关于行人生成方法的一些思考

本周看到的方法有PG2，BodyROI7、《Soft-Gated Warping-GAN for Pose-Guided Person Image Synthesis》和《Unsupervised Person Image Synthesis in Arbitrary Poses》。下面是他们的整体框架及对比效果。

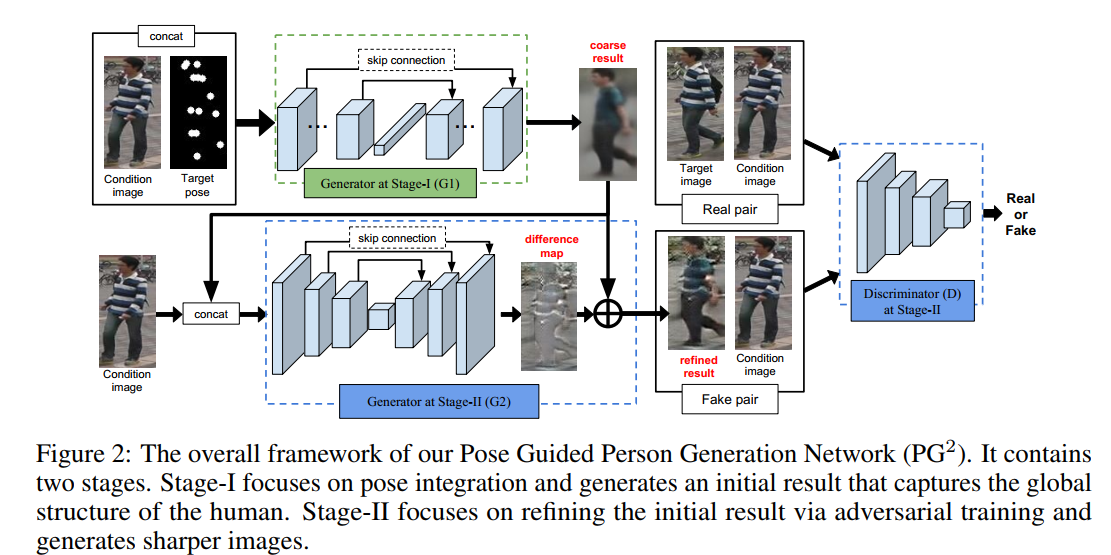


图1 PG2的结构

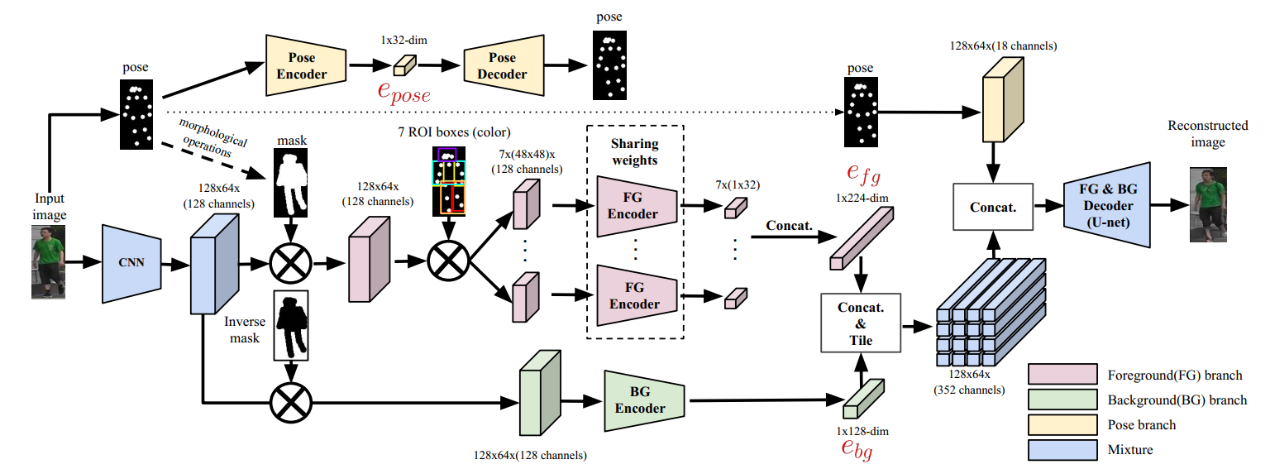


图2 BodyROI7的结构

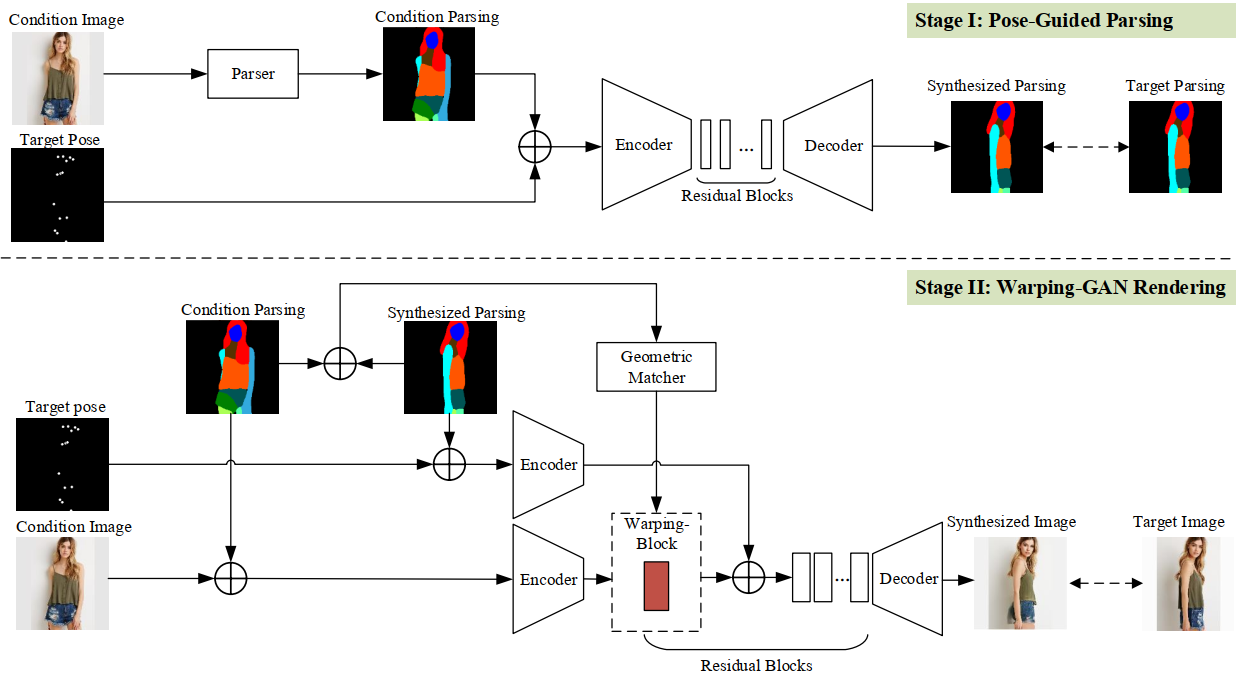


