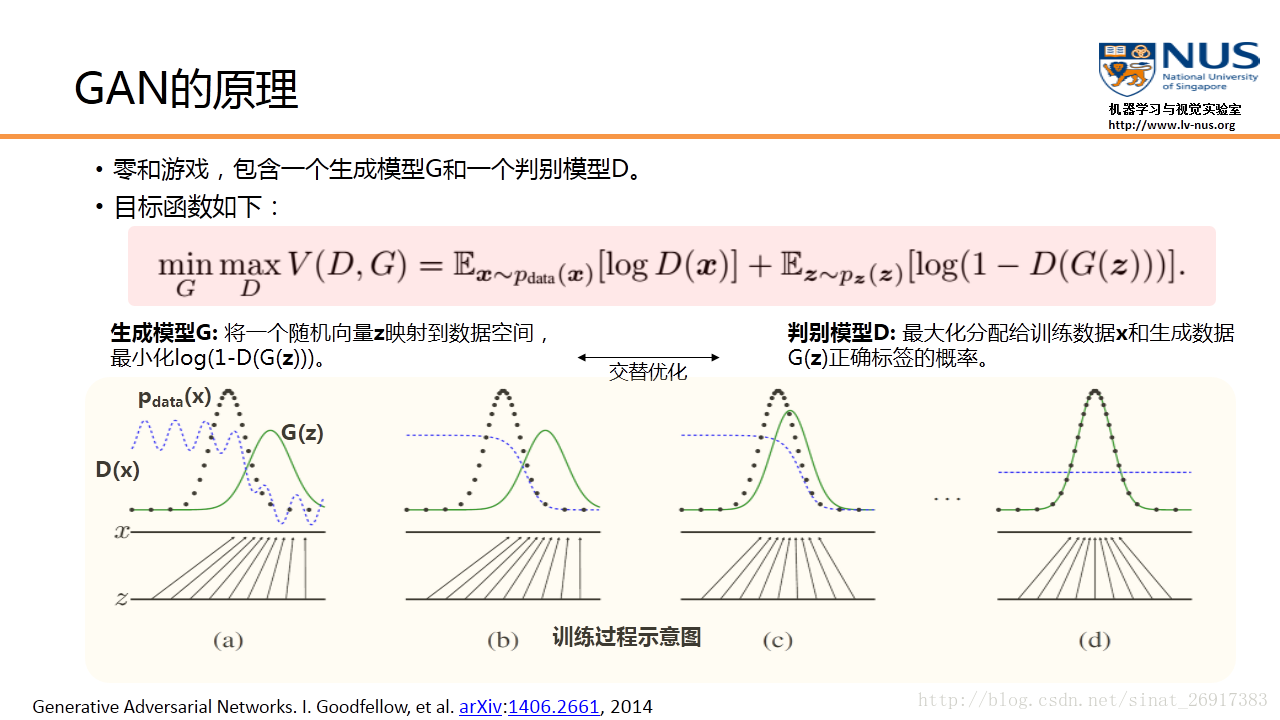
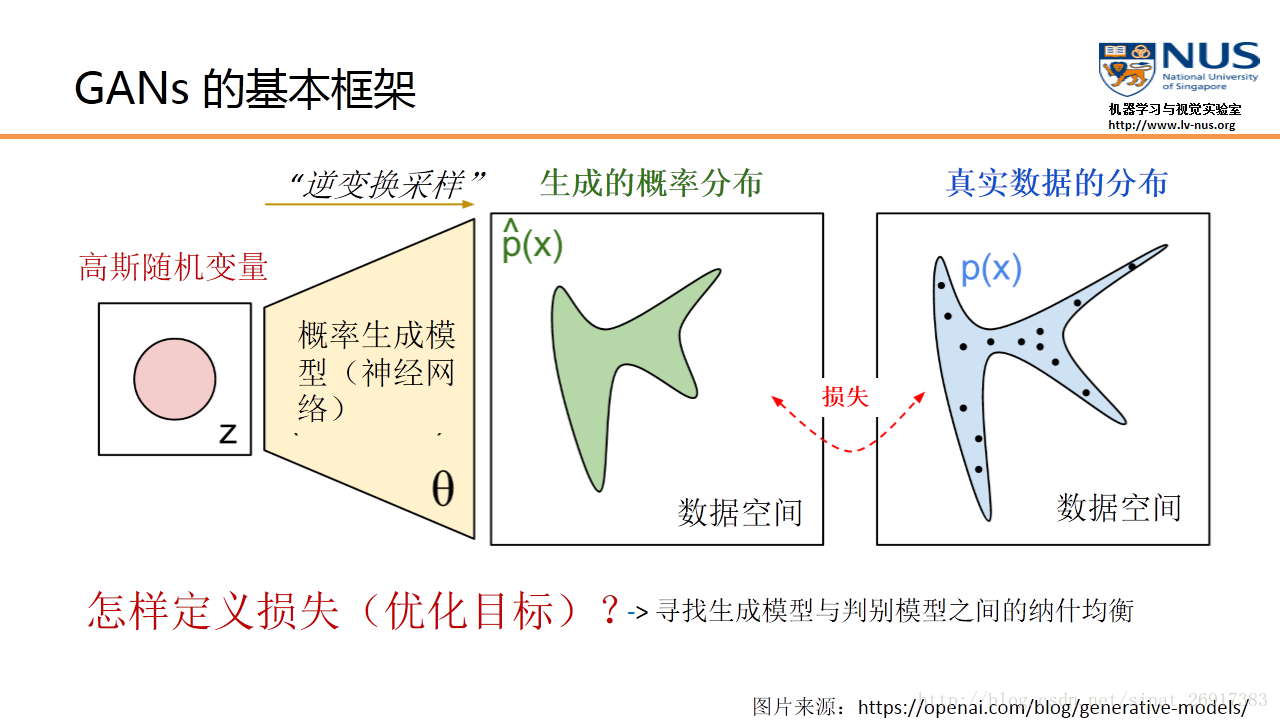
1. 原理：GAN启发自博弈论中的纳什平衡，包括了两个模型，一个生成模型（G）和一个判别模型（D）。真正的学习过程就是生成模型（G）和判别模型（D）之间的竞争过程，随机从真实样本和由生成模型（G）生成的“假样本”中取一个，让判别模型去判断是否为真。



