论文题目

Unsupervised Representation with Deep Convolutional Generative Adversarial Networks [[PDF]](https://arxiv.org/pdf/1511.06434.pdf) (将GAN于CNN结合起来)

解决问题

本文将CNN扩展到了无监督学习领域，并且由于GAN训练中存在1、不稳定，生成器最后经常产生无意义的输出 2、黑箱模型，多层GAN的中间层没有办法可视化。论文中对模型做了一系列修改和限制，使得训练变得更稳定，并且提出了DCGANs，可视化了中间的卷积层。

本文的最大贡献：

1. 将CNN与GAN结合在一起提出了DCGANs，使用DCGANs从大量的无标记数据（图像、语音）学习到有用的特征，相当于利用无标记数据初始化DCGANs的生成器和判别器的参数，在用于有监督场景，比如，图像分类。
2. 表示学习representation learning的工作：尝试理解和可视化GAN是如何工作的，多层的GAN的中间表示intermediate representation 是什么。
3. 给出了一些稳定训练DCGANs的guidelines。

创新

- the all convolutional net . 将 deteministic spatial pooling function （such as: maxpooling）with strided convolutions，使得网络可以学习其自己的 spatial downsampling。我们利用这种方法到我们的generator 当中，允许其学习自己的 spatial upsampling，and discriminator 。
- the trend towards eliminating fully connected layers on top of convolutional features. 作者发现：global average pooling 增强了模型的稳定性，但是损害了收敛的速度。A middle ground of directly connecting the highest convolutional features to the input and output respectively of the generatively of the generator and discriminator worked well. 具体网络如下图所示:

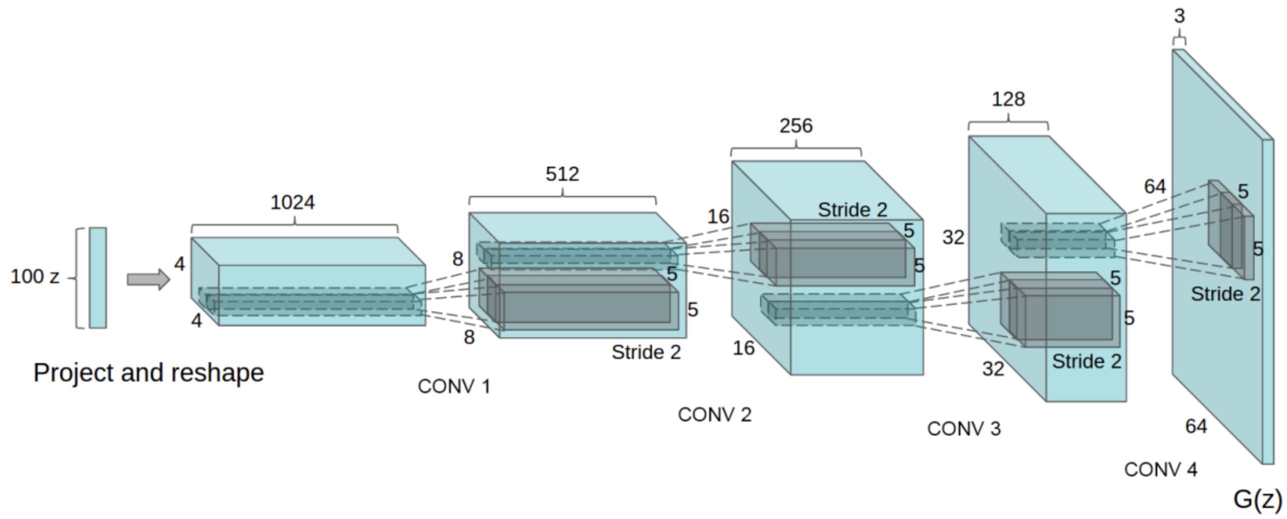


Figure 1: DCGAN generator used for LSUN scene modeling. A 100 dimensional uniform distribution Z is projected to a small spatial extent convolutional representation with many feature maps. A series of four fractionally-strided convolutions (in some recent papers, these are wrongly called deconvolutions) then convert this high level representation into a 64×64 pixel image. Notably, no fully connected or pooling layers are used.

