

### **Prerequisites: Recap on VAE**



ELBO Loss (Variational Lower Bound)

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

Reparameterization Trick

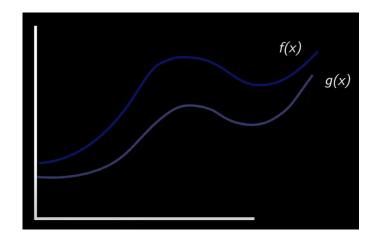
$$egin{aligned} \mathbf{z} &\sim q_{\phi}(\mathbf{z}|\mathbf{x}^{(i)}) = \mathcal{N}(\mathbf{z}; oldsymbol{\mu}^{(i)}, oldsymbol{\sigma}^{2(i)} oldsymbol{I}) \ \mathbf{z} &= oldsymbol{\mu} + oldsymbol{\sigma} \odot oldsymbol{\epsilon}, ext{where } oldsymbol{\epsilon} \sim \mathcal{N}(0, oldsymbol{I}) \end{aligned}$$
; Reparameterization trick.

#### [Note]

KL Divergence is always non-negative!

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

= difference between two distributions





# Why?

```
Expanding VAE Loss (ELBO: Evidence Lower Bound)
                                              Note: KL Divergence
DKL (90(ZIX) 11po(ZIX))
                                                 DEL (p(x) || q(x)) = I p(x) log \frac{p(x)}{q(x)} dx
        = \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
        = fq 6(z1x) log q o(z1x) po(x) dz
        = Jap(z1x) logpo(x) dz + Jap(z1x) log ap(z1x) dz
                                                                      Note: Conditional Expectation
                                                                         E(x1x=y)= Jafx11/21y)dx
        = log po(x) + Jq (z/x) log qo(z/x) dz
                                                                                 = try) Jxfx, (x,y)dx
        = log po(x) + Jq o(z/x) log a o(z/x) dz > DKL (qodz/x) lipo(z))
         = logpola + f q o(z(x) & log q o(z(x) - log po(x) & dz
          = logpo(x) + Ezngo(zlx) [log a o (zlx) - logpo(x|z) 3 dz
           - logpo(N)+ DK (qolz1211polz)) - Ez-qolz12) [logpo(N2)]
   :. DKL(qo(zlx) (po(zvx)) = log po(x)+ DKL(qp(zlx)) qo(z))- Eznqo(zlx) log po(x)z)
Minimize: DKL (qo(z(x))1/20(z(x)) - log po(x)
            = minimize difference = maximize likelihood of generating x
between approx L true
posterior distributions
        - LUATE = + leg pola)-Dec (galzin lipotzin) & log pola)
```





#### **Overview**



# **Learning Manifold(=Distribution) of Data**

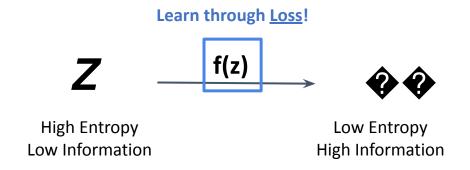




#### **Overview**



# **Learning Manifold(=Distribution) of Data**



Recap:

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

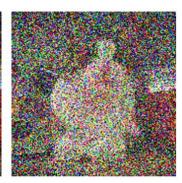


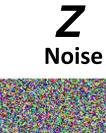
#### **Overview**



















Forward Process

Fixed Encoder!

Reverse Process

Where NN comes in! ( $\theta$ )

Noise를 얼만큼 빼서 이미지를 만들자





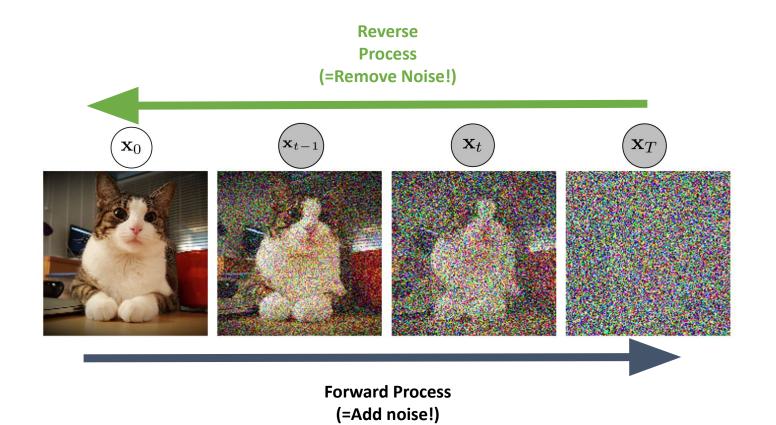
Reverse Process (=Remove Noise!)



Forward Process (=Add noise!)

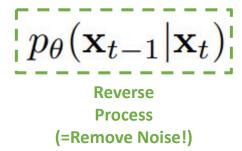


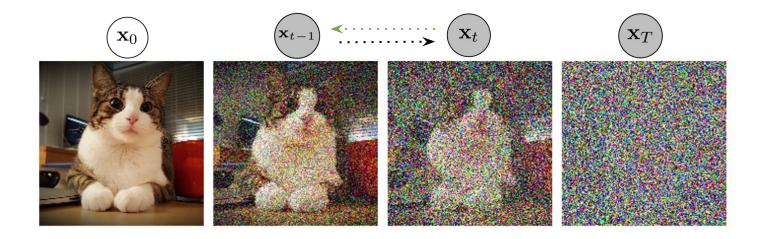










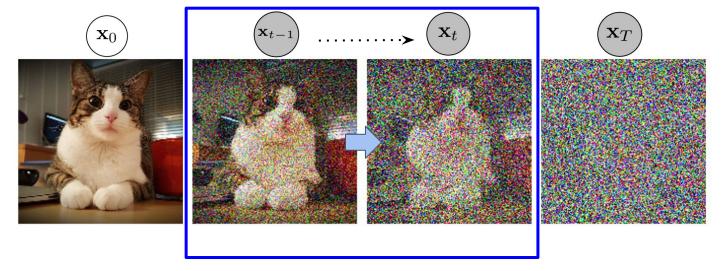


Forward Process (=Add noise!)

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





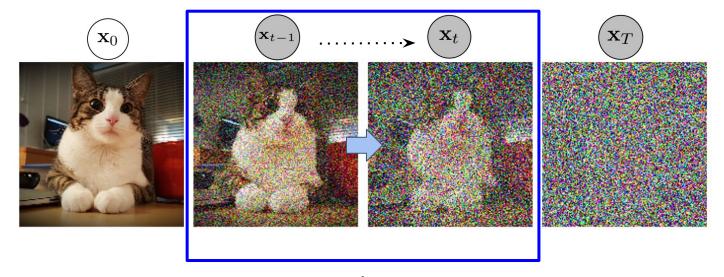


Forward Process (=Add noise!)

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







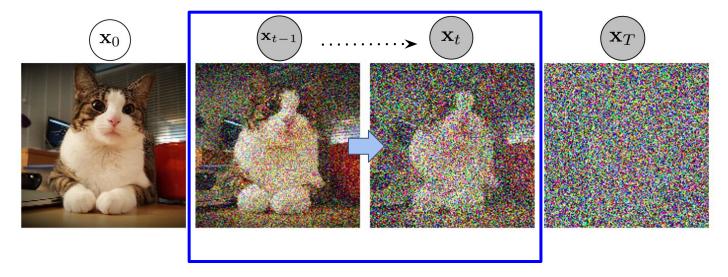
Forward Process (=Add noise!)

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

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







Forward Process (=Add noise!)

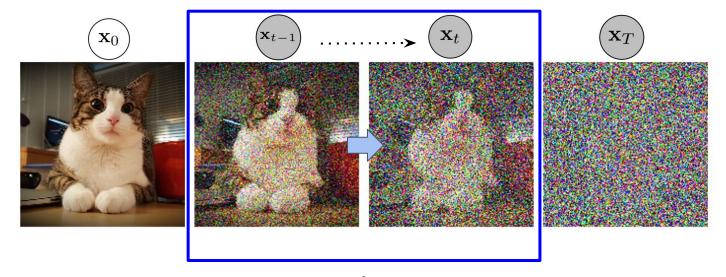
$$q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

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

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







Forward Process (=Add noise!)

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

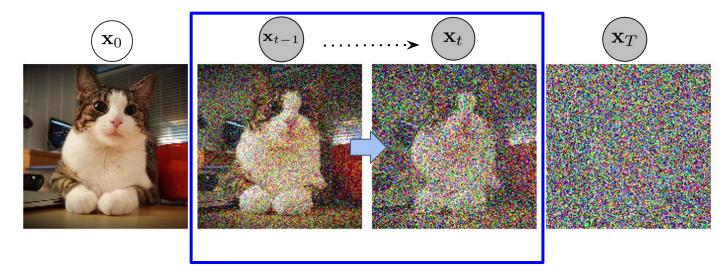
$$egin{aligned} \mathbf{z} &\sim q_{\phi}(\mathbf{z}|\mathbf{x}^{(i)}) = \mathcal{N}(\mathbf{z}; oldsymbol{\mu}^{(i)}, oldsymbol{\sigma}^{2(i)} oldsymbol{I}) \ \mathbf{z} &= oldsymbol{\mu} + oldsymbol{\sigma} \odot oldsymbol{\epsilon}, ext{ where } oldsymbol{\epsilon} \sim \mathcal{N}(0, oldsymbol{I}) \end{aligned}$$
 ; Reparameterization trick.

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

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







Forward Process (=Add noise!)

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

$$egin{aligned} \mathbf{z} &\sim q_{\phi}(\mathbf{z}|\mathbf{x}^{(i)}) = \mathcal{N}(\mathbf{z}; oldsymbol{\mu}^{(i)}, oldsymbol{\sigma}^{2(i)}oldsymbol{I}) \ \mathbf{z} &= oldsymbol{\mu} + oldsymbol{\sigma} \odot oldsymbol{\epsilon}, ext{ where } oldsymbol{\epsilon} \sim \mathcal{N}(0, oldsymbol{I}) \end{aligned}$$
 ; Reparameterization trick.

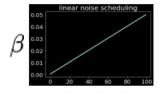
Korea Institute of Science and Technology

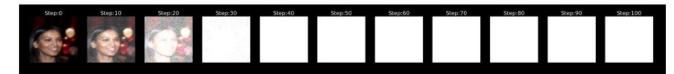
$$x_t = \boxed{a} * x_{t-1} + \boxed{b} * noise \qquad x_t = \boxed{1 - \beta_t} * x_{t-1} + \boxed{\sqrt{\beta_t}} * noise \qquad q(\mathbf{x}_{1:T}|\mathbf{x}_0) \coloneqq \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad q(\mathbf{x}_t|\mathbf{x}_{t-1}) \coloneqq \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I})}$$



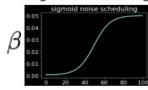
- Overview
  - · Linear, Quad, Sigmoid, Cosine, ...

