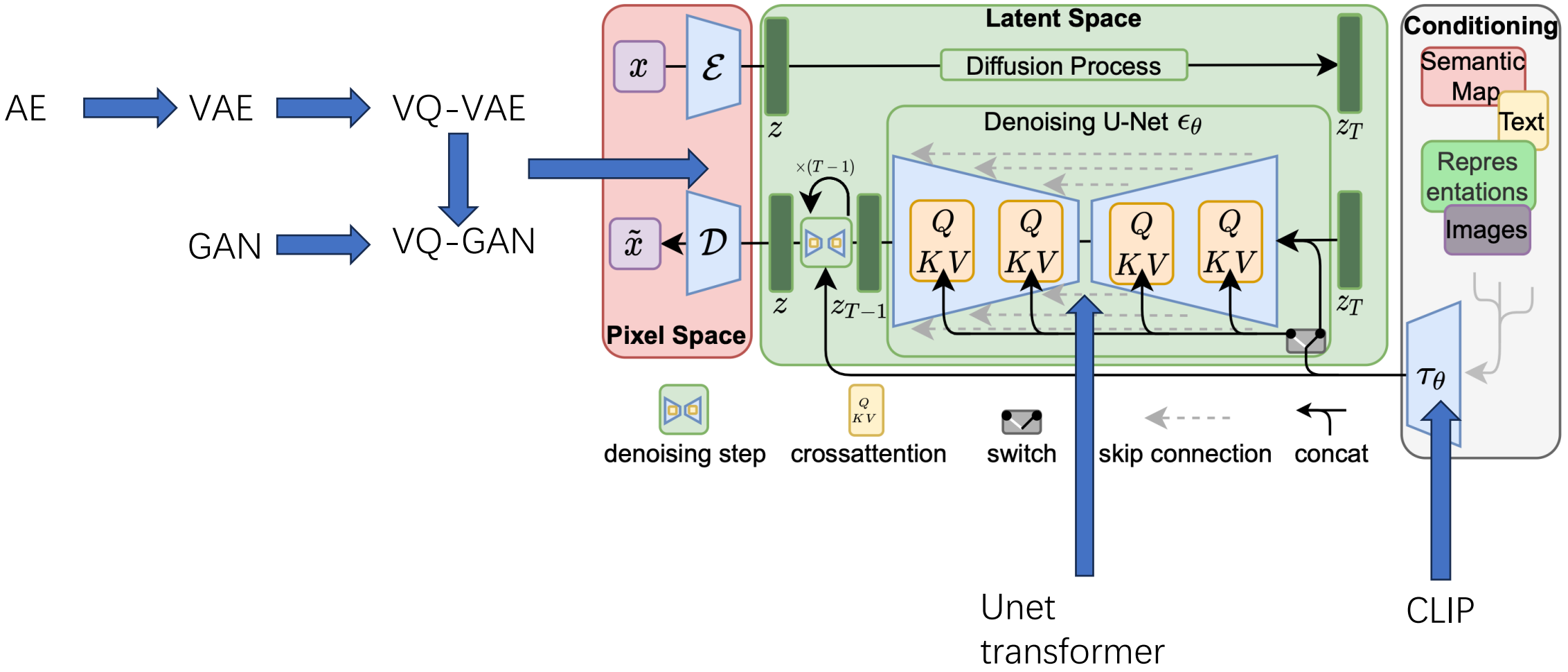
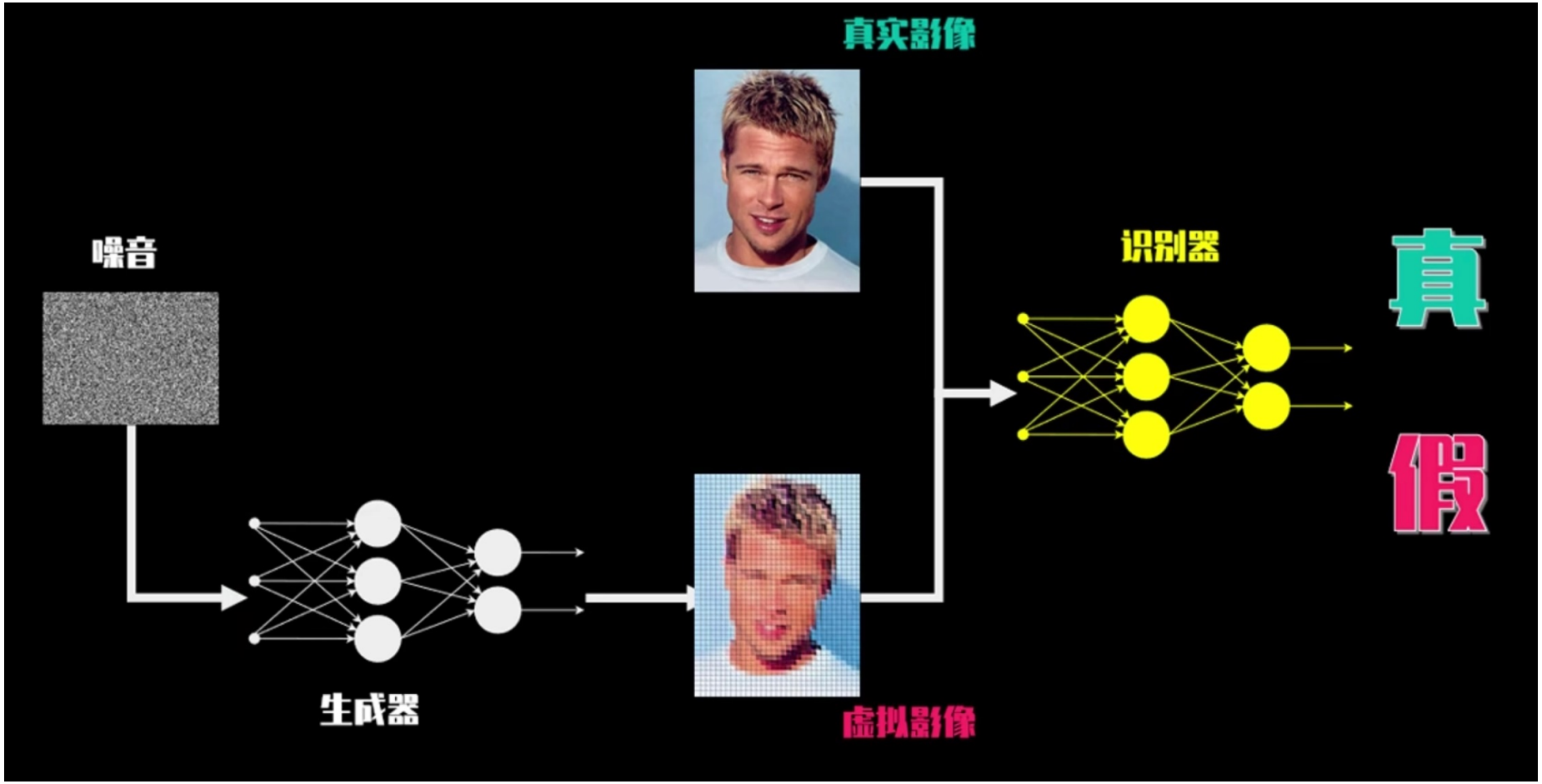
LDM预备知识

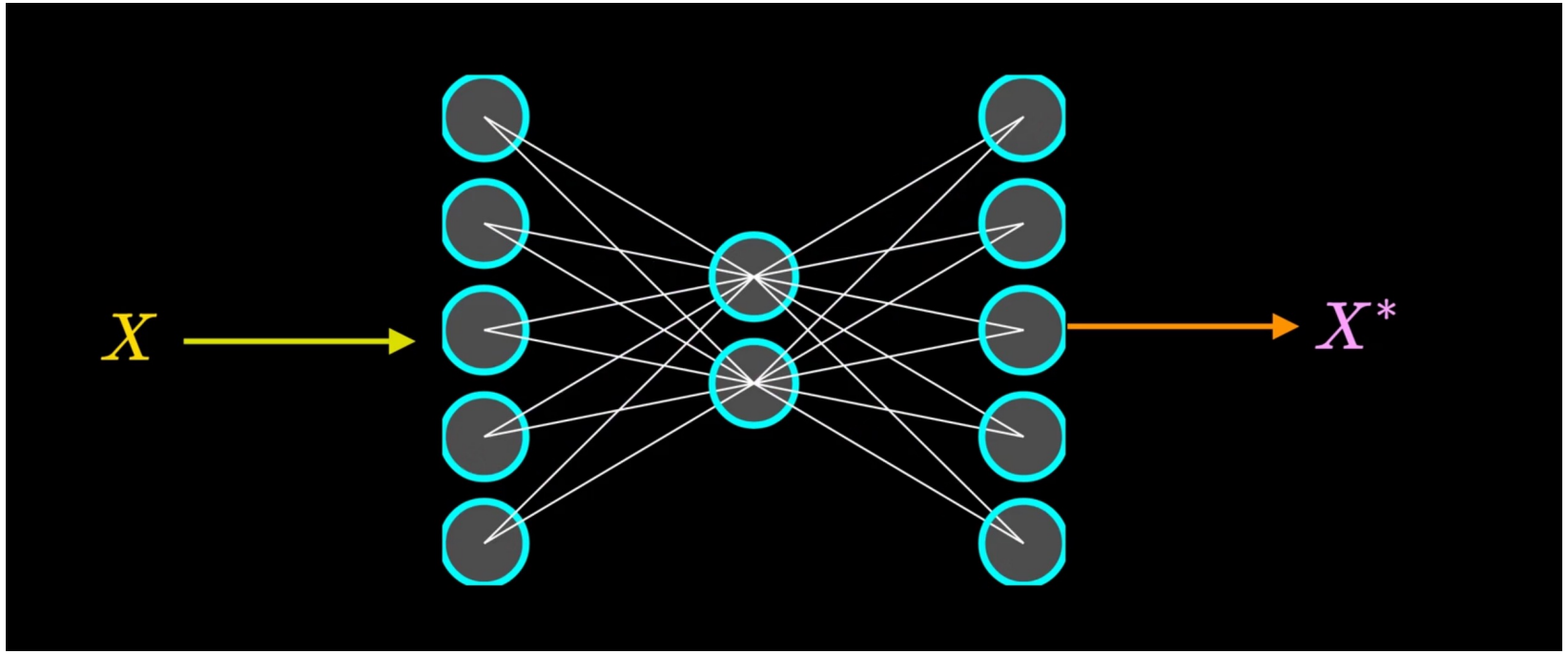
LDM -Latent Diffusion Models -潜在扩散模型



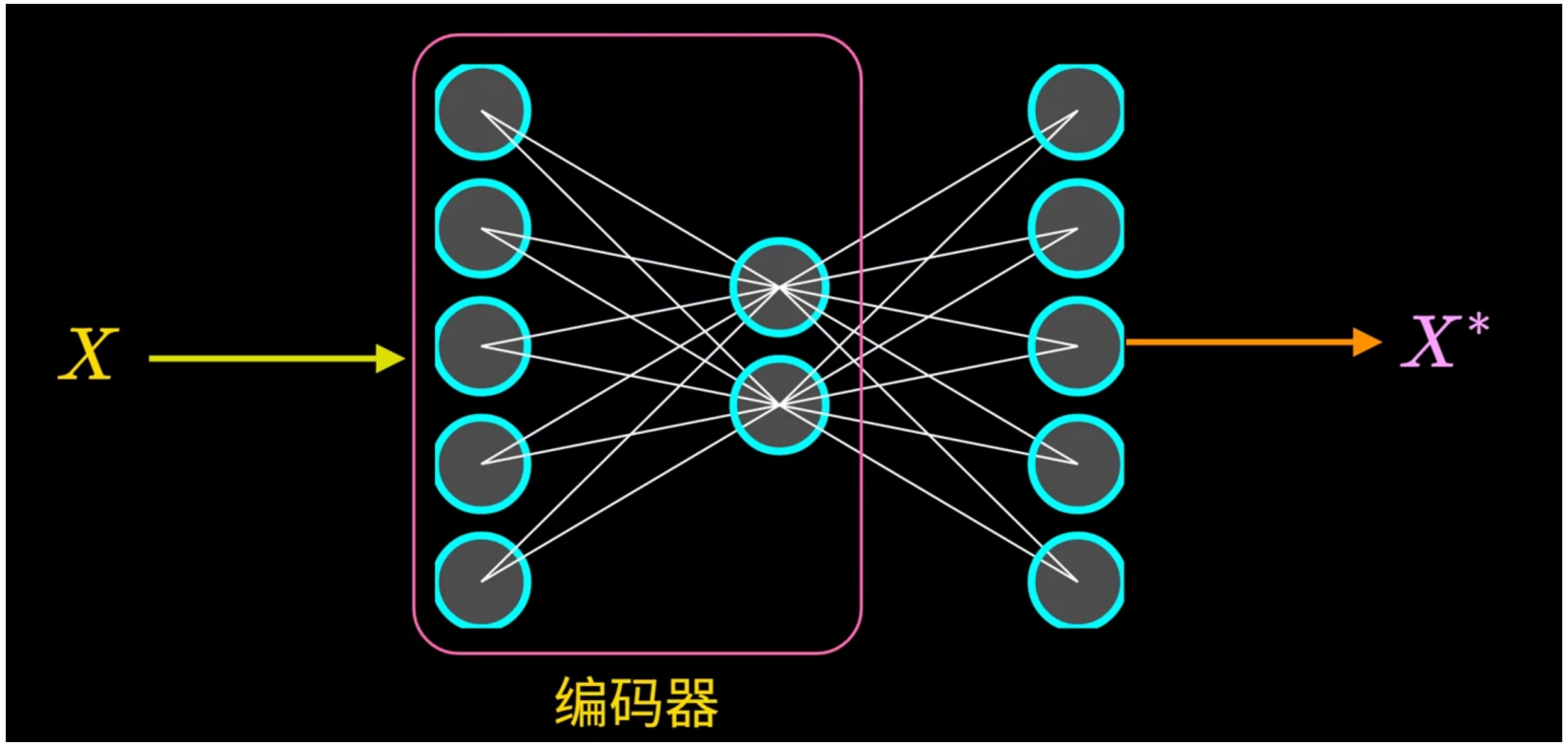
GAN –Generative Adversarial Nets -生成对抗网络



