参看：

论文：[Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks](https://arxiv.org/abs/1511.06434)

网站：<http://blog.csdn.net/solomon1558/article/details/52573596>

DCGAN是一种无监督学习的网络结构，通过DCGAN得到好的图像表示之后，可将图像表示结果以及GAN的特征提取结构应用到有监督学习当中。传统的GAN训练不够稳定，输出的图片往往没有意义。DCGAN就对上述情况做了改进。

1. DCGAN的贡献
2. 对网络结构做了约束，使得训练过程更加稳定，生成更高分辨率的图像，称这类结构为DCGAN；
3. 用训练好的判别模型进行图片分类任务，取得了与其他无监督算法相当的结果；
4. 将GAN的filters进行了可视化处理，表示特定filter要学习画特定的目标；
5. 生成模型具有向量算数的运算性能；
6. 网络结构的改进
7. 全卷积网络(all convolutional nets)

判别模型：使用带步长的卷机网络(strided convolutions)代替池化层，是网络能够学习自己的空间下采样；

生成模型：使用微步幅卷积(fractional strided)，使网络学习自己的空间上采样。

1. 在卷积特征上消除全连接层

去掉卷积层后的全连接层，用全局平均池化曾代替(global average pooling)，这种方法增加了模型稳定性，但是损害了收敛速度。生成模型的输入是100维服从均匀分布的噪声向量，通过全连接层投影到1024维向量，再reshape成1024个通道的4\*4 feature map，作为卷积层的输入，经过四个卷积层得到一个64\*64\*3的图片输出。在这过程中没有池化层和全连接层的介入。

生成模型的激活函数：输出层使用tanh激活，其他层使用ReLU激活，如此使模型收敛速度更快；

判别模型的激活函数：leaky ReLU激活函数；

1. Batch Normalization(0均值1方差)

解决因糟糕的初始化引起的训练问题，使梯度能传播到更深的层次。

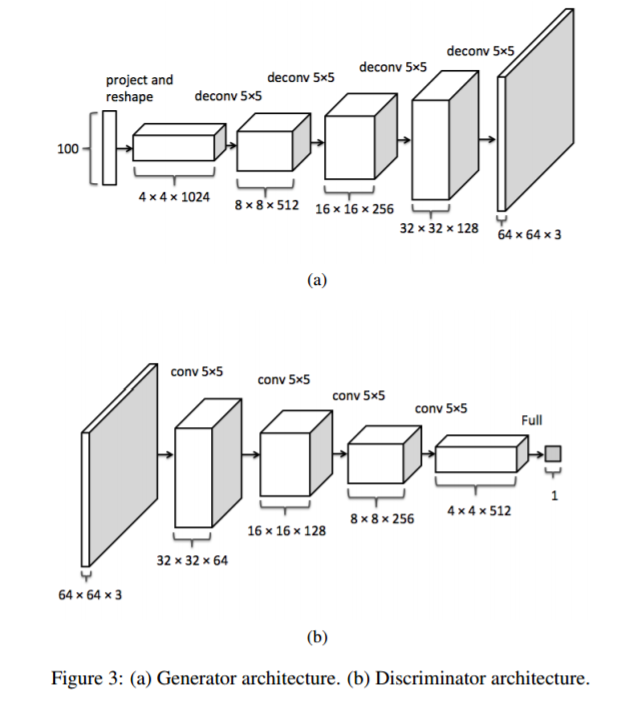
生成模型：除输出层外的其他层都要做batchnorm；

判别模型：除输入层外提他层都要做batchnorm；

1. 稳定的DCGAN结构描述

* 对于生成模型，用fractional-strided convolutions替代池化层，对于判别模型，用strided convolutions替代池化层；
* 对生成模型(除输出层)和判别模型(除输入层)的每一层做batchnorm；
* 去掉所有全连接层；
* 生成模型除最后一层外，每一层用ReLU做激活，最后一层用Tanh做激活；
* 判别模型的所有层做leakyReLU激活；