图(a)中黑色大点虚线P(x)是真实的数据分布，绿线G(z)是通过生成模型产生的数据分布（输入是均匀分布变量z，输出是绿色的曲线）。蓝色的小点虚线D(x)代表判别函数。   
在图(a)中，我们可以看到，绿线G(z)分布和黑色P(x)真实分布，还有比较大的差异。这点也反映在蓝色的判别函数上，判别函数能够准确的对左面的真实数据输入，输出比较大的值。对右面虚假数据，产生比较小的值。但是随着训练次数的增加，图（b）和图（c）反映出，绿色的分布在逐渐靠近黑色的分布。到图（d），产生的绿色分布和真实数据分布已经完全重合。这时，判别函数对所有的数据（无论真实的还是生成的数据），输出都是一样的值，已经不能正确进行分类。G成功学习到了数据分布，这样就达到了GAN的训练和学习目的。

1. 运行机制：
2. 生成模型：一个最朴素的GAN模型，实际上是将一个随机变量（可以是高斯分布，或0到1之间的均匀分布），通过参数化的概率生成模型（通常是用一个神经网络模型来进行参数化），进行概率分布的逆变换采样，从而得到一个生成的概率分布（图中绿色的分布模型）。其训练目标，就是要最小化判别模型D的判别准确率。

