背要求。当背部向内弯曲时,胫骨平面的高度将会低于标准高度,使得髌骨角明显减小。
2. **动作到位。**通过追踪底部状态下深度角  $R_d$  识别。如果深度角接近  $10^\circ$  或更小,我们认为该动作满足动作到位要求。这里我们选择  $10^\circ$  而非负值,是因为相对真实髌关节位置,从 2D 图像中提取的髌关节关键点存在一定的偏移。
3. **不能前倾。**通过追踪站起状态下杠铃距离指数  $I_b$  识别。如果杠铃距离指数在  $-0.05$  到  $0.05$  范围内波动,我们认为该动作满足不能前倾的要求。正如我们在 2.1 节中讨论那样,杠铃和脚中点之间距离偏移会导致举重者失去平衡。由于杠铃距离  $D_b$  会随着相机和举重者之间的距离发生变化,因此我们需要用该距离与举重者身高的比值。
4. **膝盖脚趾共线。**通过追踪全状态下膝盖距离指数  $I_k$  识别。如果膝盖距离指数达到  $0.05$  或者更大,

我们认为该动作满足膝盖脚趾共线的要求。这里我们不直接使用膝盖距离  $D_k$  来估计该误差的原因与  $I_b$  相同。

### 3 GymBeats

本小节描述我们的 GymBeats 框架的系统设计和反馈机制。在 3.1 节中,我们描述系统结构,并详细介绍 GymBeats 每个组件功能。在 3.2 节中,我们重点展示预处理组件如何处理实际环境中噪声和遮挡问题。在 3.3 节中,我们将说明如何实现实时语音反馈和运动后的可视化报告反馈。

#### 3.1 系统设计

图 5 显示了 GymBeats 框架的架构。我们的框架是一个端云协同系统,其数据和代码分布在云端服务器(蓝色方块)和用户移动设备(绿色方块)上。框架中包含三个基础处理阶段:骨架提取阶段,上下文感知识别阶段和评估反馈阶段。在骨骼提取阶段,系统会处理传感器数据,提取出锻炼者的身体关键点位置以及一些基本运动特征。上下文感知阶段会通过提取出的信息验证动作的正确性,并将两种不同形式的反馈结果反馈到用户的智能手机上。我们详细描述执行的每个步骤(下列数字与图 5 中的数字一一对应)

- ① 举重者在锻炼过程中通过固定的智能手机录制杠铃深蹲动作的视频。视频中的每一帧都会发送到云端服务器中进行处理。
- ② 我们在云服务器中使用 OpenPose<sup>[47]</sup> 框架识别举重者的姿势,并生成一组身体关键点。OpenPose 是基于深度神经网络 DNN 的实时多人 2D 姿势估计框架,集成了许多最先进的视觉技术<sup>[48,49,50,51]</sup>。
- ③ 拍摄过程中,可能会存在噪声和遮挡,因此我们对这些原始身体关键点进行了预处理,以保证运动纠正过程的精确性。
- ④ 我们进一步从预处理后的身体关键点中提取表 2 中描述的度量值。

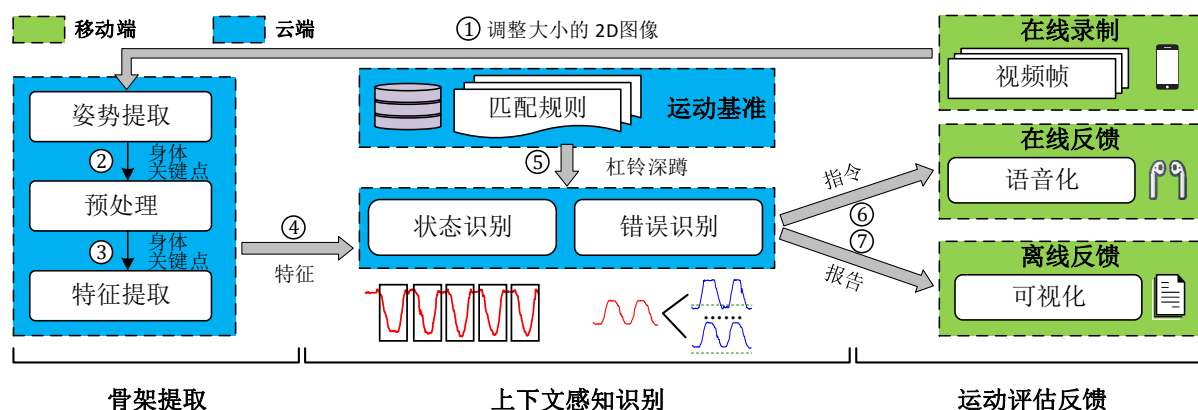


图 5 GymBeats 框架总体架构

- ⑤ 我们通过将每帧中提取的特征与运动标准库中相应运动的动作规则进行匹配，从而实现特定运动的状态和错误识别，以此验证每个动作的正确性。
- ⑥ 对于每一帧，我们都会提供实时的语音反馈，从而帮助举重者实时纠正其动作。
- ⑦ 完成所有动作后，我们将会把可视化运动报告发送到智能手机上，显示锻炼中的各个性能指标。

