

DDIM

零、引入

1. $X_t = \sqrt{\sigma_t} X_{t-1} + \sqrt{1-\sigma_t} \varepsilon \sim N(0, 1)$ 中, why $\sigma_t \rightarrow 1$ 且 $\sigma_t < 1$?

Ans. 确相邻的加噪过程, 添加的噪音的方差很小, 从而保证加噪前都是正态分布.

2. why $\bar{\sigma}_T \rightarrow 0$?

Ans. $X_t = \sqrt{\sigma_T} X_0 + \sqrt{1-\sigma_T} \varepsilon \approx \varepsilon \sim N(0, 1)$, 这样我们才能从标准正态分布取一个随机噪音并基于此开始降噪.

$\Rightarrow T$ 必须要大. \Rightarrow 通过减小 T 来加速模型 is impossible.

3. 是否可以跳步?

预测目标(噪音的分布): $P(X_{t-1}|X_t, X_0) \leftarrow$ 由马尔可夫性及得
 \Rightarrow 不能跳步.

一、DDIM 的去马尔可夫化.

0. 思想: 寻找一个新的 Non-Markov 的分布 $P(x_s|x_k, x_0)$, 其中 $s < k-1$. (跳步)

* \Rightarrow 设一个不循^连Markov 的采样方程 (但有限制 $P(x_t|x_0)$ 遵循训练过程的公式), 通过待定系数经求出对应的采样方程.

$$P(x_s|x_k, x_0) \sim N\left(\sqrt{\bar{\sigma}_s} x_0 + \frac{\sqrt{1-\bar{\sigma}_s} \bar{\sigma}^2}{\sqrt{1-\bar{\sigma}_k}} (x_k - \sqrt{\bar{\sigma}_k} x_0), \bar{\sigma}^2 I\right)$$

$(s \leq k-1)$

1. 由于训练过程我们没有用 $P(x_k|x_s, x_0)$ ^① 而是用的 $P(x_t|x_0)$ ^②, 所以即使①变化了, 只要②没变, 则模型依然可以用.

二、标准差的选取.

$$\bar{\sigma} = \eta \cdot \sigma_{\text{DDPM}} \Rightarrow \eta = 1 \Rightarrow \text{DDPM}$$

$\eta = 0 \Rightarrow \text{DDIM}$ (效果最好, 但以牺牲多样性为代价)
 (图的质量)

三、思考.

1. 多样性只源自最开始随机噪声 x_T 的初始化, i.e. $t=1 \rightarrow t=0$
这个路径是固定的, 类似于 GAN 中的 Latent space, 因此该过程
可以被用作图像编码 → 迁移学习.

2. 只是改了采样方法, 并没有修改训练方法. 我们降噪过程的
目标: $\underbrace{P(x_{t-1} | x_t, x_0)}_{\text{DDPM}} \rightarrow \underbrace{P(x_s | x_k, \cancel{x_0})}_{\text{DDIM}}$, 为什么 DDPM 直接使
用 DDIM 的采样方法依然有效?

Ans. 由于 UNet 预测的是噪声 ϵ ($\epsilon \rightarrow x_0 \rightarrow P_{\text{DDPM}}$), 所以即使
分布变了, 我们依然用到 x_0 , 此时 ϵ 依旧有效. 模型找的
是 噪声 ϵ 和当前 x_T 之间的关系, 因此依然有效.

3. 当 DDPM 中 $\sigma = 0$ 时, 效果很差, 而
DDIM 中, $\sigma = 0$ 时, 效果最好. why?

Ans. 在 DDIM 中, 我们假设 $P(x_s | x_k, x_0) \sim N(kx_0 + mx_k, \sigma^2)$ 自由
随机变量, 通过解二元一次方程组, 得 k 和 m (即用 σ 表示),
经过实验发现 $\sigma = 0$ 效果最好.

DDPM 中, σ 通过贝叶斯公式, 其中三个已知的概率分布, 推导出来的, 其本身就是正确的.

① 而 DDIM 中的 σ 是不确定的, 当 DDPM 中 $\sigma = 0$, 其采样过程
变成 deterministic, 即向着均值采样, 此过程忽略了方差
的影响导致期望误差累积大 (T 很大).

4. 训练过程依赖 $P(x_t | x_0)$, 则贝叶斯公式相关的两项不能变,
只能变 $P(x_k | x_s, x_0)$