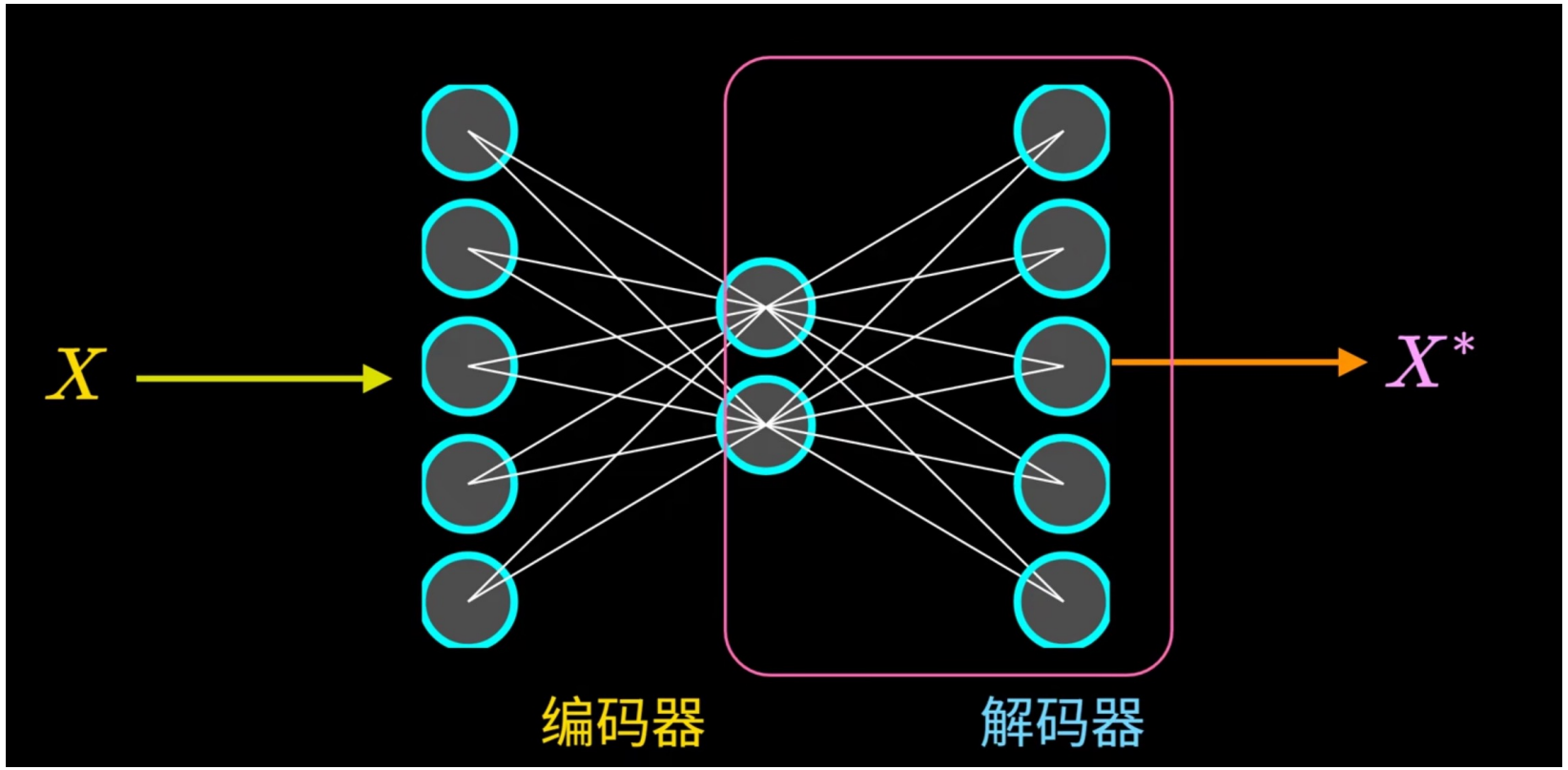
AE –autoencoder - 自编码器



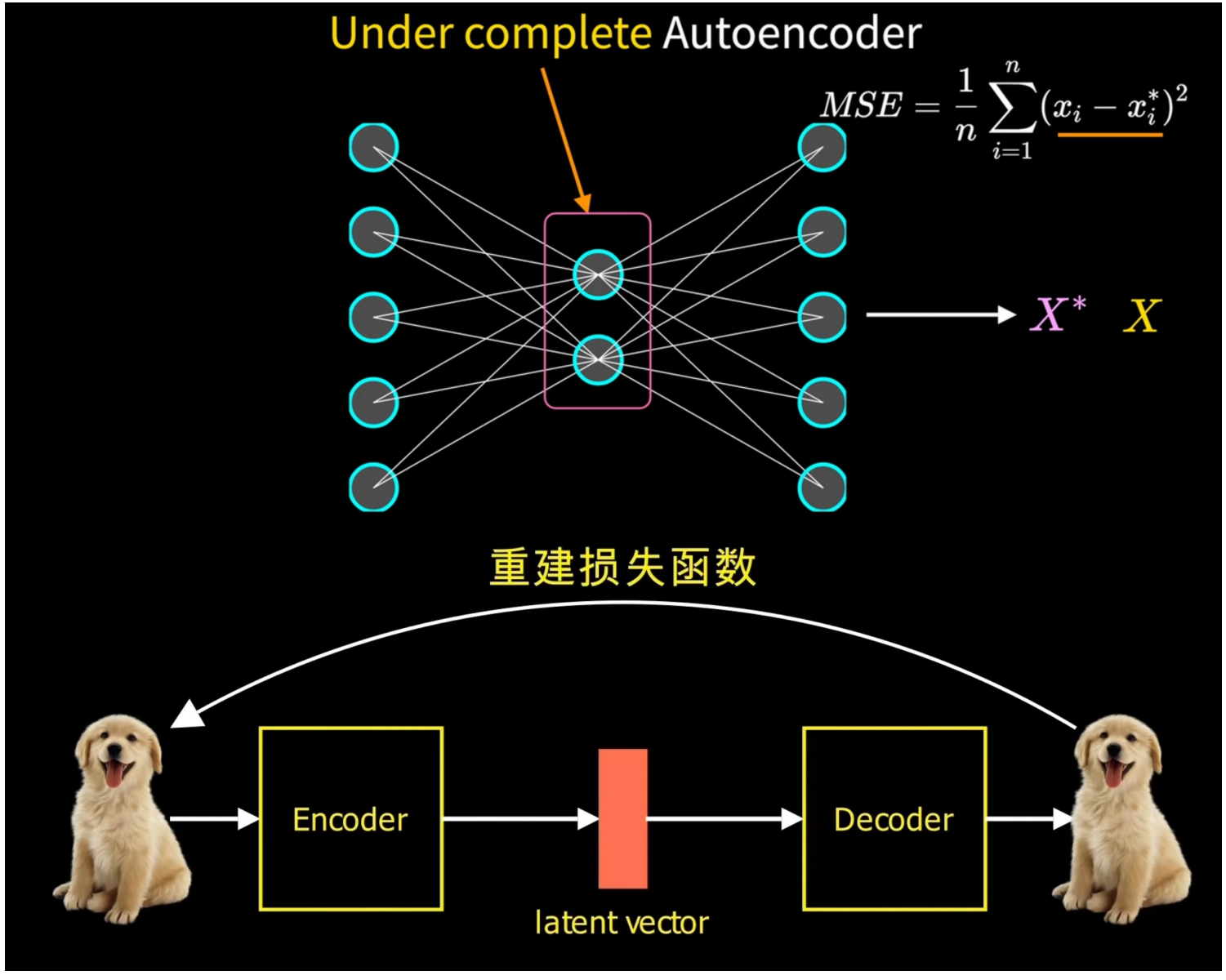
AE –autoencoder - 自编码器



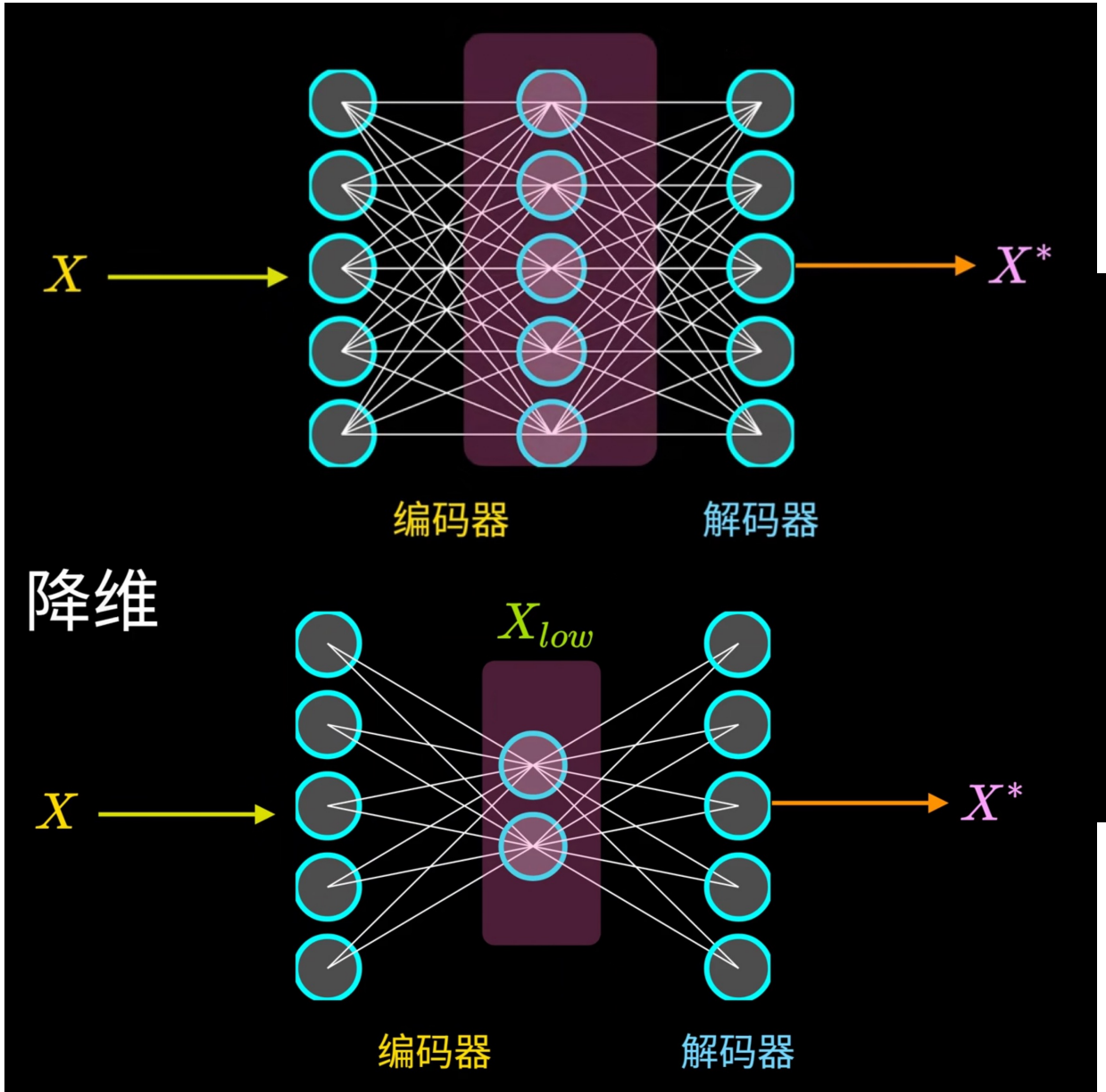
AE –autoencoder - 自编码器

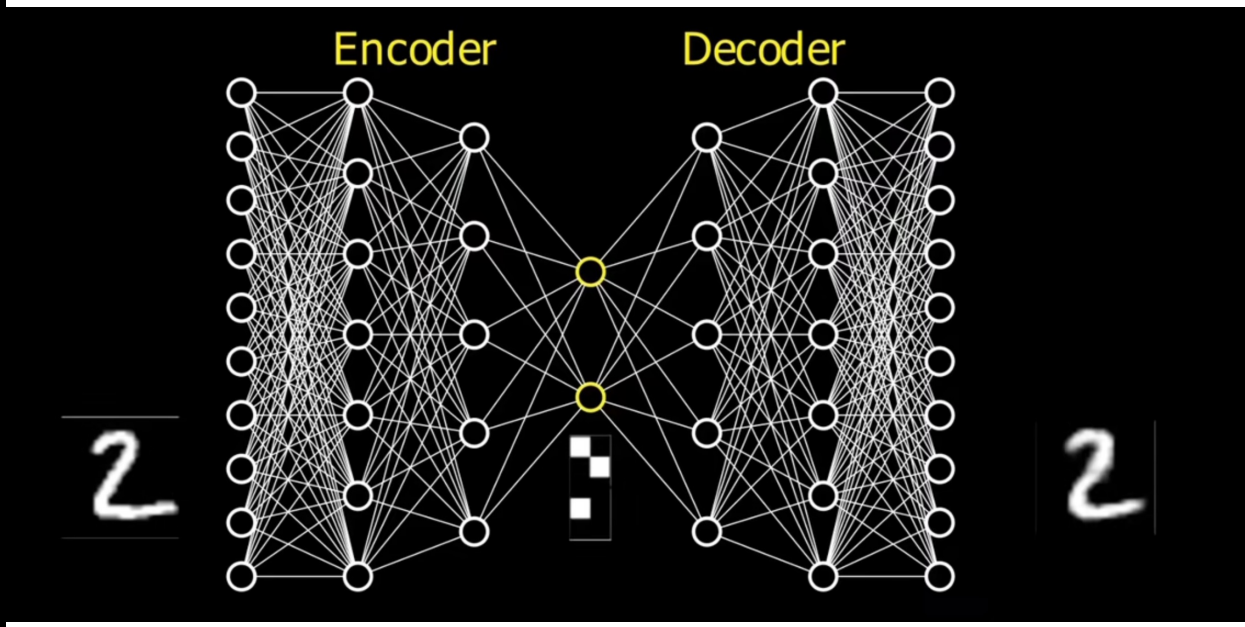


AE –autoencoder - 自编码器

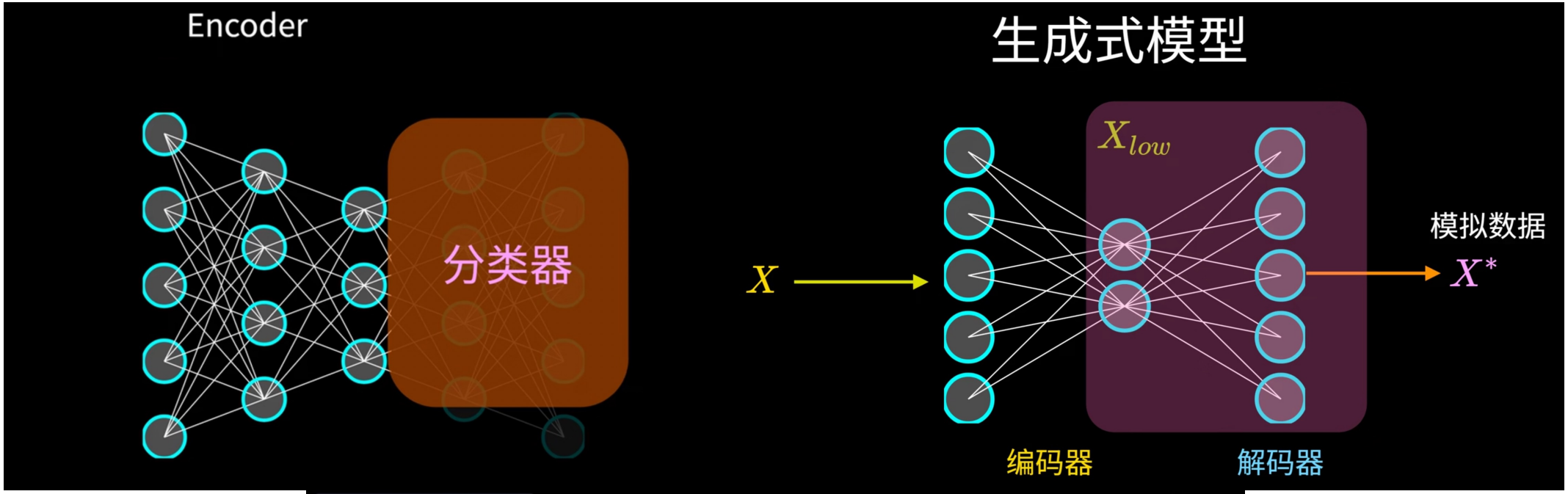


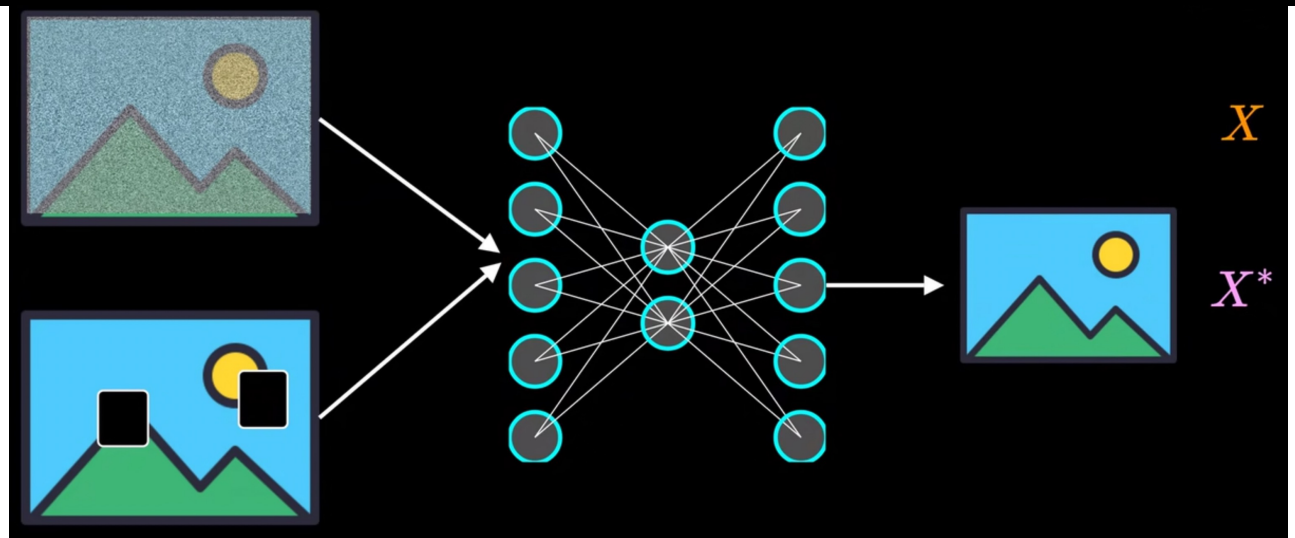
AE –autoencoder - 自编码器



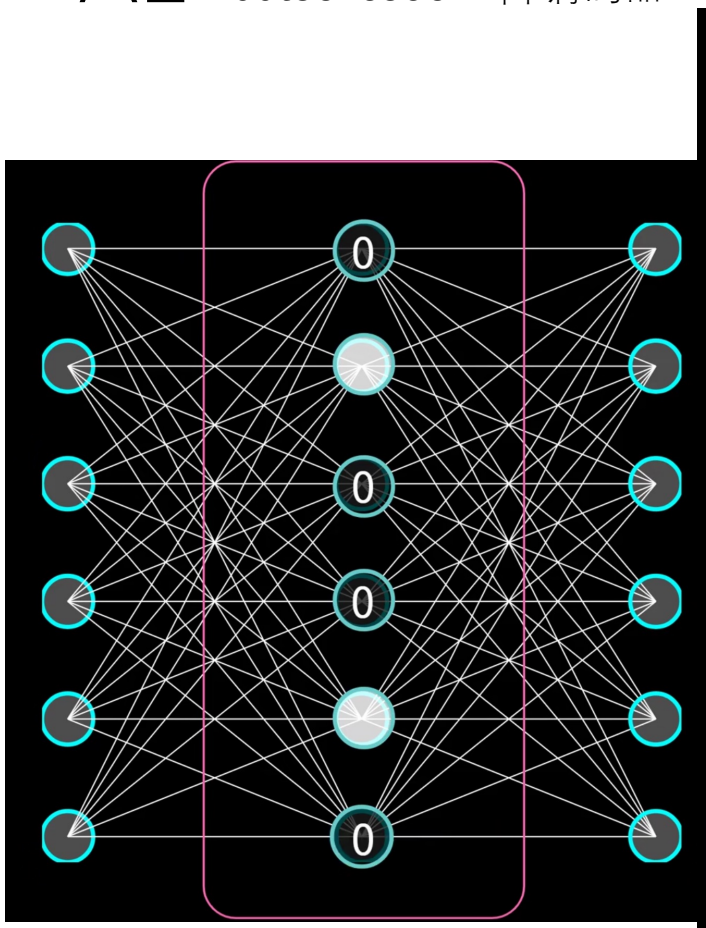


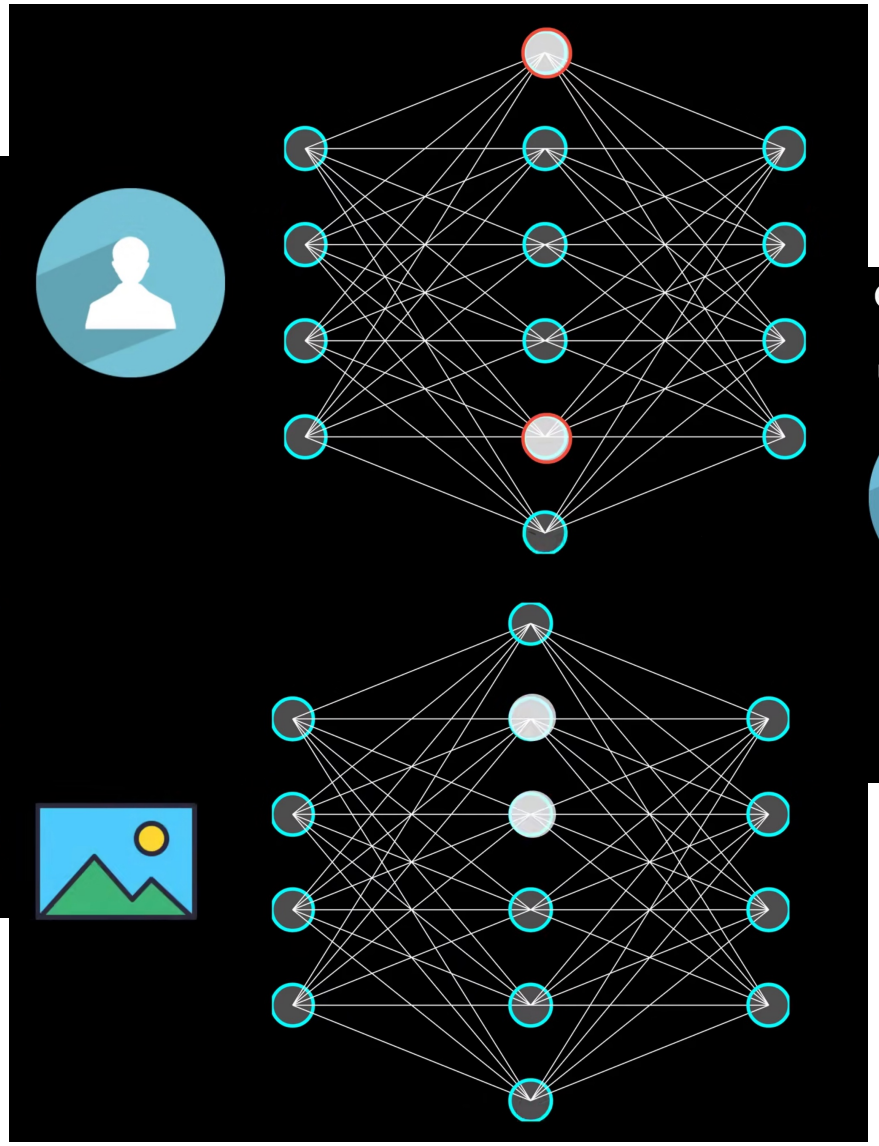
AE –autoencoder - 自编码器

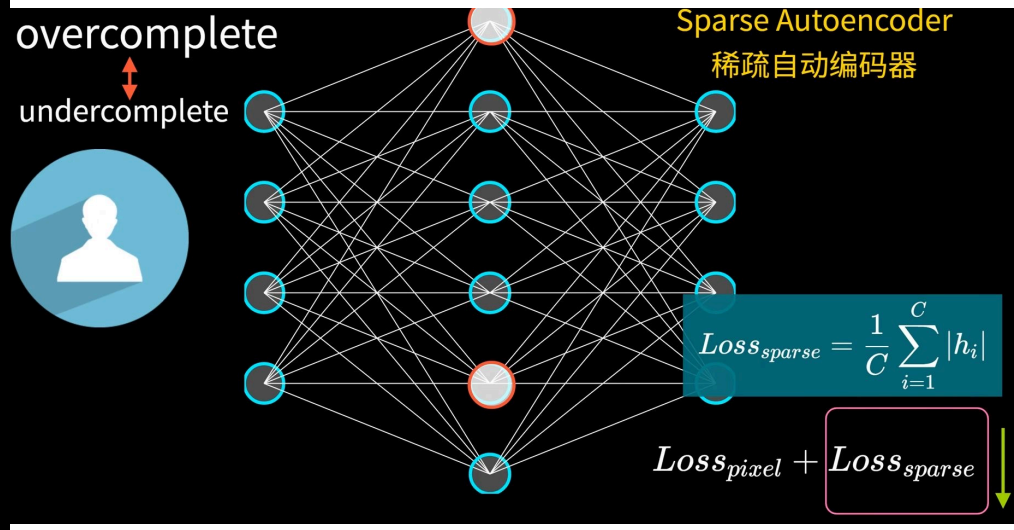




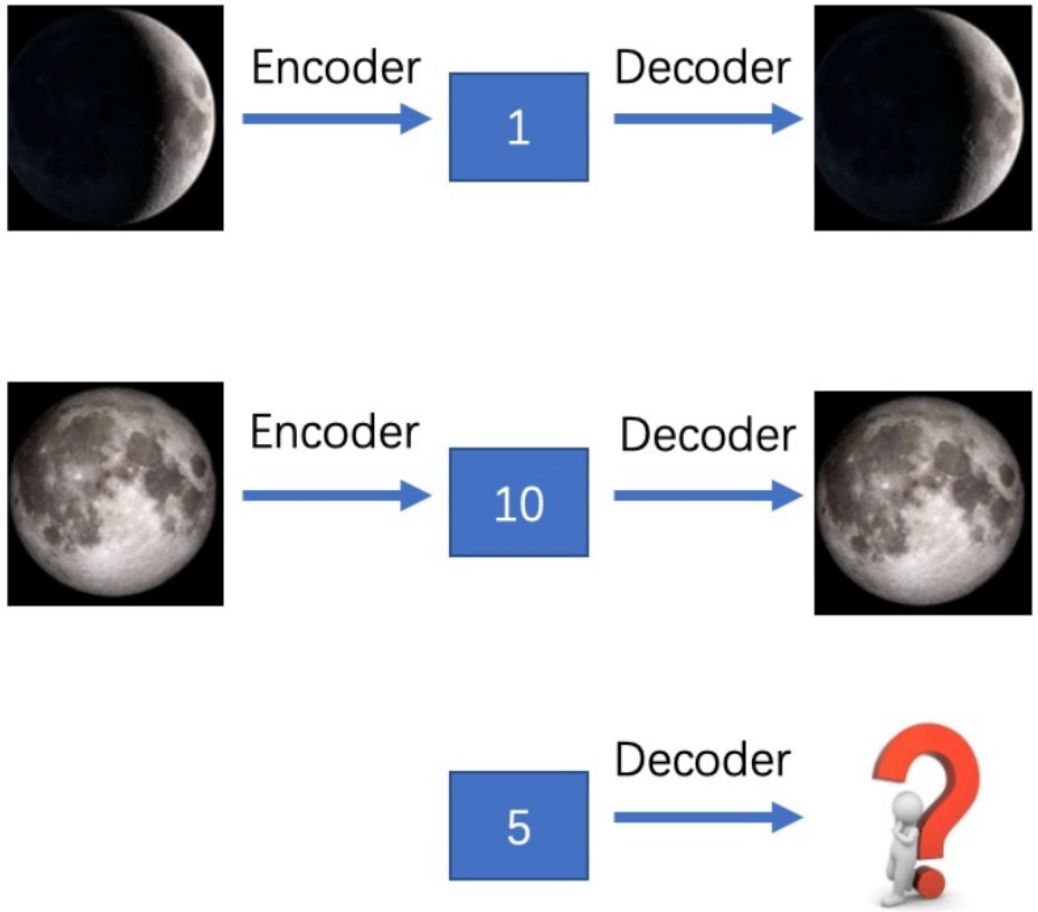
AE –autoencoder - 自编码器



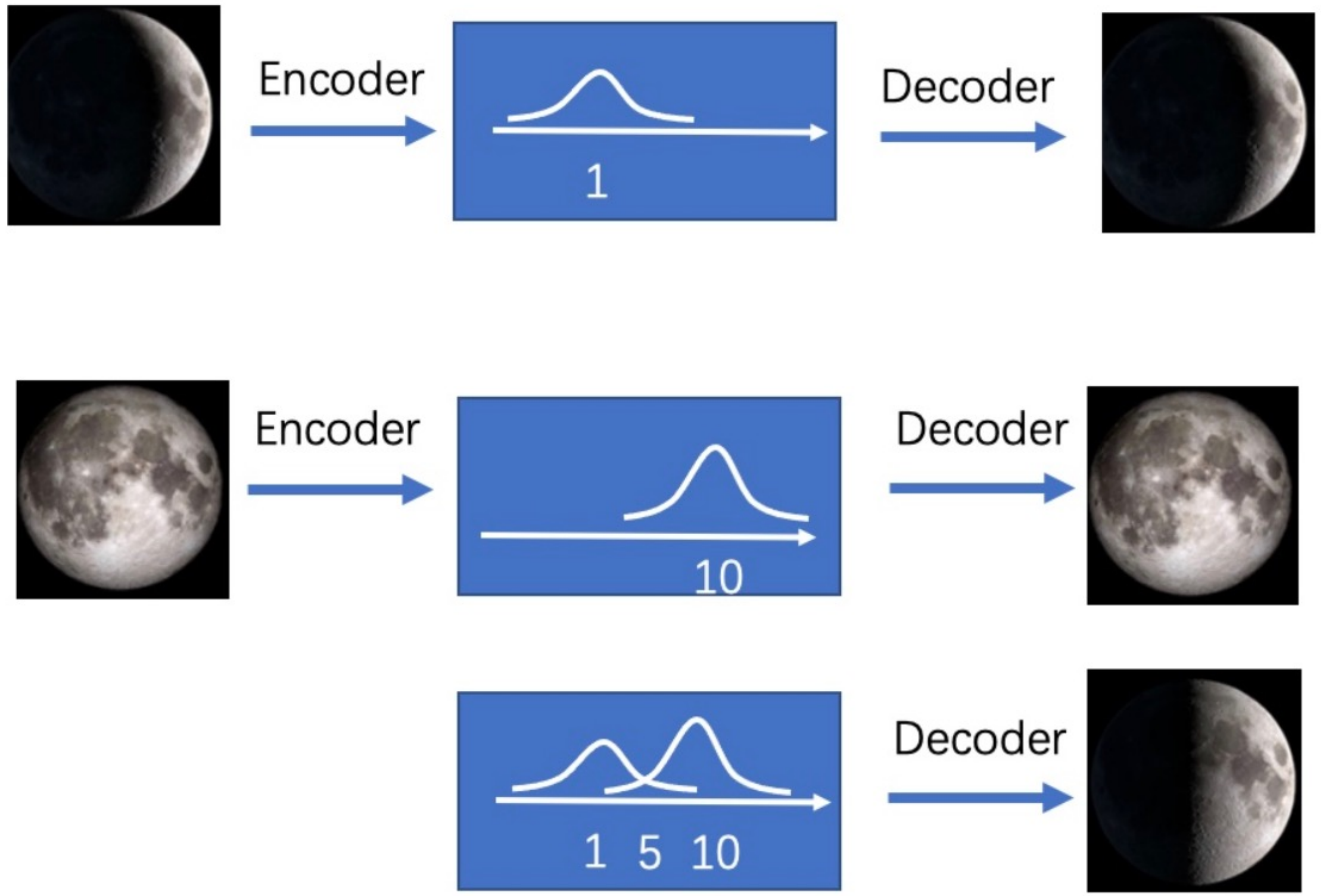




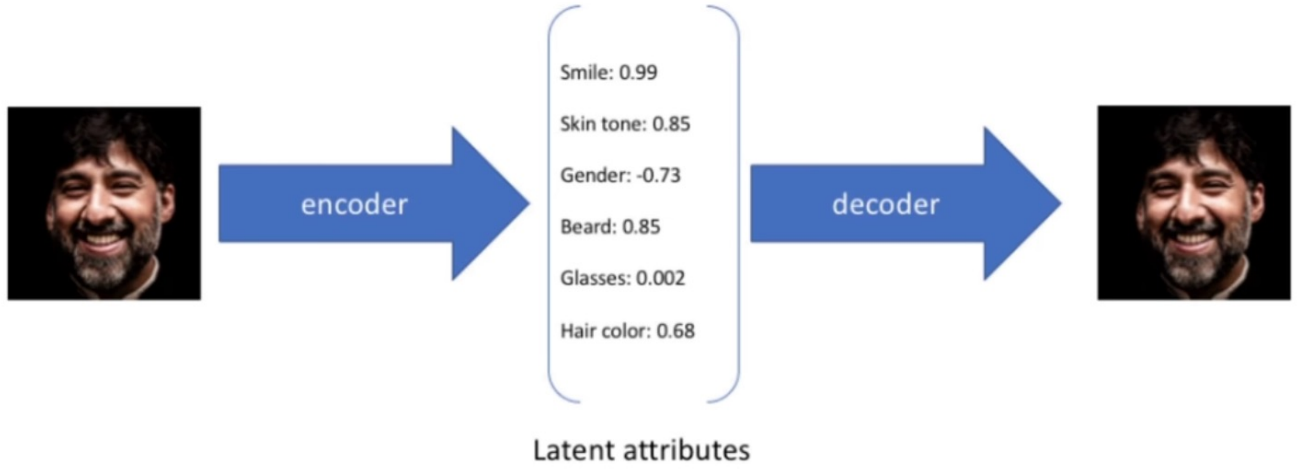
VAE -Variational Autoencoder -变分自编码器

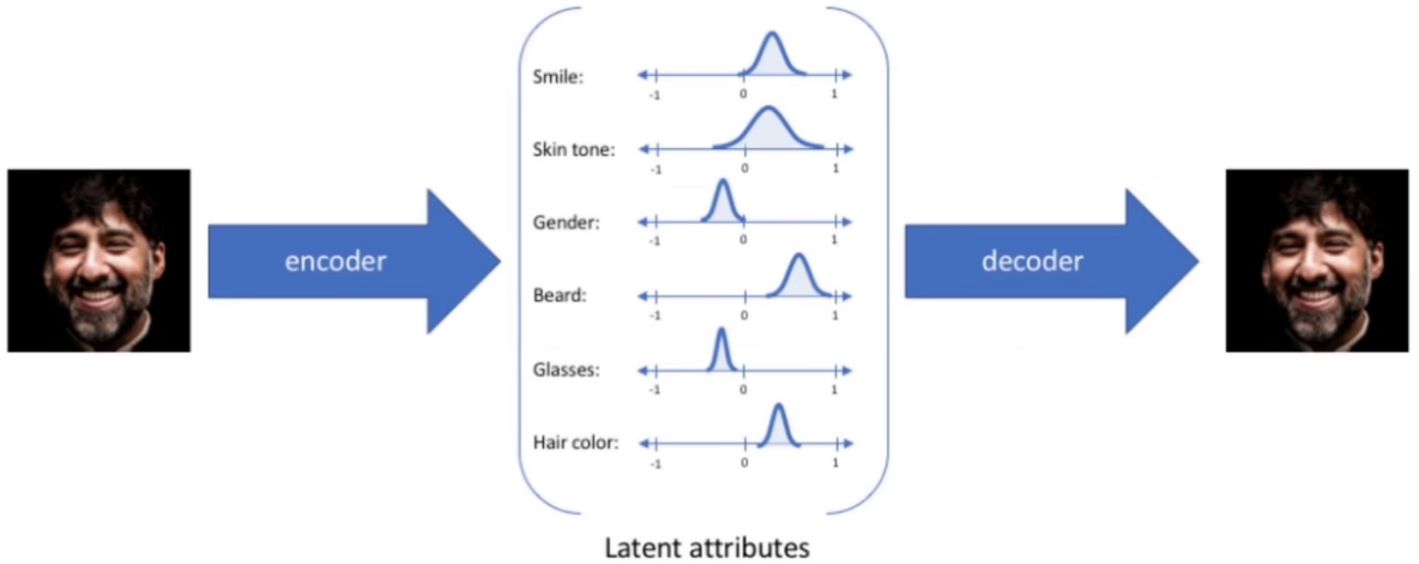


VAE -Variational Autoencoder -变分自编码器

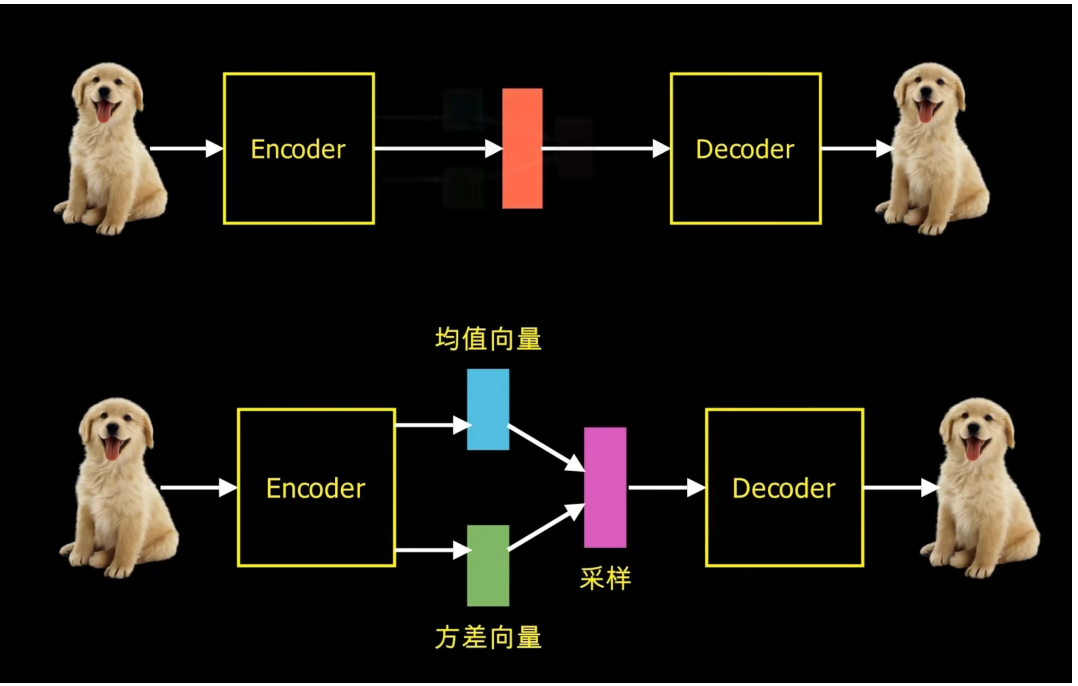


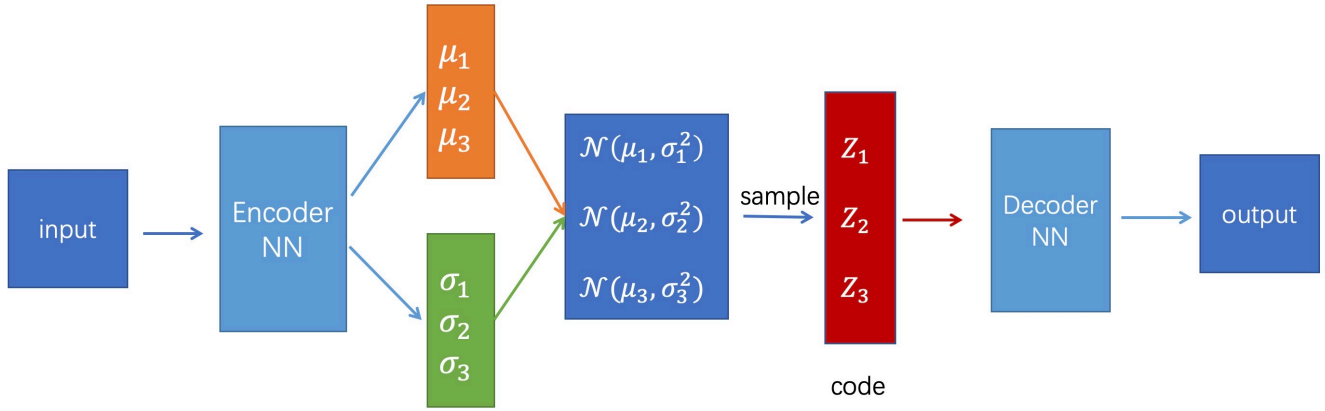
VAE -Variational Autoencoder -变分自编码器

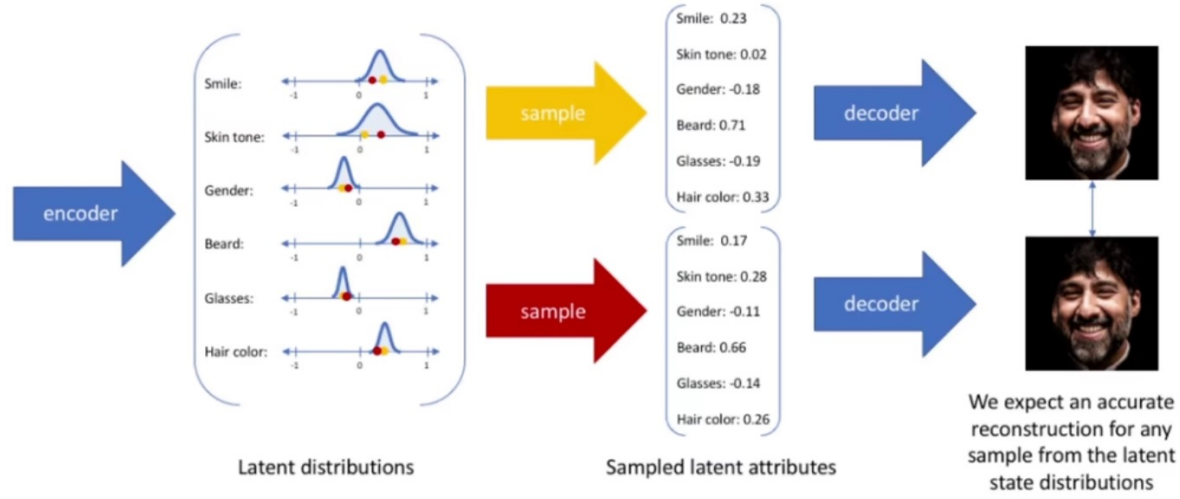




VAE -Variational Autoencoder -变分自编码器

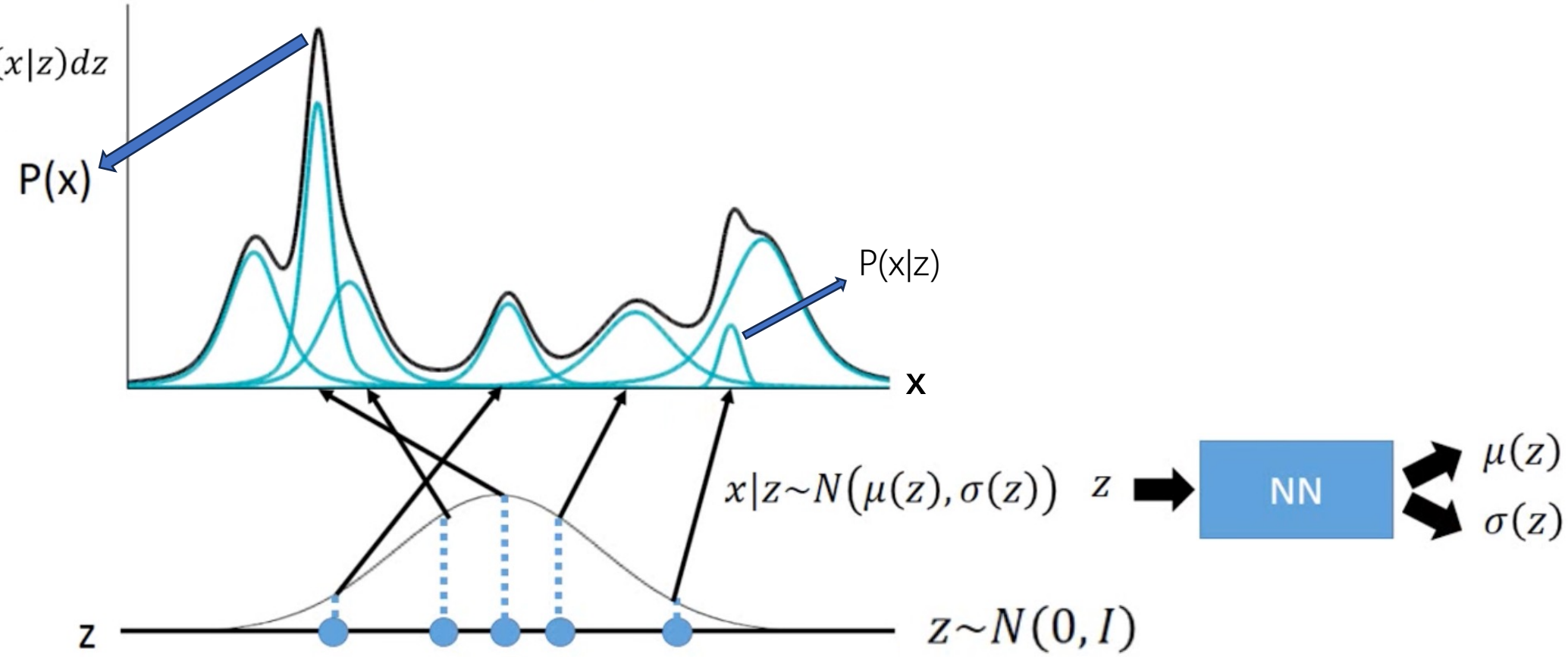




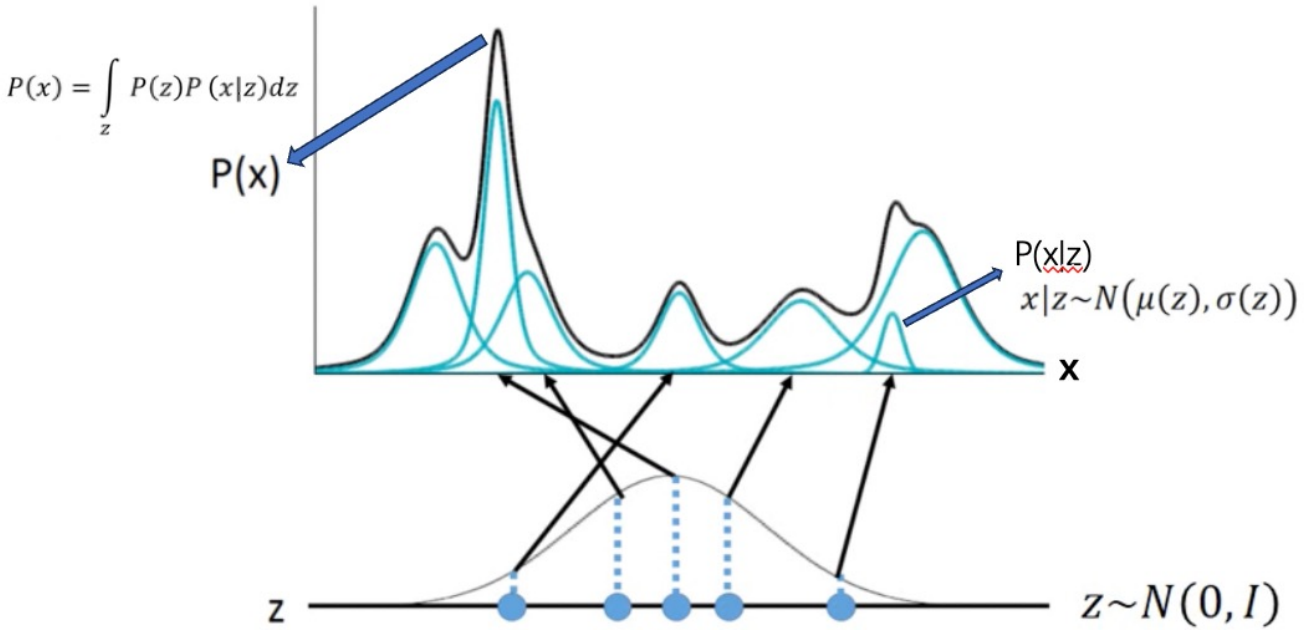


VAE -Variational Autoencoder -变分自编码器

$$P(x) = \int_z P(z)P(x|z)dz$$

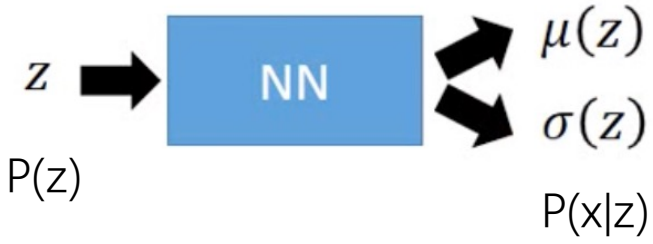


VAE -Variational Autoencoder -变分自编码器



$$P(x) = \int_z P(z)P(x|z)dz$$

$$L = \sum_x logP(x)$$



$$L = \sum_x logP(x)$$

$$logP(x) = \int_z q(z|x) logP(x) dz$$

q(z|x) can be any distribution

$$\int q(z|x)dz = 1$$

$$P(z,x) = P(z|x)P(x)$$

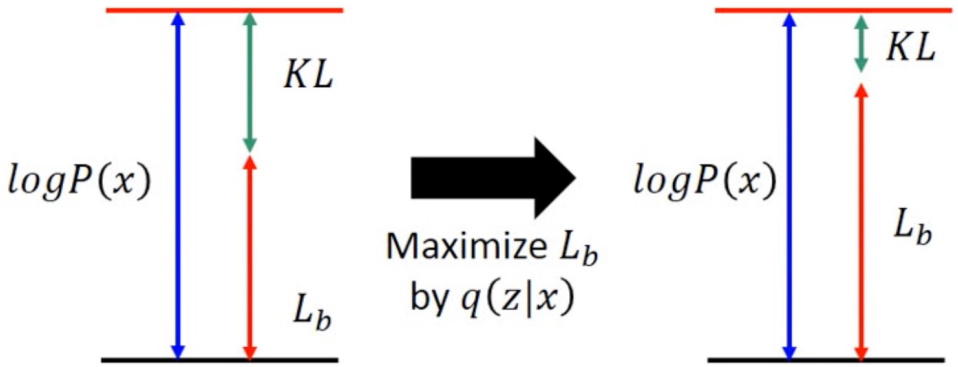
$$= \int_z q(z|x) log\left(\frac{P(z,x)}{P(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\frac{q(z|x)}{P(z|x)}\right) dz$$

$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz}_{L_b} + \underbrace{\int_z q(z|x) log\left(\frac{q(z|x)}{P(z|x)}\right) dz}_{KL(q(z|x)||P(z|x))} \geq \underbrace{\int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz}_{L_b}$$

$$logP(x) = L_b + KL(q(z|x)||P(z|x))$$

$$L_b = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz$$

$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z)}{q(z|x)}\right) dz}_{-KL(q(z|x)||P(z))} + \underbrace{\int_z q(z|x) logP(x|z) dz}_{E_{q(z|x)}[logP(x|z)]}$$



$$logP(x) = L_b + KL\big(q(z|x)||P(z|x)\big)$$

