

# DDIM

## 零、引入

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

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

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

Ans.  $X_t = \sqrt{\bar{\alpha}_T} X_0 + \sqrt{1-\bar{\alpha}_T} \epsilon \approx \epsilon \sim N(0,1)$ , 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 (噪声的分布):  $P(X_{t-1} | X_t, X_0) \leftarrow$  由马尔可夫性可得  $\Rightarrow$  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)

$\Rightarrow$  设一个不遵循 Markov 的采样过程 (但有限制  $P(X_t | X_0)$  遵循训练过程的公式), 通过待定系数法求出对应的采样方程.

$$P(X_s | X_k, X_0) \sim N\left(\sqrt{\alpha_s} X_0 + \frac{\sqrt{1-\alpha_s-\beta^2}}{\sqrt{1-\alpha_k}} (X_k - \sqrt{\alpha_k} X_0), \beta^2 I\right) \quad (s \leq k-1)$$

1. 由于训练过程我们没有用  $P(X_k | X_s, X_0)$  而是用的  $P(X_t | X_0)$ , 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.

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

$\eta = 0 \Rightarrow DDIM$  (效果最好, 但以牺牲多样性为代价) (图的效果)

### 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.  $t=T \rightarrow t=0$ . 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  $\rightarrow$  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:  $\underbrace{P(x_{t-1}|x_t, x_0)}_{\text{DDPM}} \rightarrow \underbrace{P(x_s|x_k, x_0)}_{\text{DDIM}}$ , 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  $\varepsilon$  ( $\varepsilon \rightarrow x_0 \rightarrow P_{\text{DDPM}}$ ), 所以即便分布变了, 我们依然用到  $x_0$ , 此时  $\varepsilon$  依旧有效. 模型找的是 噪声  $\varepsilon$  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. ① DDIM 中, 我们假设  $P(x_s|x_k, x_0) \sim N(kx_0 + m x_k, \sigma^2)$ , 自由随机变量, 通过解二元一次方程组, 得  $k$  和  $m$  (即用  $\sigma$  表示), 经过实验发现  $\sigma=0$  效果最好.

DDPM 中,  $\sigma$  通过贝叶斯方程, 其中三个已知的概率分布, 推导出来的, 其本身就是正确的.

② 而 DDIM 中的  $\sigma$  是不确定的, 当 DDPM 中  $\sigma=0$ , 其采样过程变成 deterministic, 即向着均值  $\mu$  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  $P(x_k|x_s, x_0)$