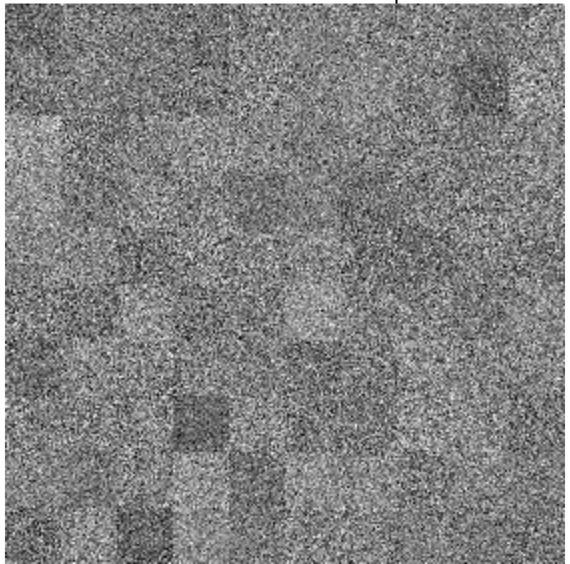
01

Diffusion



Prerequisites: Recap on VAE

- ELBO Loss (Variational Lower Bound)

$$-L_{\mathrm{VAE}} = \log p_\theta(\mathbf{x}) - D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z}|\mathbf{x})) \leq \log p_\theta(\mathbf{x})$$

[Note]
KL Divergence is always non-negative!

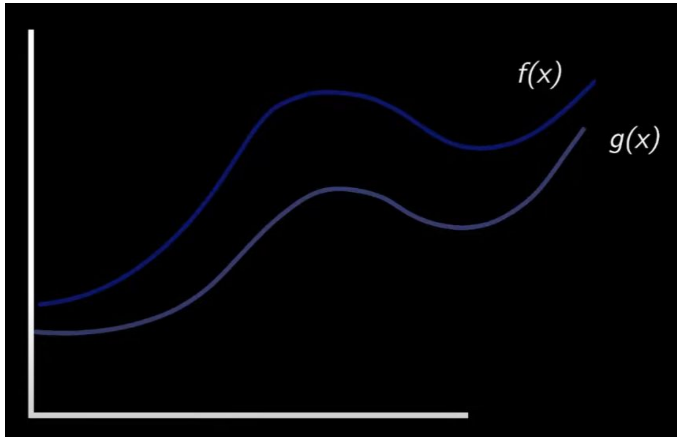
$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

= difference between two distributions

- Reparameterization Trick

$$\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x}^{(i)}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{2(i)} \boldsymbol{I})$$

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I}) \qquad \text{; Reparameterization trick.}$$



Korea Institute of Science and Technology

Why?

Expanding VAE Loss (ELBO : Evidence Lower Bound)

Note: KL Divergence

$D_{KL}(p(x)\|q(x)) = \int p(x) \log \frac{p(x)}{q(x)} dx$

$$D_{KL}(q_\phi(z|x)\|p_\theta(z|x))$$

$$= \int q_\phi(z|x) \log \frac{q_\phi(z|x)}{p_\theta(z|x)} dz$$

$$= \int q_\phi(z|x) \log \frac{q_\phi(z|x)}{p_\theta(z,x)/p_\theta(x)} dz$$

$$= \int q_\phi(z|x) \log \frac{q_\phi(z|x)\, p_\theta(x)}{p_\theta(z,x)} dz$$

$$= \int q_\phi(z|x) \left\{ \log p_\theta(x) + \log \frac{q_\phi(z|x)}{p_\theta(z,x)} \right\} dz$$

$$= \int q_\phi(z|x) \log p_\theta(x)\, dz + \int q_\phi(z|x) \log \frac{q_\phi(z|x)}{p_\theta(z,x)} dz$$

Note: Conditional Expectation

$\mathbb{E}(X|Y=y) = \int x f_{X|Y}(x|y) dx$

$= \frac{1}{f_Y(y)} \int x f_{X,Y}(x,y) dx$

$$= \log p_\theta(x) + \int q_\phi(z|x) \log \frac{q_\phi(z|x)}{p_\theta(z,x)} dz$$

$$= \log p_\theta(x) + \int q_\phi(z|x) \log \frac{q_\phi(z|x)}{p_\theta(x|z)\, p_\theta(z)} dz \rightarrow D_{KL}(q_\phi(z|x)\|p_\theta(z))$$

$$= \log p_\theta(x) + \int q_\phi(z|x) \left\{ \log \frac{q_\phi(z|x)}{p_\theta(z)} - \log p_\theta(x|z) \right\} dz$$

$$= \log p_\theta(x) + \mathbb{E}_{z\sim q_\phi(z|x)} \left[\log \frac{q_\phi(z|x)}{p_\theta(z)} - \log p_\theta(x|z) \right\} dz$$

$$= \log p_\theta(x) + D_{KL}(q_\phi(z|x)\|p_\theta(z)) - \mathbb{E}_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)]$$

$$\therefore D_{KL}(q_\phi(z|x)\|p_\theta(z|x)) = \log p_\theta(x) + D_{KL}(q_\phi(z|x)\|q_\theta(z)) - \mathbb{E}_{z\sim q_\phi(z|x)} \log p_\theta(x|z)$$

Objective

Minimize: $\underbrace{D_{KL}(q_\phi(z|x)\|p_\theta(z|x))}_{\text{= minimize difference between approx \& true posterior distributions}} - \underbrace{\log p_\theta(x)}_{\text{= maximize likelihood of generating } x}$

$D_{KL} \geq 0$

$$-L_{VAE} = +\log p_\theta(x) - D_{KL}(q_\phi(z|x)\|p_\theta(z|x)) \leq \log p_\theta(x)$$

↑ Lower Bound!

Overview

Learning Manifold(=Distribution) of Data

$z \xrightarrow{f(z)}$ ��

High Entropy
Low Information

Low Entropy
High Information

Overview

Learning Manifold(=Distribution) of Data

Learn through Loss!

$$Z \xrightarrow{\boxed{f(z)}} ��$$

High Entropy
Low Information

Low Entropy
High Information

Recap:

$$-L_{\mathrm{VAE}} = \log p_\theta(\mathbf{x}) - D_{\mathrm{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z}|\mathbf{x})) \le \log p_\theta(\mathbf{x})$$

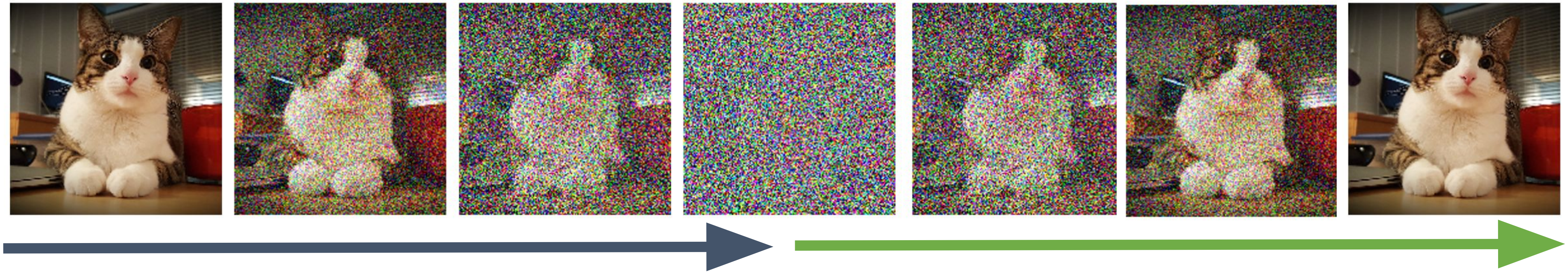
Korea Institute of
Science and Technology

Overview

�� Image

Z
Noise

�� Image



Forward Process

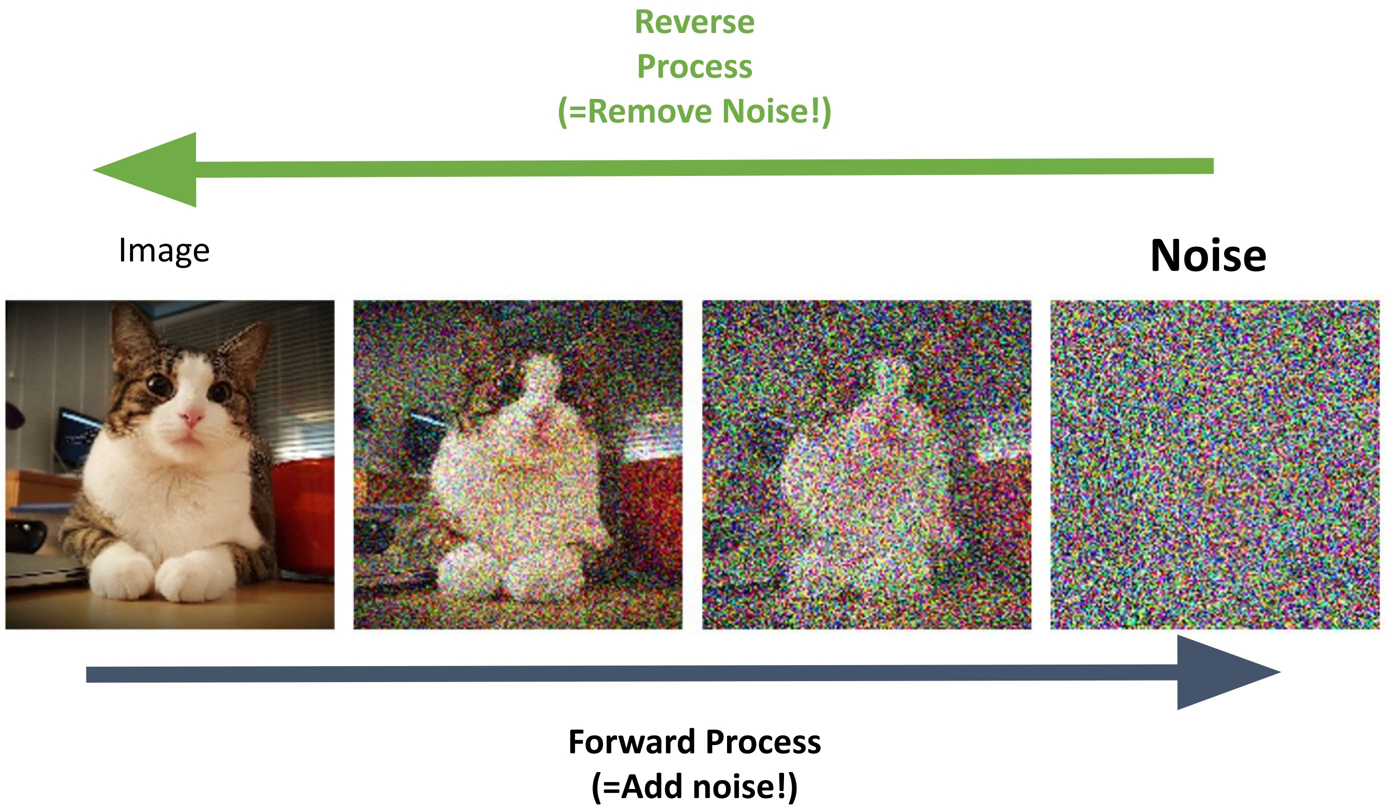
Fixed Encoder!

Reverse Process

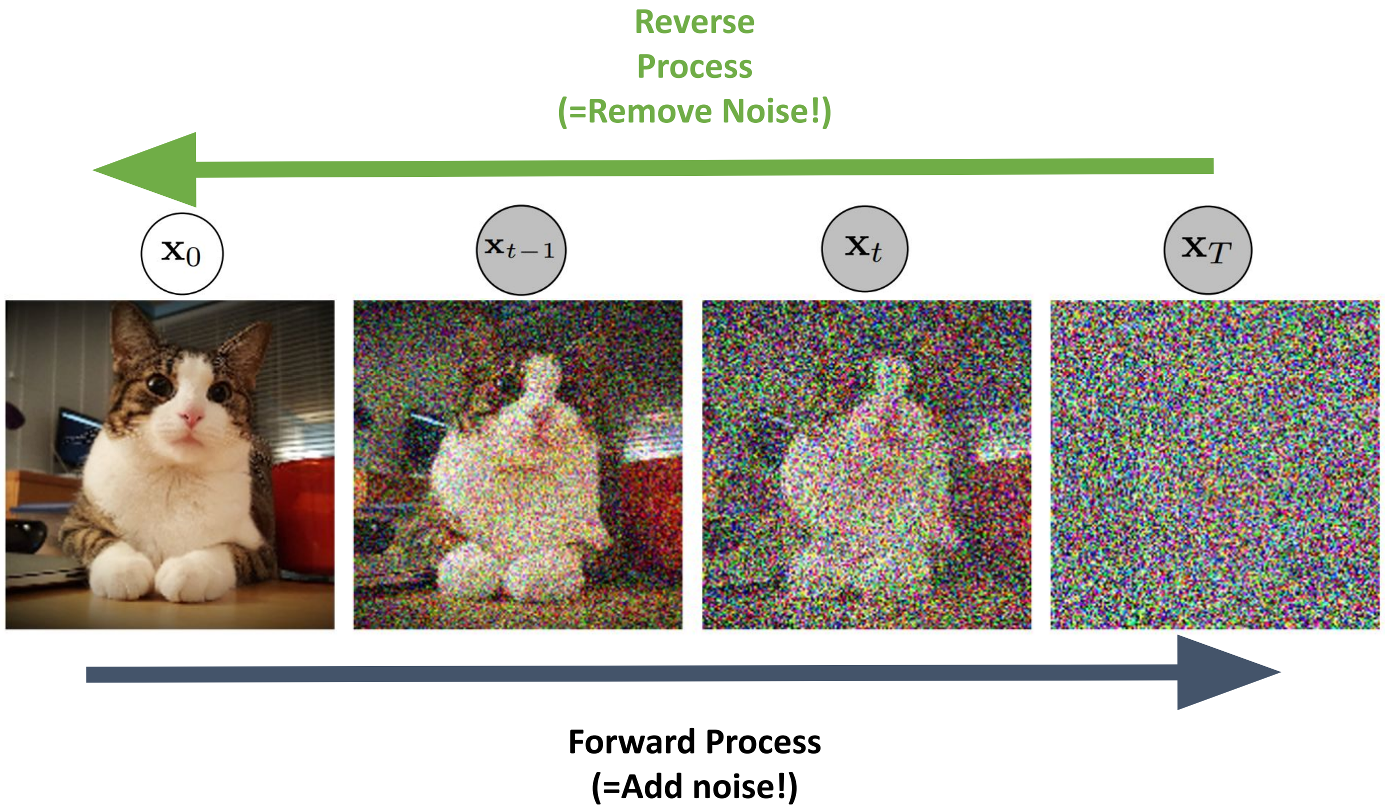
Where NN comes in! (θ)

Noise를 얼만큼 빼서 이미지를 만들자

Concept


Reverse
Process
(=Remove Noise!)
Image
Noise
Forward Process
(=Add noise!)

Concept

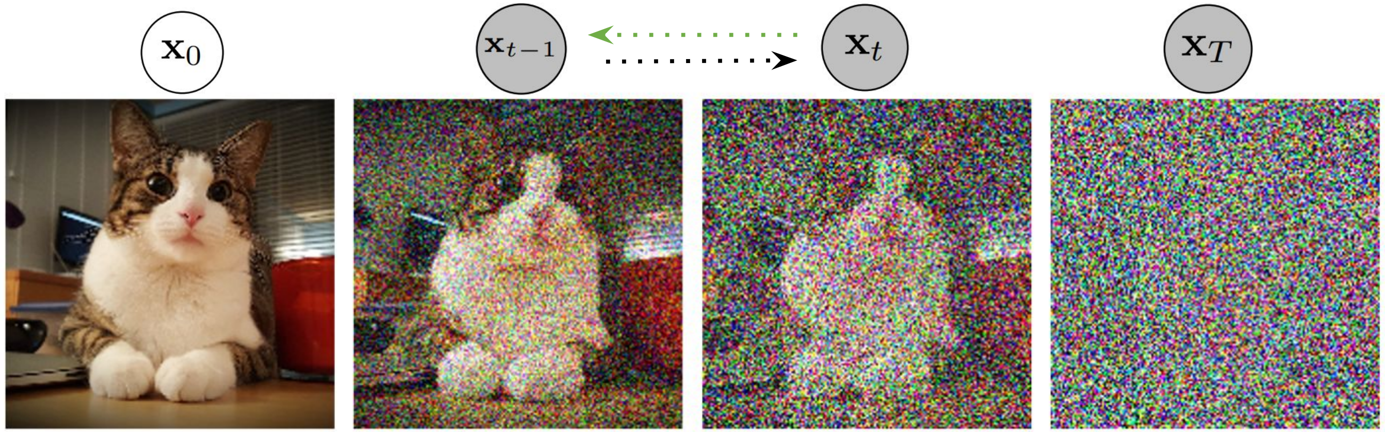


Reverse
Process
(=Remove Noise!)
$\mathbf{x}_0$
$\mathbf{x}_{t-1}$
$\mathbf{x}_t$
$\mathbf{x}_T$
Forward Process
(=Add noise!)