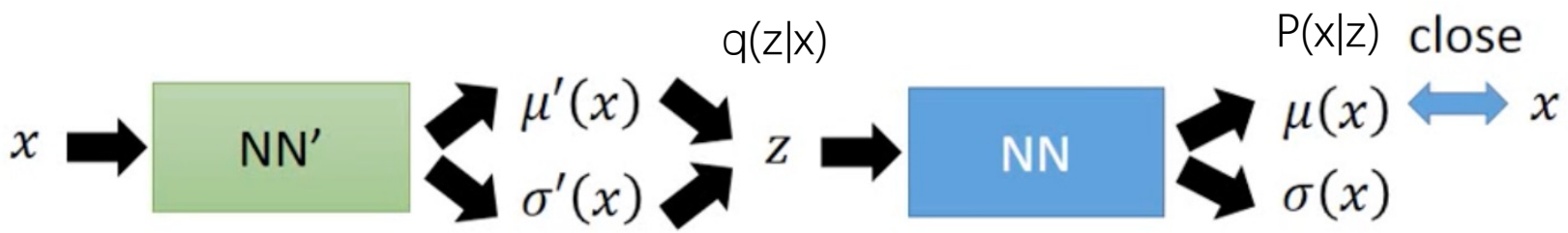
$$L_b = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz$$

$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z)}{q(z|x)}\right) dz}_{-KL\left(q(z|x)||P(z)\right)} + \underbrace{\int_z q(z|x) logP(x|z) dz}_{E_{q(z|x)}[logP(x|z)]}$$

$$p_\theta(x|z) = N(x; \mu_\theta(z), \sigma^2 I)$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\log p_\theta(x|z) = -c\|x - \mu_\theta(z)\|^2 + d$$



$$logP(x) = L_b + KL\big(q(z|x)||P(z|x)\big)$$

$$L_b = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz$$

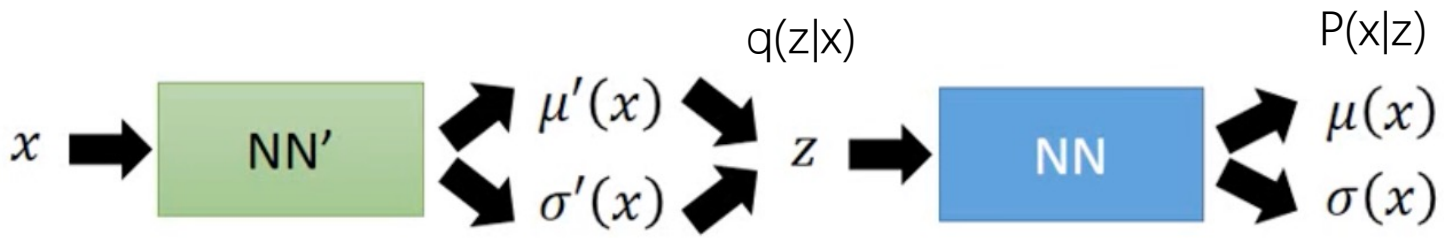
$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z)}{q(z|x)}\right) dz}_{-KL\big(q(z|x)||P(z)\big)} + \underbrace{\int_z q(z|x) logP(x|z) dz}_{E_{q(z|x)}[logP(x|z)]}$$

$$q_\varphi(z|x) = N(z; \mu_\varphi(x), \Sigma_\varphi(x))$$

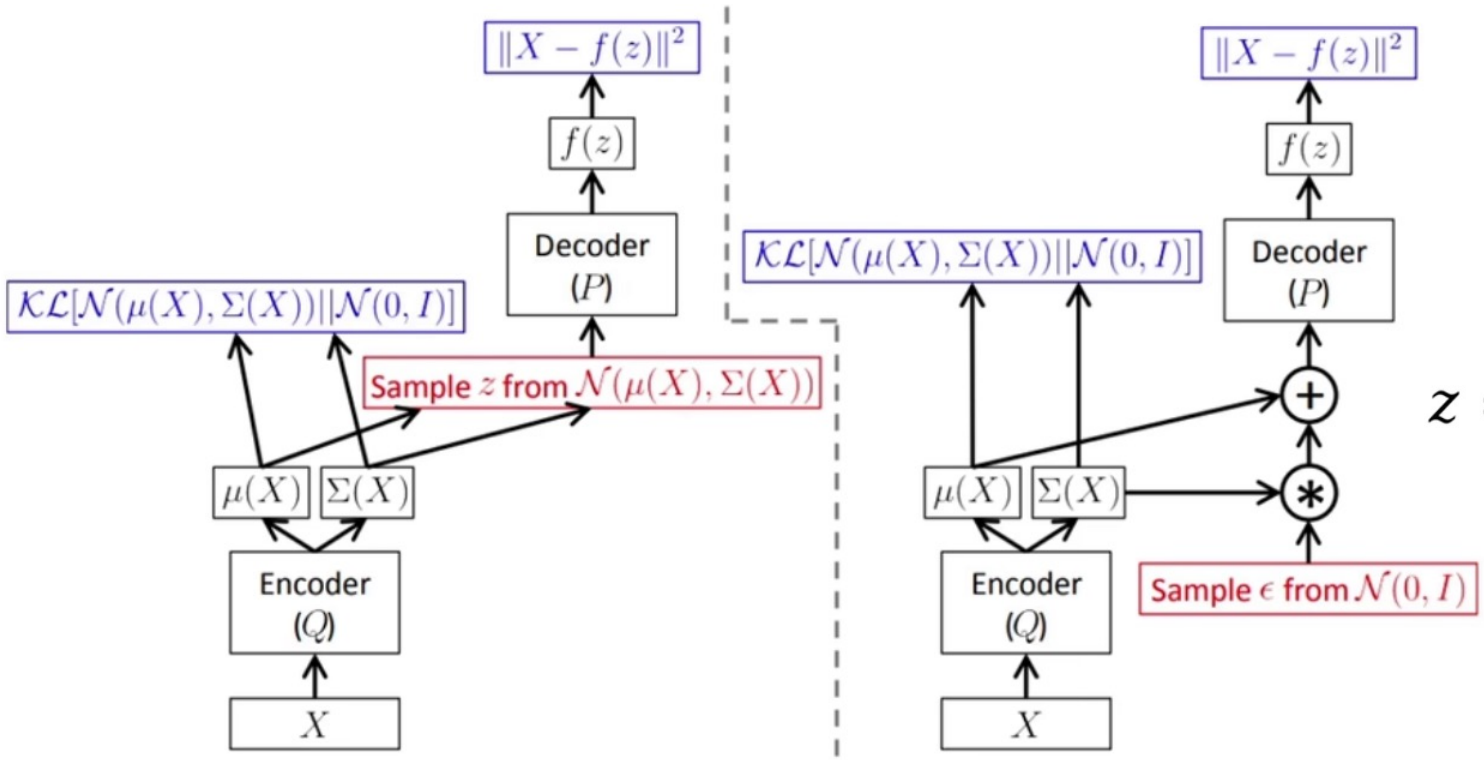
$$p(z) = N(z; 0, I)$$

$$D(N(\mu_0, \Sigma_0)\|N(\mu_1, \Sigma_1)) = \frac{1}{2}(\mathrm{tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^{\mathrm{T}}\Sigma_1^{-1}(\mu_1 - \mu_0) - k + \log(\frac{\det\Sigma_1}{\det\Sigma_0}))$$

$$D(q_\varphi(z|x)\|p(z)) = \frac{1}{2}(\mathrm{tr}(\Sigma_\varphi(x)) + \mu_\varphi(x)^{\mathrm{T}}\mu_\varphi(x) - k - \log(\det\Sigma_\varphi(x)))$$ k is the dimension



VAE -Variational Autoencoder -变分自编码器



$$z = \mu_\varphi(x) + \Sigma_\varphi^{1/2}(x)\varepsilon$$

$$logP(x) = L_b + KL\big(q(z|x)||P(z|x)\big)$$

$$L_b = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz$$

$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z)}{q(z|x)}\right) dz}_{-KL\big(q(z|x)||P(z)\big)} + \underbrace{\int_z q(z|x) logP(x|z) dz}_{E_{q(z|x)}[logP(x|z)]}$$

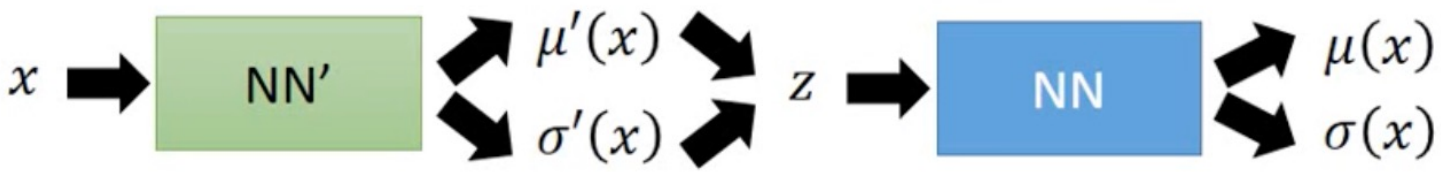
$$= \underbrace{E_{q_\phi(z|x)}\big[\log\big(p_\theta(x|z)\big)\big]}_{\text{Reconstruction Loss}} - \underbrace{D_{KL}(q_\phi(z|x)||p_\theta(z))}_{\text{Regularization Loss}}$$

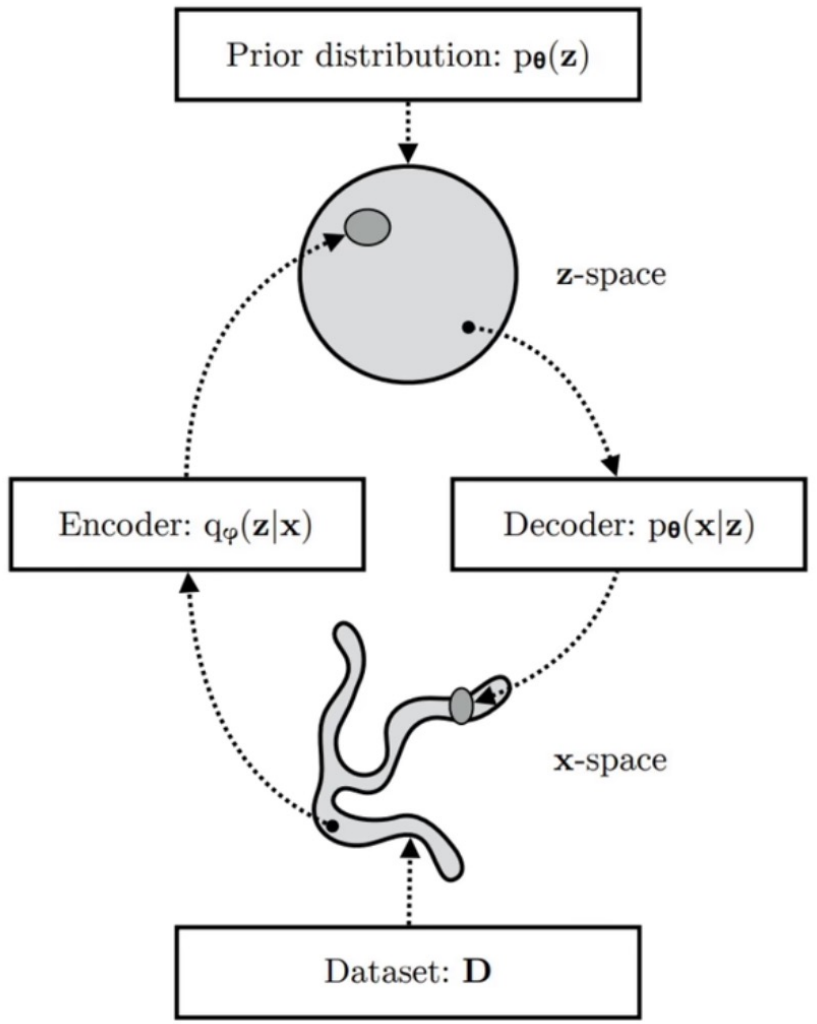
P(z)

P(x|z)

q(z|x)

P(x|z)



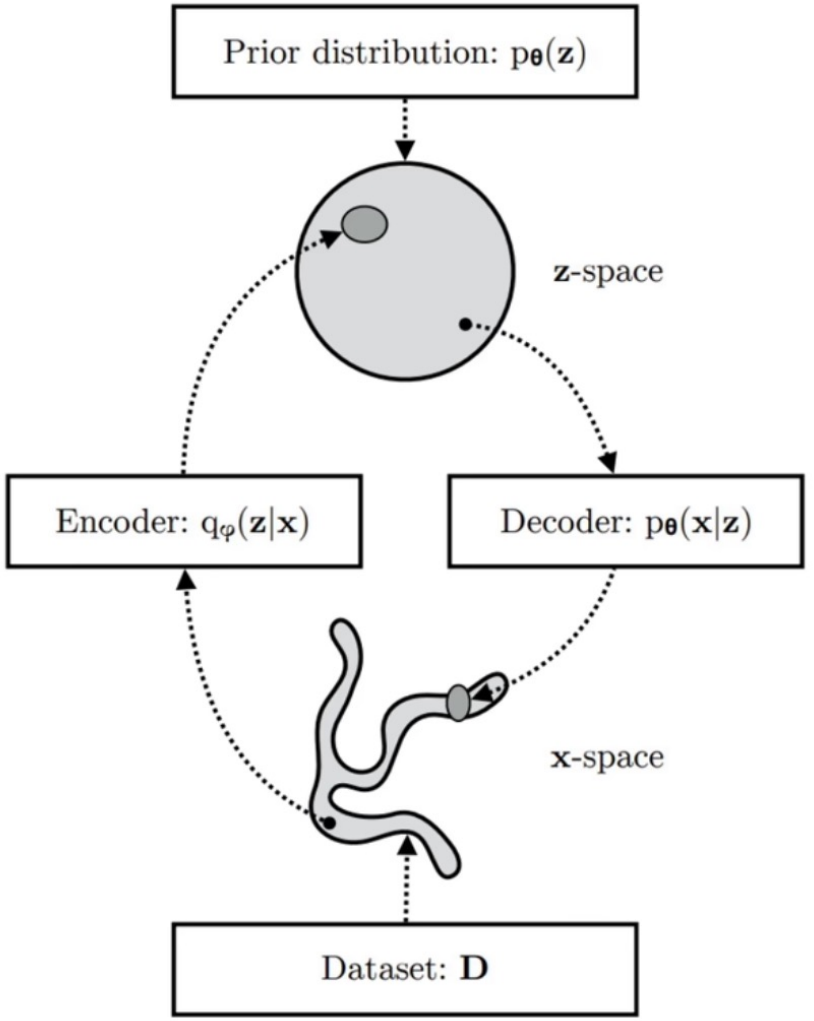


$$logP(x) = L_b + KL\big(q(z|x)||P(z|x)\big)$$

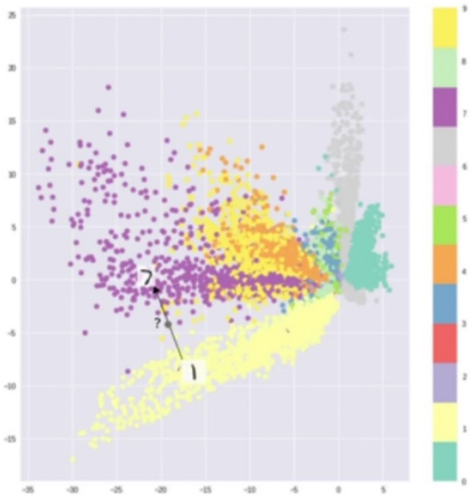
$$L_b = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz$$

