

生成对抗网络

Steven Tang

诞生

- 生成对抗网络是由Ian Goodfellow等人在论文《Generative Adversarial Networks》中提出的,
- 论文的地址是: arxiv.org/abs/1406.2661

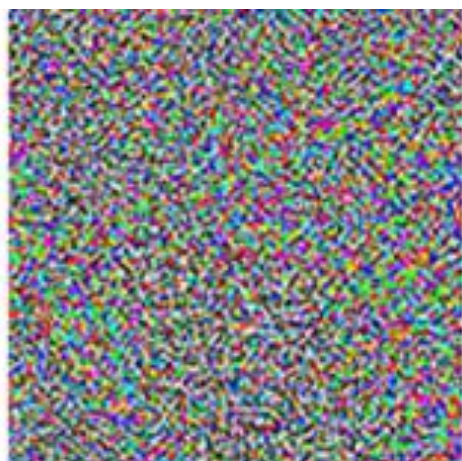
对抗样本(adversarial examples)



"panda"

57.7% confidence

+ ϵ



=



"gibbon"

99.3% confidence

生成式对抗网络

- 生成式模型 (generative model)
- 判别式模型 (discriminative model)
- 生成模型G: 捕捉样本数据的分布, 用服从某一分不 (均匀分布, 高斯分布) 的噪声 z 生成一个类似真实训练数据的样本, 追求效果是越像真实的越好。
- 判别模型D: 是一个二分类器, 估计一个样本来自训练数据 (而非生成数据) 的概率, 如果样本来自真实的训练数据, D输出大概率, 否则, D输出小概率。

博弈论-纳什均衡

囚徒困境

| A\B | 坦白 | 抵赖 |
|-----|--------|--------|
| 坦白 | -8, -8 | 0, -10 |
| 抵赖 | -10, 0 | -1, -1 |

生成模型发展历史

生成方法和判别方法

机器学习方法可以分为生成方法和判别方法，所学到的模型分别称为生成式模型和判别式模型。生成方法通过观测数据学习样本与标签的联合概率分布 $P(X, Y)$ ，训练好的模型能够生成符合样本分布的新数据，它可以用于有监督学习和无监督学习。判别方法由数据直接学习决策函数 $f(X)$ 或者条件概率分布 $P(Y|X)$ 作为预测的模型，即判别模型。

早期深层生成模型

深度产生式模型 的 深度信念网络 (DBN)。DBN是由一组受限玻尔兹曼机(RBMs)堆叠而成的深度生成式网络，它的核心部分是贪婪的、逐层学习的算法，这种算法可以最优化深度置信网络的权重。以无监督方式预训练的生成式模型 (DBN) 可以提供良好的初始点，然后通过有监督的反向传播算法微调权值

GAN

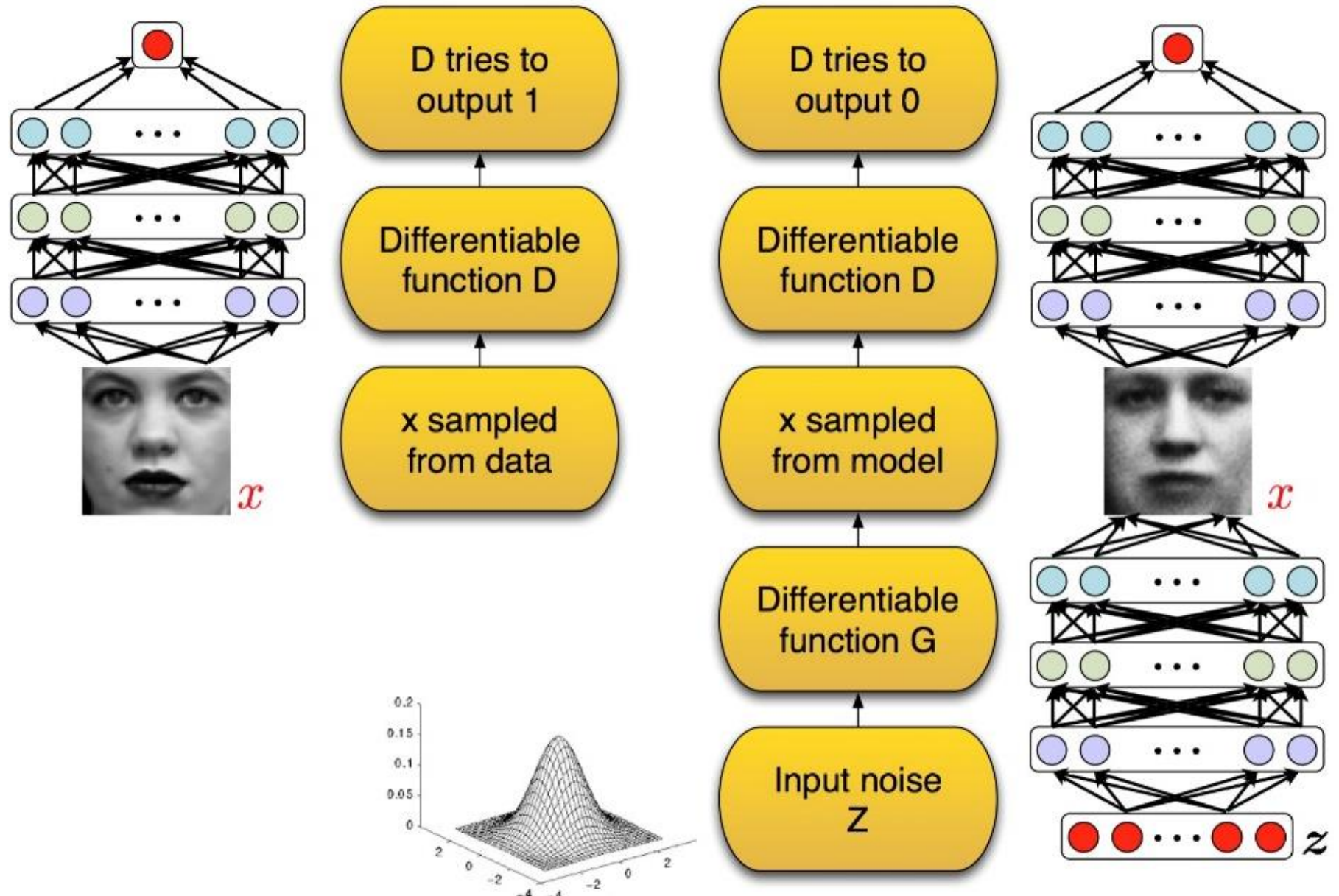
生成对抗网络，由两个网络组成，即生成器和判别器，生成器用来建立满足一定分布的随机噪声和目标分布的映射关系，判别器用来区别实际数据分布和生成器产生的数据分布。

