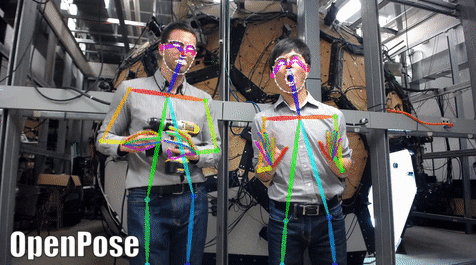
19.6.25周报 张天意

Neural Body Fitting: Unifying Deep Learning and Model Based Human Pose and Shape Estimation（3DV 2018）

目前已经有一些成功的工作：生成人体关键点，棒状表示模型（个人感觉就是指openpose）



本文作者提出更具挑战性的任务：estimating the parameters of a detailed statistical human body model from a single image

**传统方法**需要一个差不多初始化模型，然后把初值优化到最终结果（不需要3d训练数据——带3d动作标注的图片）

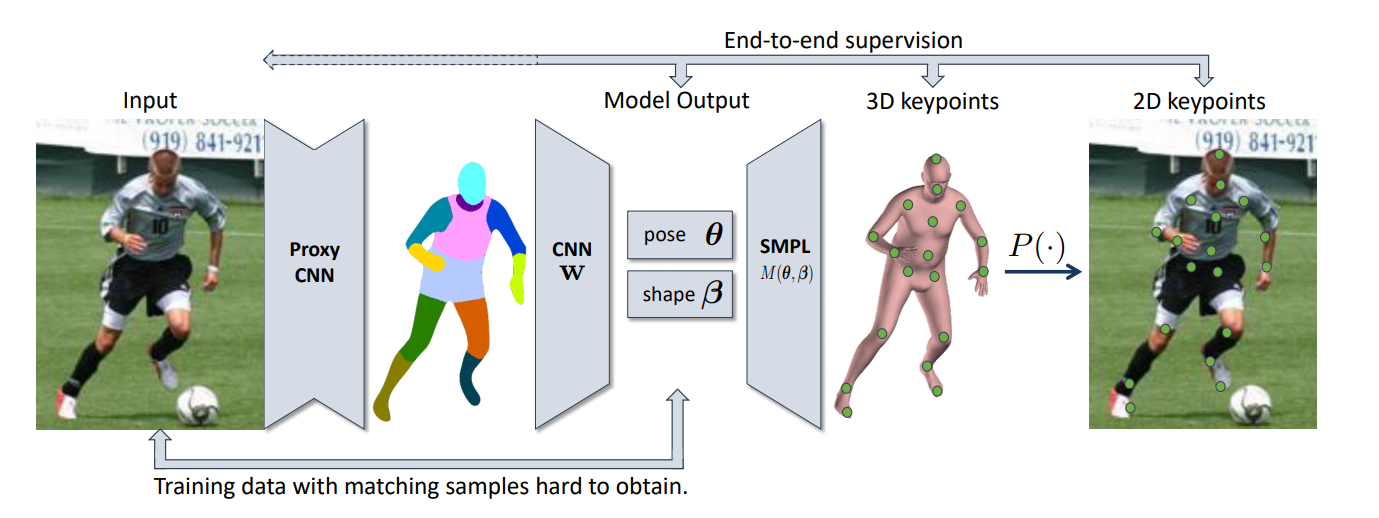
CNN就是forward prediction models，就不需要initialization，但是需要3d姿态标注，不像2d标注好获得

他们近期的工作通过把重建出的模型投影回2d空间更新损失函数，就可以使用2d标注了

本文的\*\*目的\*\*：To analyze the importance of such components

components: image--(CNN,3d notation trained)-->smpl model(hybird params)-->image--(reproject)-->2d notation for CNN training

要形成闭环（loop）



NBF = 一个包含统计身体模型的CNN

两种监督模式：full 3d supervision和weak 2d sup，bottom-up top-down的方法，使得NBF既不需要初始化模型也不需要3d标注的训练数据

