



北京航空航天大學  
BEIHANG UNIVERSITY

# 生成式AI与大模型第8讲

## 扩散模型和视觉大模型

Beihang University

人工智能研究院  
黄雷

01

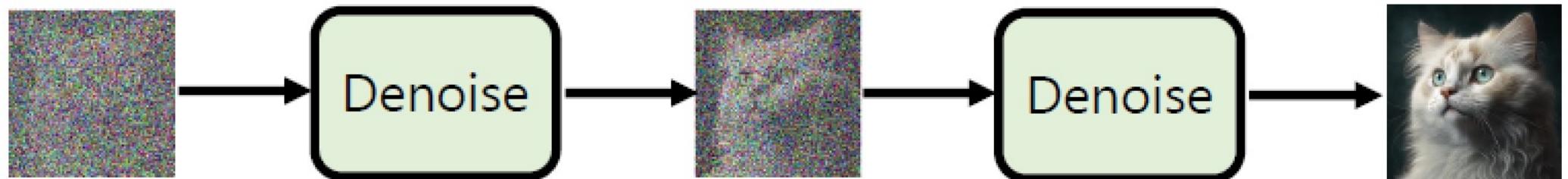
## 扩散模型

# Diffusion Model

## Forward Process

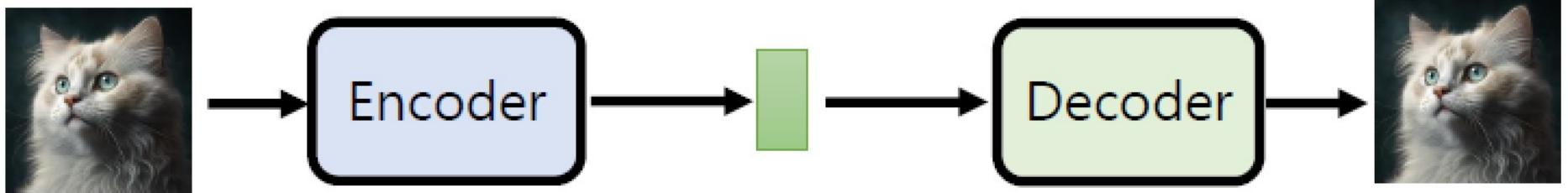


## Reverse Process

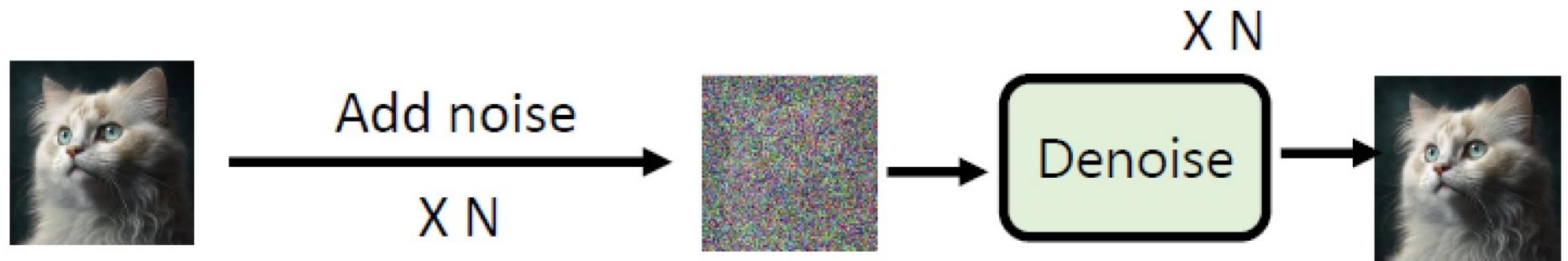


# Diffusion Model

VAE



Diffusion



# Diffusion Model

---

## Algorithm 1 Training

---

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

6: until converged
```

---

---

## Algorithm 2 Sampling

---

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

---

# Training...



$x_0$ : clean image



$\varepsilon$ : noise

---

## Algorithm 1 Training

---

- 1: **repeat**
- 2:  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$   $\leftarrow \dots$  sample clean image
- 3:  $t \sim \text{Uniform}(\{1, \dots, T\})$
- 4:  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$   $\leftarrow \dots$  sample a noise
- 5: Take gradient descent step on

$$\nabla_{\theta} \|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\theta}(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, t)\|^2$$

- 6: **until** converged

Target  
Noise

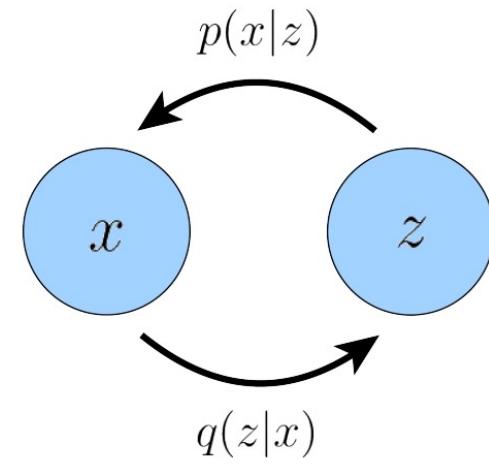
Noise  
predictor

$\bar{\alpha}_1, \bar{\alpha}_2, \dots, \bar{\alpha}_T$   
smaller

$$\begin{aligned}
\log p(\mathbf{x}) &= \log p(\mathbf{x}) \int q_{\phi}(\mathbf{z}|\mathbf{x}) dz \\
&= \int q_{\phi}(\mathbf{z}|\mathbf{x})(\log p(\mathbf{x})) dz \\
&= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p(\mathbf{x})] \\
&= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{z}|\mathbf{x})} \right] \\
&= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{p(\mathbf{x}, \mathbf{z})q_{\phi}(\mathbf{z}|\mathbf{x})}{p(\mathbf{z}|\mathbf{x})q_{\phi}(\mathbf{z}|\mathbf{x})} \right] \\
&= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{p(\mathbf{x}, \mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})} \right] + \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{q_{\phi}(\mathbf{z}|\mathbf{x})}{p(\mathbf{z}|\mathbf{x})} \right] \\
&= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{p(\mathbf{x}, \mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})} \right] + D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z}|\mathbf{x})) \\
&\geq \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{p(\mathbf{x}, \mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})} \right]
