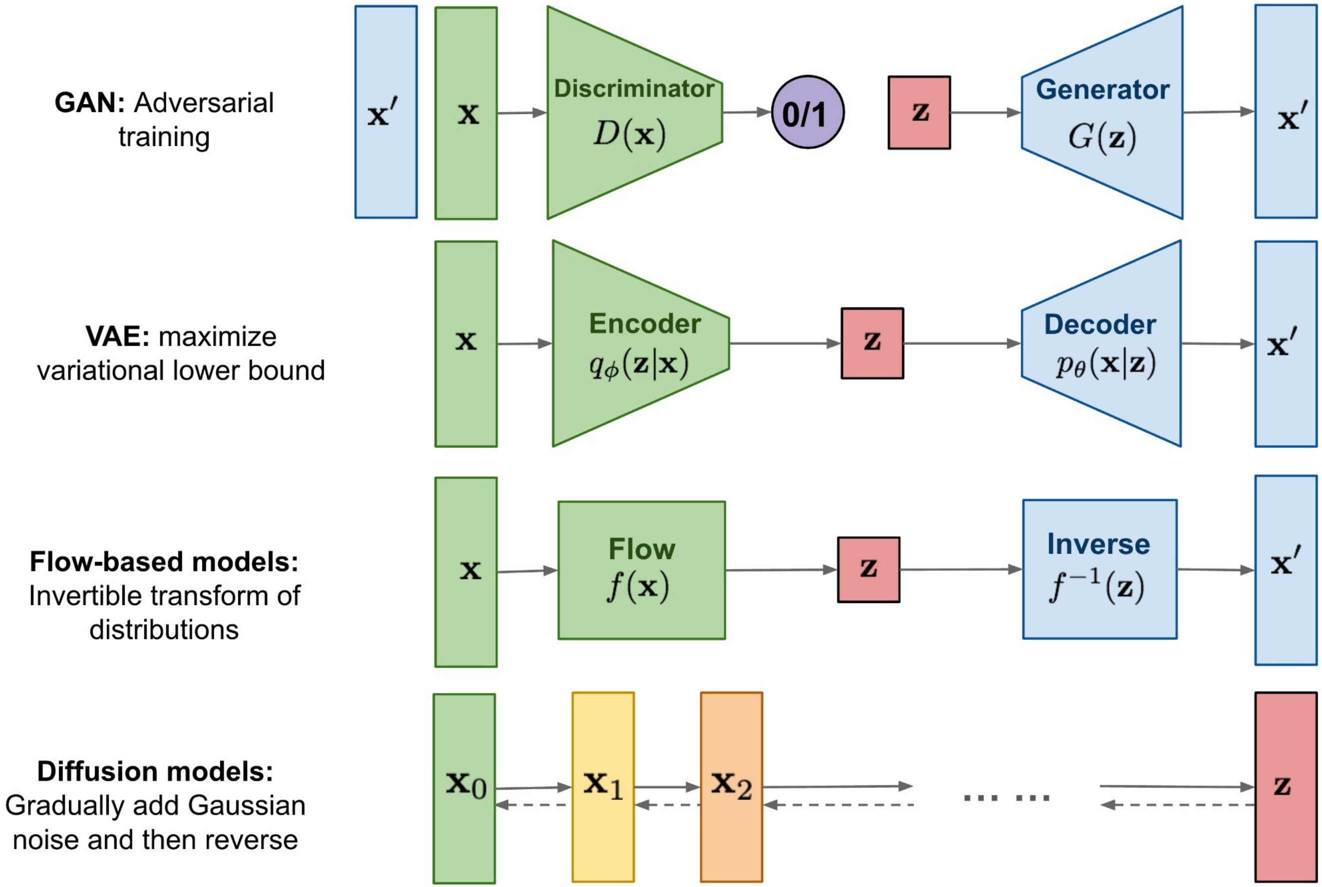
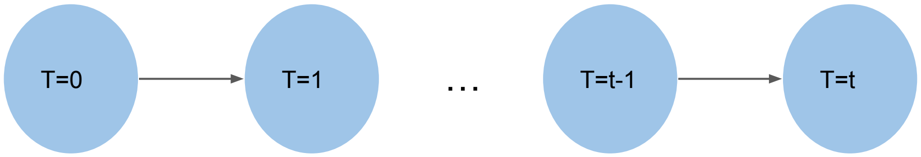
Diffusion Models



