# DDIM

参考：生成扩散模型漫谈（四）：DDIM = 高观点DDPM - 科学空间|Scientific Spaces

Insights
Training and Sampling
Accelerated Sampling
Differential Equations
Conclusion

## Insights

回顾DDPM的推导过程

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \rightarrow p(\boldsymbol{x}_t|\boldsymbol{x}_0) \rightarrow p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$$

最后的Loss由$p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$给出，然而这一项实际上仅依赖于$p(\boldsymbol{x}_t|\boldsymbol{x}_0)$，采样过程亦是如此

所以，实际上我们只需要知道$p(\boldsymbol{x}_t|\boldsymbol{x}_0)$便可以确定整个训练和推理流程，而无需了解中间加噪过程$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$

## Training and Sampling

现在我们的出发点是

$$\boldsymbol{x}_t = \bar{\alpha}_t \boldsymbol{x}_0 + \bar{\beta}_t \boldsymbol{\epsilon} \tag{1}$$

我们需要求解$p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$

不妨设

$$p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) = \kappa_t \boldsymbol{x}_t + \lambda_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}$$

注意到

$$p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0) = \int p(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) p(\boldsymbol{x}_t|\boldsymbol{x}_0) \mathrm{d}\boldsymbol{x}_t$$

其中左手边为

$$\boldsymbol{x}_{t-1} = \bar{\alpha}_{t-1}\boldsymbol{x}_0 + \bar{\beta}_{t-1}\boldsymbol{\epsilon}$$

右手边为

$$\begin{aligned}
\boldsymbol{x}_{t-1} &= \kappa_t \boldsymbol{x}_t + \lambda \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon} \\
&= \kappa_t(\bar{\alpha}_t \boldsymbol{x}_0 + \bar{\beta}_t \boldsymbol{\epsilon}_1) + \lambda_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}_2 \\
&= (\kappa_t \bar{\alpha}_t + \lambda_t)\boldsymbol{x}_0 + (\kappa_t \bar{\beta}_t \boldsymbol{\epsilon}_1 + \sigma_t \boldsymbol{\epsilon}_2) \\
&= (\kappa_t \bar{\alpha}_t + \lambda_t)\boldsymbol{x}_0 + \sqrt{\kappa_t^2 \bar{\beta}_t^2 + \sigma_t^2}\boldsymbol{\epsilon}
\end{aligned}$$

于是有

$$\begin{aligned}
\bar{\alpha}_{t-1} &= \kappa_t \bar{\alpha}_t + \lambda_t \\
\bar{\beta}_{t-1} &= \sqrt{\kappa_t^2 \bar{\beta}_t^2 + \sigma_t^2}
\end{aligned}$$

取$\sigma_t$为自由参数，解得

$$\begin{aligned}
\kappa_t &= \frac{\sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}}{\bar{\beta}_t} \\
\lambda_t &= \bar{\alpha}_{t-1} - \frac{\bar{\alpha}_t \sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}}{\bar{\beta}_t}
\end{aligned}$$

因此有

$$\boldsymbol{x}_{t-1} = \frac{\sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}}{\bar{\beta}_t}\boldsymbol{x}_t + \left(\bar{\alpha}_{t-1} - \frac{\bar{\alpha}_t \sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}}{\bar{\beta}_t}\right)\boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon} \tag{2}$$

为简化参数设置，规定

$$\begin{aligned}
\bar{\alpha}_0 = 1, \bar{\beta}_0 = 0 \\
\bar{\alpha}_t^2 + \bar{\beta}_t^2 = 1
\end{aligned}$$

虽然我们得到了采样方程(2)，但是其中的$\boldsymbol{x}_0$是未知的，因此，我们不妨利用加噪图$\boldsymbol{x}_t$来预测原图$\boldsymbol{x}_0$，这等价于一个去噪过程(利用(1))

$$\bar{\mu}(\boldsymbol{x}_t) = \frac{1}{\bar{\alpha}_t}(\boldsymbol{x}_t - \bar{\beta}_t\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t))$$

代入(2)得到

$$\begin{aligned}\boldsymbol{x}_{t-1} &= \frac{\sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}}{\bar{\beta}_t}\boldsymbol{x}_t + \frac{1}{\bar{\alpha}_t}\left(\bar{\alpha}_{t-1} - \frac{\bar{\alpha}_t\sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}}{\bar{\beta}_t}\right)(\boldsymbol{x}_t - \bar{\beta}_t\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)) + \sigma_t\boldsymbol{\epsilon} \\ &= \frac{\bar{\alpha}_{t-1}}{\bar{\alpha}_t}\left[\boldsymbol{x}_t - \left(\bar{\beta}_t + \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}}\sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}\right)\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right] + \sigma_t\boldsymbol{\epsilon}\end{aligned}$$

此时，我们训练$\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)$使用的Loss依旧为

$$\mathbb{E}_{\boldsymbol{x}_0, t, \boldsymbol{\epsilon}}\left[\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t) - \boldsymbol{\epsilon}\|^2\right]$$