\end{aligned}$$

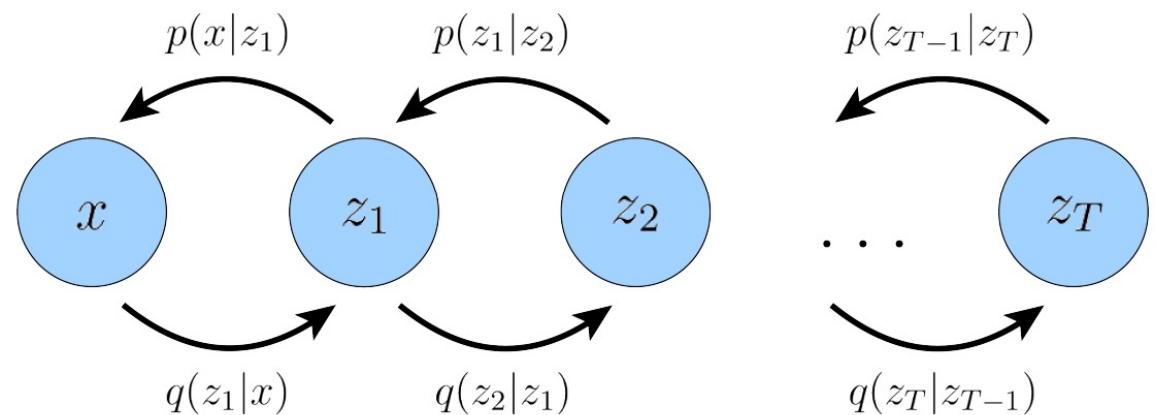
Evidence Lower  
Bound (ELBO)

$$\begin{aligned}
\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{p(\mathbf{x}, \mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})} \right] &= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{p_{\theta}(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})} \right] \\
&= \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{z})] + \mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} \left[ \log \frac{p(\mathbf{z})}{q_{\phi}(\mathbf{z}|\mathbf{x})} \right] \\
&= \underbrace{\mathbb{E}_{q_{\phi}(\mathbf{z}|\mathbf{x})} [\log p_{\theta}(\mathbf{x}|\mathbf{z})]}_{\text{reconstruction term}} - \underbrace{D_{\text{KL}}(q_{\phi}(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z}))}_{\text{prior matching term}}
\end{aligned}$$



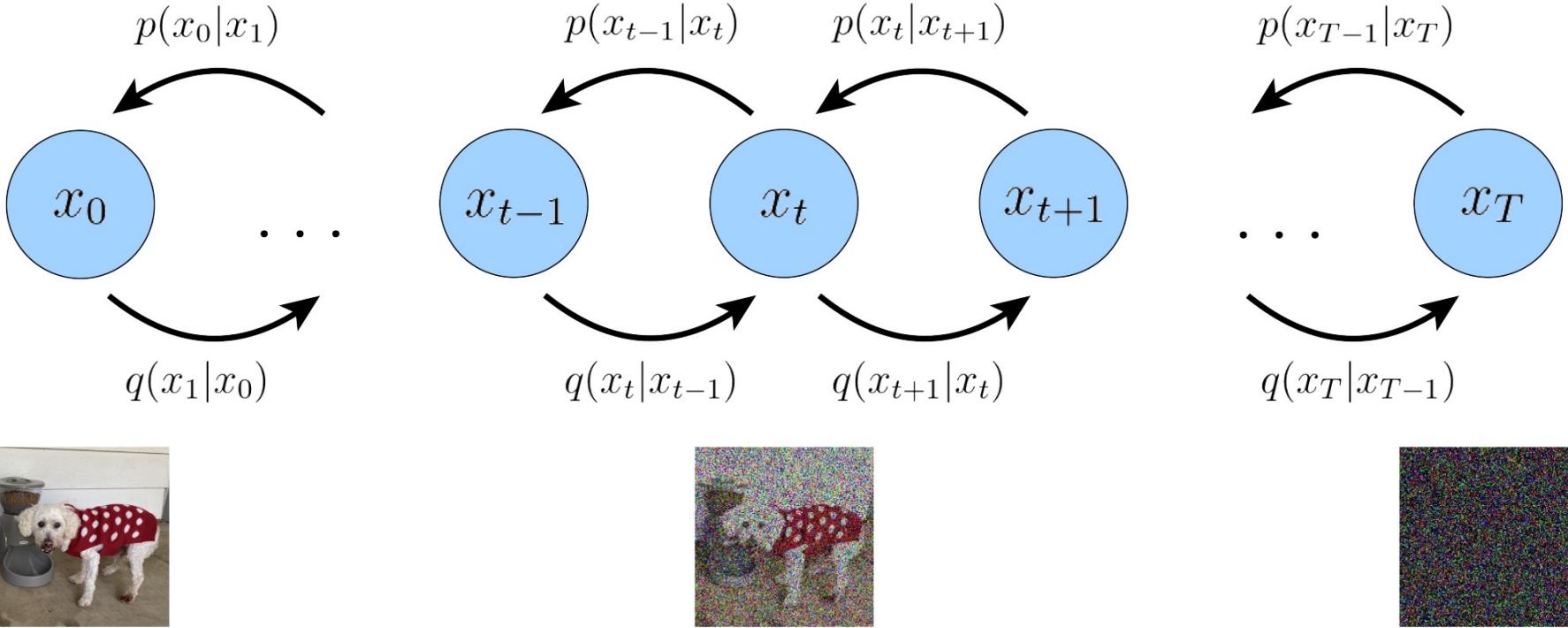
# Hierarchical Variational Autoencoders

$$\begin{aligned}
 \log p(\mathbf{x}) &= \log \int p(\mathbf{x}, \mathbf{z}_{1:T}) d\mathbf{z}_{1:T} \\
 &= \log \int \frac{p(\mathbf{x}, \mathbf{z}_{1:T}) q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})}{q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})} d\mathbf{z}_{1:T} \\
 &= \log \mathbb{E}_{q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})} \left[ \frac{p(\mathbf{x}, \mathbf{z}_{1:T})}{q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})} \right] \\
 &\geq \mathbb{E}_{q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})} \left[ \log \frac{p(\mathbf{x}, \mathbf{z}_{1:T})}{q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})} \right]
 \end{aligned}$$



$$\mathbb{E}_{q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})} \left[ \log \frac{p(\mathbf{x}, \mathbf{z}_{1:T})}{q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})} \right] = \mathbb{E}_{q_{\phi}(\mathbf{z}_{1:T} | \mathbf{x})} \left[ \log \frac{p(\mathbf{z}_T) p_{\theta}(\mathbf{x} | \mathbf{z}_1) \prod_{t=2}^T p_{\theta}(\mathbf{z}_{t-1} | \mathbf{z}_t)}{q_{\phi}(\mathbf{z}_1 | \mathbf{x}) \prod_{t=2}^T q_{\phi}(\mathbf{z}_t | \mathbf{z}_{t-1})} \right]$$

# Variational Diffusion Models



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

where,

$$p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; \mathbf{0}, \mathbf{I})$$

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

