

StackGAN

notes written by h1astro

ICCV 2017

条件增强、两阶段的GAN

核心要点

1. 现有文本到图像方法生成的样本，可以大致表达出给定的文本含义，但是图像细节和质量不佳
2. StackGAN能基于描述，生成 256×256 分辨率的照片级图像。（还不能和ProGAN比）
3. 把问题进行了分解，采用草图绘制-精细绘制 两阶段 过程
4. 阶段1的GAN根据给定的文本描述，来绘制对象的原始形状和颜色；阶段2度GAN使用文本描述和阶段1的输出作为输入，通过纠正草图中的缺陷和细节生成，来最终得到更高分辨率的图像
5. 还提出了一种条件增强方法，能够增强潜在条件流行的平滑性（latent space 不过是从词向量得来）
6. 大量实验表明，以上方法再以文本描述为条件的照片级图像生成上取得了显著进步

研究背景

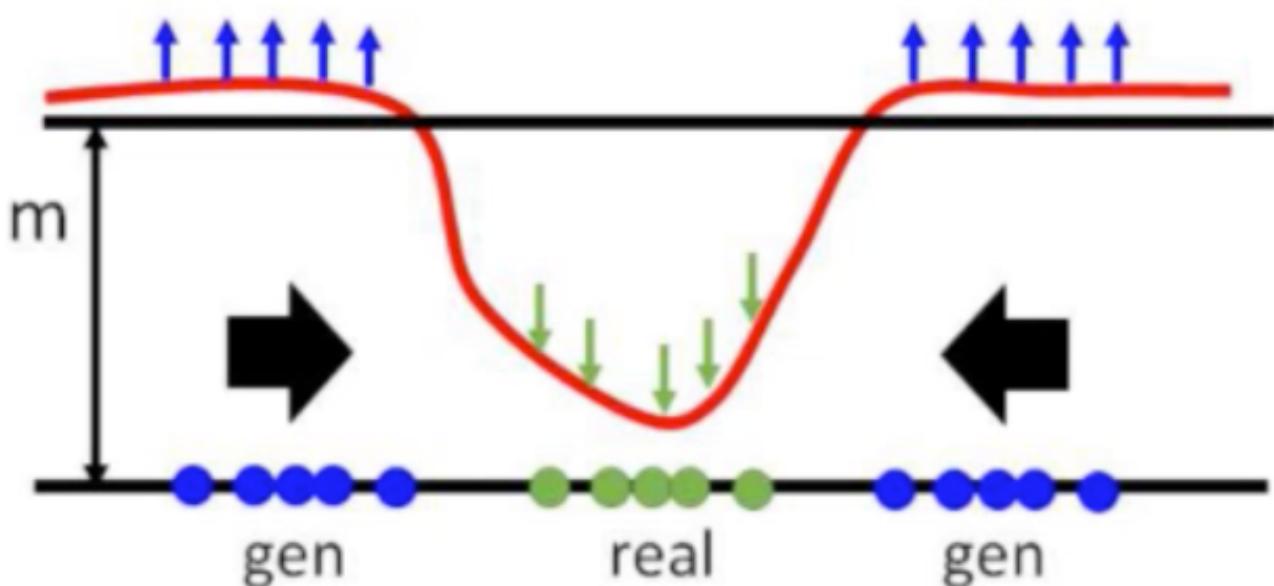
Energy-Based (EB) GAN

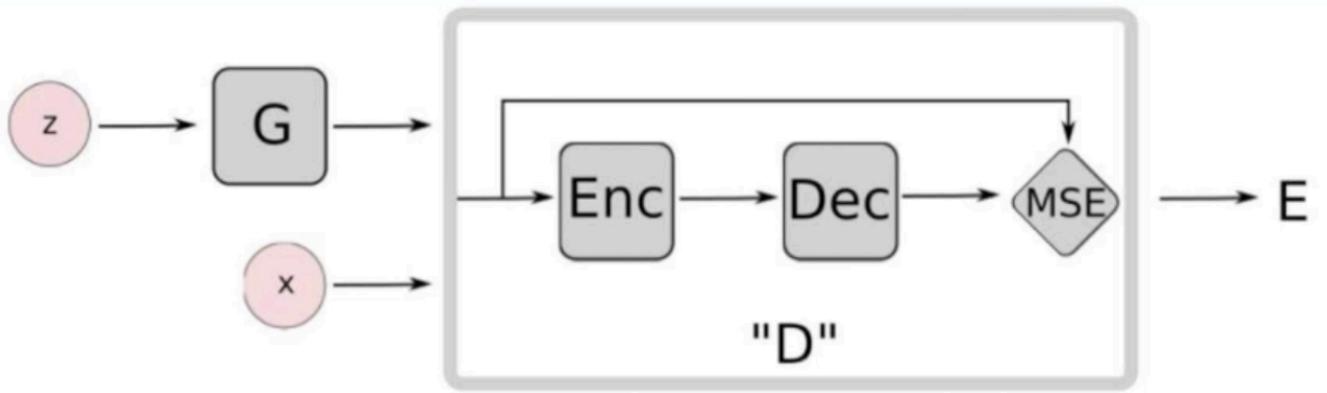
- 将判别器视为一个**energy function**，函数值（非负）越小代表data越可能是真实数据
- 使用自编码作为判别器（energy function）
- 判别器可以单独使用真实数据进行提前的预训练
- 可以基于ImageNet数据集训练，生成 256×256 分辨率的图拍呢

$$D(x) = ||Dec(Enc(x)) - x||$$