### 3.2 数据预处理

在真实运动环境中，GymBeats 通过商用设备获取到的数据可能存在三种噪声干扰。我们必须识别出这些噪声，并将其过滤。首先，我们可能需要从一群的运动的人中识别出目标用户，并在整个过程中跟踪目标的运动动作。其次，我们必须确定用户的运动方向，因为每个方向（例如正面或侧面相机）的特征提取和匹配规则都不相同。最后，遮挡是现实世界中最常见的问题。例如，当用户左侧面相机录制视频时，身体关键点将从右肩膀和右膝盖处提取。但是，当用户下蹲时，杠铃片会遮盖用户的小腿，这会降低该部位关键点提取的准确性。

对于前两种噪声，我们可以通过估计用户的相对位置并分析提取到的人体关键点的分布来确定目标。这是利用了运动时目标通常固定在一个位置，并仅沿着垂直方向移动的先验知识来进行判断。

对于遮挡，我们基于提取的关键点置信度，特定运动的先验信息和用户方向采用自适应的关键点预测方法。具体来说，当身体关键点模糊或不可见时，DNN 模型生成的置信度将会很小，甚至为零。因此，当置信度将至某个阈值以下时，我们认为关键点无效。由于杠铃只覆盖部分身体关键点（例如下蹲时的肩膀关键点），因此我们可以利用其他高置信度的关键点信息来预测被遮挡的关键点位置。此外，我们选

择与用户方向相同的身体关键点作为主要参考依据。例如，当智能手机位于用户左侧时，我们认为从身体左侧部位提取的关键点更为可信。

### 3.3 运动反馈

根据 2.1 节中讨论的需求调查，我们确定专业的运动评估反馈是帮助用户改善运动姿势的关键要素之一。因此，GymBeats 提供两种反馈：

#### (1) 语音反馈

- 运动时识别出错误姿势
- 满足实时性约束

#### (2) 可视化反馈

- 运动后精确显示身体姿势正确性
- 轻松跟踪历史锻炼数据

**语音反馈。**基于上下文感知识别阶段的生成结果，我们的语音反馈以语音指令形式生成。例如，当识别出运动动作不到位时，GymBeats 会发送一条语音指令到用户手机，从而提醒其需要再向下蹲一点。收到指令后，本地智能手机上的应用程序将调用文本语音引擎（Android Text to Sound, TTS）接口来发出语音提示。

在运动期间，为了确保安全性和人机交互的友好性，GymBeats 需要能够实时识别错误并对错误做出足够迅速的响应。假如用户在杠铃深蹲时接收到过时的语音命令，并按照命令指示进行动作，则很可能会导致受伤。不幸的是，基于 DNN 的 2D 姿态提取模型需要的强大计算能力，但是智能手机中有限的硬件资源并不能予以支撑。如图 6 所示，在本地智能手机上处理一帧时，端到端延迟高达 15.7 秒。此外，分辨率为 1920\*1080 的原始 RGB 图像的大小为 5.93MB，在每秒 30 帧（FPS）的处理频率下会浪费过多的带宽。



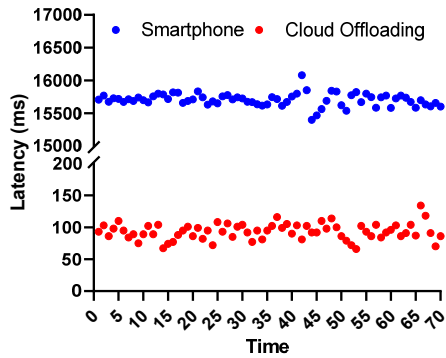


图 6 从发送一个视频帧到接收到一条语音反馈的端到端延迟

为了实现实时语音看反馈,我们需要减少每帧的处理时间并缩短两帧之间的间隔时间(即提高帧速率)。一方面,我们将姿势提取的所有计算工作转移到功能强大的远程服务器上,并将端到端的等待时间显著减少到 93.2 毫秒。另一方面,我们观察到姿势提取模型所需的图像分辨率固定为  $300 \times 300$ 。在此基础上,我们首先使用 OpenCV 库<sup>[52]</sup>将原始图像尺寸从  $1920 \times 1080$  调整为  $300 \times 300$  分辨率,然后将原始图像压缩为 40.76KB jpeg 格式的图像。如图 7 所示,我们在本地智能手机上以 30fps 的速率发送视频帧,并测量手机发送帧的带宽和在远程服务器中接收帧的速率。从图中我们可以看到调整大小操作可以将带宽从 28.1Mbps (蓝点) 显著减少到 7.4Mbps (红点),并将帧速率从 4.3fps (蓝线) 增加到 24.5fps (红线)。

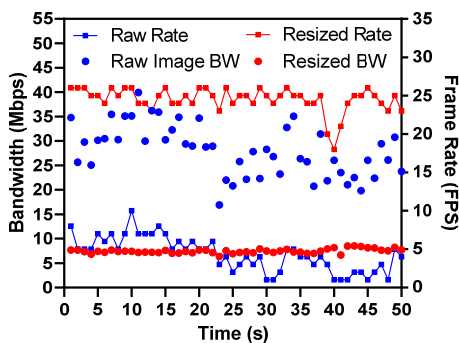


图 7 本地手机发送视频帧带宽和在云端服务器接收帧率可视化反馈。我们的可视化反馈以视频报告的形式生成,并在用户运动完后可以下载。具体而言,在处理完每个视频帧并在上下文感知识别阶段生成结果之后,我们将结果附加到每个视频帧上,并将其存储在云中。图 8 显示了视频报告中的两个帧,我们在每帧左上角四分之一处显示用户当前运动状态和错误识别结果。我们还添加了辅助线,以帮助用户纠正错误姿势,例如标准的重心 COM 线和杠铃的真实轨迹。用户可以浏览包含这些诊断信息的完整视频以及动作明显出错的关键视频帧。