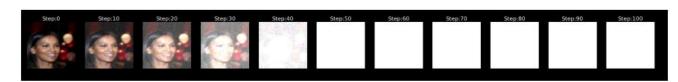
#### ✓ Linear scheduling



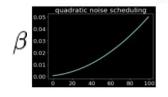


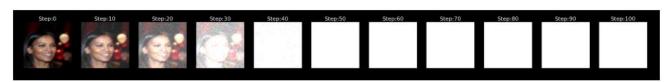
✓ Sigmoid scheduling

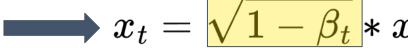




✓ Quadratic scheduling









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





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

Reverse **Process** 

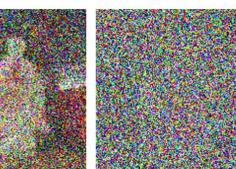


**Image** 









Noise

#### **Forward Process**

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





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

Reverse **Process** 

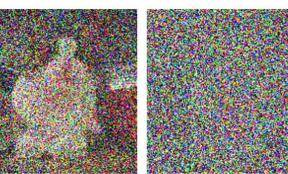


**Image** 









Noise

**Forward Process** 

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





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

Reverse **Process** 

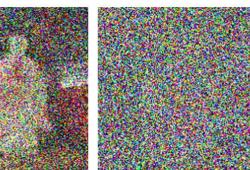


**Image** 





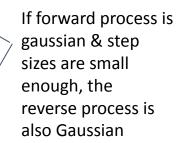




Noise

#### **Forward Process**

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





Entire reverse process to learn with NN  $oldsymbol{ heta}$ 

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

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

Reverse **Process** 

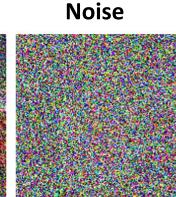


Image









**Forward Process** 

**Forward Process** given image (x0)

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

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



Entire reverse process to learn with NN  $oldsymbol{ heta}$ 

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

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

Reverse **Process** 

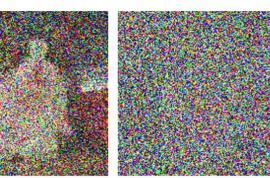


Image









Noise

**Forward Process** 

**Forward Process** given image (x0)

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

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





to learn with NN  $\theta$ 

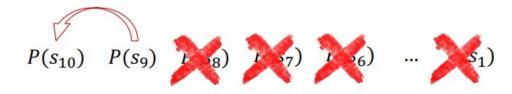
Entire reverse process

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

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

#### 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



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

**Forward Process** given image (x0)

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

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



#### **Additional Notations**

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

Reverse Process

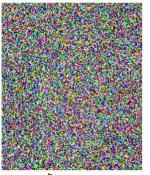


Image Noise









Forward Process

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



 $eta_t$  : Degree of noise added at time step t



#### **Additional Notations**

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

Reverse **Process** 

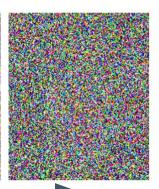


**Image** 









Noise

**Forward Process** 

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



 $oldsymbol{eta_t}$  : Degree of noise added at time step t

$$lpha_t = 1 - eta_t$$



#### **Additional Notations**

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

Reverse **Process** 



**Image** 









Noise

**Forward Process** 

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



$$eta_t$$
 : Degree of noise added at time step t

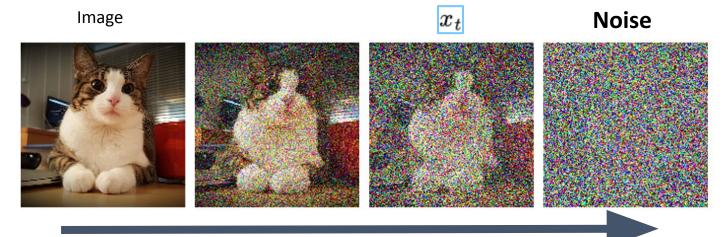
$$\alpha_t = 1 - \beta_t$$

$$\stackrel{-}{lpha_t} = \prod_{i=1}^t a_i$$



#### **Forward Process**





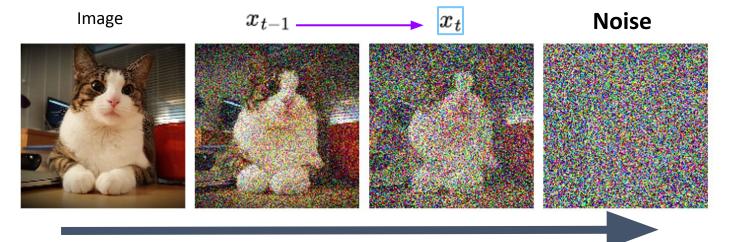
Can't we directly sample  $x_t$  at any time step t?

# Forward Process

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







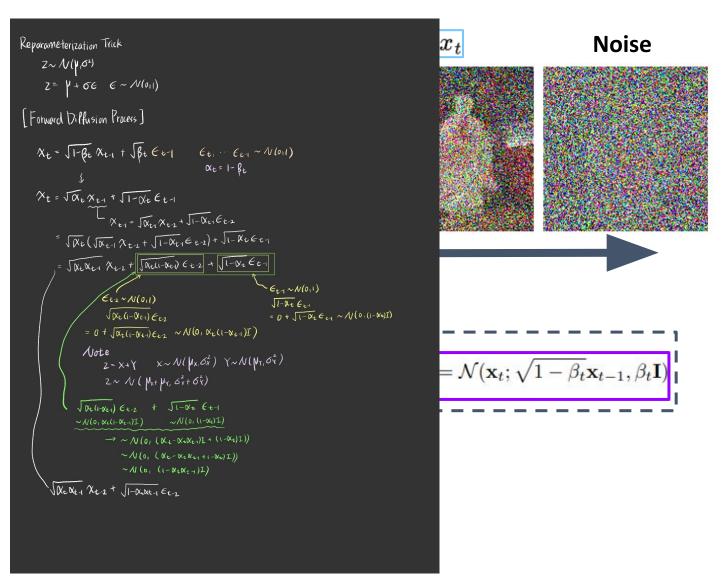
Can't we directly sample  $x_t$  at any time step t?

# Forward Process

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







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





at any time step *t*?

```
TREAL XE-2 + JI-XXXX-1 Et-2
Reparameterization Trick
                                                                                                                                                                                                                                                                                                                            [ xt-2 = \( \times \times \tag{1-\times \tan
                Z~ N(1,02)
                 Z= V+66 E~ N(011)
                                                                                                                                                                                                                                                                                          = \sqrt{\alpha_{t}\alpha_{t1}} \left( \sqrt{\alpha_{t2}} \alpha_{t-3} + \sqrt{1-\alpha_{t2}} \varepsilon_{t-3} \right) + \sqrt{1-\alpha_{t}\alpha_{t1}} \varepsilon_{t-2}
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Can't we directly sample
                                                                                                                                                                                                                                                                                                Vat X+1 (X+2 X+3+ (X+X+1 (1-X+2) E+3+ J1-X+X+1 E+2
  [Forward Dilfusion Process]
                                                                                                                                                                                                                                                                                                                                                             ~N(0, Q+Q+, (1-Q+2)) ~N(0, 1-Q+Q+1)
           Xt = 11-βt Xt-1 + 1βt Et-1 Et ... Et - ~ NO11)
αt = 1- βt
            \lambda_t = \sqrt{\alpha_t} \chi_{t-1} + \sqrt{1-\chi_t} \epsilon_{t-1}
                                                                                                                                                                                                                                                                                            Xt1= Jan Xt2+ JI-anEt2
                          = \Qt(\Qt-1 \chi_{t-2} + \sqrt{1-\alpha_{t-1}} \end{array} + \sqrt{1-\alpha_{t}}
                                                                                                                                                                                                                                                                                                                                                                                                                                              \overline{\mathbf{X}}_{t} = \overset{\mathbf{t}}{\mathbf{T}} \mathbf{X}_{t}
                                                                                                                                                                                                                                                                                         = Vatatina, Xo + VI-NERLINA, E
                                JAEREN XEZ + JAHLINEW EEZ + JI-VE EEN
                                                                                                                                                                                                                                                                                            = V Xx + VI - Xx E
                                                      (Et2~N(01)
                                                        JXE (1-0/E-1) EE-2
                                                                                                                                                                        = 0 + \sqrt{1 - \alpha_t} \epsilon_{t-1} \sim \mathcal{N}(0, (1 - \lambda \epsilon)I)
                                            = 0 + \( \alpha_t(1-\alpha_{t-1}) \epsilon_{t-2} \quad \( N(0, \alpha_t(1-\alpha_{t-1}) \in ) \)
                                                                                                                                                                                                                                                                                                              \therefore \chi_t = \sqrt{\alpha_t} \chi_0 + \sqrt{1 - \alpha_t} \epsilon
                                                        Z=X+Y \times \sim N(V_X, \sigma_X^2) Y \sim N(V_Y, \sigma_Y^2)
                                                                                                                                                                                                                                                                                                                	o we can directly sample x_{
m t} at any time step t
                                                         Z~ /v ( Mx+ Mx, 6x+ 6x)
                                        TOE (1-NE) E E-2 + JI-NE E-1
                                                            → ~ N(0, (X+-X+X+)I+(1-X+)I)
                                                                      ~ N (o, (1- 0+0++) I)
                        TREAL XEZ + JI-XXXX ELZ
```





at any time step *t*?

```
TREAL XE-2 + JI-XXXX-1 Et-2
Reparameterization Trick
                                                                                                                                                                                                                                                                                                            [ xt-2 = \( \times \times \tag{1-\times \tan
               Z~ N(1,02)
                Z= V+66 E~ N(011)
                                                                                                                                                                                                                                                                           = \sqrt{\alpha_{t}\alpha_{t1}} \left( \sqrt{\alpha_{t2}} \alpha_{t-3} + \sqrt{1-\alpha_{t2}} \varepsilon_{t-3} \right) + \sqrt{1-\alpha_{t}\alpha_{t1}} \varepsilon_{t-2}
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      Can't we directly sample
                                                                                                                                                                                                                                                                                  Vat X+1 (X+2 X+3+ (X+X+1 (1-X+2) E+3+ J1-X+X+1 E+2
  [Forward Dilfusion Process]
                                                                                                                                                                                                                                                                                                                                           Xt = 11-βt Xt-1 + 1βt Et-1 Et ... Et - ~ NO11)
αt = 1- βt
                                                                                                                                                                                                                                                                                                                                                                             > ~N(0, 1-X+X+1X+2)
            x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-x_t} \epsilon_{t-1}
                                                                                                                                                                                                                                                                             L X = J (X + X + 2 + J 1 - (X + 1 + 2)
                        = \Qt(\Qt-1 \chi_{t-2} + \sqrt{1-\alpha_{t-1}} \end{array} + \sqrt{1-\alpha_{t}}
                                                                                                                                                                                                                                                                                                                                                                                                                                   Qt = T Q;
                                                                                                                                                                                                                                                                           = Vatatina, Xo + VI-Ktatina, E
                              Jacker Xt-2 + Jack-aty Et-2 + JI-VE Et-
                                                                                                                                                                                                                                                                             = V Xx + VI - Xx E
                                                   (Et2~N(011)
                                                     JXE (1-0/E-1) EE-2
                                                                                                                                                               = 0 + JI-at E = ~ N(0, (1- XE)I)
                                         = 0 + \( \alpha_t(1-\alpha_{t-1}) \epsilon_{t-2} \quad \( N(0, \alpha_t(1-\alpha_{t-1}) \in ) \)
                                                     Z=X+Y \times \sim N(V_X, \sigma_X^2) Y \sim N(V_Y, \sigma_Y^2)
                                                                                                                                                                                                                                                                                                \rightarrow we can directly sample x_{t} at any time step t
                                                      Z~ N ( Mx+ Mx, 6x+6x)
                                      TOE (1-NE) E E-2 + JI-NE E-1
                                                         → ~ N(0, (X+-X+X+)I+(1-X+)I)
                                                                  ~ N (o, (1- 0+0++) I)
                      TREAL XEZ + JI-XXXX ELZ
```