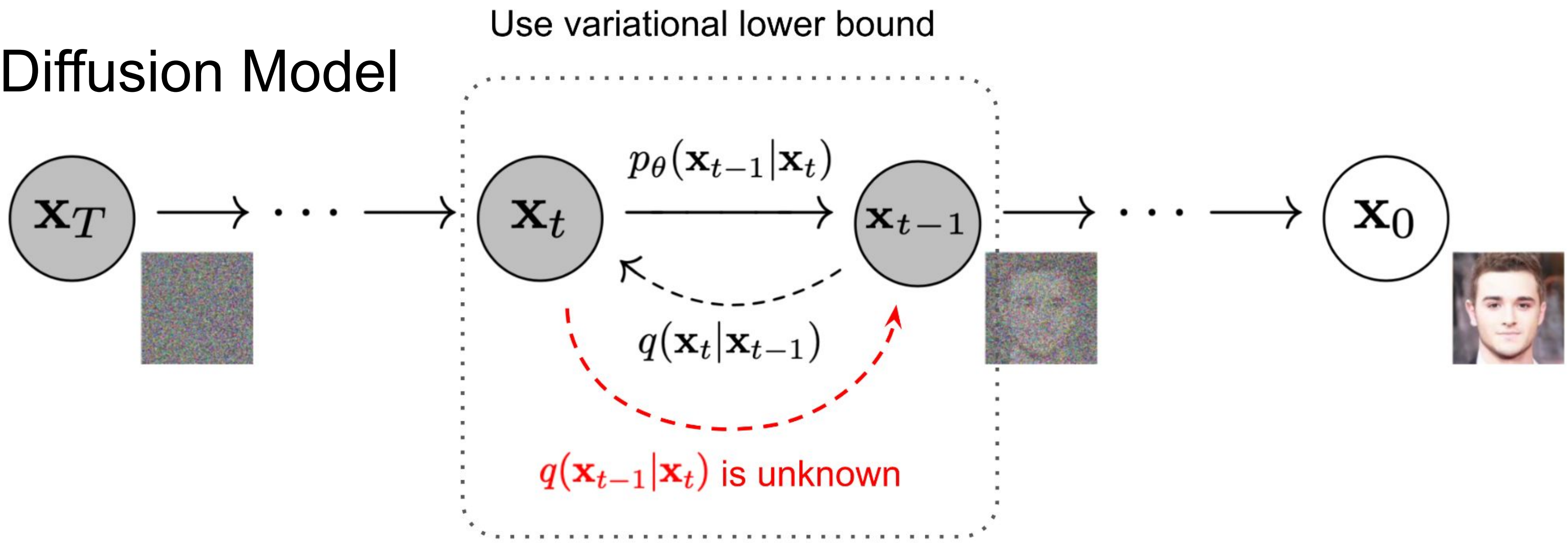
What are Diffusion Models?

Markov Chain



$$P(x_t \mid x_0, x_1, \ \ldots, x_{t-1}) = P(x_t \mid x_{t-1})$$

Diffusion Model



What are Diffusion Models?

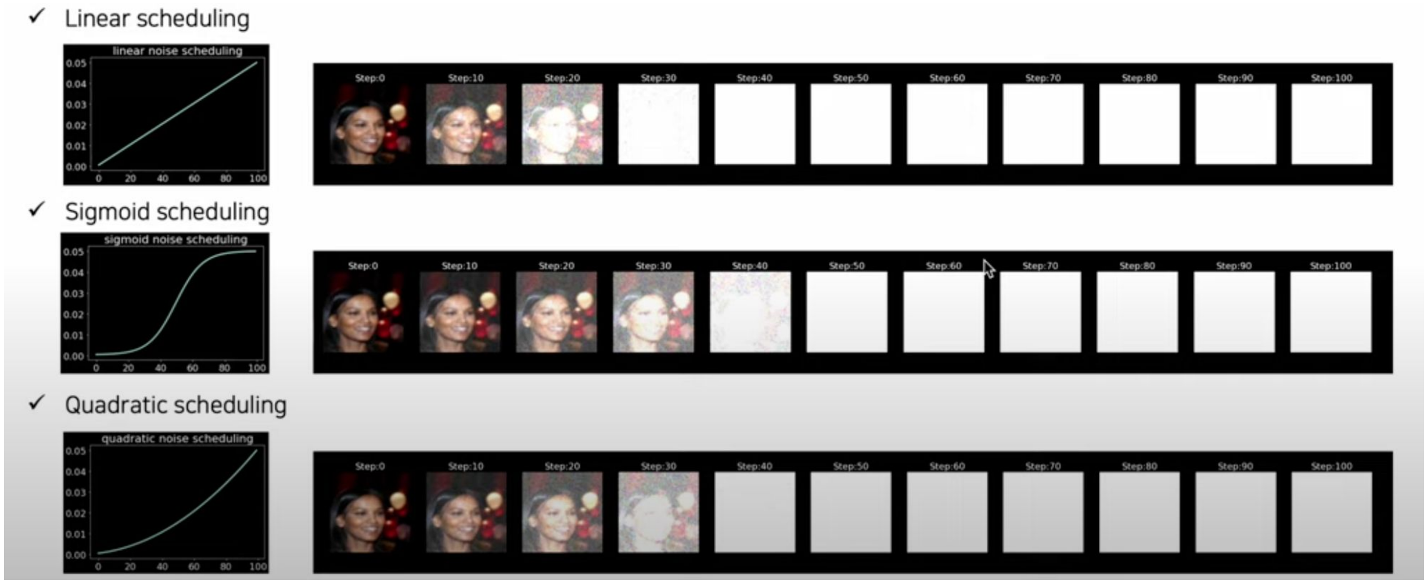
$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1},\, \beta_t I\right) \quad q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1})$$

$$p(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_{t-1}, t),\, \Sigma_\theta(x_t, t)) \quad p(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t)$$

