

对于医生：1、部分医生不重视康复；2、医生时间繁忙，无法全程指导。
对于医院：1、部分小医院没有康复科；2、大医院有床位周转率的要求。

二、研究目的

本研究的目标是基于人工智能算法开发一套基于移动终端的骨科术后下肢康复监测系统，其意义在于能够帮助患者在家中获得良好的康复训练指导，改善疾病预后，提升病人满意度，提高医院床位周转率，同时为科学研究提供临床大数据（研究意义）。

引言：

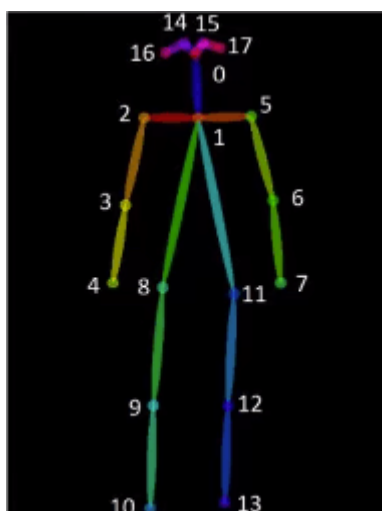
最近我在网络中看到这样一则笑话，39.5 度的司机载着 38.5 度的医生开的药给 37.5 度的患者送去，在这看似开玩笑的语气下隐藏着一个不争的事实，就是如今的医疗资源稀缺。为什么我要说这些呢，因为这和我今天要讲的主题：“智能骨科康复训练指导系统”有关。

智能骨科康复训练指导系统指的是一套能由患者自主操作的系统，这套系统能给患者做出康复指导，并对患者的康复训练进行监测，帮助患者在居家环境下也有较好的恢复效果。

除了前文所述，开发这套系统的原因有二：对于医生：1、部分医生不重视康复；2、医生时间繁忙，无法全程指导。对于医院：1、部分小医院没有康复科；2、大医院有床位周转率的要求。

接下来我将播放一小段视频，来帮助大家了解康复训练是怎样检测的。（上视频，然后对视频进行解说）大家可以看到，视频中的人物在做深蹲运动，并且人体上的某些关键点（胳膊，肩膀，膝盖，胯，脚腕）被追踪且识别了出来，很多动作都要根据这些关键点进行特征提取，将记录这些点的数值变化。

说到底，这种技术其实是人体姿态估计，它属于关键点检测和目的地跟踪，它本质是一个回归问题，要求输出关键点的坐标，



（mscoco 上的 17 个关键点）

至于某些应用要求多人检测，则需要用其他方法先将人之间分离（对人进行粗略的定位）再检测。但这里仅仅作为提及，不深入讨论（单-多）

既然人体姿态估计能解决这么多的问题，也衍生出一些新问题。那么解决这个问题的算法有那些呢？

OpenPose: (支持多人): <https://www.jianshu.com/p/3aa810b35a5d> 中文文档由美国宾夕法尼亚大学开发。

DeepCut: 先把所有关键点检测出来, 再把它们组装成一个一个的人。

Simple baselines

这个项目使用的是 Mediapipe, 它是一个在移动端和终端设备上进行边缘计算的库, 所以它的模型非常轻量化, 它运行的速度非常快, 可以在智能手机这样的廉价的设备上运行。

Mediapipe: BlazePose

专门为部署在移动设备和边缘终端设备上计算而生的, 所以它天生就是为了追求机制的速度而不是极致的准确。(16: 30)

这张图我相信很多同学已经见过很多次, 它只会再第一正的时候用一个人体检测器人脸检测器, 把人体的区域 ry (Regina 偶发 in) 检测出来, 然后再把它喂给关键点检测网络里, 输出关键点, 用上一帧得到的关键点数据, 来推算出下一帧人体可能出现的区域, 也就是上一帧的输出可以被下一帧的输入作为辅助。往后只要这个人还出现再画面中, 我们也就不需要再调用这个人体检测器了。可以推算出, 这样就可以大大加快速度, 不需要再去调用额外的 face detector。所以其实它并不能把轨迹真正的勾勒出来, 而是上一帧的输出可以被下一帧的输入作为辅助。

第二张图: 它结合两种人体姿态估计的两种主流技术, 第一种 heatmap, 热力图, 第二种是回归技术, 热力图是输出的是一张图像, 比如输出的是灰度图, 这张图和原图大小是一模一样的, 只不过里面像素某个关键点出现的概率, 比如在肩膀这热力图就会有比较高的值, 这是在高诉你这是肩膀可能存在的区域。输入图像, 输出热力图。这里有 33 个关键点, 就有各自 33 张热力图, 没个关键点给出 3 个方向的 offset 的偏移量, 所以就是 99, 它输出的是 64*64 的图, 这个图有 99 个通道, 表示的是, 这 33 个关键点的 99 个偏移量。这就是网络的第一个图。另一个头是, 直接让网络预测回归出关键点的坐标, 实现的箭头是跨层链接, 虚线箭头是不会反向传播回去的, 而输出的时候直接使用预测的结果, 能大大减少时间。

脸检测是怎样实现的:

Op: ar 准确率高, 大规模集群 gpu 但速度慢

Pck@0.2 是人体姿态估计的一个经典指标, 它表示的是预测的点的坐标和正式值之间的欧式距离小于人体躯干距离的百分之二十, 那么就判断这个点预测对了。所以看这个数据就能反映性能。

23: 30 knn

获取坐标了之后, 怎么判断做了一个深蹲呢, 举例摸脸, 只需要检测脸和手之间的距离就行了。显然深蹲要复杂得多。这时我们需要更加细节的测量: 主要节点间都进行距离的测量, 有必要的话测量角度, 用这些距离和角度作为特征就能定量一个特定的动作了, 当一个动作来的时候只需要把提取同样的特征然后把这个特征向量和训练集中原有的特征向量进行对比, 判断这个动作属于那类。

构建数据集：拍照：各个角落，完整人体，样本量越多越好

第一步：拍摄深蹲二分类数据集（也就是蹲下状态和站立状态），把每一张照片用 mediapipe 来进行关键点检测，然后构建出一个深蹲的模型，也就是说有新的照片来的时候它从而能够判断出是蹲下的还是站立的，并且能够量化出它的动作是否标准，从而实现计数。

两个端点拍照，蹲下和站立

再用 blazepose 来跑每个图片，把每张图片的特征都提取出来，再构建一个训练集。

可以用 knn 邻近算法来做。就是提取图片的特征，与训练集中与当前特征向量相差最近的向量相同。

阈值设置

总结

Blaze 意为“火焰、失火”，反映了这一系列算法是从算法设计到实现都竭力追求“极速”的特点，方便对运行速度要求比较高的 AR/VR 开发者使用。

BlazePose，一个轻量级的卷积神经网络架构，用于人体姿势估计，它是为移动设备上的实时推理量身定做的。在推理过程中，该网络为一个人产生 33 个身体关键点，并在 Pixel 2 手机上以每秒超过 30 帧的速度运行。因此它特别适合实时用例，如健身追踪和手语识别。在 Pixel 3 上 GPU 运行，BlazePose 可以达到 112 FPS(frames per second)。

BlazePose 是从达芬奇的《维特鲁威人》中得到的启发，可预测人的臀部中点、外接整个人的圆的半径以及连接肩部和臀部中点的直线的倾斜角度，共计可预测 33 个人体关键点。根据手和脚的比例和方向信息，即使是非常复杂的情况，比如特定的瑜伽姿态，其也能得到一致的追踪。