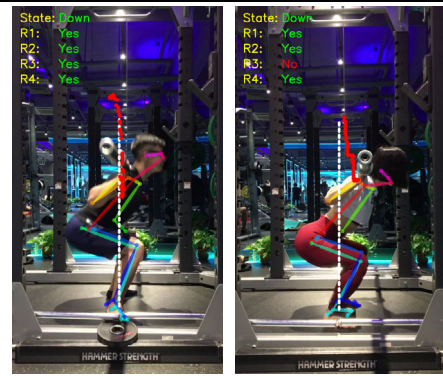


图 8 可视化视频报告反馈

## 4 实验

本小节通过一项用户研究来评估我们两种反馈机制在深蹲运动中的作用。我们分别回答了以下三个问题:

- 上下文感知运动纠正方法是否能准确识别运动状态和错误? (4.2 节)
- 在深蹲运动中,我们的两种反馈机制是否可以真正改善用户的身体姿势? (4.3 节)
- 与运动后的可视化反馈相比,用户更喜欢实时语音反馈吗? (4.4 节)

### 4.1 实验配置

我们使用商用智能手机和远程功能强大的服务器进行实验<sup>③</sup>。智能手机是配备高通骁龙 845 CPU @ 2.8GHz, 6GM RAM 和超光敏 AI 后置摄像头的最先进的 Android 手机,该摄像头可以在 120 fps 的速率下录制 1080p 视频。我们使用的无线耳机是 Sumsang Galaxy Buds, 用于发出语音指令。智能手机通过无源 5GHz 频段无线网络连接到我们实验室中功能强大的远程服务器(英特尔 Xeon®Silver 4114 CPU @ 2.20GHz, 31G RAM 和 Titan X Pro GPU)。

我们选择 15 位年龄分布在 18 至 35 岁之间的参与者,在填写同意后,使用 GymBeats 系统录制他们的运动视频。由于 GymBeats 旨在帮助初学者用户提高运动表现,因此我们选择的所有参与者都没有足够的杠铃深蹲经验。其中 10 名参与者几乎没有去过健身房,而其 5 名参与者每周去健身房花费 3 至 5 个小时进行有氧运动。所有的这 15 名参与者从未练习过杠铃深蹲,但是我们让一位经验丰富的教练在锻炼前进行了标准动作示范。参与者将在以下三种情况进行杠铃深蹲训练:(1) 无反馈,(2) 实时语音反馈,(3) 运动后可视化反馈。在确保所有其他条件相同的前提下,每个参与者在每种情况下都必须重复杠铃深蹲 5 组。在动作结束后,我们将给每位参与者一份问卷,询问他们关于这三种类型的反馈的有用性和偏好。此外,我们还将要求他们提供有关当前 GymBeats

③实验视频发布在网站 <https://www.bilibili.com/video/BV1q54y1t7FQ>

系统设计和交互模式的其他建议。

#### 4.2 状态和错误识别

本实验重点是测试 2.2 节中讨论的状态和错误识别的阈值的有效性。如图 9 所示, 我们进行了一组对照实验, 观察在运动过程中标准动作和错误动作的指标数据分布。我们首先收集了示范教练的指标数据, 作为基本事实对照组。然后我们跟踪了 15 名参与者的指标数据, 并选择了不符合这四个标准要求的动作数据。

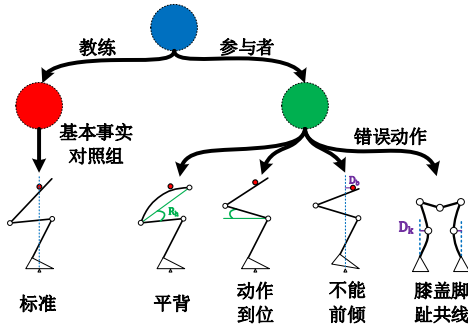


图 9 错误识别流程

图 10 展示了状态识别的指标结果。从图 10 中, 我们可以观察到 5 个虚线方形中的数据具有相同的分布, 这表明参与者进行了 5 次重复的深蹲运动。我们在第一个数据波形中标记了深蹲运动的四个状态。参与者蹲下时膝盖角速度为负, 而参与者站起时, 膝盖的角速度为正值。当膝盖角速度接近于 0 rad/s 时, 如果膝盖角接近于  $180^\circ$ , 这说明参与者处于站立状态。如果膝盖角远远小于  $180^\circ$ , 则说明参与者处于底部深蹲状态。