Diffusion Process

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \quad q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1})$$

$$x_t = \sqrt{(1-\beta_t)}x_{t-1} + \sqrt{(\beta_t)}\epsilon_{t-1},\ \epsilon_t \sim \mathcal{N}(0, I)$$



Diffusion Process

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1},\, \beta_t I\right) \quad q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1})$$

$$x_t = \sqrt{(1-\beta_t)}x_{t-1} + \sqrt{(\beta_t)}\epsilon_{t-1},\ \epsilon_t \sim \mathcal{N}(0, I)$$

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1},\ \text{where}\ \alpha_t = 1-\beta_t$$

$$x_{t-1} = \sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon_{t-2}$$

$$\vdots$$

$$x_2 = \sqrt{\alpha_2}x_1 + \sqrt{1-\alpha_2}\epsilon_1$$

$$x_1 = \sqrt{\alpha_1}x_0 + \sqrt{1-\alpha_1}\epsilon_0$$

$$\begin{aligned} x_2 &= \sqrt{\alpha_2}\left(\sqrt{\alpha_1}x_0 + \sqrt{1-\alpha_1}\epsilon_0\right) + \sqrt{1-\alpha_2}\epsilon_1 \\ &= \sqrt{\alpha_2}\sqrt{\alpha_1}x_0 + \sqrt{\alpha_2}\sqrt{1-\alpha_1}\epsilon_0 + \sqrt{1-\alpha_2}\epsilon_1 \end{aligned}$$

Diffusion Process

$$q(x_t \mid x_{t-1}) \quad \Rightarrow \quad q(x_t \mid x_0)$$

$$\begin{aligned} x_2 &= \sqrt{\alpha_2}\left(\sqrt{\alpha_1}x_0 + \sqrt{1-\alpha_1}\epsilon_0\right) + \sqrt{1-\alpha_2}\epsilon_1 \\ &= \sqrt{\alpha_2}\sqrt{\alpha_1}x_0 + \sqrt{\alpha_2}\sqrt{1-\alpha_1}\epsilon_0 + \sqrt{1-\alpha_2}\epsilon_1 \end{aligned}$$

If $\epsilon_0, \epsilon_1 \sim \mathcal{N}(0,1)$ and $\epsilon_0,\ \epsilon_1$ are indpendent , then what is variance of $\sigma_1\epsilon_0 + \sigma_2\epsilon_1$?

$$\mathbb{V}[\sigma_1\epsilon_0 + \sigma_2\epsilon_1] = \mathbb{E}\left[(\sigma_1\epsilon_0 + \sigma_2\epsilon_1)^2\right] = \mathbb{E}\left[\sigma_1^2\epsilon_0^2 + \sigma_2^2\epsilon_1^2 + 2\sigma_1\sigma_2\epsilon_0\epsilon_1\right]$$

$$= \mathbb{E}\left[\sigma_1^2\epsilon_0^2\right] + \mathbb{E}\left[\sigma_2^2\epsilon_1^2\right] + \mathbb{E}[2\sigma_1\sigma_2\epsilon_0\epsilon_1]$$

$$= \sigma_1^2 \underbrace{\mathbb{E}\left[\epsilon_0^2\right]}_{=\mathbb{V}[\epsilon_0]=1} + \sigma_2^2 \underbrace{\mathbb{E}\left[\epsilon_1^2\right]}_{=\mathbb{V}[\epsilon_1]=1} + 2\sigma_1\sigma_2 \underbrace{\mathbb{E}[\epsilon_0\epsilon_1]}_{\epsilon_0,\epsilon_1 \text{are indep.}}$$

$$= \sigma_1^2 + \sigma_2^2$$

$$\begin{aligned} x_2 &= \sqrt{\alpha_2}\sqrt{\alpha_1}x_0 + \sqrt{\alpha_2}\sqrt{1-\alpha_1}\epsilon_0 + \sqrt{1-\alpha_2}\epsilon_1 \\ &= \sqrt{\alpha_2}\sqrt{\alpha_1}x_0 + \sqrt{(1-\alpha_1)\alpha_2 + 1 - \alpha_2}\epsilon \\ &= \sqrt{\alpha_2}\sqrt{\alpha_1}x_0 + \sqrt{1-\alpha_1\alpha_2}\epsilon \end{aligned}$$