生成对抗网络基本框架

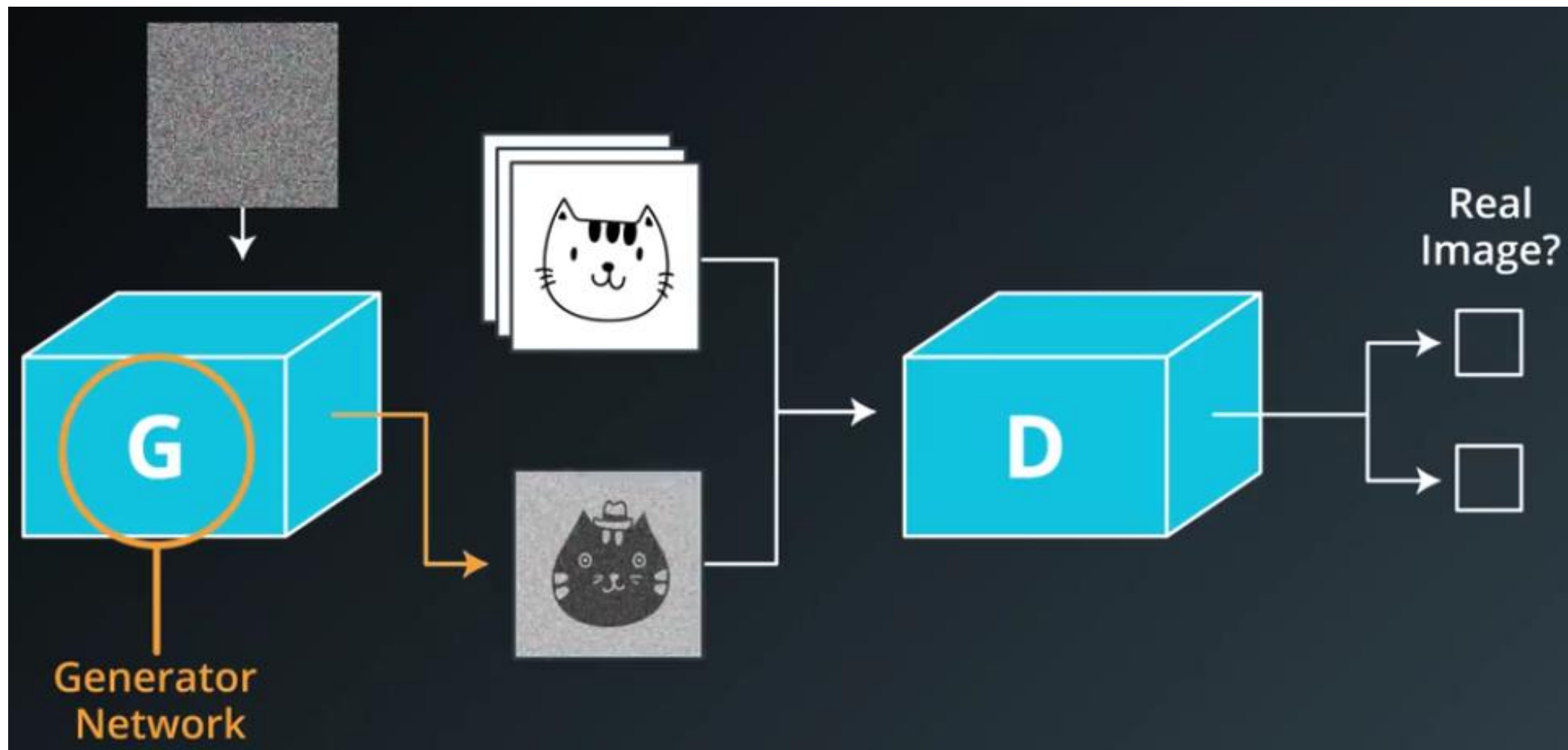
◆当固定生成网络 G 的时候，对于判别网络 D 的优化，可以这样理解：输入来自于真实数据，D 优化网络结构使自己输出 1，输入来自于生成数据，D 优化网络结构使自己输出 0；当固定判别网络 D 的时候，G 优化自己的网络使自己输出尽可能和真实数据一样的样本，并且使得生成的样本经过 D 的判别之后，D 输出高概率。