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

# Variational Diffusion Models

$$\begin{aligned}\log p(\mathbf{x}) &\geq \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \right] \\&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T) \prod_{t=1}^T p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{\prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})} \right] \\&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T) p_{\boldsymbol{\theta}}(\mathbf{x}_0|\mathbf{x}_1) \prod_{t=2}^T p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_1|\mathbf{x}_0) \prod_{t=2}^T q(\mathbf{x}_t|\mathbf{x}_{t-1})} \right] \\&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T) p_{\boldsymbol{\theta}}(\mathbf{x}_0|\mathbf{x}_1) \prod_{t=2}^T p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_1|\mathbf{x}_0) \prod_{t=2}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)} \right] \\&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p_{\boldsymbol{\theta}}(\mathbf{x}_T) p_{\boldsymbol{\theta}}(\mathbf{x}_0|\mathbf{x}_1)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \prod_{t=2}^T \frac{p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)} \right] \\&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T) p_{\boldsymbol{\theta}}(\mathbf{x}_0|\mathbf{x}_1)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \prod_{t=2}^T \frac{p_{\boldsymbol{\theta}}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}} \right]\end{aligned}$$

# Variational Diffusion Models

$$\begin{aligned}
&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T)p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \prod_{t=2}^T \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{\frac{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)q(\mathbf{x}_t|\mathbf{x}_0)}{q(\mathbf{x}_{t-1}|\mathbf{x}_0)}} \right] \\
&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T)p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)}{q(\mathbf{x}_1|\mathbf{x}_0)} + \log \frac{q(\mathbf{x}_1|\mathbf{x}_0)}{q(\mathbf{x}_T|\mathbf{x}_0)} + \log \prod_{t=2}^T \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} \right] \\
&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T)p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)}{q(\mathbf{x}_T|\mathbf{x}_0)} + \sum_{t=2}^T \log \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} \right] \\