Diffusion Process

$$
\begin{aligned}
x_3 &= \sqrt{\alpha_3} x_2 + \sqrt{1-\alpha_3}\epsilon_2 \\
&= \sqrt{\alpha_3}\left(\sqrt{\alpha_2} x_1 + \sqrt{1-\alpha_2}\epsilon_1\right) + \sqrt{1-\alpha_3}\epsilon_2 \\
&= \sqrt{\alpha_3}\sqrt{\alpha_2} x_1 + \sqrt{\alpha_3}\sqrt{1-\alpha_2}\epsilon_1 + \sqrt{1-\alpha_3}\epsilon_2 \\
&= \sqrt{\alpha_3}\sqrt{\alpha_2}\left(\sqrt{\alpha_1} x_0 + \sqrt{1-\alpha_1}\epsilon_0\right) + \sqrt{1-\alpha_2\alpha_3}\epsilon \\
&= \sqrt{\alpha_3}\sqrt{\alpha_2}\sqrt{\alpha_1} x_0 + \sqrt{\alpha_3}\sqrt{\alpha_2}\sqrt{1-\alpha_1}\epsilon_0 + \sqrt{1-\alpha_2\alpha_3}\epsilon \\
&= \sqrt{\alpha_3}\sqrt{\alpha_2}\sqrt{\alpha_1} x_0 + \sqrt{1-\alpha_1\alpha_2\alpha_3}\epsilon
\end{aligned}
$$

$$
\begin{aligned}
x_t &= \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1} \\
&= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}} x_{t-2} + \sqrt{1-\alpha_{t-1}}\epsilon_{t-2}\right) + \sqrt{1-\alpha_t}\epsilon_{t-1} \\
&\;\;\vdots \\
&= \sqrt{\alpha_t}\sqrt{\alpha_{t-1}}\ldots\sqrt{\alpha_2}\sqrt{\alpha_1} x_0 + \sqrt{1-\alpha_1\alpha_2\ldots\alpha_{t-1}\alpha_t}\epsilon \\
&= \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon \;\text{ where } \bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s
\end{aligned}
$$

Diffusion Process

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t} x_{t-1},\ \beta_t I\right) \quad q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1})$$

$$x_t = \sqrt{(1-\beta_t)} x_{t-1} + \sqrt{(\beta_t)} \epsilon_{t-1},\ \epsilon_t \sim \mathcal{N}(0, I)$$

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{(1-\bar{\alpha}_t)} \epsilon, \epsilon \sim \mathcal{N}(0, I)$$

$$q(x_t \mid x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t} x_0,\ (1-\bar{\alpha}_t) I\right)$$

$$\text{where}\ \alpha_t = 1 - \beta_t,\ \text{and}\ \bar{\alpha}_t = \prod_{t=i}^{T} \alpha_i$$

Reverse Process

$$
\begin{aligned}
p(x_{t-1} \mid x_t) &= \mathcal{N}(x_{t-1}; \mu_\theta(x_{t-1}, t),\ \Sigma_\theta(x_t, t)) \\
&= \mathcal{N}\big(x_{t-1}; \mu_\theta(x_{t-1}, t),\ \sigma_t^2 I\big)
\end{aligned}
$$

Loss

$$L_{CE} = -\mathbb{E}_{q(x_0)}[\log p_\theta(x_0)] = -\mathbb{E}_{q(x_0)}\left[\log\left(\boxed{\int p_\theta(x_{0:T})dx_{1:T}}\right)\right]$$

역으로 생각해보면, marginalization

$$= -\mathbb{E}_{q(x_0)}\left[\log\left(\int p_\theta(x_{0:T})\boxed{\frac{q(x_{1:T} \mid x_0)}{q(x_{1:T} \mid x_0)}}dx_{1:T}\right)\right]$$

같은 값으로 곱하고, 나누어도 결과는 달라지지 않음

$$= -\mathbb{E}_{q(x_0)}\left[\log\left(\int \boxed{q(x_{1:T} \mid x_0)}\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}dx_{1:T}\right)\right]$$

식을 정리하기 위해 위치를 이동

$$= -\mathbb{E}_{q(x_0)}\left[\log\left(\boxed{\mathbb{E}_{q(x_{1:T} \mid x_0)}\left[\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right]}\right)\right]$$

평균의 정의로 형태만 변경

$$\leqq -\mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right)\right]$$

Jensen’s inequality (?)

$$= \mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{q(x_{1:T} \mid x_0)}{p_\theta(x_{0:T})}\right)\right] = L_{VLB}$$

로그의 성질

Loss

$$L_{CE} = -\mathbb{E}_{q(x_0)}[\log p_\theta(x_0)] = -\mathbb{E}_{q(x_0)}\left[\log\left(\int p_\theta(x_{0:T})dx_{1:T}\right)\right]$$

$$= -\mathbb{E}_{q(x_0)}\left[\log\left(\int p_\theta(x_{0:T})\frac{q(x_{1:T} \mid x_0)}{q(x_{1:T} \mid x_0)}dx_{1:T}\right)\right]$$

$$= -\mathbb{E}_{q(x_0)}\left[\log\left(\int q(x_{1:T} \mid x_0)\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}dx_{1:T}\right)\right]$$

$$\boxed{\begin{aligned} &= -\mathbb{E}_{q(x_0)}\left[\log\left(\mathbb{E}_{q(x_{1:T} \mid x_0)}\left[\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right]\right)\right] \\ &\leqq -\mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right)\right] \end{aligned}}$$

$$= \mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{q(x_{1:T} \mid x_0)}{p_\theta(x_{0:T})}\right)\right] = L_{VLB}$$