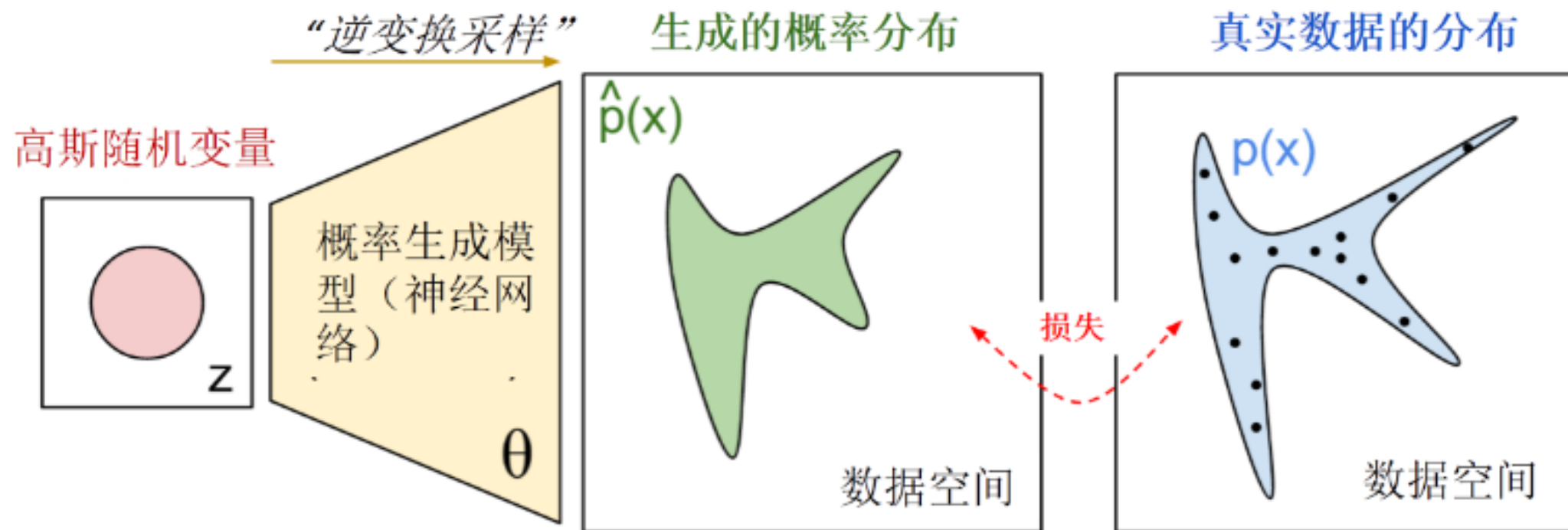
网络模型



GAN架构



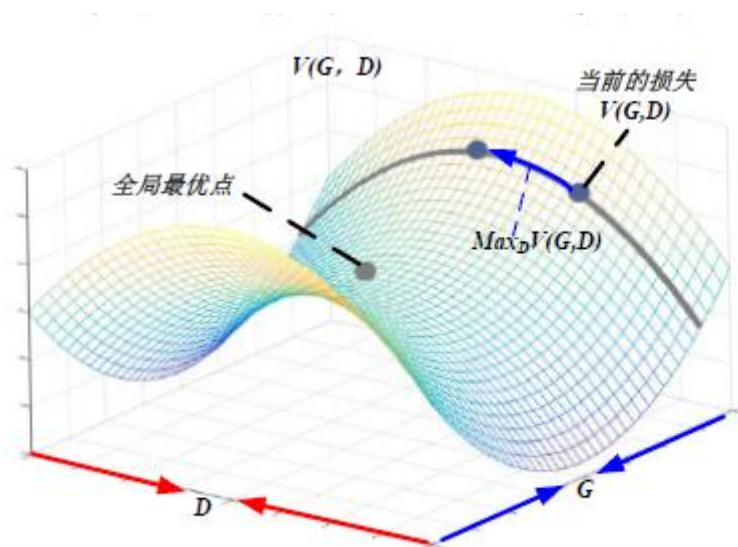
生成式对抗网络--如何定义损失



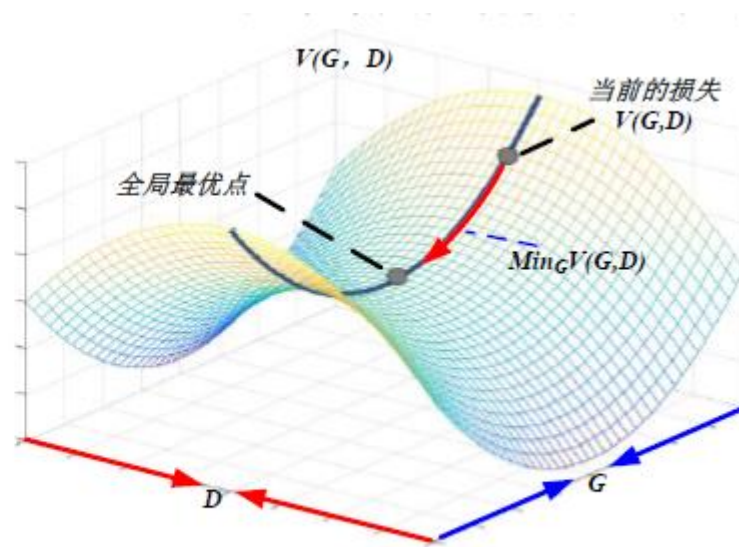
怎样定义损失（优化目标）？-> 寻找生成模型与判别模型之间的纳什均衡

优化函数的目标函数

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))].$$

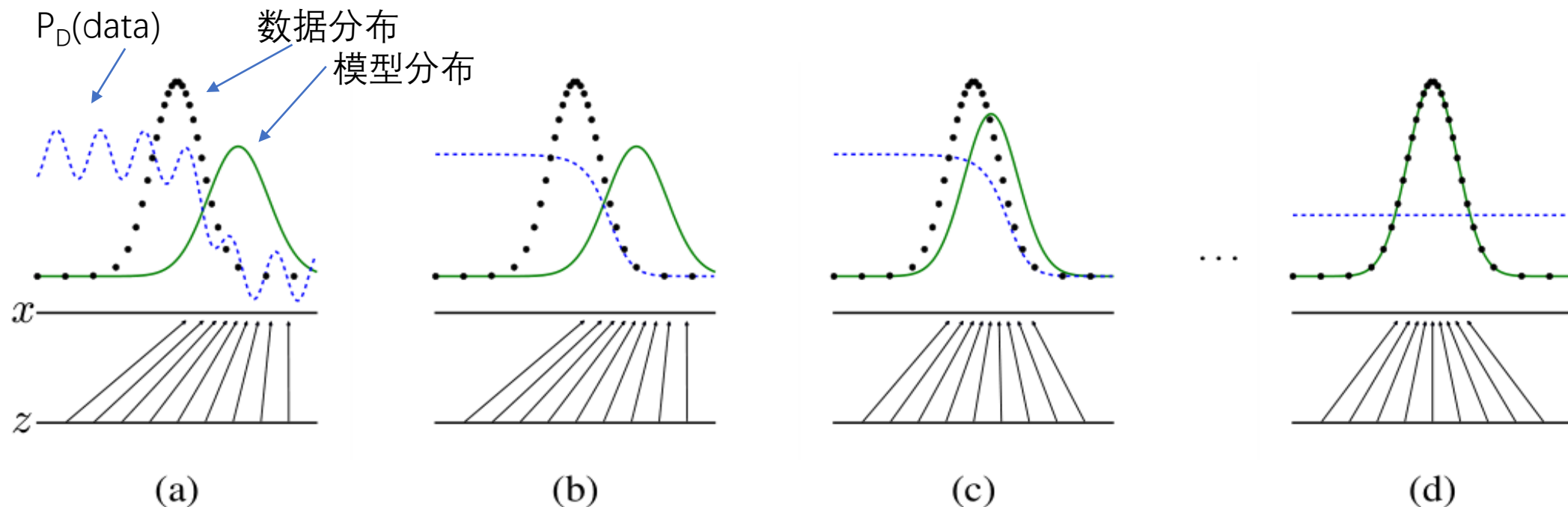


(a) D 的优化过程



(b) G 的优化过程

生成式对抗网络--训练方法



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 ，输出是绿色的线）

蓝色的小点虚线 $D(X)$ 代表判别函数

较低的水平线是 z 采样的区域，在这种情况下，上面的水平线是 X 域的一部分。向上箭头显示映射 $x=g(z)$ 如何将非均匀分布的 p_g 强加于转换后的样本上。 g 在高密度区域收缩，在 p_g 低密度区域扩展。