$$\mathcal{L}_D(x, z) = D(x) + [m - D(G(z))]^+ \text{ 不需要大于m，超过变0？}$$

$$\mathcal{L}_G(z) = D(G(z))$$





变分自编码器 (VAE)

- 希望根据离散变量集合 $[X_1, \dots, X_n]$ 得到 X 的原始分布 $p(X)$

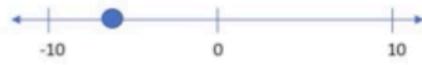
$$P(X) = \int P(X|z; \theta)P(z)dz$$

- 假设 $P(z) \sim N(0, I)$
- 对于大部分 z , $P(X|z; \theta)$ 接近 0, 不太好求解
- 改为计算 $p(Z|X)$
- 使用变分推断把对 $p(Z|X)$ 进行近似求解

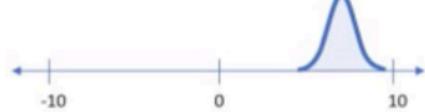
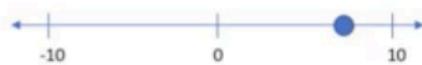
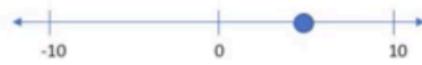
$$\mathcal{D}[Q(z)||P(z|X)] = E_{z \sim Q}[\log Q(z) - \log P(z|X)]$$



Smile (discrete value)



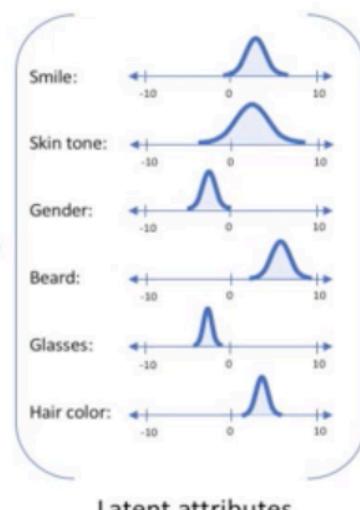
Smile (probability distribution)



vs.



