

论文题目：UNSUPERVISED REPRESENTATION

LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS ----from ICLR 2016

一、创新点

- 1、作者提出并评估了一系列卷积 GAN 体系结构拓扑上的约束条件，这些约束条件使得它们在大多数情况下可以稳定地训练。作者将这种架构称为 Deep Convolutional GANs (DCGAN)。
- 2、使用训练过的鉴别器进行图像分类任务，与其他无监督算法一起显示出竞争优势。
- 3、可视化由 GAN 学习到的过滤器，并凭经验显示特定的过滤器已经学会了绘制特定的对象。
- 4、展示了生成器具有有趣的向量属性，可以容易地处理生成的样本的许多语义质量。

二、DCGAN

作者将 CNN 应用在 GAN 网络上，基于对 CNN 以下改进：

- 1、用步 strided convolutions (discriminator) 和 fractional-strided convolution (generator) 替代 pooling 层。对于判别模型容许网络学习自己的空间下采样，对于生成模型容许学习自己空间上采样。
- 2、在 generator 和 discriminator 上都使用 batchnorm
 - (1) 解决初始化差问题
 - (2) 帮助梯度传播到每一层
 - (3) 防止 generator 把所有样本收敛在单个点
- 3、在 CNN 中移除全连接层

4、在 generator 除了输出层外的所有层使用 ReLU 激活函数，输出层采用 tanh

5、在 discriminator 的所有层上使用 LeakyReLU

具体如下图所示：

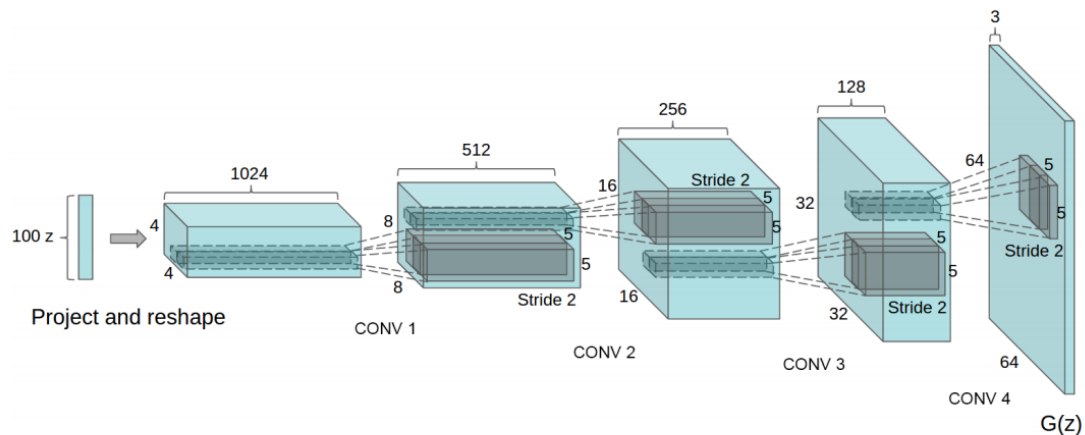


Figure 1: DCGAN generator used for LSUN scene modeling. A 100 dimensional uniform distribution Z is projected to a small spatial extent convolutional representation with many feature maps. A series of four fractionally-strided convolutions (in some recent papers, these are wrongly called deconvolutions)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

三、实验

实验环境：

- Python 2.7 or Python 3.3+（我使用的是 python3.5）
- Tensorflow 0.12.1
- SciPy
- pillow