因为光照、衣服、杂乱的背景都不想要，专注于pose和shape，所以用处理后的image代替原始rgb image

结论：

1. 12-body-part的分割就包含了足够的shape和pose信息

2. 这种处理后图像的方法比起用原图，效果有竞争力，更简单，训练数据利用率更高

3. 分割质量可以很大程度上预测三维重建（fit）的质量

Networks：

Segmentation Network：

RefineNet（基于ResNet-101）

Fitting network：

repurpose a ResNet-50 network pretrained on ImageNet

replace the final pooling layer with a single fully-connected layer that outputs the 10 shape and 216 pose parameters

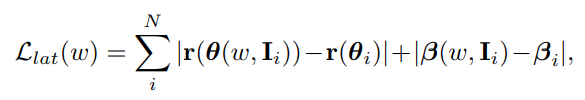
Loss func：

3D latent parameter loss

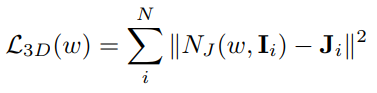
R ：the vectorized rotation matrices of the 24 parts of the body

I ：colour-coded part segmentation map

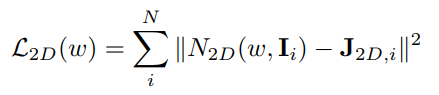
W ：CNN weights



关节点坐标：



投影后的2d关节点坐标

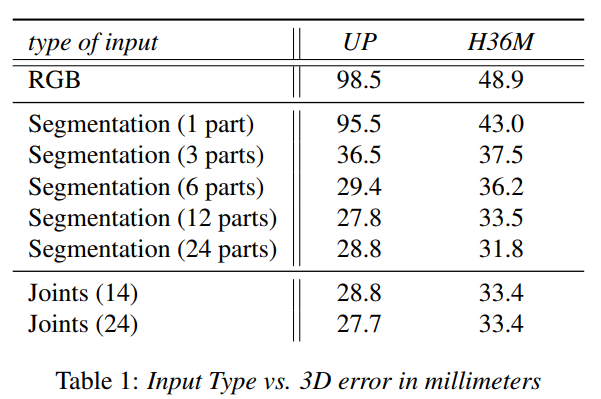


D ：dataset with 2d or 3d notation

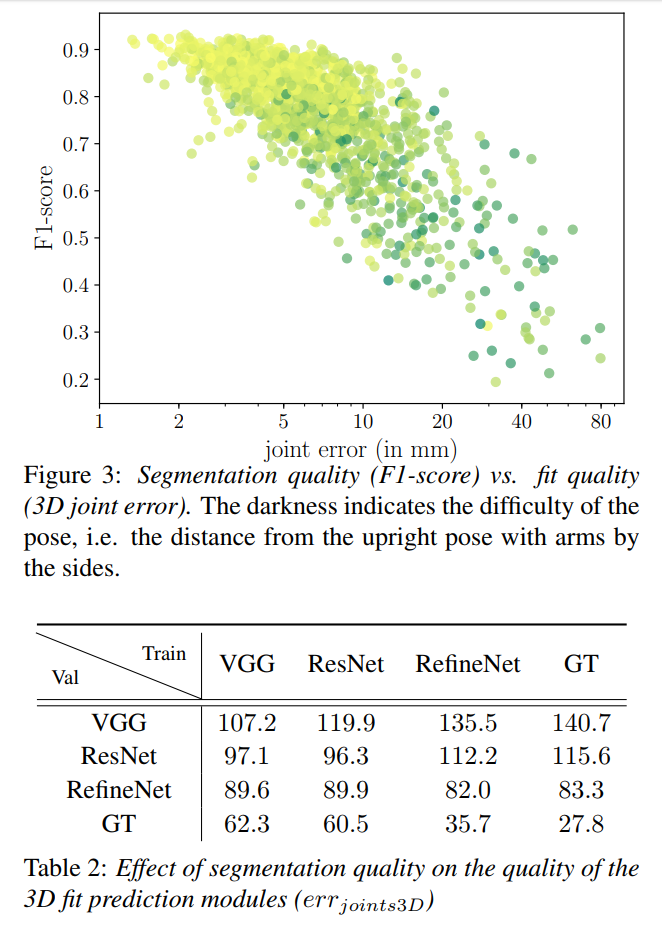


Evaluation & analysis

三个数据集UP-3D，HumanEva-I，Human3.6M



We observe that explicit **part representations** (part segmentations or joint heatmaps) are more useful for 3D shape/pose estimation compared to **RGB images and plain silhouettes**.



