

# Diffusion扩散模型

## 前向加噪(Forward Diffusion Process)

前向加噪就是按照噪声调度对原图不断增加高斯噪声，最终变为纯噪声。该过程是一个马尔科夫过程，每一次加噪只和上次有关，每一步的递推公式为

$$x_t = \sqrt{1 - \beta_t} \cdot x_{t-1} + \sqrt{\beta_t} \cdot \epsilon_t$$

- $\beta_t$ : 第  $t$  步的噪声强度，称为“噪声调度表”(一般很小)
- $\epsilon_t \sim \mathcal{N}(0, I)$ : 标准高斯噪声

常见的闭式表达：为了加速训练，我们通常用闭式公式直接得到  $x_t$ :

$$x_t = \sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon$$

- $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$ : 原图的保留因子 (越来越小)
- $\epsilon \sim \mathcal{N}(0, I)$ : 加的一次性噪声

每步都只在上一张图的基础上加一点噪声，直到变成高斯噪声。

实际上训练的时候是一次就把噪声加上去

## 反向去噪(Reverse Denosing Process)

模型  $p_\theta$  给出了一个条件概率分布，它告诉你：

如果现在有一张加了噪声的图  $x_t$ ，那么前一时刻的图  $x_{t-1}$  的可能性分布是怎样的

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

使用预测噪声的方法来重构  $x_{t-1}$   
反向均值

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \cdot \hat{\epsilon}_\theta(x_t, t) \right)$$

预测  $x_{t-1}$  伪代码

```

for  $t = T, T-1, \dots, 1$  do
     $z \sim \mathcal{N}(0, I)$ , if  $t > 1$ ; else  $z = 0$ 
     $x_{t-1} = \mu_\theta(x_t, t) + \sigma_t \cdot z$ 
end for
return  $x_0$ 

```

其中  $\sigma_t$  是可设定的标准差。

预测混入的噪音长什么样子。

训练目标是最小化预测噪声和真实噪声之间的均方误差：

$$\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} \cdot x_0 + \sqrt{1 - \bar{\alpha}_t} \cdot \epsilon, t)\|^2$$

- $x_0$  为真实图像
- $t \sim \text{Uniform}(1, T)$  为采样的时间步
- $\epsilon \sim \mathcal{N}(0, I)$  为真实噪声

去噪过程是多次去噪的，单步去噪很难建模整个数据分布，学不到东西。多次去噪，每一步只需要处理少量噪声，训练稳定，逐步采样，可以更好训练和收敛

## 模型生成图片质量评估

### FID

生成图像的特征分布与真实图片的特征分布之间的差距

将其输入预先训练好的CNN图像分类模型，将其输出的特征分布计算Fréchet 距离 (Wasserstein-2 距离)：

$$\text{FID} = \|\boldsymbol{\mu}_r - \boldsymbol{\mu}_g\|^2 + \text{Tr}\left(\boldsymbol{\Sigma}_r + \boldsymbol{\Sigma}_g - 2(\boldsymbol{\Sigma}_r\boldsymbol{\Sigma}_g)^{\frac{1}{2}}\right)$$

FID越小越好