最近对生成模型有了一点小思考，实际上还是比较简单的问题，实验一下可能看得更清楚。于是写了这篇文章。这里说的生成模型主要指两类：GANs和VAEs。

**1. Why generative model?**

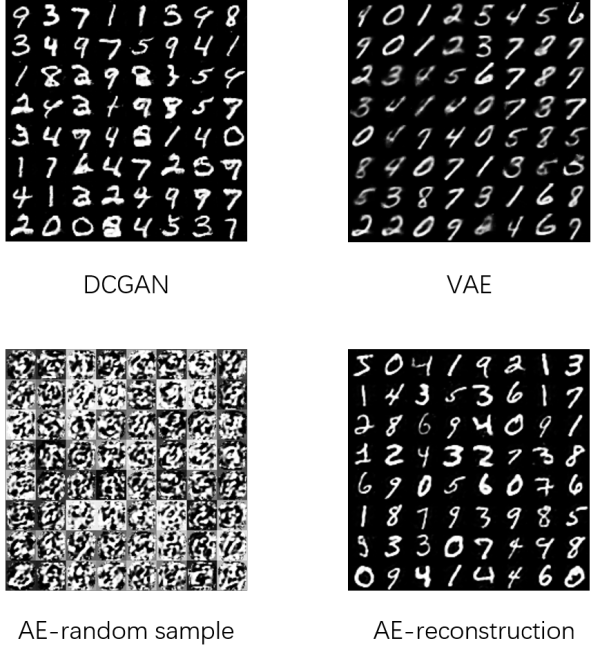
上周跟人聊起过GAN（generative adversarial network）和VAE（variational autoencoder），然后对方问了一句，为什么不直接用AE（autoencoder），而要用那么复杂的模型呢？很惭愧地说，关注了这么久生成模型，我没审视过这个问题。

我们知道，autoencoder能够在保持数据的可恢复性的基础上对原始数据进行压缩，也就是说，它能够对数据进行编码。如果我们随机产生一些编码，然后执行decoder，它能生成跟原始数据类似的样本出来吗？

如果这个想法可行的话，那autoencoder也是一类生成模型了。

我们通过实验观察来说明这个问题。

以toy example MNIST数据集为例，GAN我们采用DCGAN[1]（架构与DCGAN文中提出的一致），VAE采用最初版本的VAE[2]，autoencoder的encoder和decoder分别采用单层全连接网络，encoder的激活函数采用softplus，decoder的激活函数采用tanh。每个方法均训练25个epoch，实验效果如下：

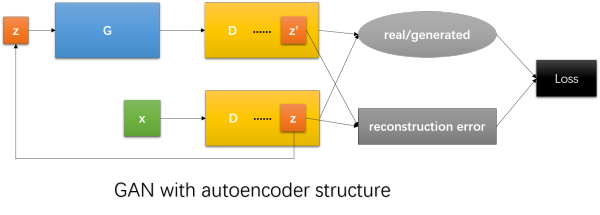


从以上结果可以很清楚地看出，生成模型泛化能力更好，随机的输入也能产生质量还不错的样本。而autoencoder在重构样本上能力很强，重构的样本质量很高，但是随机的输入并没有得到像样的样本，而是一团噪声。

当然，对于why generative model这个问题，还有很多其他原因，这不在本文的讨论范围之内。

**2. AE vs generative model**

为什么autoencoder不能对随机的输入产生像样的样本？VAE也是autoencoder的结构，这表明不是结构的问题。为了看清楚这个问题，我对GAN做了一些修改，“设计”了一种autoencoder结果的GAN，姑且称为AE-GAN吧。具体来说，它的结构如下：



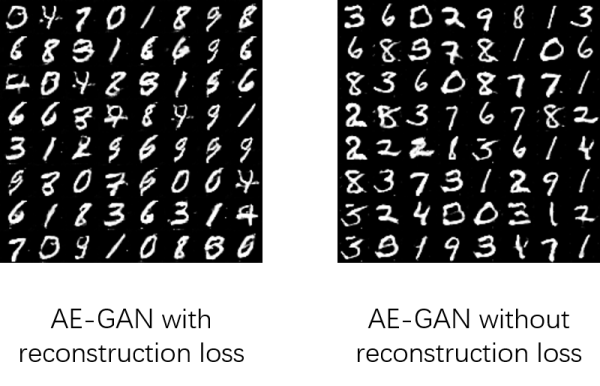
这个GAN可以拆成encoder和decoder两部分：

decoder：D网络将图像x编码为z

encoder：z作为隐变量输入G网络生成图像x'

它的loss分为两部分，第一部分还是general GAN的loss，第二部分是重构误差，可以是\Delta (x,x')，也可以是\Delta (z,z')，上面的结构图给出的是后者。以下实验也采用后者。