Representation Trick

$$2 \sim \mathcal{N}(\mu, \sigma)$$
 $2 = \frac{1}{2} + 66 = 6 \sim \mathcal{N}(\kappa_1)$ 

[Finance Different Praces]

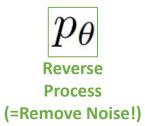
 $\chi_{-} = \sqrt{\log \chi_{+1}} + \sqrt{\log \kappa_{+1}} +$ 

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

$$x_t = \sqrt{rac{1}{lpha_t x_0}} + \sqrt{1-lpha_t \epsilon}$$







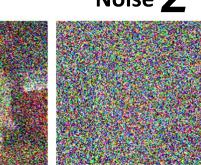












For VAE, minimize:

Noise  $oldsymbol{Z}_{L_{VAE}} = -log(p_{ heta}(x)) \leq -log(p_{ heta}(x)) + D_{KL}(q_{ heta}(z|x)||p_{ heta}(z|x))$ 

**Forward Process** (=Add noise!)







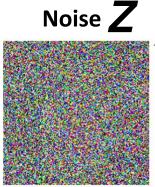












 $x_0$ 

Forward Process (=Add noise!)





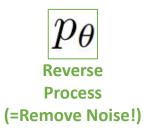
Extend to Diffusion

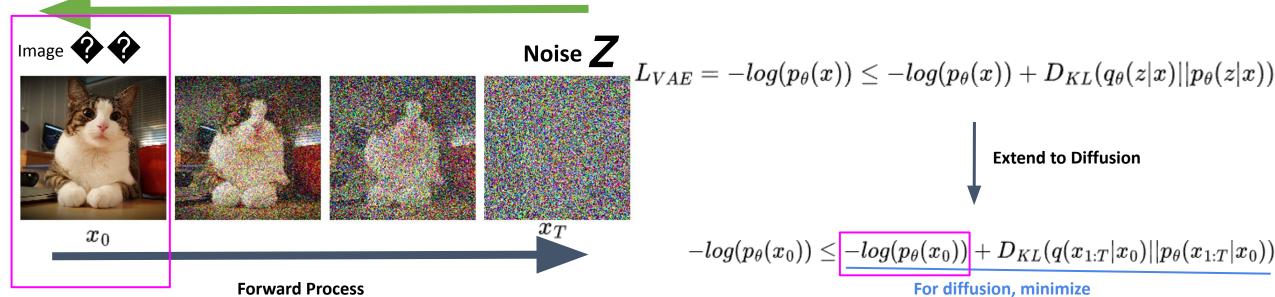
$$-log(p_{ heta}(x_0)) \leq \underline{-log(p_{ heta}(x_0)) + D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0))}$$

For diffusion, minimize









**Extend to Diffusion** 

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

For diffusion, minimize

Maximize the likelihood of generating real data

(=Add noise!)





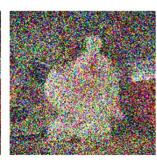


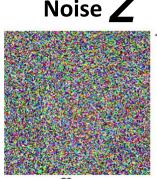




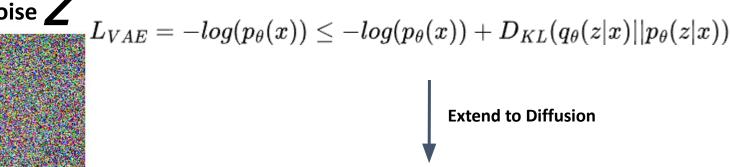
 $x_0$ 







Noise **Z** 



Extend to Diffusion

**Forward Process** (=Add noise!)



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

For diffusion, minimize

also minimize the difference between the real and estimated posterior distributions



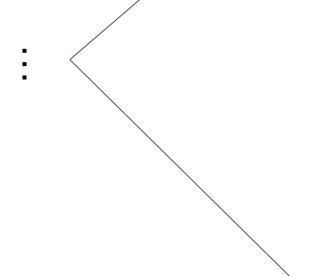


$$-log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0))$$

#### **Background:**

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

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





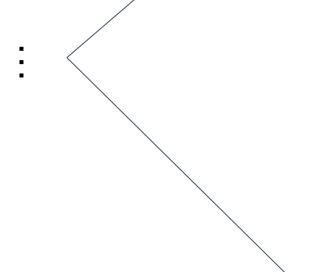
$$-log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0))$$

#### **Background:**

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

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

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

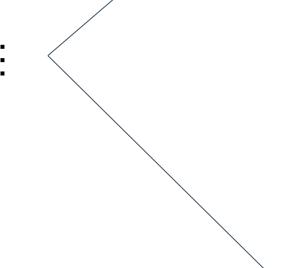




$$-log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0))$$

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

$$egin{align} -log(p_{ heta}(x_0)) & \leq -log(p_{ heta}(x_0)) + egin{align} D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \ & = \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)} dx_{1:T} \ & p_{ heta}(x_{1:T}|x_0) \end{pmatrix} \end{array}$$





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



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

$$-log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + egin{array}{c} D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T} \ &p_{ heta}(x_{1:T}|x_0) \ &= rac{p_{ heta}(x_0|x_{1:T})p_{ heta}(x_{1:T})}{p_{ heta}(x_0)} \end{array}$$
 Bayesian Rule  $= rac{p_{ heta}(x_0|x_{1:T})p_{ heta}(x_{1:T})}{p_{ heta}(x_0)}$ 



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



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



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



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

$$-log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0))$$
  $= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T}$   $p_{ heta}(x_{1:T}|x_0)$  Bayesian Rule  $= rac{p_{ heta}(x_0|x_{1:T})p_{ heta}(x_{1:T})}{p_{ heta}(x_0)}$   $= rac{p_{ heta}(x_0,x_{1:T})}{p_{ heta}(x_0)}$   $= rac{p_{ heta}(x_0,x_{1:T})}{p_{ heta}(x_0)}$ 



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



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

$$-log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + egin{array}{c} D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \\ = \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T} \\ p_{ heta}(x_{1:T}|x_0) \\ = \frac{p_{ heta}(x_0|x_{1:T})p_{ heta}(x_{1:T})}{p_{ heta}(x_0)} \\ = rac{p_{ heta}(x_0,x_{1:T})}{p_{ heta}(x_0)} \\ = \frac{p_{ heta}(x_0,x_{1:T})}{p_{ heta}(x_0)} \\ = \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)p_{ heta}(x_0)}{p_{ heta}(x_{0:T})}dx_{1:T} \end{aligned}$$





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

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

$$-log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + egin{aligned} D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T} \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)p_{ heta}(x_0)}{p_{ heta}(x_{0:T})}dx_{1:T} \end{aligned}$$



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



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

$$egin{aligned} -log(p_{ heta}(x_0)) &\leq -log(p_{ heta}(x_0)) + egin{aligned} D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T} \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)p_{ heta}(x_0)}{p_{ heta}(x_{0:T})}dx_{1:T} \ &= \int q(x_{1:T}|x_0)[lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0))]dx_{1:T} \end{aligned}$$

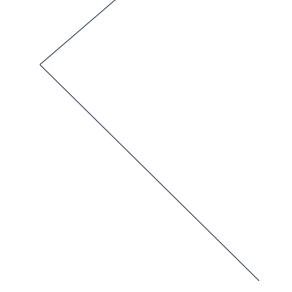


$$-log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0))$$



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

$$egin{aligned} -log(p_{ heta}(x_0)) &\leq -log(p_{ heta}(x_0)) + egin{aligned} D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T} \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)p_{ heta}(x_0)}{p_{ heta}(x_{0:T})}dx_{1:T} \ &= \int q(x_{1:T}|x_0)[lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0))]dx_{1:T} \end{aligned}$$





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



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

$$egin{aligned} -log(p_{ heta}(x_0)) &\leq -log(p_{ heta}(x_0)) + egin{aligned} D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T} \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)p_{ heta}(x_0)}{p_{ heta}(x_{0:T})}dx_{1:T} \ &= \int q(x_{1:T}|x_0)[lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0))]dx_{1:T} \ &= lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0)) \end{aligned}$$



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

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

$$egin{aligned} -log(p_{ heta}(x_0)) &\leq -log(p_{ heta}(x_0)) + egin{aligned} D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T} \ &= \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)p_{ heta}(x_0)}{p_{ heta}(x_{0:T})}dx_{1:T} \ &= \int q(x_{1:T}|x_0)[lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0))]dx_{1:T} \ &= lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0)) \ &- log(p_{ heta}(x_0)) + lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0)) \end{aligned}$$





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

$$egin{align*} -log(p_{ heta}(x_0)) & \leq -log(p_{ heta}(x_0)) + D_{KL}(q(x_{1:T}|x_0)||p_{ heta}(x_{1:T}|x_0)) \ & = \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{1:T}|x_0)}dx_{1:T} \ & = \int q(x_{1:T}|x_0)lograc{q(x_{1:T}|x_0)p_{ heta}(x_0)}{p_{ heta}(x_{0:T})}dx_{1:T} \ & = \int q(x_{1:T}|x_0)[lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0))]dx_{1:T} \ & = lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + lograc{q(x_{0:T}|x_0)}{p_{ heta}(x_0)} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + lograc{q(x_{0:T}|x_0)}{p_{ heta}(x_0)} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + lograc{q(x_{0:T}|x_0)}{p_{ heta}(x_0)} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + log rac{q(x_0)}{p_{ heta}(x_0)} + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \leq -log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) + log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \ & -log(p_{ heta}(x_0)) \ & -l$$