Concept

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$$

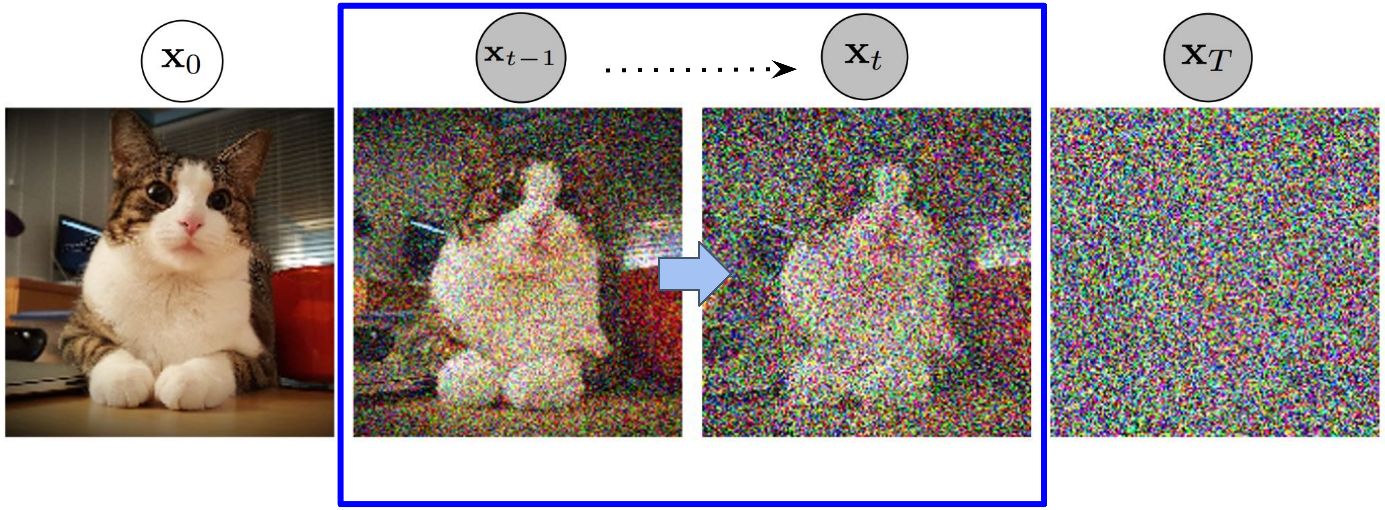
Reverse
Process
(=Remove Noise!)



Forward Process
(=Add noise!)

$$q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

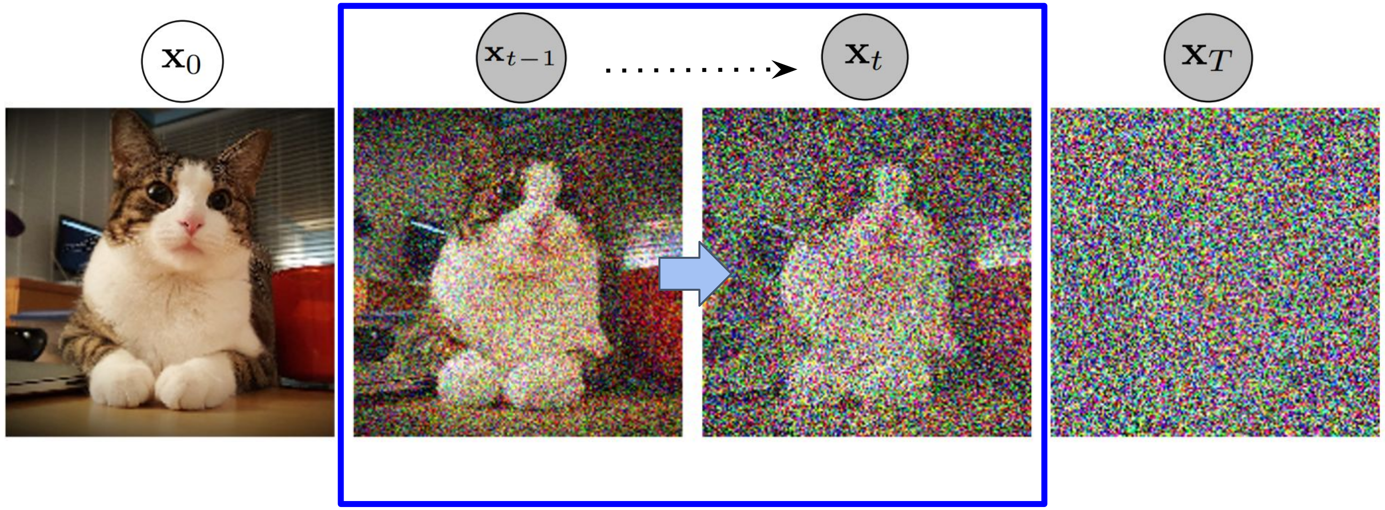
Concept



Forward Process
(=Add noise!)

$$\boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1})}$$

Concept

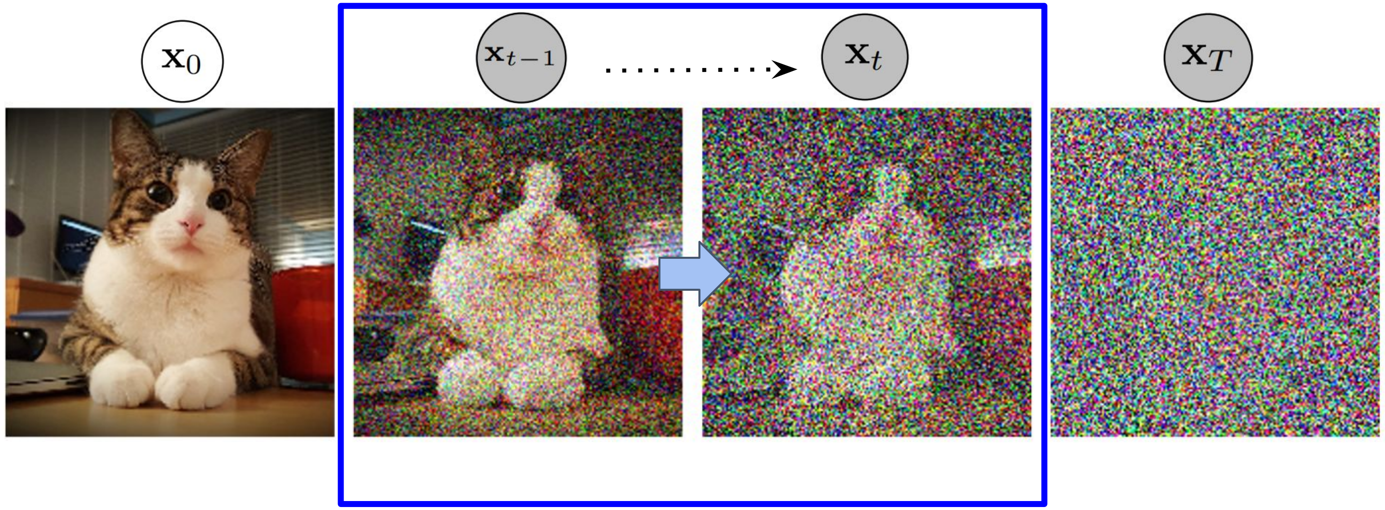


Forward Process
(=Add noise!)

$$\boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1})}$$

$$x_t = a * x_{t-1} + b * noise$$

Concept



Forward Process
(=Add noise!)

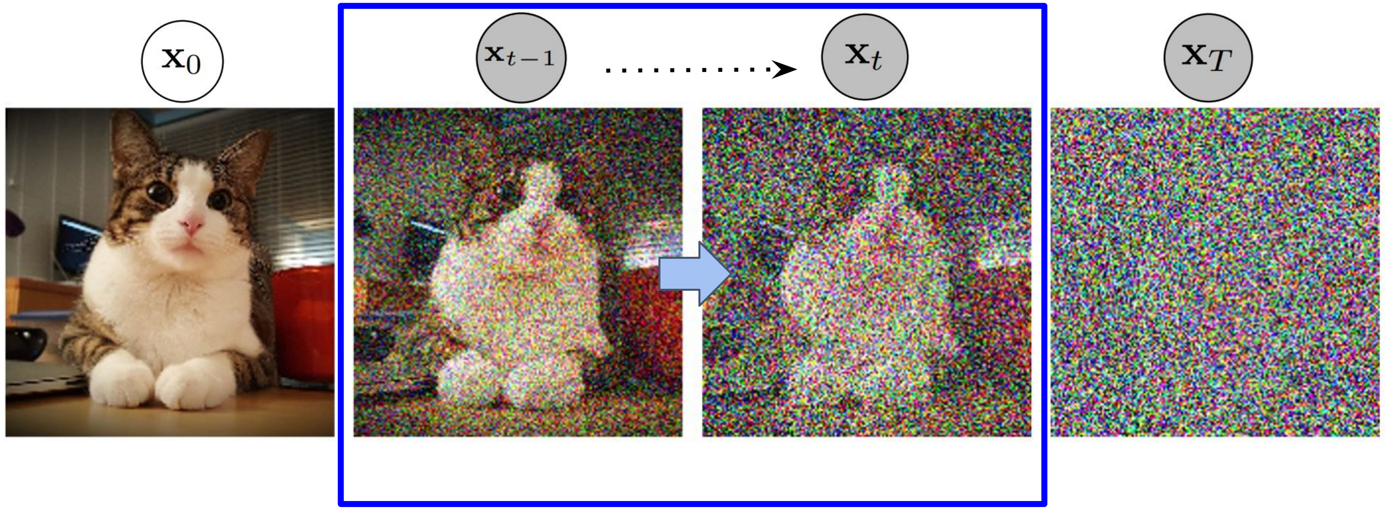
$$\boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1})}$$

$$x_t = a * x_{t-1} + b * noise$$

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Korea Institute of Science and Technology

Concept



Forward Process
(=Add noise!)

$$\boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1})}$$

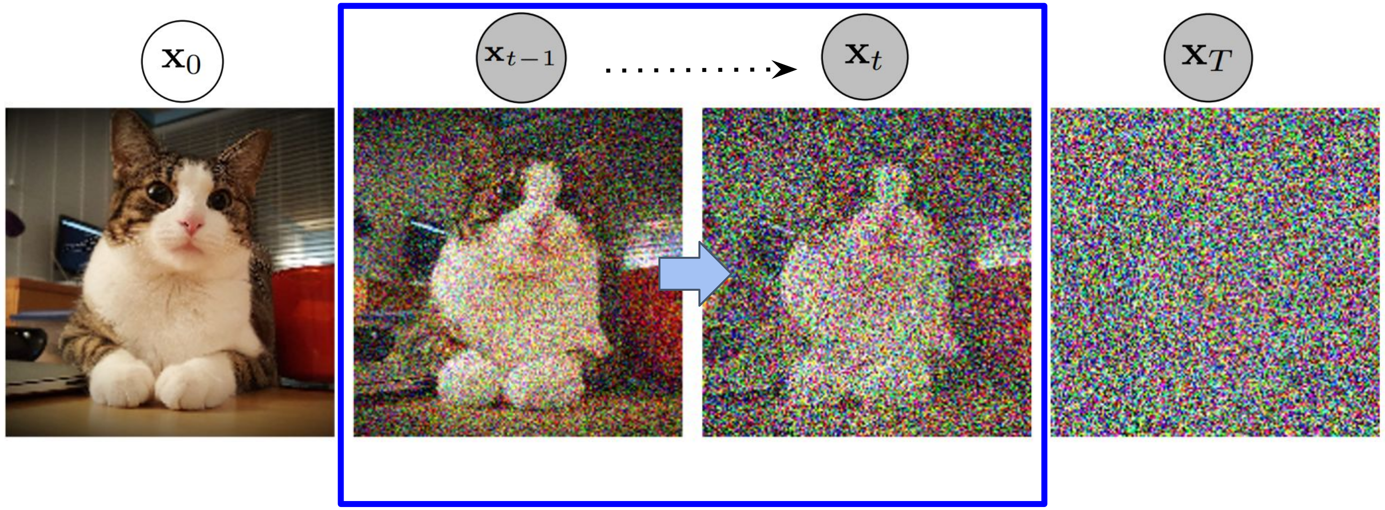
$\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x}^{(i)}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{2(i)}\boldsymbol{I})$

$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$; Reparameterization trick.

$$x_t = a * x_{t-1} + b * noise$$

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad \boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})}$$

Concept



Forward Process
(=Add noise!)

$$q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

$\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x}^{(i)}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}^{(i)}, \boldsymbol{\sigma}^{2(i)}\boldsymbol{I})$

$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$; Reparameterization trick.

$$x_t = \boxed{a} * x_{t-1} + \boxed{b} * noise \longrightarrow x_t = \boxed{\sqrt{1-\beta_t}} * x_{t-1} + \boxed{\sqrt{\beta_t}} * noise$$

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \boxed{\sqrt{1-\beta_t}}\mathbf{x}_{t-1}, \boxed{\beta_t}\mathbf{I})$$

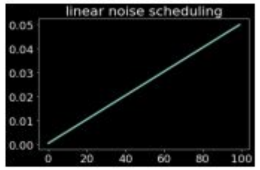
Korea Institute of Science and Technology

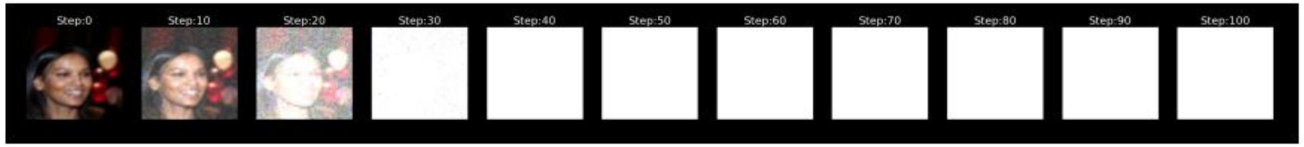
Concept

- Overview
  - Linear, Quad, Sigmoid, Cosine, ...

✓ Linear scheduling

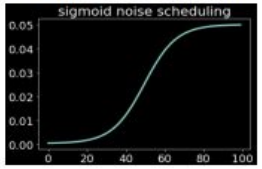
β

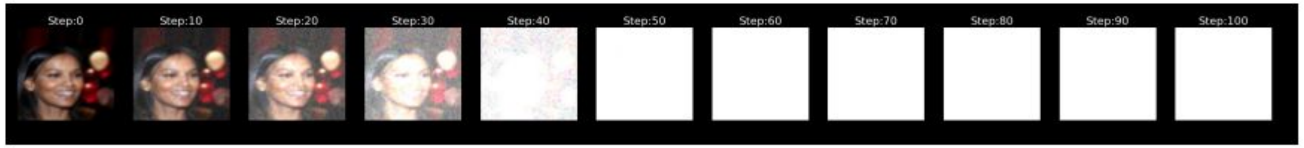




✓ Sigmoid scheduling

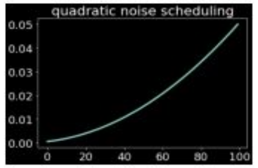
β

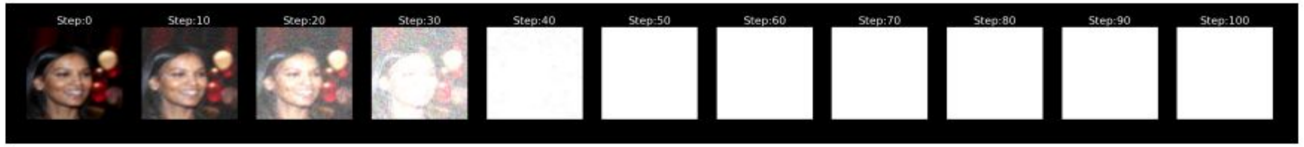




✓ Quadratic scheduling

β





$$x_t = \boxed{a} * x_{t-1} + \boxed{b} * noise \longrightarrow x_t = \boxed{\sqrt{1-\beta_t}} * x_{t-1} + \boxed{\sqrt{\beta_t}} * noise$$

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Korea Institute of Science and Technology

Concept

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Reverse
Process

Image

Noise

Forward
Process

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Concept

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Reverse Process

Image

Noise

Forward Process

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad \boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})}$$

Korea Institute of Science and Technology

Concept

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Reverse Process

Image

Noise

Forward Process

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

If forward process is gaussian & step sizes are small enough, the reverse process is also Gaussian

Concept

Entire reverse process to learn with NN θ

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Reverse Process

Image

Noise

Forward Process

Forward Process given image (x0)

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Concept

Entire reverse process to learn with NN θ

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T}\boxed{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Reverse Process

Image

Noise

Forward Process

Forward Process given image (x0)

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T}\boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1}),}$$

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Concept

Entire reverse process to learn with NN θ

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T}\boxed{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

❑ Markov Chain

- Markov 성질을 갖는 이산 확률과정

  ✓ Markov 성질 : "특정 상태의 확률(t+1)은 오직 현재(t)의 상태에 의존한다"

  ✓ 이산 확률과정 : 이산적인 시간(0초, 1초, 2초, ..) 속에서의 확률적 현상

$$P(s_{10}) \quad P(s_9) \quad \cancel{P(s_8)} \quad \cancel{P(s_7)} \quad \cancel{P(s_6)} \quad \cdots \quad \cancel{P(s_1)}$$

$$\boxed{P[s_{t+1}|s_t] = \mathrm{P}[s_{t+1}|s_1, .., s_t]}$$

Forward Process given image (x0)

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T}\boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1}),} \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Additional Notations

