

目录

1	Karras 的 EDM	2
1.1	扩散模型的通用加噪公式	2
1.2	一步写出 PF-ODE	4
1.3	通用确定性采样	5

1 Karras 的 EDM

1.1 扩散模型的通用加噪公式

1. Flow Matching

Flow Matching 的一步加噪公式:

$$\mathbf{x}_t = (1 - t)\mathbf{x}_0 + t\varepsilon$$

其中 \mathbf{x}_0 是原始图像, ε 是高斯噪声, $\varepsilon \sim \mathcal{N}(0, \mathbf{I})$

写成概率分布形式

$$p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left((1 - t)\mathbf{x}_0, t^2\mathbf{I}\right)$$

2. Score Matching

Score Matching 的一步加噪公式:

$$\mathbf{x}_t = \mathbf{x}_0 + \sigma_t\varepsilon, \quad \text{其中}\sigma_t\text{表示}\sigma\text{是}t\text{的函数}$$

写成概率分布形式

$$p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_0, \sigma_t^2\mathbf{I}\right)$$

3. DDPM 与 DDIM

DDPM 与 DDIM 的一步加噪公式:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\varepsilon$$

写成概率分布形式

$$p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)\mathbf{I}\right)$$

4. 通用加噪公式

对比上述三个概率形式的加噪公式, 可以将这三个式子建模为:

$$p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left(s(t)\mathbf{x}_0, s^2(t)\sigma^2(t)\mathbf{I}\right)$$

上述的通用形式对应一个随机微分方程 (SDE), 这个方程的解可以描述 \mathbf{x}_t 的分布的变化.

$$d\mathbf{x}_t = f(\mathbf{x}_t, t)dt + g(t)d\mathbf{w}$$

根据 DDPM 和 SMLD 的推导结果有: $f(\mathbf{x}_t, t) = f(t)\mathbf{x}_t$, 并且 $f(t) : \mathbb{R}^1 \rightarrow \mathbb{R}^1$

因此此时的 SDE 变为:

$$d\mathbf{x}_t = f(t)\mathbf{x}_t dt + g(t)d\mathbf{w}$$

现在是已知 $f(t), g(t)$, 推导出通用加噪公式中的 $s(t), \sigma(t)$, 即正态分布的期望和方差. 经过一系列的推导可以得到:

$$\begin{cases} s(t) = e^{\int_0^t f(r)dr} \\ s^2(t) = e^{\int_0^t 2f(r)dr} \\ \frac{1}{s^2(t)} = e^{\int_0^t -2f(r)dr} \\ \sigma(t) = \int_0^t \frac{g^2(r)}{s^2(r)} dr \end{cases}$$

此时意味着只要知道如下一个 SDE 的 $f(t), g(t)$:

$$d\mathbf{x}_t = f(t)\mathbf{x}_t dt + g(t)d\mathbf{w}$$

便可以得到一个扩散模型的加噪公式:

$$p(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left(s(t)\mathbf{x}_0, s^2(t)\sigma^2(t)\mathbf{I}\right)$$

加噪公式中的 \mathbf{x}_0 是原始图像, $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}_0, 0)$

同样的, 只要知道 $s(t), \sigma(t)$ 也可以推导出 $f(t), g(t)$

至此, 扩散模型大一统!

1.2 一步写出 PF-ODE

这个 section 想要实现只要给出一个通用的加噪公式:

$$p(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}\left(s(t)\mathbf{x}_0, s^2(t)\sigma^2(t)\mathbf{I}\right)$$

哪怕 SDE 里的 $f(\cdot), g(\cdot)$ 不知道, 都可以直接写出对应的 PF-ODE.

写出 PF-ODE 后, 就可以根据 Euler 方法求解, 从一个噪声图像 \mathbf{x}_t 求出原始图像 \mathbf{x}_0 .