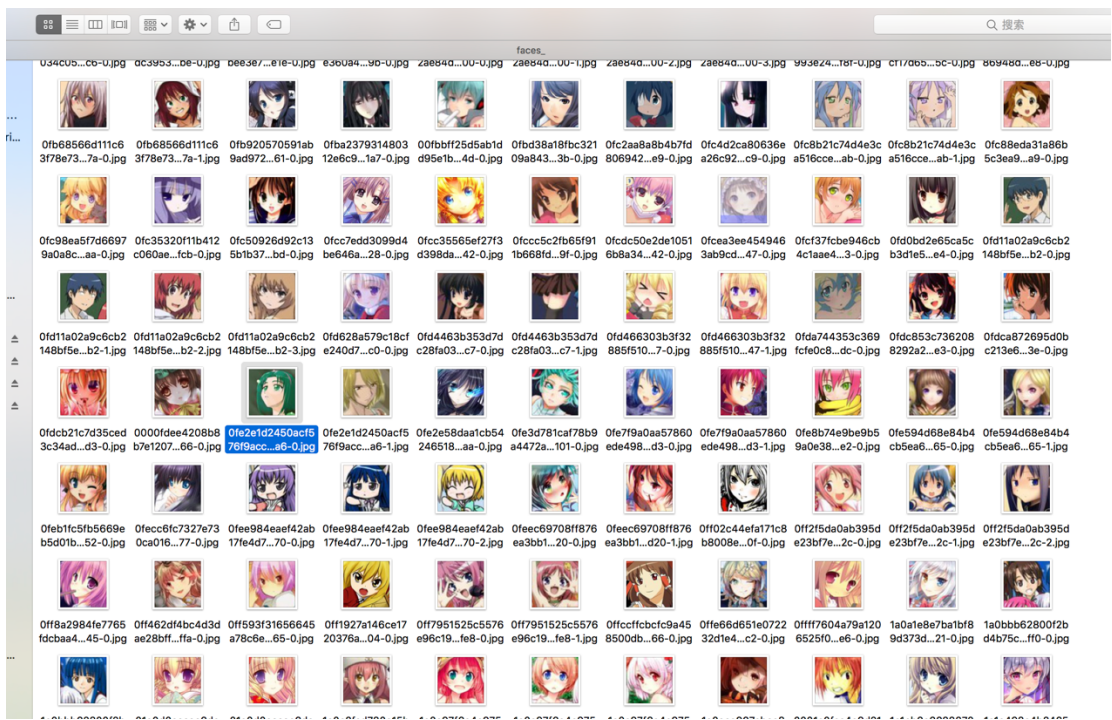
DCGAN 网络由 generator（生成器）和 discrimination（判别器）组成。在代码中，Generator（生成器）通过反卷积网络将特征向量 reshape 成 $64 \times 64 \times 3$ 的特征图最后使用 tanh 激活函数输出。discrimination(判别器)通过四层卷积神经网络（除输出层所有层都使用 relu 激活函数）输出判别结果。训练模型中通过随机梯度下降法（SGD）不断迭代降低 loss，每 100 个批次输出图片结果。在网络

上找了动漫图片作为训练集，通过 25 次迭代训练（通过 GAN 网络生成的图像）如下：



左图为迭代一次的结果，右图为迭代第 25 次的结果

训练之后的效果还是不错的。训练集如下：



作者给出的对比实验

Table 1: CIFAR-10 classification results using our pre-trained model. Our DCGAN is not pre-trained on CIFAR-10, but on Imagenet-1k, and the features are used to classify CIFAR-10 images.

Model	Accuracy	Accuracy (400 per class)	max # of features units
1 Layer K-means	80.6%	63.7% ($\pm 0.7\%$)	4800
3 Layer K-means Learned RF	82.0%	70.7% ($\pm 0.7\%$)	3200
View Invariant K-means	81.9%	72.6% ($\pm 0.7\%$)	6400
Exemplar CNN	84.3%	77.4% ($\pm 0.2\%$)	1024
DCGAN (ours) + L2-SVM	82.8%	73.8% ($\pm 0.4\%$)	512

Table 2: SVHN classification with 1000 labels

Model	error rate
KNN	77.93%
TSVM	66.55%
M1+KNN	65.63%
M1+TSVM	54.33%
M1+M2	36.02%
SWWAE without dropout	27.83%
SWWAE with dropout	23.56%
DCGAN (ours) + L2-SVM	22.48%
Supervised CNN with the same architecture	28.87% (validation)

四、总结

作者提出了一套更稳定的架构来训练生成对抗网络，并且给出对抗网络学习监督学习和生成建模的良好图像表示的证据。作者在训练过程中也提出自己遇到的问题未解决：有一些形式的模型不稳定-随着模型的训练时间更长，他们有时会把一个子集的滤波器转换成一个单一振荡模式。

改进点：

- 1、解决训练过程中模型不稳定的问题。
- 2、可以将这个框架扩展到视频（用于帧预测）和音频（用于语音合成的预先训练的特征）等其他领域。
- 3、对学习的潜在空间的性质的进一步研究。