简介

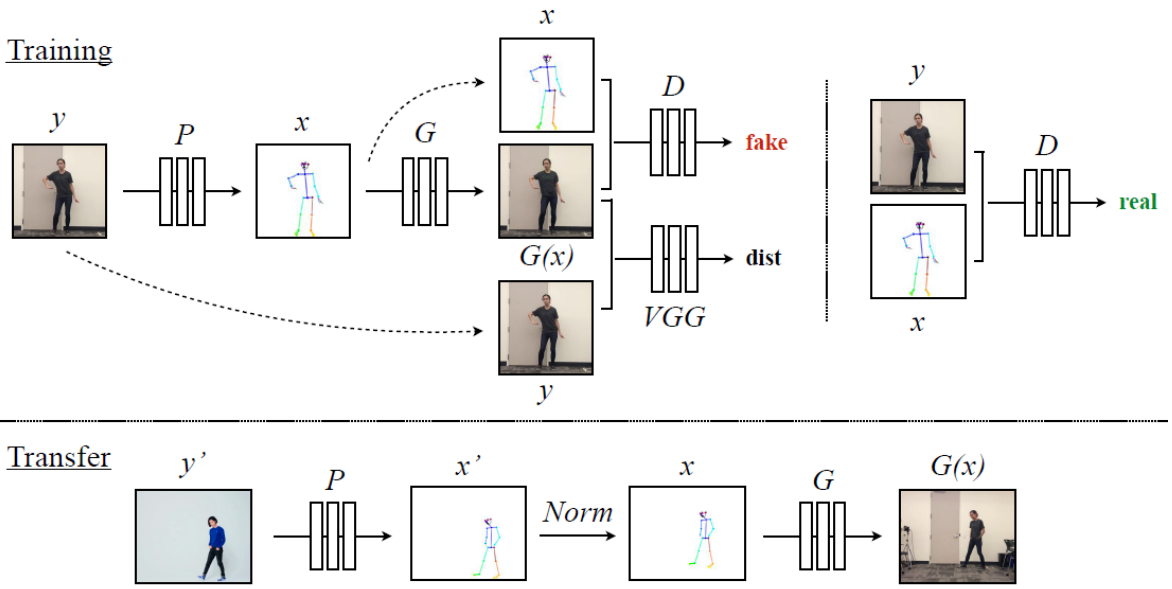
本文主要是通过人体keypoints进行motion transfer形成视频，方法主要框架是GAN，同时考虑了source和targe的keypoint alignment问题、video soomth问题和face synthesis的问题。

首先对于target video，从每一帧提取出pose stick figure和target person image对。这样我们就获取了supervised aligned data。可以监督学习到从keypoint到image的变换。接下来再将source的motion给tranfer过来，就可以实现dance的动作。

为了得到更好的效果，作者还做出了两点改进：为了提高temporal smoothness,作者将每个帧的预测与之前时间步骤的预测进行比较。为了提高脸的真实性在，作者用了专门的GAN训练生成目标人的脸。

方法总览

- pose detection
- global pose normalization
- mapping from normalized pose stick figures to the target subject



训练过程

假设y是从original target video里面提取的frame，用一个姿势检测器提取出对应的pose stick figure $x = P(y)$。在训练阶段用corresponding (x, y) paris来学习到一个映射G，其作用是将将pose stick x映射为真是的图像。这里作者用了adversatial training with discriminator D并且和**a perceptual reconstruction loss dist using a pretrained VGGNet**。D用来分辨G(x)产生的fake image pairs和原始视频里的real image paris。

迁移过程

迁移过程是将source frame y'经过pose detector得到对应的pose stick figure x'。由于和target的figure大小不匹配，因此需要做一个global pose normalization Norm将其对齐得到x。最后将x送入我们已经训练好的G 得到G(x)即是source y'对应的姿势。

姿势估计和归一化

姿势估计

这里和前面讲的一样

Global pose normalization

其实就是在source和target之间做线性变换（scale and translation，缩放和平移）