skeleton

概要

- 实现了：输入：野外单一图像。输出：3d joint positions +3dangular pose(bone rotation)+去皮skined动画
- 怎么做：两步：1回归：回归骨骼的旋转得到坐标（通过考虑骨骼结构）2.细化：交叉热力图（堆叠热力图中的xy and zy坐标）
- 用到的数据集：MPⅡ+人工标注3d信息+骨骼旋转；human3.6，包含了三维位置和骨骼旋转信息（但主题较少，10个左右）

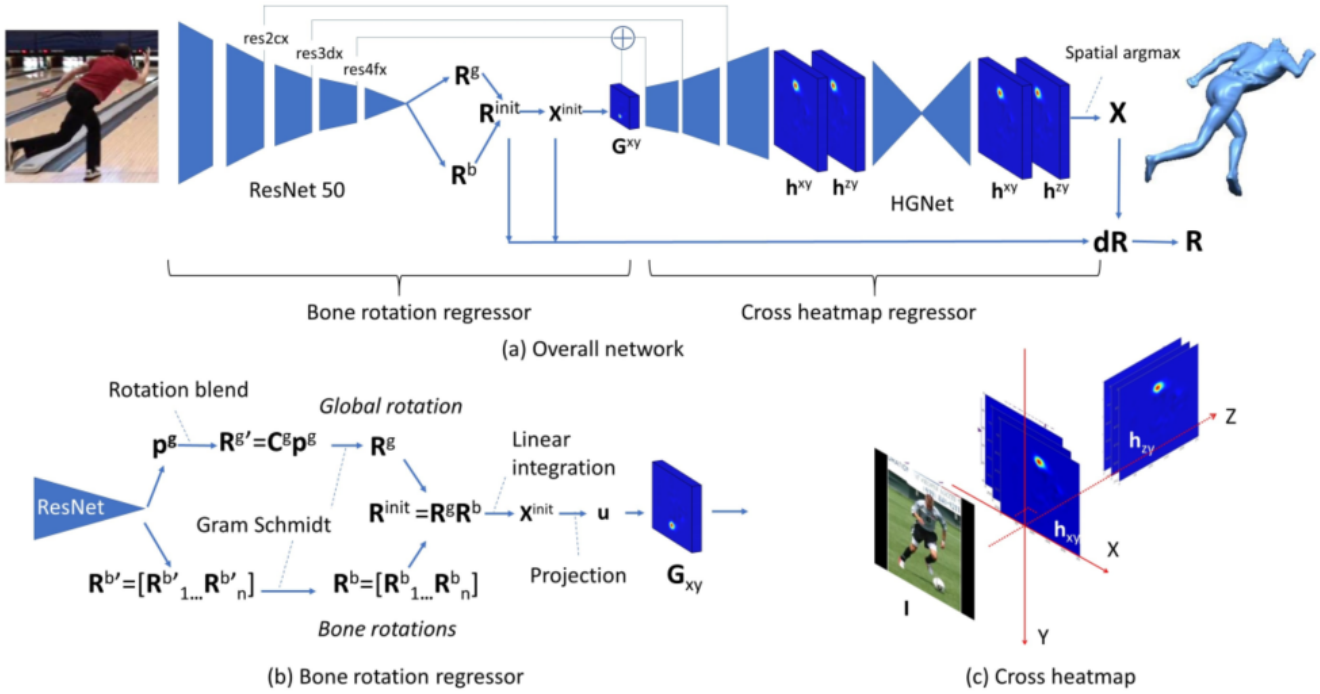
背景

- CNN：detect **2d positions** accrurately: **how to** achieve **accruately**:represent 2d joint locations as **heatmaps** and iteratively refines them by **context infomation**
- 3d 主要挑战：
  - how to represent 3d pose:使用热图效果不错（例如体积和2D热图+深度）表示3D姿态。CNN的高**非线性？？？？？？？？？？？？？？？？**不利于学习到3D joint 定位。而使用 **计算机动画和生物力学**有利于预测3D joint 位置+骨骼旋转角度姿态（如关节角度+节段旋转）
  - data sarcity（数据集缺乏）：3D缺乏，特别是3D骨骼角度位姿很难标注。最常见：MoCapsystem+rgb摄像机同时使用，但只在部分场景下有用。因此 **只从3D位姿数据集**中获取信息精确定位3D JOINT很困难