图3 《Soft-Gated Warping-GAN for Pose-Guided Person Image Synthesis》的结构

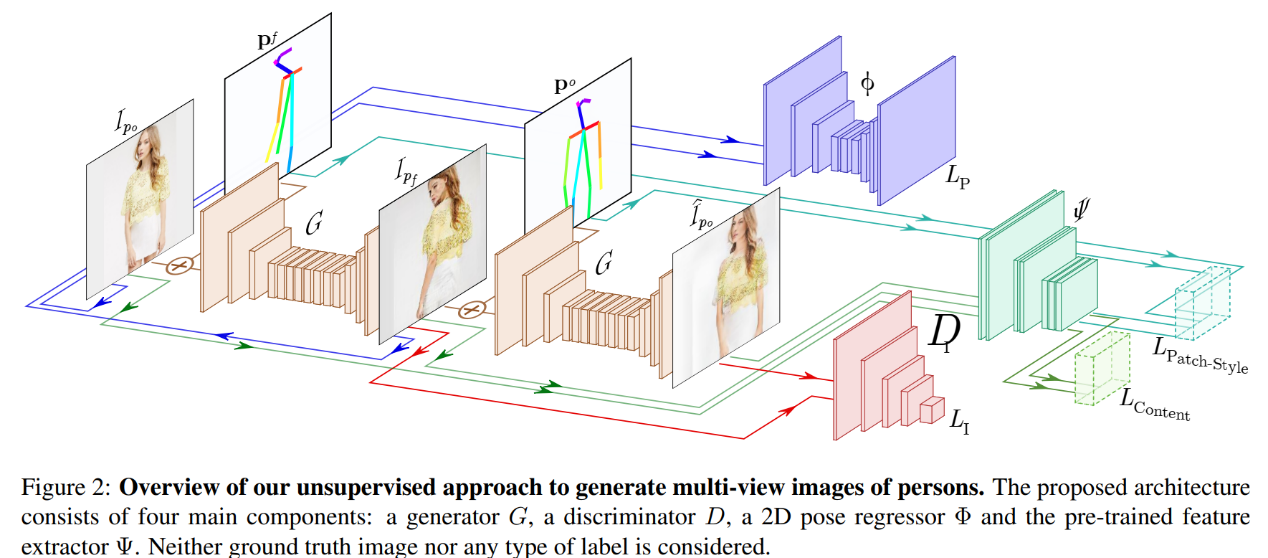


图4 《Unsupervised Person Image Synthesis in Arbitrary Poses》的结构

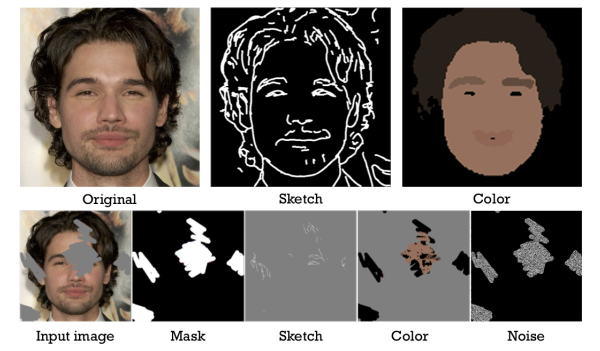
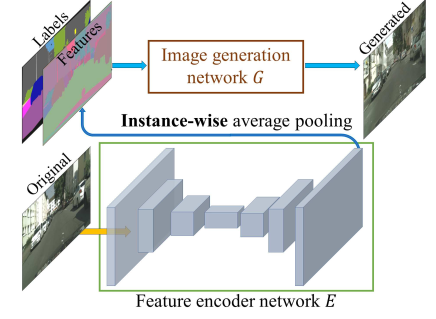




图5 PG2、BodyROI7和Soft-Gated的效果对比图

除了第四篇使用的是CycleGAN的思路以外，PG2、BodyROI7和Soft-Gated方法都是由Pix2Pix方法发展而来的，而其改进思路概括的来说都是增加约束信息辅以能够合理利用这些信息的网络结构，让GAN更容易更准确地生成图像。其中，PG2的方法最为朴素，除了pose以外没补充其他约束信息，而之后的BodyROI7和Soft-Gated则分别补充了【ROI7 boxes、mask、background】信息和【语义分割】信息。

从pix2pixHD加入边缘检测图能够有效部分边缘模糊错误的结果，近年在图像修复任务上SC-FEGAN集成各种先验约束信息的结果，以及中科大古纾旸学长的《Mask Guided Portrait Editing》先训练语义特征自编码器再GAN的结果来看，补充并提取参照图像的特征作为约束信息貌似是近些年图像到图像生成类任务的主流思路。