Jensen's inequallity

given f is convex function, $x_1, x_2, .., x_N$ is domain of function f
$f(x_1, x_2, \ldots, x_N) \leq f(x_1) + f(x_2) + \ldots + f(x_N)$

$$= -\mathbb{E}_{q(x_0)}\left[\log\left(\mathbb{E}_{q(x_{1:T} \mid x_0)}\left[\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right]\right)\right]$$

$$= -\mathbb{E}_{q(x_0)}\left[\log\left(\int q(x_{1:T} \mid x_0)\left[\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right]dx_{1:T}\right)\right]$$

$$\leqq -\mathbb{E}_{q(x_0)}\left[\int q(x_{1:T} \mid x_0)\log\left(\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right)dx_{1:T}\right]$$

$$= \int q(x_0)\int q(x_{1:T} \mid x_0)\log\left(\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right)dx_{1:T}dx_0$$

$$= \int q(x_0)q(x_{1:T} \mid x_0)\log\left(\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right)dx_{0:T}$$

$$= \int q(x_{1:T})\log\left(\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right)dx_{0:T}$$

$$= -\mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{p_\theta(x_{0:T})}{q(x_{1:T} \mid x_0)}\right)\right]$$

Loss

$$L_{VLB} = \mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{q(x_{1:T} \mid x_0)}{p_\theta(x_{0:T})}\right)\right]$$

$$= \mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{\boxed{\prod_{t=1}^{T} q(x_t \mid x_{t-1})}}{p_\theta(x_T)\boxed{\prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t)}}\right)\right]$$

Markov chain과 bayesian rule

$$= \mathbb{E}_{q(x_{0:T})}\left[\boxed{-\log p_\theta(x_T)} + \log\left(\frac{\prod_{t=1}^{T} q(x_t \mid x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1} \mid x_t)}\right)\right]$$

로그의 성질

$$= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \log\left(\frac{\boxed{\prod_{t=2}^{T}} q(x_t \mid x_{t-1})}{\boxed{\prod_{t=2}^{T}} p_\theta(x_{t-1} \mid x_t)}\right) + \boxed{\log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}}\right]$$

t=1을 로그의 성질을 이용하여 다시 정리

$$= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \boxed{\sum_{t=2}^{T}}\log\left(\frac{q(x_t \mid x_{t-1})}{p_\theta(x_{t-1} \mid x_t)}\right) + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right]$$

로그의 성질

Loss

$$
\begin{aligned}
L_{VLB} &= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{q(x_t \mid x_{t-1})}{p_\theta(x_{t-1} \mid x_t)}\right) + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{\boxed{q(x_t \mid x_{t-1}, x_0)}}{p_\theta(x_{t-1} \mid x_t)}\right) + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{1}{p_\theta(x_{t-1} \mid x_t)} \boxed{\frac{q(x_t, x_{t-1}, x_0)}{q(x_{t-1}, x_0)}}\right) + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{1}{p_\theta(x_{t-1} \mid x_t)} \frac{q(x_t, x_{t-1}, x_0)}{q(x_{t-1}, x_0)} \boxed{\frac{q(x_t, x_0)}{q(x_t, x_0)}}\right) + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{1}{p_\theta(x_{t-1} \mid x_t)} \frac{q(x_t, x_{t-1}, x_0)}{\boxed{q(x_t, x_0)}} \frac{q(x_t, x_0)}{\boxed{q(x_{t-1}, x_0)}}\right) + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{\boxed{q(x_{t-1} \mid x_t, x_0)}}{p_\theta(x_{t-1} \mid x_t)} \boxed{\frac{q(x_t \mid x_0)}{q(x_{t-1} \mid x_0)}}\right) + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right]
\end{aligned}
$$