    解决方法：骨骼结构+构建了新的数据集，人工标注（对mocap数据集的加工）
- 本文主要用了骨骼变换网络（骨架网）即加入了 **骨骼结构做判断，即热图预测+skeletonnet**
  - 利用骨架网的步骤：1.回归骨骼旋转+考虑骨骼结构得到初始解 2.细化：在初始解的基础上，利用CNN热图回归细化
- 贡献总结：
  - end-to-end cnn网络预测 **关节位置+骨骼旋转**
  - 提出 **骨骼旋转回归器**预测3D位置（通过使用3*3变换矩阵），提出gram schmidt正交层将任意线性变换转化为旋转
  - 解决 represention问题：**3D交叉热图=xy热图（2d joint）+zy热图** 优于体积热图
  - 建立了野外的三维人体姿态数据集

网络结构

skeleton transformer networks



两部分组成：

- a 骨骼旋转回归 gxy 得到初步解，尊重骨骼结构不会有类似左右混淆的大错误
- b 交叉热图回归 hxy+hzy

1.bone rotation regressor

- 全局旋转（转换为**分类**问题 坐站躺）

  p^g是全局**分类**结果（坐、躺.....），旋转混合 使用了施密特正交化来获得正交矩阵。

  得到的是3维关键点（由2D关键点）的位置，之后才是进一步细化什么的

  这里用了3*3rotation matrices 来得到一个初始解

  因为是分类问题，所以用的是交叉熵

$$\mathcal{L}_{\mathrm{RotG}} = L_{cls}(\mathbf{p}^{\mathrm{g}}, \bar{\mathbf{p}}^{\mathrm{g}})$$

  其中 **标签p-g是通过对数据集进行k均值聚类来得到的**，output pg是通过RESNET网络得到的

  p-g是独热编码，

- 相对于某一根参照的骨骼各部分的旋转（**回归问题**）

  R^b'意味着局部回归问题（局部**相较于整体的位置信息），这里的Rb'是向量组合（向量数目=Bones的数目）**

  回归问题，用的是MSE

$$\mathcal{L}_{\mathrm{RotB}} = \sum_{i}^{n} ||\mathrm{vec}(\mathbf{R}_i^{\mathrm{b}'}) - \mathrm{vec}(\bar{\mathbf{R}}_i^{\mathrm{b}})||_2^2$$

- 结合：结合前 都要施密特正交化，然后用$R^{init} = R^g R^b$即全局变化与局部的信息乘积得到绝对旋转，

  如何得到3D关键点位置：把这个绝对旋转应用在**静止状态下的原始骨向量上（即回归问题上的各个向量）**

2.cross heatmap regressor交叉热图回归

目的：细化 关节三维位置投影到xy平面（3D-2D），沙漏模块用来计算交叉热图（交叉热图用来连接两个热图），叠加xy yz的热图，就可以得到xyz坐标

其中h^{xy}表示2D joints在image space(xy space)，h^{yz}表示在zy space上的

回归问题，用MSE（热力图的误差）：

$$\mathcal{L}_{\text{hm}} = \sum_{i}^{m} \sum_{j,k} ||\mathbf{h}^{\text{xy}}_{(j,k)} - \bar{\mathbf{h}}^{\text{xy}}_{(j,k)}||_2^2 + \sum_{i}^{m} \sum_{j,k} ||\mathbf{h}^{\text{zy}}_{(j,k)} - \bar{\mathbf{h}}^{\text{zy}}_{(j,k)}||_2^2$$

注意这里还有标签hxy-和hzy-，m是关键点数

注意看图，得到了以后先投影得到hxy hzy然后沙漏网络计算交叉热图，最后根据交叉热图来得到X（3维坐标细化），用X乘以之前第一部分得到的绝对旋转得到最后的旋转量R

