

简介

商汤的新作，最牛逼的一点是开源了新的3D数据集。本文主要工作：

- 利用多角度的2D pose构造了一个3D label generator，并且利用geometric searching来refine这些3D关键点。
- 建立了一个大型的数据集，有400,000张in-the-wild数据和3D kp gt
- 在这些数据上训练了一个baseline模型

作者用unity toolbox生成了超过2500种活动的2/3D pose，远超过Human3.6m当中的15种。然后在这些数据的基础上训练了一个2D-to-3D model。

模型

模型主要有三个模块：

- stereoscopic view synthesis subnetwork 这个模块的作用主要是从一张2D的pose生成另一个视角的2D视图。
- 3D pose reconstruction subnetwork: 利用多角度的2D pose数据估计其coarse 3d pose。
- Geometric search scheme: 对最后的3D pose进行refine

这个模型最大的好处在于其输入是2D pose，也就是说，可以用生成数据的pose（比如unity），只要这些生成的数据符合人体的解剖学结构即可。其中的geometric search的主要原理也是3D pose可以投影回2D和对应的gt比较。

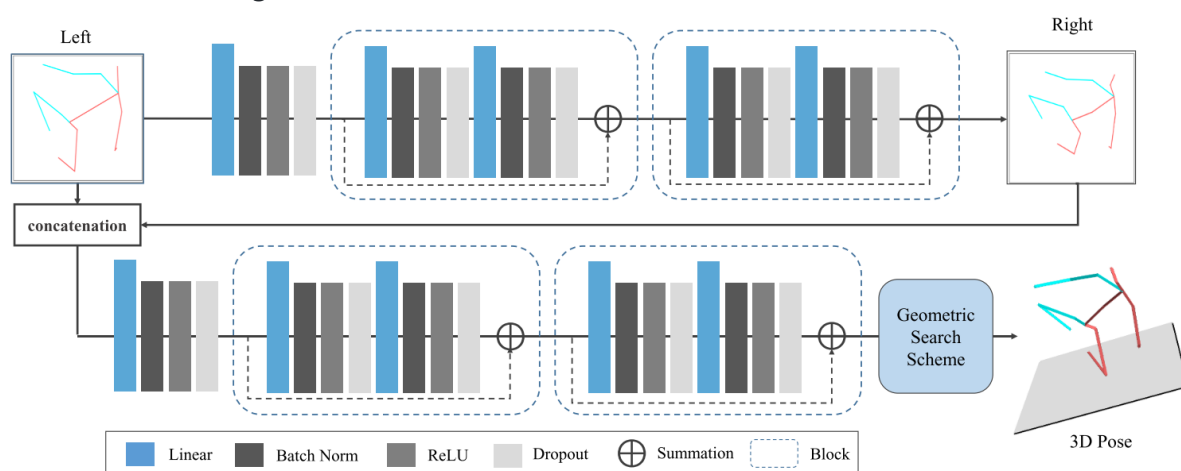


Figure 2. Architecture of the 3D label generator. The generator consists of stereoscopic view synthesis subnetwork, 3D pose reconstruction subnetwork, and a geometric search scheme. Given the 2D pose from the left viewpoint, stereoscopic view synthetic subnetwork aims to generate the 2D pose from the right viewpoint. 3D pose reconstruction subnetwork utilizes the multi-view 2D poses to estimate a coarse 3D human pose. Geometric search scheme is applied to further refine the predicted 3D human pose.

stereoscopic view synthesis subnetwork

主要是输入的预处理，从一个left-view的kp和3D kp得到另一个right-view的kp。方法就是讲3D kp沿着x轴往右移动一些（500mm），保持y和z轴坐标不变。然后根据相机参数再讲kp re-map到

2D。

其pred模型是通过几个linear + BN + ReLU + Dropout block然后Max-norm constraint和残差连接组成的。

3D Pose Reconstruction Subnetwork

模型的结构和上一个相同。

输入是left-view kp和生成的right-view kp的concat。然后直接回归出3D kp。

Geometric Search Scheme

确保生成的3d kp投影回2D可以做到zero-pix erro

需要注意3D pose预测的结果是根据盆骨位置aligned的。所以模型的3D pose输出其实是依照hip的relative depth。

实验

实验比较复杂，因为首先作者先用Human3.6m + unity训练了自己的3个模块，这样就得到了一个2d-to-3d的模型，然后作者用

- Leeds Sports Pose dataset (LSP)
- MPII human pose dataset (MPII)
- Ai Challenger dataset for 2D human pose estimation (Ai-Challenger)

这几个数据集的2D图像通过模型生成了400,000(320,000训练8000测试)张in the wild的3D label。

所以实验分为两个方面，一是验证作者的2 view 2d kp concat的这套（3个）子网络是否work，另一个就是观察这个数据集的一些qualitive和quantitive的效果。