实验结果如下（25 epochs）：



可以看到，不管是否带有重构误差项，AE-GAN都能够产生质量还可以的样本。

这再次表明，autoencoder的结构并不是导致autoencoder不能对随机输入产生decent的样本的原因。

不是结构的问题，那是什么问题呢？

其实这个问题很好回答，是目标函数。生成模型的目标函数要么是似然函数，要么是分布相似性的度量，两者都是衡量生成样本的分布质量的指标。而autoencoder的目标函数是重构误差，也就是，它不管样本分布是怎样的，只要能够比较好地压缩并重构样本就行了。

那么，接下来就产生了两个问题：

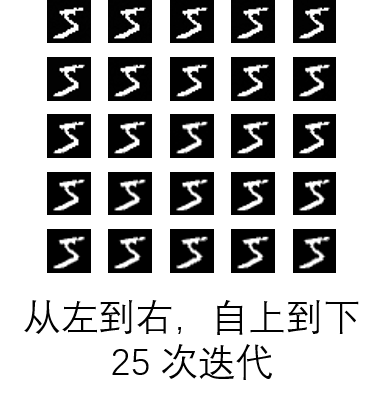
1. Autoencoder是否在优化的时候也自动考虑了样本的分布以使重构误差尽可能地小？
2. LSGAN（least square GAN）的目标函数是平方误差，它的目标函数并没有显式地衡量生成样本的分布质量如何，为什么也能对随机输入生成decent的样本？

对于问题1，考虑样本的分布应该能够引导重构误差向更小的方向发展，但是，神经网络具有非常多的鞍点，算法完全可能陷入了其他局部最优解。关于这个问题，我还在做实验，以后再做进一步的讨论。

至于问题2，实际上它优化的是Pearson \chi^2 散度，也是分布之间相似性的一种度量。更进一步地，我们似乎可以认为，是『对抗式』训练让模型获得了逼近真实样本分布的能力。

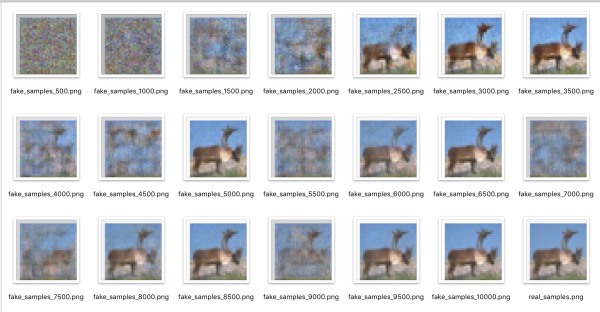
**3. Single sample case**

我们知道，一般来说GANs需要大量的样本才能得到较好的结果。而VAE则没有这个要求，单样本依然能够学得很好，并且也只会产生该样本，25次迭代产生的图像肉眼难以发现显著差别，大概是模型过拟合了，直接记住了样本。



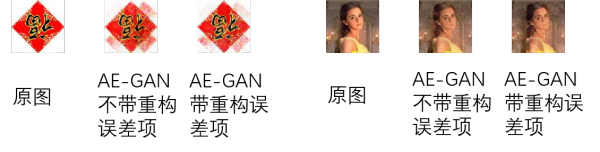
我们回过头来看看GANs，大部分GAN在单样本（or 少量样本）的情况下都会失败，这是为什么呢？我们观察到，以散度为分布相似性衡量标准的GAN一般都会失效，这个很好理解，D很快就达到了最优（第一次迭代时，G只会产生噪声，D的任务是区分噪声和真实图像，这是一个非常简单的任务），导致G没有梯度更新，这在WGAN中已经有证明。

那么，WGAN呢？WGAN的证明中指出，即使D接近最优，G依然有足够强的梯度，实际中我们也能发现，它对单样本也是有效的。这个我没有做实验，GAN讨论群里有个童鞋跑了实验，这里我直接贴(dao)出(qu)他的实验结果。可以观察到WGAN在这种情况下也不稳定。



还有什么GAN在单样本情形下也能work吗？

答案是有的。我们上面提出的AE-GAN就是其中一种。实验中，我们比较了带重构误差项和不带重构误差项两种情形，发现它们都能够work。说明原因不在于重构误差项，而在于AE-GAN的设计上。



对于不带重构误差项的AE-GAN，它的目标函数与general GAN一样，都是JS散度，也会出现梯度消失。然而，当迭代次数足够多时，它能正确收敛，这跟我们设计的网络结构有关，将真实图像提取的特征作为G的输入，当D比较强时，G也会增强，D不容易直接达到最优。

最后，我们指出，**单样本的情形在实际中并！没！有！什！么！用！**尽管如此，它却能简单有效地检验模型是否容易出现梯度消失。

**参考文献**

1. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.
2. Kingma D P, Welling M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.