encoder



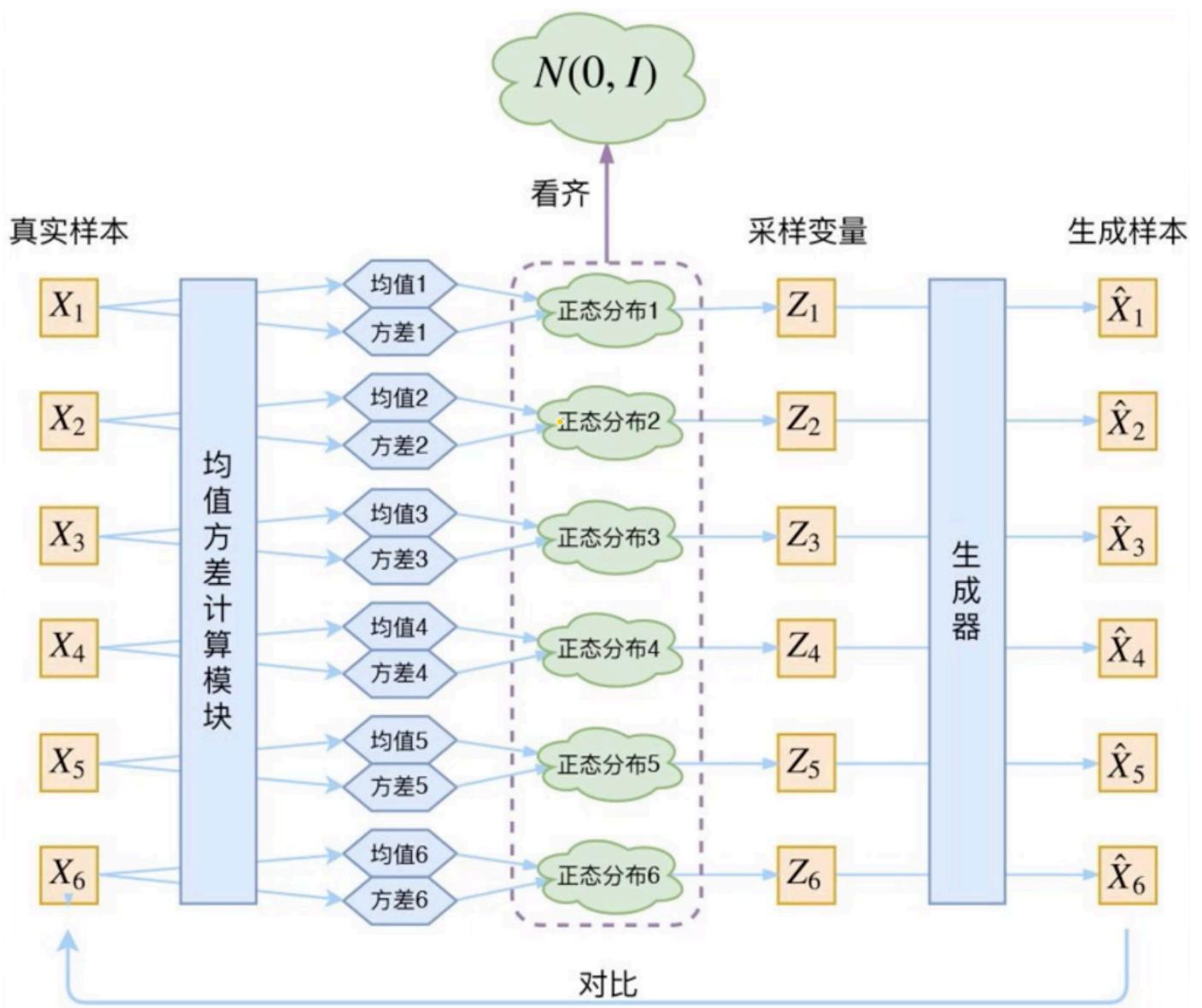
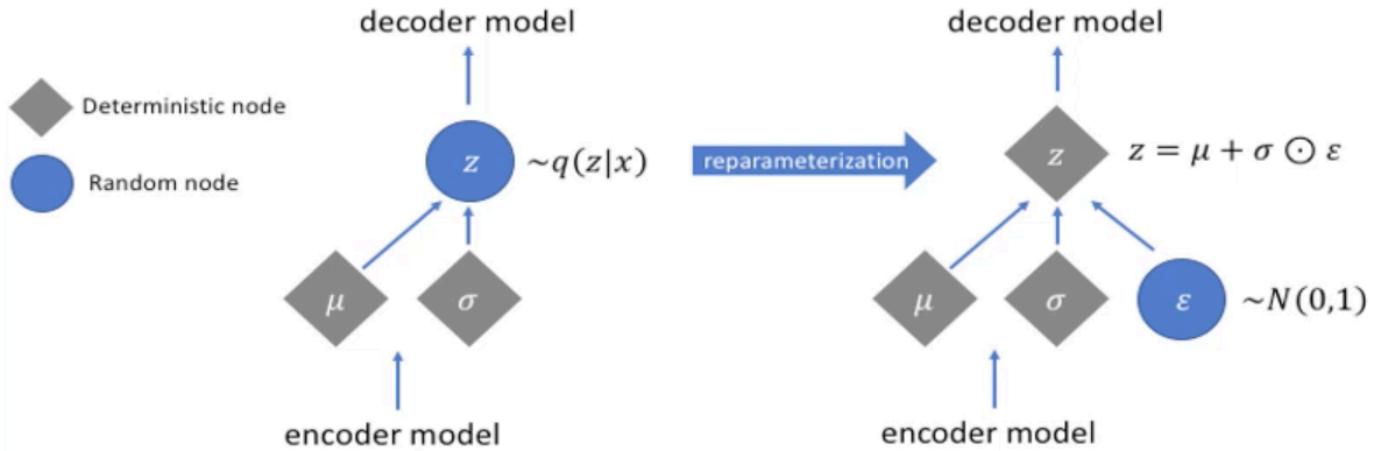
decoder