$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z)}{q(z|x)}\right) dz}_{-KL\big(q(z|x)||P(z)\big)} + \underbrace{\int_z q(z|x) logP(x|z) dz}_{E_{q(z|x)}[logP(x|z)]}$$

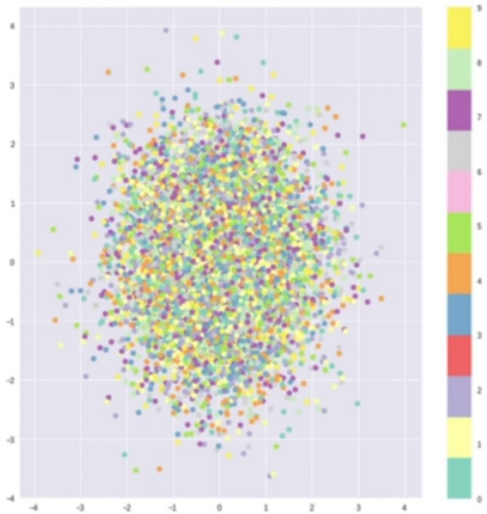
$$= \underbrace{E_{q_\phi(z|x)}\big[\log\big(p_\theta(x|z)\big)\big]}_{\text{Reconstruction Loss}} - \underbrace{D_{KL}(q_\phi(z|x)||p_\theta(z))}_{\text{Regularization Loss}}$$



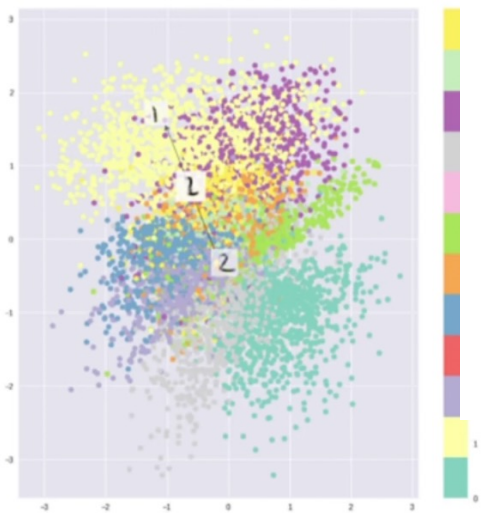
Only reconstruction loss



Only KL divergence

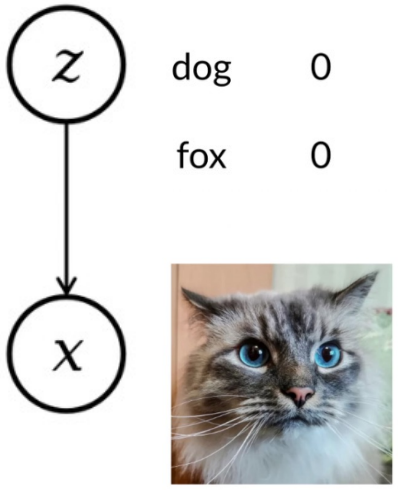


Combination

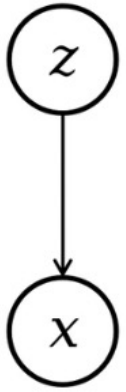


VAE -Variational Autoencoder -变分自编码器

变分推断



cat	1
dog	0
fox	0



$$z \backsim p(z) = \begin{cases} e^{-z}, z \geq 0 \\ 0 \quad , z < 0 \end{cases} = e^{-z} I(z \geq 0)$$

$$x \backsim p(x|z) = N(x, \mu = z, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}(x-z)^2)}$$

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x|z)p(z)}{\int_z p(x,z)\,dz}$$

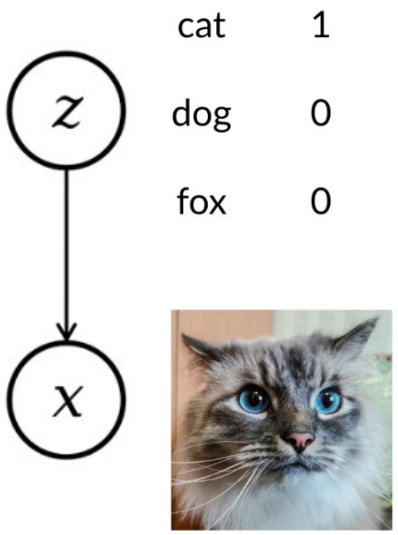
$$q_\theta(z) \approx p(z|x)$$

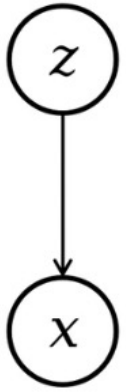
$$p(x) = \int_0^\infty p(x,z)\,dz = \int_0^\infty e^{-z} \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}(x-z)^2)}\,dz$$

the integral has no closed-form solution

thus the posterior has no closed-form solution

VAE -Variational Autoencoder -变分自编码器





$$z \backsim p(z) = \begin{cases} e^{-z}, z \geq 0 \\ 0 \quad , z < 0 \end{cases} = e^{-z} I(z \geq 0)$$

$$x \backsim p(x|z) = N(x, \mu = z, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}(x-z)^2)}$$

$$q_\theta(z) \approx p(z|x)$$

$$D(q_\theta(z)||p(z|x)) = E_{z\sim q}[\log \frac{q_\theta(z)}{p(z|x)}] = E_{z\sim q}[\log q_\theta(z) - \log p(z|x)]$$

$$= E_{z\sim q}[\log q_\theta(z) - \log \frac{p(z,x)}{p(x)}]$$

$$= E_{z\sim q}[\log q_\theta(z) - \log p(z,x)] + \log p(x)$$

$$\log p(x) = E_{z\sim q}[\log p(z,x) - \log q_\theta(z)] + D(q_\theta(z)||p(z|x))$$

$$\log p(x) \geq E_{z\sim q}[\log p(z,x) - \log q_\theta(z)] \equiv \mathcal{L}_q$$

$$q_\theta(z) \approx p(z|x)$$

$$D(q_\theta(z)\|p(z|x)) = E_{z\sim q}[\log\frac{q_\theta(z)}{p(z|x)}] = E_{z\sim q}[\log q_\theta(z) - \log p(z|x)]$$

$$= E_{z\sim q}[\log q_\theta(z) - \log\frac{p(z,x)}{p(x)}]$$

$$= E_{z\sim q}[\log q_\theta(z) - \log p(z,x)] + \log p(x)$$

$$\log p(x) = E_{z\sim q}[\log p(z,x) - \log q_\theta(z)] + D(q_\theta(z)\|p(z|x))$$

$$\log p(x) \geq E_{z\sim q}[\log p(z,x) - \log q_\theta(z)] \equiv \mathcal{L}_q$$

$$q_\theta(z) \approx p(z|x)$$

$$q_\theta(z) = \begin{cases} \theta e^{-\theta z}, z \geq 0 \\ 0 \quad , z < 0 \end{cases} = \theta e^{-\theta z} I(z \geq 0)$$

$$z \backsim p(z) = \begin{cases} e^{-z}, z \geq 0 \\ 0 \quad , z < 0 \end{cases} = e^{-z} I(z \geq 0)$$

$$x \backsim p(x|z) = N(x, \mu = z, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}(x-z)^2)}$$

$$\mathcal{L}_q = E_{z\sim q}[\log p(z,x) - \log q_\theta(z)]$$

$$= E_{z\sim q}[\log\frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}(x-z)^2)} e^{-z} I(z \geq 0) - \log\theta e^{-\theta z} I(z \geq 0)]$$

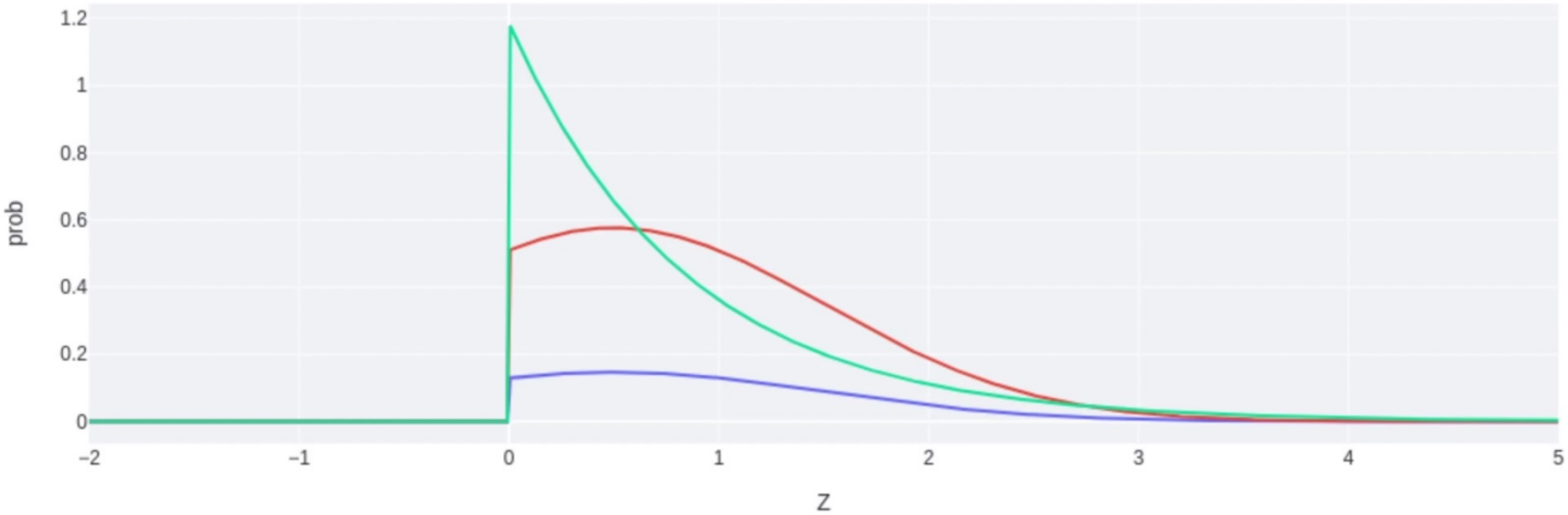
$$= E_{z\sim q}[-\frac{1}{2}z^2 + (x - 1 + \theta)z + C] = -\frac{1}{\theta^2} + \frac{x - 1 + \theta}{\theta} - \log\theta + C$$

$$\frac{\partial \mathcal{L}_q}{\partial \theta} = \frac{2}{\theta^3} - \frac{0.5}{\theta^2} - \frac{1}{\theta} = 0$$

↑

expectation of exponential distribution $E_\theta[z^n] = \frac{n!}{\theta^n}$

$x = 1.5$



Joint $p(x|z)p(z)$

'True Posterior' $p(z|x)$

Surrogate Posterior $q_\theta(z)$

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x|z)p(z)}{\int_z p(x,z)\,dz}$$

$$q_\theta(z) \approx p(z|x)$$

$$q_\theta(z) = \begin{cases} \theta e^{-\theta z}, z \geq 0 \\ 0 \quad , z < 0 \end{cases} = \theta e^{-\theta z} I(z \geq 0)$$

$$z \backsim p(z) = \begin{cases} e^{-z}, z \geq 0 \\ 0 \quad , z < 0 \end{cases} = e^{-z} I(z \geq 0)$$

$$x \backsim p(x|z) = N(x, \mu = z, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}(x-z)^2)}$$

$$\mathcal{L}_q = E_{z\sim q}[\log p(z,x) - \log q_\theta(z)]$$

$$= E_{z\sim q}[\log \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}(x-z)^2)} e^{-z} I(z \geq 0) - \log \theta e^{-\theta z} I(z \geq 0)]$$

$$= E_{z\sim q}[-\frac{1}{2}z^2 + (x-1+\theta)z + C] = -\frac{1}{\theta^2} + \frac{x-1+\theta}{\theta} - \log\theta + C$$

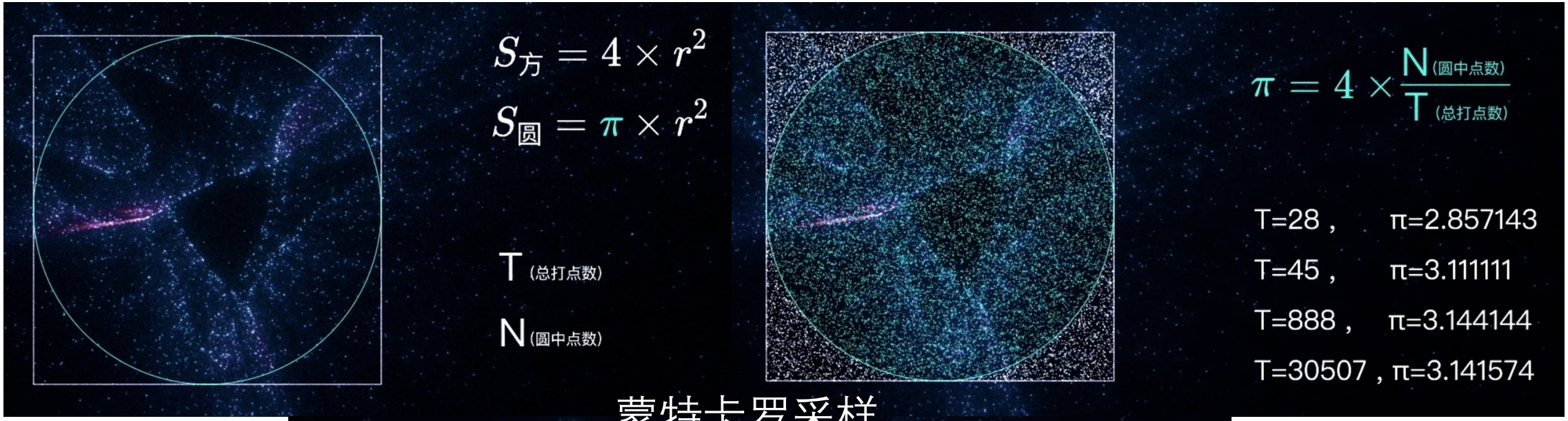
$$\frac{\partial \mathcal{L}_q}{\partial \theta} = \frac{2}{\theta^3} - \frac{0.5}{\theta^2} - \frac{1}{\theta} = 0$$

↑

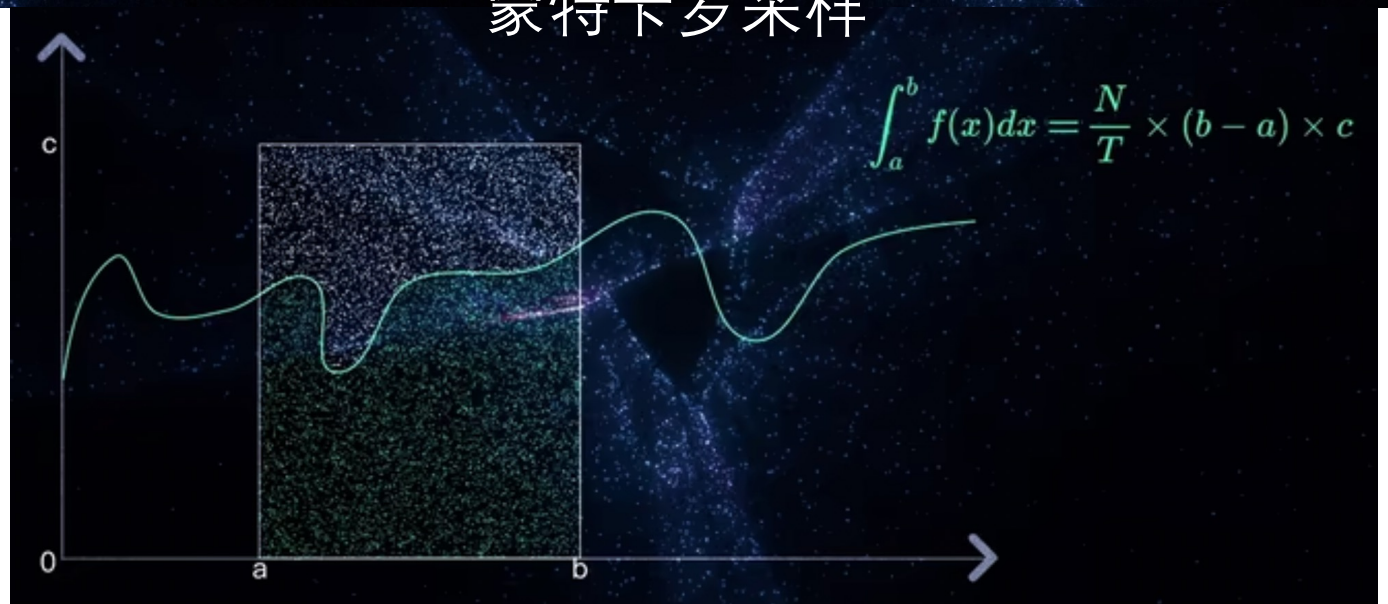
expectation of exponential distribution $E_\theta[z^n] = \frac{n!}{\theta^n}$

$x = 1.5$

VAE -Variational Autoencoder -变分自编码器



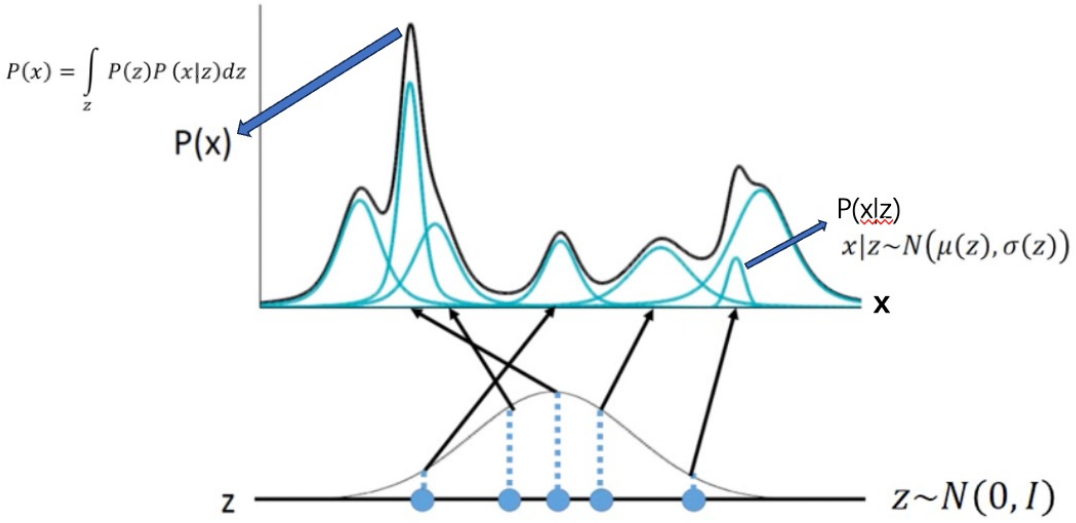
蒙特卡罗采样



VAE -Variational Autoencoder -变分自编码器

P(z)

P(x|z)

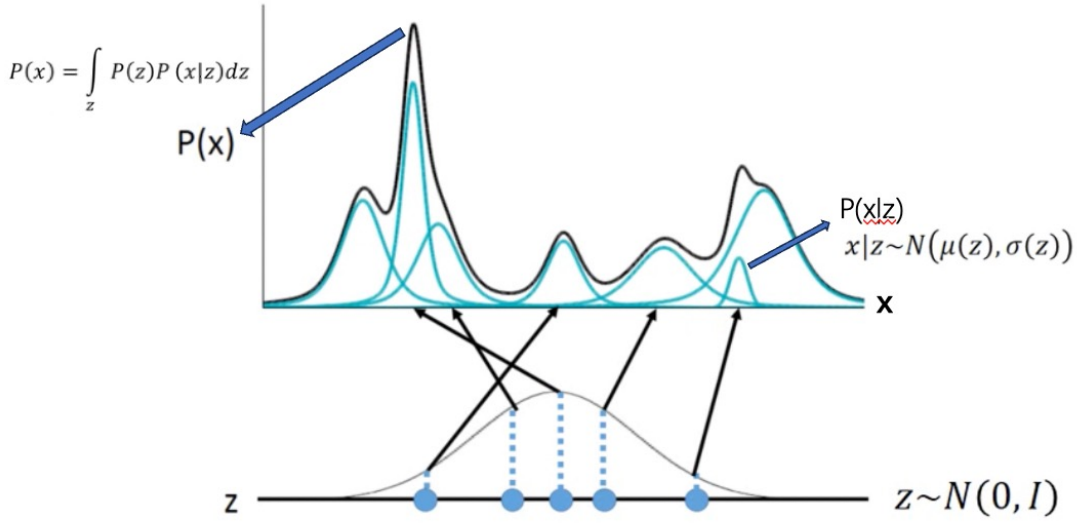


$$p(z) = \mathcal{N}(0, I) \qquad p_\theta(X \mid z) = \mathcal{N}(X \mid \mu_i'(z;\theta), \sigma_i'^2(z;\theta) * I)$$

$$p_\theta(X) = \int_z p_\theta(X \mid z)p(z)dz \approx \frac{1}{m}\sum_{j=1}^{m} p_\theta(X \mid z_j)$$

1) 从 $p(z)$ 中多次采样 $z_1, z_2, \ldots z_m$
2) 根据 $p(x|z;\theta)$ 计算 $x_1, x_2, \ldots, x_m$
3) 求 x 的均值 $\frac{1}{m}\sum_{j=1}^{m} x_j$

VAE -Variational Autoencoder -变分自编码器

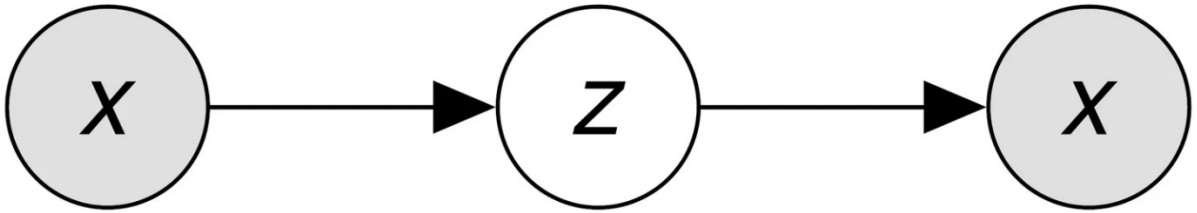


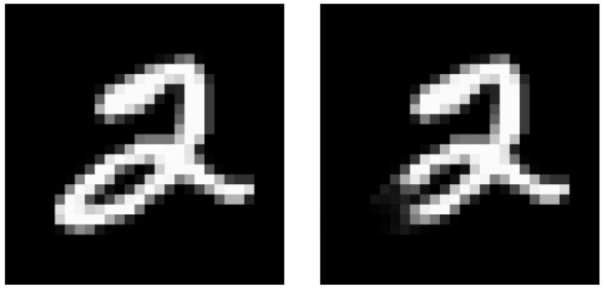
$$p(z) = \mathcal{N}(0, I) \qquad p_\theta(X \mid z) = \mathcal{N}(X \mid \mu_i'(z;\theta), \sigma_i'^2(z;\theta) * I)$$

$$p_\theta(X) = \int_z p_\theta(X \mid z)p(z)dz$$

$$\approx \frac{1}{m}\sum_{j=1}^{m} p_\theta(X \mid z_j)$$

$q_\theta(z \mid X)$ ➡ $p_\theta(z \mid X)$ $\qquad p_\theta(X \mid z)$





$$p_\theta(z \mid x_i) = \frac{p_\theta(x_i \mid z)p(z)}{p_\theta(x_i)}$$

$$= \frac{p_\theta(x_i \mid z)p(z)}{\int_{\hat{z}} p_\theta(x_i \mid \hat{z})p(\hat{z})d\hat{z}}$$

VAE -Variational Autoencoder -变分自编码器

$$L = \sum_x logP(x)$$

$$logP(x) = \int_z q(z|x) logP(x) dz$$

q(z|x) can be any distribution

$$= \int_z q(z|x) log\left(\frac{P(z,x)}{P(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\frac{q(z|x)}{P(z|x)}\right) dz$$

$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz}_{L_b} + \underbrace{\int_z q(z|x) log\left(\frac{q(z|x)}{P(z|x)}\right) dz}_{KL(q(z|x)||P(z|x))} \geq \underbrace{\int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz}_{L_b}$$

$$logP(x) = L_b + KL(q(z|x)||P(z|x))$$

$$L_b = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz$$

$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z)}{q(z|x)}\right) dz}_{-KL(q(z|x)||P(z))} + \underbrace{\int_z q(z|x) logP(x|z) dz}_{E_{q(z|x)}[logP(x|z)]}$$