Variational Lower bound for diffusion

 $-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{q(x_{1:T}|x_0)}$ 



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

#### **Variational Lower bound for diffusion**

$$p_{\theta}(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^{T} p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t) \qquad q(\mathbf{x}_{1:T} | \mathbf{x}_0) := \prod_{t=1}^{T} q(\mathbf{x}_t | \mathbf{x}_{t-1})$$

reverse process (=remove noise)

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

forward process (=add noise)



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

Reverse Process

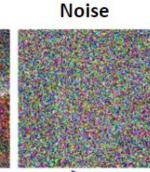


Image









#### Forward **Process**

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



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

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

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





$$-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{ig|p_{ heta}(x_{0:T})ig|}$$

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

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





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

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

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





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

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

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

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





$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \end{aligned}$$





$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \end{aligned}$$





$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{aligned} \log rac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \ -log(p(x_T))) + log(rac{\prod_{t=1}^T q(x_t|x_{t-1})}{\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}) \end{aligned}$$



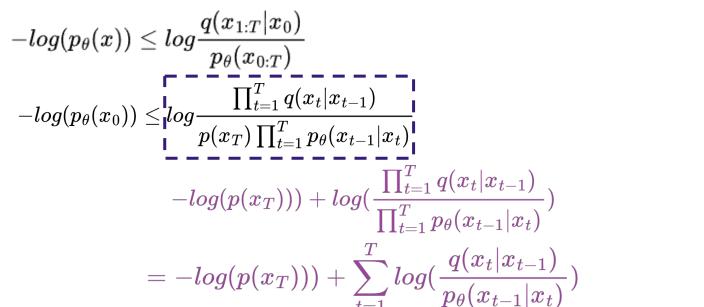


$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{aligned} \log rac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \ -log(p(x_T))) + log(rac{\prod_{t=1}^T q(x_t|x_{t-1})}{\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}) \ & = -log(p(x_T))) + \sum_{t=1}^T log(rac{q(x_t|x_{t-1})}{p_{ heta}(x_{t-1}|x_t)}) \end{aligned}$$





Training 
$$log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})}$$



$$egin{aligned} & t=1 & p_{ heta}(x_{t-1}|x_{t}) \ & = -log(p(x_{T}))) + \sum_{t=0}^{T} log(rac{q(x_{t}|x_{t-1})}{p_{ heta}(x_{t-1}|x_{t})}) + log(rac{q(x_{1}|x_{0})}{p_{ heta}(x_{0}|x_{1})}) \end{aligned}$$





$$\begin{split} -log(p_{\theta}(x)) & \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) & \leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ & -log(p(x_{T}))) + log(\frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{\prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}) \\ & = -log(p(x_{T}))) + \sum_{t=1}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$







$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \left| log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})} \right| \\ &- log(p(x_{0})) + log(\frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{\prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}) \\ &= -log(p(x_{T}))) + \sum_{t=1}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) + log(\frac{q(x_{t}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) & \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) & \leq \left[ log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})} \right] \\ & -log(p(x_{T}))) + log(\frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{\prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}) \\ & = -log(p(x_{T}))) + \sum_{t=1}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$

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

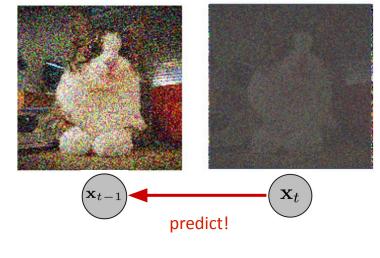
predict!





$$\begin{split} -log(p_{\theta}(x)) & \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) & \leq \left| log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})} \right| \\ & -log(p(x_{T}))) + log(\frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{\prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}) \\ & = -log(p(x_{T}))) + \sum_{t=1}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$

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



= High Variance :(





$$\begin{split} -log(p_{\theta}(x)) & \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) & \leq \left| log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})} \right| \\ & -log(p(x_{T}))) + log(\frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{\prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}) \\ & = -log(p(x_{T}))) + \sum_{t=1}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$

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





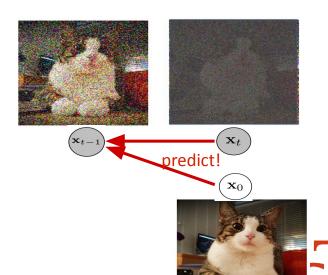






$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left[lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}
ight] & q(x_0) \ -log(p(x_T)) & + log(rac{\prod_{t=1}^T q(x_t|x_{t-1})}{\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}) \ & = -log(p(x_T))) + \sum_{t=1}^T log(rac{q(x_t|x_{t-1})}{p_{ heta}(x_{t-1}|x_t)}) \ & = -log(p(x_T))) + \sum_{t=2}^T log(rac{q(x_t|x_{t-1})}{p_{ heta}(x_{t-1}|x_t)}) + log(rac{q(x_1|x_0)}{p_{ heta}(x_0|x_1)}) \end{aligned}$$

$$egin{aligned} q(x_t|x_{t-1}) &= rac{q(x_{t-1}|x_t)q(x_t)}{q(x_{t-1})} \ &= rac{q(x_{t-1}|x_t, x_0)q(x_t|x_0)}{q(x_{t-1}|x_0)} \end{aligned}$$





$$\begin{aligned} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \left| log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})} \right| \\ &- log(p(x_{0})) + log(\frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{\prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}) \\ &= -log(p(x_{T}))) + \sum_{t=1}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{aligned}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \left| log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})} \right| \\ &- log(p(x_{T}))) + log(\frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{\prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}) \\ &= -log(p(x_{T}))) + \sum_{t=1}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) + log(\frac{q(x_{t}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \left| log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})} \right| \\ &- log(p(x_{T}))) + log(\frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{\prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}) \\ &= -log(p(x_{T}))) + \sum_{t=1}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{t-1})}{p_{\theta}(x_{t-1}|x_{t})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})q(x_{t}|x_{0})}{p_{\theta}(x_{t-1}|x_{0})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$



$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})q(x_{t}|x_{0})}{p_{\theta}(x_{t-1}|x_{0})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$



$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = -log(p(x_T))) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)q(x_t|x_0)}{p_{ heta}(x_{t-1}|x_t)q(x_{t-1}|x_0)}) + log(rac{q(x_1|x_0)}{p_{ heta}(x_0|x_1)}) \end{aligned}$$



$$\begin{split} -log(p_{\theta}(x)) & \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) & \leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ & = -log(p(x_{T}))) + \underbrace{\sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})q(x_{t}|x_{0})}{p_{\theta}(x_{t-1}|x_{0})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})})}_{p_{\theta}(x_{0}|x_{1})} \\ & = -log(p(x_{T}))) + \underbrace{\sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + \sum_{t=2}^{T} log(\frac{q(x_{t}|x_{0})}{p_{\theta}(x_{t-1}|x_{0})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ \end{split}$$



$$\begin{split} -log(p_{\theta}(x)) & \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) & \leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})q(x_{t}|x_{0})}{p_{\theta}(x_{t-1}|x_{t})q(x_{t-1}|x_{0})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + \sum_{t=2}^{T} log\frac{q(x_{t}|x_{0})}{q(x_{t-1}|x_{0})} + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})q(x_{t}|x_{0})}{p_{\theta}(x_{t-1}|x_{0})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + \sum_{t=2}^{T} log\frac{q(x_{t}|x_{0})}{q(x_{t-1}|x_{0})} + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + log\frac{q(x_{T}|x_{0})}{q(x_{1}|x_{0})} + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) & \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) & \leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})q(x_{t}|x_{0})}{p_{\theta}(x_{t-1}|x_{t})q(x_{t-1}|x_{0})}) + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + \sum_{t=2}^{T} log\frac{q(x_{t}|x_{0})}{q(x_{t-1}|x_{0})} + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ & = -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + \underline{log}\frac{q(x_{T}|x_{0})}{q(x_{1}|x_{0})} + log(\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log (\frac{q(x_{t-1}|x_{t},x_{0})q(x_{t}|x_{0})}{p_{\theta}(x_{t-1}|x_{0})}) + log (\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log (\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + \sum_{t=2}^{T} log \frac{q(x_{t}|x_{0})}{q(x_{t-1}|x_{0})} + log (\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log (\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + \underline{log} \frac{q(x_{T}|x_{0})}{q(x_{1}|x_{0})} + log (\frac{q(x_{1}|x_{0})}{p_{\theta}(x_{0}|x_{1})}) \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log (\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + \underline{log}(q(x_{T}|x_{0})) - logp_{\theta}(x_{0}|x_{1}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + log(q(x_{T}|x_{0})) - logp_{\theta}(x_{0}|x_{1}) \\ &= log(q(x_{T}|x_{0})) - log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + log(q(x_{T}|x_{0})) - logp_{\theta}(x_{0}|x_{1}) \\ &= \underline{log(q(x_{T}|x_{0})) - log(p(x_{T})))} + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + log(q(x_{T}|x_{0})) - logp_{\theta}(x_{0}|x_{1}) \\ &= \underline{log(q(x_{T}|x_{0})) - log(p(x_{T})))} + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \\ &= \underline{log(\frac{q(x_{T}|x_{0})}{p(x_{T})})} + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + log(q(x_{T}|x_{0})) - logp_{\theta}(x_{0}|x_{1}) \\ &= log(q(x_{T}|x_{0})) - log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \\ &= \boxed{log(\frac{q(x_{T}|x_{0})}{p(x_{T})})} + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \end{split}$$





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + log(q(x_{T}|x_{0})) - logp_{\theta}(x_{0}|x_{1}) \\ &= log(q(x_{T}|x_{0})) - log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \\ &= \boxed{log(\frac{q(x_{T}|x_{0})}{p(x_{T})})} + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \end{split}$$

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





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + log(q(x_{T}|x_{0})) - logp_{\theta}(x_{0}|x_{1}) \\ &= log(q(x_{T}|x_{0})) - log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \\ &= \boxed{log(\frac{q(x_{T}|x_{0})}{p(x_{T})})} + \sum_{t=2}^{T} \boxed{log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1})} \end{split}$$

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





$$\begin{split} -log(p_{\theta}(x)) &\leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ -log(p_{\theta}(x_{0})) &\leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ &= -log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) + log(q(x_{T}|x_{0})) - logp_{\theta}(x_{0}|x_{1}) \\ &= log(q(x_{T}|x_{0})) - log(p(x_{T}))) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1}) \\ &= \boxed{log(\frac{q(x_{T}|x_{0})}{p(x_{T})})} + \sum_{t=2}^{T} \boxed{log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{0}|x_{1})} \\ \boxed{D_{KL}(q(x_{T}|x_{0})||p(x_{T}))} &D_{KL}(q(x_{t-1}|x_{t},x_{0})||p_{\theta}(x_{t-1}|x_{t})) \end{split}$$