A. P_g 和 P_{data} 相似， D 是部分精确的分类器

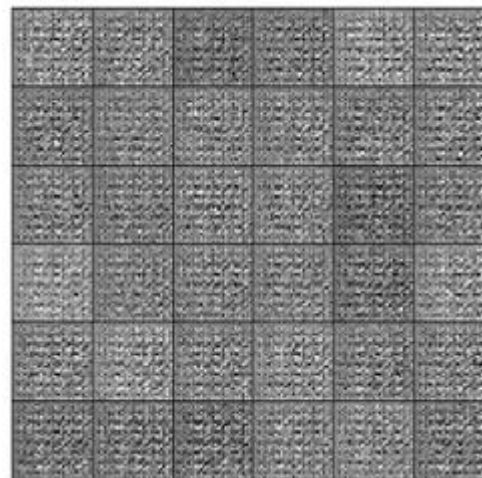
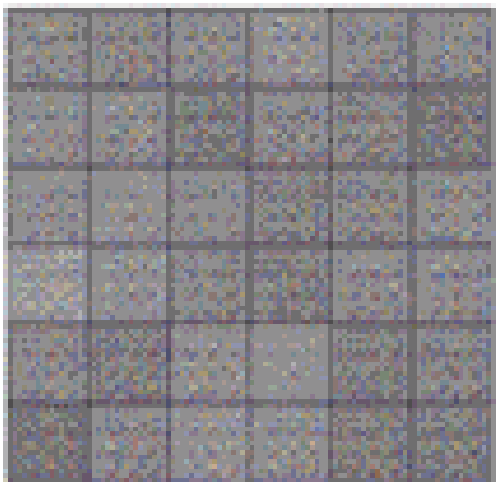
B. D 被训练以区分样本和数据，并收敛到

$$D(x) = \frac{P_{\text{data}}(x)}{P_{\text{data}}(x) + P_g(x)}$$

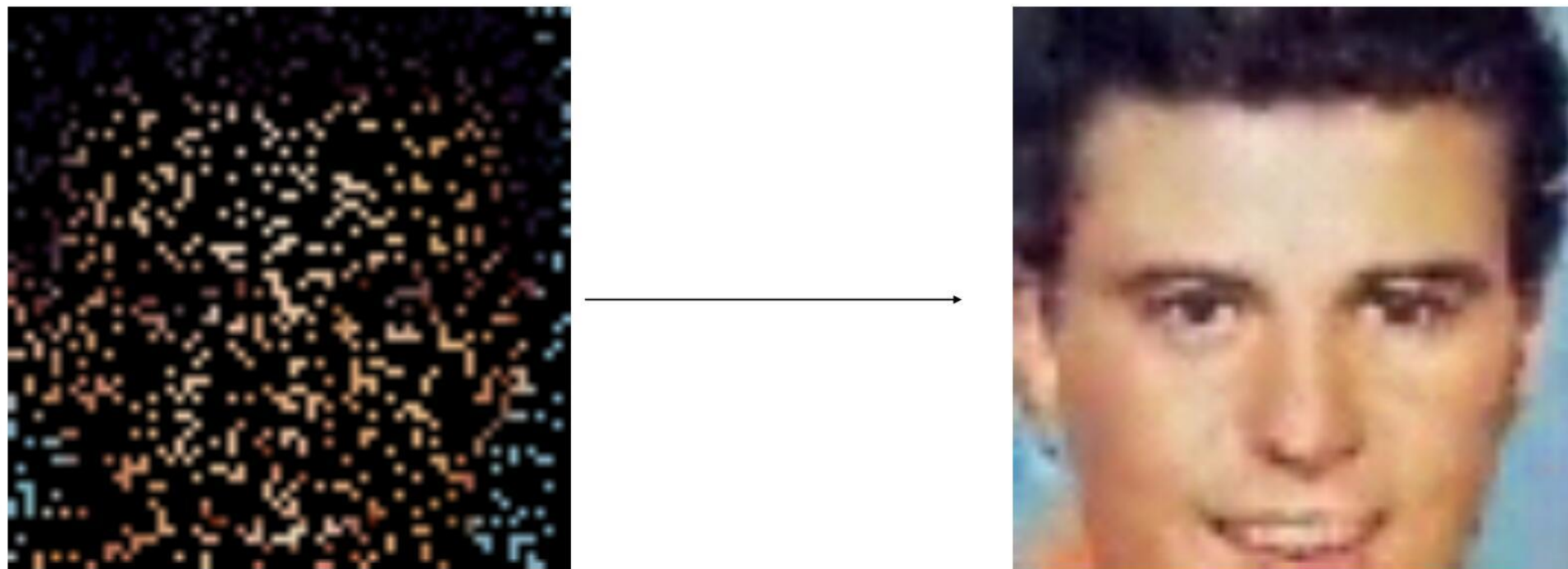
C. 在更新 g 之后， d 的梯度引导 $g(z)$ 流向更有可能被归类为数据的区域。

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。 $P_g = P_{\text{data}}$ ，判别器无法区分这两个分布，此时 $D(x) = 1/2$