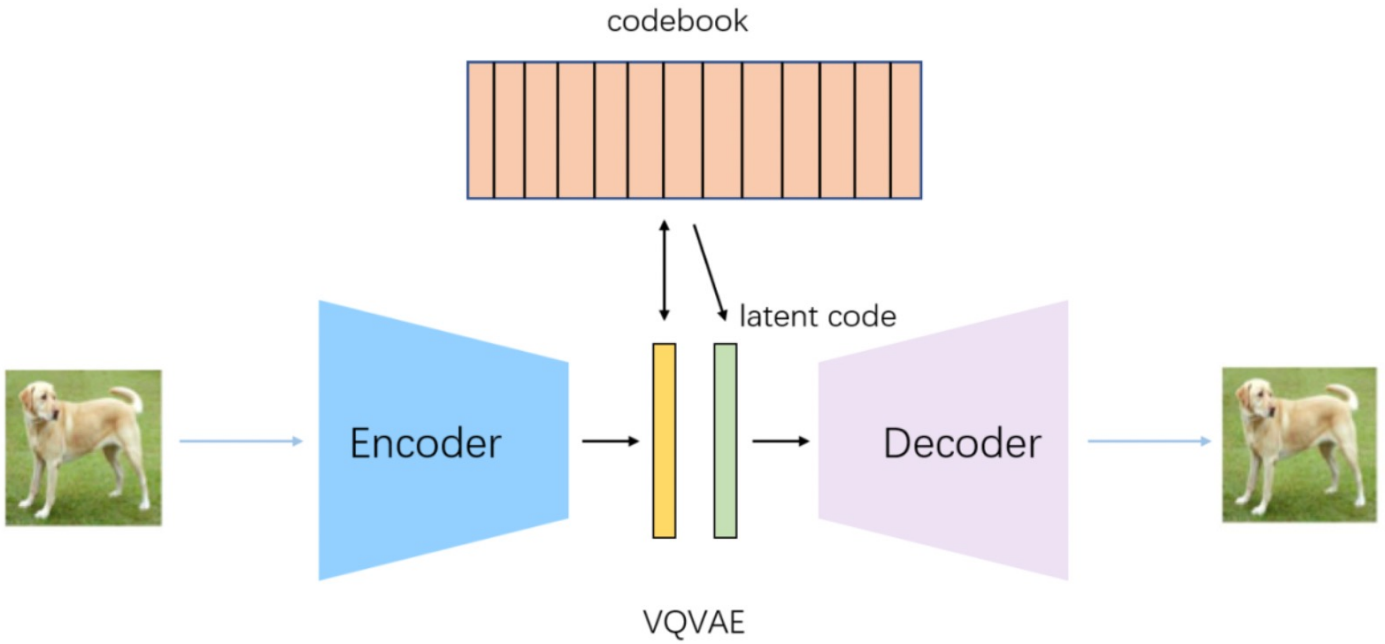
$$\begin{aligned}
&KL(q_\theta(z|x)||p(z|x)) \\
&= \int q_\theta(z|x) \ln \frac{q_\theta(z|x)}{p(z|x)} \mathrm{d}z \\
&= \mathbb{E}_{z\sim q_\theta(z|x)}[\ln \frac{q_\theta(z|x)}{p(z|x)}] \\
&= \mathbb{E}_{z\sim q_\theta(z|x)}[\ln q_\theta(z|x) - \ln p(z|x)] \\
&= \mathbb{E}_{z\sim q_\theta(z|x)}[\ln q_\theta(z|x) - \ln \frac{p(x|z)p(z)}{p(x)}] \\
&= \mathbb{E}_{z\sim q_\theta(z|x)}[\ln q_\theta(z|x) - \ln p(z) - \ln p(x|z)] + \ln p(x) \\
&= KL(q_\theta(z|x)||p(z)) - \mathbb{E}_{z\sim q_\theta(z|x)}[\ln p(x|z)] + \ln p(x)
\end{aligned}$$

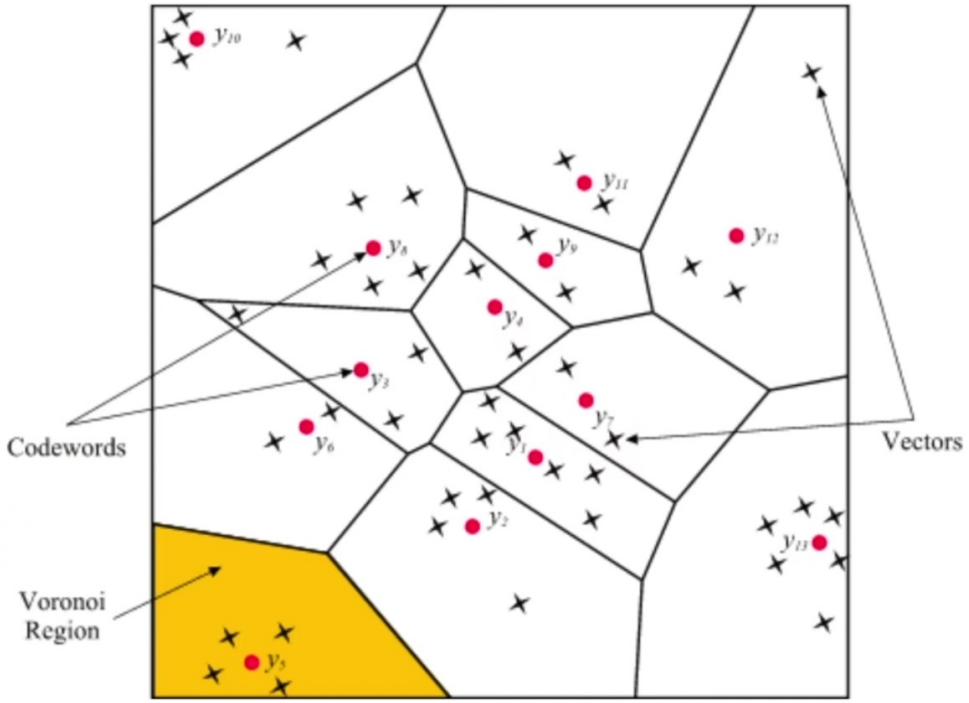
$$\ln p(x) - KL(q_\theta(z|x)||p(z|x)) = \mathbb{E}_{z\sim q_\theta(z|x)}[\ln p(x|z)] - KL(q_\theta(z|x)||p(z))$$

VQ-VAE -Vector Quantised Variational AutoEncoder –向量量化变分自动编码器

Posterior Collapse

Have a powerful decoder, but latents are ignored.



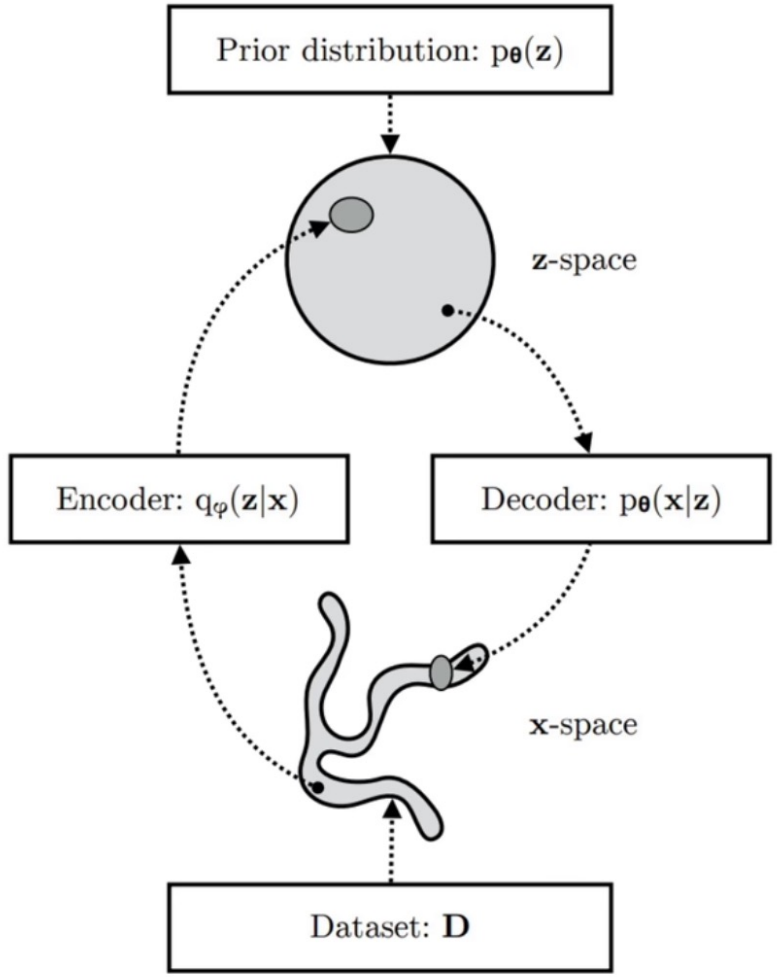


$$logP(x) = L_b + KL\big(q(z|x)||P(z|x)\big)$$

$$L_b = \int_z q(z|x) log\left(\frac{P(z,x)}{q(z|x)}\right) dz = \int_z q(z|x) log\left(\frac{P(x|z)P(z)}{q(z|x)}\right) dz$$

$$= \underbrace{\int_z q(z|x) log\left(\frac{P(z)}{q(z|x)}\right) dz}_{-KL\big(q(z|x)||P(z)\big)} + \underbrace{\int_z q(z|x) logP(x|z) dz}_{E_{q(z|x)}[logP(x|z)]}$$

$$= \underbrace{E_{q_\phi(z|x)}\big[\log\big(p_\theta(x|z)\big)\big]}_{\text{Reconstruction Loss}} - \underbrace{D_{KL}(q_\phi(z|x)||p_\theta(z))}_{\text{Regularization Loss}}$$



$$q_\phi(z|x) \simeq q_\phi(z) = \mathcal{N}(a, b)$$

$$\mathcal{L} = \mathcal{L}_1 + \alpha\mathcal{L}_2$$

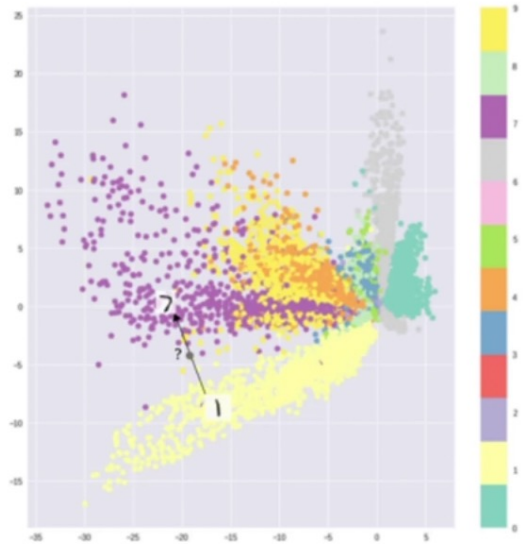
$$logP(x) = L_b + KL\big(q(z|x)||P(z|x)\big)$$

$$= \underbrace{E_{q_\phi(z|x)}\big[\log\big(p_\theta(x|z)\big)\big]}_{\text{Reconstruction Loss}} - \underbrace{D_{KL}(q_\phi(z|x)||p_\theta(z))}_{\text{Regularization Loss}}$$

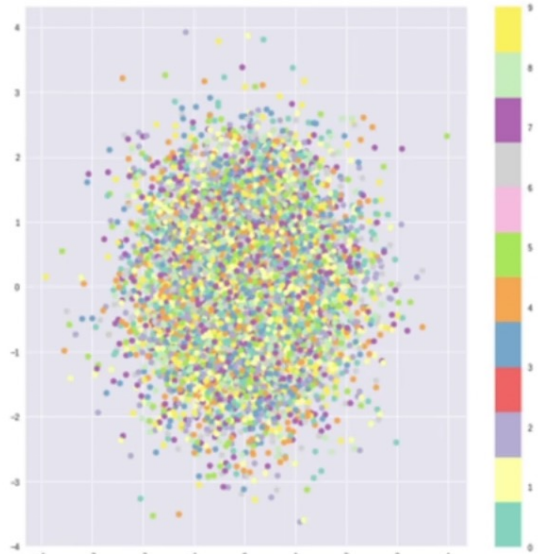
$$\mathcal{C} = \mathcal{C}_1 + \alpha\mathcal{C}_2$$

$$q_\phi(z|x) \simeq q_\phi(z) = \mathcal{N}(a, b)$$

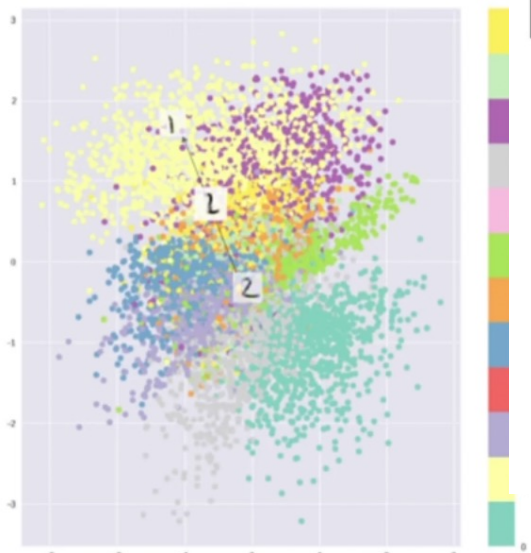
Only reconstruction loss

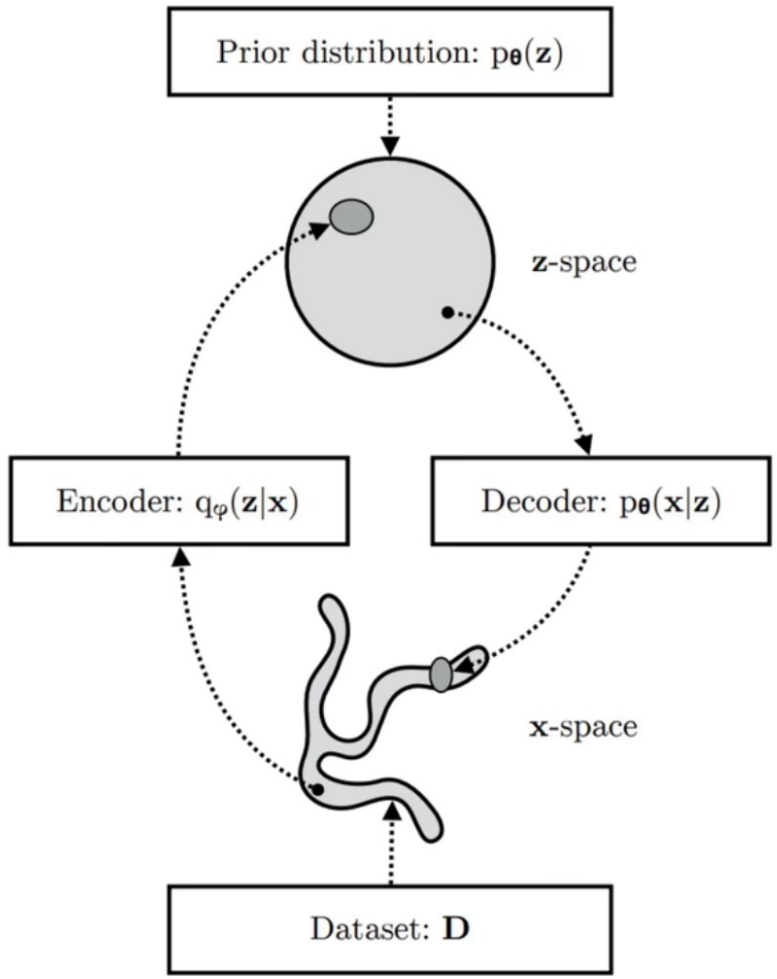


Only KL divergence

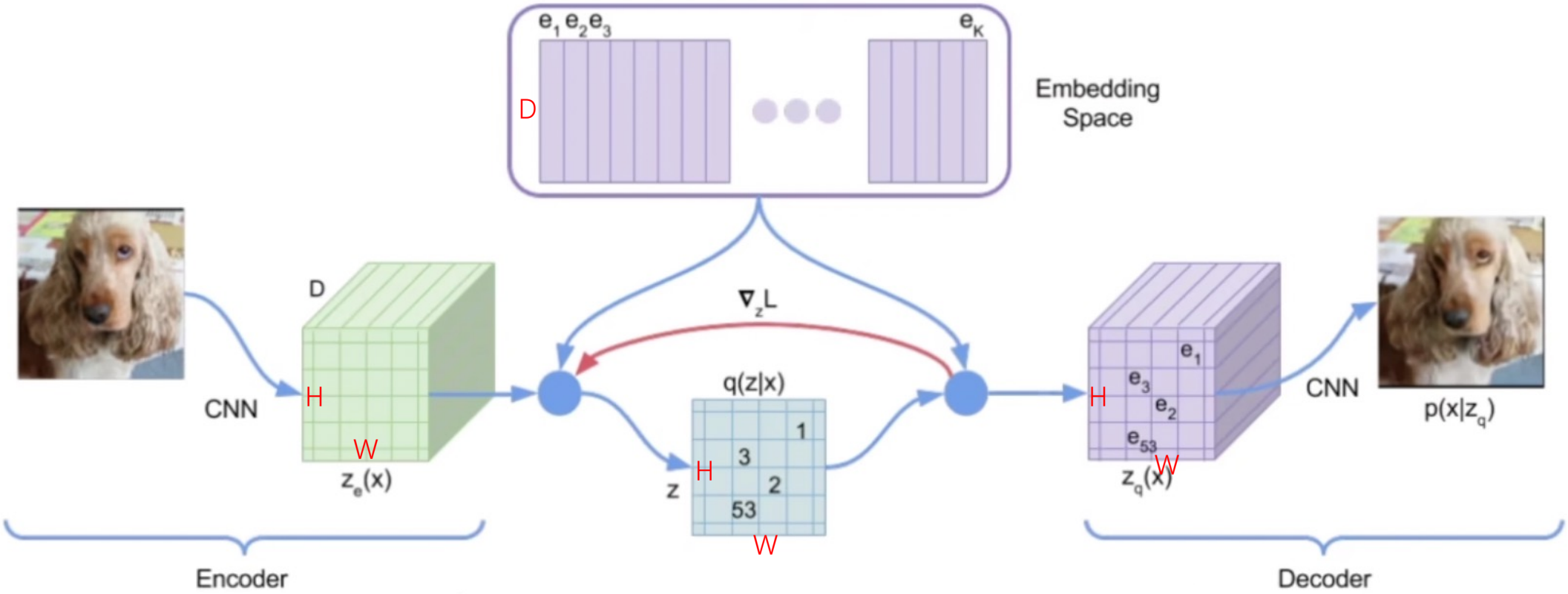


Combination





VQ-VAE -Vector Quantised Variational AutoEncoder –向量量化变分自动编码器



$$z_q(x) = e_k, \quad \text{where} \quad k = \text{argmin}_j \|z_e(x) - e_j\|_2$$

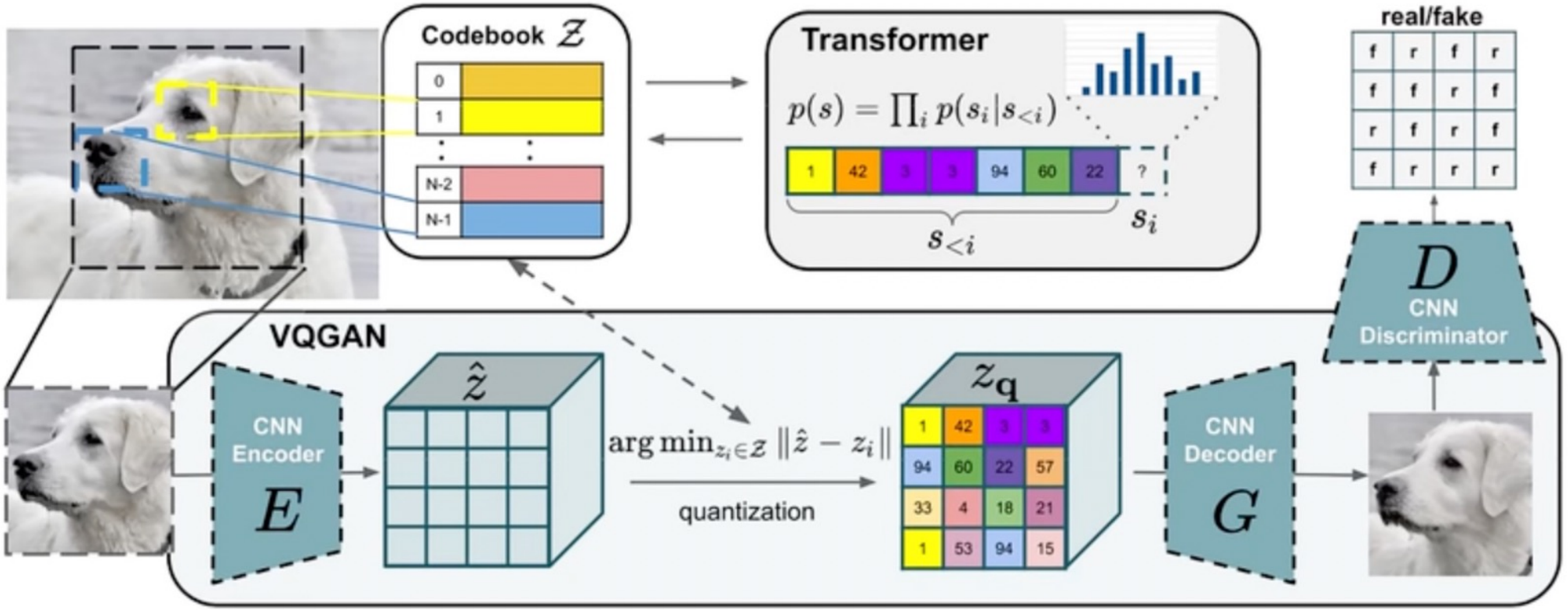
$$L = \underbrace{\log p(x|z_q(x))}_{\text{reconstruction loss}} + \underbrace{\|\text{sg}[z_e(x)] - e\|_2^2}_{\text{VQ loss}} + \underbrace{\beta\|z_e(x) - \text{sg}[e]\|_2^2}_{\text{commitment loss}}$$

For enc & dec

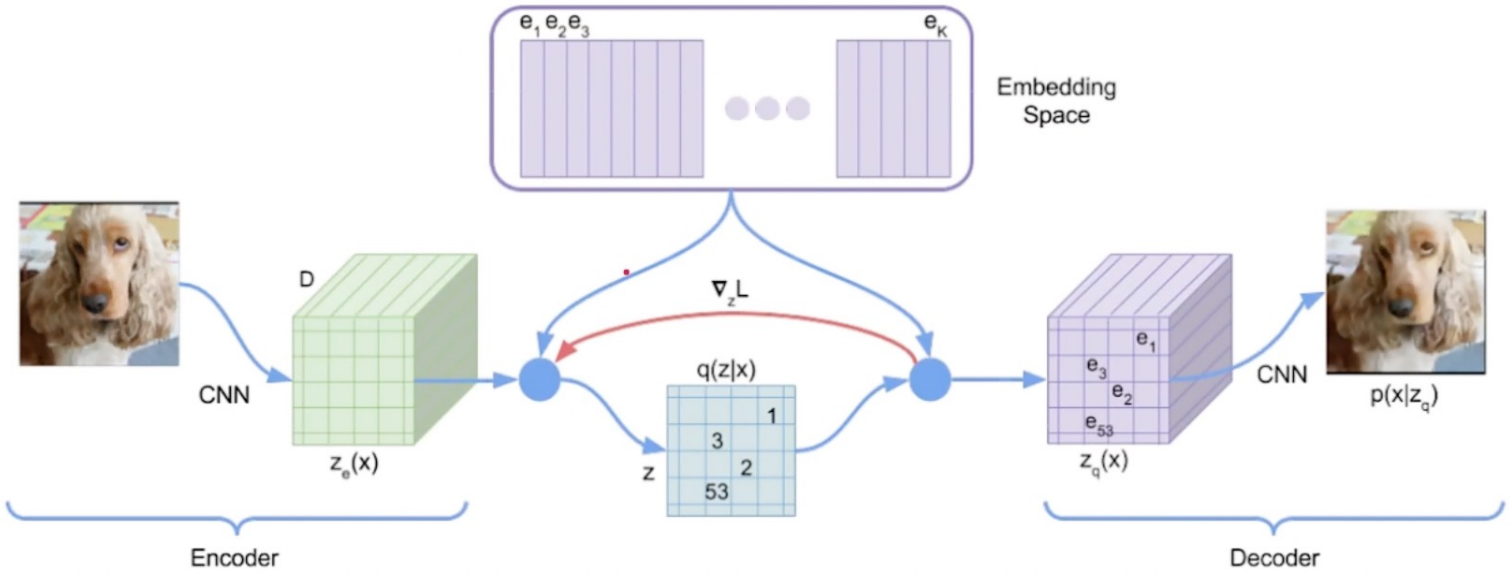
For e

For enc

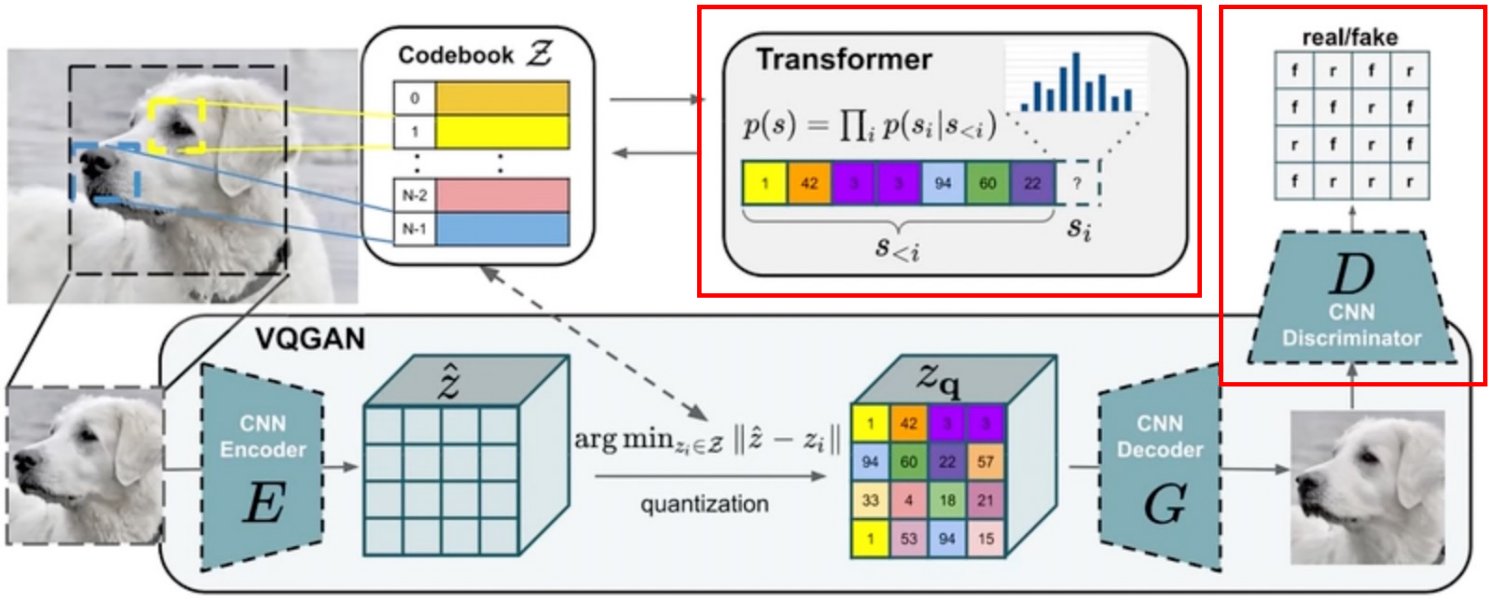
VQ-GAN -Vector Quantized Generative Adversarial Network -向量量化生成对抗网络