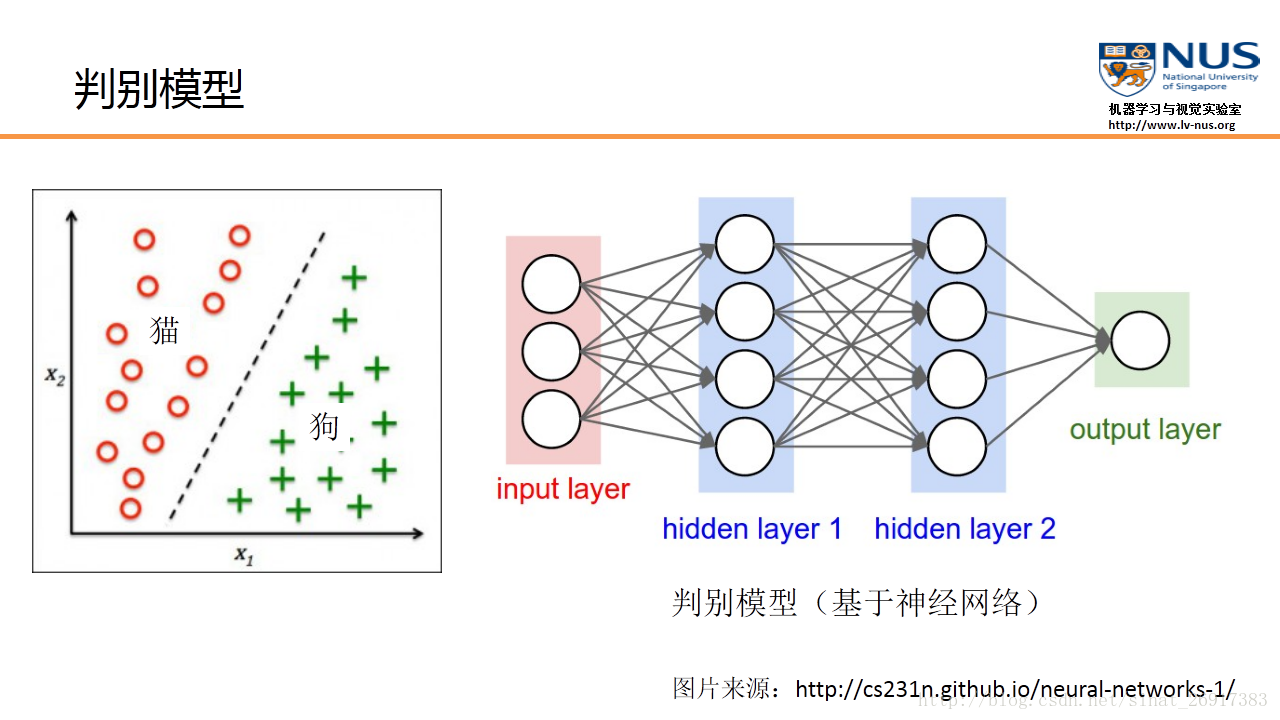


**生成模型就是一种寻找分布最优参数的过程，而这些参数更新不是来自于数据样本本身（不是对数据的似然性进行优化），而是来自于判别模型D的一个反传梯度**   
通过优化目标，使得我们可以调节概率生成模型的参数\theta，从而使得生成的概率分布和真实数据分布尽量接近。   
但是这里的分布参数不再跟传统概率统计一样了，这些参数保存在一个黑盒中：最后所学到的一个数据分布Pg(G)，没有显示的表达式。它只是一个黑盒子一样的映射函数：输入是一个随机变量，输出是我们想要的一个数据分布。