&= \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} [\log p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)] + \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T)}{q(\mathbf{x}_T|\mathbf{x}_0)} \right] + \sum_{t=2}^T \mathbb{E}_{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \left[ \log \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} \right] \\
&= \mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)} [\log p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)] + \mathbb{E}_{q(\mathbf{x}_T|\mathbf{x}_0)} \left[ \log \frac{p(\mathbf{x}_T)}{q(\mathbf{x}_T|\mathbf{x}_0)} \right] + \sum_{t=2}^T \mathbb{E}_{q(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{x}_0)} \left[ \log \frac{p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)} \right] \\
&= \underbrace{\mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)} [\log p_{\theta}(\mathbf{x}_0|\mathbf{x}_1)]}_{\text{reconstruction term}} - \underbrace{D_{\text{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p(\mathbf{x}_T))}_{\text{prior matching term}} - \sum_{t=2}^T \underbrace{\mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)} [D_{\text{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) \parallel p_{\theta}(\mathbf{x}_{t-1}|\mathbf{x}_t))]}_{\text{denoising matching term}}
\end{aligned}$$

# Optimization object by conduction

$$\mathbb{E}_{q(x_1|x_0)}[\log P(x_0|x_1)] - KL(q(x_T|x_0)||P(x_T))$$

$$- \sum_{t=2}^T \mathbb{E}_{q(x_t|x_0)}[KL(q(x_{t-1}|x_t, x_0)||P(x_{t-1}|x_t))]$$



$x_0$

.....

$$q(x_t|x_0)$$

$$q(x_{t-1}|x_0)$$



$x_{t-1}$



$x_t$

已知  
Gaussian Gaussian

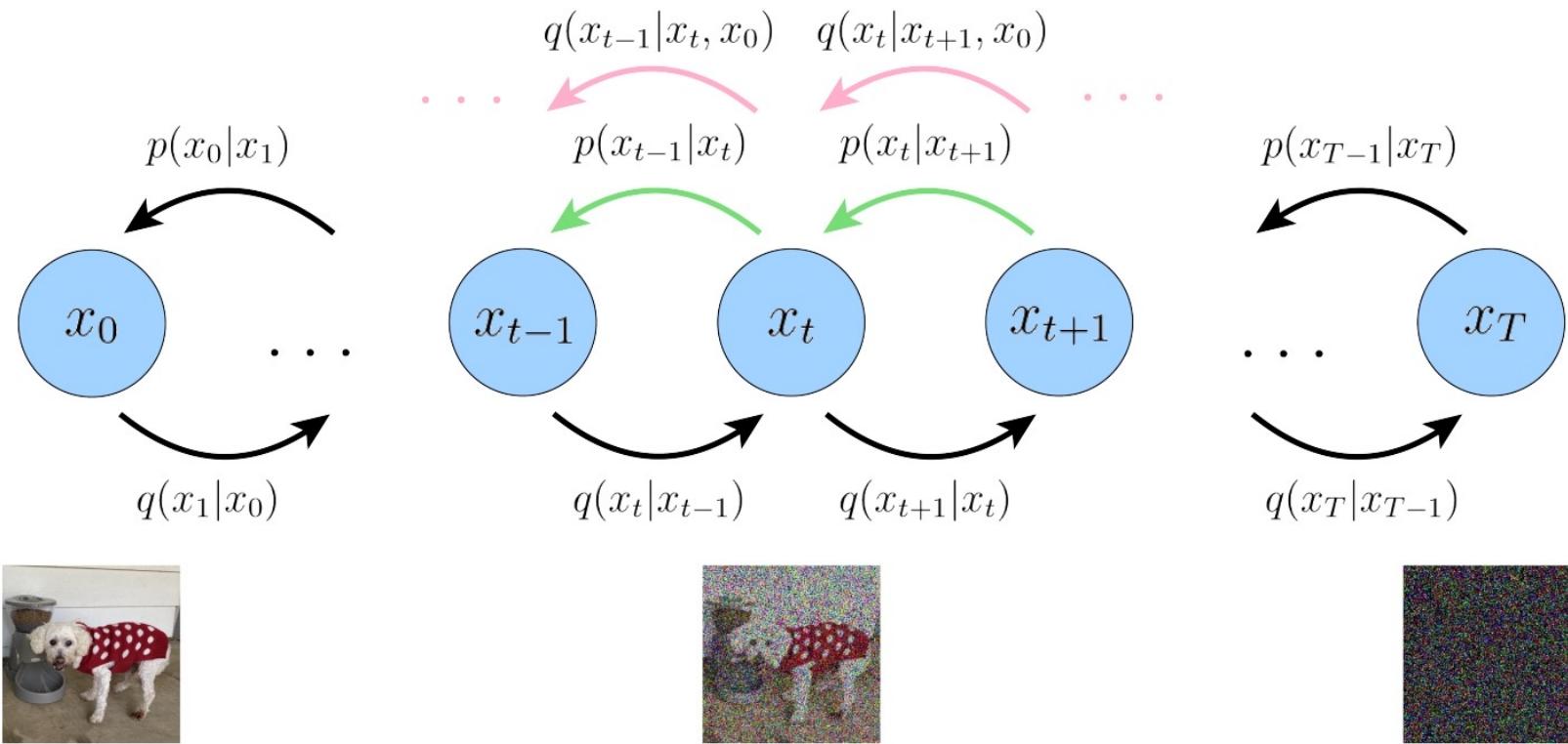
$$q(x_{t-1}|x_t, x_0)$$

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

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

已知  
Gaussian

# Optimization object by conduction



$$\arg \min_{\theta} \mathbb{E}_{t \sim U\{2, T\}} \left[ \mathbb{E}_{q(\mathbf{x}_t | \mathbf{x}_0)} [D_{\text{KL}}(q(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathbf{x}_0) \parallel p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t))] \right]$$

# Obtain the target by calculation

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

$$\epsilon \sim \mathcal{N}(\mathbf{0}, I)$$

 $x_{t-1}$ 

$$= \sqrt{1 - \beta_t}$$

 $x_t$ 

$$+ \sqrt{\beta_t}$$

 $\beta_1, \beta_2, \dots, \beta_T$ 

$$q(x_t|x_0)$$

 $x_0$ 

+



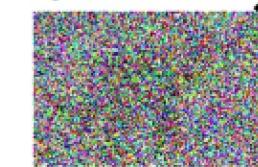
+