VQ-GAN -Vector Quantized Generative Adversarial Network -向量量化生成对抗网络

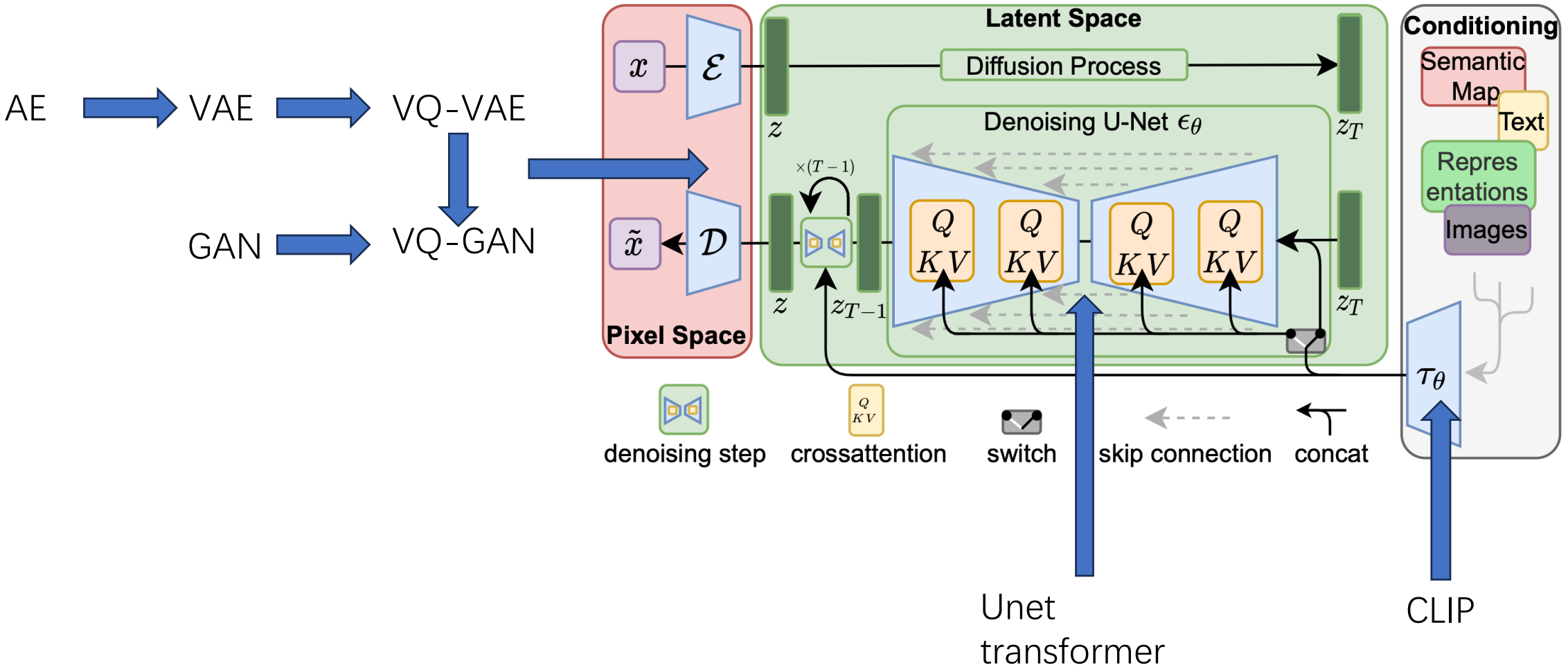


$$\mathcal{L}_{\text{VQ}}(E, G, \mathcal{Z}) = \|x - \hat{x}\|^2 + \|\text{sg}[E(x)] - z_{\mathbf{q}}\|_2^2 + \|\text{sg}[z_{\mathbf{q}}] - E(x)\|_2^2.$$

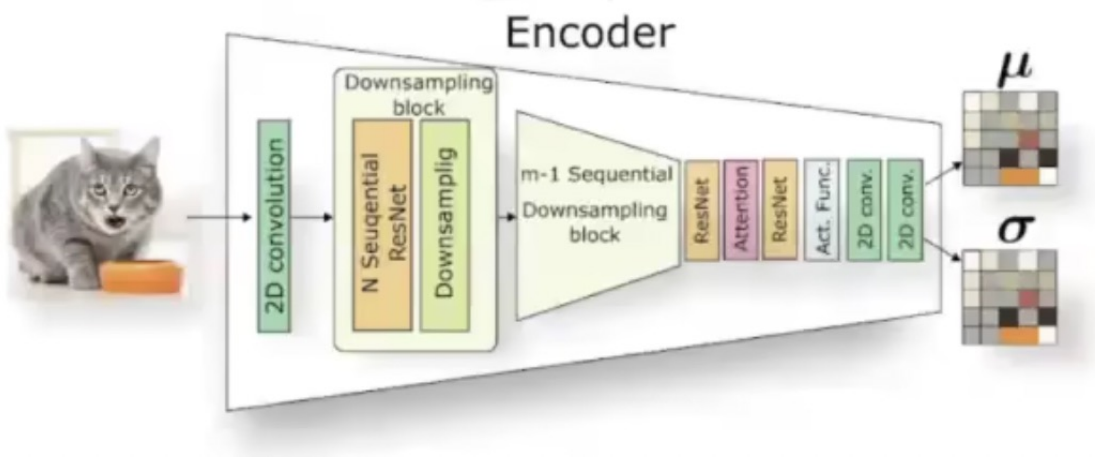


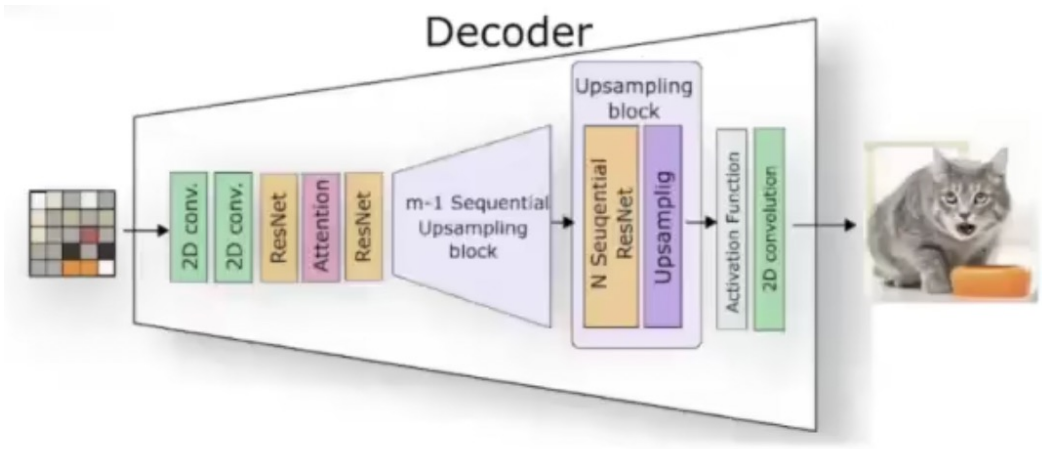
$$\mathcal{Q}^* = \arg\min_{E,G,\mathcal{Z}} \max_D \mathbb{E}_{x\sim p(x)} \Big[\mathcal{L}_{\text{VQ}}(E, G, \mathcal{Z}) + \lambda \mathcal{L}_{\text{GAN}}(\{E, G, \mathcal{Z}\}, D)\Big]$$

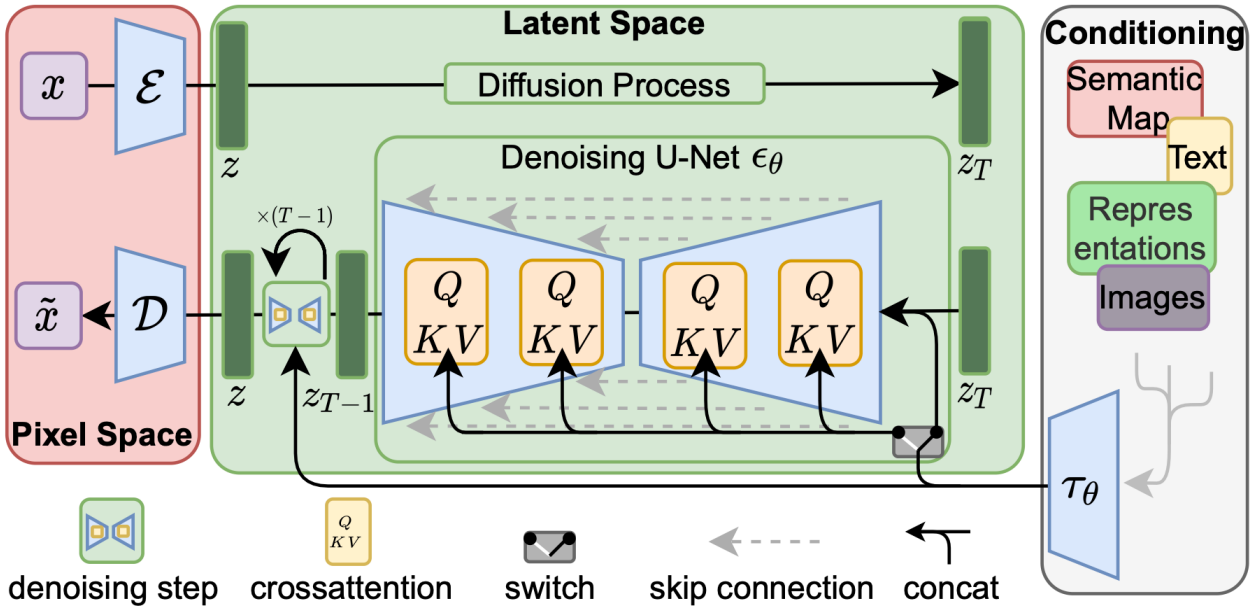
LDM -Latent Diffusion Models -潜在扩散模型



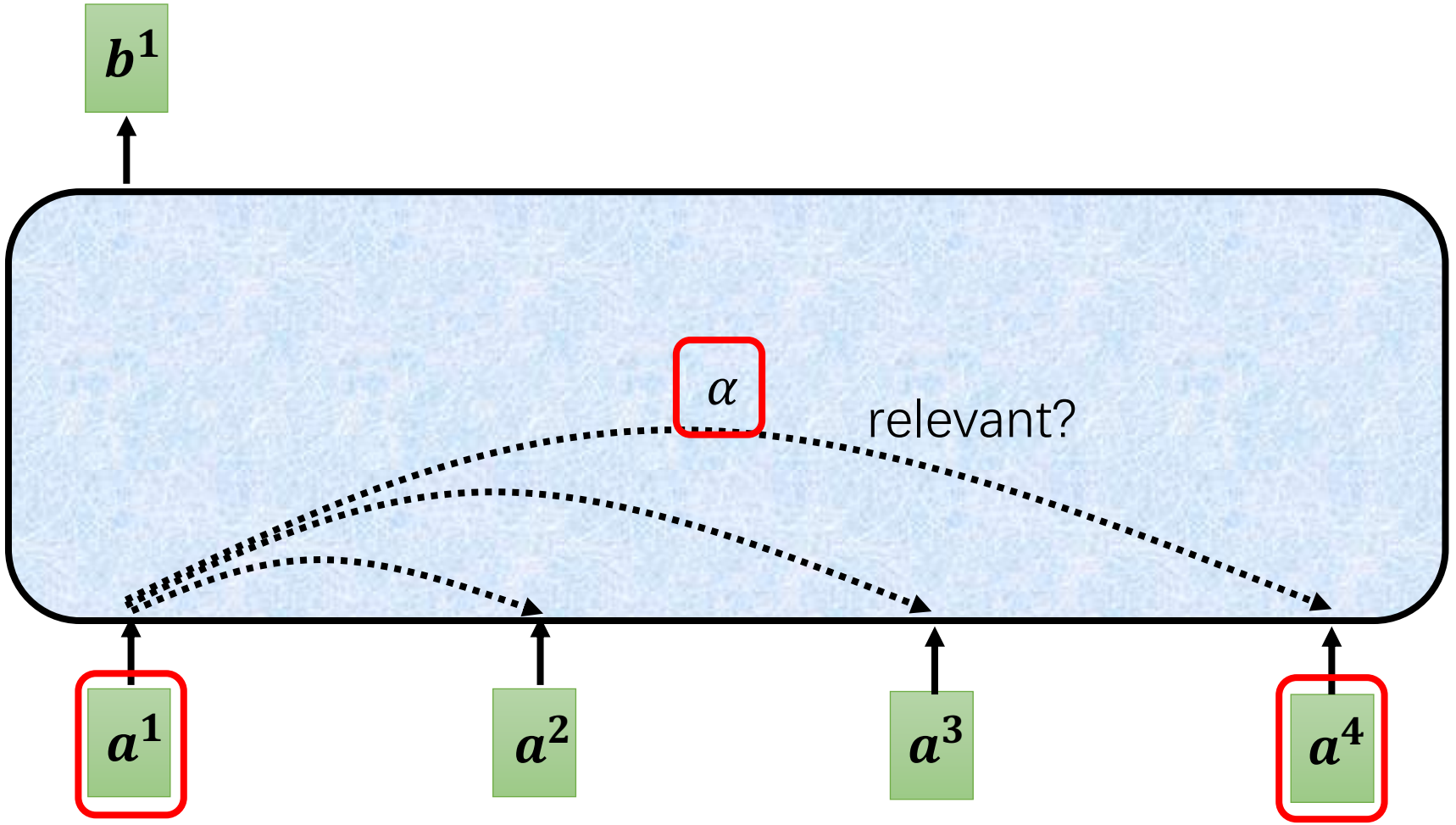
LDM -Latent Diffusion Models -潜在扩散模型





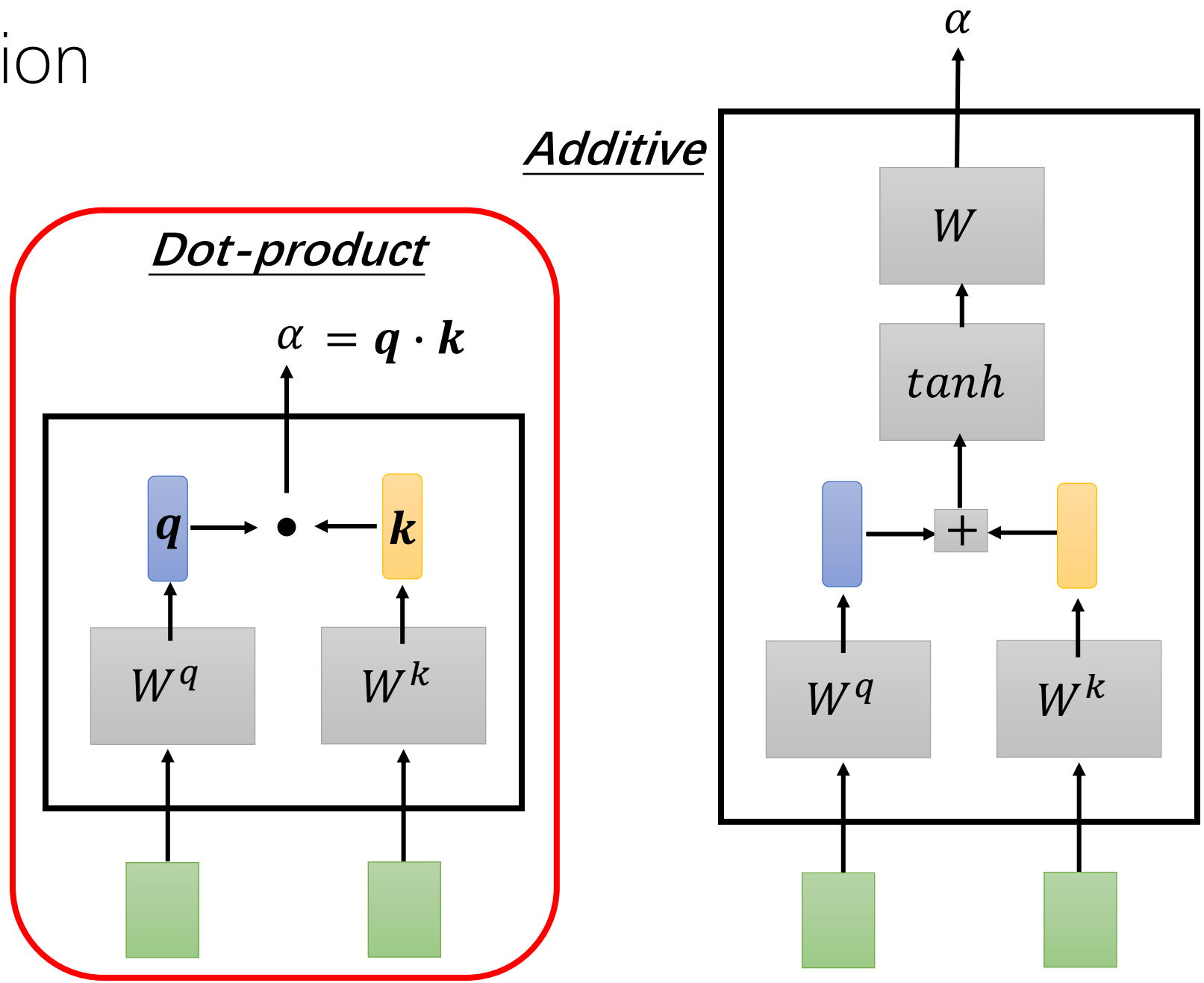


Self-attention



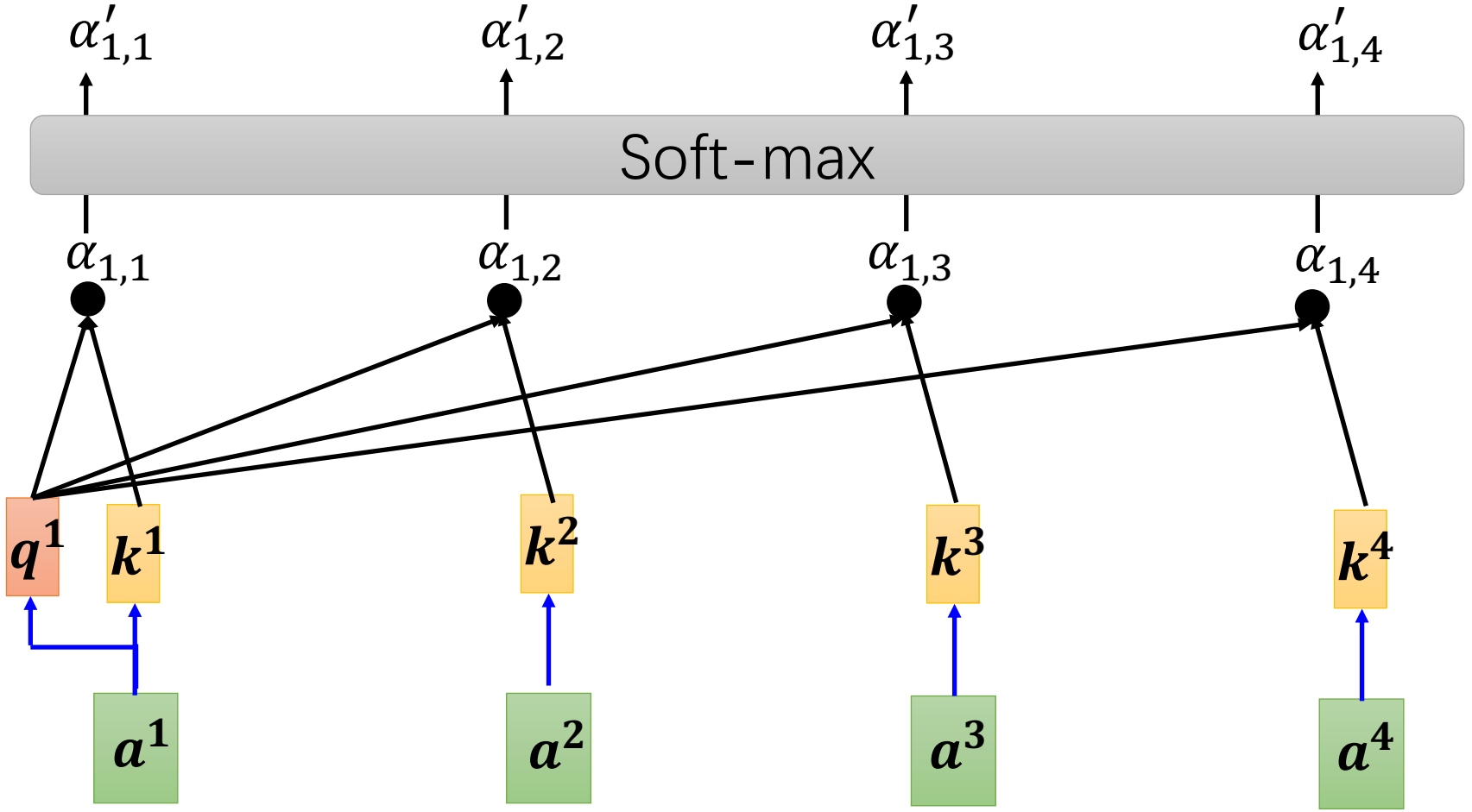
Find the relevant vectors in a sequence

Self-attention



Self-attention

$$\alpha'_{1,i} = exp(\alpha_{1,i}) / \sum_j exp(\alpha_{1,j})$$



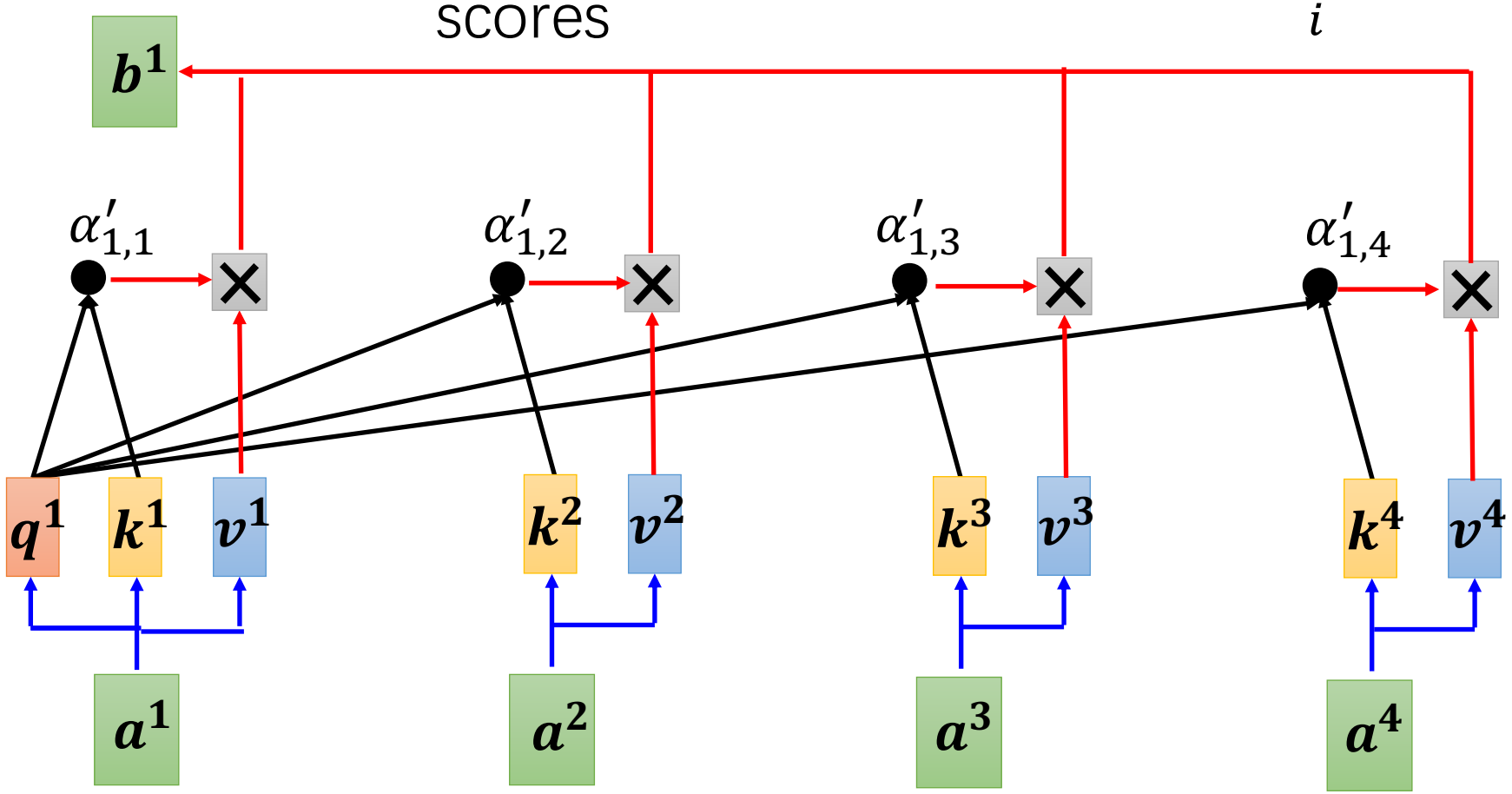
$q^1 = W^q a^1$ $k^2 = W^k a^2$ $k^3 = W^k a^3$ $k^4 = W^k a^4$

$k^1 = W^k a^1$

Self-attention

Extract information based on attention scores

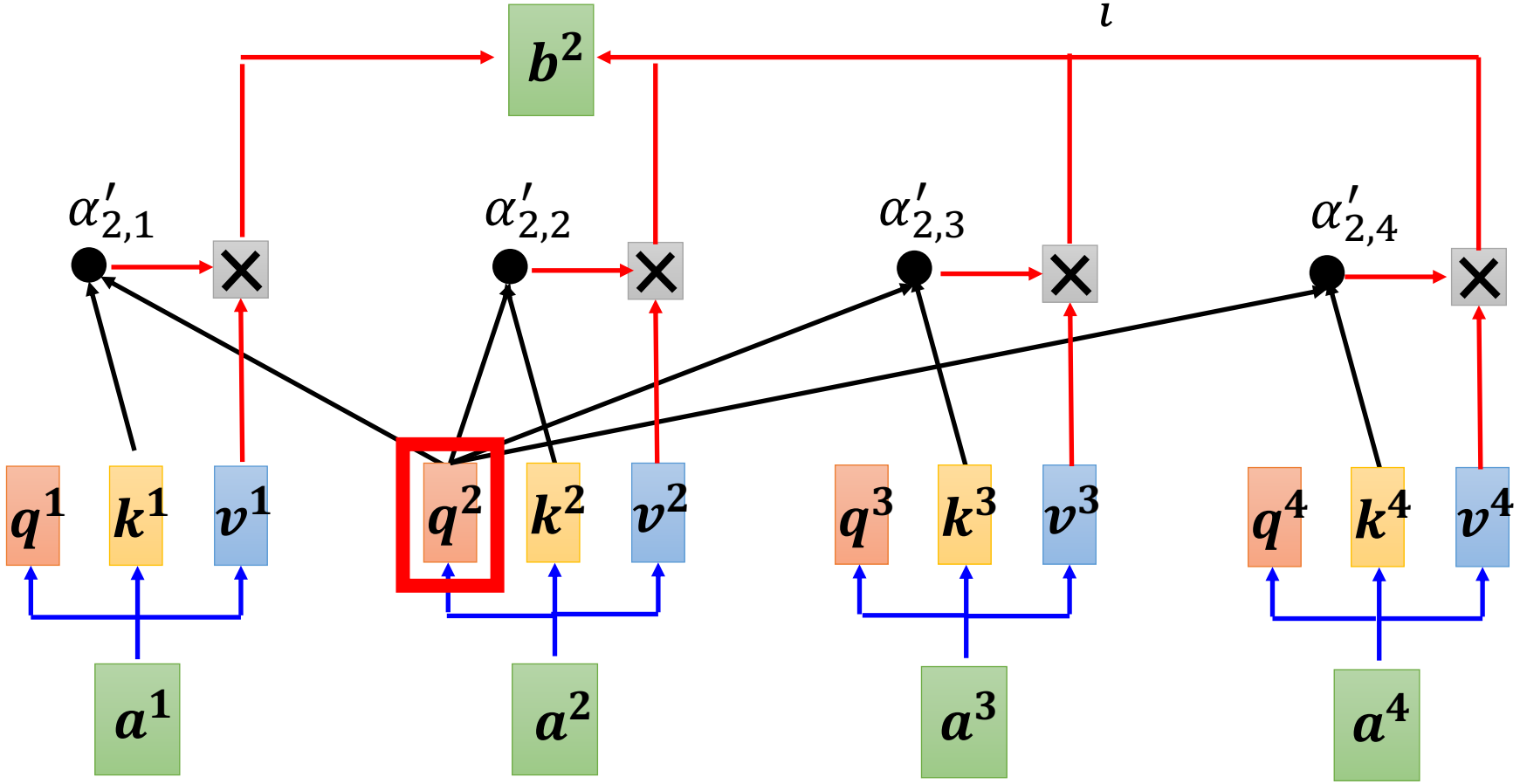
$$\boldsymbol{b^1} = \sum_i \alpha'_{1,i} \boldsymbol{v^i}$$



$$\boldsymbol{v^1} = W^v \boldsymbol{a^1} \qquad \boldsymbol{v^2} = W^v \boldsymbol{a^2} \qquad \boldsymbol{v^3} = W^v \boldsymbol{a^3} \qquad \boldsymbol{v^4} = W^v \boldsymbol{a^4}$$