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Reverse Process

Image

Noise

Forward Process

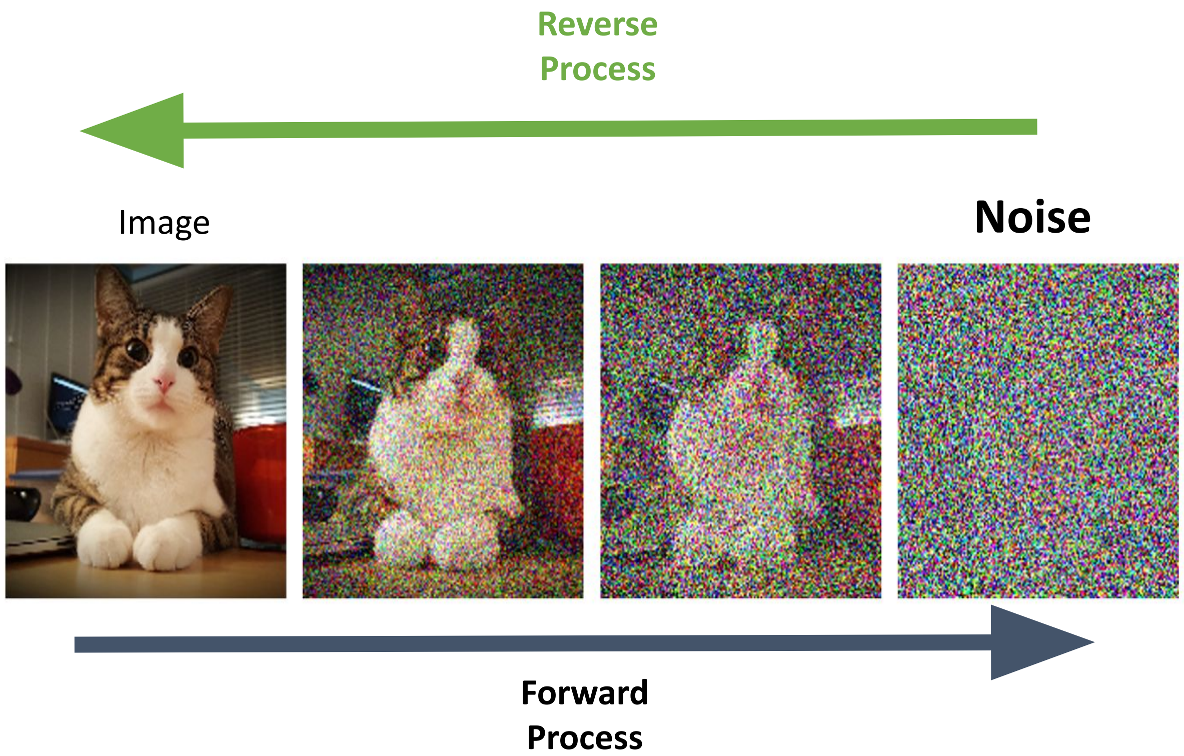
$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

β_t : Degree of noise added at time step t

Korea Institute of Science and Technology

Additional Notations

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

β_t : Degree of noise added at time step t

$$\alpha_t = 1 - \beta_t$$

Additional Notations

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1};\boldsymbol{\mu}_\theta(\mathbf{x}_t,t),\boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t))$$

Reverse Process

Image

Noise

Forward Process

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t;\sqrt{1-\beta_t}\mathbf{x}_{t-1},\beta_t\mathbf{I})$$

β_t : Degree of noise added at time step t

$$\alpha_t = 1 - \beta_t$$

$$\bar{\alpha}_t = \prod_{i=1}^{t} a_i$$

Korea Institute of Science and Technology

Forward Process

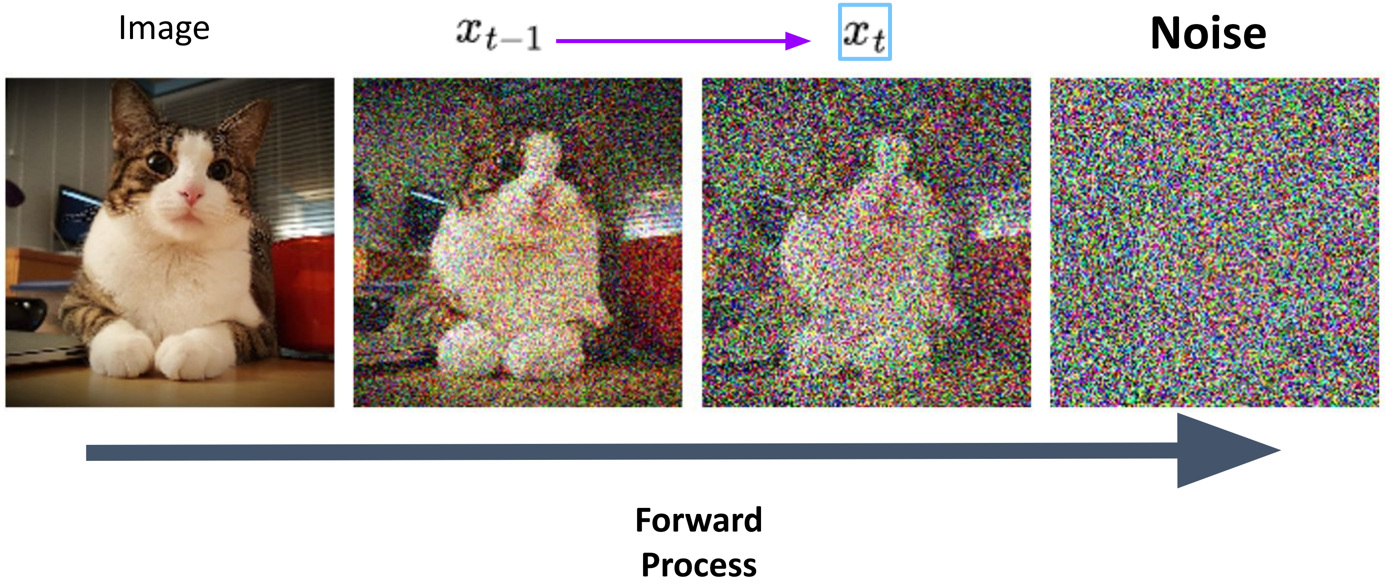
Image $\boxed{x_t}$ **Noise**

Forward Process

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

Can’t we directly sample $\boxed{x_t}$ at any time step t?

Direct Forward Process



$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad \boxed{q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})}$$

Can't we directly sample $\boxed{x_t}$ at any time step *t*?

Direct Forward Process

Reparameterization Trick

$z \sim \mathcal{N}(\mu, \sigma^2)$

$z = \mu + \sigma\epsilon \quad \epsilon \sim \mathcal{N}(0,1)$

[Forward Diffusion Process]

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

$\alpha_t = 1 - \beta_t$

$x_t = \sqrt{\alpha_t}\, \underbrace{x_{t-1}} + \sqrt{1-\alpha_t}\, \epsilon_{t-1}$

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

$= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, \epsilon_{t-2}\right) + \sqrt{1-\alpha_t}\, \epsilon_{t-1}$

$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \boxed{\boxed{\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}} + \boxed{\sqrt{1-\alpha_t}\, \epsilon_{t-1}}}$

$\epsilon_{t-2} \sim \mathcal{N}(0,1)$

$\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}$

$= 0 + \sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2} \sim \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})I)$

$\epsilon_{t-1} \sim \mathcal{N}(0,1)$

$\sqrt{1-\alpha_t}\, \epsilon_{t-1}$

$= 0 + \sqrt{1-\alpha_t}\, \epsilon_{t-1} \sim \mathcal{N}(0, (1-\alpha_t)I)$

Note

$Z = X + Y \quad X \sim \mathcal{N}(\mu_X, \sigma_X^2) \quad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

$Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

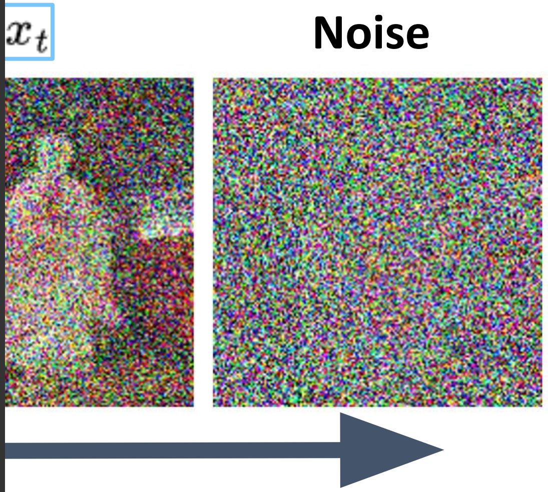
$\underbrace{\underset{\sim \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})I)}{\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}} + \underset{\sim \mathcal{N}(0, (1-\alpha_t)I)}{\sqrt{1-\alpha_t}\, \epsilon_{t-1}}}$

$\rightarrow \sim \mathcal{N}(0, (\alpha_t - \alpha_t\alpha_{t-1})I + (1-\alpha_t)I)$

$\sim \mathcal{N}(0, (\alpha_t - \alpha_t\alpha_{t-1} + 1 - \alpha_t)I)$

$\sim \mathcal{N}(0, (1-\alpha_t\alpha_{t-1})I)$

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



$= \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$

Can't we directly sample x_t at any time step *t*?

Direct Forward Process

Reparameterization Trick

$z \sim \mathcal{N}(\mu, \sigma^2)$

$z = \mu + \sigma\epsilon \quad \epsilon \sim \mathcal{N}(0,1)$

[Forward Diffusion Process]

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

$\alpha_t = 1-\beta_t$

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

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

$= \sqrt{\alpha_t}(\sqrt{\alpha_{t-1}}\,x_{t-2} + \sqrt{1-\alpha_{t-1}}\,\epsilon_{t-2}) + \sqrt{1-\alpha_t}\,\epsilon_{t-1}$

$= \sqrt{\alpha_t\alpha_{t-1}}\,x_{t-2} + \boxed{\boxed{\sqrt{\alpha_t(1-\alpha_{t-1})}\,\epsilon_{t-2}} + \boxed{\sqrt{1-\alpha_t}\,\epsilon_{t-1}}}$

$\epsilon_{t-2} \sim \mathcal{N}(0,1)$

$\sqrt{\alpha_t(1-\alpha_{t-1})}\,\epsilon_{t-2}$

$= 0 + \sqrt{\alpha_t(1-\alpha_{t-1})}\,\epsilon_{t-2} \sim \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})I)$

$\epsilon_{t-1} \sim \mathcal{N}(0,1)$

$\sqrt{1-\alpha_t}\,\epsilon_{t-1}$

$= 0 + \sqrt{1-\alpha_t}\,\epsilon_{t-1} \sim \mathcal{N}(0, (1-\alpha_t)I)$

Note

$Z = X + Y \quad X \sim \mathcal{N}(\mu_X, \sigma_X^2) \quad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

$Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

$\underbrace{\underset{\sim\mathcal{N}(0,\,\alpha_t(1-\alpha_{t-1})I)}{\sqrt{\alpha_t(1-\alpha_{t-1})}\,\epsilon_{t-2}} + \underset{\sim\mathcal{N}(0,\,(1-\alpha_t)I)}{\sqrt{1-\alpha_t}\,\epsilon_{t-1}}}$

$\rightarrow \sim \mathcal{N}(0, (\alpha_t - \alpha_t\alpha_{t-1})I + (1-\alpha_t)I)$

$\sim \mathcal{N}(0, (\alpha_t - \alpha_t\alpha_{t-1} + 1 - \alpha_t)I)$

$\sim \mathcal{N}(0, (1-\alpha_t\alpha_{t-1})I)$

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

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

$x_{t-2} = \sqrt{\alpha_{t-2}}\,x_{t-3} + \sqrt{1-\alpha_{t-2}}\,\epsilon_{t-3} \qquad \epsilon = \varepsilon$

$= \sqrt{\alpha_t\alpha_{t-1}}(\sqrt{\alpha_{t-2}}\,x_{t-3} + \sqrt{1-\alpha_{t-2}}\,\epsilon_{t-3}) + \sqrt{1-\alpha_t\alpha_{t-1}}\,\epsilon_{t-2}$

$= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}\,x_{t-3} + \sqrt{\alpha_t\alpha_{t-1}(1-\alpha_{t-2})}\,\epsilon_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}}\,\epsilon_{t-2}$

$\sim \mathcal{N}(0, \alpha_t\alpha_{t-1}(1-\alpha_{t-2})) \qquad \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1})$

$\sim \mathcal{N}(0, \alpha_t\alpha_{t-1} - \alpha_t\alpha_{t-1}\alpha_{t-2}) \qquad \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1})$

$\rightarrow \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1}\alpha_{t-2})$

$= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}\,x_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}}\,\epsilon_{t-3}$

$\vdots$

$= \sqrt{\alpha_t\alpha_{t-1}\cdots\alpha_1}\,x_0 + \sqrt{1-\alpha_t\alpha_{t-1}\cdots\alpha_1}\,\epsilon \qquad \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

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

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

→ we can directly sample x_t at any time step t

Can't we directly sample x_t at any time step *t*?

Direct Forward Process

Reparameterization Trick

$z \sim \mathcal{N}(\mu, \sigma^2)$

$z = \mu + \sigma\epsilon \quad \epsilon \sim \mathcal{N}(0,1)$

[Forward Diffusion Process]

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

$\alpha_t = 1 - \beta_t$

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

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

$= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, \epsilon_{t-2}\right) + \sqrt{1-\alpha_t}\, \epsilon_{t-1}$

$= \sqrt{\alpha_t \alpha_{t-1}}\, x_{t-2} + \boxed{\boxed{\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}} + \boxed{\sqrt{1-\alpha_t}\, \epsilon_{t-1}}}$

$\epsilon_{t-2} \sim \mathcal{N}(0,1)$

$\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}$

$= 0 + \sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2} \sim \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})I)$

$\epsilon_{t-1} \sim \mathcal{N}(0,1)$

$\sqrt{1-\alpha_t}\, \epsilon_{t-1}$

$= 0 + \sqrt{1-\alpha_t}\, \epsilon_{t-1} \sim \mathcal{N}(0, (1-\alpha_t)I)$

Note

$Z = X + Y \quad X \sim \mathcal{N}(\mu_X, \sigma_X^2) \quad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

$Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

$\underbrace{\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}}_{\sim \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})I)} + \underbrace{\sqrt{1-\alpha_t}\, \epsilon_{t-1}}_{\sim \mathcal{N}(0, (1-\alpha_t)I)}$

$\rightarrow \sim \mathcal{N}(0, (\alpha_t - \alpha_t\alpha_{t-1})I + (1-\alpha_t)I)$

$\sim \mathcal{N}(0, (\alpha_t - \alpha_t\alpha_{t-1} + 1 - \alpha_t)I)$

$\sim \mathcal{N}(0, (1-\alpha_t\alpha_{t-1})I)$

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

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

$x_{t-2} = \sqrt{\alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_{t-2}}\, \varepsilon_{t-3} \qquad \epsilon = \varepsilon$

$= \sqrt{\alpha_t\alpha_{t-1}}\left(\sqrt{\alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_{t-2}}\, \varepsilon_{t-3}\right) + \sqrt{1-\alpha_t\alpha_{t-1}}\, \epsilon_{t-2}$

$= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}\, x_{t-3} + \underline{\sqrt{\alpha_t\alpha_{t-1}(1-\alpha_{t-2})}\, \varepsilon_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}}\, \varepsilon_{t-2}}$

$\sim \mathcal{N}(0, \alpha_t\alpha_{t-1}(1-\alpha_{t-2})) \qquad \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1})$

$\sim \mathcal{N}(0, \alpha_t\alpha_{t-1} - \alpha_t\alpha_{t-1}\alpha_{t-2}) \qquad \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1})$

$\rightarrow \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1}\alpha_{t-2})$

$= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}}\, \epsilon_{t-3}$

$\vdots$

$= \sqrt{\alpha_t\alpha_{t-1}\cdots\alpha_1}\, x_0 + \sqrt{1-\alpha_t\alpha_{t-1}\cdots\alpha_1}\, \epsilon \qquad \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$

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

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

→ we can directly sample x_t at any time step t

Can't we directly sample x_t at any time step *t*?

