

《人体姿态估计与分析在人机交互中的应用》 进度总结 （3月5日）

目前已完成任务

- 了解 OpenPose模型 的原理以及结构，并能基于CMU训练好的两个 OpenPose模型
 - 一个基于COCO数据集，精度高，位点更多，速度稍慢
 - 一个基于MPI数据集，速度快，精度略低
- 基于 OpenPose模型 ，开始了跌倒检测的应用开发，分别尝试使用两种方法进行跌倒行为的判断：
 - 基于学习：找到了一个UR的跌倒检测数据集和Fall Detection Dataset数据集，利用普通的DNN进行训练
 - 基于规则：自己定义了一套摔倒检测的算法，简单测试后发现在背景不复杂环境下效果不错（图像帧中只存在一个人）

遇到的问题

跌倒检测中的问题

基于学习的方法

在基于学习的方法中，采用何种训练网络进行训练是我遇到的第一个问题。我之前的设计是采用普通的DNN网络，网络的输入为身体各位点归一化（在图像中的百分比*100）后的坐标，对于不能检测到的点，坐标设为了负数。但是这样训练的结果效果一般，在测试集上准确率为70%。

导致这一结果的因素可能有一下几个：

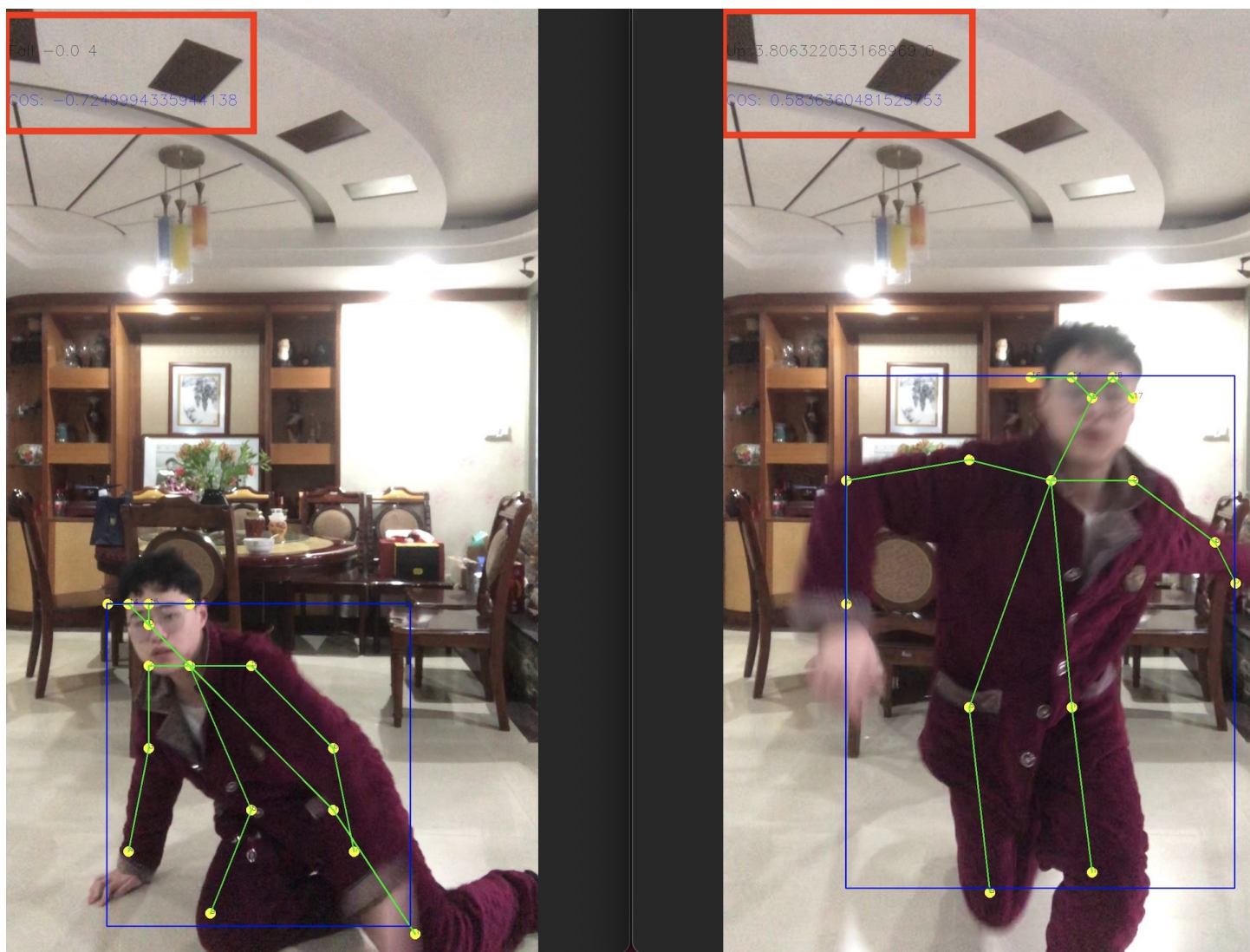
- OpenPose 参数没调整好，包括阈值等参数。
- 训练集本身的问题，由于训练集我一开始用的是Fall Detection Dataset数据集，该数据集分辨率过小（320*280），因此基于该训练集训练对于图中人的一些姿势难以估计准确，出现了人没有被检测到而旁边的沙发被错误检测到的情况，如下图：