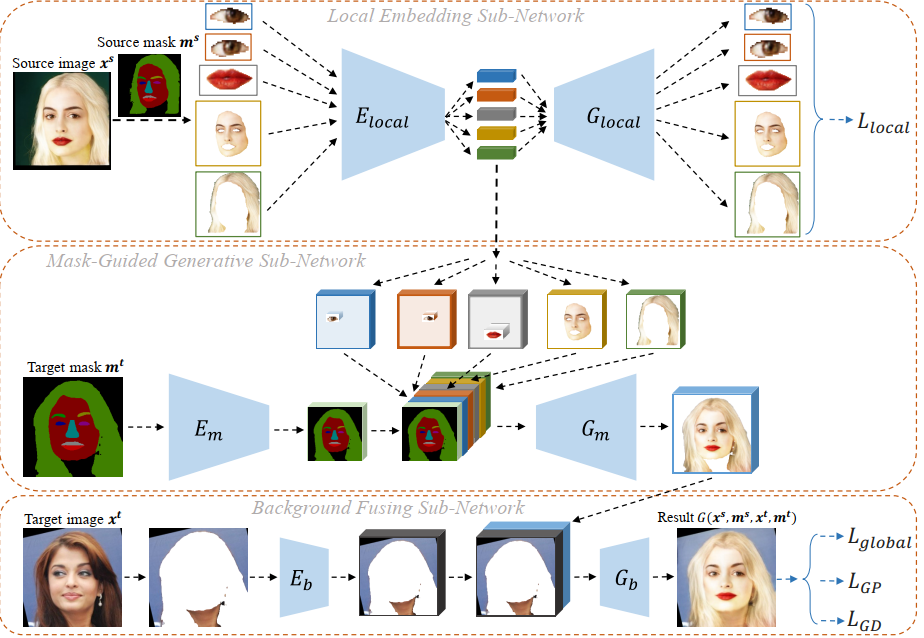
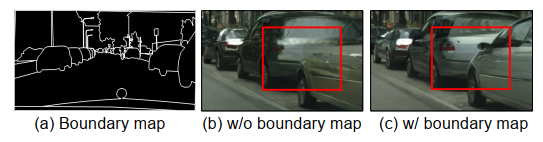
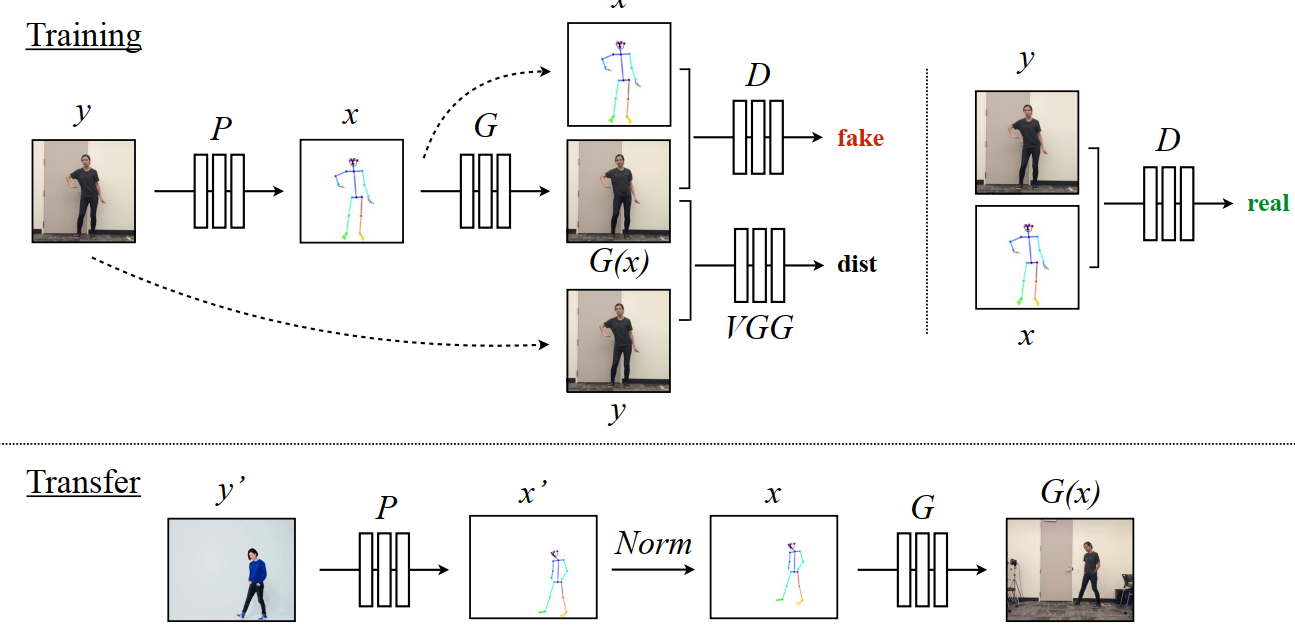