Korea Institute of Science and Technology

Direct Forward Process

Reparameterization Trick

$z \sim \mathcal{N}(\mu, \sigma^2)$

$z = \mu + \sigma\epsilon \quad \epsilon \sim \mathcal{N}(0,1)$

[Forward Diffusion Process]

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

$\alpha_t = 1 - \beta_t$

$\Downarrow$

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

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

$= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\, x_{t-2} + \sqrt{1-\alpha_{t-1}}\, \epsilon_{t-2}\right) + \sqrt{1-\alpha_t}\, \epsilon_{t-1}$

$= \sqrt{\alpha_t\alpha_{t-1}}\, x_{t-2} + \boxed{\boxed{\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}} + \boxed{\sqrt{1-\alpha_t}\, \epsilon_{t-1}}}$

$\epsilon_{t-2} \sim \mathcal{N}(0,1)$

$\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}$

$= 0 + \sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2} \sim \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})I)$

$\epsilon_{t-1} \sim \mathcal{N}(0,1)$

$\sqrt{1-\alpha_t}\, \epsilon_{t-1}$

$= 0 + \sqrt{1-\alpha_t}\, \epsilon_{t-1} \sim \mathcal{N}(0, (1-\alpha_t)I)$

Note

$Z = X + Y \quad X \sim \mathcal{N}(\mu_X, \sigma_X^2) \quad Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$

$Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

$\underbrace{\sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2}}_{\sim \mathcal{N}(0, \alpha_t(1-\alpha_{t-1})I)} + \underbrace{\sqrt{1-\alpha_t}\, \epsilon_{t-1}}_{\sim \mathcal{N}(0, (1-\alpha_t)I)}$

$\rightarrow \sim \mathcal{N}(0, (\alpha_t - \alpha_t\alpha_{t-1})I + (1-\alpha_t)I)$

$\sim \mathcal{N}(0, (\alpha_t - \alpha_t\alpha_{t-1} + 1 - \alpha_t)I)$

$\sim \mathcal{N}(0, (1-\alpha_t\alpha_{t-1})I)$

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

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

$x_{t-2} = \sqrt{\alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_{t-2}}\, \epsilon_{t-3} \qquad \epsilon = \varepsilon$

$= \sqrt{\alpha_t\alpha_{t-1}}\left(\sqrt{\alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_{t-2}}\, \epsilon_{t-3}\right) + \sqrt{1-\alpha_t\alpha_{t-1}}\, \epsilon_{t-2}$

$= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}\, x_{t-3} + \sqrt{\alpha_t\alpha_{t-1}(1-\alpha_{t-2})}\, \varepsilon_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}}\, \varepsilon_{t-2}$

$\sim \mathcal{N}(0, \alpha_t\alpha_{t-1}(1-\alpha_{t-2})) \qquad \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1})$

$\sim \mathcal{N}(0, \alpha_t\alpha_{t-1} - \alpha_t\alpha_{t-1}\alpha_{t-2}) \qquad \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1})$

$\rightarrow \sim \mathcal{N}(0, 1-\alpha_t\alpha_{t-1}\alpha_{t-2})$

$= \sqrt{\alpha_t\alpha_{t-1}\alpha_{t-2}}\, x_{t-3} + \sqrt{1-\alpha_t\alpha_{t-1}\alpha_{t-2}}\, \epsilon_{t-3}$

$\vdots$

$= \sqrt{\alpha_t\alpha_{t-1}\cdots\alpha_1}\, x_0 + \sqrt{1-\alpha_t\alpha_{t-1}\cdots\alpha_1}\, \epsilon \qquad \bar{\alpha}_t = \prod_{i=1}^{t}\alpha_i$

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

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

$\rightarrow$ we can directly sample x_t at any time step t

Can't we directly sample $\boxed{x_t}$ at any time step t?

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

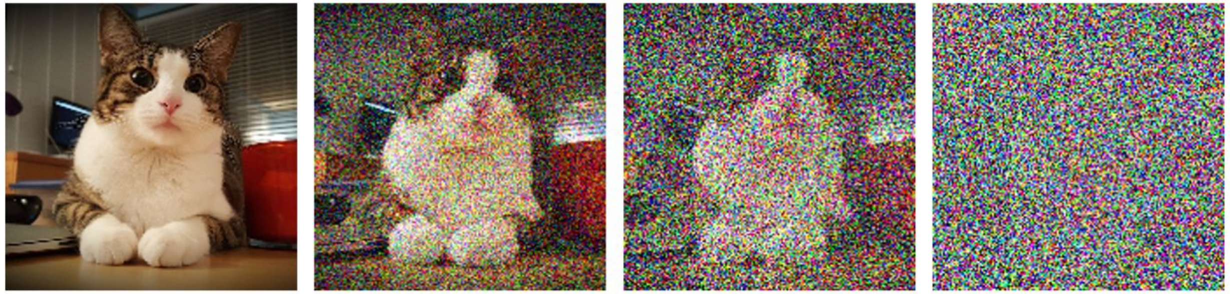
Training

$\boxed{p_\theta}$

Reverse
Process
(=Remove Noise!)

Image ��

Noise Z



Forward Process
(=Add noise!)

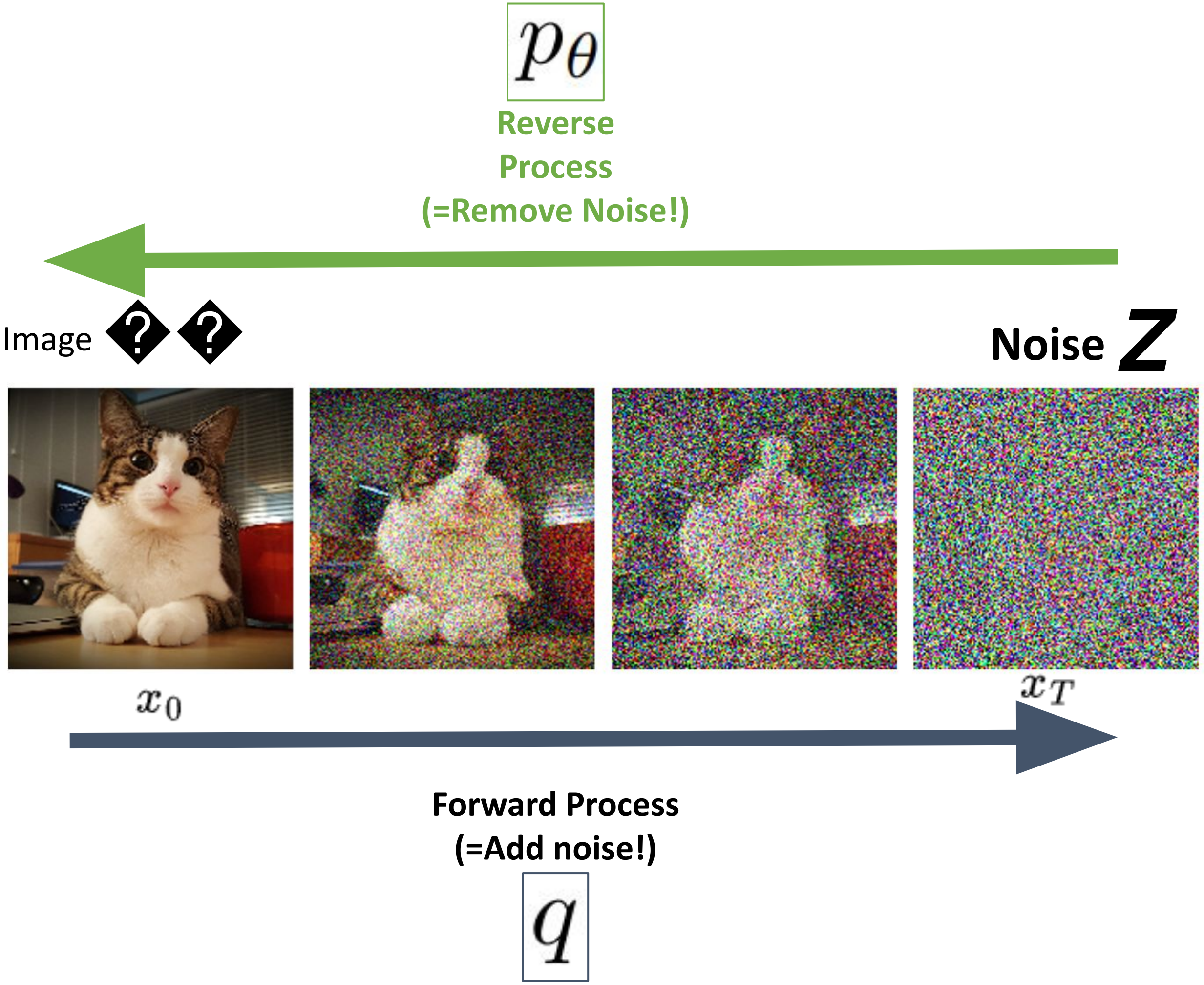
$\boxed{q}$

For VAE, minimize:

$$L_{VAE} = -log(p_\theta(x)) \leq -log(p_\theta(x)) + D_{KL}(q_\theta(z|x)||p_\theta(z|x))$$

Korea Institute of
Science and Technology

Training



$$L_{VAE} = -log(p_\theta(x)) \leq -log(p_\theta(x)) + D_{KL}(q_\theta(z|x)||p_\theta(z|x))$$

Extend to Diffusion

$$-log(p_\theta(x_0)) \leq \underbrace{-log(p_\theta(x_0)) + D_{KL}(q(x_{1:T}|x_0)||p_\theta(x_{1:T}|x_0))}_{\text{For diffusion, minimize}}$$

Training

p_θ

Reverse Process (=Remove Noise!)

Image ��

Noise Z

x_0 x_T

Forward Process (=Add noise!)

q

Maximize the likelihood of generating real data

$$L_{VAE} = -log(p_\theta(x)) \leq -log(p_\theta(x)) + D_{KL}(q_\theta(z|x)||p_\theta(z|x))$$

Extend to Diffusion

$$-log(p_\theta(x_0)) \leq \underbrace{\boxed{-log(p_\theta(x_0))} + D_{KL}(q(x_{1:T}|x_0)||p_\theta(x_{1:T}|x_0))}_{\text{For diffusion, minimize}}$$

Korea Institute of Science and Technology

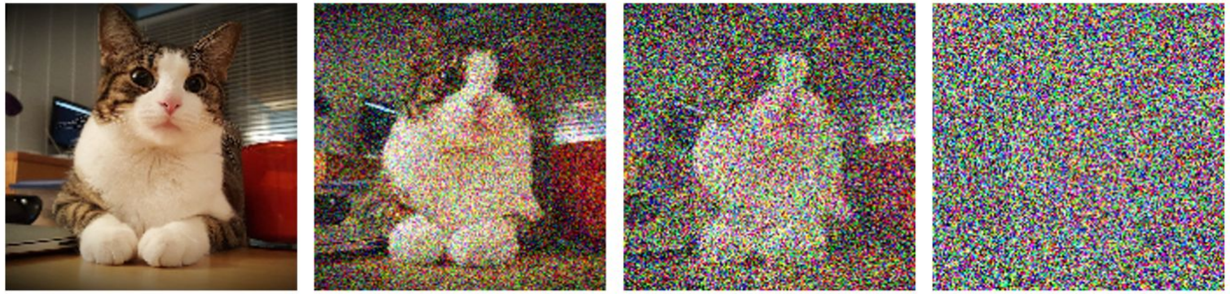
Training

p_θ

Reverse
Process
(=Remove Noise!)

Image ��

Noise Z



x_0 x_T

Forward Process
(=Add noise!)

q

$$L_{VAE} = -log(p_\theta(x)) \leq -log(p_\theta(x)) + D_{KL}(q_\theta(z|x)||p_\theta(z|x))$$

Extend to Diffusion

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

For diffusion, minimize

also minimize the difference between the real and estimated posterior distributions

Korea Institute of Science and Technology

Training

Background:

$$\boxed{D_{KL}(p||q) = \int_x p(x)log\frac{p(x)}{q(x)}dx}$$

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

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

⋮

Training

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

Background:

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

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

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

⋮

Korea Institute of Science and Technology

Training

Background:

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

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

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

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

$$p_\theta(x_{1:T}|x_0)$$

⋮

Korea Institute of Science and Technology

Training

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

Background:

$$D_{KL}(p||q) = \int_x p(x)log\frac{p(x)}{q(x)}dx$$

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

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

$$p_\theta(x_{1:T}|x_0)$$

Bayesian Rule

$$= \frac{p_\theta(x_0|x_{1:T})p_\theta(x_{1:T})}{p_\theta(x_0)}$$

⋮

Korea Institute of Science and Technology

Training

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

⋮

Background:

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

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

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

$$p_\theta(x_{1:T}|x_0)$$

Bayesian Rule

$$= \frac{p_\theta(x_0|x_{1:T}) p_\theta(x_{1:T})}{p_\theta(x_0)}$$

$$= \frac{p_\theta(x_0, x_{1:T})}{p_\theta(x_0)}$$

Training

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

⋮

Background:

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

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

$$= \int q(x_{1:T}|x_0) log \frac{q(x_{1:T}|x_0)}{\boxed{p_\theta(x_{1:T}|x_0)}} dx_{1:T}$$

$$p_\theta(x_{1:T}|x_0)$$

Bayesian Rule

$$= \frac{p_\theta(x_0|x_{1:T})p_\theta(x_{1:T})}{p_\theta(x_0)}$$

$$= \frac{p_\theta(x_0, x_{1:T})}{p_\theta(x_0)}$$

$$= \frac{p_\theta(x_{0:T})}{p_\theta(x_0)}$$

Korea Institute of Science and Technology

Training

Background:

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

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

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

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

$$p_\theta(x_{1:T}|x_0)$$

Bayesian Rule

$$= \frac{p_\theta(x_0|x_{1:T})p_\theta(x_{1:T})}{p_\theta(x_0)}$$

$$= \frac{p_\theta(x_0, x_{1:T})}{p_\theta(x_0)}$$

$$= \frac{p_\theta(x_{0:T})}{p_\theta(x_0)}$$

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

⋮

Korea Institute of Science and Technology

Training

Background:

$$\boxed{D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx}$$

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

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

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

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

⋮

Training

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

Background:

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

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

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

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

$$= \int q(x_{1:T}|x_0) [log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} + log(p_\theta(x_0))] dx_{1:T}$$

⋮

Training

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

⋮

Background:

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

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

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

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

$$= \int q(\boxed{x_{1:T}}|x_0)[log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} + log(p_\theta(x_0))]d\boxed{x_{1:T}}$$

Korea Institute of Science and Technology

Training

Background:

$$D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx$$

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

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

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

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

$$= \int q(x_{1:T}|x_0) [log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} + log(p_\theta(x_0))] dx_{1:T}$$

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

$\vdots$

Training

Background:

$$\boxed{D_{KL}(p||q) = \int_x p(x) log \frac{p(x)}{q(x)} dx}$$

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

⋮

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

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

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

$$= \int q(x_{1:T}|x_0) [log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} + log(p_\theta(x_0))] dx_{1:T}$$

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

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

Training

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

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

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

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

$$= \int q(x_{1:T}|x_0) [log \frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})} + log(p_\theta(x_0))] dx_{1:T}$$

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

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

⋮

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

Variational Lower bound for diffusion

Korea Institute of Science and Technology

Training

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

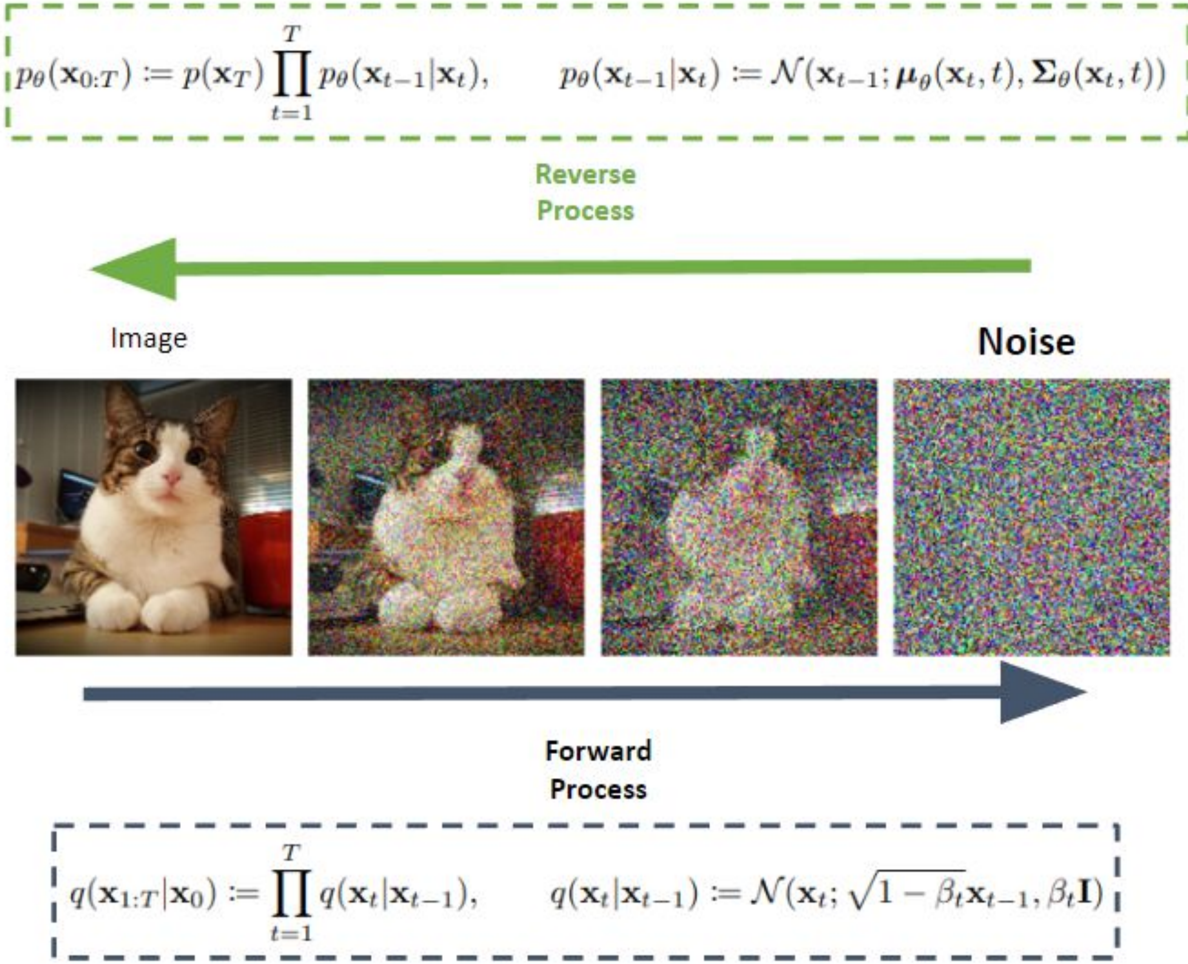
Variational Lower bound for diffusion

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$$

reverse process
(=remove noise)