即得到了一个位置坐标X和一个旋转量R

X和R的损失函数，注意这里也有个标签X-和R-

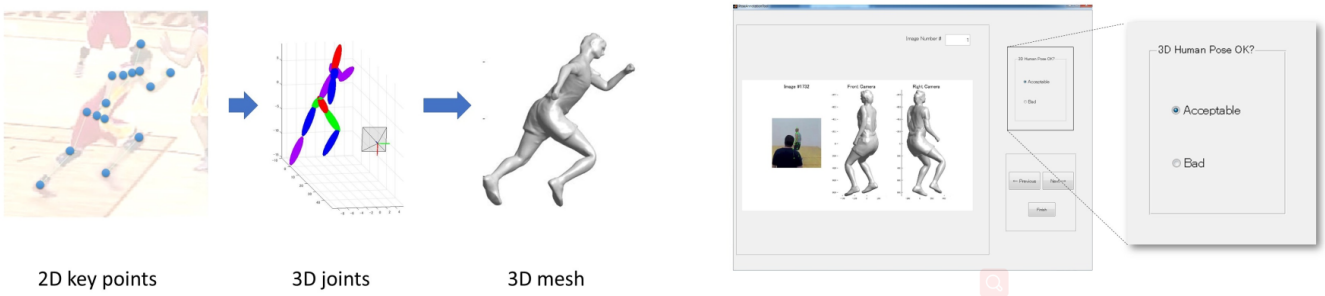
$$\mathcal{L}_{\text{pos}} = \sum_{i}^{m} ||\mathbf{x}_i - \bar{\mathbf{x}}_i||_2^2, \quad \mathcal{L}_{\text{Rot}} = \sum_{i}^{n} ||\text{vec}(\mathbf{R}_i) - \text{vec}(\bar{\mathbf{R}}_i)||_2^2$$

对于得到的R和X，用线性混合skining可以得到3D网格，这个过程是一个**线性矩阵乘法**

3.损失函数

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{RotG}} + \alpha\mathcal{L}_{\text{RotB}} + \beta\mathcal{L}_{\text{Rot}} + \gamma\mathcal{L}_{\text{pos}} + \lambda\mathcal{L}_{\text{hm}}$$