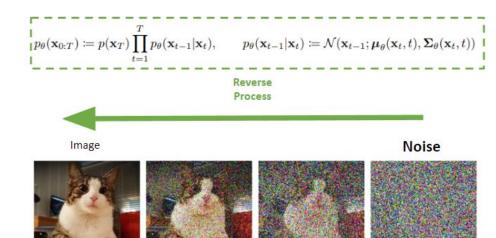


$$-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} - log(p_{ heta}(x_0)) \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} - log(p_{ heta}(x_0)) \leq lograc{\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} - log(p_{ heta}(x_0)|x_1) - log(p_{ heta}(x_0)|x_1) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - log(p_{ heta}(x_0)|x_1) + \sum_{t=2}^T log(q(x_t|x_t)|x_t) + \sum_{t=2}^T log(q(x_t|x_t)|x_t) + log(p_{ heta}(x_t|x_t)) + log(p_{ heta}(x_t|x_t|x_t)) + log(p_{ heta}(x_t|x_t)) + log(p_{ heta}(x_t|x_t)) + l$$



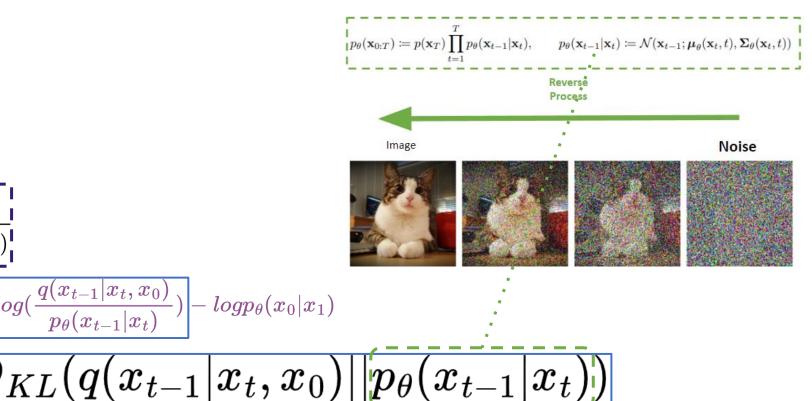
 $\left|D_{KL}(q(x_T|x_0)||p(x_T))
ight|$ 

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{aligned} lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \ & = egin{aligned} log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T egin{aligned} log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \end{aligned}$$



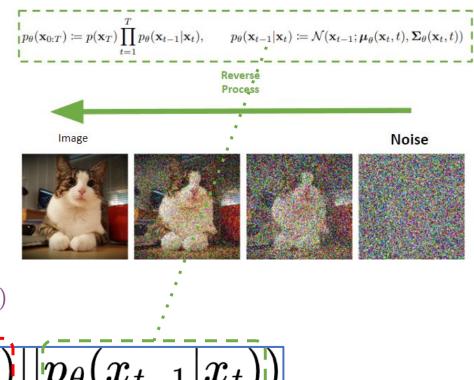


$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T \left| log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) 
ight| - logp_{ heta}(x_0|x_1) \ D_{KL}(q(x_T|x_0)) | p(x_T)) 
ight| D_{KL}(q(x_T|x_0)) | p(x_T) 
ight) | p$$





$$-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} - log(p_{ heta}(x_0)) \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} - log(p_{ heta}(x_0)) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) - logp_{ heta}(x_0|x_1)$$

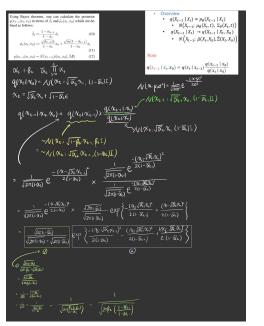


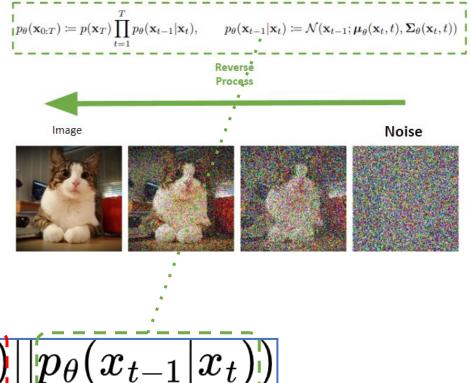


$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left|lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}
ight| \ & = \left|log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) 
ight| \end{aligned}$$

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

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



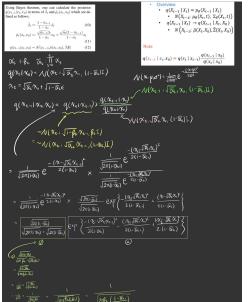


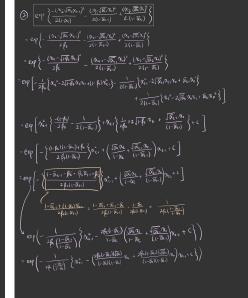


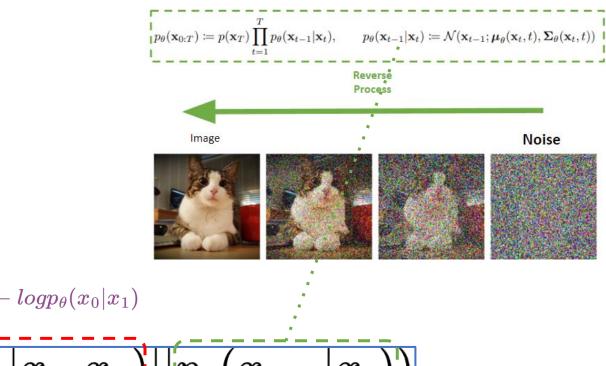
$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left[lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}
ight] \ & = \left[log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}
ight] \end{aligned}$$

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

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





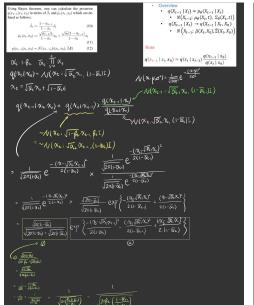




$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{aligned} & \prod_{t=1}^T q(x_t|x_{t-1}) \ & p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t) \end{aligned} \ & = egin{aligned} log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T egin{aligned} log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)} \end{aligned}$$

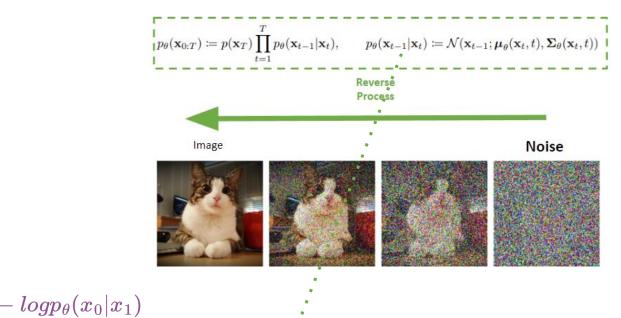
$$D_{KL}(q(x_T|x_0)||p(x_T))$$

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



```
\begin{split} & \left[ \bigotimes_{i} \left( \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right) + \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right] \\ & = \bigotimes_{i} \left\{ \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \\ & = \bigotimes_{i} \left\{ \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} + \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \\ & = \bigotimes_{i} \left\{ \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i}, \chi_{i})^{2} + \sum_{i} (\lambda_{i}, \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} + \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i}, \chi_{i})^{2} + \sum_{i} (\lambda_{i}, \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} + \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i}, \chi_{i})^{2} + \sum_{i} (\lambda_{i}, \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} + \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i}, \chi_{i})^{2} + \sum_{i} (\lambda_{i}, \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i}, \chi_{i})^{2} + \sum_{i} (\lambda_{i}, \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i}, \chi_{i})^{2} + \sum_{i} (\lambda_{i}, \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i}, \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i} \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ - \frac{(\lambda_{i} - \sqrt{\lambda_{i}} \chi_{i})^{2}}{2(1 - 2\lambda_{i})} \right\} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i} \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i} \chi_{i})^{2} \right] \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i} \chi_{i})^{2} \right] \right\} \\ & = \bigotimes_{i} \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i} \chi_{i})^{2} \right] \left\{ \lambda_{i}^{2} - 2 \left[ \sum_{i} (\lambda_{i} \chi_{i})^{2} \right] \right\}
```

```
\begin{split} &= \exp\left(-\frac{1}{2\rho\left(\frac{|\Sigma_{\infty}|}{|\Sigma_{\infty}|}\right)} \frac{1}{2} \lambda_{t, t}^{2} - \left(\frac{2\rho\left(|\Sigma_{\infty}|}{|\Sigma_{\infty}|}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||\Sigma_{\infty}||
```

