$$Loss_{VAE} = D_{KL}\left(q_\phi(z|x) \parallel p_\theta(z|x)\right) - log p_\theta(x) \quad \Rightarrow \quad Loss_{DDPM} = D_{KL}\left(q_\phi(x_{1:T}|x_0) \parallel p_\theta(x_{1:T}|x_0)\right) - log p_\theta(x_0)$$

$$-log(p_\theta(x_0)) \leq -log(p_\theta(x_0)) + \underline{D_{KL}(q(x_{1:T}|x_0)||p_\theta(x_{1:T}|x_0))}$$

$$\downarrow$$

$$log(\underline{\frac{q(x_{1:T}|x_0)}{p_\theta(x_{1:T}|x_0)}}$$

$$\downarrow$$

$$\frac{p_\theta(x_{0:T})}{p_\theta(x_0)} \leftarrow \frac{p_\theta(x_0, x_{1:T})}{p_\theta(x_0)} \leftarrow \frac{p_\theta(x_0|x_{1:T})p_\theta(x_{1:T})}{p_\theta(x_0)}$$

$$log(\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}) \rightarrow log(\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}) + log(p_\theta(x_0))$$

$$-log(p_\theta(x_0)) \leq -log(p_\theta(x_0)) + log(\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}) + log(p_\theta(x_0))$$

---

(3)

$$-log(p_\theta(x_0)) \leq log(\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})})$$

$$
\begin{aligned}
log(\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}) &= \log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}) \\
&= -\log(p(x_T)) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}) \\
&= -\log(p(x_T)) + \sum_{t=1}^{T} log(\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}) \qquad t=1 \\
&= -\log(p(x_T)) + \boxed{\sum_{t=2}^{T}} log(\frac{q(x_t|x_{t-1})}{p_\theta(x_{t-1}|x_t)}) + log(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)})
\end{aligned}
$$

$$
\begin{aligned}
q(x_t|x_{t-1}) &= \frac{q(x_{t-1}|x_t)\,q(x_t)}{q(x_{t-1})} \\
&\Rightarrow \frac{q(x_{t-1}|x_t, x_0)\,q(x_t|x_0)}{q(x_{t-1}|x_0)}
\end{aligned}
$$

$$
= -\log(p(x_T)) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)}\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}) + log(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)})
$$

$$
= -\log(p(x_T)) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_t,\ x_0)}{p_\theta(x_{t-1}|x_t)}) + \boxed{\sum_{t=2}^{T} log(\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)})} + log(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)})
$$

T=4

$$
\sum_{t=2}^{4} log(\frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}) = log(\prod_{t=2}^{4} \frac{q(x_t|x_0)}{q(x_{t-1}|x_0)}) = log(\frac{\cancel{q(x_2|x_0)}\,\cancel{q(x_3|x_0)}\,\boxed{q(x_4|x_0)}}{\boxed{q(x_1|x_0)}\,\cancel{q(x_2|x_0)}\,\cancel{q(x_3|x_0)}})
$$

$$
= -\log(p(x_T)) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_t,\ x_0)}{p_\theta(x_{t-1}|x_t)}) + log(\frac{q(x_T|x_0)}{q(x_1|x_0)}) + log(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)})
$$

$$
= \boxed{-\log(p(x_T))} + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_t,\ x_0)}{p_\theta(x_{t-1}|x_t)}) + \boxed{log(\frac{q(x_T|x_0)}{q(x_1|x_0)}) + log(\frac{q(x_1|x_0)}{p_\theta(x_0|x_1)})}
$$

$$
\log(\frac{q(x_T|x_0)}{p(x_T)}) \qquad \boxed{\log(q(x_T|x_0))} - \log(p_\theta(x_0|x_1))
$$

$$
= \log(\frac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_t,\ x_0)}{p_\theta(x_{t-1}|x_t)}) - \log(p_\theta(x_0|x_1))
$$

(5)

$$= D_{KL}(q(x_T|x_0)\,||\,p(x_T)) + \sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t,\ x_0)\,||\,p_\theta(x_{t-1}|x_t)) - \log(p_\theta(x_0|x_1))$$