让我们回顾一下正态概率密度函数：   
这里写图片描述   
正态随机变量服从的分布就称为正态分布，记作 ，读作X服从正态分布

1. 判别模型：判别模型D的训练目的就是要尽量最大化自己的判别准确率。当这个数据被判别为来自于真实数据时，标注 1，自于生成数据时，标注 0。

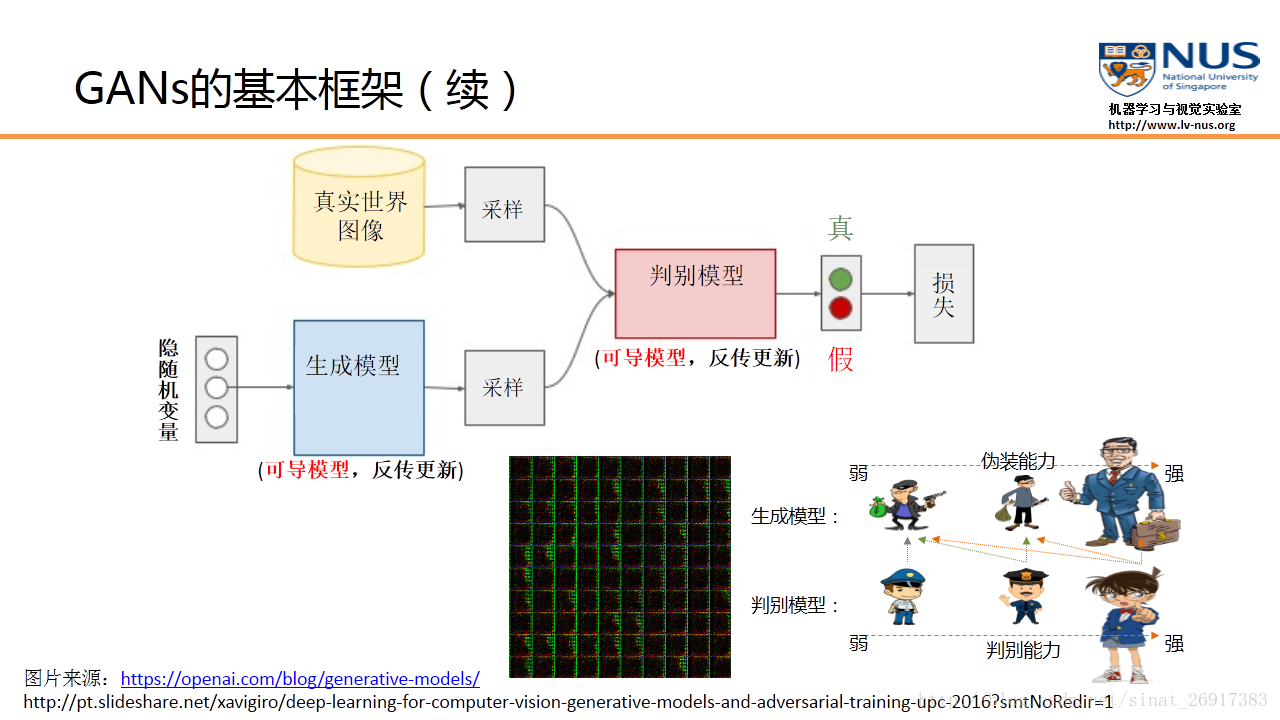
GAN创新性地引入了一个判别模型（常用的有支持向量机和多层神经网络）。它的优化过程就是在寻找生成模型和判别模型之间的一个纳什均衡



GAN引入一个判别模型（常用的有支持向量机和多层神经网络）。它的优化过程就是在寻找生成模型和判别模型之间的一个纳什均衡。

1. 目标函数——生成模型和判别模型的均衡

如果我们把生成模型比作是一个伪装者的话，那么判别模型就是一个警察的角色。伪装者的目的，就是通过不断的学习来提高自己的伪装能力，从而使得自己提供的数据能够更好地欺骗这个判别模型。而判别模型则是通过不断的训练来提高自己判别的能力，能够更准确地判断数据来源究竟是哪里。



**在训练过程中，GAN采用了一种非常直接的交替优化方式，它可以分为两个阶段：**

* 第一个阶段：固定判别模型D，然后优化生成模型G，使得判别模型的准确率尽量降低。
* 第二个阶段：固定生成模型G，来提高判别模型的准确率。

1. GAN的全局最优解和收敛性
2. **GAN是存在全局最优解的。**这个全局最优解可以通过一些简单的分析得到。首先，如果固定G，那么D的最优解就是一个贝叶斯分类器。将这个最优解形式带入，可以得到关于G的优化函数。简单的计算可以证明，当产生的数据分布与真实数据分布完全一致时，这个优化函数达到全局最小值。
3. **关于GAN的收敛性。**如果G和D的学习能力足够强，两个模型可以收敛。但在实际中，GAN的优化还存在诸如不稳定等一些问题。如何平衡两个模型在训练中是一个很重要的问题。
4. GAN的应用实例