- 训练模型的问题。由于训练时只采用了普通的DNN，且DNN的输入为十几个身体位点的坐标，这样的模型不能很好地体现姿态之间的关联的特征，因此可能训练的模型需要调整。

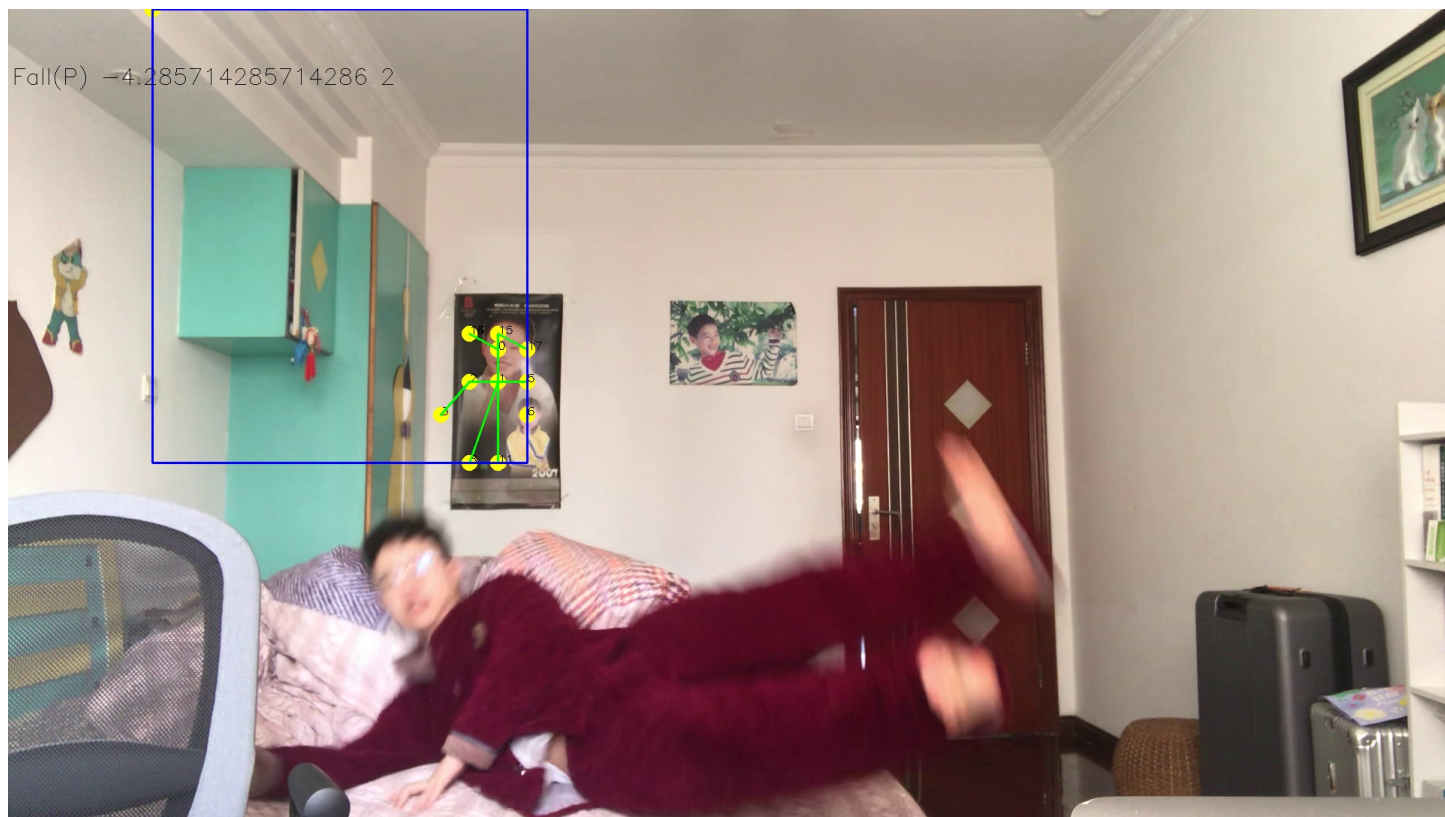
基于规则的方法

在基于规则的方法中，经过测试，在单人情况下检测良好，可以判断出摔倒以及站起，如下图：



但在背景复杂的环境下检测就会遇到问题。问题出现在我自己录制的一段摔倒视频中。

该视频在我的房间里拍摄，背景在墙上有我本人的照片。一开始我用CMU基于 **MPI数据集** 训练出的 **caffe**模型进行姿态估计。该模型得出的效果并不好，经常有很多我身上的位点识别不出来，因此我改用了基于 **coco数据集** 训练出的**caffe**模型进行姿态估计。这个模型得出的精度较高，但是会出现对于同一位点，墙上照片位点的置信度高于移动时我身上的置信度（移动时有些帧我会变得模糊），这会影响基于规则的摔倒检测的判断。问题检测如下图：