.....



+

 $x_t$

# Obtain the target by calculation

 $x_1$ 

$$= \boxed{\sqrt{1 - \beta_1}} \quad \text{[Original Image]} + \sqrt{\beta_1} \quad \text{[Noise Image]}$$

 $x_0$  $x_2$ 

$$= \sqrt{1 - \beta_2} \quad \text{[Original Image]} + \sqrt{\beta_2} \quad \text{[Noise Image]}$$

 $x_1$  $\sim \mathcal{N}(0, I)$ 

Ind.

 $x_2$ 

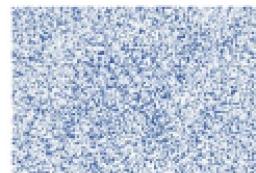
$$= \sqrt{1 - \beta_2} \quad \sqrt{1 - \beta_1} \quad \text{[Original Image]} + \sqrt{1 - \beta_2} \quad \sqrt{\beta_1} \quad \text{[Noise Image]}$$

 $x_0$ 

$$+ \sqrt{1 - \beta_2} \quad \sqrt{\beta_1}$$



$$+ \sqrt{\beta_2}$$



# Obtain the target by calculation

 $x_2$ 

$$= \sqrt{1 - \beta_2} \sqrt{1 - \beta_1}$$

 $x_0$ 

$$\sim \mathcal{N}(\mathbf{0}, I)$$

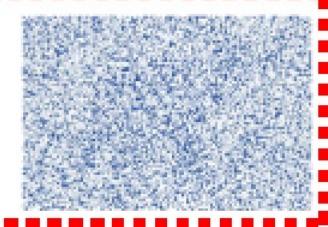


$$\sim \mathcal{N}(\mathbf{0}, I)$$

$$+ \sqrt{1 - \beta_2} \sqrt{\beta_1}$$

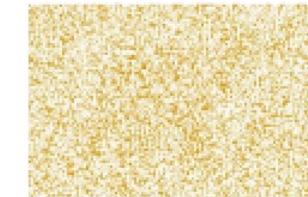


$$+ \sqrt{\beta_2}$$



$$\sim \mathcal{N}(\mathbf{0}, I)$$

$$+ \sqrt{1 - (1 - \beta_2)(1 - \beta_1)}$$



# Obtain the target by calculation

$$q(x_t|x_0)$$

$$\beta_1, \beta_2, \dots, \beta_T$$



$$\sim \mathcal{N}(\mathbf{0}, I)$$

$$\alpha_t = 1 - \beta_t$$

$$\bar{\alpha}_t = \alpha_1 \alpha_2 \dots \alpha_t$$



$$= \sqrt{1 - \beta_1}$$



$$+ \sqrt{\beta_1}$$



$$= \sqrt{1 - \beta_2}$$



$$+ \sqrt{\beta_2}$$



⋮

⋮

⋮

$$= \sqrt{1 - \beta_t}$$



$$+ \sqrt{\beta_t}$$



|||



$$= \boxed{\sqrt{1 - \beta_1} \dots \sqrt{1 - \beta_t}}$$

$$\sqrt{\bar{\alpha}_t}$$



+

$$\boxed{\sqrt{1 - (1 - \beta_1) \dots (1 - \beta_t)}}$$

$$\sqrt{1 - \bar{\alpha}_t}$$



# Math.....

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_0)q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)} \quad (71)$$

$$= \frac{\mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I})\mathcal{N}(\mathbf{x}_{t-1}; \sqrt{\alpha_{t-1}}\mathbf{x}_0, (1-\bar{\alpha}_{t-1})\mathbf{I})}{\mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})} \quad (72)$$

$$\propto \exp \left\{ - \left[ \frac{(\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_{t-1})^2}{2(1-\alpha_t)} + \frac{(\mathbf{x}_{t-1} - \sqrt{\alpha_{t-1}}\mathbf{x}_0)^2}{2(1-\bar{\alpha}_{t-1})} - \frac{(\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_0)^2}{2(1-\bar{\alpha}_t)} \right] \right\} \quad (73)$$