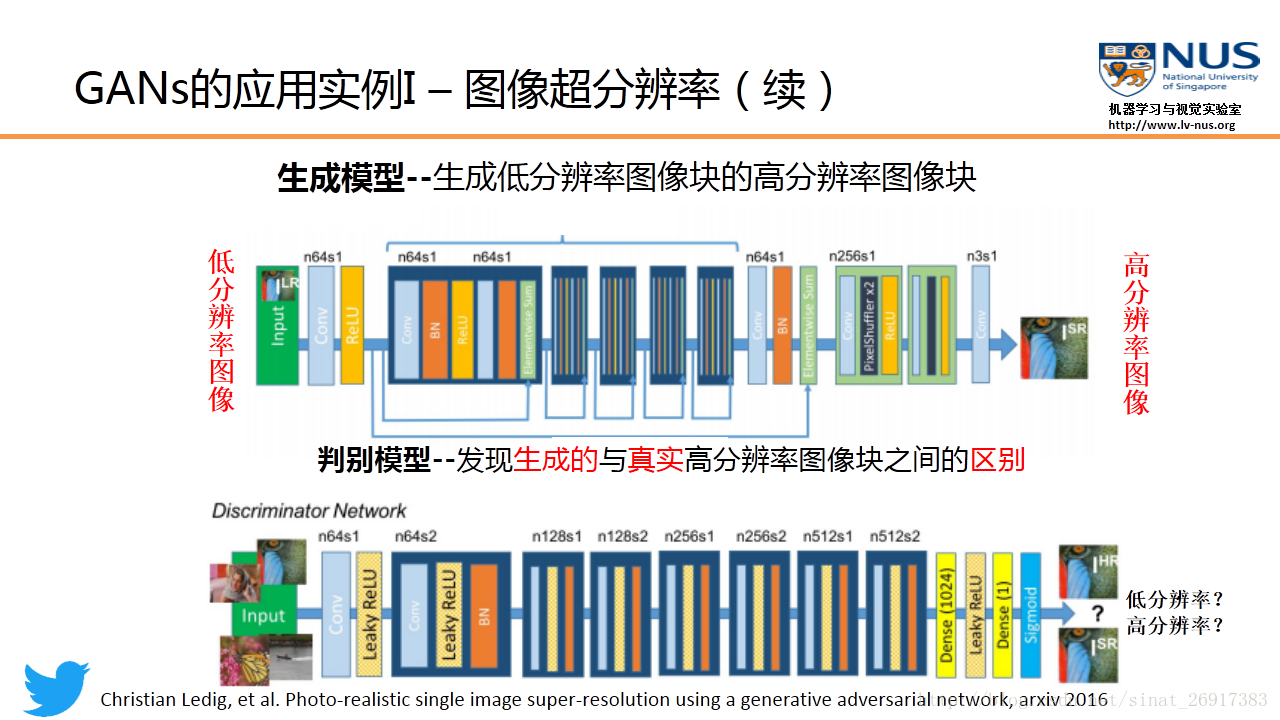
因为内部对抗训练的机制，GAN可以解决一些传统的机器学习中所面临的数据不足的问题，可以应用在半监督学习、无监督学习、多视角、多任务学习的任务中。最近有一些工作已经将其成功应用在强化学习中，来提高强化学习的学习效率。