摔倒检测的应用场景是针对独居老人或儿童，因此需要在复杂场景下进行单人检测。

对于此问题，我尝试进行多人检测，但是多人检测后的 **单人目标跟踪** 又会产生新的问题。

第二个方法是利用背景消去法，对于这一方法我还需要进一步的学习和咨询。我采取两种背景消除法：

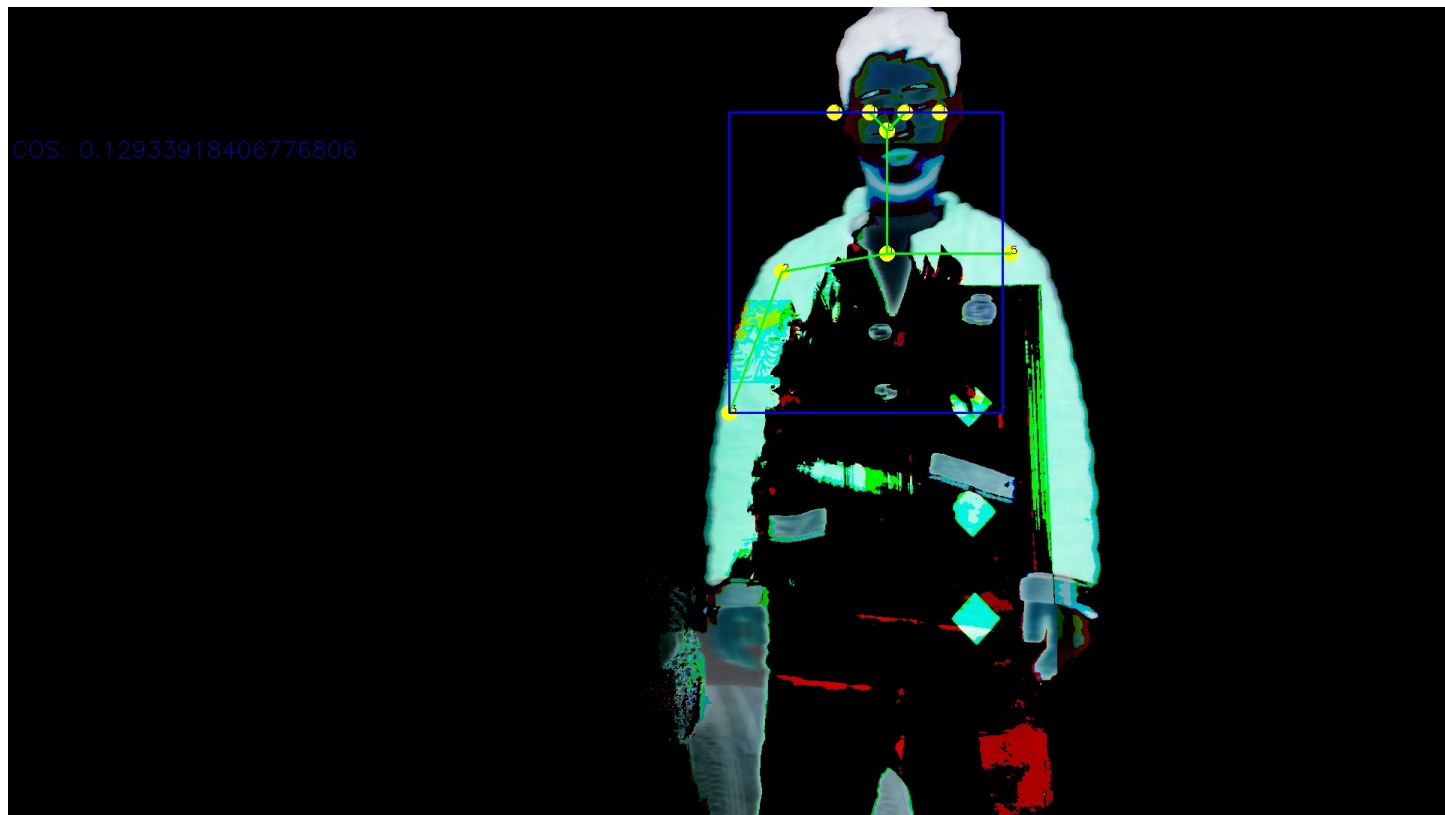
- 利用OpenCV的 **BackgroundSubtractorMOG2** 函数进行背景减除，将得到的掩码呈上原图进行处理，再进行姿态估计。但这样的处理可能会导致下图这种情况，导致无法正确进行姿态估计：



- 利用平均背景法得到平均背景，再进行减除，利用图像差的二值掩码乘上原图。这种方法能得到很好的背景信息，但是后续处理还是遇到了问题，需要进一步学习的请教。得到的平均背景效果图如下：



但减除后效果如下：



此处应该是我对于减除后图像的二值化输出结果理解有误，需要进行调整。