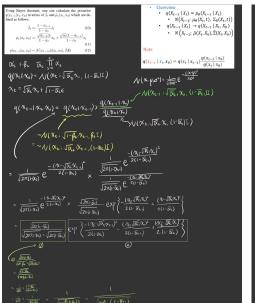




$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \end{aligned}$$

$$D_{KL}(q(x_Tert x_0)ertert p(x_T))$$

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



```
\exp\left[-\frac{1}{2\beta_{k}}\Big\{\chi_{\xi_{k}}^{2}-2\sqrt{L\beta_{k}}\chi_{\xi_{k}}\chi_{\xi_{k}}+(L-\beta_{k})\tilde{\chi_{\xi_{k}}}\Big\}-\frac{1}{2(L-\overline{\chi_{k}})}\Big\{\chi_{\xi_{k}}^{2}-2\sqrt{\overline{\chi_{\xi_{k}}}}\chi_{\xi_{k}}\chi_{0}+\overline{\chi_{\xi_{k}}}\tilde{\chi_{\xi_{k}}}\chi_{0}\Big\}
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     + 1 { ( 1 - 1 ) { ( 1 - 2 ) [ 1 - 2 ] [ 1 - 2 ] [ 1 - 2 ] [ 1 - 2 ] }
= exp \( \lambda_{k+1}^2 \right\) \frac{-(1-\frac{1}{4}\ell)}{2\beta_k} - \frac{1}{2(1-\beta_k)} \right\} + \lambda_{k+1} \right\} \frac{\pmu}{2\beta_k} \times 2\sqrt{1-\beta_k} \lambda_k + \frac{\beta_k \cdot N_k}{(1-\beta_k)} \right\} + C \\ \]
                                                                                                                                                                           \left\{ \begin{array}{c} \left\{ 1 - \widetilde{\Omega}_{t+1} - \beta \varepsilon + \beta_t \widetilde{\Omega}_{t+1} + \beta_t \widetilde{\Omega}_{t+1} + \beta_t \widetilde{\Omega}_{t+1} + \left( \frac{\widetilde{\mathcal{N}}_t \widetilde{\mathcal{N}}_t}{1 - \widetilde{\mathcal{N}}_t} + \frac{\widetilde{\mathcal{N}}_t \widetilde{\mathcal{N}}_t}{1 - \widetilde{\mathcal{N}}_t} \right) \widetilde{\mathcal{N}}_{t+1} + \varepsilon \right\} \\ \\ \left\{ \begin{array}{c} 2 \beta_t (|r\widetilde{\mathcal{N}}_{t+1}|) \\ -\widetilde{\mathcal{N}}_t (|r\widetilde{\mathcal{N}}_{t+1}|) \end{array} \right\} \widetilde{\mathcal{N}}_{t+1}^{-1} + \left( \frac{\widetilde{\mathcal{N}}_t \widetilde{\mathcal{N}}_t}{1 - \widetilde{\mathcal{N}}_t} + \frac{\widetilde{\mathcal{N}}_t \widetilde{\mathcal{N}}_t}{1 - \widetilde{\mathcal{N}}_t} \right) \widetilde{\mathcal{N}}_{t+1} + \varepsilon \right] \\ \\ \left\{ \begin{array}{c} 2 \beta_t (|r\widetilde{\mathcal{N}}_t|) \\ -\widetilde{\mathcal{N}}_t (|r\widetilde{\mathcal{N}}_t|) \end{array} \right\} \widetilde{\mathcal{N}}_{t+1}^{-1} + \left( \frac{\widetilde{\mathcal{N}}_t \widetilde{\mathcal{N}}_t}{1 - \widetilde{\mathcal{N}}_t} + \frac{\widetilde{\mathcal{N}}_t \widetilde{\mathcal{N}}_t}{1 - \widetilde{\mathcal{N}}_t} \right) \widetilde{\mathcal{N}}_{t+1} + \varepsilon \right] \\ \\ \left\{ \begin{array}{c} 2 \beta_t (|r\widetilde{\mathcal{N}}_t|) \\ -\widetilde{\mathcal{N}}_t (|r\widetilde{\mathcal{N}}_t|) \end{array} \right\} \widetilde{\mathcal{N}}_{t+1}^{-1} + \left( \frac{\widetilde{\mathcal{N}}_t \widetilde{\mathcal{N}}_t}{1 - \widetilde{\mathcal{N}}_t} + \frac{\widetilde{\mathcal{N}}_t \widetilde{\mathcal{N}}_t}{1 - \widetilde{\mathcal{N}}_t} \right) \widetilde{\mathcal{N}}_{t+1} + \varepsilon \right] \\ \\ \left\{ \begin{array}{c} 2 \beta_t (|r\widetilde{\mathcal{N}}_t|) \\ -\widetilde{\mathcal{N}}_t (|r\widetilde{\mathcal{N}}_t|) \end{array} \right\} \widetilde{\mathcal{N}}_{t+1}^{-1} + \varepsilon \right\} \widetilde{\mathcal{N}}_{t+1}^{-1} + \varepsilon \widetilde{\mathcal{N}}_t^{-1} 
        \exp\left(-\frac{1}{\frac{1}{2\beta_{k}}\left(\frac{1-\widetilde{\beta_{k+1}}}{1-\widetilde{\beta_{k+1}}}\right)} \left\{ \chi_{k+1}^{2} - \frac{2\beta_{k}(k-\widetilde{\beta_{k}})}{1-\widetilde{\beta_{k}}} \left( \frac{\widetilde{\beta_{k}}, \gamma_{k}}{1-\widetilde{\beta_{k+1}}} + \frac{\sqrt{\widetilde{\beta_{k}}}, \gamma_{k}}{1-\widetilde{\beta_{k+1}}} \right) \chi_{k+1} + C \right\} \right)
= \exp\left(-\frac{1}{\frac{2\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)}{2\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)}}\frac{1}{2}\chi_{x_{0}}^{2} - \left(\frac{\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}} + \frac{2\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}\right)\chi_{x_{0}} + \frac{2}{2}\frac{\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}\right)\chi_{x_{0}} + \frac{2}{2}\frac{\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}} + \frac{2}{2}\frac{\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}\right)\chi_{x_{0}} + \frac{2}{2}\frac{\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}} + \frac{2}{2}\frac{\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}} + \frac{2}{2}\frac{\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}} + \frac{2}{2}\frac{\beta_{0}\left(\frac{|x-y|_{0}}{2}\right)}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}}{\left(\frac{|x-y|_{0}}{2}\right)}\chi_{x_{0}}}
```

$$\begin{split} &= \exp\left(-\frac{1}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)}\frac{\lambda_{k+1}}{(|\mathcal{R}|)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{(|\mathcal{R}|)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{(|\mathcal{R}|)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{(|\mathcal{R}|)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{(|\mathcal{R}|)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)}\frac{\lambda_{k+1}}{(|\mathcal{R}|)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R}|}\right)(|\mathcal{R}|)}\frac{\lambda_{k+1}}{2\rho\left(\frac{2\pi}{|\mathcal{R$$

$$\tilde{\beta}_{t} \coloneqq \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_{t}} \beta_{t}$$

$$\tilde{\mu}_{t}(x_{t}, x_{0}) \coloneqq \frac{\sqrt{\bar{\alpha}_{t-1}} \beta_{t}}{1 - \bar{\alpha}_{t}} x_{0} + \frac{\sqrt{\alpha_{t}} (1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_{t}} x_{t}$$

$$(11)$$

$$(12)$$

**Process** 

 $p_{\theta}(\mathbf{x}_{0:T}) \coloneqq p(\mathbf{x}_T) \prod p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t),$ 

Image

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



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

Noise

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T \left| log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) 
ight| - logp_{ heta}(x_0|x_1) \ D_{KL}(q(x_T|x_0)) | p(x_T) 
ight) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D$$

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

**Process** 

Noise



Image







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

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

$$\tilde{\beta}_{t} \coloneqq \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_{t}} \beta_{t}$$

$$\tilde{\mu}_{t}(x_{t}, x_{0}) \coloneqq \frac{\sqrt{\bar{\alpha}_{t-1}} \beta_{t}}{1 - \bar{\alpha}_{t}} x_{0} + \frac{\sqrt{\alpha_{t}} (1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_{t}} x_{t}$$

$$(11)$$

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



$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T \left| log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) 
ight| \ D_{KL}(q(x_T|x_0)) \left| p(x_T) 
ight) 
ight| D_{KL}(q(x_T|x_0)) \left| p(x_T|x_0) 
ight| D_{KL}(q(x_T|x_0)) \left| p(x_T|x_$$

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

Reverse Process

\*



Image







Noise

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

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

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

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

$$x_t = \sqrt{rac{1}{lpha_t x_0}} + \sqrt{1-lpha_t \epsilon}$$
 Korea

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T \left| log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) 
ight| - logp_{ heta}(x_0|x_1) \ D_{KL}(q(x_T|x_0)) | p(x_T)) 
ight| D_{KL}(q(x_T|x_0)) | p(x_T) 
ight) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D_{KL}(q$$

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

Process

Image

Noise









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

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

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

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

$$x_t = \sqrt{rac{1}{lpha_t x_0}} + \sqrt{1-rac{1}{lpha_t \epsilon}}$$
 Korea Institute of Science and Technology

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left[lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}
ight] \ & = \left[log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T \left[log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1)
ight] \ D_{KL}(q(x_T|x_0)||p(x_T)) & D_{KL}(q(x_T|x_0)||p(x_T)) \end{aligned}$$

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

Process

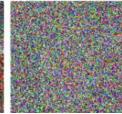
Image











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

$$ilde{\mu_t} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon)$$

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

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

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

$$x_t = \sqrt{rac{1}{lpha_t x_0}} + \sqrt{1-lpha_t \epsilon}$$

Korea Institute of Science and Technology

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T \left| log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) 
ight| - logp_{ heta}(x_0|x_1) \ D_{KL}(q(x_T|x_0)) | p(x_T) 
ight) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D$$

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

Process Process

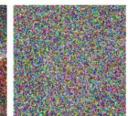
Image











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

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

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

$$q(x_{t-1}|x_t, x_0) = \begin{array}{c} \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I) \end{array}$$
(11)

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T \left| log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) 
ight| - logp_{ heta}(x_0|x_1) \ D_{KL}(q(x_T|x_0)) | p(x_T)) 
ight| D_{KL}(q(x_T|x_0)) | p(x_T) 
ight) = 0$$

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

Process

Image











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

$$\tilde{\beta}_{t} \coloneqq \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_{t}} \beta_{t}$$

$$\tilde{\mu}_{t}(x_{t}, x_{0}) \coloneqq \frac{\sqrt{\bar{\alpha}_{t-1}} \beta_{t}}{1 - \bar{\alpha}_{t}} x_{0} + \frac{\sqrt{\alpha_{t}} (1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_{t}} x_{t}$$

$$\vdots$$

$$q(x_{t-1} | x_{t}, x_{0}) = \sqrt[\mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_{t}}} (x_{t} - \frac{\beta_{t}}{\sqrt{1 - \bar{\alpha}_{t}}} \epsilon), \tilde{\beta}_{t} I)}$$

$$(12)$$

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T \left| log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) 
ight| - logp_{ heta}(x_0|x_1) \ D_{KL}(q(x_T|x_0)) | p(x_T)) 
ight| D_{KL}(q(x_T|x_0)) | p(x_T) 
ight) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D_{KL}(q$$

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

Reverse Process

Image

Noise









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

$$\tilde{\beta}_{t} := \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_{t}} \beta_{t}$$

$$\tilde{\mu}_{t}(x_{t}, x_{0}) := \frac{\sqrt{\bar{\alpha}_{t-1}} \beta_{t}}{1 - \bar{\alpha}_{t}} x_{0} + \frac{\sqrt{\alpha_{t}} (1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_{t}} x_{t}$$

$$q(x_{t-1} | x_{t}, x_{0}) = \sqrt[\mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_{t}}} (x_{t} - \frac{\beta_{t} - \bar{\alpha}_{t}}{\sqrt{1 - \alpha_{t}}}) \tilde{\beta}_{t} I)}$$

$$(11)$$

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left| lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight| \ & = \left| log(rac{q(x_T|x_0)}{p(x_T)}) 
ight| + \sum_{t=2}^T \left| log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) 
ight| - logp_{ heta}(x_0|x_1) \ D_{KL}(q(x_T|x_0)) | p(x_T)) 
ight| D_{KL}(q(x_T|x_0)) | p(x_T) 
ight) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D_{KL}(q(x_T|x_0)) | p(x_T|x_0) \ D_{KL}(q(x_T|x_0)) \ D_{K$$

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

Reverse Process

Image











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

$$\mathcal{N}(x_{t-1}; rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}} rac{ ilde{\epsilon_{ heta}(x_t)}, ilde{eta_t}I)}{\sqrt{1-lpha_t}}$$