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

forward process
(=add noise)



Training

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

Variational Lower bound for diffusion

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$$

reverse process
(=remove noise)

$$\boxed{q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})}$$

forward process
(=add noise)

Korea Institute of Science and Technology

Training

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

Variational Lower bound for diffusion

$$\boxed{p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T}p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}$$

reverse process
(=remove noise)

$$\boxed{q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T}q(\mathbf{x}_t|\mathbf{x}_{t-1})}$$

forward process
(=add noise)

Training

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

Variational Lower bound for diffusion

$$\boxed{p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}$$

reverse process
(=remove noise)

$$\boxed{q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})}$$

forward process
(=add noise)

Korea Institute of Science and Technology

Training

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

Variational Lower bound for diffusion

$$\boxed{p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)}$$

reverse process
(=remove noise)

$$\boxed{q(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})}$$

forward process
(=add noise)

$$-log(p_\theta(x_0)) \leq log\frac{\boxed{\prod_{t=1}^{T} q(x_t|x_{t-1})}}{\boxed{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

Training

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

$$-log(p_\theta(x_0)) \leq log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

Training

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

$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}$$

Korea Institute of Science and Technology

Training

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

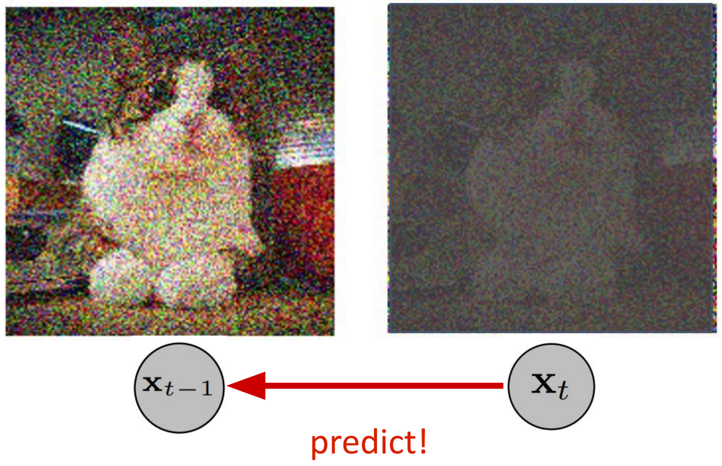
$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}$$



Training

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

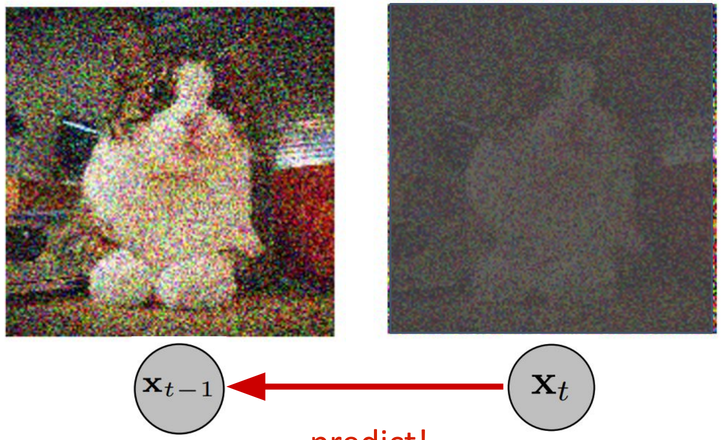
$$-log(p_\theta(x_0)) \leq log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}$$



predict!

= High
Variance :(

Korea Institute of
Science and Technology

Training

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

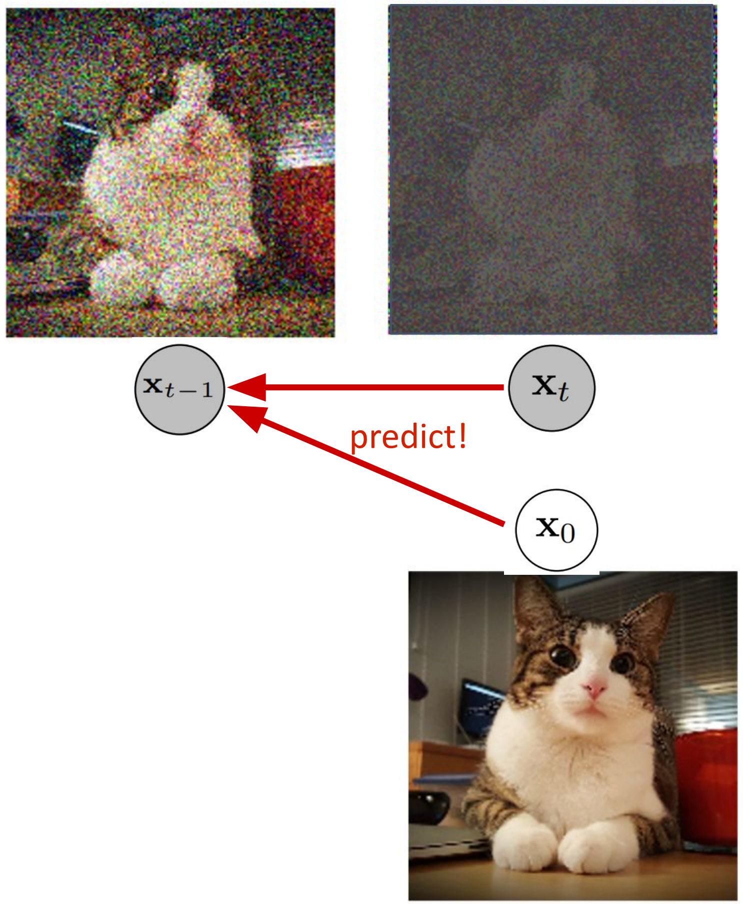
$$-log(p_\theta(x_0)) \leq log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}$$



Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

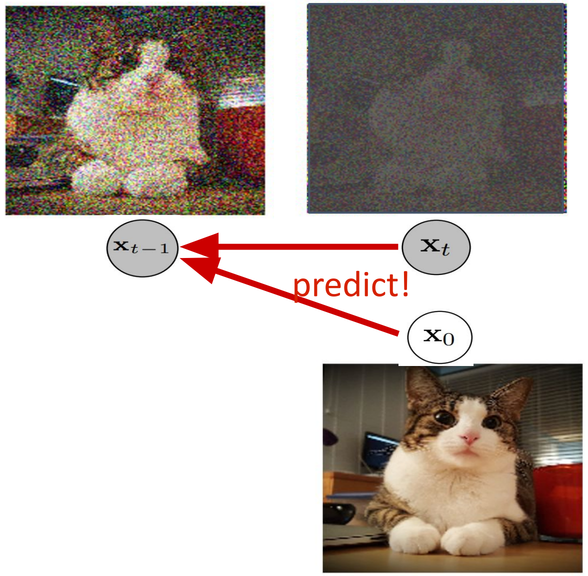
$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}$$

$$= \frac{q(x_{t-1}|x_t,\boxed{x_0})q(x_t|\boxed{x_0})}{q(x_{t-1}|\boxed{x_0})}$$



Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}$$

$$= \frac{q(x_{t-1}|x_t, x_0)q(x_t|x_0)}{q(x_{t-1}|x_0)}$$

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T)) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}$$

$$= \frac{q(x_{t-1}|x_t, x_0)q(x_t|x_0)}{q(x_{t-1}|x_0)}$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

$$-log(p(x_T))) + log(\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)})$$

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

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

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

$$q(x_t|x_{t-1}) = \frac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})}$$

$$= \frac{q(x_{t-1}|x_t, x_0)q(x_t|x_0)}{q(x_{t-1}|x_0)}$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

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

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

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

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

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

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

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

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

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

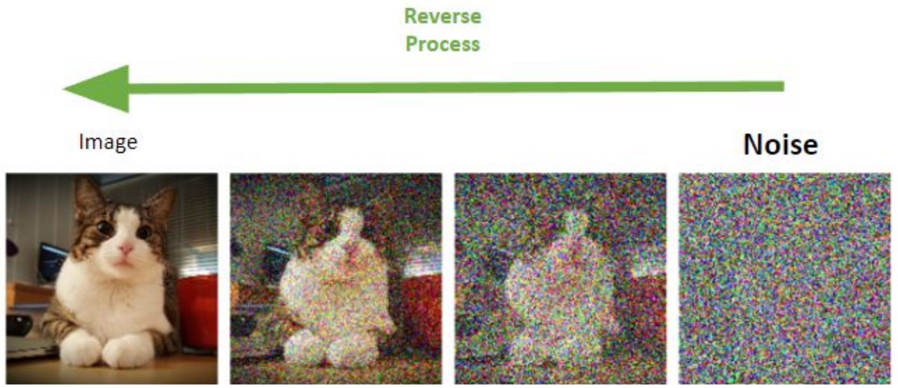
$$= \boxed{log(\frac{q(x_T|x_0)}{p(x_T)})} + \sum_{t=2}^{T} \boxed{log(\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)})} - logp_\theta(x_0|x_1)$$

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \quad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



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

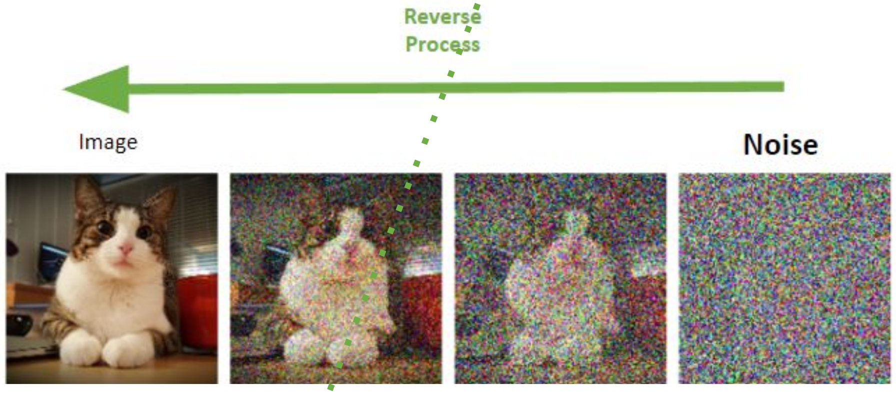
$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \quad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



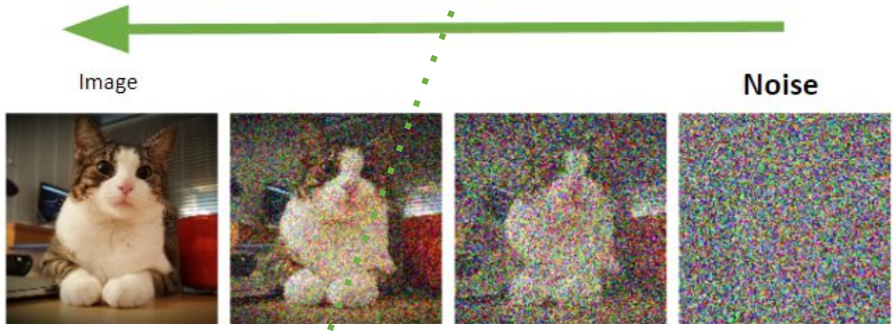
$$-log(p_\theta(x)) \leq log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$$

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1};\boldsymbol{\mu}_\theta(\mathbf{x}_t,t),\boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t))$$



Training

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

$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

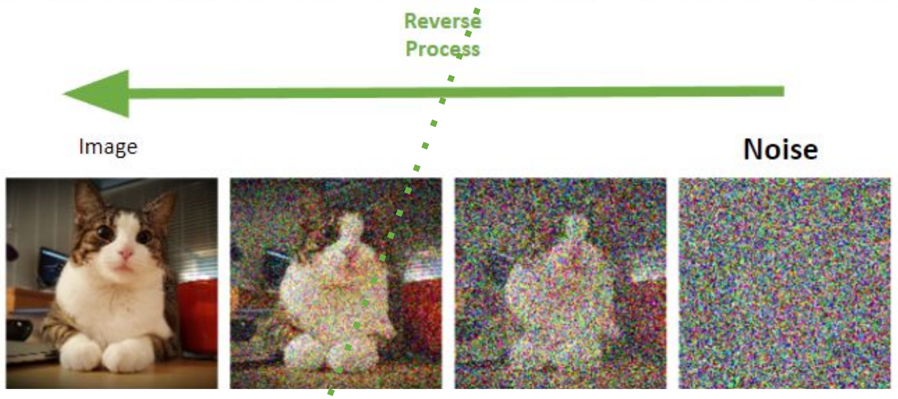
$$= \boxed{log(\frac{q(x_T|x_0)}{p(x_T)})} + \sum_{t=2}^{T}\boxed{log(\frac{q(x_{t-1}|x_t,x_0)}{p_\theta(x_{t-1}|x_t)})} - logp_\theta(x_0|x_1)$$

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(\boxed{q(x_{t-1}|x_t,x_0)}||\boxed{p_\theta(x_{t-1}|x_t)})}$$

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

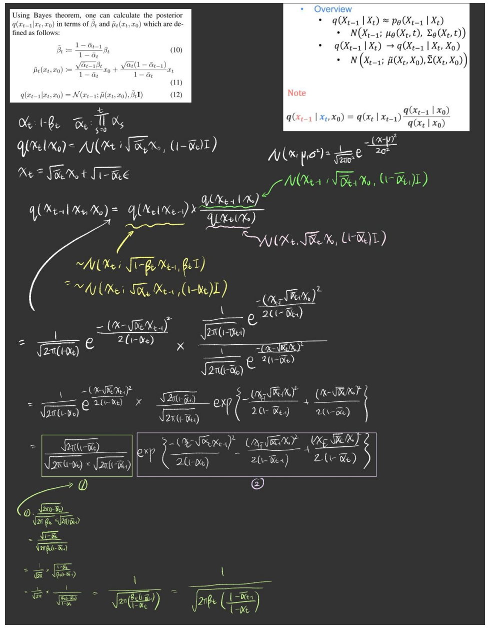


$$-log(p_\theta(x)) \leq log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$$

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

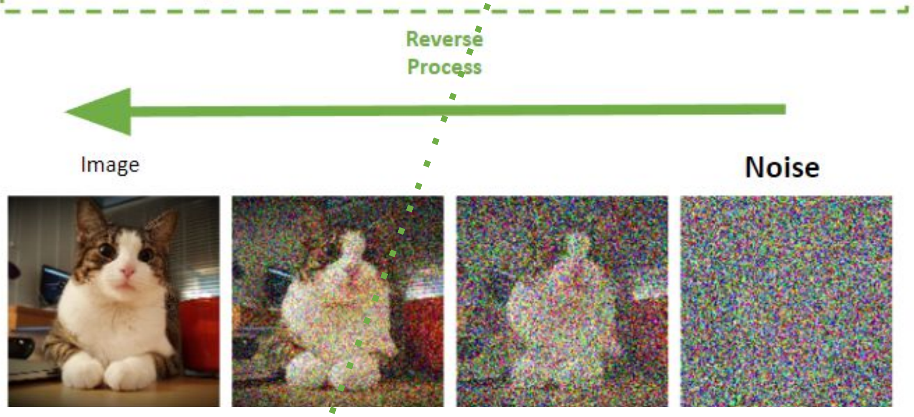
$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$



Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

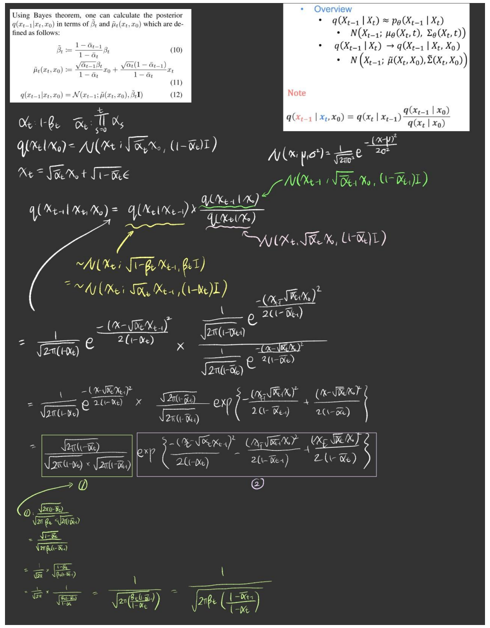


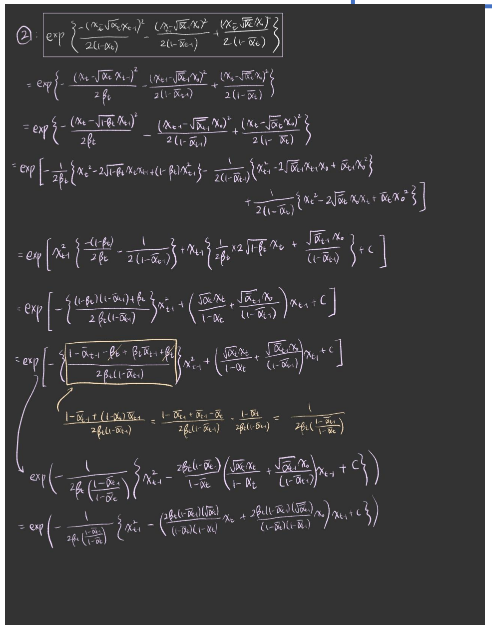
$$-log(p_\theta(x)) \leq log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$$