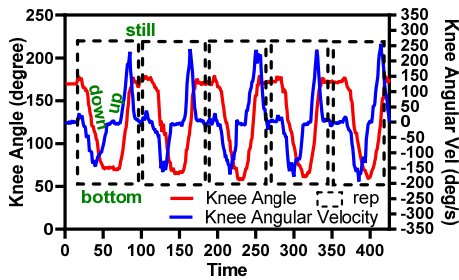


图 10 状态识别度量值测量

图 11 (a-d) 分别展示了错误动作的指标值分布, 这些动作分别无法满足平背, 动作到位, 不能前倾以及膝盖脚跟贴共线的要求。蓝色波形表示参与者的标准指标分布, 而红色波形表示参与者的错误分布。我们是用绿色虚线表示每个指标中的阈值, 从而识别动作是否满足要求。可以清晰地看到, 参与者背部向后弯曲时, 髋部角低于  $60^\circ$  (图 11a)。当参与者深蹲不到位时, 深度角则超过了  $10^\circ$  (图 11b)。当参与者前倾时, 杠铃距离指数超过 0.05 或者低于 -0.05 (图 11c)。当参与者膝盖和脚跟贴不处于一条直线时, 膝盖

距离指数超过 0.05 (图 11d)。

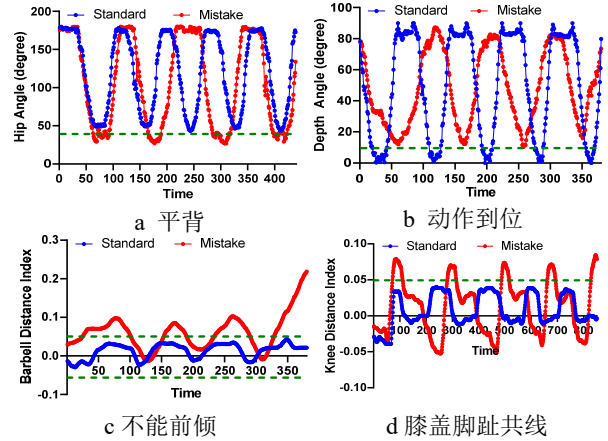


图 11 错误识别度量值测量

**发现 1:** 上下文感知方法对于杠铃深蹲细粒度的状态识别和错误识别来说十分必要。这种运动纠正方法也可以很好地应用到其他运动类型中。

#### 4.3 运动质量提高

为了验证 GymBeats 在深蹲运动中改善身体姿势的效果。我们统计了 15 名参与者在无反馈, 实时语音反馈和运动后可视化报告反馈条件下的动作指标。具体来说, 每位参与者总共进行 9 次深蹲运动。其中前 3 次运动在无反馈下进行, 中间 3 次运动在语音反馈下进行, 最后 3 次运动在可视化报告反馈下进行。我们仅关注特定状态下的指标数据, 例如平背错误识别中深蹲底部状态的髋部角度量, 因为这些数据对于帮助我们准确分析身体姿势的改善至关重要。

图 12 展示了我们深蹲试验中的指标分布。对于平背错误识别 (图 12a), 我们观察到实验数据分布分别是无反馈 ( $M = [34.2, 33, 32.8]$ ), 实时语音反馈 ( $M = [56, 56.4, 53.6]$ ) 和运动后可视化反馈 ( $M = [36.2, 43.3, 52.7]$ )。对于动作到位错误识别 (图 12b), 我们观察到实验数据分布分别是无反馈 ( $M = [6.9, 7.5, 13.6]$ ), 实时语音反馈 ( $M = [3.4, 1.2, 1.2]$ ) 和运动后可视化反馈 ( $M = [10.4, 5.1, 1.4]$ )。对于不能前倾错误识别 (图 12c), 我们观察到实验数据分布分别是无反馈 ( $M = [0.06, 0.04, 0.09]$ ), 实时语音反馈 ( $M = [0.02, 0.03, 0.01]$ ) 和运动后可视化反馈 ( $M = [0.11, 0.03, 0.01]$ )。最后, 对于膝盖脚跟贴共线错误识别 (图 12d), 我们观察到实验数据分布分别是无反馈 ( $M = [0.05, 0.06, 0.06]$ ), 实时语音反馈 ( $M = [0.03, 0.04, 0.04]$ ) 和运动后可视化反馈 ( $M = [0.06, 0.05, 0.04]$ )。