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

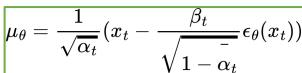
$$\tilde{\mu}_t(x_t, x_0) \coloneqq \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t \tag{11}$$

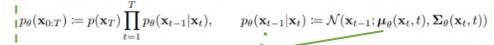
$$i q(x_{t-1}|x_t, x_0) = \frac{\mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}, \tilde{\beta_t}I)}{\sqrt{1 - \alpha_t}}$$
(11)



$$-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \ .$$

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$





**Process** 

Image

Noise









$$= log(rac{q(x_T|x_0)}{q(x_T|x_0)}) + \sum_{i=1}^{T}$$

$$= oxed{log(rac{q(x_T|x_0)}{p(x_T)})} +$$

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

$$-logp_{ heta}(x_0|x_1)$$

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

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

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

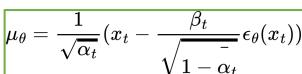
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t$$

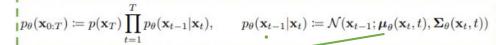
$$\mathbf{i} \ q(x_{t-1}|x_t, x_0) = \frac{\mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}, \tilde{\beta_t}I)}{\sqrt{1 - \alpha_t}}$$
(12)

$$\mathcal{N}(x_{t-1}; rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}} rac{oldsymbol{\hat{\epsilon}_{ heta}}(x_t)}{\sqrt{1-lpha_t}})$$

$$-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}$$

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$





**Process** 

Image











$$= \overline{log(rac{q(x_T|x_0)}{p(x_T)})} +$$

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

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

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

$$\epsilon - \epsilon_{ heta}(x_t)$$

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

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

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

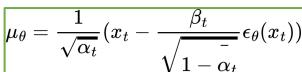
$$(10)$$

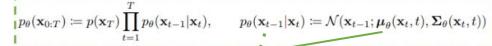
$$q(x_{t-1}|x_t, x_0) = \begin{array}{c} \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I) \end{array} (11)$$



$$-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}$$

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$





**Process** 

Image











$$= oxed{log(rac{q(x_T|x_0)}{p(x_T)})} + oxed{}$$

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

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

$$\epsilon - \epsilon_{ heta}(x_t)$$

$$=rac{eta_t^2}{2\sigma_t^2lpha_t(1-lpha_t)}||\epsilon-\epsilon_ heta(x_t)||^2$$

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

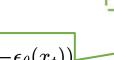
$$\tilde{\mu}_t(x_t, x_0) \coloneqq \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t$$

$$q(x_{t-1}|x_t, x_0) = \begin{array}{c} \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I) \end{array} (11)$$



$$-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}$$

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$



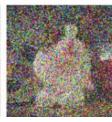
**Process** 

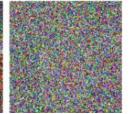
 $p_{\theta}(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t), \qquad p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t), \boldsymbol{\Sigma}_{\theta}(\mathbf{x}_t, t))$ 

Noise









$$p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)$$

$$= iggl[ log(rac{q(x_T|x_0)}{p(x_T)}) +$$

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

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

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

$$\epsilon - \epsilon_{ heta}(x_t)$$

$$rac{eta_t^2}{2\sigma_t^2lpha_t(1-lpha_t)}||\epsilon-\epsilon_ heta(x_t)||^2$$

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

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

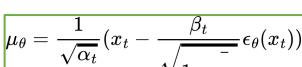
$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I)$$
(11)

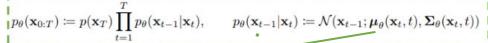
$$x_t = \sqrt{rac{1}{lpha_t x_0}} + \sqrt{1-lpha_t \epsilon}$$



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

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$





**Process** 

#### Noise









$$-log(p_{ heta}(x_0)) \leq egin{aligned} & rac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \end{aligned}$$

$$= iggl[ log(rac{q(x_T|x_0)}{p(x_T)}) iggr] +$$

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

$$-logp_{ heta}(x_0|x_1)$$

# $\left|D_{KL}(q(x_T|x_0)||p(x_T)) ight|$

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

$$\epsilon - \epsilon_{ heta}(x_t)$$

$$=rac{eta_t^2}{2\sigma_t^2lpha_t(1-lpha_t)}||\epsilon-\epsilon_ heta(x_t)||^2$$

$$=rac{eta_t^2}{2\sigma_t^2lpha_t(1-lpha_t)}||\epsilon-\epsilon_ heta(\sqrt{rac{-}{lpha_t}}x_0+\sqrt{1-lpha_t}\epsilon,t)||^2$$

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

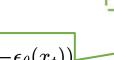
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t$$

$$q(x_{t-1}|x_t, x_0) = \begin{array}{c} \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I) \end{array} (11)$$



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

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$



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

**Process** 

#### Noise









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

$$= \overline{log(rac{q(x_T|x_0)}{p(x_T)})} +$$

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

$$- log p_{ heta}(x_0|x_1)$$

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

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

$$\epsilon - \epsilon_{ heta}(x_t)$$

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

$$=rac{eta_t^2}{2\sigma_t^2lpha_t(1-lpha_t)}||\epsilon-\epsilon_ heta(\sqrt{rac{-}{lpha_t}}x_0+\sqrt{1-lpha_t}\epsilon,t)||^2$$

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

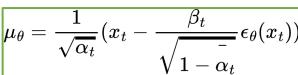
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t$$

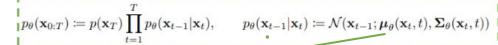
$$q(x_{t-1}|x_t, x_0) = \begin{array}{c} \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I) \end{array} (11)$$



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

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$





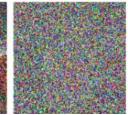
**Process** 











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

$$= egin{aligned} log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=1}^T & \sum_{$$

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

$$D_{KL}(q(x_Tert x_0)ertert p(x_T))$$

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

$$\epsilon - \epsilon_{ heta}(x_t)$$

$$= ||\epsilon - \epsilon_{\theta}(\sqrt{\frac{1}{\alpha_{t}}x_{0}} + \sqrt{1 - \alpha_{t}\epsilon_{t}}, t)||^{2}$$

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

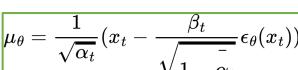
$$\tilde{\mu}_t(x_t, x_0) := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t$$

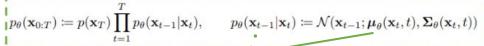
$$q(x_{t-1}|x_t, x_0) = \begin{array}{c} \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I) \end{array} (11)$$



$$-log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} - log(p_{ heta}(x_0)) \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)}$$

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$





**Process** 

Noise









$$= log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{i=1}^{n} log(rac{q(x_T|x_0)}{p(x_T)})$$

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

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

# $\| q(x_{t-1}|x_t,x_0) \| \| p_{ heta}(x_{t-1}|x_t) \|$

$$\epsilon - \epsilon_{ heta}(x_t)$$

$$= ||\epsilon - \epsilon_{ heta}(\sqrt{lpha_{t}}x_{0} + \sqrt{1 - lpha_{t}}\epsilon, t)||^{2} \ = ||\epsilon - \epsilon_{ heta}(x_{t})||^{2}$$

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

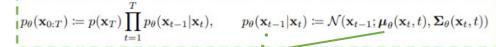
$$\tilde{\mu}_t(x_t, x_0) \coloneqq \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t$$

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I)$$
(11)



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

$$\mu_{ heta} = rac{1}{\sqrt{lpha_t}}(x_t - rac{eta_t}{\sqrt{1-lpha_t}}\epsilon_{ heta}(x_t))$$

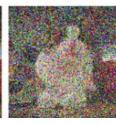


**Process** 

#### Noise









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

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

$$\left|-logp_{ heta}(x_0|x_1)
ight|$$

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

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

$$\epsilon - \epsilon_{ heta}(x_t)$$

$$egin{array}{ll} &= ||\epsilon - \epsilon_{ heta}(\sqrt{lpha_t}x_0 + \sqrt{1 - lpha_t}\epsilon, t)||^2 \ &= ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{array}$$

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

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

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}}\epsilon), \tilde{\beta_t}I)$$
(11)



$$egin{align*} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{align*} & \prod_{t=1}^T q(x_t|x_{t-1}) \ & p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t) \ & = egin{align*} & log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T egin{align*} & log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \ & D_{KL}(q(x_T|x_0)||p(x_T)) \end{bmatrix} & D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ & = ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{aligned}$$



$$egin{align*} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{align*} & \prod_{t=1}^T q(x_t|x_{t-1}) \ & p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t) \ & = egin{align*} & log(rac{q(x_{T}|x_0)}{p(x_T)}) + \sum_{t=2}^T egin{align*} & log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - egin{align*} & logp_{ heta}(x_0|x_1) \ & p_{ heta}(x_t,x_0) \mid p(x_T) \ & p_{ heta}(x_{t-1}|x_t,x_0) \mid p_{ heta}(x_{t-1}|x_t) \ & = |\epsilon - \epsilon_{ heta}(x_t)||^2 \ & = |\epsilon - \epsilon_{ heta}(x_t)||$$



$$egin{align*} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{align*} & rac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \ & = egin{align*} & log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \ & D_{KL}(q(x_T|x_0)||p(x_T)) \end{pmatrix} & D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ & = ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{aligned}$$

$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

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

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{aligned} rac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \ & = egin{aligned} log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \end{aligned}$$

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

$$egin{aligned} D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ &= ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{aligned}$$



$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

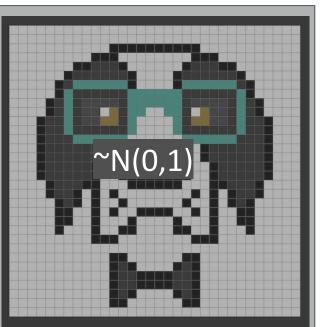
$$\delta_{+}(x) = \begin{cases} \infty & \text{if } x = 1\\ x + \frac{1}{255} & \text{if } x < 1 \end{cases} \quad \delta_{-}(x) = \begin{cases} -\infty & \text{if } x = -1\\ x - \frac{1}{255} & \text{if } x > -1 \end{cases}$$

100

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{aligned} rac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \ & = egin{aligned} log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \end{aligned}$$

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

$$egin{aligned} D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ &= ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{aligned}$$



$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

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

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \end{aligned}$$

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

$$= egin{aligned} log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \end{aligned}$$

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

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

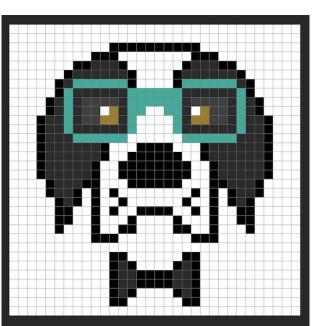
$$= ||\epsilon - \epsilon_{\theta}(x_t)||^2$$