- Batch Normalizaiton，which stabilizes learning by normalizing the input to each unit to have zero mean and unit variance。但是，直接对所有的 layer 都使用这种技术，就会出现问题：resulted in sample oscillation and model instability 。这种困难是通过 不对 generator output layer 和 discriminator input layer 采用这种方法，就行了。

实验细节

者在三个数据集上进行了训练，分别是：Large-scale Scene Understanding （LSUN），Image Net-1k and Faces dataset 。实验结果如下：

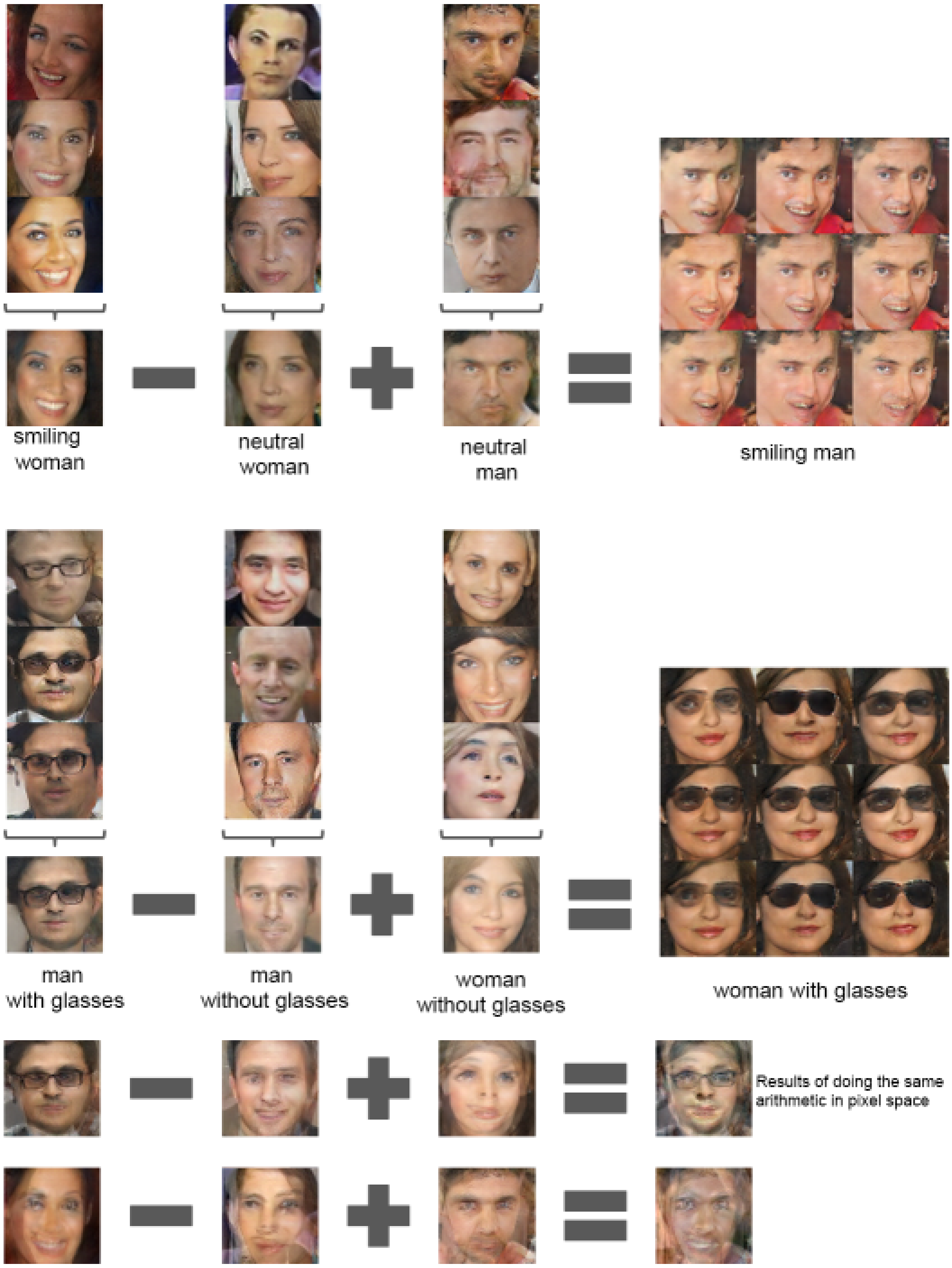
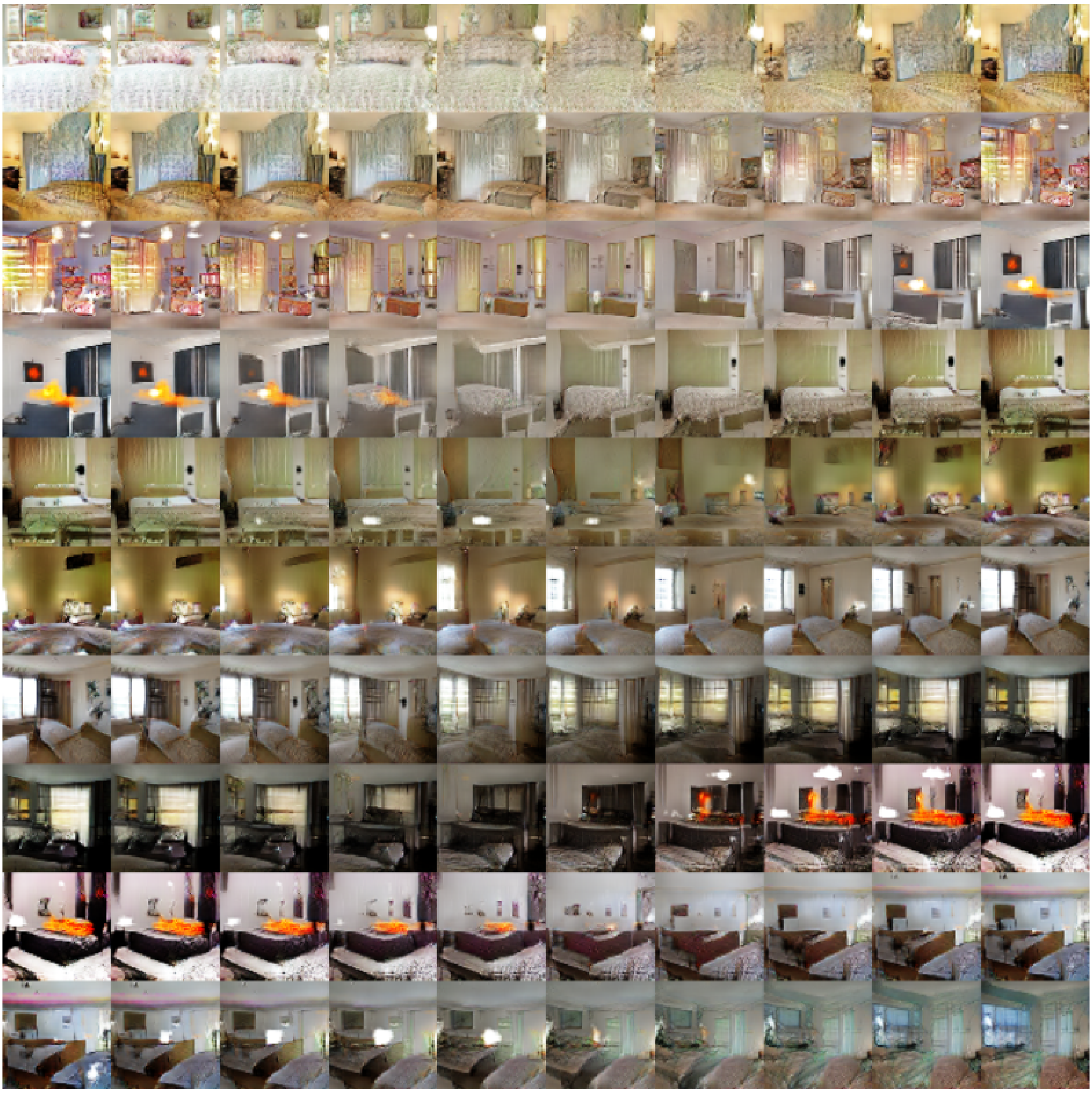


Figure 7: Vector arithmetic for visual concepts. For each column, the Z vectors of samples are averaged. Arithmetic was then performed on the mean vectors creating a new vector Y. The center sample on the right hand side is produce by feeding Y as input to the generator. To demonstrate the interpolation capabilities of the generator, uniform noise sampled with scale +-0.25 was added to Y to produce the 8 other samples. Applying arithmetic in the input space (bottom two examples) results in noisy overlap due to misalignment.



总结

作者在训练生成对抗网络上提出了一套更稳定的架构，以及给出足够的证据表明在监督学习和生成模型上对抗网络可以为图像学习到很好的特征表示。但仍然存在一些形式的模型不稳定性：随着模型训练时间的增长，有一些filter会趋向于不稳定。个人感觉这篇论文看的不是很懂，可能还是要根据代码来学。