其中

$$\boldsymbol{x}_t = \bar{\alpha}_t\boldsymbol{x}_0 + \bar{\beta}_t\boldsymbol{\epsilon}$$

也就是说，DDIM并不改变DDPM的训练过程

$\sigma_t$的选择是自由的，DDIM原文特别强调了$\sigma_t = 0$这种情况，此时从噪声到图像的过程是确定性的(对应一个ODE)

## Accelerated Sampling

注意到，DDIM去除了$\boldsymbol{x}_t$对$\boldsymbol{x}_{t-1}$的单步依赖关系(去除了Markov假设)，因此我们得以作“跳步”采样

所谓“跳步”，是指当我们拟合好一个长序列之后，我们可以跳过中间的某些时间步而得到一个短序列采样。具体来说，假设我们拟合了$[1, 2, \ldots, T]$上的DDPM模型(对应时间步的去噪模型)，设$\tau = [\tau_1, \tau_2, \ldots, \tau_{\dim_\tau}]$为$[1, 2, \ldots, T]$的子列，则我们同样完成了对$\tau$对应的DDPM的拟合！！！

于是，我们可以用$[1, 2, \ldots, T]$来作训练而用$\tau$来作采样，采样步数自然就从T步减少到$\dim_\tau$步，此时的采样方程变为

$$\boldsymbol{x}_{\tau_{i-1}} = \frac{\bar{\alpha}_{\tau_{i-1}}}{\bar{\alpha}_{\tau_i}}\left[\boldsymbol{x}_{\tau_i} - \left(\bar{\beta}_{\tau_i} + \frac{\bar{\alpha}_{\tau_i}}{\bar{\alpha}_{\tau_{i-1}}}\sqrt{\bar{\beta}_{\tau_{i-1}}^2 - \sigma_{\tau_i}^2}\right)\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{\tau_i}, \tau_i)\right] + \sigma_{\tau_i}\boldsymbol{\epsilon}$$

之所以用$[1, 2, \ldots, T]$来训练而不是直接训练$\tau$对应的DDPM，是为了增加泛化能力

如果从SDE的角度来理解加速采样其实更简单，训练$T$步DDPM相当于拟合$[0,1]$上的对数概率密度梯度场，在得到梯度场的拟合后我们就可以用SDE/ODE的数值解法来作采样，数值解的离散步数即采样步数

## Differential Equations

来研究一下$\sigma_t \equiv 0$的情形

此时，采样方程化为

$$\frac{\boldsymbol{x}_t}{\bar{\alpha}_t} - \frac{\boldsymbol{x}_{t-1}}{\bar{\alpha}_{t-1}} = \left(\frac{\bar{\beta}_t}{\bar{\alpha}_t} - \frac{\bar{\beta}_{t-1}}{\bar{\alpha}_{t-1}}\right) \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)$$

令$T \to +\infty$，得到

$$\frac{\mathrm{d}}{\mathrm{d}s} \frac{\boldsymbol{x}_s}{\bar{\alpha}(s)} = \boldsymbol{\epsilon}_\theta(\boldsymbol{x}(s), s) \frac{\mathrm{d}}{\mathrm{d}s} \frac{\bar{\beta}(s)}{\bar{\alpha}(s)} \tag{3}$$

其中$s \in [0,1]$，此即DDIM对应的ODE

以DDPM默认参数为例

$$T = 1000, \alpha_t = \sqrt{1 - \frac{0.02t}{T}}$$

则有以下估计

$$\log \bar{\alpha}_t = \sum_{i=1}^{t} \log \alpha_i = \sum_{i=1}^{t} \log \sqrt{1 - \frac{0.02i}{T}} \approx \frac{1}{2} \sum_{i=1}^{t} \left(-\frac{0.02i}{T}\right) = -\frac{0.005t(t+1)}{T}$$

于是有

$$\bar{\alpha}_t \approx \exp\left(-\frac{0.005t(t+1)}{T}\right) \approx \exp\left(-\frac{5t^2}{T^2}\right)$$

取$s = \frac{t^2}{T^2} \in [0,1]$，有

$$\bar{\alpha}(s) = \exp(-5s) = e^{-5s}$$

对应有

$$\bar{\beta}(s) = \sqrt{1 - e^{-10s}}$$

代入(3)有

$$\frac{\mathrm{d}\boldsymbol{x}_s}{\mathrm{d}s} = 5\left(\frac{\boldsymbol{\epsilon}_\theta(\boldsymbol{x}(s), T\sqrt{s})}{\sqrt{1-e^{-10s}}} - \boldsymbol{x}(s)\right)$$

## Conclusion

DDIM去掉DDPM的Markov加噪过程，增加从核心点$p(\boldsymbol{x}_t|\boldsymbol{x}_0)$入手，获得了更一般的采样方程并允许在不改变训练过程的情形下作加速采样，最后成功与ODE建立联系