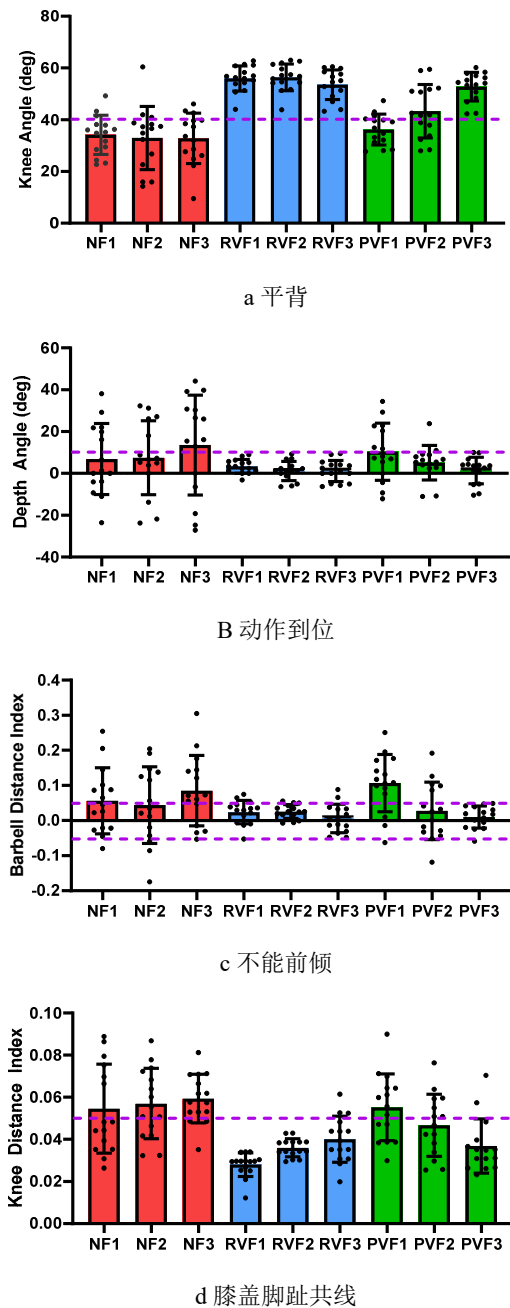


图 12 无经验参与者在无反馈 (NF), 实时语音反馈 (RVF) 和运动后可视化反馈 (PVF) 三种情况下的上下文感知度量值分布。紫色虚线表示每个度量值的阈值

通过对比每个度量值与紫色虚线所表示的阈值, 显然, 实时语音反馈 RVF 是纠正无经验参与者错误身体姿势的最佳方法。这是因为参与者在发现错误后可以立即纠正。另一方面, 运动后可视化反馈 PVF 在每次运动结束后起到作用。参与者在看到过报告后, 他们可以找到姿势错误的地方, 并在下一组运动中更正错误。请注意, 即使参与者从 PVF 意识到了自己所有的姿势错误, 也不一定能在下一组中全部纠正。

**发现 2:** 参与者可以从 GymBeats 中的实时语音反馈和运动后可视化反馈中受益。这两个反馈之间的区别

在于参与者是否实时获取到运动纠正信息。

#### 4.4 反馈偏好

我们分别对 5 位教练和 15 位参与者进行主观评价偏好的问卷调查。问卷的形式为针对每个反馈的 5 分李克特量表 (1 分表示非常低, 5 分表示非常高), 该调查表的结果如图 13 所示。

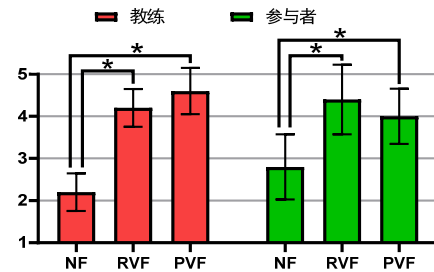


图 13 教练与无经验参与者对不同反馈的主观评价有用性总结。教练和无经验参与者对 RVF 和 PVF 偏好各有不同

对于教练, 就有效性而言, 三种反馈中均值最高的是运动后可视化反馈 ( $M = 4.6; SD = 0.5$ ), 其次分别是实时语音反馈 ( $M = 4.2; SD = 0.4$ ) 和无反馈 ( $M = 2.2; SD = 0.4$ )。单项方差分析测试显示主要影响数据为 ( $F_{2,12} = 35.43; p < 0.00001$ )。我们通过 Tukey's HSD 测试分别识别 NFV-NF 和 PVF-NF 对的差异性。数据显示有经验的教练更喜欢 PVF 而不是 RVF, 这是因为他们熟悉标准的深蹲动作, 因此 RVF 几乎没有帮助。而更直观的运动后可视化报告反馈 PVF 对于他们而言更为有效。

对于无经验参与者而言, 结果却并不一样。三种反馈中均值最高的是实时语音反馈 ( $M = 4.4; SD = 0.8$ ), 其次分别是运动后可视化反馈 ( $M = 4; SD = 0.6$ ) 和无反馈 ( $M = 2.8; SD = 0.8$ )。同时, 单项方差分析测试显示主要影响数据为 ( $F_{2,42} = 18.2; p < 0.00001$ )。与教练数据相似, Tukey's HSD 测试表明, 与 NF 相比, RVF 和 PVF 都被认为有效。

显然, 教练和无经验参与者对于不同反馈的主观性偏好各有不同。首先, 由于无经验参与者缺乏深蹲练习的经验, 他们希望 GymBeats 可以实时纠正他们的错误, 因此他们更喜欢 RVF 而不是 PVF。其次, 参与者的三种反馈的标准差和无反馈的均值都大于教练。这是因为参与者缺乏专业知识, 并没有意识到深蹲运动中非标准动作的危害性。最后, 参与者之间的个体差异比教练之间的个体差异数值更显著, 这导致三种反馈的接受范围变化幅度更大。

**发现 3:** 对于无经验参与者, 实时反馈对于改善运动表现更为有效, 而运动后反馈对于有经验的教练而言更为有效。与无反馈情况相比, 另两种类型反馈都认为是有效的。

## 5 讨论与局限性

### ● 可扩展性

目前, GymBeats 仍处于概念验证阶段, 只在深蹲运动中得到验证。未来的工作计划是提供丰富的运动动作库平台, 以满足不同人群的各种运动需求。我们相信我们的上下文感知运动纠正方法和基于云的框架可以很好地应用到其他运动中去。因为我们的方法和真实环境中教练通过自己丰富运动知识和肉眼观察来纠正学员的身体姿势的方法是类似的。但是, 我们方法的一个局限性是要求用户针对不同的运动方式放置相机位置可能不同。例如, 引体向上和俯卧撑运动要求相机放置的高度是完全不同的。

