

# 生成穿着定制服装的高分辨率时装模特形象

Gokhan Yildirim Nikolay Jetchev Roland Vollgraf Urs Bergmann  
Zalando Research

fgokhan.yildirim,nikolay.jetchev,roland.vollgraf,urs.bergmann@zalando.de

## 摘要

可视化服装是购物的重要组成部分。在组合时尚物品的方面,可用于生成限于预定一套服装的图像。在本文中,我们通过生成在输入身体姿势下穿着定制服装的时尚模特的高分辨率图像来提升这些视觉效果。我们展示了我们的方法不仅可以将一个生成的服装的风格和姿势转移到另一个,还可以创建人体和服装的逼真图像。

## 1. 介绍

时尚电子商务平台通过搜索和个性化使服装购物变得简单。可以进一步增强用户体验的想法是可视化人体上的服装。以前的研究主要集中在已经存在的时装模特图像上替换服装 [5,2], 或者通过使用姿势和服装颜色作为输入条件从头开始生成低分辨率图像 [8]。在本文中,我们专注于生成穿着所需服装和姿势的时装模特的高分辨率图像。

近年来,生成对抗性网络 (GAN) [1]的进步使得能够通过隐式生成建模对真实图像进行采样。其中一个改进是 Style GAN [7],它建立在使用 Progressive GAN [6]生成高分辨率图像的基础上,通过自适应实例归一化 (AdaIN) [4]对其进行修改。在本文中,我们在两种设置下在模特服装姿势图像的数据集上使用和修改 Style GAN: 我们首先在一组时装模特图像上训练原始 Style GAN, 并表明我们可以将一个生成时装模特的服装颜色和身体姿势转移到另一个。其次,我们修改 Style GAN 以适应服装和人体姿势的生成过程。这使我们能够在不同的身体姿势和类型下快速发挥定制服装的作用。

## 2. 服装数据集

我们使用拥有大约 380K 条件的专有图像数据集。我们的数据集中的每个条目都包含一个时装模特

穿着具有一定身体姿势的衣服。服装由一组最多 6 个物件组成。为了获得身体姿势,我们使用深度姿势估计器提取 16 个关键点 [10]。在图 1 中,我们可以从我们的数据集中可视化一些样本。时装模特上的红色标记代表了提取的关键点。模特和物件图像都具有  $1024 \times 768$  像素的分辨率。



图 1: 来自我们数据集的样本 (红色标记代表提取的关键点)。

## 3. 实验

无条件版本的 Style GAN 的流程图如图 2 (a) 所示。我们有 18 个生成器层,它们接收用于自适应实例规范化的样式向量的仿射变换副本。判别器与原版 Style GAN 完全相同。我们在四个 NVIDIA V100 GPU 上训练这个网络大约四周,训练了 160 个 epoch。

在条件版本中,我们使用编码网络修改 Style GAN,如图 2 (b) 所示。该网络的输入是六个物件图像 (总共 18 个通道) 和一个 16 通道热图图像,它是从 16 个关键点计算出来的。物件图像与固定排序连接,以便跨服装实现语义一致性。我们可以在图 1 中看到这种排序。如果一个服装没有关于特定语义类别的物件,它将填充一个空的灰色图像。编码网络创建 512 维向量,其与潜在向量连接以产生样式向量。该模型也训练了四周 (训练 115 个 epoch)。条件模型中的判别器使用单独的网络来计算输入物件的嵌入和