$$= \exp \left\{ - \frac{1}{2} \left[ \frac{(\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_{t-1})^2}{1-\alpha_t} + \frac{(\mathbf{x}_{t-1} - \sqrt{\alpha_{t-1}}\mathbf{x}_0)^2}{1-\bar{\alpha}_{t-1}} - \frac{(\mathbf{x}_t - \sqrt{\alpha_t}\mathbf{x}_0)^2}{1-\bar{\alpha}_t} \right] \right\} \quad (74)$$

$$= \exp \left\{ - \frac{1}{2} \left[ \frac{(-2\sqrt{\alpha_t}\mathbf{x}_t\mathbf{x}_{t-1} + \alpha_t\mathbf{x}_{t-1}^2)}{1-\alpha_t} + \frac{(\mathbf{x}_{t-1}^2 - 2\sqrt{\alpha_{t-1}}\mathbf{x}_{t-1}\mathbf{x}_0)}{1-\bar{\alpha}_{t-1}} + C(\mathbf{x}_t, \mathbf{x}_0) \right] \right\} \quad (75)$$

$$\propto \exp \left\{ - \frac{1}{2} \left[ - \frac{2\sqrt{\alpha_t}\mathbf{x}_t\mathbf{x}_{t-1}}{1-\alpha_t} + \frac{\alpha_t\mathbf{x}_{t-1}^2}{1-\alpha_t} + \frac{\mathbf{x}_{t-1}^2}{1-\bar{\alpha}_{t-1}} - \frac{2\sqrt{\alpha_{t-1}}\mathbf{x}_{t-1}\mathbf{x}_0}{1-\bar{\alpha}_{t-1}} \right] \right\} \quad (76)$$

$$= \exp \left\{ - \frac{1}{2} \left[ \left( \frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\bar{\alpha}_{t-1}} \right) \mathbf{x}_{t-1}^2 - 2 \left( \frac{\sqrt{\alpha_t}\mathbf{x}_t}{1-\alpha_t} + \frac{\sqrt{\alpha_{t-1}}\mathbf{x}_0}{1-\bar{\alpha}_{t-1}} \right) \mathbf{x}_{t-1} \right] \right\} \quad (77)$$

$$= \exp \left\{ - \frac{1}{2} \left[ \frac{\alpha_t(1-\bar{\alpha}_{t-1}) + 1-\alpha_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})} \mathbf{x}_{t-1}^2 - 2 \left( \frac{\sqrt{\alpha_t}\mathbf{x}_t}{1-\alpha_t} + \frac{\sqrt{\alpha_{t-1}}\mathbf{x}_0}{1-\bar{\alpha}_{t-1}} \right) \mathbf{x}_{t-1} \right] \right\} \quad (78)$$

$$= \exp \left\{ - \frac{1}{2} \left[ \frac{\alpha_t - \bar{\alpha}_t + 1 - \alpha_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})} \mathbf{x}_{t-1}^2 - 2 \left( \frac{\sqrt{\alpha_t}\mathbf{x}_t}{1-\alpha_t} + \frac{\sqrt{\alpha_{t-1}}\mathbf{x}_0}{1-\bar{\alpha}_{t-1}} \right) \mathbf{x}_{t-1} \right] \right\} \quad (79)$$

$$= \exp \left\{ - \frac{1}{2} \left[ \frac{1 - \bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})} \mathbf{x}_{t-1}^2 - 2 \left( \frac{\sqrt{\alpha_t}\mathbf{x}_t}{1-\alpha_t} + \frac{\sqrt{\alpha_{t-1}}\mathbf{x}_0}{1-\bar{\alpha}_{t-1}} \right) \mathbf{x}_{t-1} \right] \right\} \quad (80)$$

$$= \exp \left\{ - \frac{1}{2} \left( \frac{1 - \bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})} \right) \left[ \mathbf{x}_{t-1}^2 - 2 \frac{\left( \frac{\sqrt{\alpha_t}\mathbf{x}_t}{1-\alpha_t} + \frac{\sqrt{\alpha_{t-1}}\mathbf{x}_0}{1-\bar{\alpha}_{t-1}} \right)}{\frac{1 - \bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}} \mathbf{x}_{t-1} \right] \right\} \quad (81)$$