生成对抗网络效果

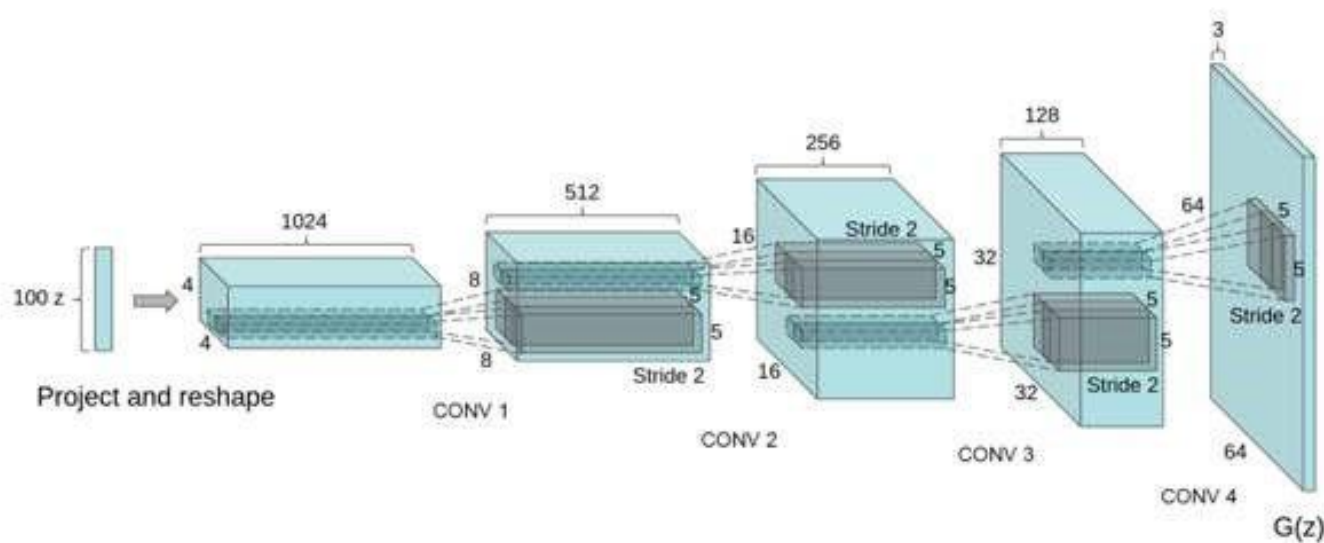


在大量噪声中恢复出一张人脸



(Yeh et al., 2016)

DCGAN (Deep Convolutional Generative Adversarial Network)



- 所有的pooling层使用步幅卷积(判别网络)和微步幅度卷积(生成网络)进行替换。
- 在生成网络和判别网络上使用批处理规范化。
- 对于更深的架构移除全连接隐藏层。
- 在生成网络的所有层上使用ReLU激活函数，除了输出层使用Tanh激活函数。
- 在判别网络的所有层上使用LeakyReLU激活函数。

<https://arxiv.org/pdf/1511.06434.pdf>

哪些是真的衣服？





horse → zebra

Summer ↔ Winter



真实图像



重建图像



金发 ↑

爆炸头 ↑

微笑 ↑

男人 ↑

http://blog.csdn.net/sir_chai



http://blog.csdn.net/sir_chai



a pizza on a plate at a restaurant



oranges on a table next to a liquor bottle



a) Generic description:

A group of people are sitting around a living room together. One of the men is wearing black sleeve shirt and blue pants. A man is sitting next to the wooden table. A man and woman are sitting on a couch. There is a brown wooden table in the room.

http://blog.csdn.net/sir_chai

模糊图像



生成模型
(神经网络)

判别模型
(神经网络)

高分辨率图像？
低分辨率图像？

模糊图像



SRResNet结果图






SRGAN结果图



高分辨率图像



| Caption | Image |
|---|--|
| <p>this vibrant red bird has a pointed black beak</p> |  |
| <p>this bird is yellowish orange with black wings</p> |  |
| <p>the bright blue bird has a white colored belly</p> |  |



(a)



(b)



(c)



(d)

http://blog.csdn.net/sinat_26917383

viewpoint



shape



texture



人脸生成的发展



[Goodfellow et al., 2014]
University of Montreal



[Radford et al., 2015]
Facebook AI Research



[Roth et al., 2017]
Microsoft and ETHZ



[Karras et al., 2018]
NVIDIA

BigGAN（最强的生成网络）

这是GAN在图像生成方面的最新发展。谷歌的一名实习生和谷歌的DeepMind部门的两名研究人员发表了一篇题为“Large Scale GAN Training for High Fidelity Natural Image Synthesis”的论文, 可在arxiv.org/abs/1809.11096上找到

BigGan

- 通过2-4倍的增加参数量（增加channel），8倍的扩大batchsize，可以使GAN获得最大的性能提升。
- 通过使用截断技巧（truncation trick），可以使得训练更加平稳，但是需要在多样性和逼真度之间做平衡。
- 通过现存的和其他新颖的各种技术的集合，可以保证训练的平稳性，但是精度也会随之下降，需要在性能和训练平稳性之间做平衡。

BigGan



Figure 1: Class-conditional samples generated by our model.