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t))}$$

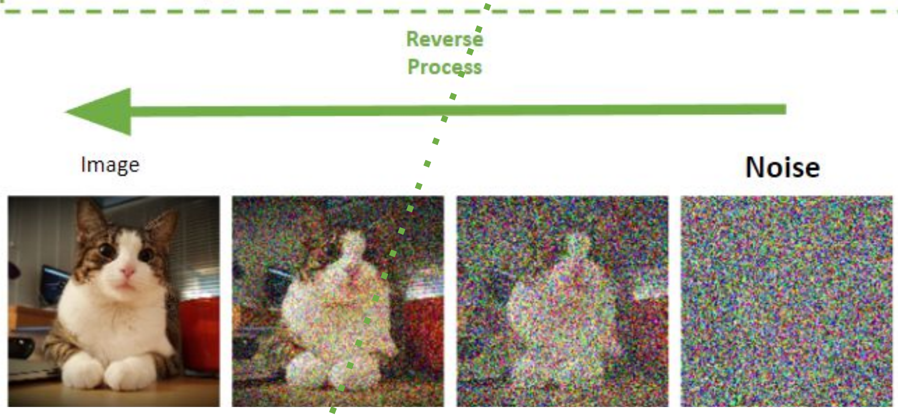




Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1};\boldsymbol{\mu}_\theta(\mathbf{x}_t,t),\boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t))$$

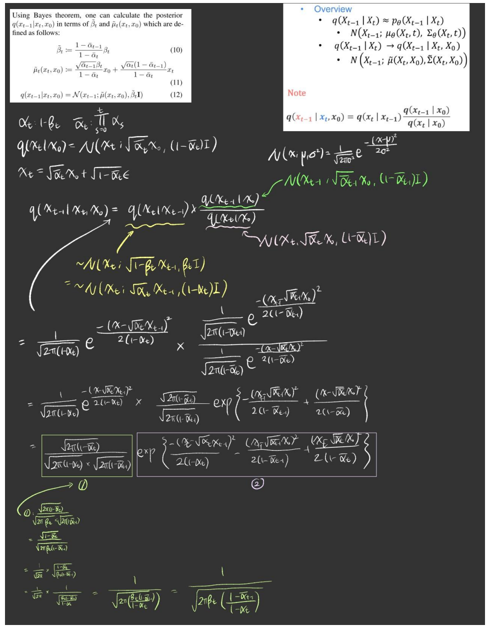


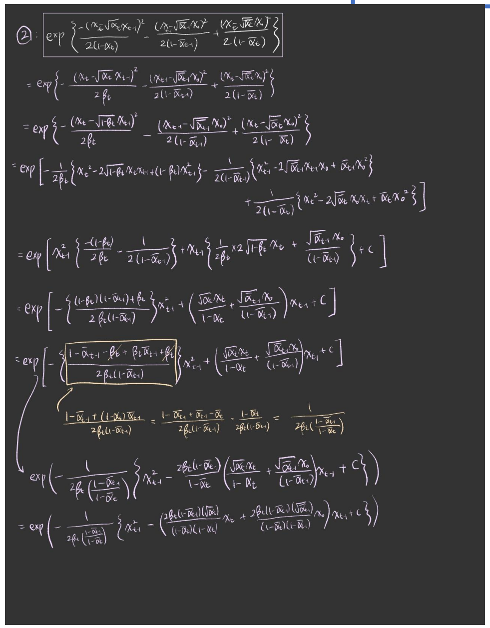
$$-log(p_\theta(x)) \le log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$$

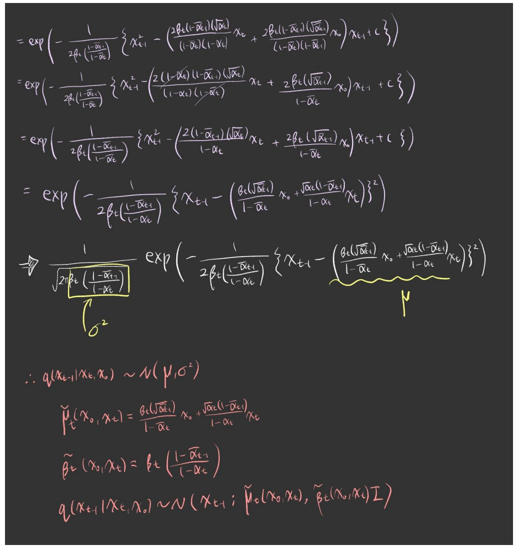
$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(\boxed{q(x_{t-1}|x_t,x_0)}||\boxed{p_\theta(x_{t-1}|x_t)})}$$



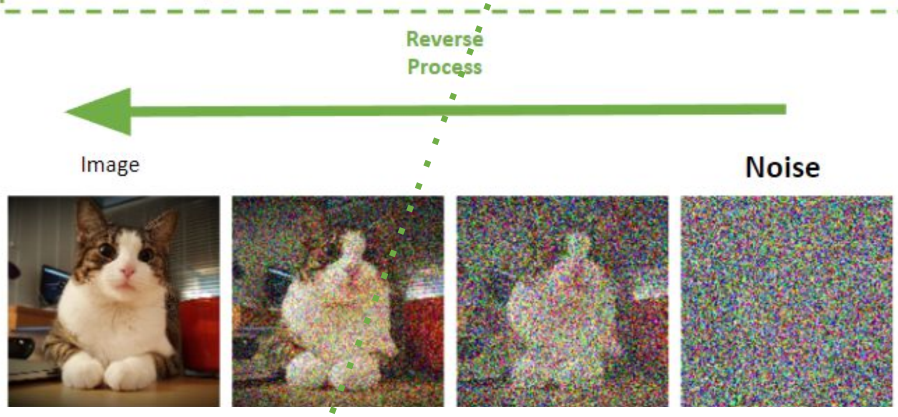




Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

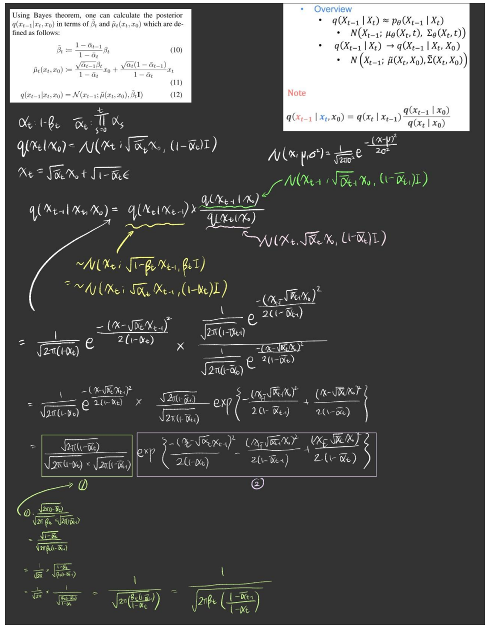


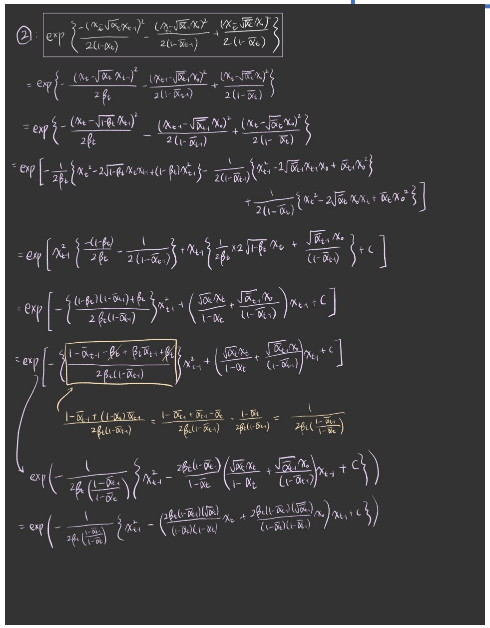
$$-log(p_\theta(x)) \le log\frac{q(x_{1:T}|x_0)}{p_\theta(x_{0:T})}$$

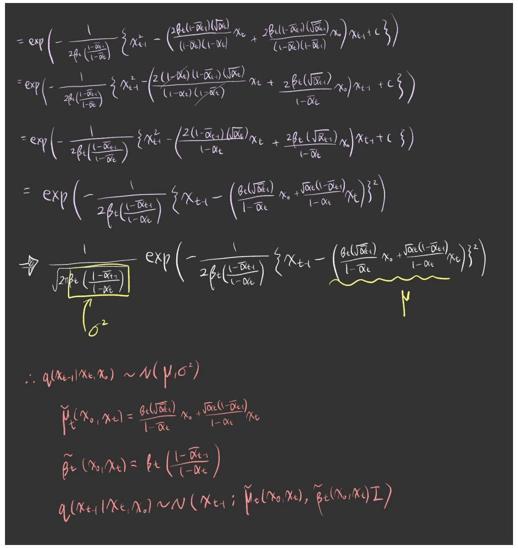
$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$







$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \qquad (10)$$

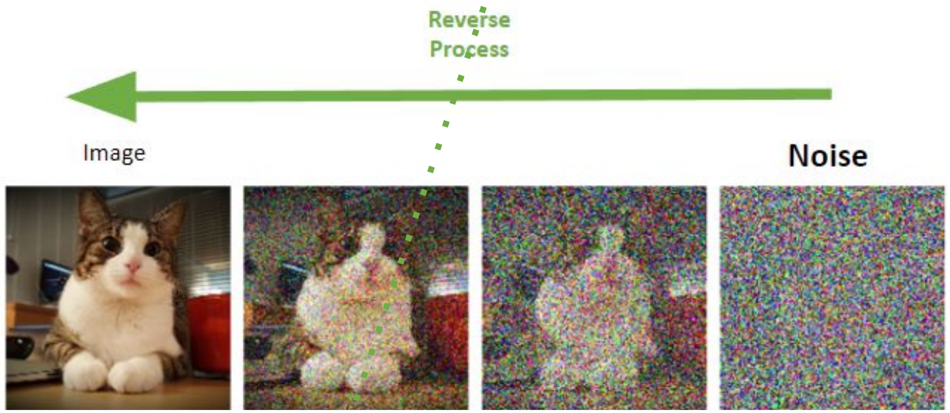
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \qquad (11)$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t\mathbf{I}) \qquad (12)$$

Korea Institute of Science and Technology

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Training



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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

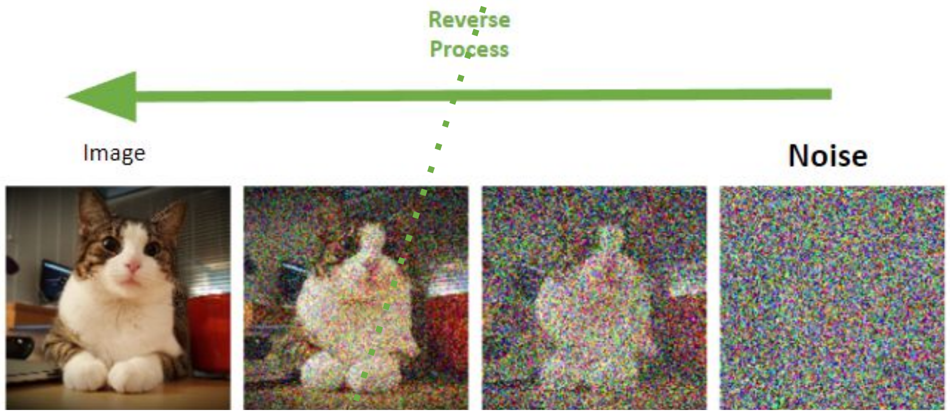
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t\mathbf{I}) \tag{12}$$

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(\boxed{q(x_{t-1}|x_t, x_0)}||\boxed{p_\theta(x_{t-1}|x_t)})}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

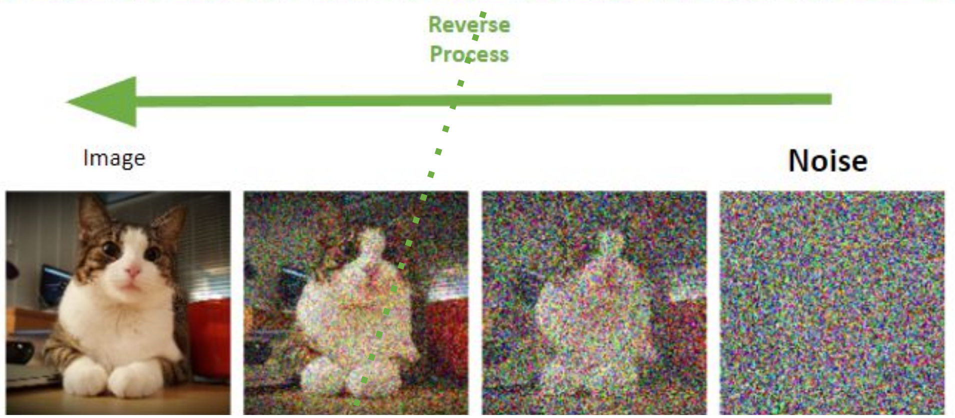
$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t\mathbf{I}) \tag{12}$$

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

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



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

$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boxed{x_0} + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

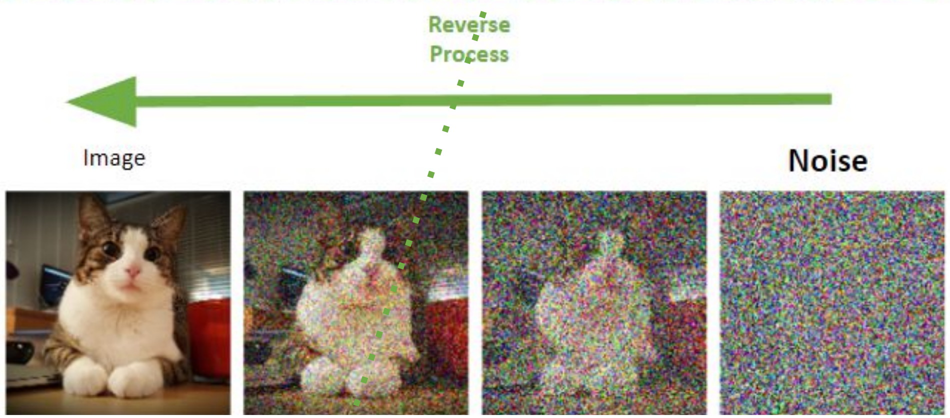
$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t\mathbf{I}) \tag{12}$$

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

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$D_{KL}(\boxed{q(x_{t-1}|x_t, x_0)}||\boxed{p_\theta(x_{t-1}|x_t)})$$

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

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t\mathbf{I}) \tag{12}$$

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

Korea Institute of Science and Technology

Training

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

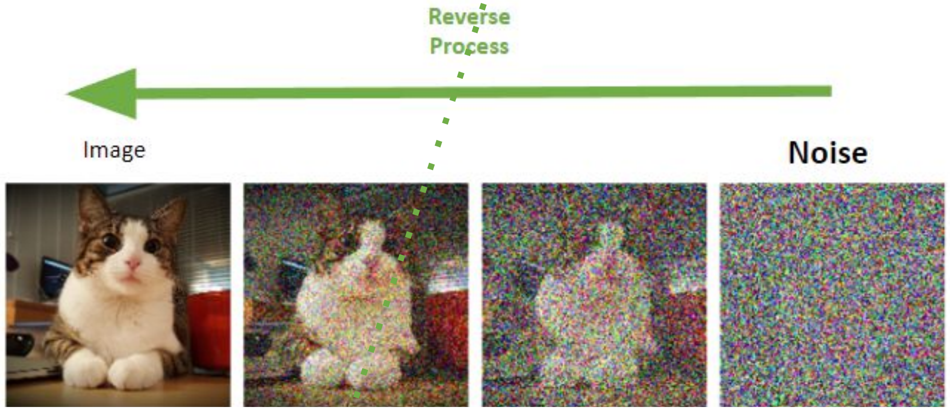
$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(\boxed{q(x_{t-1}|x_t,x_0)}||\boxed{p_\theta(x_{t-1}|x_t)})}$$

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t,t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t,t))$$



$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

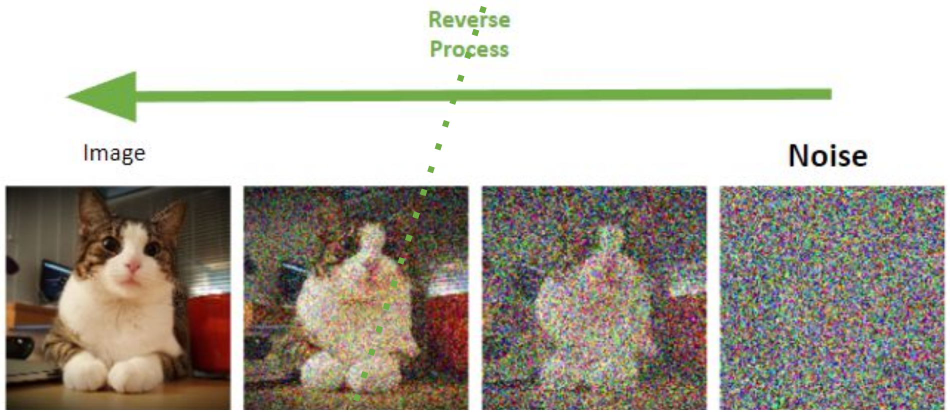
$$\tilde{\mu}_t(x_t,x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

$$q(x_{t-1}|x_t,x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I) \tag{12}$$