Markov chain property: 이전 시점 이외의 조건부 확률에는 영향을 받지 않으므로, x_0를 추가하여도 식은 동일

Loss

$$
\begin{aligned}
L_{VLB} &= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1} \mid x_t, x_0)}{p_\theta(x_{t-1} \mid x_t)} \frac{q(x_t \mid x_0)}{q(x_{t-1} \mid x_0)}\right) + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1} \mid x_t, x_0)}{p_\theta(x_{t-1} \mid x_t)}\right) + \boxed{\sum_{t=2}^{T} \log\left(\frac{q(x_t \mid x_0)}{q(x_{t-1} \mid x_0)}\right)} + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[-\log p_\theta(x_T) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1} \mid x_t, x_0)}{p_\theta(x_{t-1} \mid x_t)}\right) + \boxed{\log\left(\frac{q(x_T \mid x_0)}{q(x_1 \mid x_0)}\right)} + \log \frac{q(x_1 \mid x_0)}{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[\log\left(\boxed{\frac{q(x_T \mid x_0)}{p_\theta(x_T)}}\right) + \sum_{t=2}^{T} \log\left(\frac{q(x_{t-1} \mid x_t, x_0)}{p_\theta(x_{t-1} \mid x_t)}\right) - \log \boxed{p_\theta(x_0 \mid x_1)}\right] \\
&= \mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{q(x_T \mid x_0)}{p_\theta(x_T)}\right)\right] + \mathbb{E}_{q(x_{0:T})}\left[\sum_{t=2}^{T} \log\left(\frac{q(x_{t-1} \mid x_t, x_0)}{p_\theta(x_{t-1} \mid x_t)}\right)\right] - \mathbb{E}_{q(x_{0:T})}[\log p_\theta(x_0 \mid x_1)]
\end{aligned}
$$

Loss

$$L_{VLB} = \mathbb{E}_{q(x_{0:T})}\left[\log\left(\frac{q(x_T\,|\,x_0)}{p_\theta(x_T)}\right)\right] + \mathbb{E}_{q(x_{0:T})}\left[\sum_{t=2}^{T}\log\left(\frac{q(x_{t-1}\,|\,x_t,x_0)}{p_\theta(x_{t-1}\,|\,x_t)}\right)\right] - \mathbb{E}_{q(x_{0:T})}[\log p_\theta(x_0\,|\,x_1)]$$

$$= \underbrace{\mathbb{E}_{q(x_{0:T})}[D_{KL}(q(x_T\,|\,x_0)||p_\theta(x_T))]}_{L_T} + \sum_{t=2}^{T}\underbrace{\mathbb{E}_{q(x_{0:T})}[D_{KL}(q(x_{t-1}\,|\,x_t,x_0)||p_\theta(x_{t-1}\,|\,x_t))]}_{L_{t-1}} - \underbrace{\mathbb{E}_{q(x_{0:T})}[\log p_\theta(x_0\,|\,x_1)]}_{L_0}$$

$$L_T = \mathbb{E}_{q(x_{0:T})}[D_{KL}(q(x_T\,|\,x_0)||p_\theta(x_T))]$$

해당 논문에서, diffusion process에는 diffusion rate(**β**)의 값이 상수로 정의되어 학습 파라미터가 없으므로 실제 구현에서는 무시

$$L_0 = \mathbb{E}_{q(x_{0:T})}[\log p_\theta(x_0\,|\,x_1)]$$

이미지 데이터 [0, 255]의 정수를 선형 변환을 통해 [-1, 1]의 실수로 변환하였다고 가정. 따라서 이런 변환 과정을 스케일링 되지 않은 원본 이미지를 얻는 과정이라 생각하여 실제 샘플링 과정에서는 x_1까지만 계산

$$L_{t-1} = \mathbb{E}_{q(x_{0:T})}[D_{KL}(q(x_{t-1}\,|\,x_t,x_0)||p_\theta(x_{t-1}\,|\,x_t))]$$

실제 학습에서 사용될 Loss 함수

$$L_{t-1} = \mathbb{E}_{q(x_{0:T})}[D_{KL}(q(x_{t-1} \mid x_t, x_0) || p_\theta(x_{t-1} \mid x_t))]$$

$$q(x_{t-1} \mid x_t, x_0) \sim \mathcal{N}\left(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I\right)$$

$$q(x_{t-1} \mid x_t, x_0) \propto \exp\left(-\frac{1}{2}\frac{(x_t - \tilde{\mu}(x_t, x_0))^2}{\tilde{\beta}_t}\right)$$

$$\begin{aligned}
& q(x_{t-1} \mid x_t, x_0) \\
&= q(x_t \mid x_{t-1}, x_0)\frac{q(x_{t-1} \mid x_0)}{q(x_t \mid x_0)} \\
&\propto \exp\left(-\frac{1}{2}\left(\frac{(x_t - \sqrt{\alpha_t}x_{t-1})^2}{\beta_t} + \frac{(x_{t-1} - \sqrt{\bar{\alpha}_{t-1}}x_0)^2}{1-\bar{\alpha}_{t-1}} - \frac{(x_t - \sqrt{\alpha_t}x_0)^2}{1-\bar{\alpha}_{t-1}}\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\frac{x_t^2 - 2\sqrt{\alpha_t}x_t x_{t-1} + \alpha_t x_{t-1}^2}{\beta_t} + \frac{x_{t-1}^2 - 2\sqrt{\bar{\alpha}_{t-1}}x_0 x_{t-1} - \bar{\alpha}_{t-1}x_0^2}{1-\bar{\alpha}_{t-1}} - \frac{x_t^2 + 2\sqrt{\bar{\alpha}_t}x_0 x_t - \bar{\alpha}_t x_0^2}{1-\bar{\alpha}_{t-1}}\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)x_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t + \frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)x_{t-1} + \left(\frac{1}{\beta_t} - \frac{1}{1-\bar{\alpha}_{t-1}}\right)x_t^2 + \frac{2\sqrt{\bar{\alpha}_t}}{1-\bar{\alpha}_t}x_t x_0 - \left(\frac{\bar{\alpha}_{t-1}}{1-\bar{\alpha}_{t-1}} + \frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\right)x_0^2\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)x_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t + \frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)x_{t-1} + C(x_t, x_0)\right)\right)
\end{aligned}$$