结果



(a) 3D mesh reconstruction　　(b) annotation interface

注意a图，由2D关键点的输入，得到3D关键点的位姿，再得到3D网格图形。b图是标注的系统

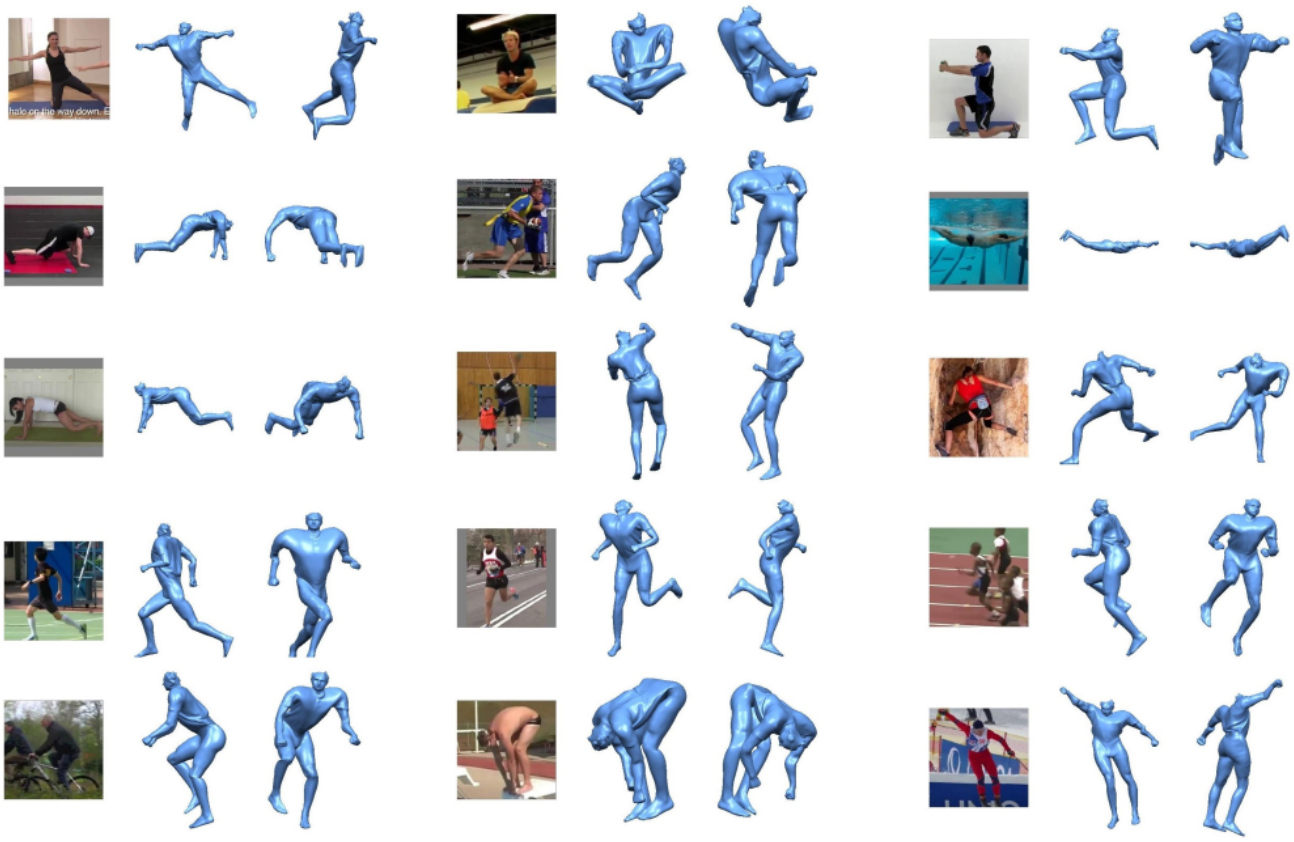


Fig. 3. Some results on in-the-wild images.

最后得到的是一个3DMesh可以旋转

数据集的补充

注释存储在 `RELEASE` 具有以下字段的matlab结构中

- `.annolist(imgidx)` - 图像注释 `imgidx`
  - `.image.name` - 图像文件名
  - `.annorect(ridx)` - 一个人的身体注释 `ridx`
    - `.x1, .y1, .x2, .y2` - 头部矩形的坐标
    - `.scale` - 人体尺度为200像素高度
    - `.objpos` - 图像中粗糙的人体位置
    - `.annopoints.point` - 以人为中心的身体关节注释
      - `.x, .y` - 关节的坐标
      - `id` - 关节id（0 - r踝关节，1 - r膝关节，2 - r臀部，3 - l臀部，4 - l膝关节，5 - l踝关节，6 - 骨盆，7 - 胸部，8 - 上颈部，9 - 头顶部，10 - r手腕，11 - r肘，12 - 肩膀，13 - 肩膀，14 - l肘，15 - l手腕）
      - `is_visible` - 联合可见性
  - `.vidx` - 视频索引 `video_list`
  - `.frame_sec` - 视频中的图像位置，以秒为单位
- `img_train(imgidx)` - 培训/测试图像分配
- `single_person(imgidx)` -包含矩形ID `ridx` 的*充分分离*的个体
- `act(imgidx)` - 图像的活动/类别标签 `imgidx`
  - `act_name` - 活动名称
  - `cat_name` - 分类名称
  - `act_id` - 活动ID
- `video_list(videoidx)` - 指定YouTube提供的视频ID。要在youtube上观看视频，请访问https://www.youtube.com/watch?v=video_list(videoidx)

是2D的，skeleton标注了三维位置和骨骼旋转，三维位置是skeleton通过PMP方法，得到了 camera pose, scale and 3D joint positions as a combination of PCA basis that is constructed fromMocap database.

骨骼旋转是From the resulting 3D joint positions, rotations of bones are obtained based on a method which is conceptually similar non-rigid surface deformation techniques。具体方法：Specifically, the skeleton in the rest shape is fitted to the PMP result by balancing the rigidity of bones, the smoothness between bone rotations and the position constraints to attract the skeleton to them.

通过这两步计算得到的标注不一定准确，因此选了一个人工验证的方法，看标注与真实姿势的差距（如何看到差距，可视化来得到差距），如果得到的姿势与真实的全局姿势超过30度，那么标为不可接受，抛弃不用来训练

个人总结

精度

1、靠着多种标记进行全监督。

2、RESNET**粗略**得到3维坐标和骨旋转向量R，然后3D投影到2D，得到热图，利用hourglass计算交叉热图**细化**精度

训练集得到

通过一个粗略的PMP计算得到3维坐标，然后人工去筛选合格的训练集。

用了两个数据集去训练（HM3.6(缺点，主题较少)，MPII（缺点：lable没有3D信息，计算标注，人工去除不合格））