Number of pixels

$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

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

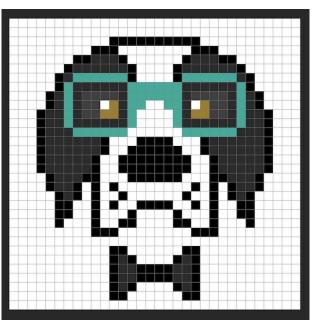
$$\delta_{+}(x) = \begin{cases} \infty & \text{if } x = 1\\ x + \frac{1}{255} & \text{if } x < 1 \end{cases} \qquad \delta_{-}(x) = \begin{cases} -\infty & \text{if } x = -1\\ x - \frac{1}{255} & \text{if } x > -1 \end{cases}$$



$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{aligned} rac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \ & = egin{aligned} log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \end{aligned}$$

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

$$egin{aligned} D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ &= ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{aligned}$$



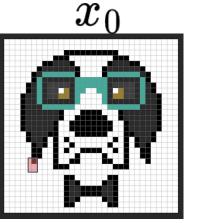
Number of pixels
Border range

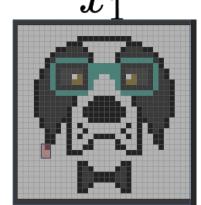
$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

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

114

$$egin{align*} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left[ lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight] \ & = \left[ log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T \left[ log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) 
ight] \ D_{KL}(q(x_T|x_0)||p(x_T)) & D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ & = ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{aligned}$$





Number of pixels
Border range

$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

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



$$egin{align*} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{align*} & \prod_{t=1}^T q(x_t|x_{t-1}) \ & p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t) \ & = egin{align*} & log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T egin{align*} & log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \ & D_{KL}(q(x_T|x_0)||p(x_T)) \ & D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ & = ||\epsilon - \epsilon_{ heta}(x_t)||^2 \ & \end{pmatrix}$$

0.45

Number of pixels
Border range

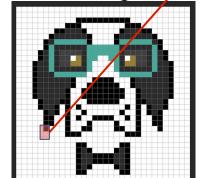
$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

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

$$egin{align*} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq \left[ lograc{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} 
ight] \ & = \left[ log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T \left[ log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) 
ight] \ & D_{KL}(q(x_T|x_0)||p(x_T)) & D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ & = ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{aligned}$$

 $x_0$ 

 $x_0^{
m predicted}$ 



Number of pixels Border range

$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

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



$$egin{align*} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{bmatrix} Iog & \prod_{t=1}^T q(x_t|x_{t-1}) \ p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t) \end{bmatrix} \ & = egin{bmatrix} log(rac{q(x_{T}|x_0)}{p(x_T)}) + \sum_{t=2}^T egin{bmatrix} log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \end{bmatrix} \ & D_{KL}(q(x_T|x_0)) \|p(x_T)) \end{bmatrix} \ & = \|\epsilon - \epsilon_{ heta}(x_t)\|^2 \end{aligned}$$

 $x_0$ 

0.65

 $x_0^{
m predicted}$ 

0.55

Number of pixels

$$p_{\theta}(\mathbf{x}_{0}|\mathbf{x}_{1}) = \prod_{i=1}^{D} \int_{\delta_{-}(x_{0}^{i})}^{\delta_{+}(x_{0}^{i})} \mathcal{N}(x; \mu_{\theta}^{i}(\mathbf{x}_{1}, 1), \sigma_{1}^{2}) dx$$

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

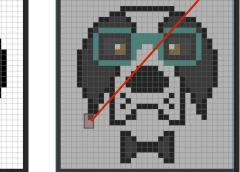
$$\delta_{+}(x) = \begin{cases} \infty & \text{if } x = 1\\ x + \frac{1}{255} & \text{if } x < 1 \end{cases} \qquad \delta_{-}(x) = \begin{cases} -\infty & \text{if } x = -1\\ x - \frac{1}{255} & \text{if } x > -1 \end{cases}$$

$$egin{aligned} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq rac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{ heta}(x_{t-1}|x_t)} \end{aligned}$$

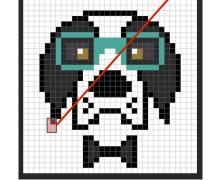
$$D_{KL}(q(x_T|x_0)||p(x_T))$$

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

$$x_0$$
  $x_1$ 



# predicted



### **Algorithm 2** Sampling

1: 
$$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

2: **for** 
$$t = T, ..., 1$$
 **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  $x_0$ 

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

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

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

0.55

Number of pixels

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

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

$$egin{align*} -log(p_{ heta}(x)) & \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ -log(p_{ heta}(x_0)) & \leq egin{align*} & \prod_{t=1}^T q(x_t|x_{t-1}) \ & p(x_T) \prod_{t=1}^T p_{ heta}(x_{t-1}|x_t) \ & = egin{align*} & log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^T egin{align*} & log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - egin{align*} & logp_{ heta}(x_0|x_1) \ & p_{ heta}(x_t) \|p(x_T) \end{pmatrix} \ & D_{KL}(q(x_T|x_0)) \|p(x_T)) \end{bmatrix} \ & = \|\epsilon - \epsilon_{ heta}(x_t)\|^2 \end{aligned}$$





$$egin{aligned} & egin{aligned} egin{aligned} & egin{$$





$$egin{aligned} & -log(p_{ heta}(x)) \leq lograc{q(x_{1:T}|x_0)}{p_{ heta}(x_{0:T})} \ & -log(p_{ heta}(x_0)) \leq egin{aligned} & \int_{t=1}^{T} q(x_t|x_{t-1}) \ & p(x_T) \prod_{t=1}^{T} p_{ heta}(x_{t-1}|x_t) \ & = egin{aligned} & log(rac{q(x_T|x_0)}{p(x_T)}) + \sum_{t=2}^{T} log(rac{q(x_{t-1}|x_t,x_0)}{p_{ heta}(x_{t-1}|x_t)}) - logp_{ heta}(x_0|x_1) \ & D_{KL}(q(x_T|x_0)|p(x_T)) & D_{KL}(q(x_{t-1}|x_t,x_0)||p_{ heta}(x_{t-1}|x_t)) \ & = ||\epsilon - \epsilon_{ heta}(x_t)||^2 \end{aligned}$$





$$\begin{array}{l} \text{nimize!} \\ -log(p_{\theta}(x)) \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ \\ -log(p_{\theta}(x_{0})) \leq \overline{bog \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ \\ = \overline{bog(\frac{q(x_{T}|x_{0})}{p(x_{T})})} + \sum_{t=2}^{T} \overline{bog(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})})} - \overline{bogp_{\theta}(x_{0}|x_{0})} \\ D_{KL}(q(x_{T}|x_{0})|p(x_{T})) \\ = |\epsilon - \epsilon_{\theta}(x_{t})||^{2} \\ \end{array}$$





$$\begin{array}{l} \text{nimize!} \\ -log(p_{\theta}(x)) \leq log \frac{q(x_{1:T}|x_{0})}{p_{\theta}(x_{0:T})} \\ \\ -log(p_{\theta}(x_{0})) \leq \boxed{log \frac{\prod_{t=1}^{T} q(x_{t}|x_{t-1})}{p(x_{T}) \prod_{t=1}^{T} p_{\theta}(x_{t-1}|x_{t})}} \\ \\ = \boxed{log(\frac{q(x_{T}|x_{0})}{p(x_{T})}) + \sum_{t=2}^{T} log(\frac{q(x_{t-1}|x_{t},x_{0})}{p_{\theta}(x_{t-1}|x_{t})}) - logp_{\theta}(x_{t-1}|x_{t})} \\ D_{KL}(q(x_{T}|x_{0}), p(x_{T})) \boxed{D_{KL}(q(x_{t-1}|x_{t},x_{0})) ||p_{\theta}(x_{t-1}|x_{t}))} \\ = ||\epsilon - \epsilon_{\theta}(x_{t})||^{2} \\ \end{array}$$



$$\begin{aligned} & \text{nimize!} \\ & -log(p_{\theta}(x)) \leq log\frac{q(x_{1:T}|x_0)}{p_{\theta}(x_{0:T})} \\ & -log(p_{\theta}(x_0)) \leq \boxed{log\frac{\prod_{t=1}^T q(x_t|x_{t-1})}{p(x_T)\prod_{t=1}^T p_{\theta}(x_{t-1}|x_t)}} \end{aligned}$$

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



### **Algorithms**



### **Algorithm 1** Training

### 1: repeat

- 2:  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
- 3:  $t \sim \text{Uniform}(\{1,\ldots,T\})$
- 4:  $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 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

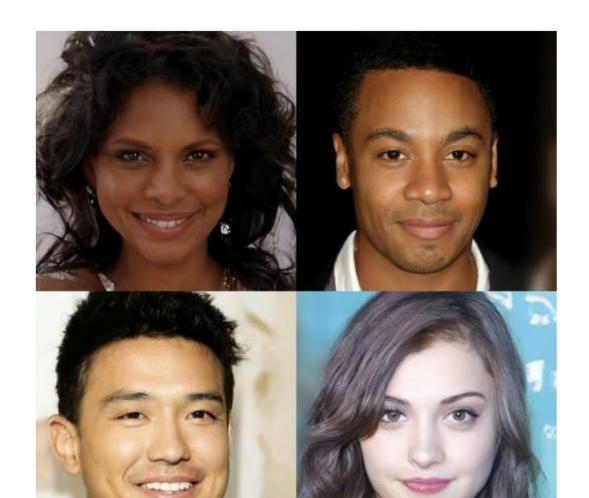
- 1:  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 2: **for** t = T, ..., 1 **do**
- 3:  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  if t > 1, else  $\mathbf{z} = \mathbf{0}$

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  $x_0$

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













### **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)





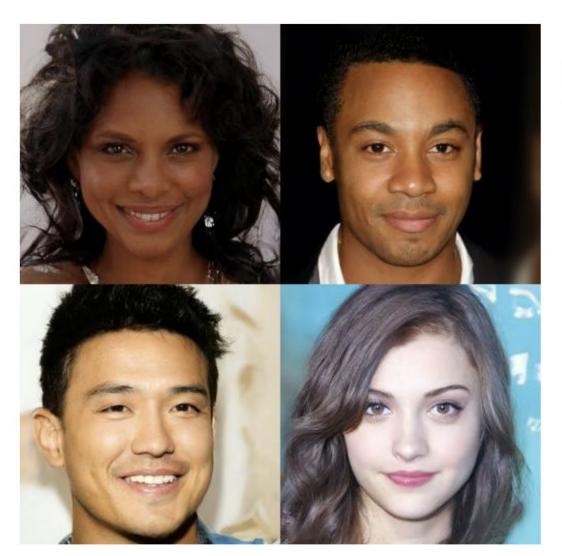


### **Paper List**

- 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)







### **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)