回顾上一个 section 得到的 SDE, 通过一个超级复杂的推导过程, 可以通过 section 得到的 SDE 得到对应的反向去噪 ODE(概率流常微分方程 PF-ODE):

$$d\mathbf{x}_t = \left[f(t)\mathbf{x}_t - \frac{1}{2}g^2(t)\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \right] dt$$

首先根据 $f(t), g(t)$ 和 $\sigma(t), s(t)$ 的关系可得:

$$\begin{cases} s(t) = e^{\int_0^t f(r)dr} \\ \ln(s(t)) = \int_0^t f(r)dr \end{cases} \implies f(t) = \frac{\dot{s}(t)}{s(t)}$$

$$\begin{cases} \sigma(t) = \sqrt{\int_0^t \frac{g^2(r)}{s^2(r)} dr} \\ \sigma^2(t) = \int_0^t \frac{g^2(r)}{s^2(r)} dr \end{cases} \implies g(t) = s(t)\sqrt{2\sigma(t)\dot{\sigma}(t)}$$

看似已经完成了目标但是在反向去噪的 PF-ODE 中还有一个梯度项 $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$, 也就是说现在最大的问题是 $p_t(\mathbf{x}_t)$ 是未知的, 只要知道了 $p_t(\mathbf{x}_t)$, 就可以直接采样了!

于是经过了漫长的推导过程:

$$p_t(\mathbf{x}_t) = s^{-d}(t) [p_{data} * \mathcal{N}(0, \sigma^2(t)\mathbf{I})] \left(\frac{\mathbf{x}_t}{s(t)} \right)$$

上式中 $*$ 表示卷积, d 是空间的维数. 此时把得到的式子全部带入到反向去噪的 PF-ODE:

$$\begin{aligned} d\mathbf{x}_t &= \left[f(t)\mathbf{x}_t - \frac{1}{2}g^2(t)\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \right] dt \\ &= \left[\frac{\dot{s}(t)}{s(t)}\mathbf{x}_t - s^2(t)\sigma(t)\dot{\sigma}(t)\nabla_{\mathbf{x}_t} \log \left(s^{-d}(t) [p_{data} * \mathcal{N}(0, \sigma^2(t)\mathbf{I})] \left(\frac{\mathbf{x}_t}{s(t)} \right) \right) \right] dt \\ &= \left[\frac{\dot{s}(t)}{s(t)}\mathbf{x}_t - s^2(t)\sigma(t)\dot{\sigma}(t) \left(\nabla_{\mathbf{x}_t} \log s^{-d}(t) + \nabla_{\mathbf{x}_t} \log p \left(\frac{\mathbf{x}_t}{s(t)}; \sigma(t) \right) \right) \right] dt \\ &= \left[\frac{\dot{s}(t)}{s(t)}\mathbf{x}_t - s^2(t)\sigma(t)\dot{\sigma}(t)\nabla_{\mathbf{x}_t} \log p \left(\frac{\mathbf{x}_t}{s(t)}; \sigma(t) \right) \right] dt \end{aligned}$$

上式中, 仅仅依赖 $s(t), \sigma(t)$ 就可以直接写出 PF-ODE, 而不显示依赖 $f(t), g(t)$, 此时可以根据 Euler 等数值方法进行确定性采样.

但是此时还是要预测 score.

1.3 通用确定性采样

1.2 中的问题是要预测 score, 各种扩散模型在这一步的操作如下

$$\begin{cases} \text{SMLD} : s_\theta \longrightarrow \nabla_{\mathbf{x}_t}(\log(\cdot)) \\ \text{DDPM/DDIM} : \varepsilon_\theta \longrightarrow \varepsilon \\ \text{FM/RF} : v_\theta \longrightarrow v \end{cases}$$

这个 section 的目的是想把上面这些方法也统一起来. 用一个 D_θ 作为去噪网络. D_θ 的输入是噪声, 和时间步.

更换符号:

D_θ 的输入 $\mathbf{x} = \mathbf{y} + \mathbf{n}$, \mathbf{y} 是原始图像, \mathbf{n} 为噪声, σ 是噪声强度, 和时间等价, 因此训练的 Loss 为:

$$\mathbb{E}_{\mathbf{y} \sim p_{data}, \mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})} \|D_\theta(\mathbf{y} + \mathbf{n}; \sigma) - \mathbf{y}\|_2^2$$

去噪网络最理想的状态:

$$D_\theta(\mathbf{y} + \mathbf{n}; \sigma) = \mathbf{y}$$