### ● 个人隐私性

将用户的图像传输到云服务器尽情运动纠正是会引起个人隐私问题。为了解决这个问题, 一种解决方案是在本地运行整个骨架提取过程。例如, 苹果公司近几年提出一种新的学习框架 CoreML<sup>[53]</sup>。该框架可以轻易将机器学习模型集成到手机应用程序中。在此基础上, 我们就可以同时满足实时性和隐私性的要求。但是这种新的框架只在配备高端硬件资源的 iPhone X 及其以后系列中采用。这会限制 GymBeats 的应用场景。因此, 个人隐私性和通用性之间存在着权衡取舍。

### ● 自定义运动计划生成

定制符合自己的运动计划对于初学者来说非常重要。因为通常无经验的用户会倾向于选择超过自己负重上限的重量来燃烧更多的脂肪或者更快的锻炼肌肉。错误的选择最终会导致错误的动作, 进一步对用户的身体造成损伤。更糟糕的是, 他们在运动时并不会意识到自己的错误。因此 GymBeats 的未来工作之一就是为初学者提供正确的指导, 根据运动中的每个动作反馈提供合适的重量建议。此外, 我们还会提供刺激不同肌肉部位的运动组合, 建议初学者在哪一天进行哪一类的运动, 从而提高初学者的锻炼效率。

### ● 运动游戏扩展

对于初学者而言, 理想的运动应用程序不仅可以提供专业的建议, 还可以带来乐趣。由于 GymBeats 允许使用商用设备进行运动检测和分析, 因此一个有前景并且可行的想法是使用我们的框架来实现类似 XBOX 的健身运动游戏平台。通过某些屏幕镜像技术 (例如 AirPlay<sup>[54]</sup>和 Google Chromecast<sup>[55]</sup>), 我们可以将智能手机屏幕与电视共享。因此, 云服务器, 智能手机和电视的硬件组合可以实现廉价的健身游戏平台搭建。在此基础上, 我们可以将许多流行的 Kinect 健身游戏移植到该平台。例如 Your Shape Fitness Evolved<sup>[56]</sup>和 UFC Personal Trainer<sup>[57]</sup>。

## 4 总结

我们提出了一种新颖的基于端云协同的运动纠正框架 GYMBEATS, 该框架可以使用智能手机和笔记本电脑等商用设备进行运动纠正。具体来说, 我们描述了现有解决方案存在的不足之处, 并提出了一种细粒度的上下文感知运动纠正方法, 包括状态识别和错误识别。在此基础上, 我们设计了基于端云协同的概念验证原型, 可以生成实时语音和运动后的可视化反馈。GYMBEATS 共有三个处理阶段, 包括骨骼提取, 上下文感知识别和评估反馈阶段。我们与 15 位参与者一起评估了 GYMBEATS。结果表明, 实时语音和运动后的可视化反馈都显著改善了用户在深蹲运动中的身体姿势。我们相信, GYMBEATS 将在智能手机上建立一个有前景的健身生态系统, 并为人们的整体健康和福祉做出贡献。

## 参考文献

- [1] Asus Xtion Pro [EB/OL]. [https://www.asus.com/3D-Sensor/Xtion\\_PRO/](https://www.asus.com/3D-Sensor/Xtion_PRO/): ASUS Corporation, 2020
- [2] Microsoft Kinect [EB/OL]. <https://developer.microsoft.com/en-us/windows/kinect/>: Microsoft Corporation, 2020
- [3] Vicon [EB/OL]. <https://www.vicon.com/>: Vicon Corporation, 2020
- [4] OptiTrack [EB/OL]. <https://www.optitrack.com/>: OptiTrack Corporation, 2020
- [5] Sunny C, David W M, Tammy T, et al. Activity sensing in the wild: a field trial of ubifit garden[C] // Computer Human Interaction (CHI), Florence, Italy, 2008: 1797-1806
- [6] Han D, Longfei S, Zheng Y, et al. FEMO: A Platform for Free-weight Exercise Monitoring with RFIDs[C] // ACM International Conference on Embedded Networked Sensor Systems (SenSys), Seoul, South Korea, 2015: 141-154
- [7] Tianshi C, Zidong D, Ninghui S, et al. DianNao: a small-footprint high-throughput accelerator for ubiquitous machine-learning[C] // International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), Salt Lake City, UT, USA, 2014: 269-284
- [8] Yunji C, Tao L, Shaoli L, et al. DaDianNao: A Machine-Learning Supercomputer[C] // Annual IEEE/ACM International Symposium on Microarchitecture (MICRO), Cambridge, United Kingdom, 2014: 609-622
- [9] Zidong D, Robert F, Tianshi C, et al. ShiDianNao: shifting vision processing closer to the sensor[C] // International Symposium on Computer Architecture (ISCA), Portland, OR, USA, 2015: 92-104
- [10] Dao-Fu L, Tianshi C, Shaoli L, et al. PuDianNao: A Polyvalent Machine Learning Accelerator[C] // International Conference on Architectural