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

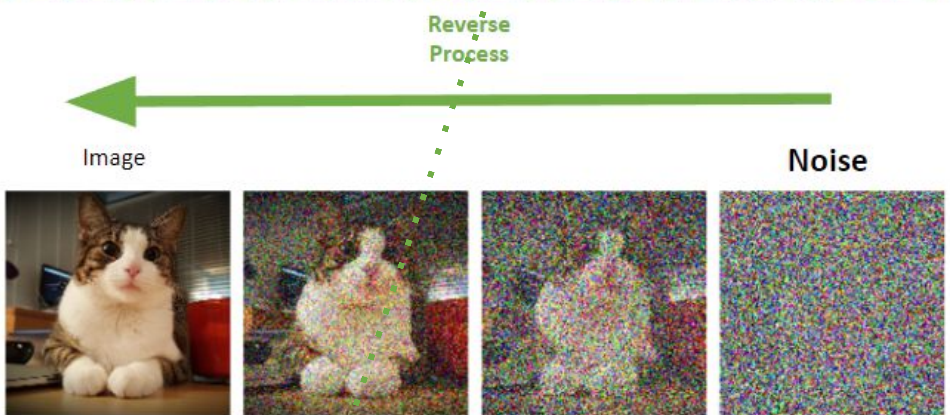
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

$$q(x_{t-1}|x_t, x_0) = \boxed{\mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I)} \tag{12}$$

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

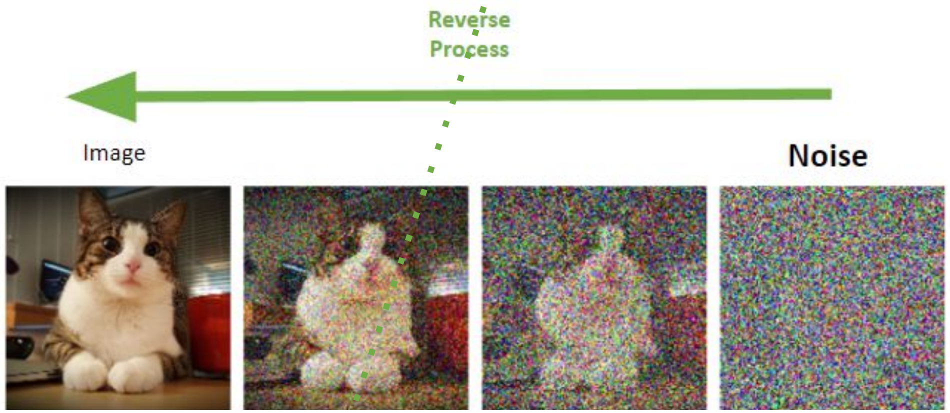
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

$$q(x_{t-1}|x_t, x_0) = \boxed{\mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I)} \tag{12}$$

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))} \qquad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

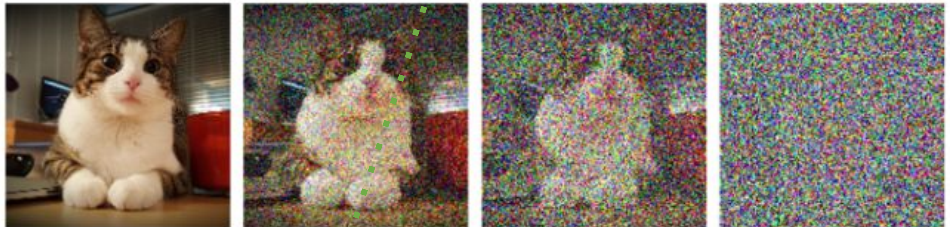
$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I) \tag{12}$$

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

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

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



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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I) \tag{12}$$

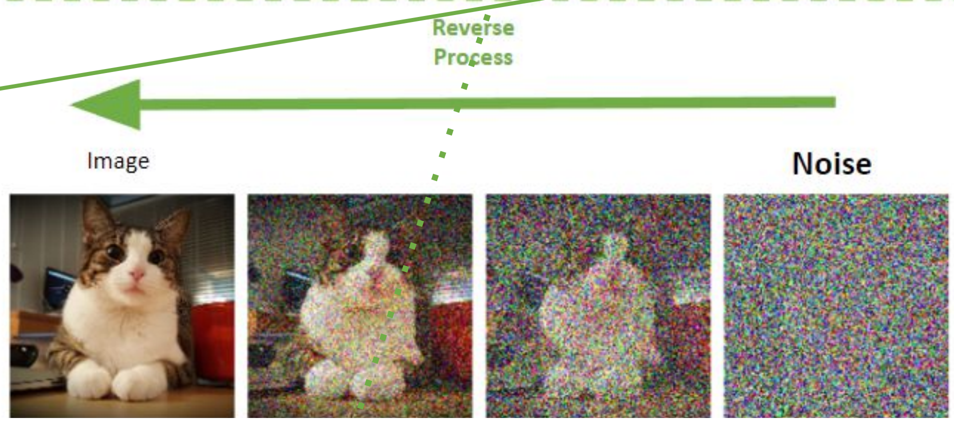
$$\boxed{\mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t)), \tilde{\beta}_t I)}$$

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

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



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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\boxed{\epsilon - \epsilon_\theta(x_t)}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \qquad (10)$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \qquad (11)$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I) \qquad (12)$$

Korea Institute of Science and Technology

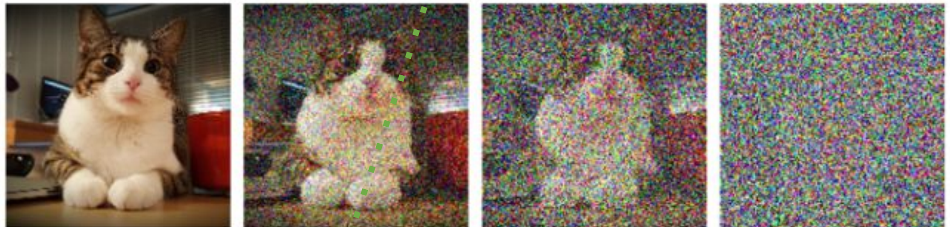
Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Reverse Process

Image

Noise



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

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\epsilon - \epsilon_\theta(x_t)$$

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

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \qquad (10)$$

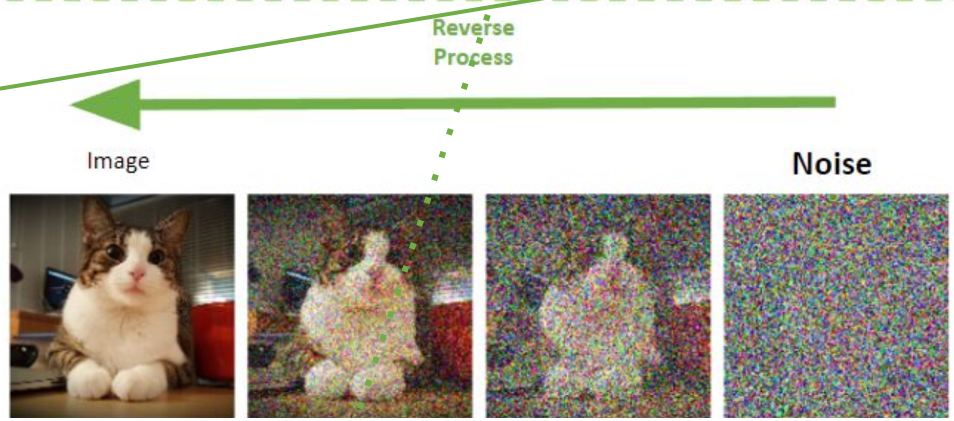
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \qquad (11)$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}\left(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right), \tilde{\beta}_t I\right) \qquad (12)$$

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

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



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

$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \qquad (10)$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \qquad (11)$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I) \qquad (12)$$

$$\epsilon - \epsilon_\theta(x_t)$$

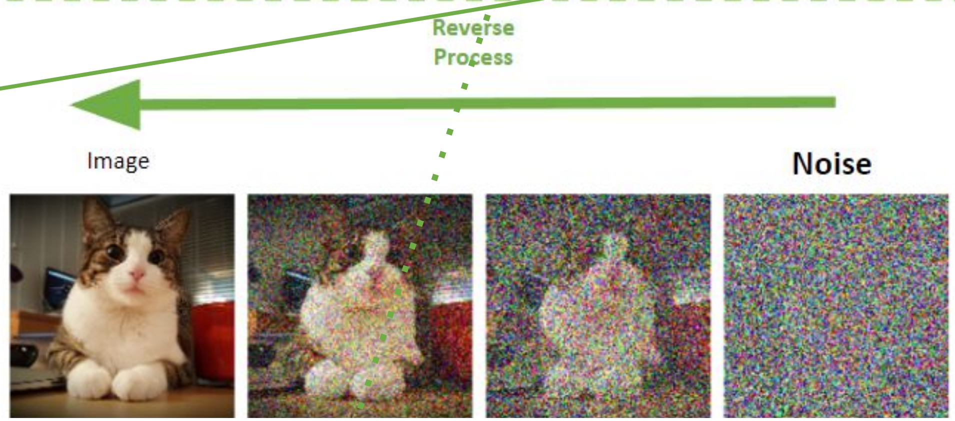
$$\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}||\epsilon - \epsilon_\theta(x_t)||^2$$

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

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$



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

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(\boxed{q(x_{t-1}|x_t, x_0)}||\boxed{p_\theta(x_{t-1}|x_t)})}$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I) \tag{12}$$

$$\epsilon - \epsilon_\theta(x_t)$$

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

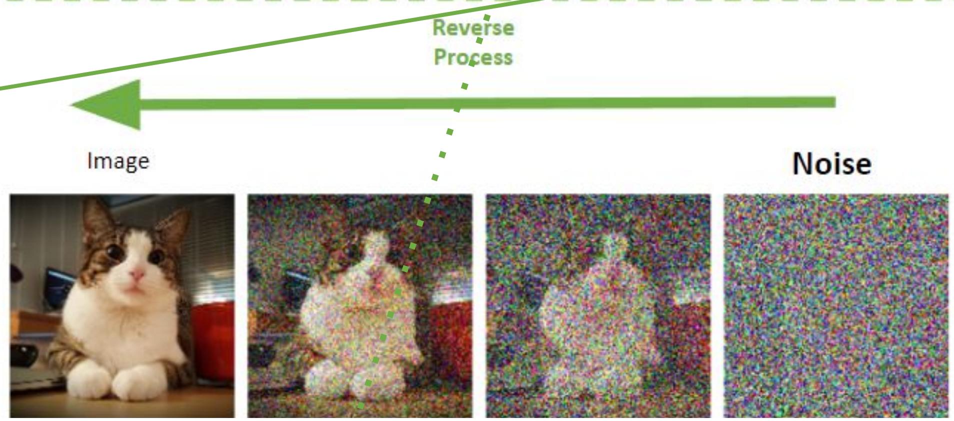
$$= \boxed{\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}}||\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)||^2$$

Korea Institute of Science and Technology

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

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



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

$$-log(p_\theta(x_0)) \le log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

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

$$D_{KL}(q(x_T|x_0)||p(x_T))$$

$$D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \qquad (10)$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \qquad (11)$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}\left(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right), \tilde{\beta}_t I\right) \qquad (12)$$

$$\epsilon - \epsilon_\theta(x_t)$$

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

$$= \frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\bar{\alpha}_t)}\;1\;||\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)||^2$$

Korea Institute of Science and Technology

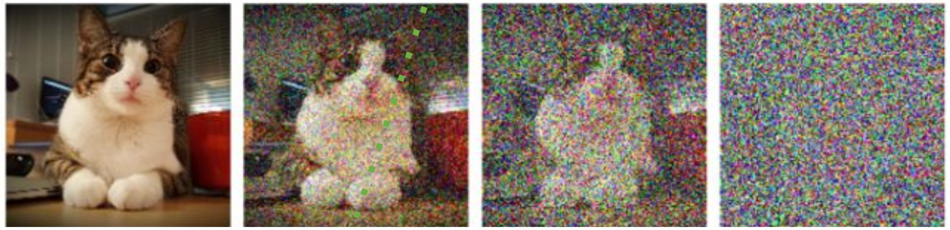
Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Reverse Process

Image

Noise



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

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

$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$D_{KL}(q(x_T|x_0)||p(x_T)) \qquad D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$$

$$\epsilon - \epsilon_\theta(x_t)$$

$$= \quad ||\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)||^2$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \qquad (10)$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \qquad (11)$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon), \tilde{\beta}_t I) \qquad (12)$$

Korea Institute of Science and Technology

Training

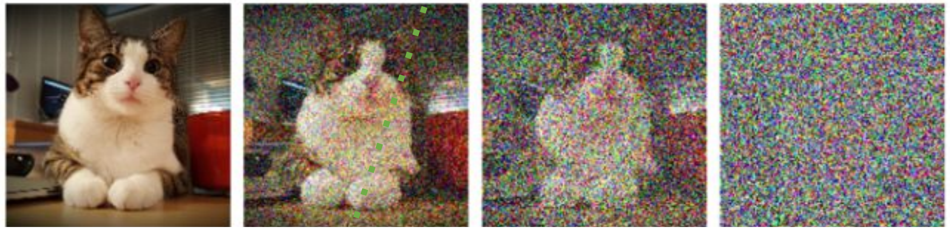
$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

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

Reverse Process

Image

Noise



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

$$-log(p_\theta(x_0)) \leq log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

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

$$D_{KL}(q(x_T|x_0)||p(x_T)) \qquad D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \tag{10}$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \tag{11}$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}\left(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right), \tilde{\beta}_t I\right) \tag{12}$$

$$\epsilon - \epsilon_\theta(x_t)$$

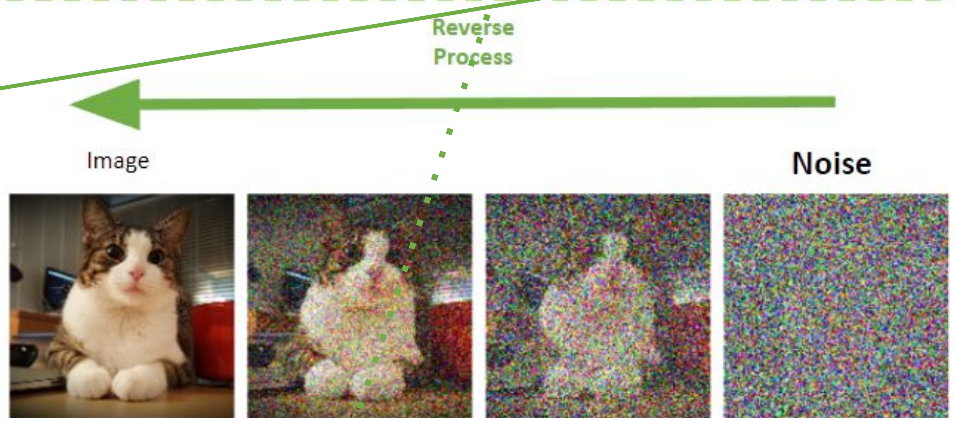
$$= ||\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)||^2$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

Training

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T)\prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

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



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

$$-log(p_\theta(x_0)) \le log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}$$

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

$$D_{KL}(q(x_T|x_0)||p(x_T))$$

$$D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$$

$$\epsilon - \epsilon_\theta(x_t)$$

$$= ||\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)||^2$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

$$\tilde{\beta}_t := \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \quad (10)$$

$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t \quad (11)$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}\left(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon\right), \tilde{\beta}_t I\right) \quad (12)$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

Korea Institute of Science and Technology

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{D} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x;\mu_\theta^i(\mathbf{x}_1,1),\sigma_1^2)\,dx$$

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

Training

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

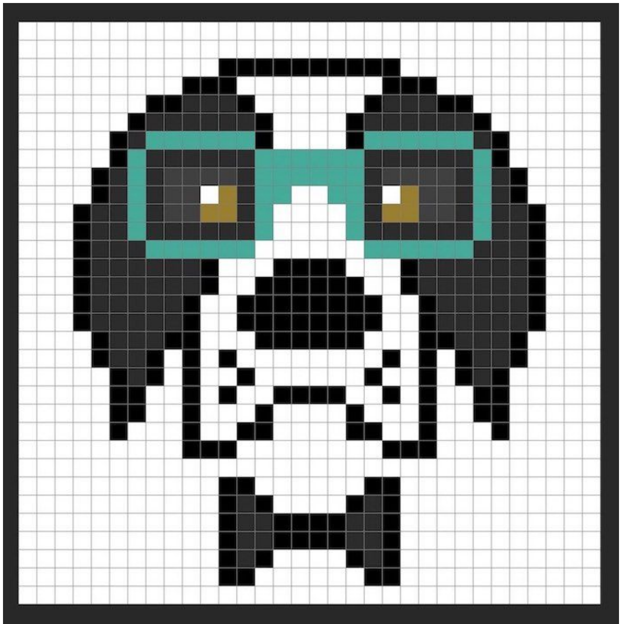
$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$



$$\boxed{\begin{gathered} p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{D} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x; \mu_\theta^i(\mathbf{x}_1, 1), \sigma_1^2)\, dx \\ \delta_+(x) = \begin{cases} \infty & \text{if } x = 1 \\ x + \frac{1}{255} & \text{if } x < 1 \end{cases} \qquad \delta_-(x) = \begin{cases} -\infty & \text{if } x = -1 \\ x - \frac{1}{255} & \text{if } x > -1 \end{cases} \end{gathered}}$$