BigGan

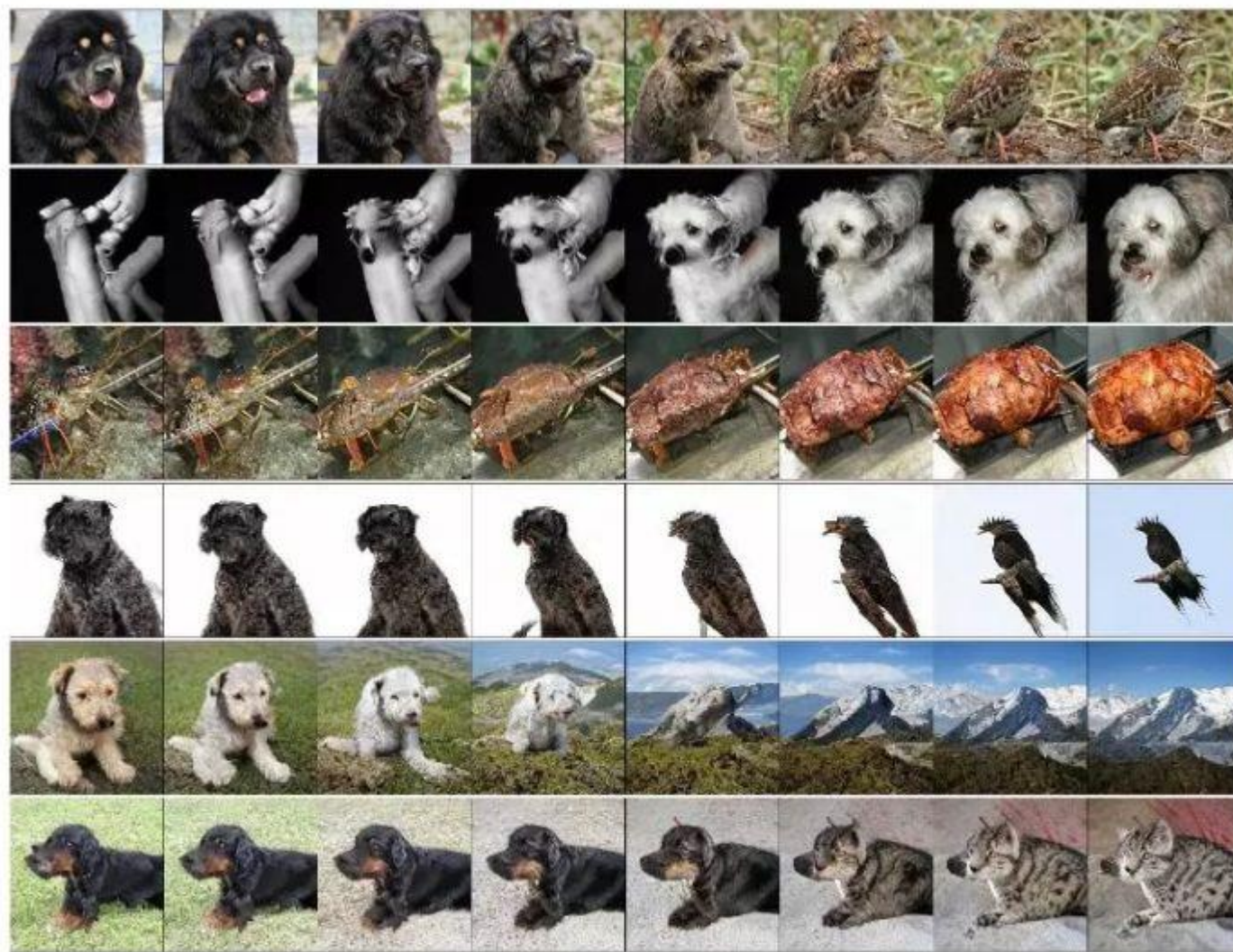


Figure 9: Interpolations between c with z held constant. Pose semantics are frequently maintained between endpoints (particularly in the final row). Row 2 demonstrates that grayscale is encoded in the joint z, c space, rather than in z .

Inception Score

Inception Score 是主要考虑这两个方面的：

1. 清晰度：把生成的图片 x 输入 Inception V3 中，将输出 1000 维的向量 y ，向量的每个维度的值对应图片属于某类的概率。对于一个清晰的图片，它属于某一类的概率应该非常大，而属于其它类的概率应该很小（这个假设本身是有问题的，有可能有些图片很清晰，但是具体属于哪个类却是模棱两可的）。用专业术语说， $p(y|x)$ 的熵应该很小（熵代表混乱度，均匀分布的混乱度最大，熵最大）。

Inception Score

Inception Score 是主要考虑这两个方面的：

2. 多样性： 如果一个模型能生成足够多样的图片，那么它生成的图片在各个类别中的分布应该是平均的，假设生成了 10000 张图片，那么最理想的情况是，1000 类中每类生成了 10 张。转换成术语，就是生成图片在所有类别概率的边缘分布 $p(y)$ 熵很大（均匀分布）。

$$\hat{p}(y) = \frac{1}{N} \sum_{i=1}^N p(y|\mathbf{x}^{(i)}) \quad (1)$$

$$\mathbf{IS}(G) = \exp\left(\mathbb{E}_{\mathbf{x} \sim p_g} D_{KL}(p(y|\mathbf{x}) || p(y))\right) \quad (2)$$