Self-attention

$$b^2 = \sum_i \alpha'_{2,i} v^i$$

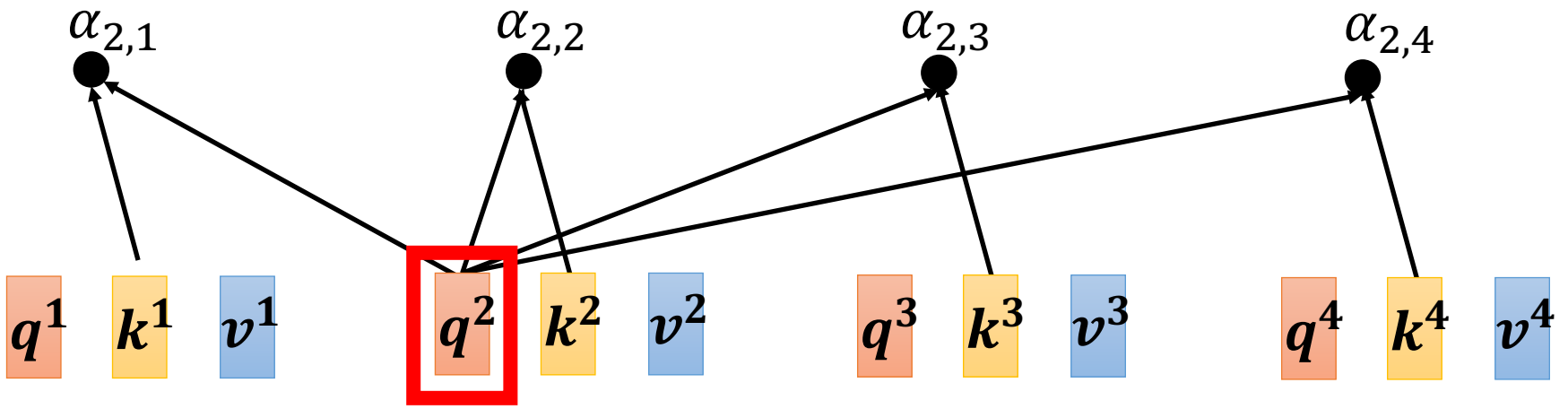


Self-attention

$\alpha_{1,1} = \boldsymbol{k^1}\ \boldsymbol{q^1}$ $\alpha_{1,2} = \boldsymbol{k^2}\ \boldsymbol{q^1}$

$\alpha_{1,3} = \boldsymbol{k^3}\ \boldsymbol{q^1}$ $\alpha_{1,4} = \boldsymbol{k^4}\ \boldsymbol{q^1}$

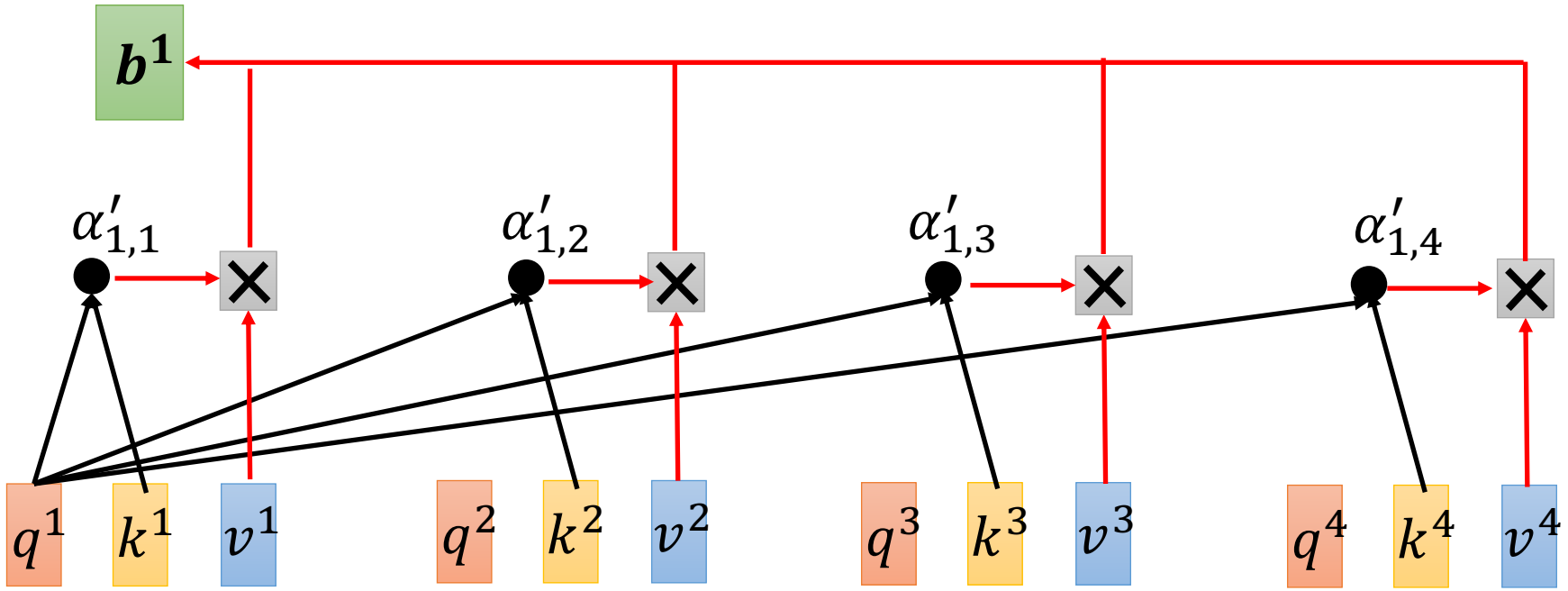
$$\begin{bmatrix}\alpha_{1,1}\\ \alpha_{1,2}\\ \alpha_{1,3}\\ \alpha_{1,4}\end{bmatrix} = \begin{bmatrix}\boldsymbol{k^1}\\ \boldsymbol{k^2}\\ \boldsymbol{k^3}\\ \boldsymbol{k^4}\end{bmatrix} \boldsymbol{q^1}$$



$$\begin{bmatrix}\alpha'_{1,1} & \alpha'_{2,1} & \alpha'_{3,1} & \alpha'_{4,1}\\ \alpha'_{1,2} & \alpha'_{2,2} & \alpha'_{3,2} & \alpha'_{4,2}\\ \alpha'_{1,3} & \alpha'_{2,3} & \alpha'_{3,3} & \alpha'_{4,3}\\ \alpha'_{1,4} & \alpha'_{2,4} & \alpha'_{3,4} & \alpha'_{4,4}\end{bmatrix} = A' \quad \text{softmax} \quad \begin{bmatrix}\alpha_{1,1} & \alpha_{2,1} & \alpha_{3,1} & \alpha_{4,1}\\ \alpha_{1,2} & \alpha_{2,2} & \alpha_{3,2} & \alpha_{4,2}\\ \alpha_{1,3} & \alpha_{2,3} & \alpha_{3,3} & \alpha_{4,3}\\ \alpha_{1,4} & \alpha_{2,4} & \alpha_{3,4} & \alpha_{4,4}\end{bmatrix} = A$$

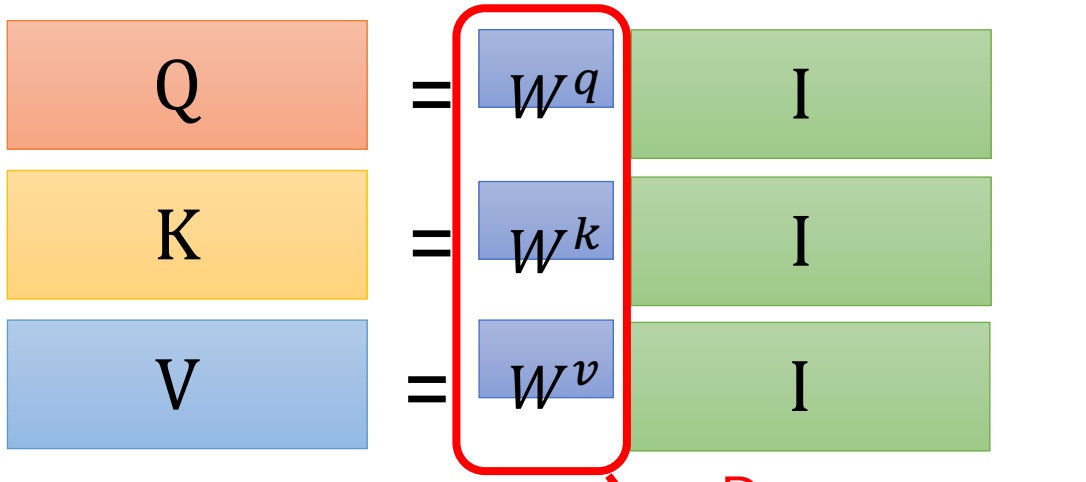
$$\begin{bmatrix}\boldsymbol{k^1}\\ \boldsymbol{k^2}\\ \boldsymbol{k^3}\\ \boldsymbol{k^4}\end{bmatrix} = K^T \qquad \begin{bmatrix}\boldsymbol{q^1} & \boldsymbol{q^2} & \boldsymbol{q^3} & \boldsymbol{q^4}\end{bmatrix} = Q$$

Self-attention

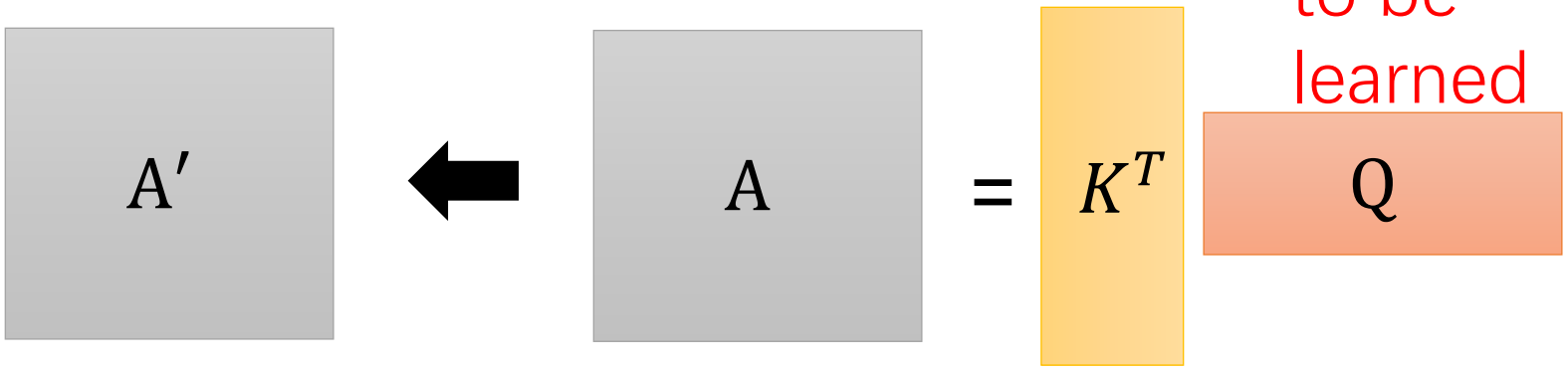


$$\underbrace{b^1\,b^2\,b^3\,b^4}_{O} = \underbrace{v^1\,v^2\,v^3\,v^4}_{V} \begin{bmatrix} \alpha'_{1,1} & \alpha'_{2,1} & \alpha'_{3,1} & \alpha'_{4,1} \\ \alpha'_{1,2} & \alpha'_{2,2} & \alpha'_{3,2} & \alpha'_{4,2} \\ \alpha'_{1,3} & \alpha'_{2,3} & \alpha'_{3,3} & \alpha'_{4,3} \\ \alpha'_{1,4} & \alpha'_{2,4} & \alpha'_{3,4} & \alpha'_{4,4} \end{bmatrix}_{A'}$$

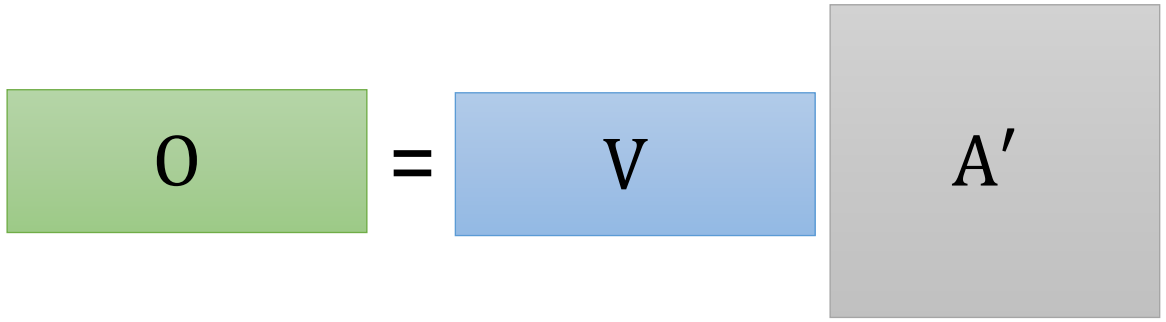
Self-attention



Parameters to be learned



Attention Matrix

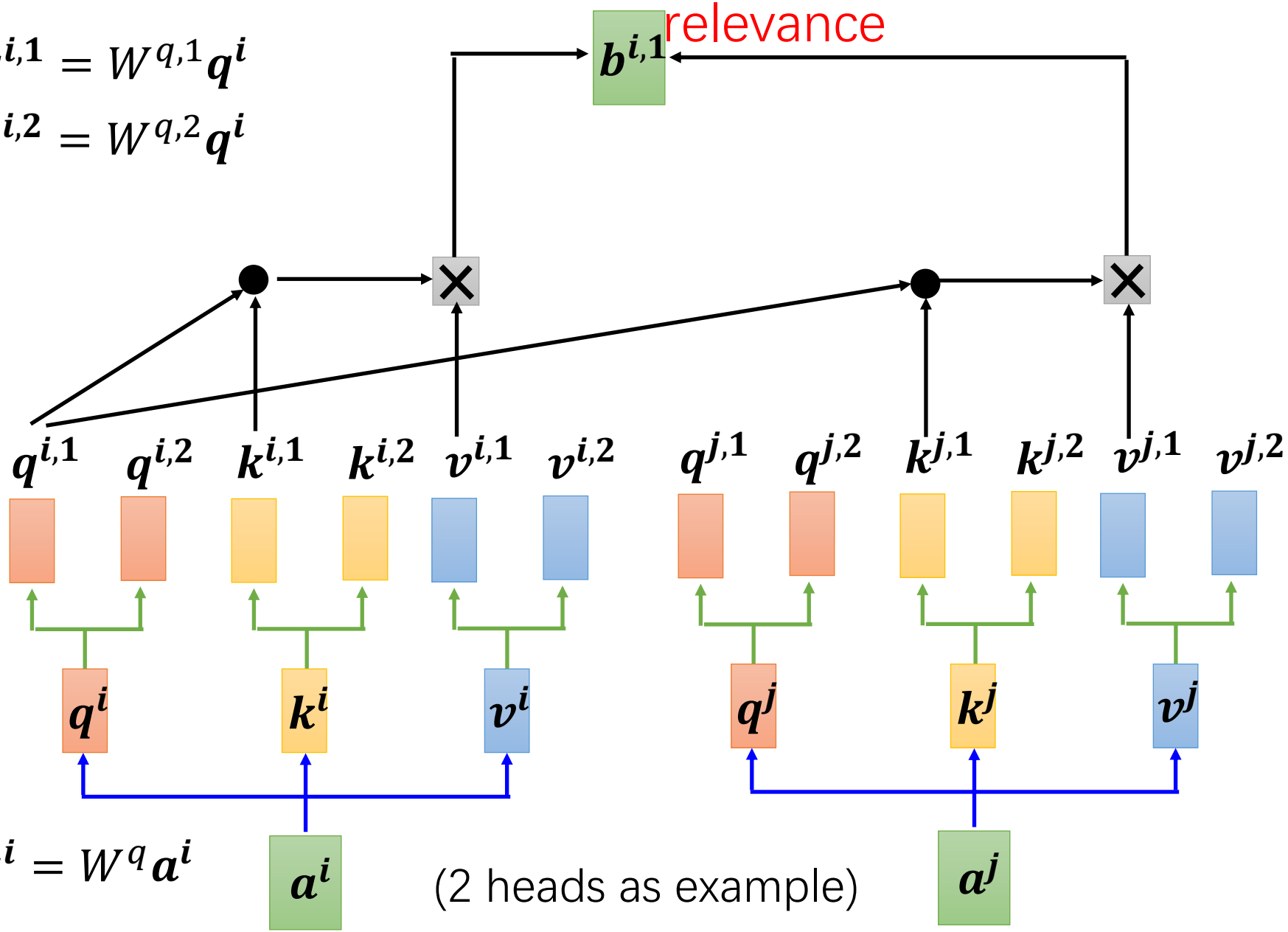


Multi-head Self-attention

Different types of relevance

$$q^{i,1} = W^{q,1} q^i$$

$$q^{i,2} = W^{q,2} q^i$$



$$q^i = W^q a^i$$

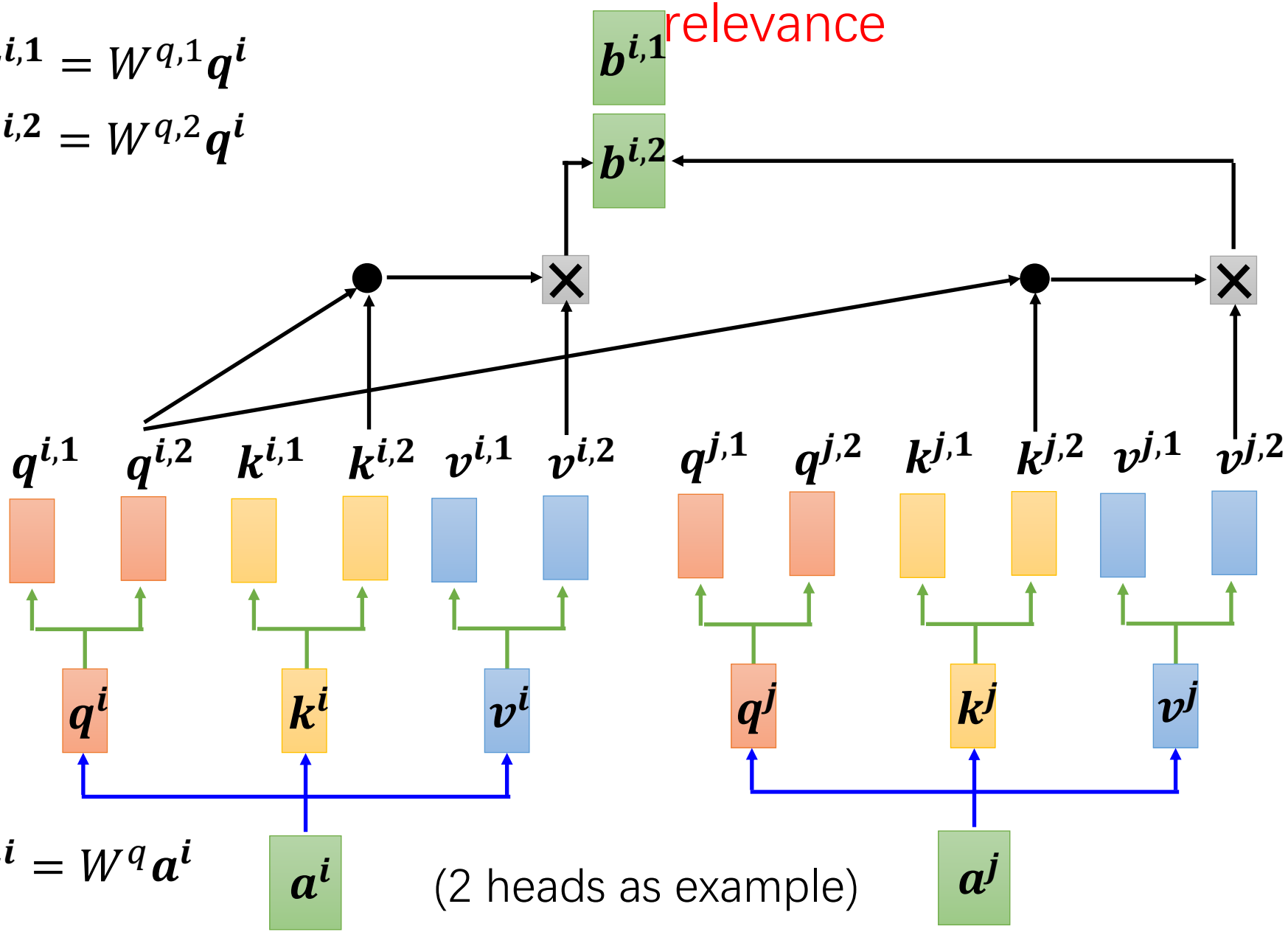
(2 heads as example)

Multi-head Self-attention

Different types of relevance

$$q^{i,1} = W^{q,1} q^i$$

$$q^{i,2} = W^{q,2} q^i$$

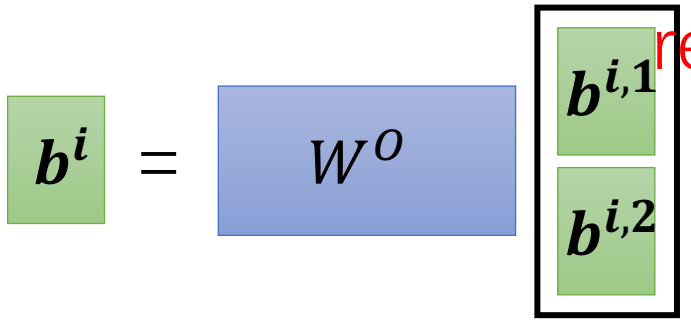


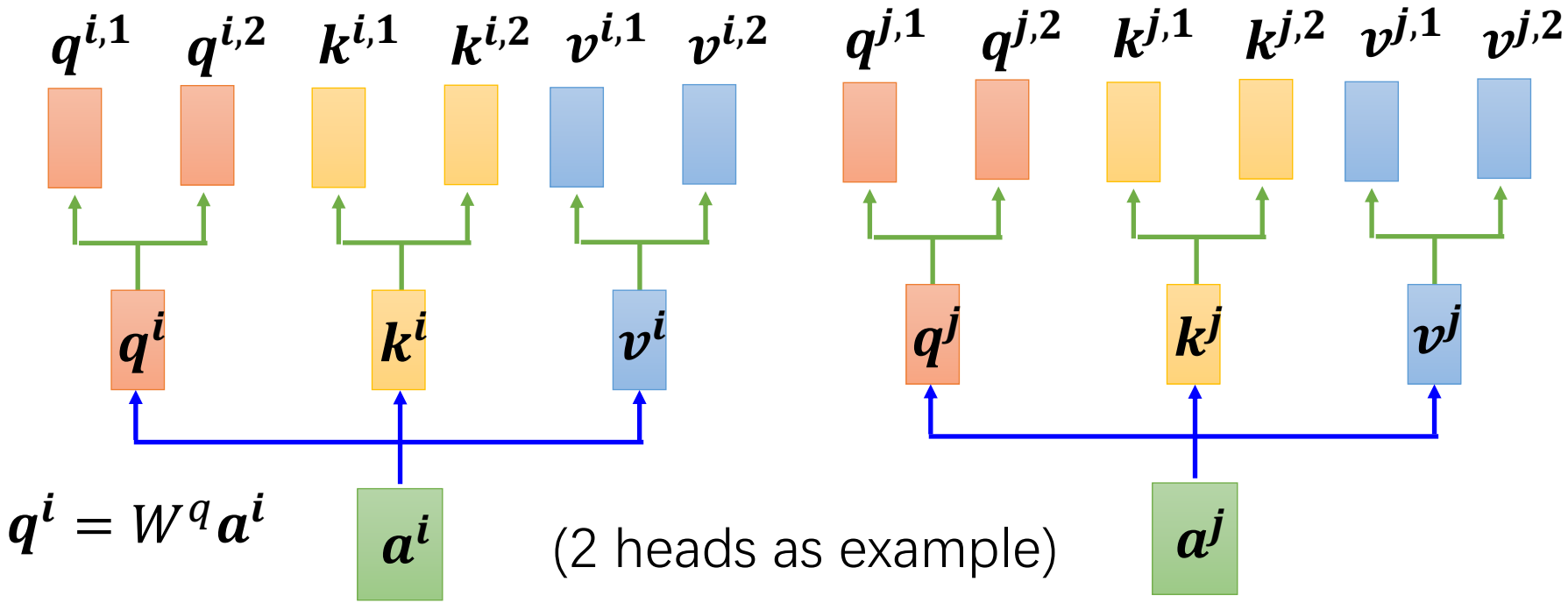
$$q^i = W^q a^i$$

(2 heads as example)