$$L_{t-1} = \mathbb{E}_{q(x_{0:T})}[D_{KL}(q(x_{t-1} \mid x_t, x_0) || p_\theta(x_{t-1} \mid x_t))]$$

$$q(x_{t-1} \mid x_t, x_0) \sim \mathcal{N}\left(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t I\right)$$

$$q(x_{t-1} \mid x_t, x_0) \propto \exp\left(-\frac{1}{2}\frac{(x_t - \tilde{\mu}(x_t, x_0))^2}{\tilde{\beta}_t}\right)$$

$$\begin{aligned}
&q(x_{t-1} \mid x_t, x_0) \\
&\propto \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)x_{t-1}^2 - \left(\frac{2\sqrt{\alpha_t}}{\beta_t}x_t + \frac{2\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)x_{t-1} + C(x_t, x_0)\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\frac{\left((x_{t-1}) - \frac{\left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)}{\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)}\right)^2}{1/\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)} + C(x_t, x_0)\right)\right)
\end{aligned}$$

$$\tilde{\beta}_t = \frac{1}{\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}} = \frac{\beta_t(1-\bar{\alpha}_{t-1})}{\alpha_t - \alpha_t\bar{\alpha}_{t-1} + \beta_t} = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$$

$$\begin{aligned}
\tilde{\mu}_t(x_t, x_0) &= \frac{\left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)}{\left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)} = \left(\frac{\sqrt{\alpha_t}}{\beta_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}x_0\right)\left(\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\right)\beta_t \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0
\end{aligned}$$

$$L_{t-1} = \mathbb{E}_{q(x_{0:T})}\left[D_{KL}\left(q(x_{t-1} \mid x_t, x_0) \| p_\theta(x_{t-1} \mid x_t)\right)\right]$$

$$\begin{aligned}
\tilde{\mu}_t(x_t, x_0) &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}\boxed{\left(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_t\right)} \\
&= \left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}\right)x_t - \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}\left(\sqrt{1-\bar{\alpha}_t}\epsilon_t\right) \\
&= \left(\frac{\sqrt{\alpha_t}\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_{t-1}) + \sqrt{\bar{\alpha}_{t-1}}\beta_t}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\right)x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}}\epsilon_t \\
&= \left(\frac{\sqrt{\bar{\alpha}_{t-1}}((1-\beta_t)(1-\bar{\alpha}_{t-1}) + \beta_t)}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\right)x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}}\epsilon_t \\
&= \left(\frac{\sqrt{\bar{\alpha}_{t-1}}(1-\bar{\alpha}_{t-1} + \beta_t\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\right)x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}}\epsilon_t \\
&= \left(\frac{\sqrt{\bar{\alpha}_{t-1}}(1-(1-\beta_t)\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\right)x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}}\epsilon_t \\
&= \left(\frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\right)x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}}\epsilon_t \\
&= \left(\frac{\sqrt{\bar{\alpha}_{t-1}}(1-\bar{\alpha}_t)}{(1-\bar{\alpha}_t)\sqrt{\bar{\alpha}_t}}\right)x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}}\epsilon_t \\
&= \left(\frac{1}{\sqrt{\alpha_t}}\right)x_t - \frac{\beta_t}{\sqrt{\alpha_t}\sqrt{1-\bar{\alpha}_t}}\epsilon_t \\
&= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t\right)
\end{aligned}$$

NOTE

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{(1-\bar{\alpha}_t)}\epsilon, \epsilon \sim \mathcal{N}(0, I)$$

$$q(x_t \mid x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t}x_0,\ (1-\bar{\alpha}_t)I\right)$$

where $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{t=i}^{T}\alpha_i$

Loss

$$
\begin{aligned}
L_{t-1} &= \mathbb{E}_{q(x_{0:T})}[D_{KL}(q(x_{t-1} \mid x_t, x_0)||p_\theta(x_{t-1} \mid x_t))] \\
&= \mathbb{E}\left[\frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2\right] \\
&= \mathbb{E}\left[\frac{1}{2\sigma_t^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t\right) - \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)\right)\right\|^2\right] \\
&= \mathbb{E}\left[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\sigma_t^2}\|\epsilon_t - \epsilon_\theta(x_t, t)\|^2\right] \\
&= \mathbb{E}\left[\frac{(1-\alpha_t)^2}{2\alpha_t(1-\bar{\alpha}_t)\sigma_t^2}\left\|\epsilon_t - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, t\right)\right\|^2\right]
\end{aligned}
$$

$$
L_{t-1}^{\text{simple}} = \mathbb{E}\left[\left\|\epsilon_t - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t, t\right)\right\|^2\right]
$$