- Support for Programming Languages and Operating Systems (ASPLOS), Istanbul, Turkey, 2015: 369-381
- [11] Don S E, Denys J C M, Lóic D, et al. GymSoles: Improving Squats and Dead-Lifts by Visualizing the User's Center of Pressure[C] // Computer Human Interaction (CHI), Glasgow, Scotland, UK, 2019: 174
- [12] OptiTrack Accessories [EB/OL]. <https://www.optitrack.com/accessories/>: OptiTrack Corporation, 2020
- [13] Jamie S, Andrew W F, Mat C, et al. Real-time human pose recognition in parts from single depth images[C] // Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 2011: 1297-1304
- [14] Edwin W T, Peijiang Y. Recognition of Yoga poses through an interactive system with Kinect based on confidence value[C] // International Conference on Advanced Robotics and Mechatronics (ICARM), Singapore, 2018: 606-611
- [15] Fraser A, Tovi G, Justin M, et al. YouMove: enhancing movement training with an augmented reality mirror[C] // ACM Symposium on User Interface Software and Technology (UIST), St. Andrews, UK, 2013: 311-320
- [16] Caleb C, Gene M P. Correcting Exercise Form Using Body Tracking [C] // CHI Extended Abstracts, San Jose, CA, USA, 2016: 3028-3034
- [17] Samanta R, Gabriella B, Marco K. Comparison of Different Sets of Features for Human Activity Recognition by Wearable Sensors[J]. Sensors, 2018, 18(12): 4189
- [18] Christina Z M, Yan T L, Queenie T K S, et al. Towards Wearable Comprehensive Capture and Analysis of Skeletal Muscle Activity during Human Locomotion[J]. Sensors, 2019, 19(1): 195
- [19] Aras Y, Billur B. Activity Recognition Invariant to Sensor Orientation with Wearable Motion Sensors [J]. Sensors, 2017, 17(8): 1838
- [20] Daqian Y, Jian H, Xikai T, et al. A Wearable Activity Recognition Device Using Air-Pressure and IMU Sensors[J]. IEEE Access, 2019, 7: 6611-6621
- [21] Arif R A, Hongnian Y, Michael V. An Automatic Gait Feature Extraction Method for Identifying Gait Asymmetry Using Wearable Sensors[J]. Sensors, 2018, 18(2): 676
- [22] Ik-Hyun Y, Jong-Hoon Y, Joseph A Z, et al. Biomechanical Gait Variable Estimation Using Wearable Sensors after Unilateral Total Knee Arthroplasty[J]. Sensors, 2018, 18(5): 1577
- [23] Yongjia Z, Suiping Z. Wearable Device-Based Gait Recognition Using Angle Embedded Gait Dynamic Images and a Convolutional Neural Network [J]. Sensors, 2019, 17(3): 478
- [24] Samer M, Allou S, Latifa O, et al. Recognition of gait cycle phases using wearable sensors[J]. Robotics and Autonomous Systems, 2016, 75: 50-59
- [25] Sumit M, Tapas M, M. J D. Wearable Device-Based Gait Recognition Using Angle Embedded Gait Dynamic Images and a Convolutional Neural Network [J]. Sensors, 2017, 17(1): 130
- [26] Duarte D, João P S C. Wearable Health Devices - Vital Sign Monitoring, Systems and Technologies[J]. Sensors, 2018, 18(8): 2414
- [27] A. S. A, O. S. A, A. A. Z, et al. Based Multiple Heterogeneous Wearable Sensors: A Smart Real-Time Health Monitoring Structured for Hospitals Distributor[J]. IEEE Access, 2019, 7: 37269-37323
- [28] Sunder A K, Aria G P, Feri S, et al. Contextual activity based Healthcare Internet of Things, Services, and People (HioTSP): An architectural framework for healthcare monitoring using wearable sensors[J]. Computer Networks, 2018, 145: 190-206
- [29] Keng-hao C, Mike Y. C, John F. C. Tracking Free-Weight Exercises[C] // Ubiquitous Computing (UbiComp), Innsbruck, Austria, 2007: 19-37
- [30] Heli K, Pekka S. Recognizing gym exercises using acceleration data from wearable sensors[C] // IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Orlando, FL, USA, 2014: 321-328
- [31] Sunny C, David W. M, Tammy T, et al. Activity sensing in the wild: a field trial of ubifit garden[C] // Computer Human Interaction (CHI), Florence, Italy, 2008: 1797-1806
- [32] Oberlander K. Inverse Kinematics: Joint Considerations and the Maths for Deriving Anatomical Angles. Inertial Measurement Unit (IMU) Technology[R]. 2015
- [33] Thomas S, Jörg R, Thomas S. IMU-Based Joint Angle Measurement for Gait Analysis[J]. Sensors, 2014, 14(4): 6891-6909
- [34] Chenguang S, Bo-Jhang H, Mani B S. MiLift: Efficient Smartwatch-Based Workout Tracking Using Automatic Segmentation[J]. IEEE Transactions on Mobile Computing, 2018, 17(7): 1609-1622
- [35] Hawkar O, Anurag S, Min-Hyung C, et al. VRInsole: An unobtrusive and immersive mobility training system for stroke rehabilitation[C] // International Conference on Wearable and Implantable Body Sensor Networks (BSN), Las Vegas, NV, USA, 2018: 5-8
- [36] Abu-Faraj Z O, Harris G F, Abler J H, et al. A Holter-type microprocessor-based rehabilitation instrument for acquisition and storage of plantar pressure data in children with cerebral palsy[J]. IEEE Transactions on Rehabilitation Engineering, 1996, 4(1): 33-38
- [37] Stacy J M B, Ari Y. B, Donna M S, et al. Gait Analysis Using a Shoe-Integrated Wireless Sensor System[J]. IEEE Transactions on Information Technology in Biomedicine, 2008, 12(4): 413-423
- [38] Postolache O, Girão P S, Pereira J M D, et al. Wearable system for gait assessment during physical rehabilitation process[C] // International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania, 2015: 321-326
- [39] Eskofier B M, Lee S I, Baron M, et al. An overview of smart shoes in the internet of health things: gait and mobility assessment in health promotion and disease monitoring [J]. Applied Sciences, 2017, 7(10): 986
- [40] Chunyu W, Yizhou W, Alan L Y. An Approach to Pose-Based Action Recognition[C] // Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 2013: 915-922
- [41] Diogo C. L, David P, Hedi T. 2D/3D Pose Estimation and Action Recognition using Multitask Deep Learning[C] // Computer Vision and



- Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018: 5137-5146
- [42] Vijay K, Anoop M N, Manohar P, et al. Pose-Aware Person Recognition[C] // Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017: 6797-6806
- [43] Glassman G. The crossfit training guide[EB/OL]. <http://library.crossfit.com/free/pdf/CFJ/Seminars/TrainingGuide/012013-S-Dy.pdf>: Crossfit Corporation, 2020
- [44] Rippetoe M, Kilgore L. Starting Strength Basic Barbell Training[M]. Wichita Falls/Texas/USA: Aasgaard Company Publishers, 2001
- [45] Lorenzetti S, Gülay T, Stoop M, et al. Comparison of the angles and corresponding moments in the knee and hip during restricted and unrestricted squats[J]. The Journal of Strength & Conditioning Research, 2012, 26(10): 2829-2836
- [46] Schoenfeld B J. Squatting kinematics and kinetics and their application to exercise performance[J]. The Journal of Strength & Conditioning Research, 2010, 24(12): 3497-3506
- [47] CMU OpenPose Project[EB/OL]. <https://github.com/CMU-Perceptual-Computing-Lab/openpose/>: CMU, 2020
- [48] Zhe C, Gines H, Tomas S, et al. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields[C] // CoRR abs/1812.08008, 2018
- [49] Zhe C, Tomas S, Shih-En W, et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields [C] // Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017: 1302-1310
- [50] Tomas S, Hanbyul J, Iain A. M, et al. Hand Keypoint Detection in Single Images Using Multiview Bootstrapping [C] // Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017: 4645-4653
- [51] Shih-En W, Varun R, Takeo K, et al. Convolutional Pose Machines [C] // Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016: 4724-4732
- [52] The OpenCV Project [EB/OL]. <https://opencv.org/>: 2020
- [53] The Apple Core ML Framework[EB/OL]. <https://developer.apple.com/machine-learning/core-ml/>: Apple Corporation, 2020
- [54] The Apple AirPlay Application[EB/OL]. <https://www.apple.com/airplay/>: Apple Corporation, 2020
- [55] The Google Chromecast Application[EB/OL]. <https://store.google.com/product/chromecast>: Google Corporation, 2020
- [56] The Your Shape: Fitness Evolved Game[EB/OL]. <https://marketplace.xbox.com/en-US/Product/Your-Shape-Fitness-Evolved/66acd000-77fe-1000-9115-d8025553084f/>: Microsoft Corporation, 2020
- [57] The UFC Personal Trainer Game[EB/OL]. <https://marketplace.xbox.com/en-US/Product/UFC-Personal-Trainer/66acd000-77fe-1000-9115-d8025451085b/>: Microsoft Corporation, 2020

## Design and implementation of cloud-edge based exercise correction system with commodity devices

Yuan Xu<sup>\*\*\*</sup> Zhigang Liu<sup>\*\*\*</sup> Sa Wang<sup>\*</sup> Bao Yungang<sup>\*</sup>

(<sup>\*</sup>Institute of Computing Technology, Chinese Academy of Science, Beijing 100190, China)

(<sup>\*\*</sup>University of Chinese Academy of Sciences, Beijing 100049)

### Abstract

The correct execution of exercises is essential to enhance workout quality and prevent various bodily injuries. Existing solutions rely on either expensive depth-based optical sensors (e.g. kinetic), or complex motion tracking sensors (e.g. Inertial Measurement Units (IMU), force-plates). None of them are suitable for normal users to adopt everywhere and every time. This paper introduces an exercise correction prototype, GymBeats, and shows for the first time that commodity devices such as smartphones and laptops can be used to correct body posture while users do exercise alone at the gym or home. Specifically, we propose a novel context-aware exercise correction methodology, including state recognition and mistake recognition. We also design a proof-of-concept cloud-based system and use a typical full-range-of-motion-exercise (i.e. barbell squats) to describe our methodology. GymBeats was evaluated with 15 users in two feedbacks: 1) real-time voice feedback 2) visual report feedback. Our results show that GymBeats can significantly

improve body posture in a low- cost and convenient way.

**Key words:** exercise correction, commodity device sensor, skeleton extraction, context-aware recognition, real-time voice feedback, post-exercise visual feedback