- 上述优化项经推导可改写为

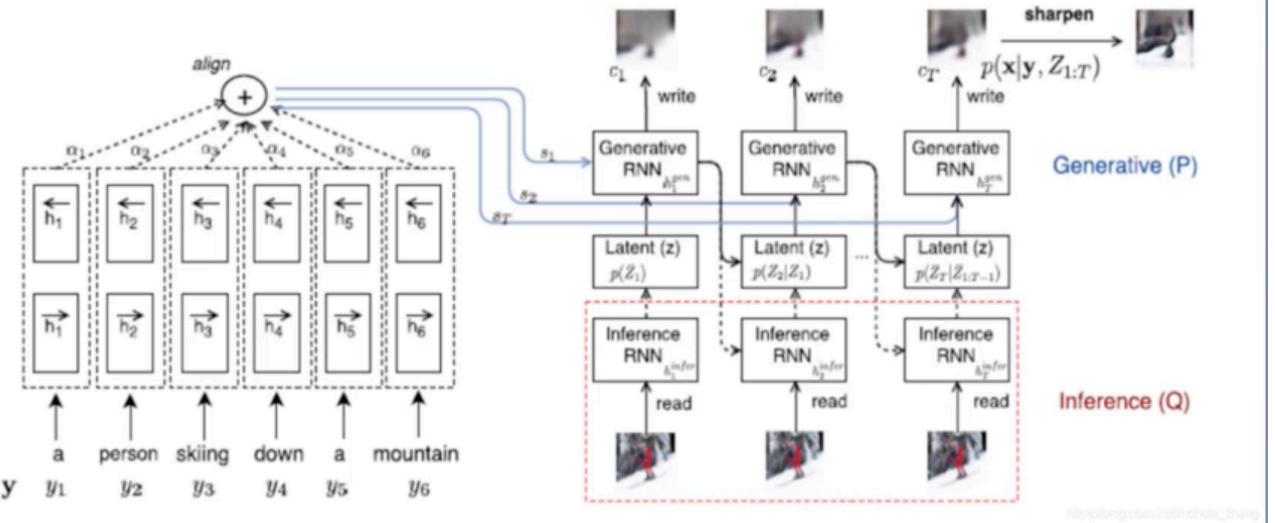
$$E_{z \sim Q} [\log P(X|z) - D[Q(z|X)||P(z)]]$$

- $P(X|z)$ 通过Decoder来学习
- 令 $Q(z|X)$ 为高斯分布，通过Encoder来学习
- 为了可导，从 $Q(z|X)$ 采样时，使用重参数化技巧