生成模型和判别模型的结构如下：



这个下采样卷积计算是不是有问题？stride是2，（64-5）是个奇数，不管怎么加padding都没法整除2？

1. 实验

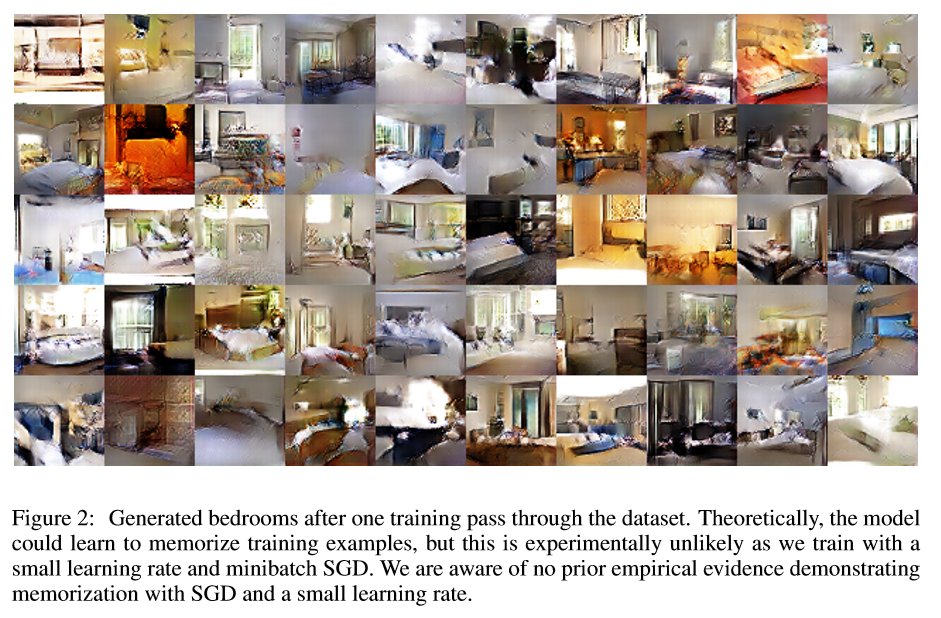
三种数据库：Large-scale Scene Understanding(LSUN)，Imagenet-1K，Faces dataset；

参数设定(要做实验还原，所以关注一下细节)：训练图片无需预处理，只是tanh激活函数会将特征限定到[-1,1]之间；Mini-batch尺寸维128，训练用stochastic gradient descent(随机梯度下降)；所有权重以正态分布N(0,0.022)初始化；在LeakyReLU激活函数中，slope of the leak设为0.2；用带有可调超参数的Adam optimizer替代momentum来加速训练；学习率由0.001调整为0.0002；将momentum的term由0.9调整为0.5可以避免训练不稳定？不是将Momentum替代了嘛？。

1. LSUN

超过3百万张图片训练样本，无额外增加数据。小的学习率和mini-batch SGD使得模型不会记忆训练样本，训练不会过拟合(学习的模型过于复杂，限制度高，对于训练样本可以百分之百输出正确结果，对于测试样本却不然)，展示出的样本完全通过G生成。

下图展示了一个训练周期后生成的图片，图片逼真度比较高：



下图展示了在模型收敛后，G生成的图片，出现了视觉是上的欠拟合，由于训练周期次数多，噪声纹理多次与样本交织，使得效果变差：那就是模型训练完毕之后效果反而差了？什么道理？



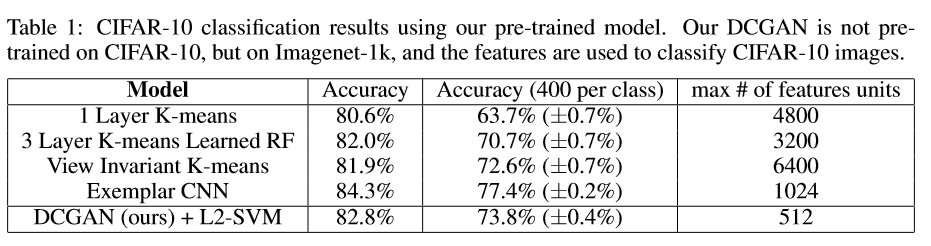
4.1是用deduplication的方法改进fig2中的memorizing问题，不甚明白，而且不是说实验上来说fig2没有记忆训练样本嘛？