첫번째 텀은 학습하는게 아니라 상수여서 무시

$$\sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t,\ x_0)\,||\,p_\theta(x_{t-1}|x_t)) - \log(p_\theta(x_0|x_1))$$

$$= \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I) \qquad = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \beta I)$$

$$\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0 \qquad \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\cdot\beta_t$$

$$\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boxed{\mathbf{x}_0}$$

$$x_t = \sqrt{\bar{\alpha}_t}\ x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon$$

$$\boxed{x_0} = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon)$$

$$\tilde{\mu}_t = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon)$$

$$= \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\epsilon\right)$$

$$\sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t,\ x_0)\,||\,p_\theta(x_{t-1}|x_t)) - \log(p_\theta(x_0|x_1))$$

$$= \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I) \qquad = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \beta I)$$

$$\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\epsilon\right)$$

Loss function

$$L_t = \frac{1}{2\sigma_t^2}||\,\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)||^2$$

$$L_t = \frac{1}{2\sigma_t^2}||\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right) - \mu_\theta(x_t,t)||^2$$

$$\mu_\theta(x_t,t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t)\right)$$

$$\frac{1}{2\sigma_t^2}||\frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right) - \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t)\right)||^2$$

$$\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}||\epsilon - \epsilon_\theta(x_t,t)||^2 \longrightarrow ||\epsilon - \epsilon_\theta(x_t,t)||^2$$

P(xt-1|xt)

$$\mathcal{N}(x_{t-1};\overbrace{\mu_\theta(x_t,t)}^{\mu_\theta(x_t,t) = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t)\right)},\underbrace{\Sigma_\theta(x_t,t)}_{\beta_t})$$

$$\mathcal{N}\left(x_{t-1};\ \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t)\right),\beta_t\right)$$

Reparameterization Trick:

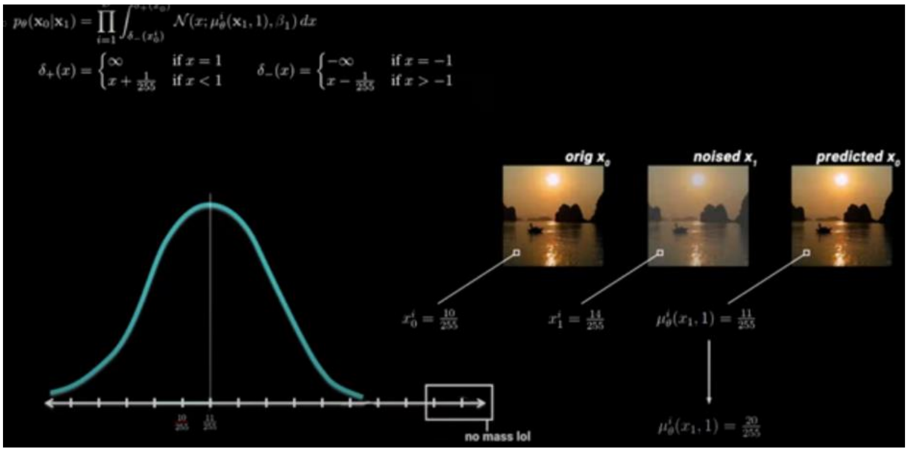
$$\mathcal{N}(\mu,\sigma^2) = \mu + \sigma\cdot\epsilon$$

$$\boxed{x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t,t)\right) + \boxed{\sqrt{\beta_t}}\epsilon}$$

$$L_{VLB} = \underbrace{\sum_{t=2}^{T} D_{KL}(q(x_{t-1}|x_t,\ x_0)\,||\,p_\theta(x_{t-1}|x_t))}_{||\epsilon - \epsilon_\theta(x_t,t)||^2} - \log(p_\theta(x_0|x_1))$$

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{D}\int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)}\mathcal{N}(x;\mu_\theta^i(\mathbf{x}_1,1),\beta_1)\,dx$$

$$\delta_+(x) = \begin{cases}\infty & \text{if } x = 1\\ x + \frac{1}{255} & \text{if } x < 1\end{cases} \qquad \delta_-(x) = \begin{cases}-\infty & \text{if } x = -1\\ x - \frac{1}{255} & \text{if } x > -1\end{cases}$$


$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x; \mu_\theta^i(\mathbf{x}_1, 1), \beta_1)\, dx$
$\delta_+(x) = \begin{cases} \infty & \text{if } x = 1 \\ x + \frac{1}{255} & \text{if } x < 1 \end{cases}$
$\delta_-(x) = \begin{cases} -\infty & \text{if } x = -1 \\ x - \frac{1}{255} & \text{if } x > -1 \end{cases}$
orig x_0
noised x_1
predicted x_0
$x_0^i = \frac{10}{255}$
$x_1^i = \frac{14}{255}$
$\mu_\theta^i(x_1, 1) = \frac{11}{255}$
$\mu_\theta^i(x_1, 1) = \frac{20}{255}$
$\frac{10}{255}$
$\frac{11}{255}$
no mass lol