1. 图像分辨率

图像超分辨率的目的，是将一个低分辨率的模糊图像，进行某种变换，得到一个高分辨率的带有丰富细节的清晰图像。

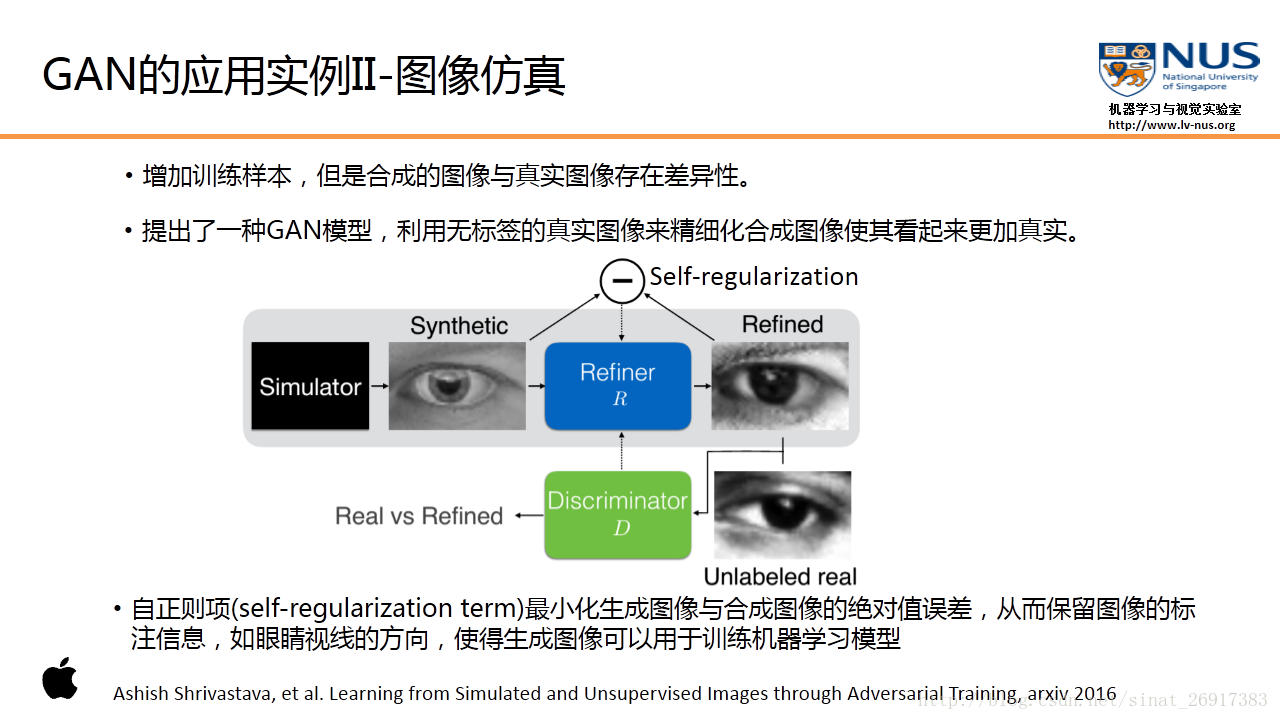


在 Twitter 这篇论文中，他们用一个16个残差块的网络来参数化生成模型。而判别模型使用的是一个VGG网络。这个实验结果也说明了使用GAN模型能够得到更好的结果。与以往基于深度学习模型做图像超分辨率的结果相比的话（比如SRResNet等），我们可以看到GAN的结果图能够提供更丰富的细节。这也就是GAN做图像生成时的一个显著优点，即能够提供更锐利的数据细节。



1. 图片合成

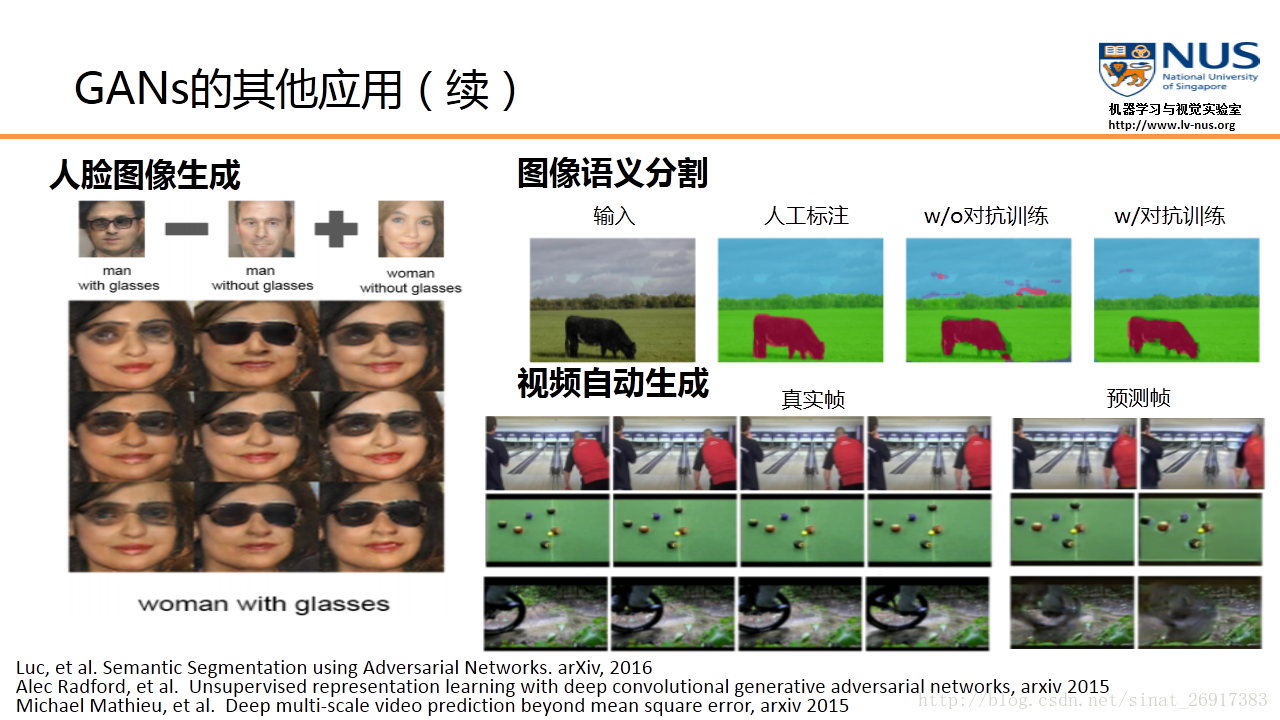
Apple发表的第一篇AI论文，要解决的问题，就是如何使得模拟的数据更加逼真，与真实图像的差异性尽量小。这篇论文中使用了类似GAN的框架，将模拟器（Simulator）产生的虚拟数据作为输入，通过一个叫做改进器（Refiner）的模型（对应生成模型）来产生改进后的虚拟数据。再同样的，使用一个判别器，来判断所产生的图像是真实的，还是虚拟的 。