1. Faces

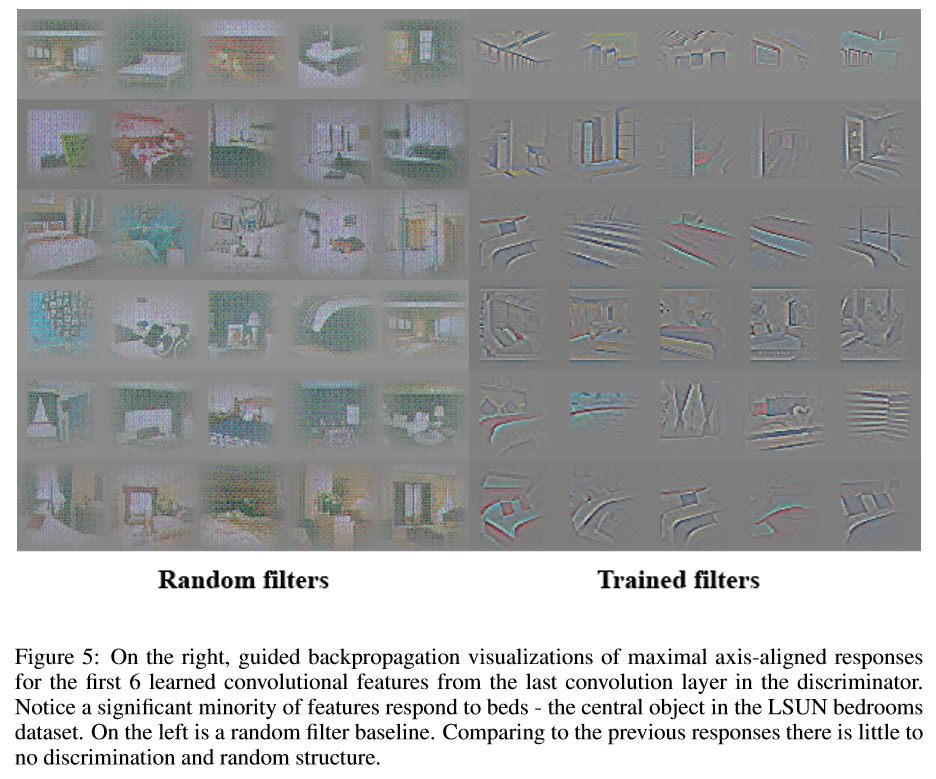
全部是现代人脸，350,000face boxes做训练。无额外增加数据。

1. Imagenet-1K

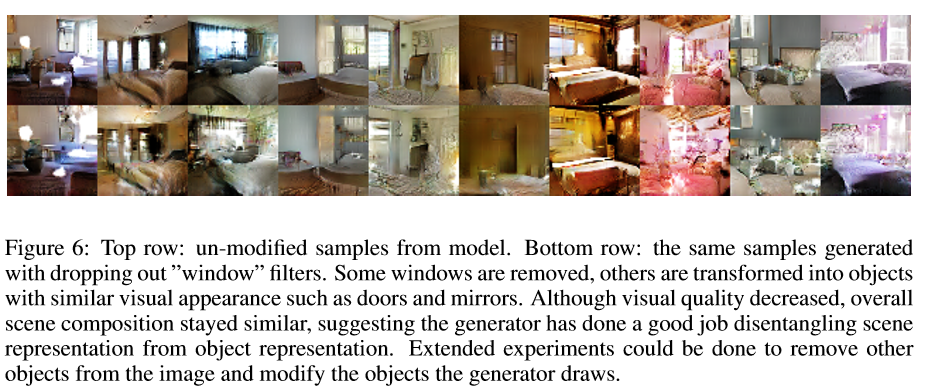
32\*32 mini-resized center crops images，无额外的数据增加。在该数据集上训练DCGAN后，将DCGAN中的discriminator作为特征提取模型，放到有监督模型中做CIFAR-10的分类任务，不在CIDAR-10上微调，直接提取特征，送入L2-SVM分类器中训练分类器，得到分类结果准确度。结果显示，discriminator的特征提取效果具有鲁棒性，它在imagenet-1k上训练，未经过微调直接提取CIFSR-10的特征做分类，结果仍好于K-means分类方式。虽然discriminator中每一层的feature map不超过512，但多层加起来使得最后一层的特征输出为28672维，仍然很多。



1. 可视化
2. 可视化判别模型的filters，可以看到这些训练的filter(右侧)只对特定结构活跃，如床，窗户：



1. 去掉个别物体的filters，如去掉windows的filters，结果显示生成模型仍能生成拥有类似场景的图片，只是窗户被摸出了，或者替换上了其他类似类型的物体。因此生成模型是将场景的生成和物体的生成分开来做的，这是一个优点。



1. 利用产生模型做人脸合成，还可以做视角变更等(无具体叙述，看看实验吧)