

Unsupervised Person Image Synthesis in Arbitrary Poses

2018/9/13 王诺

本文作者提出了一种使用生成对抗性学习来合成任意姿势、人物的真实感图像的新方法。给定人的输入图像和由 2D 骨架图表示的期望姿势，本文提出的模型可呈现同一人在新姿势下的图像，该图像包含输入图像中可见部分的新颖视图和虚拟出的输入图像中未看到的部分。最近已有人给出了通过监督学习实现该方法的方法，即，在训练期间将新姿势下的真实图像提供给网络。而本文提出了基于无监督学习的方法，将问题分为两个子任务来解决这一具有挑战性的情况。首先，在没有指定姿势的真实图像条件下，将生成器的输出和原始图像对应的 2D 骨架图作为生成器的输入，此时生成器的理想输出图像应为原始图像，因此可以此为目标对生成器进行训练。其次，本文作者设计了新颖、合理的损失函数，其中包含对图像内容和风格的要求，旨在产生高质量的图像。文章中有大量实验数据表明，通过本文提出的模型得到的图像在外观上与通过完全监督的方法获得的图像非常接近。

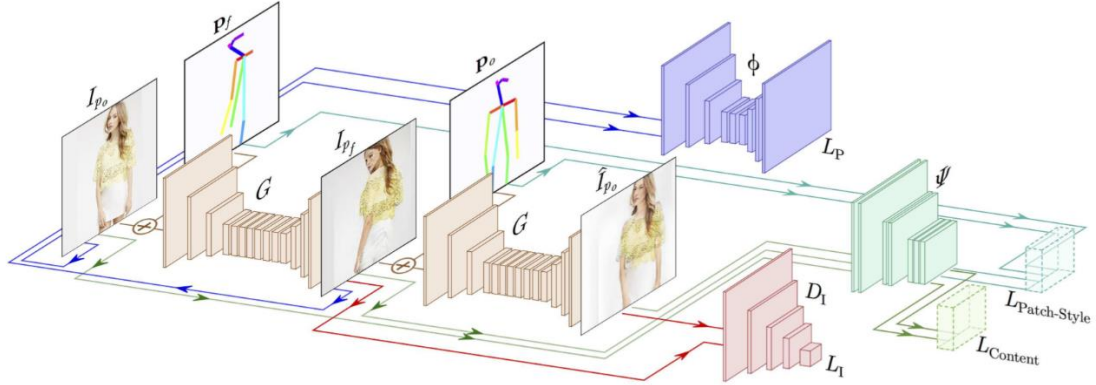


图 1

图 1 展示了本文所提出的算法的基本架构，其中包含一个生成器 G ：生成指定人物、姿势对应的图像；一个鉴别器 D_1 ：鉴别图像为真实图像还是生成的图像，鉴别器的输入为图像经处理得到的 26×26 的矩阵；姿势识别器 ϕ （已预先训练好）：对图像对应的姿势进行判定；特征提取器 ψ （已预先训练好）：含有两个子功能，分别对图像的内容特征和风格特征进行提取。

$$\begin{aligned}\mathcal{L} = & \mathcal{L}_I(G, D_I, I_{p_o}, \mathbf{p}_f) + \lambda_P \mathcal{L}_P(G, D_P, I_{p_f}, \mathbf{p}_f, \mathbf{p}_o) \\ & + \mathcal{L}_I(G, D_I, I_{p_f}) + \lambda_P \mathcal{L}_P(G, D_P, \hat{I}_{p_o}, \mathbf{p}_o) \\ & + \lambda_{Id} \mathcal{L}_{Id} + \lambda_P \mathcal{L}_\Phi(I, \mathbf{p}_o)\end{aligned}$$

图 2

图 2 展示了损失函数的定义，包含了对抗性损失、风格一致性损失、身份一致性损失和内容一致性损失。

Method	SSIM	IS
Our Approach	0.747	2.97
Ma <i>et al.</i> NIPS'2017 [16]	0.762	3.09
Zhao <i>et al.</i> ArXiv'2017 [35]	0.620	3.03
Sohn <i>et al.</i> NIPS'2015 [29]*	0.580	2.35
Mirza <i>et al.</i> ArXiv'2014 [19]*	0.590	2.45

图 3

图 3 从数据方面说明了本文提出的无监督方法可与存在的基于监督学习的方法达到相似的效果。



图 4

图 4 展示了使用本文提出的方法进行处理后的结果图像。