1. 图像到图像的翻译+文本到图像的翻译

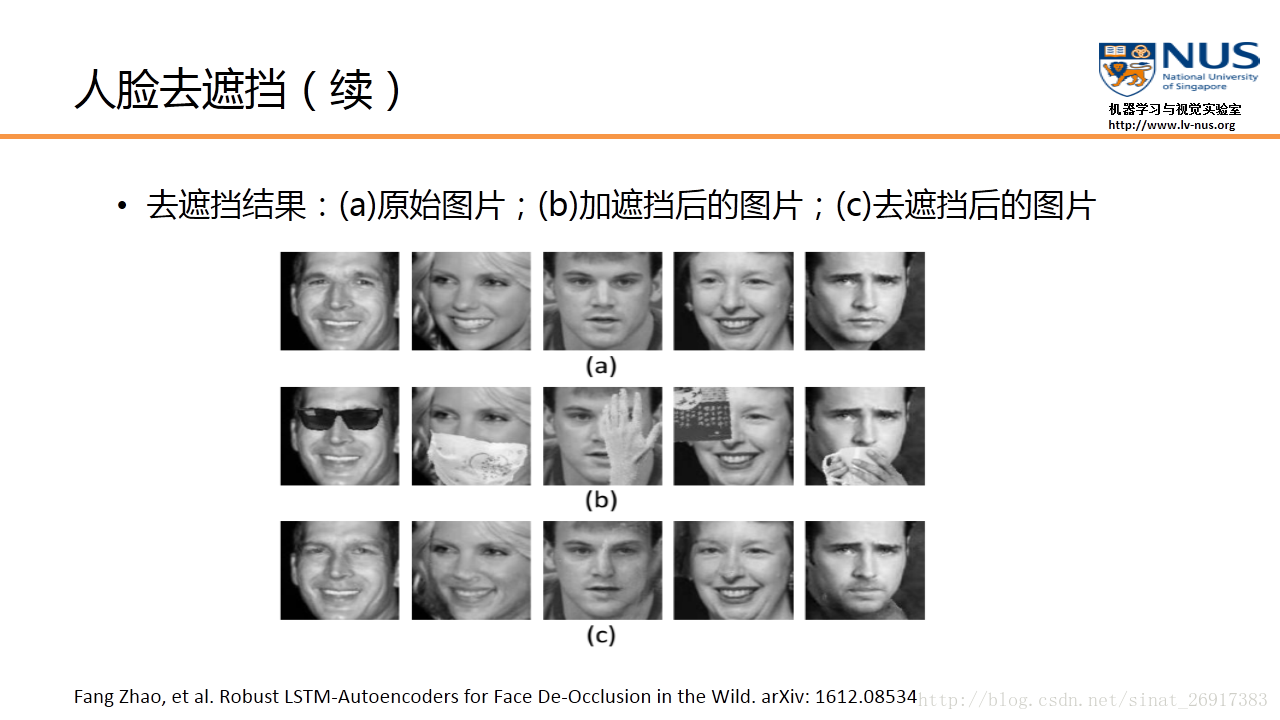
比如说将语义标注图、灰度图或边缘图作为GAN的输入，那么我们希望它输出能够和输入图一致的真实图像，例如这里的街景图和彩色图。

文本到图像的翻译：GAN的输入是一个描述图像内容的一句话，比如“一只有着粉色的胸和冠的小鸟”，那么所生成的图像内容要和这句话所描述的内容相匹配。



1. 人脸去遮挡

新加坡国立大学【机器学习与视觉实验室】研究项目。保持人的身份信息的GAN模型，实验结果证明，这个模型不仅能够检测和去掉在人脸上的遮挡，同时还能保持人的身份信息，从而提高人脸的识别准确率。