$$= \exp \left\{ - \frac{1}{2} \left( \frac{1 - \bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})} \right) \left[ \mathbf{x}_{t-1}^2 - 2 \frac{\left( \frac{\sqrt{\alpha_t}\mathbf{x}_t}{1-\alpha_t} + \frac{\sqrt{\alpha_{t-1}}\mathbf{x}_0}{1-\bar{\alpha}_{t-1}} \right) (1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{x}_{t-1} \right] \right\} \quad (82)$$

$$= \exp \left\{ - \frac{1}{2} \left( \frac{1}{\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t}} \right) \left[ \mathbf{x}_{t-1}^2 - 2 \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\mathbf{x}_t + \sqrt{\alpha_{t-1}}(1-\alpha_t)\mathbf{x}_0}{1 - \bar{\alpha}_t} \mathbf{x}_{t-1} \right] \right\} \quad (83)$$

$$\propto \mathcal{N}(\mathbf{x}_{t-1}; \underbrace{\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\mathbf{x}_t + \sqrt{\alpha_{t-1}}(1-\alpha_t)\mathbf{x}_0}{1 - \bar{\alpha}_t}}_{\mu_q(\mathbf{x}_t, \mathbf{x}_0)}, \underbrace{\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})\mathbf{I}}{1 - \bar{\alpha}_t}}_{\Sigma_q(t)}) \quad (84) \quad \text{https://arxiv.org/pdf/2208.11970.pdf}$$

# KL Distance

$$\mathbb{E}_{q(x_1|x_0)}[\log P(x_0|x_1)] - KL(q(x_T|x_0)||P(x_T))$$

$$- \sum_{t=2}^T \mathbb{E}_{q(x_t|x_0)}[KL(q(x_{t-1}|x_t, x_0)||P(x_{t-1}|x_t))]$$

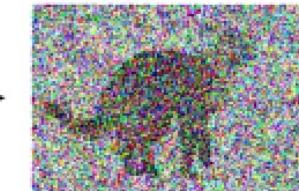


.....

Gaussian



.....



Mean

Variance

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

$$\frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \beta_t I$$

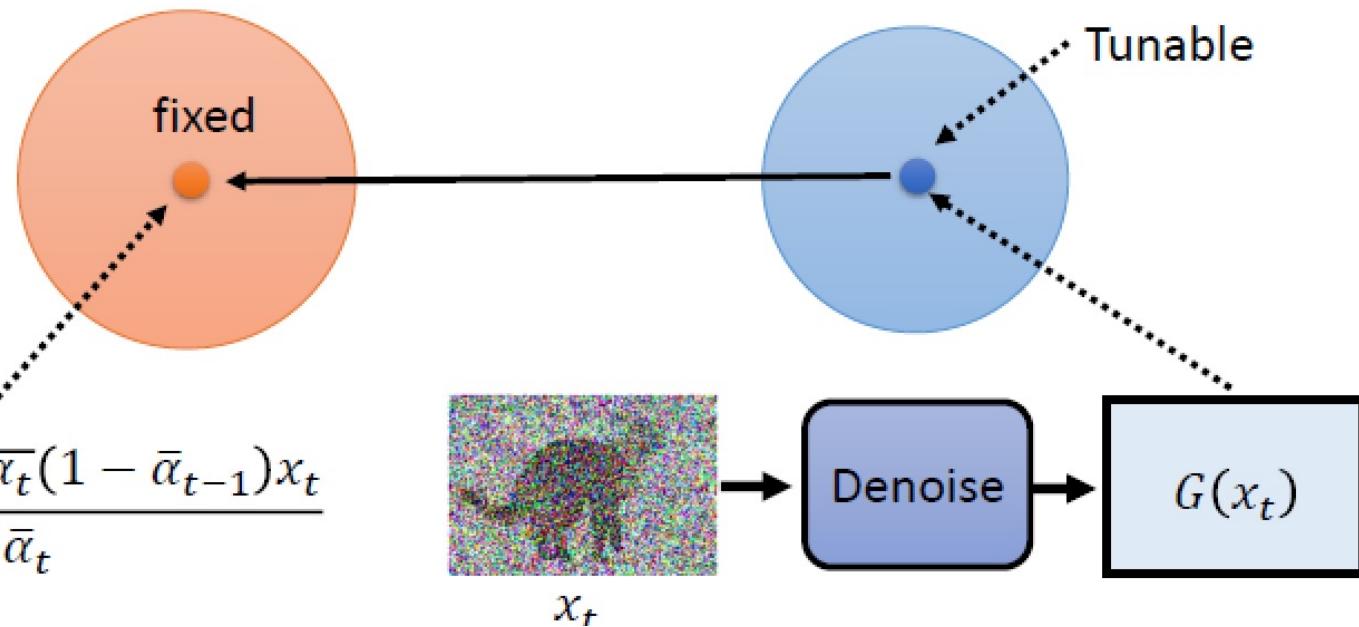
# Mean match

$$\mathbb{E}_{q(x_1|x_0)}[\log P(x_0|x_1)] - KL(q(x_T|x_0) || P(x_T))$$

$$- \sum_{t=2}^T \mathbb{E}_{q(x_t|x_0)}[KL(q(x_{t-1}|x_t, x_0) || P(x_{t-1}|x_t))]$$

How to minimize  
KL divergence?

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



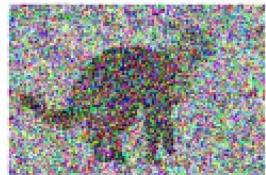
# Problem?

$$\mathbb{E}_{q(x_1|x_0)}[\log P(x_0|x_1)] - KL(q(x_T|x_0)||P(x_T))$$

$$-\sum_{t=2}^T \mathbb{E}_{q(x_t|x_0)}[KL(q(x_{t-1}|x_t, x_0)||P(x_{t-1}|x_t))]$$



Sample  $x_0$



Sample  $x_t$

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

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

Training

# Problem?

$$\mathbb{E}_{q(x_1|x_0)}[\log P(x_0|x_1)] - KL(q(x_T|x_0)||P(x_T))$$