Inception Score 存在的一些问题：

IS大，不一定生成的图片就真实；

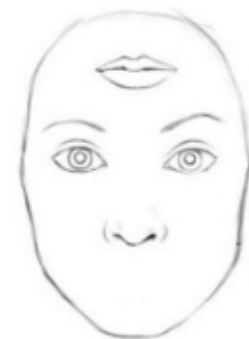
分类模型参数的轻微变动将影响 IS；

使用 IS 时，分类模型和生成模型应该在同一个数据集上训练；

通常计算 IS 的方法是有漏洞的：估计的样本数据量太小，导致同一堆数据，分割的份数不同算出的 IS 不同；

以 IS 为优化目标会导致产生对抗样本；

IS 无法反映生成模型过拟合情况。



BigGan 不足



GAN创造力

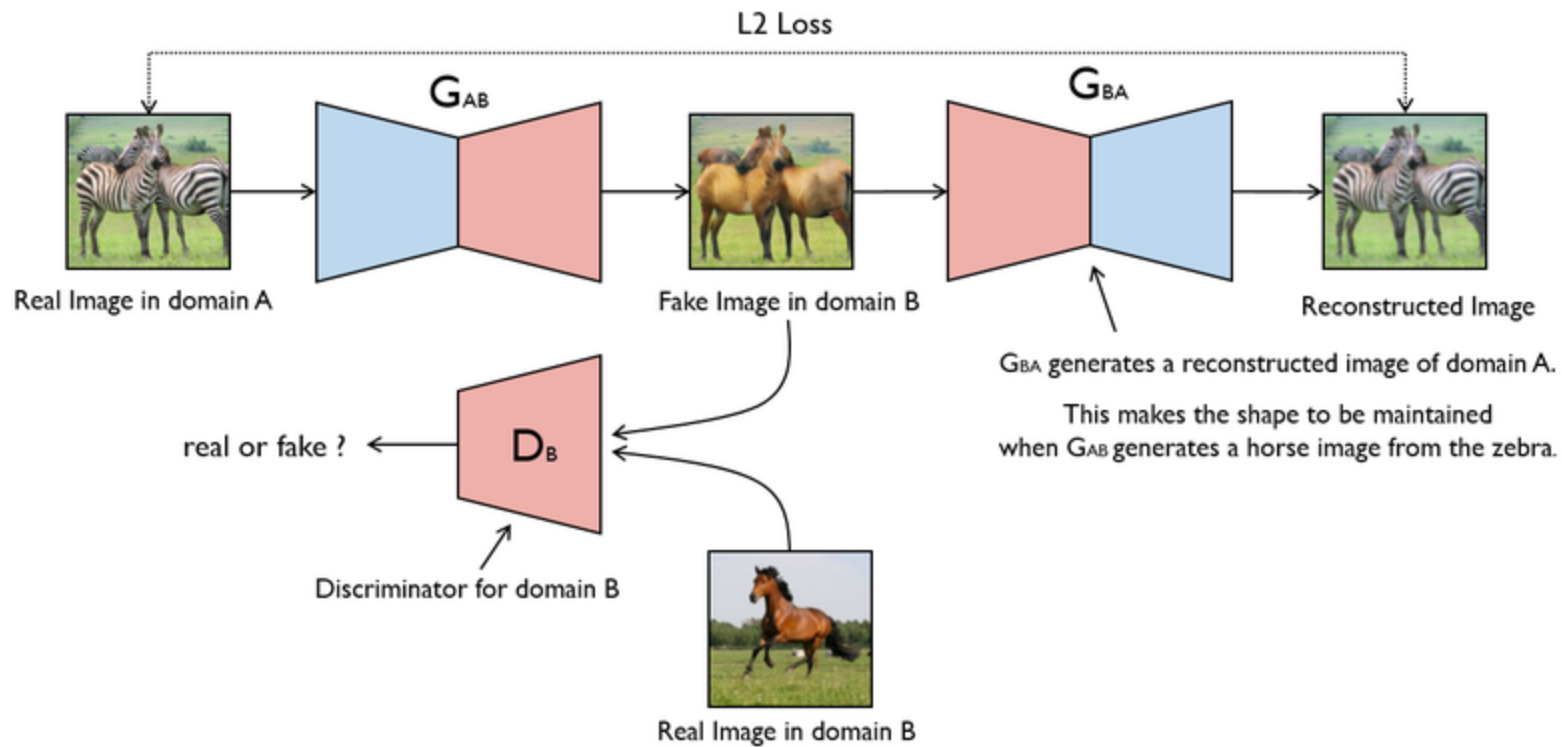


CycleGan真正的新秀

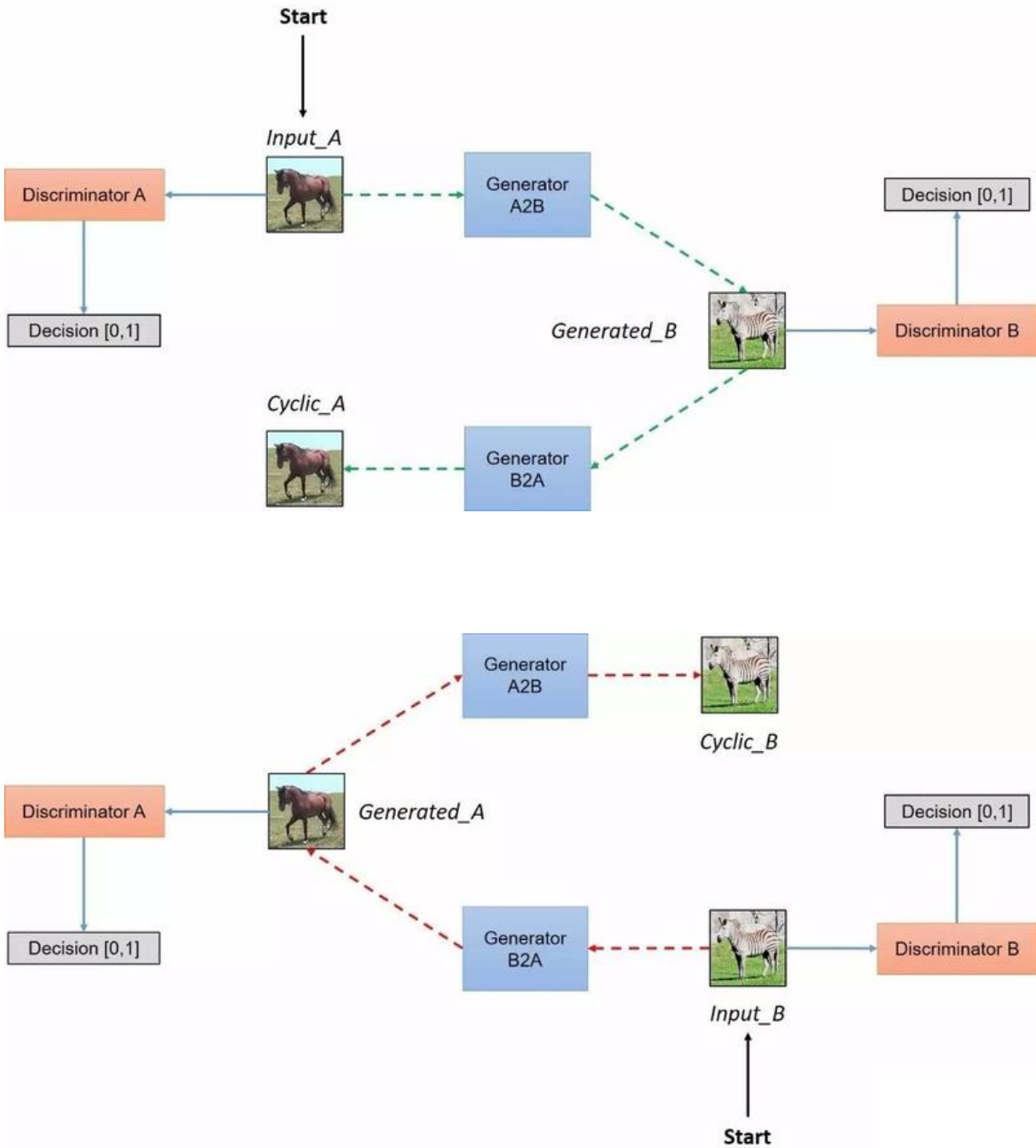


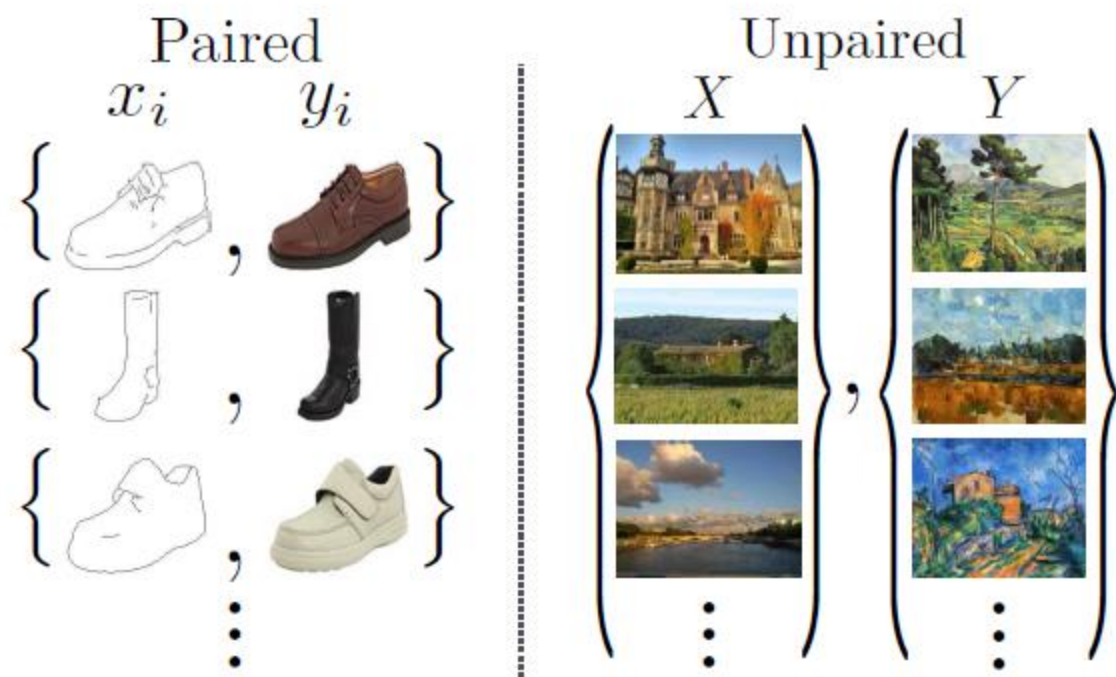
<https://arxiv.org/abs/1703.10593>

单向GAN



双向GAN示意图





$$x \rightarrow G(x) \rightarrow F(G(x)) \approx x.$$

$$y \rightarrow F(y) \rightarrow G(F(y)) \approx y.$$

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] \\ + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(F(y)) - y\|_1].$$

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D_Y(y)] \\ + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(1 - D_Y(G(x)))],$$