热图, 然后用于计算最终得分[9]。

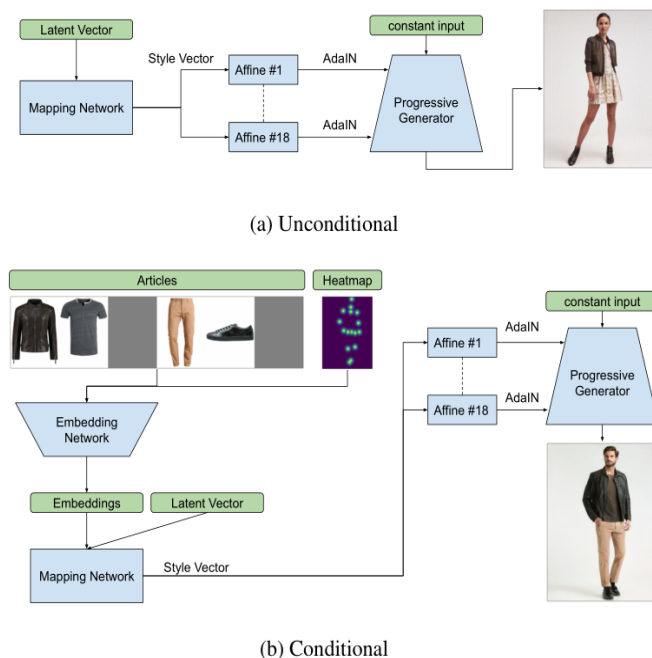


图 2: 我们 (a) 无条件和 (b) 有条件 GAN 的流程图。

### 3.1. 无条件的

在图 3 中, 我们说明了由无条件模型生成的图像。正如我们所看到的, 不仅物品, 而且人体部位在 1024 768 像素的最大分辨率下实际生成。

在训练期间, 可以通过切换某些层的样式向量来使发生器正规化。这具有将信息从一个生成的图像传送到另一个图像的效果。在图 4 中, 我们举例说明了信息传输的两个例子。首先, 我们将相同的源样式向量广播到生成器的第 13 到 18 层 (在图 2 中的仿射变换之前), 它将源服装的颜色传递给目标生成的图像, 如图 4 所示。如果我们复制源样式向量到早期图层, 这将传输源姿势。在表 1 中, 我们显示了我们广播源和目标样式向量的层, 以实现所需的传输效果。

	Color Transfer	Pose Transfer
Source	13-18	1-3
Target	1-12	4-18

表 1: 用于广播样式向量的图层。

### 3.2. 有条件的

在训练我们的条件模型之后, 我们可以输入一组待预览的物件和一个姿势来可视化一个服装在一个

人体上的样子, 如图 5 所示。我们在图 5 (a) 和 (b) 中使用两个不同的服装搭配, 并且四个随机挑选的身体姿势分别在图 5 (c) 和 (d) 中生成模特图像。我们可以观察到物件在生成的身体上正确呈现, 并且姿势在不同的服装中是一致的。在图 5 (e) 中, 我们通过将夹克从第一个服装添加到第二个服装, 使用定制服装可视化生成的图像。我们可以看到牛仔夹克的质地和尺寸在时装模特上正确呈现。请注意, 由于我们的数据集中存在多种相关性, 生成的模特的角度可能会根据服装和姿势而有所不同。

在我们的数据集中, 我们有各种身体类型的时装模特, 这取决于他们的性别, 身材和体重。通过提取的关键点之间的相对差异隐含地表示这种变化。我们的条件模型能够捕获和再现具有不同体型的时装模特, 如图 5 中的第四个生成图像所示。这个结果令人鼓舞, 我们的方法可能会在未来扩展到更广泛的客户通过虚拟试穿应用程序。

### 3.3. 定量结果

我们通过计算无条件和条件 GAN 的 Frechet'初始距离 (FID) 得分[3]来测量生成图像的质量。从表 2 可以看出, 无条件 GAN 产生了更高质量的图像, 可以通过比较图 3 和图 5 来观察。条件判别器还有检查输入服装和姿势是否正确生成的附加任务。这可能导致图像质量 (或“真实性”) 与直接控制生成的服装和姿势的能力之间的权衡。

	FID Score	Training Epochs
Unconditional	5.15	115
Conditional	9.63	115

Table 2: 模型的 FID 分数。

### 4. 结论

在本文中, 我们提出了两种生成时尚模特高分辨率图像的方法。首先, 我们展示了无条件 Style GAN 可以用于通过在特定层交换样式向量来在生成的图像之间传递样式/颜色和姿势。其次, 我们使用编码网络修改了 Style GAN, 这样我们就可以生成穿着定制服装的时装模特的图像。作为未来的工作, 我们计划在更具挑战性的案例中提高条件模型的图像质量和一致性, 例如生成具有复杂纹理和质地的物件。



图 3: 由无条件 Style GAN 生成的模型图像。



图 4: 将服装或身体姿势的颜色转移到不同模特的生成模型。

## 参考文献

- [1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In NIPS. 2014.
- [2] Xintong Han, Zuxuan Wu, Zhe Wu, Ruichi Yu, and Larry S. Davis. Viton: An image-based virtual try-on network. CVPR, 2017.
- [3] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In NIPS. 2017.
- [4] Xun Huang and Serge J. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. ICCV, 2017.
- [5] Nikolay Jetchev and Urs Bergmann. The conditional analogy gan: Swapping fashion articles on people images. In ICCV Workshops, 2017.
- [6] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. ICLR, 2017.
- [7] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. CVPR, 2019.
- [8] Christoph Lassner, Gerard Pons-Moll, and Peter V. Gehler. A generative model of people in clothing. In ICCV, 2017.
- [9] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for gans do actually converge? In ICML, 2018.
- [10] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In CVPR, 2019.



(a) 服装 #1

(b) 服装 #2



(c) 生成的有服装的模特 #1



(d) 生成的有服装的模特 #2



(e) 生成的有服装的模特 #2 并且服装是夹克衫 #1

图 5: 两个不同的服装 (a) 和 (b) 用于生成 (c) 和 (d) 中的模型图像。 (e) 服装 #1 的夹克被添加到服装 #2 以定制可视化。