图6 SC-FEGAN、pix2pixHD和Mask Guided Portrait Editing的约束信息添加思路

从以上的例子来看，补充约束信息已经成为了优化GAN效果的较为成熟的通用的套路了，从使用的角度来说，这类方法应该是通用的，只要辅以合适的网络结构和损失函数就应该能有所提升。对于行人图像生成任务应当也是类似的，如果我们从源图像中（Condition Image）提取更多有效特征以提供更多的约束信息，或者分割成多个子任务——比如先从pose/skeleton预测semantic segmentation或boundary map再预测human，那可能效果会更好（当然这只是一种思路）。

但是从另一个角度来看，这类方法有些过于繁琐了，不符合“大道至简”的哲学原理，若是有“集大成者”出现，其瓶颈也是可以预见的。而从《EveryBoody Dance Now》基于姿态识别的视频生成的惊艳效果来看，框架简单并不一定效果不好（虽然这篇文章图像质量高的一个重要原因在于过拟合）。



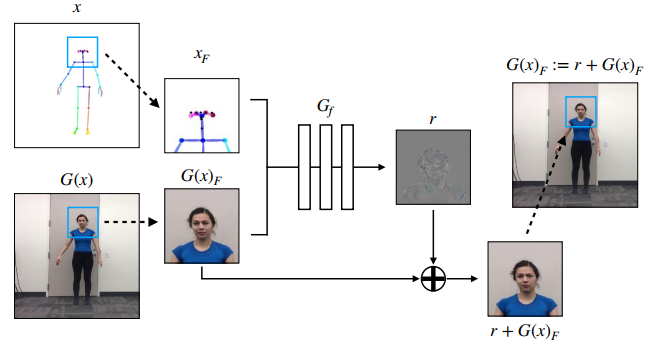
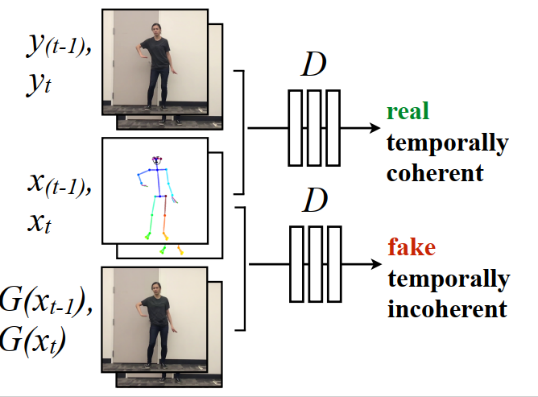
 

图6 EveryBody Dance Now的框架思路

这篇文章给我的启发是：在不过度添加约束信息、不做出划时代性结构变化的情况下，引入其他领域（方向）的方法或许也可以给这类任务带来意想不到的提升。

受去年的科创计划项目书中所提及内容的启发，对行人图像风格进行归一化、使用真实图像超分辨率/去模糊技术对图像进行质量优化与现有生成方式的复合（用于生成前或生成后都可能有价值）或许也能对该类任务有意想不到的提升。当然这些思路也存在很多的问题，归一化的难点在于“如何归一，归一成什么”，超分辨率的难点在于“真实图像的退化原因各不相同，如何泛化，高清样本如何获得”，对于这些具体问题我还没有明确的解决思路。

以上内容就是我上周学习后的一些思考和感想，受限于眼界原因想必存在不少问题，如果有不正确的地方欢迎各位斧正我的错误，谢谢！

冯汉森

2019.4.22