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ + \lambda \mathcal{L}_{\text{cyc}}(G, F),$$

Monet ↔ Photos



Monet → photo

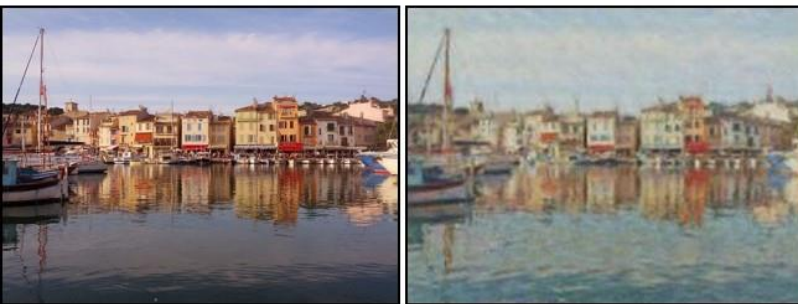


photo → Monet

Zebras ↔ Horses



zebra → horse



horse → zebra

Summer ↔ Winter



summer → winter



winter → summer



Photograph



Monet



Van Gogh



Cezanne



Ukiyo-e

镜像GAN

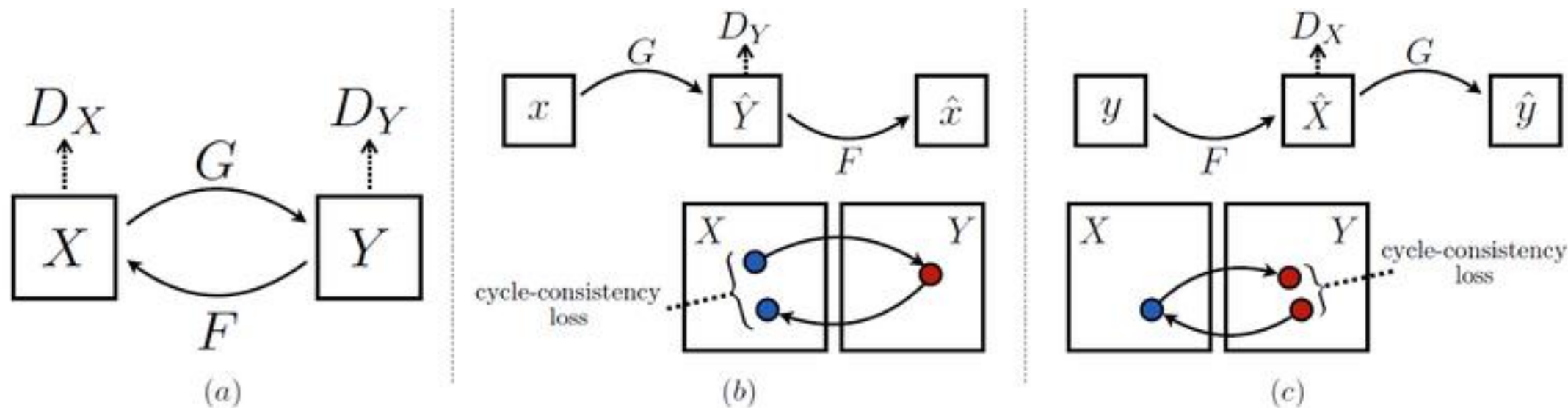


Figure 3: (a) Our model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X . D_Y encourages G to translate X into outputs indistinguishable from domain Y , and vice versa for D_X , F , and X . To further regularize the mappings, we introduce two “cycle consistency losses” that capture the intuition that if we translate from one domain to the other and back again we should arrive where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$