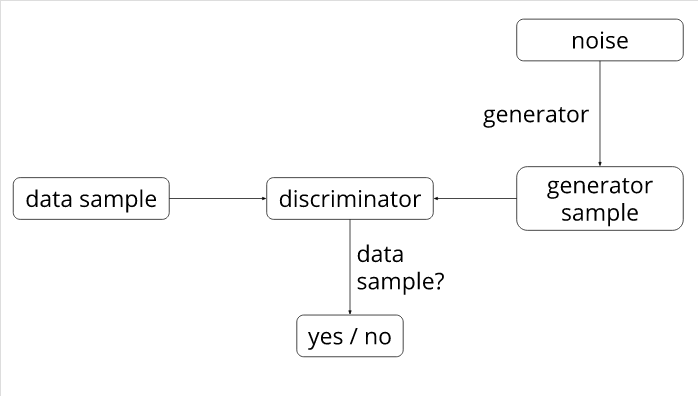
进阶篇

**Adversarial Training**

Adversarial Training是训练生成模型的一种方法。为了训练生成模型，Adversarial Training提出一种Discriminative Model(判别模型)来和生成模型产生对抗，下面来说说Generative models G(z)和 Discriminative Model D(x) 是如何相互作用的。

* 生成模型的目标是模仿输入训练数据, 通过输入一个随机噪声来产生和训练数据相似的样本；
* 判别模型的目标就是判断生成模型产生的样本和真实的输入样本之间的相似性。

其中生成模型和判别模型合起来的框架被称为GAN网络。通过下图我们来理清判别模型和生成模型之间的输入输出关系：生成模型通过输入随机噪声 z(z∼pz)产生合成样本；而判别模型通过分别输入真实的训练数据和生成模型的训练数据来判断输入的数据真实。



描述了GAN的网络结构，但它的优化目标是什么？怎么就可以通过训练使得生成模型能够产生和真实数据相似的输出？优化的目标其实很简单，简单来说就是：

* 判别模型努力的想把真实的数据预测为1，把生成的数据预测为0；
* 而生成模型的奋斗目标则为‘我’要尽力的让判别模型对‘我’生成的数据预测为1，让判别模型分不清‘我’产生的数据和真实数据之间的区别，从而达到‘以假乱真’的效果。

下面用形式化说明下如果训练GAN网络, 先定义一些参数：

| **参数** | **含义** |
| --- | --- |
| pz | 输入随机噪声 z 的分布 |
| pdata | 未知的输入样本的数据分布 |
| pg | 生成模型的输出样本的数据分布，GAN的目标就是要pg=pdatapg=pdata |

训练判别模型 D(x) 的目标：

1. 对每一个输入数据 x∼pdatax∼pdata 要使得 D(x)最大；
2. 对每一个输入数据 x≁pdatax≁pdata 要使得 D(x)最小。

训练生成模型 G(z) 的目标是来产生样本来欺骗判别模型 D, 因此目标为最大化 D(G(z))，也就是把生成模型的输出输入到判别模型，然后要让判别模型预测其为真实数据。同时，最大化 D(G(z))等同于最小化 1−D(G(z))，因为 D 的输出是介于0到1之间的，真实数据努力预测为1，否则为0。

所以把生成模型和判别模型的训练目标结合起来，就得到了GAN的优化目标：



总结一下上面的内容，GAN启发自博弈论中的二人零和博弈，在二人零和博弈中，两位博弈方的利益之和为零或一个常数，即一方有所得，另一方必有所失。GAN模型中的两位博弈方分别由生成模型和判别模型充当。生成模型G捕捉样本数据的分布，判别模型是一个二分类器，估计一个样本来自于训练数据（而非生成数据）的概率。G和D一般都是非线性映射函数，例如多层感知机、卷积神经网络等。生成模型的输入是一些服从某一简单分布（例如高斯分布）的随机噪声z，输出是与训练图像相同尺寸的生成图像。向判别模型D输入生成样本，对于D来说期望输出低概率（判断为生成样本），对于生成模型G来说要尽量欺骗D，使判别模型输出高概率（误判为真实样本），从而形成竞争与对抗。