$$- \sum_{t=2}^T \mathbb{E}_{q(x_t|x_0)}[KL(q(x_{t-1}|x_t, x_0)||P(x_{t-1}|x_t))]$$



$x_0$

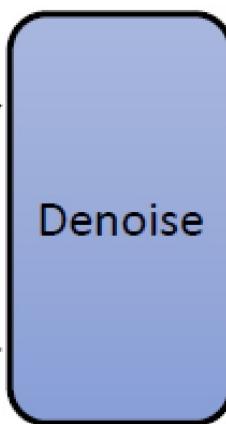


Sample  $x_t$



$x_t$

$t$



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

Inference

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

# Reparameterization



$x_0$



Sample  $x_t$

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

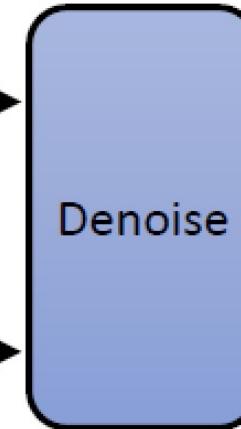
$$x_t - \sqrt{1 - \bar{\alpha}_t}\varepsilon = \sqrt{\bar{\alpha}_t}x_0$$

$$\frac{x_t - \sqrt{1 - \bar{\alpha}_t}\varepsilon}{\sqrt{\bar{\alpha}_t}} = x_0$$



$x_t$

$t$



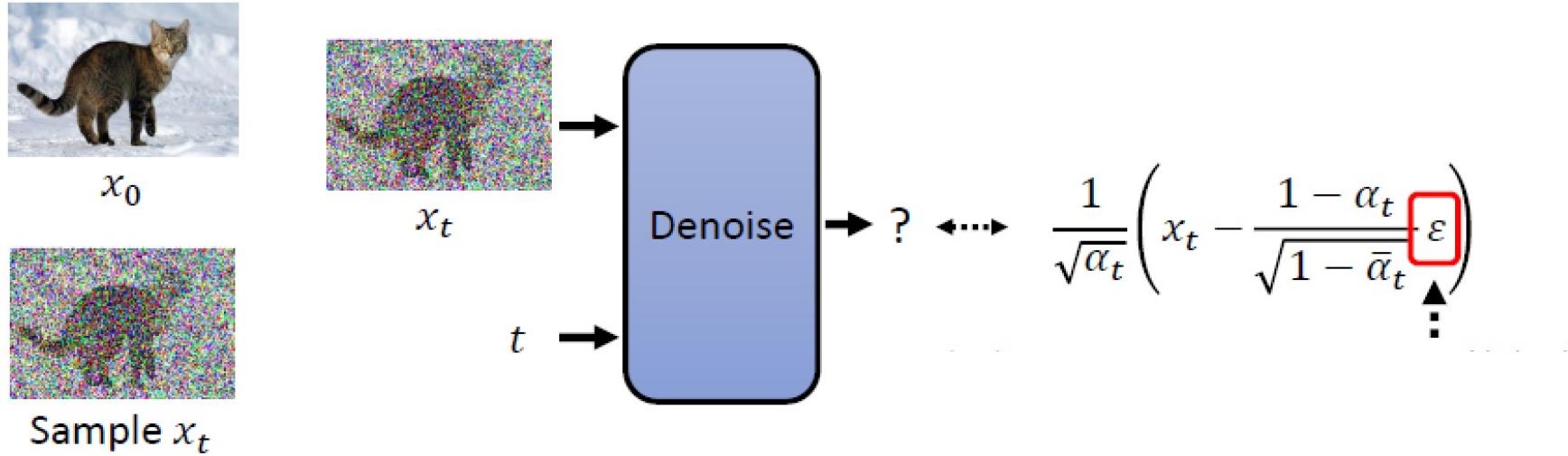
?

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

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

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

# Reparameterization



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

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

$$\frac{x_t - \sqrt{1 - \bar{\alpha}_t} \varepsilon}{\sqrt{\bar{\alpha}_t}} = x_0$$

---

## Algorithm 2 Sampling

---

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

---

02

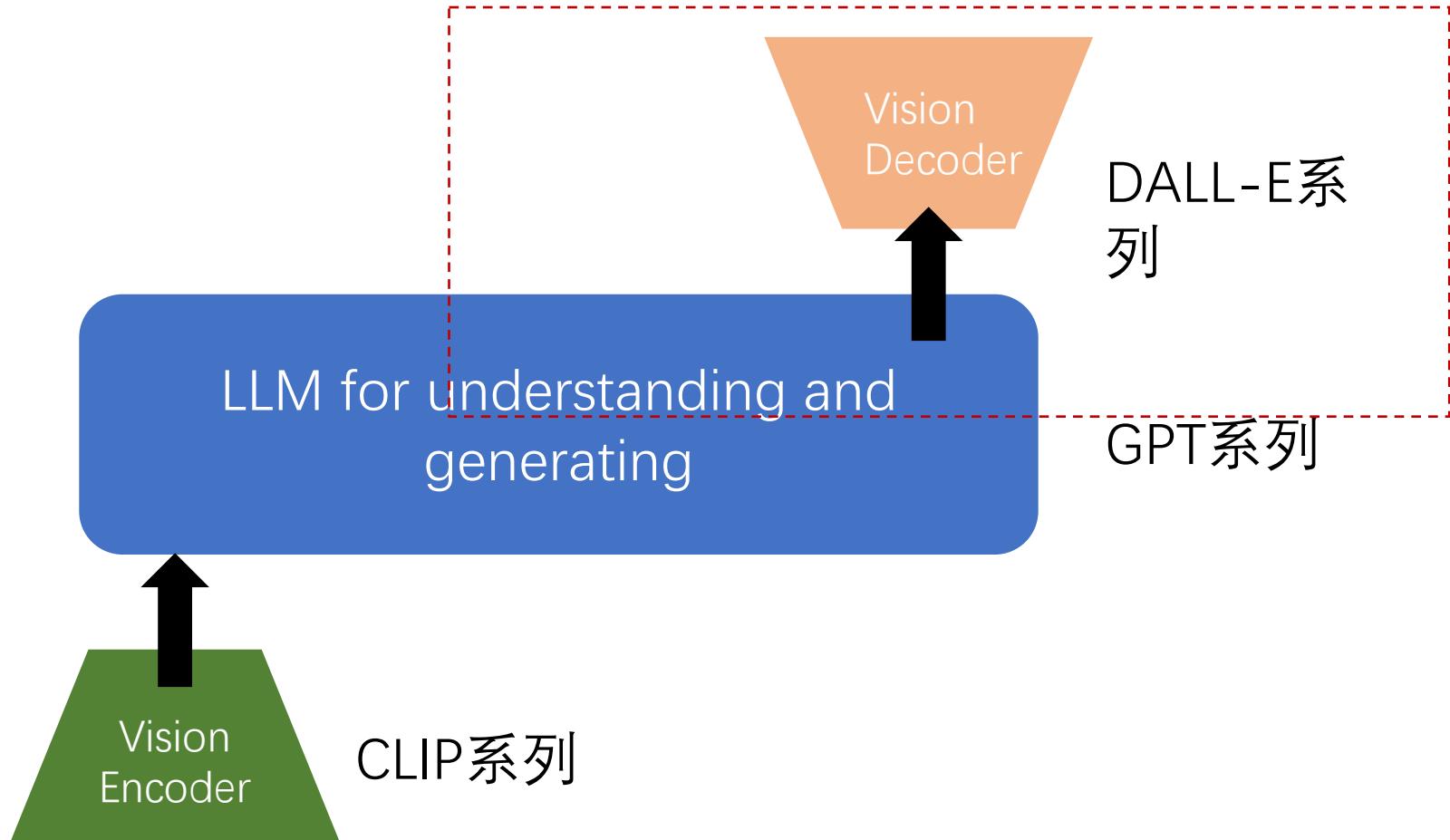
## 视觉大模型

# 多模态和视觉大模型

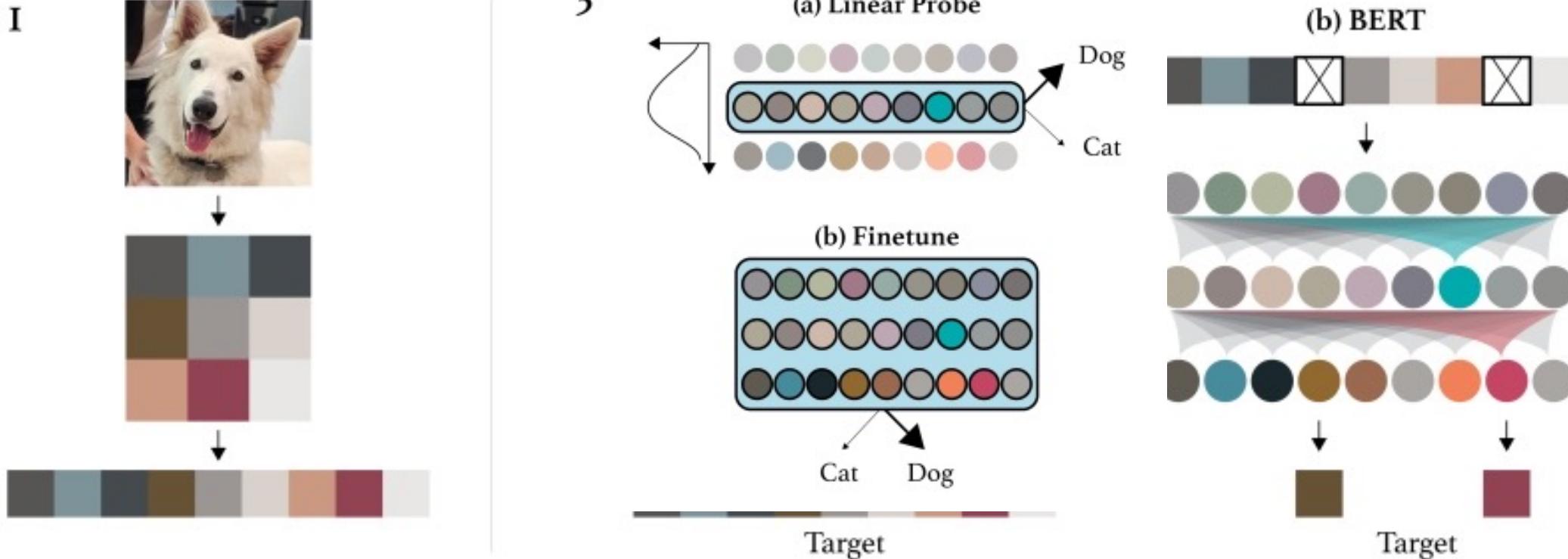
图像：

文本：

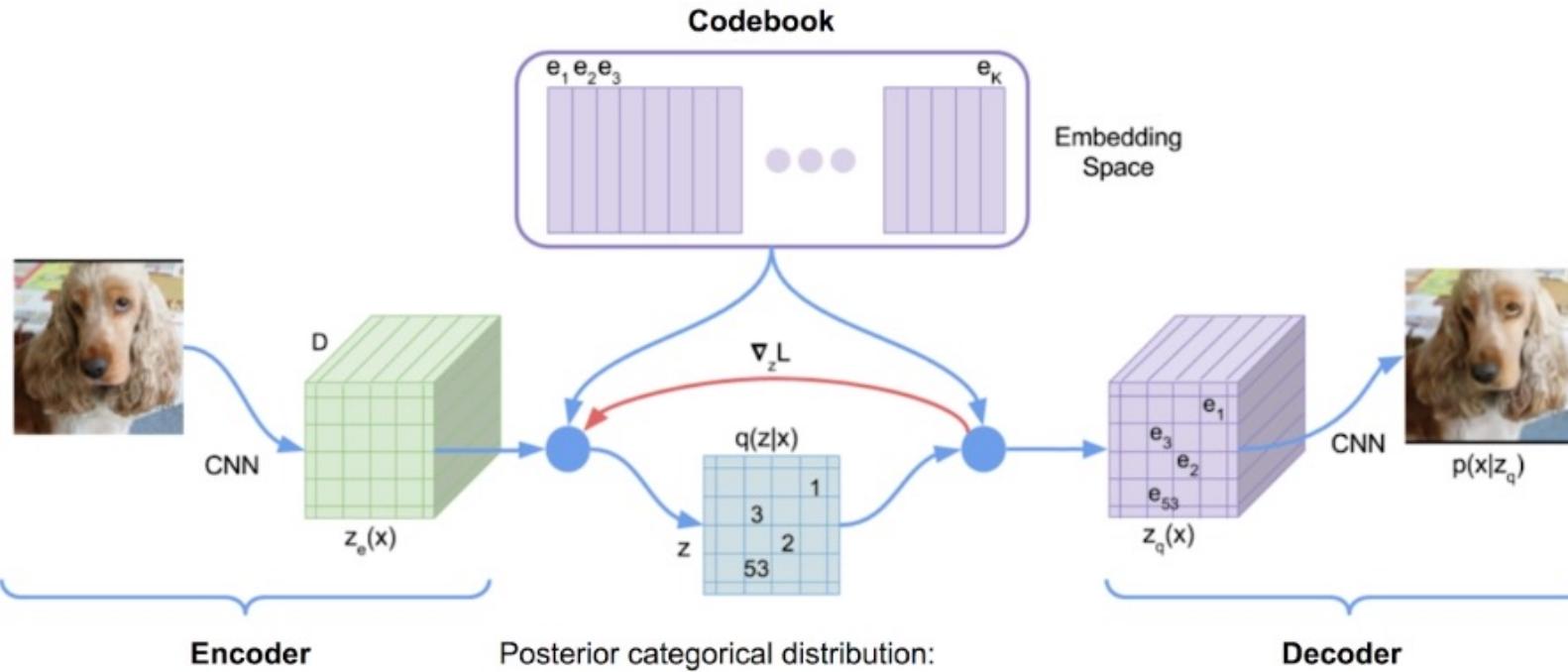
图像：



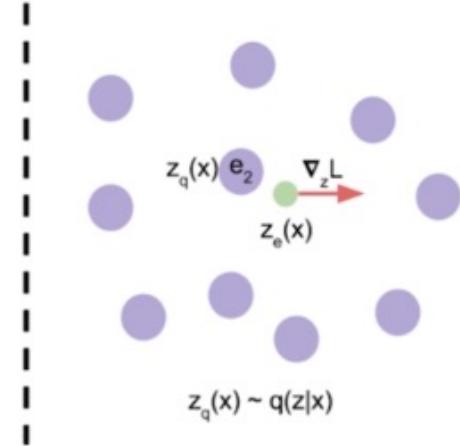
# imageGPT



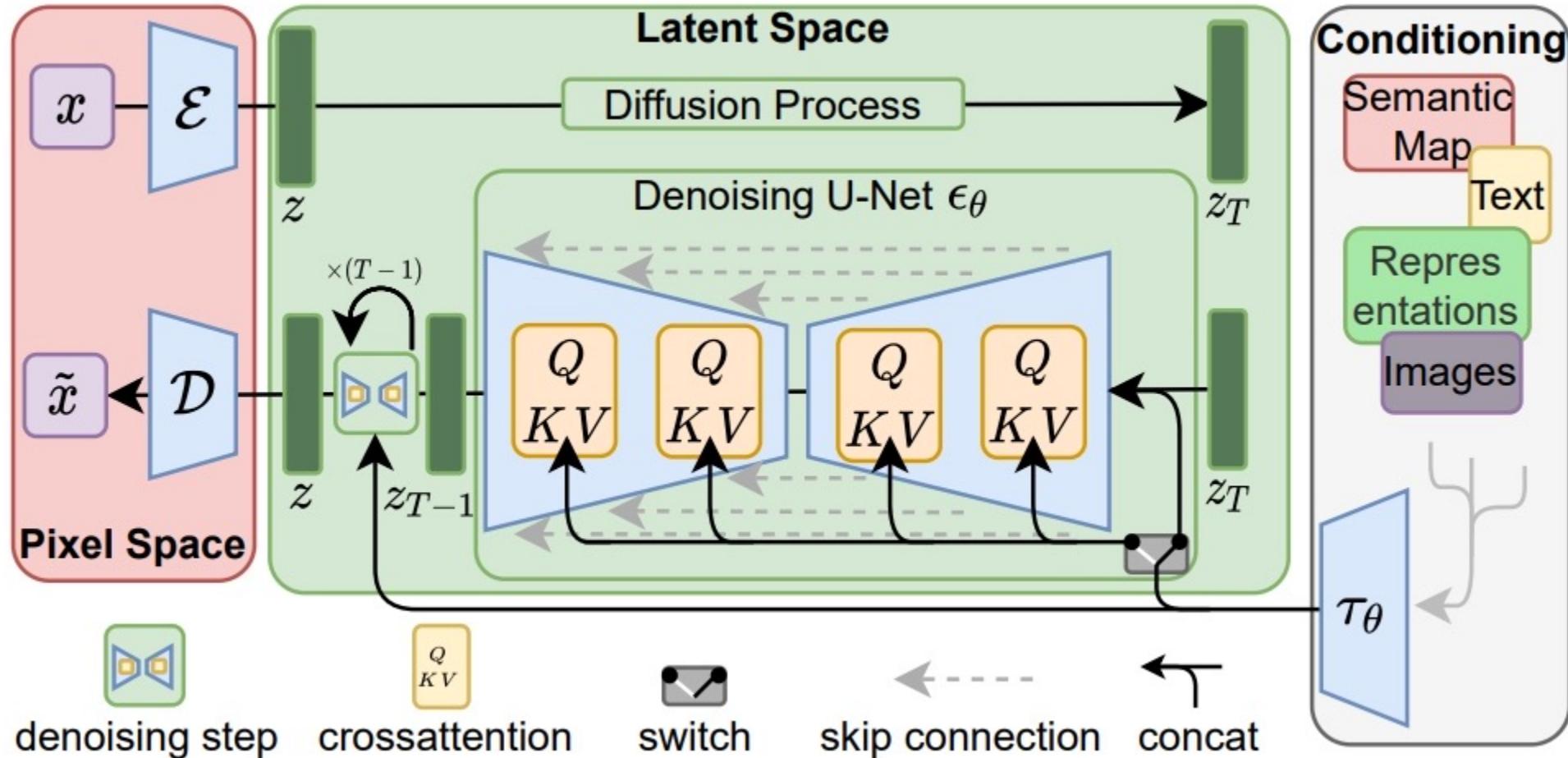
# DaLL-E



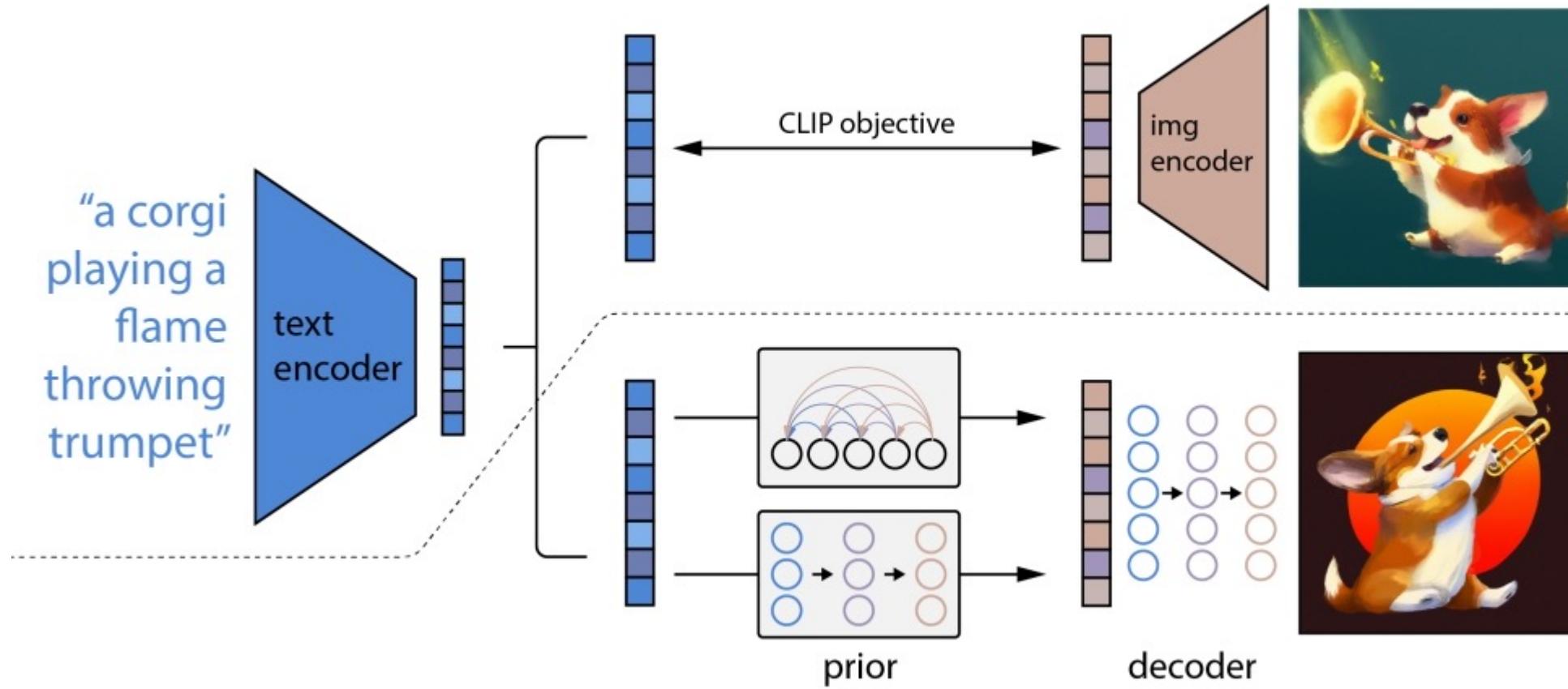
$$q(\mathbf{z} = \mathbf{e}_k | \mathbf{x}) = \begin{cases} 1 & \text{if } k = \arg \min_i \|\mathbf{z}_e(\mathbf{x}) - \mathbf{e}_i\|_2 \\ 0 & \text{otherwise.} \end{cases}$$



# Stable-Diffusion



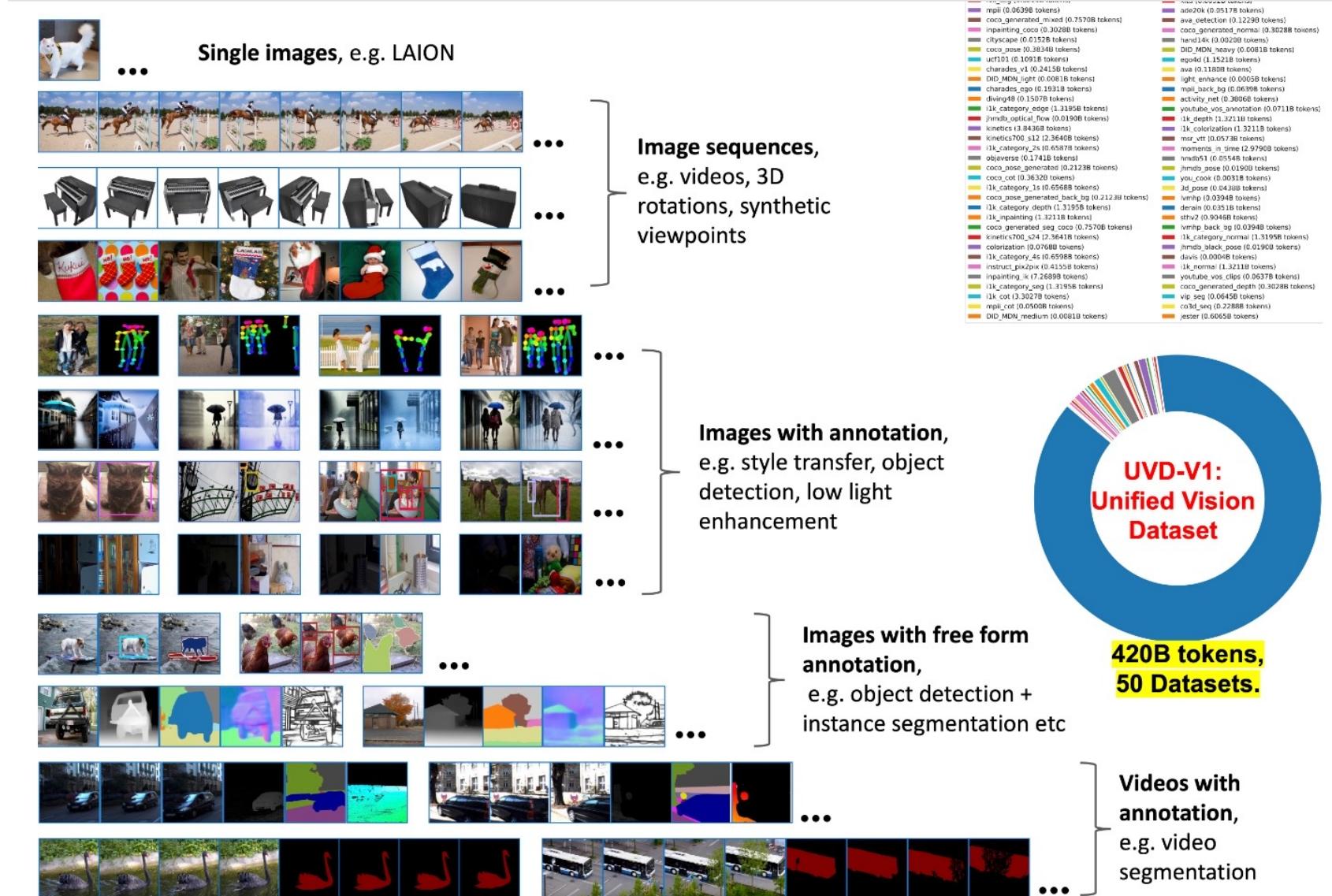
# DaLL-E2



# Autoregressive Multi-Task UVM

LVM (23.12.4)

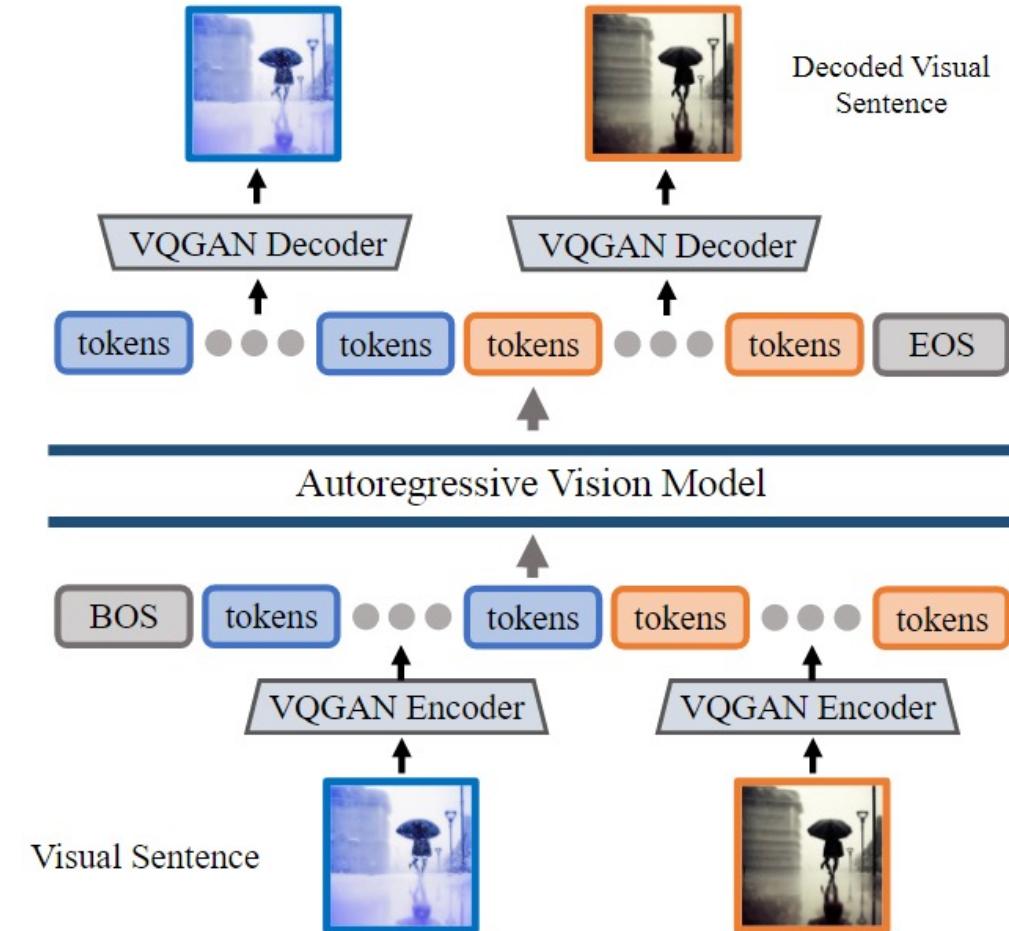
Sequential Modeling  
Enables Scalable  
Learning for Large  
Vision Models, CVPR  
2024



# Autoregressive Multi-Task UVM

- **LVM** (23.12.4)

- LVM combines autoregressive model with in-context learning. However, **this autoregressive model predicts sequences directly related to the entire image rather than visual descriptions**, distinguishing it fundamentally from pix2seq.
- LVM has the capability to capture contextual information for **out-of-domain tasks**. However, due to its lack of consideration for discrepancies across tasks, it is primarily limited to certain low-level vision tasks.



谢谢！