Training & Sampling

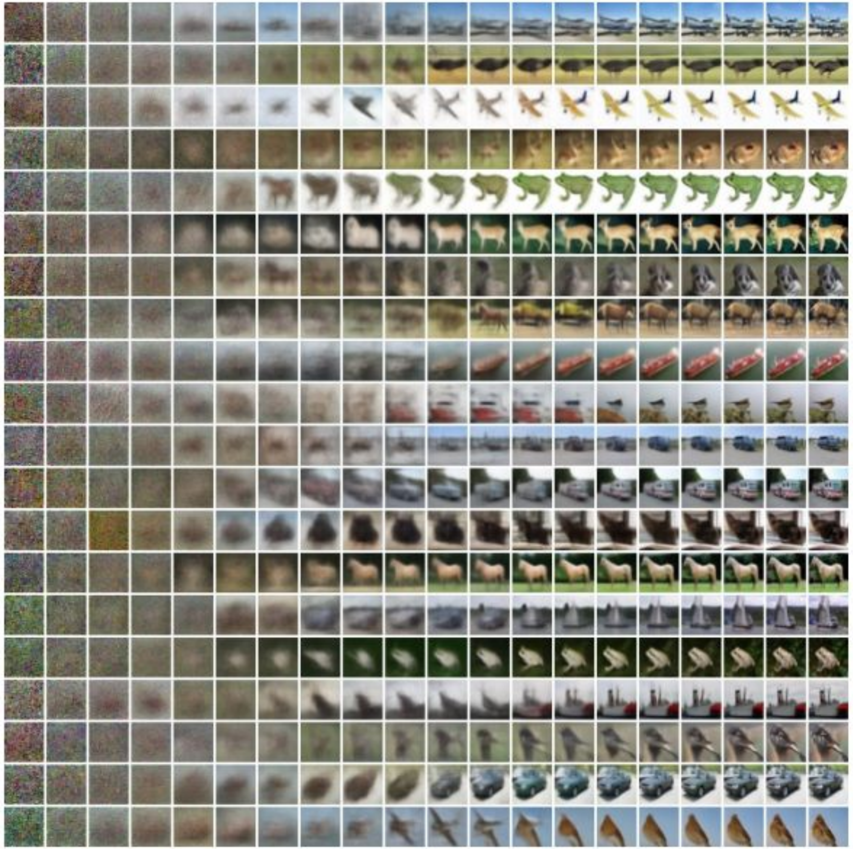
Algorithm 1 Training

repeat
  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
  $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Take gradient descent step on
  $$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
until converged

Algorithm 2 Sampling

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = T, \ldots, 1$ **do**
  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
  $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
end for
return $\mathbf{x}_0$

Denoising Diffusion Probabilistic Model의 학습 및 표본추출 (Image source: Ho et al. 2020)



Denoising Diffusion Probabilistic Model, CIFAR10 기반 unconditional progressive generation
(Image source: Ho et al. 2020)

Model	IS	FID	NLL Test (Train)
Conditional			
EBM [11]	8.30	37.9	
JEM [17]	8.76	38.4	
BigGAN [3]	9.22	14.73	
StyleGAN2 + ADA (v1) [29]	**10.06**	**2.67**	
Unconditional			
Diffusion (original) [53]			≤ 5.40
Gated PixelCNN [59]	4.60	65.93	3.03 (2.90)
Sparse Transformer [7]			**2.80**
PixelIQN [43]	5.29	49.46	
EBM [11]	6.78	38.2	
NCSNv2 [56]		31.75	
NCSN [55]	8.87 ± 0.12	25.32	
SNGAN [39]	8.22 ± 0.05	21.7	
SNGAN-DDLS [4]	9.09 ± 0.10	15.42	
StyleGAN2 + ADA (v1) [29]	$\mathbf{9.74} \pm 0.05$	3.26	
Ours (L, fixed isotropic $\mathbf{\Sigma}$)	7.67 ± 0.13	13.51	≤ 3.70 (3.69)
Ours (L_{simple})	9.46 ± 0.11	**3.17**	≤ 3.75 (3.72)

IS = Inception Score
-> 생성된 이미지로 classification이 얼마나 잘 되는지

FID = Frechet Inception Distance
-> 실제 데이터와의 확률 분포의 차이

Denoising Diffusion Probabilistic Model, 실험 결과 (Image source: Ho et al. 2020)

참고문헌

Denoising Diffusion Probabilistic Models, https://arxiv.org/pdf/2006.11239.pdf
What are Diffusion Models?, https://lilianweng.github.io/posts/2021-07-11-diffusion-models/
Denoising Diffusion Implicit Models, https://keras.io/examples/generative/ddim/