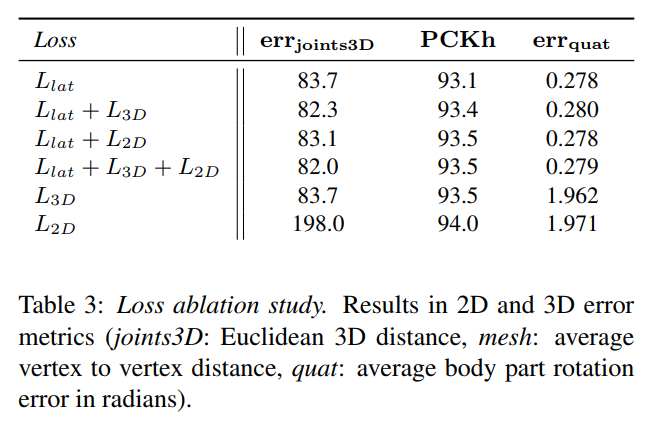
the higher the F-1 score, the lower the 3D joint error.

F-1 score代表了图像分割的质量

**57.0，53.2，67.1，100**

To determine the effect of segmentation quality on the results, we train three different part segmentation networks. Besides RefineNet, we also train two variants of DeepLab [9], based on VGG-16 [52] and ResNet-101 [14]. These networks result in **IoU scores of 67.1, 57.0, and 53.2 respectively on the UP validation set.**

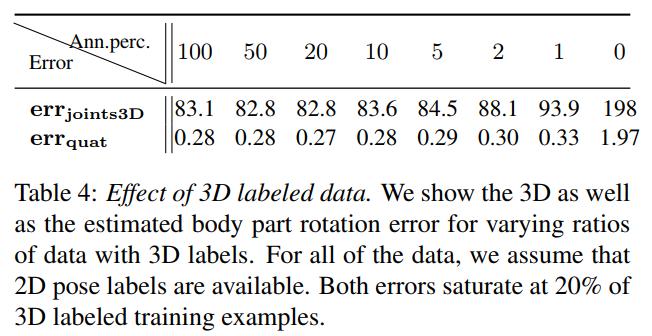
we **then train four 3D prediction networks** - one for each of the part segmentation networks, andan additional one using the ground truth segmentations.



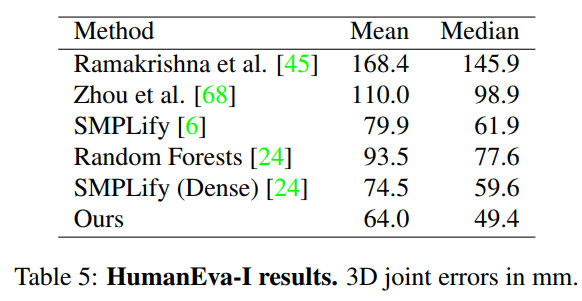
(i) errjoints3D, the Euclidean distance between ground truth and predicted SMPL joints (in mm).

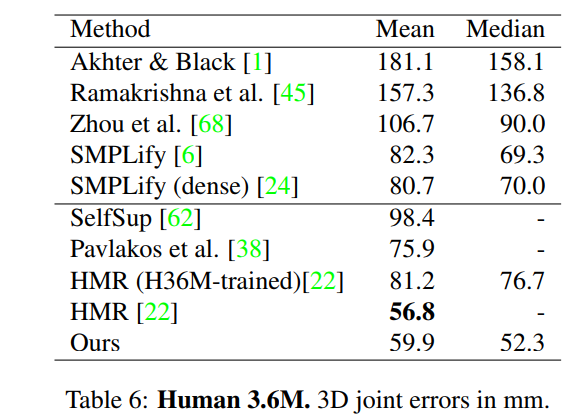
(ii) P CKh [4], the percentage of correct keypoints with the error threshold being 50% of head size, which we measure on a perexample basis.

(iii) errquat, quaternion distance error of the predicted joint rotations (in radians).



需要用多少3d标注





与其他方法的比较