Training

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

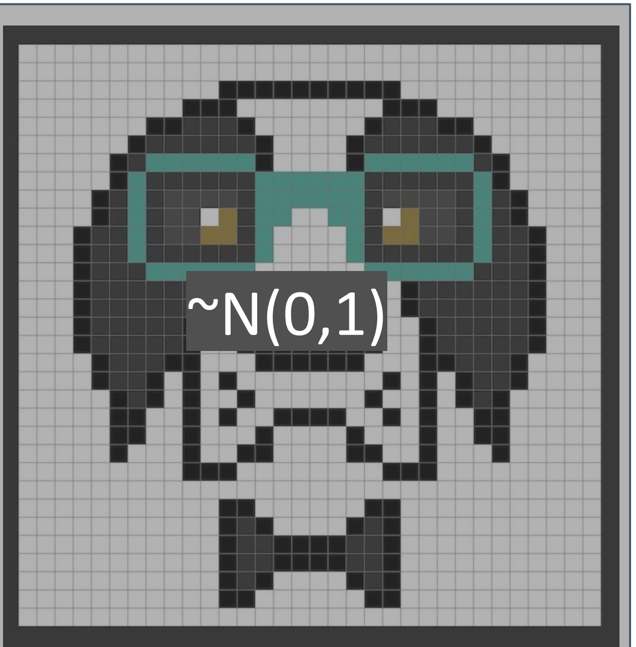
$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$



$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{D}\int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x;\mu_\theta^i(\mathbf{x}_1,1),\sigma_1^2)\,dx$$

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

Training

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

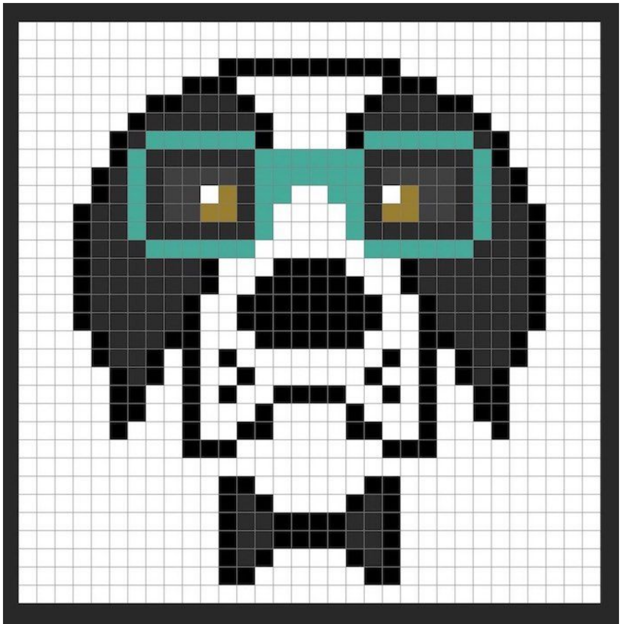
$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$



Number of pixels

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{\boxed{D}} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x; \mu_\theta^i(\mathbf{x}_1, 1), \sigma_1^2)\, dx$$

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

Training

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

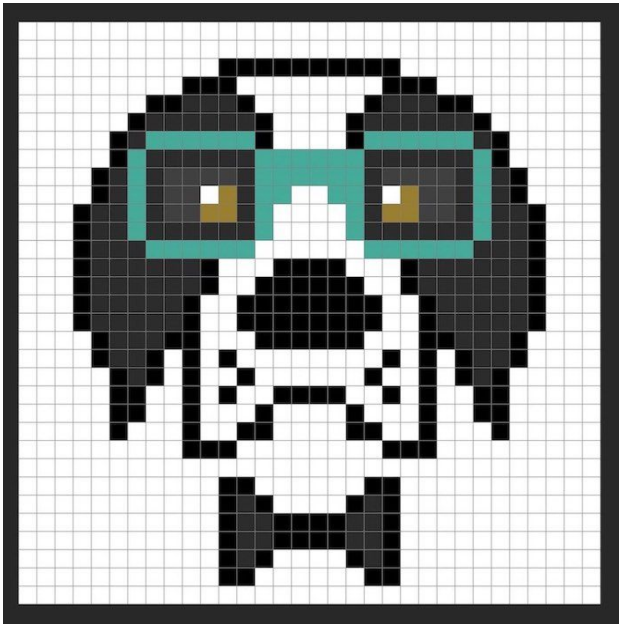
$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$



Number of pixels

Border range

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{\boxed{D}} \boxed{\int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)}} \mathcal{N}(x; \mu_\theta^i(\mathbf{x}_1, 1), \sigma_1^2)\, dx$$

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

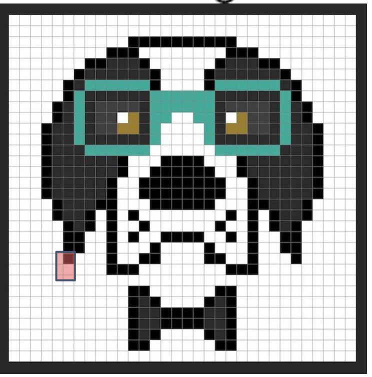
$$= \boxed{log(\frac{q(x_T|x_0)}{p(x_T)})} + \sum_{t=2}^{T} \boxed{log(\frac{q(x_{t-1}|x_t, x_0)}{p_\theta(x_{t-1}|x_t)})} - \boxed{logp_\theta(x_0|x_1)}$$

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

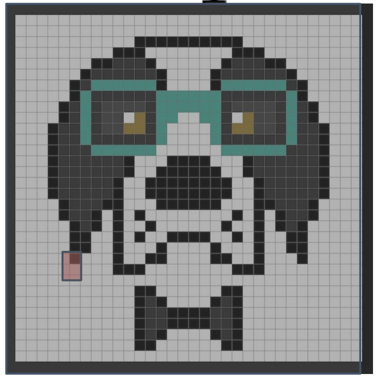
$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

x_0



x_1



Number of pixels

Border range

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{D} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x; \mu_\theta^i(\mathbf{x}_1, 1), \sigma_1^2)\, dx$$

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

Training

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

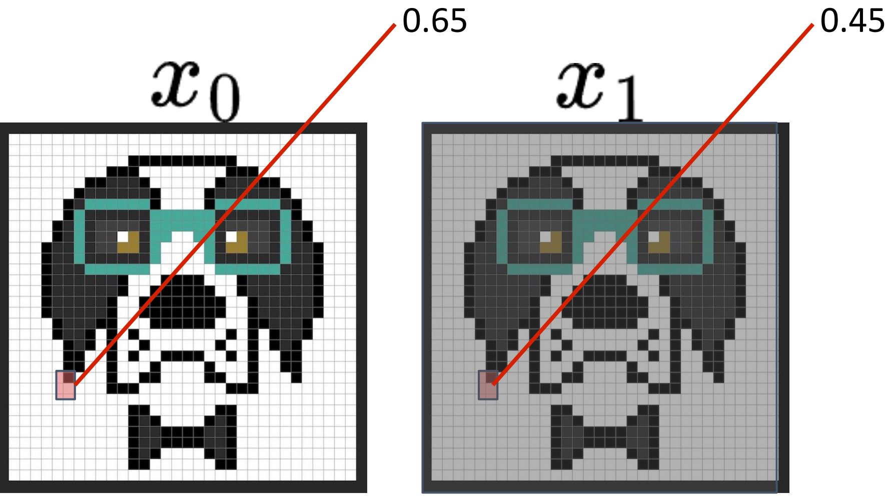
$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$



Number of pixels

Border range

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{\boxed{D}} \boxed{\int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)}} \mathcal{N}(x;\mu_\theta^i(\mathbf{x}_1,1),\sigma_1^2)\,dx$$

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

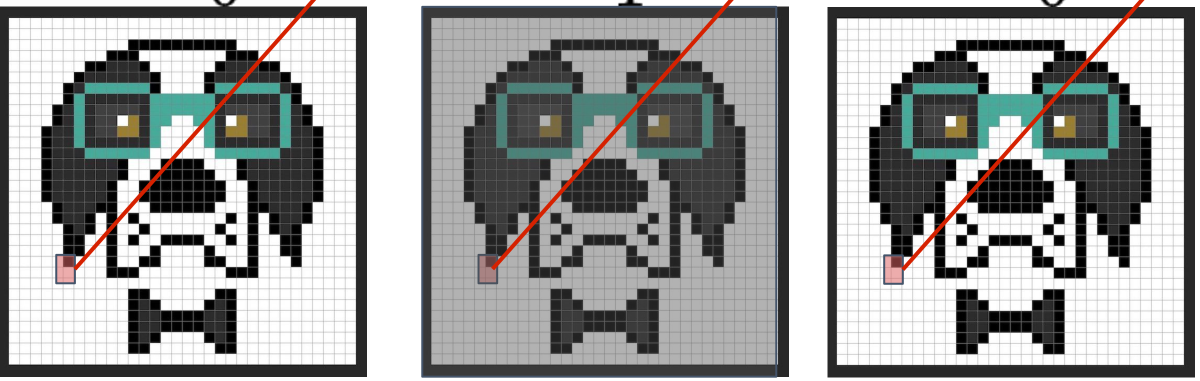
$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

0.65 0.45 predicted 0.55

x_0 x_1 x_0



Number of pixels

Border range

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{D} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x; \mu_\theta^i(\mathbf{x}_1, 1), \sigma_1^2)\, dx$$

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

Training

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

$$-log(p_\theta(x_0)) \le \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

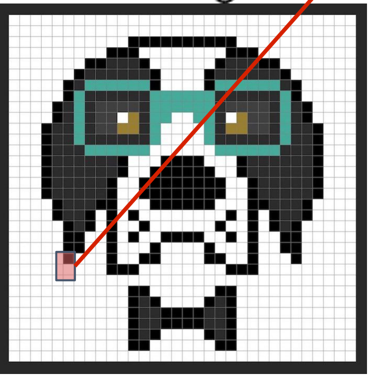
$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t,x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

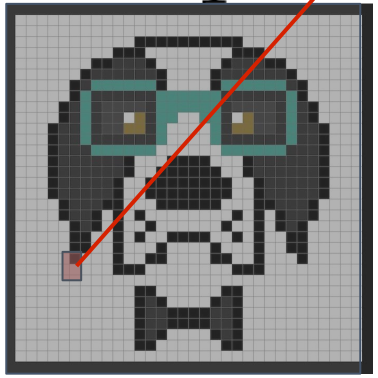
0.65

x_0



0.45

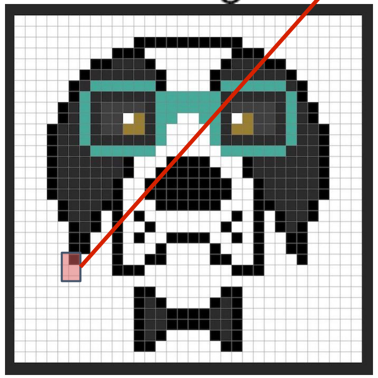
x_1



predicted

x_0

0.55



Number of pixels

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{\boxed{D}} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x;\mu_\theta^i(\mathbf{x}_1,1),\sigma_1^2)\,dx$$

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

Algorithm 2 Sampling

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

Training

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

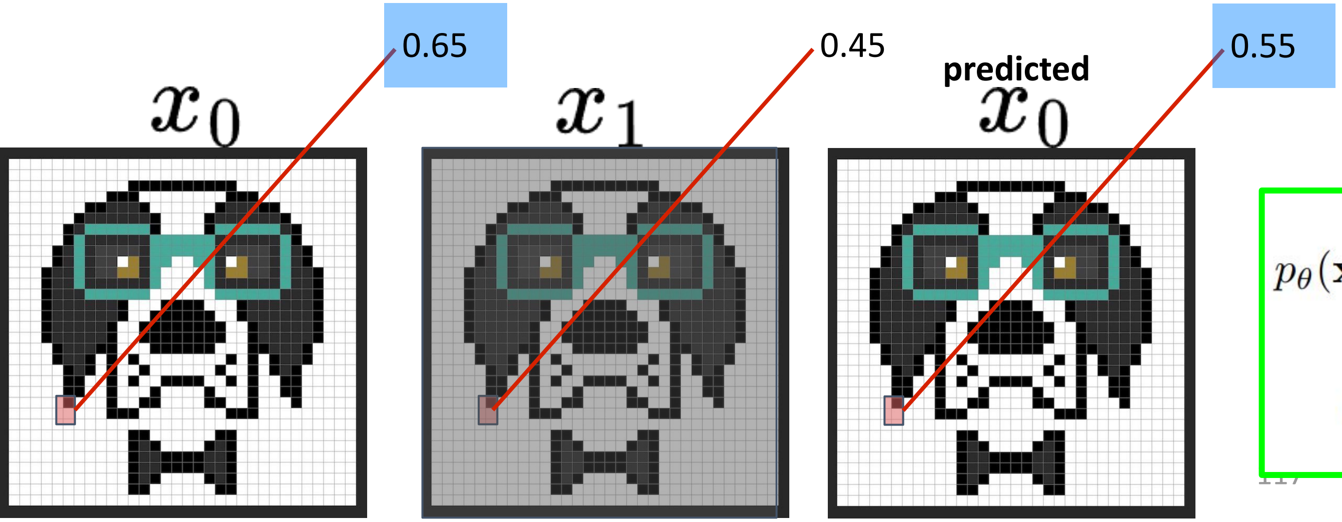
$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$



Number of pixels

$$p_\theta(\mathbf{x}_0|\mathbf{x}_1) = \prod_{i=1}^{\boxed{D}} \int_{\delta_-(x_0^i)}^{\delta_+(x_0^i)} \mathcal{N}(x; \mu_\theta^i(\mathbf{x}_1, 1), \sigma_1^2)\, dx$$

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

Training

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

Korea Institute of Science and Technology

Training

Minimize!

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

Korea Institute of Science and Technology

Training

Minimize!

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\cancel{\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}}$$

$$\boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

Training

Minimize!

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\cancel{\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}} \quad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$= ||\epsilon - \epsilon_\theta(x_t)||^2$$

Korea Institute of Science and Technology

Training

Minimize!

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

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

$$\cancel{\boxed{D_{KL}(q(x_T|x_0)||p(x_T))}} \quad \boxed{D_{KL}(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))}$$

$$\boxed{= ||\epsilon - \epsilon_\theta(x_t)||^2}$$

Korea Institute of Science and Technology

Training

Minimize!

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

$$-log(p_\theta(x_0)) \leq \boxed{log\frac{\prod_{t=1}^{T} q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)}}$$

Diffusion Loss

$$\boxed{= ||\epsilon - \epsilon_\theta(x_t)||^2}$$

$$\mathbb{E}_{t,x_0,\epsilon}[||\epsilon - \epsilon_\theta(x_t, t)||^2]$$

Korea Institute of Science and Technology

Algorithms

Algorithm 1 Training

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

Algorithm 2 Sampling

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

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t^2\mathbf{I})$$

Korea Institute of Science and Technology

Qualitative Results

Korea Institute of Science and Technology

Qualitative Results

Paper List

- Generative Modeling by Estimating Gradients of the Data Distribution (NeurIPS, 2019)
- Denoising Diffusion Probabilistic Models (DDPM, NeurIPS 2020)
- Improved Denoising Diffusion Probabilistic Models (2021)
- Denoising diffusion Implicit Models (DDIM. 2021 ICLR)
- Score-Based Generative Modeling through Stochastic Differential Equations (ICLR,2021)
- Diffusion Models Beat GANs on Image Synthesis (2021 NeurIPS)
- Tackling the Generative Learning Trilemma with Denoising Diffusion GANs (ICLR, 2022)
- High-Resolution Image Synthesis with Latent Diffusion Models (Stable Diffusion) (CVPR, 2022)
- Classifier-Free Diffusion Guidance (2022)
- Consistency Models (2023)

Qualitative Results

Paper List

- Generative Modeling by Estimating Gradients of the Data Distribution (NeurIPS, 2019)
- Denoising Diffusion Probabilistic Models (DDPM, NeurIPS 2020)
- Improved Denoising Diffusion Probabilistic Models (2021)
- Denoising diffusion Implicit Models (DDIM. 2021 ICLR)
- Score-Based Generative Modeling through Stochastic Differential Equations (ICLR,2021)
- Diffusion Models Beat GANs on Image Synthesis (2021 NeurIPS)
- Tackling the Generative Learning Trilemma with Denoising Diffusion GANs (ICLR, 2022)
- High-Resolution Image Synthesis with Latent Diffusion Models (Stable Diffusion) (CVPR, 2022)
- Classifier-Free Diffusion Guidance (2022)
- Consistency Models (2023)

Korea Institute of Science and Technology

Qualitative Results

Paper List

- Generative Modeling by Estimating Gradients of the Data Distribution (NeurIPS, 2019)
- Denoising Diffusion Probabilistic Models (DDPM, NeurIPS 2020)
- Improved Denoising Diffusion Probabilistic Models (2021)
- Denoising diffusion Implicit Models (DDIM. 2021 ICLR)
- Score-Based Generative Modeling through Stochastic Differential Equations (ICLR,2021)
- Diffusion Models Beat GANs on Image Synthesis (2021 NeurIPS)
- Tackling the Generative Learning Trilemma with Denoising Diffusion GANs (ICLR, 2022)
- High-Resolution Image Synthesis with Latent Diffusion Models (Stable Diffusion) (CVPR, 2022)
- Classifier-Free Diffusion Guidance (2022)
- Consistency Models (2023)

Korea Institute of Science and Technology