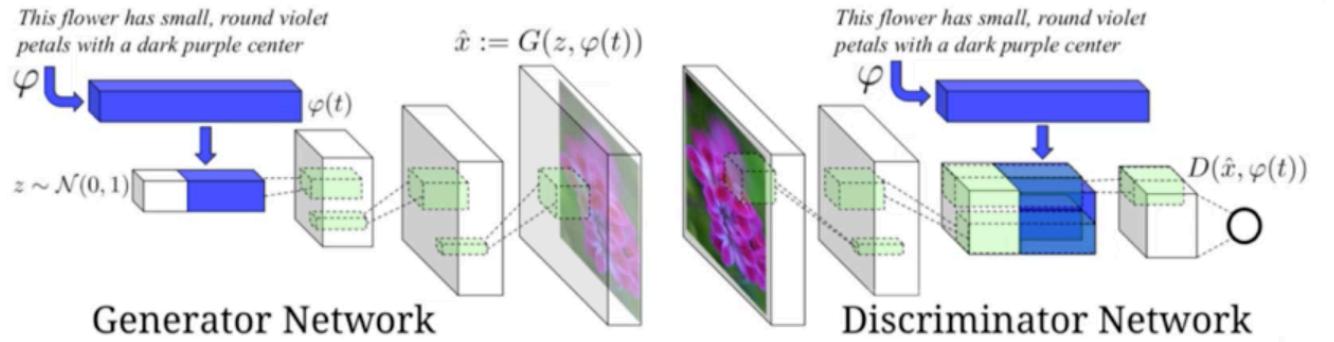
文本生成图像

- VAE
- DRAW (Deep Recurrent Attention Writer)
 - 使用循环神经网络+注意力机制
 - 依次生成一个个对象叠加在一起得到最终结果



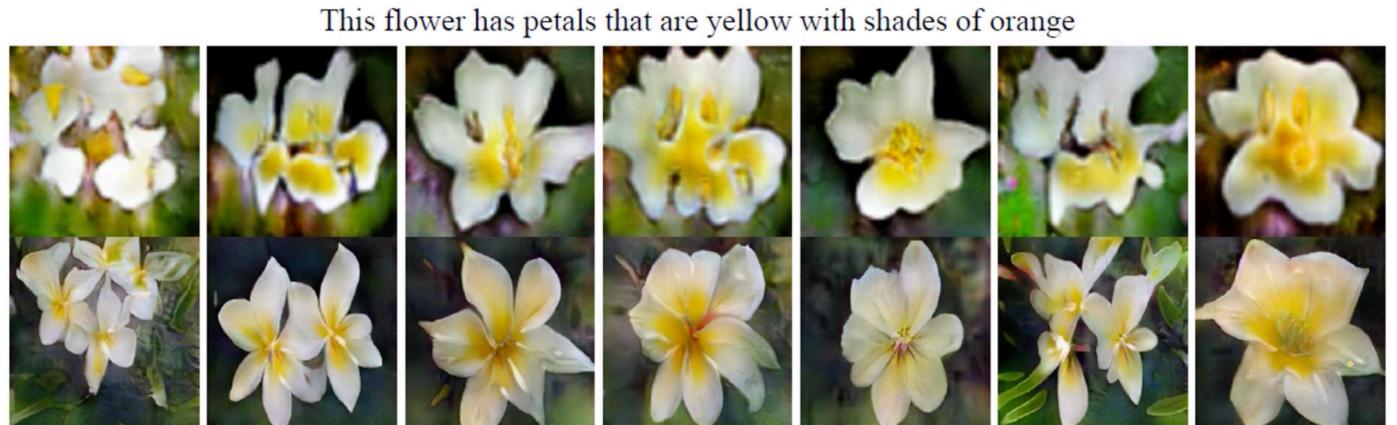
- GAN

- 在生成器中，text embedding跟随机噪声融合后一起输入到生成网络中
- 鉴别器会对错误情况进行分类，一种生成的法克图像匹配了正确的文本，另一种是真实图像但匹配了错误文本



研究成果

- 首次在文本到图像的任务中，生成了256x256分辨率的高质量图像
- 提出的条件增强方法，能增强模型的鲁棒性并提升生成效果的多样性



研究意义

- 成为了文本生成图像任务中的一个里程碑
- 基于VAE思想的条件增强方法，对之后的研究者造成了一定启发

The small bird has a red head with feathers that fade from red to gray from head to tail