Multi-head Self-attention

Different types of relevance



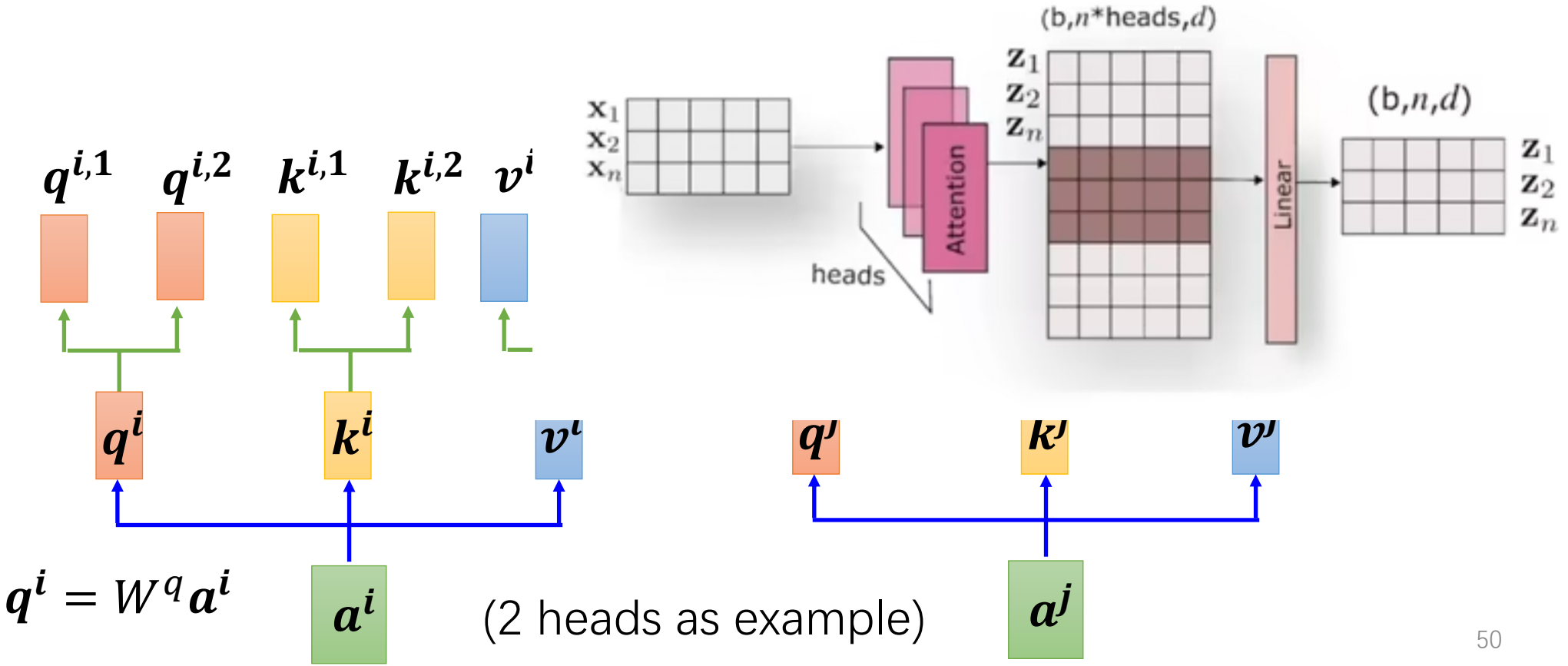


(2 heads as example)

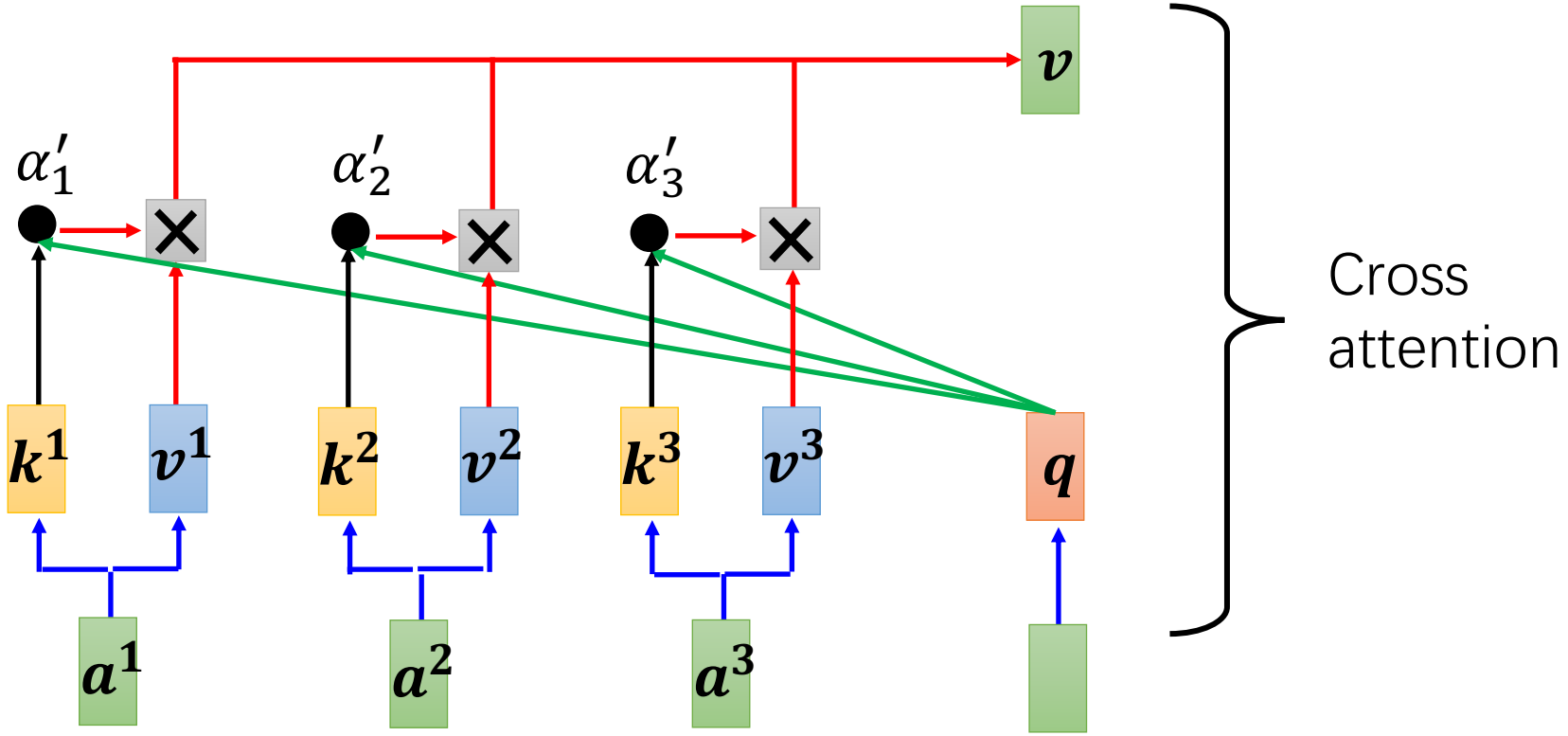
Multi-head Self-attention

Different types of relevance

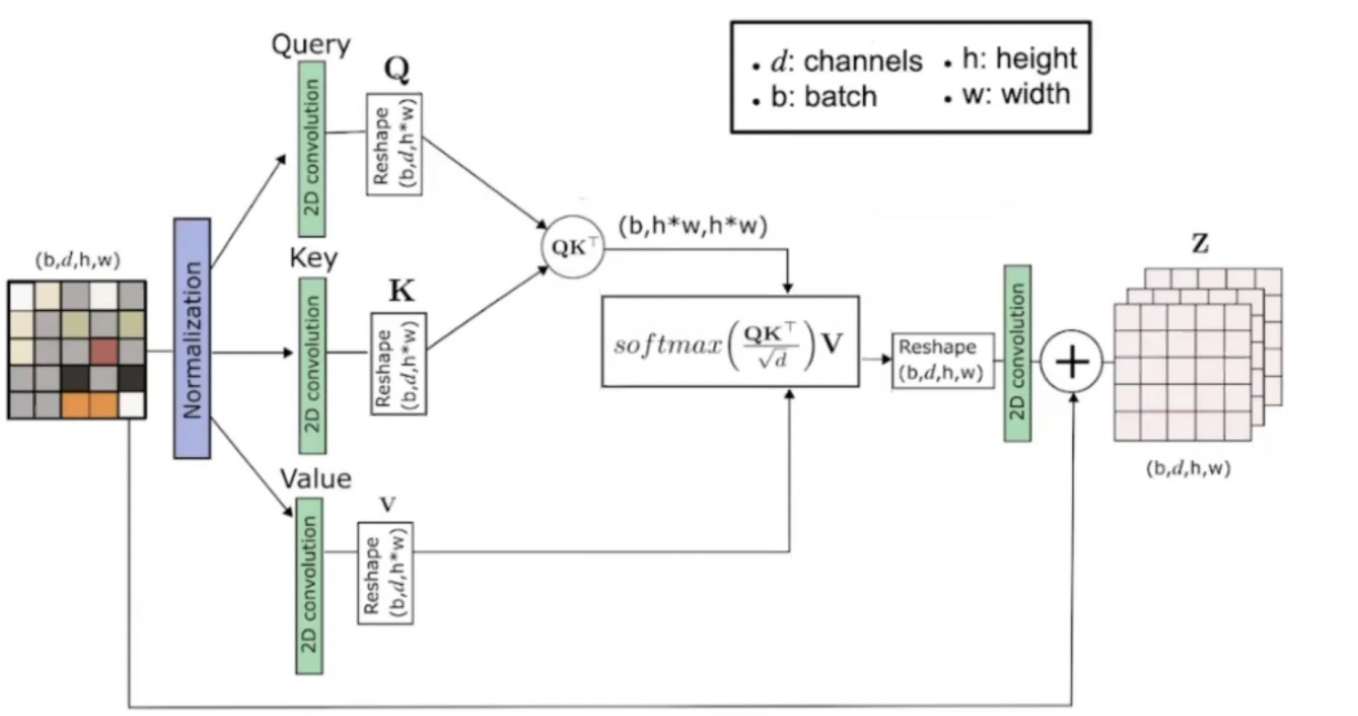
$$b^i = W^O \begin{bmatrix} b^{i,1} \\ b^{i,2} \end{bmatrix}$$

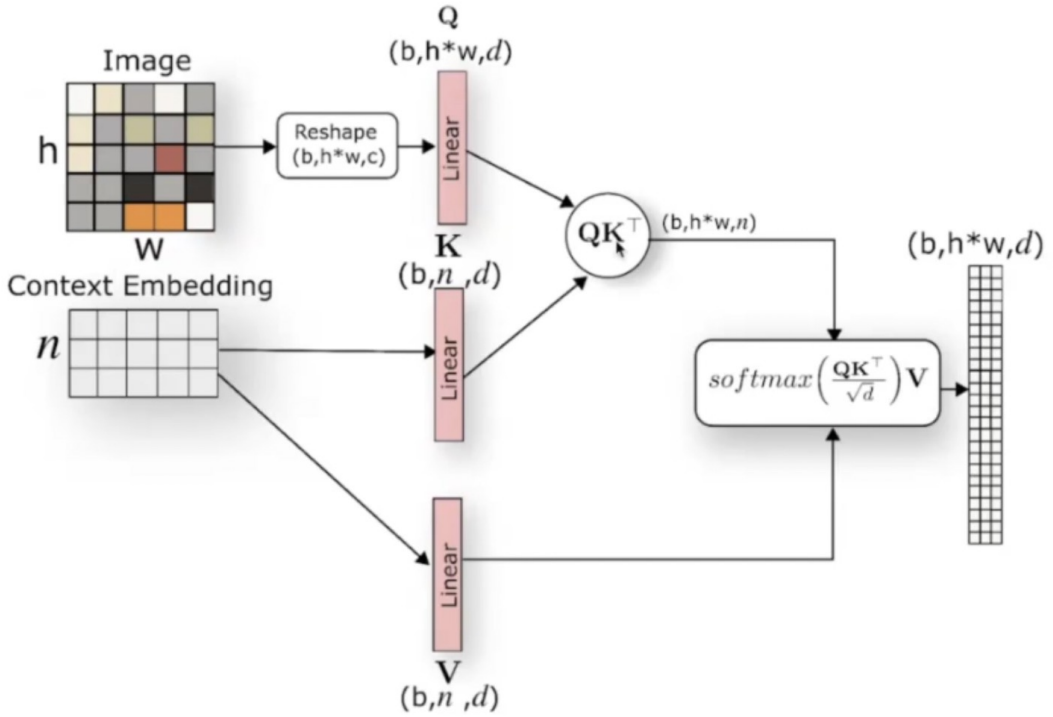


Cross Attention

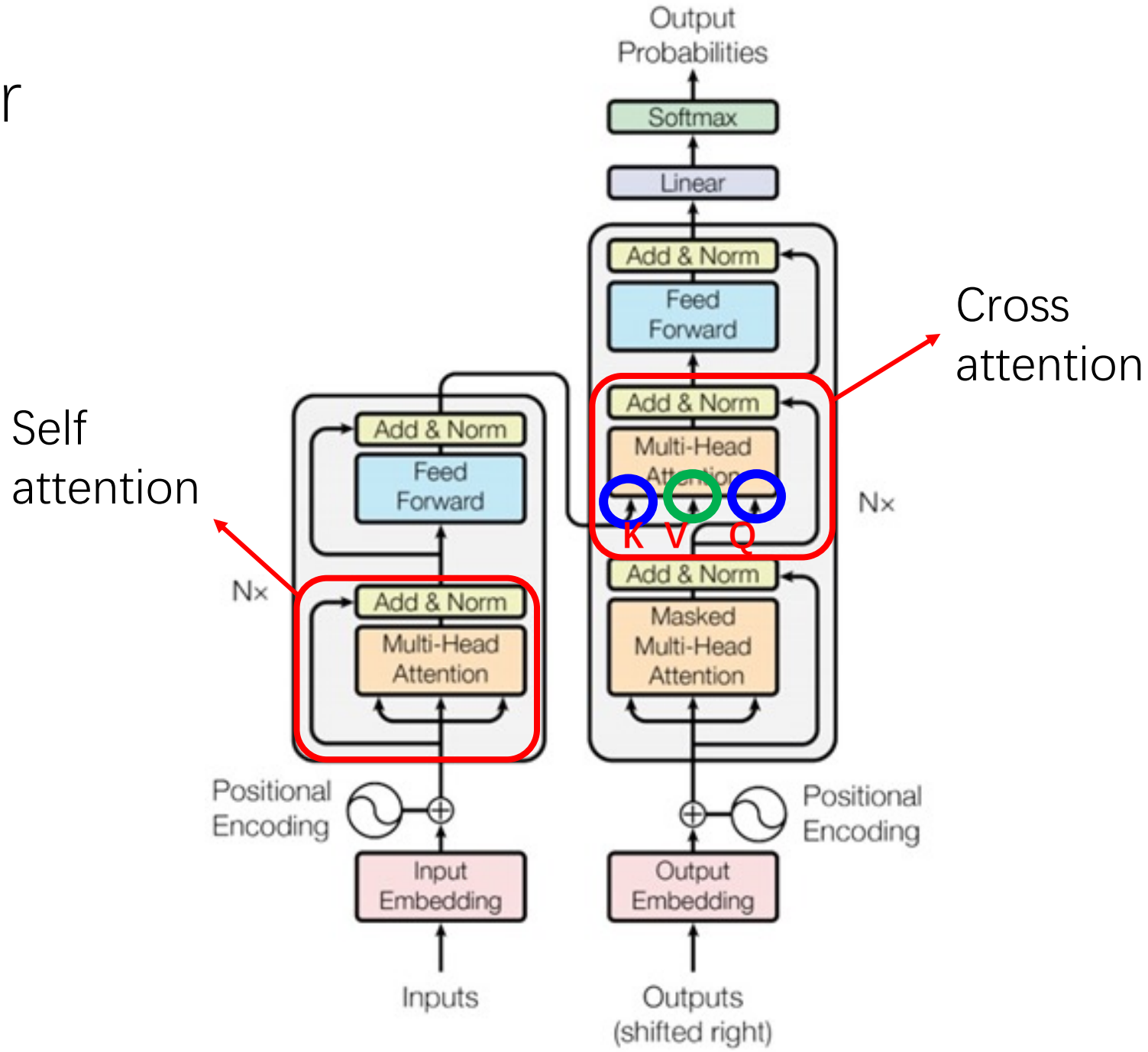


Cross Attention

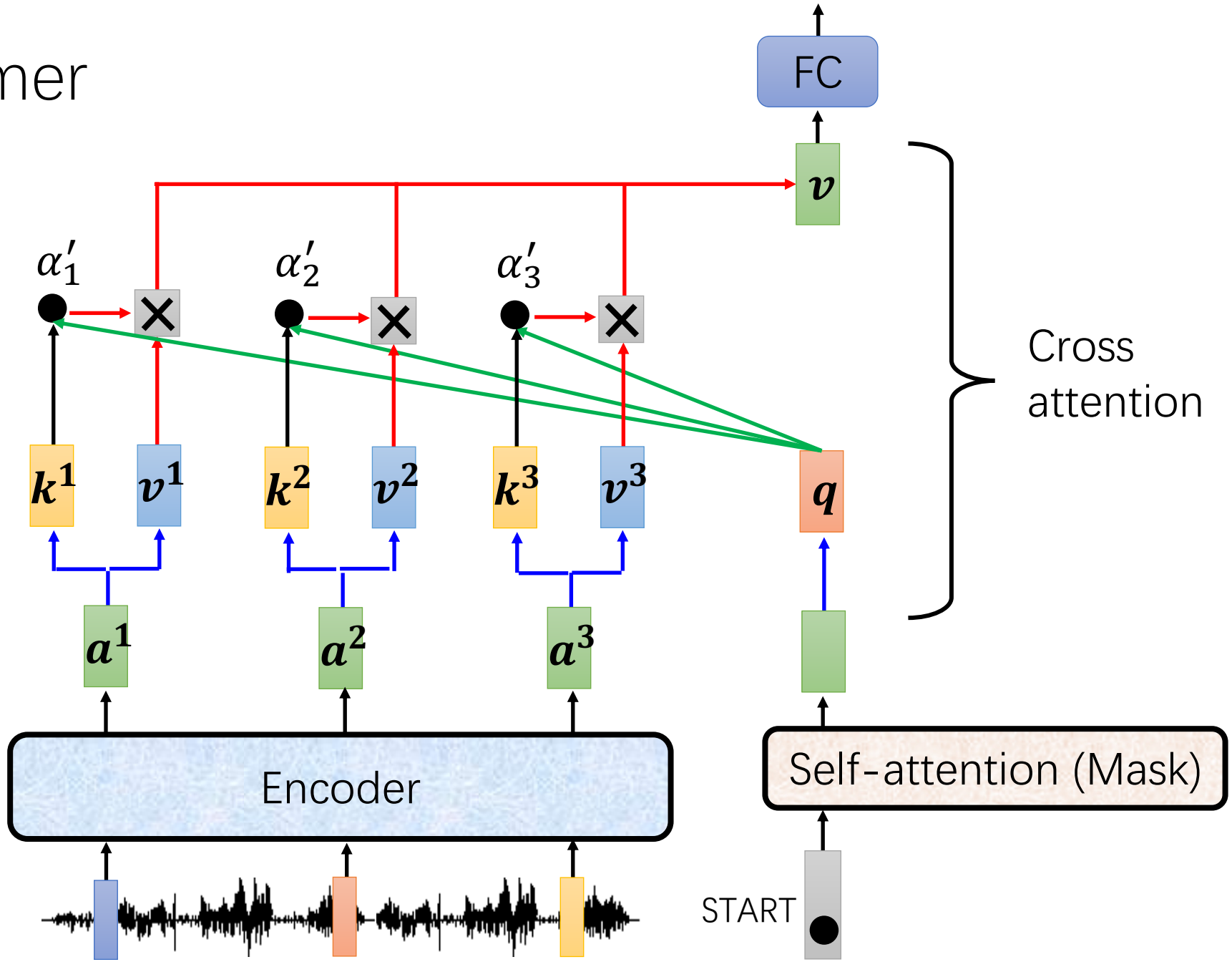




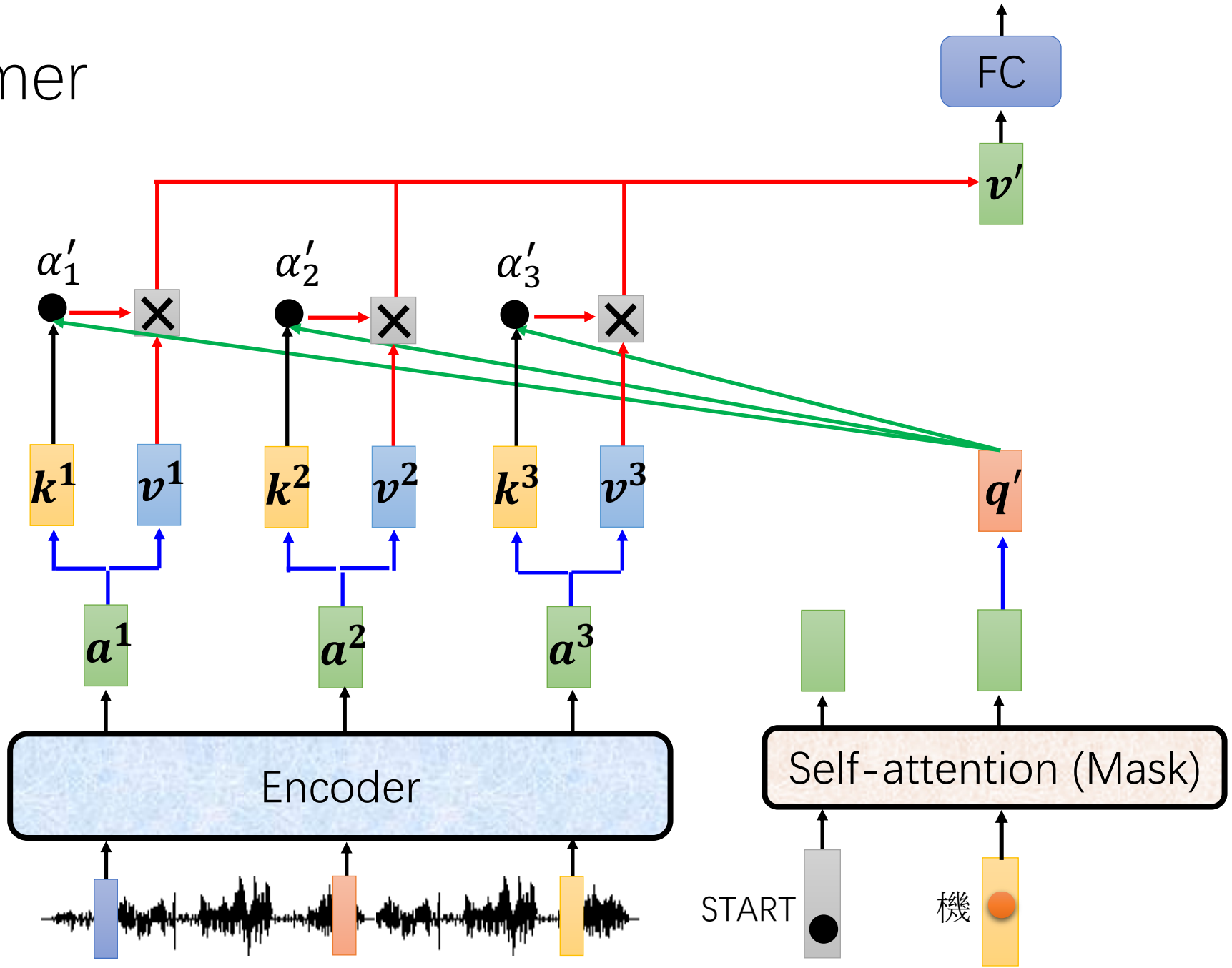
Transformer



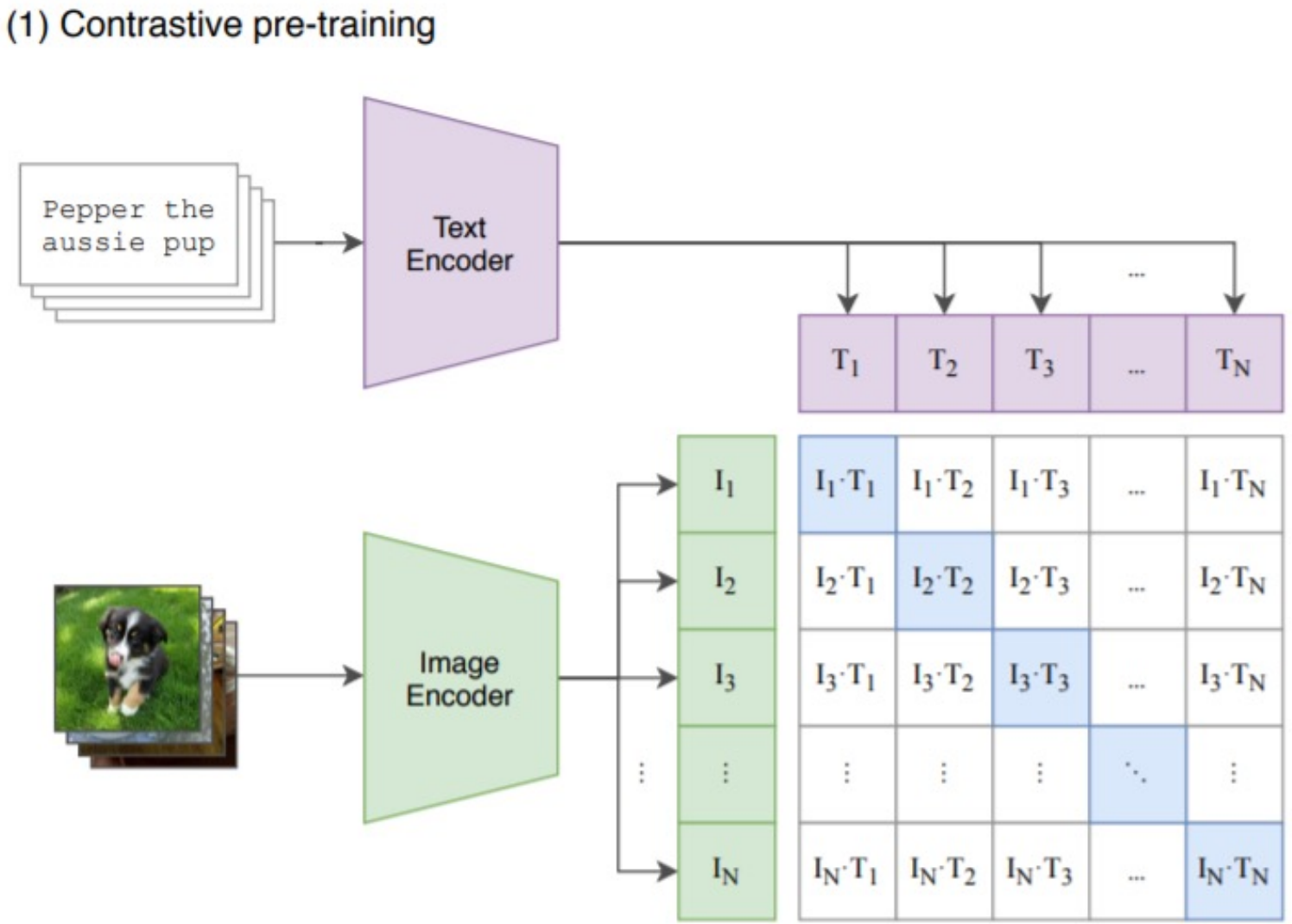
Transformer



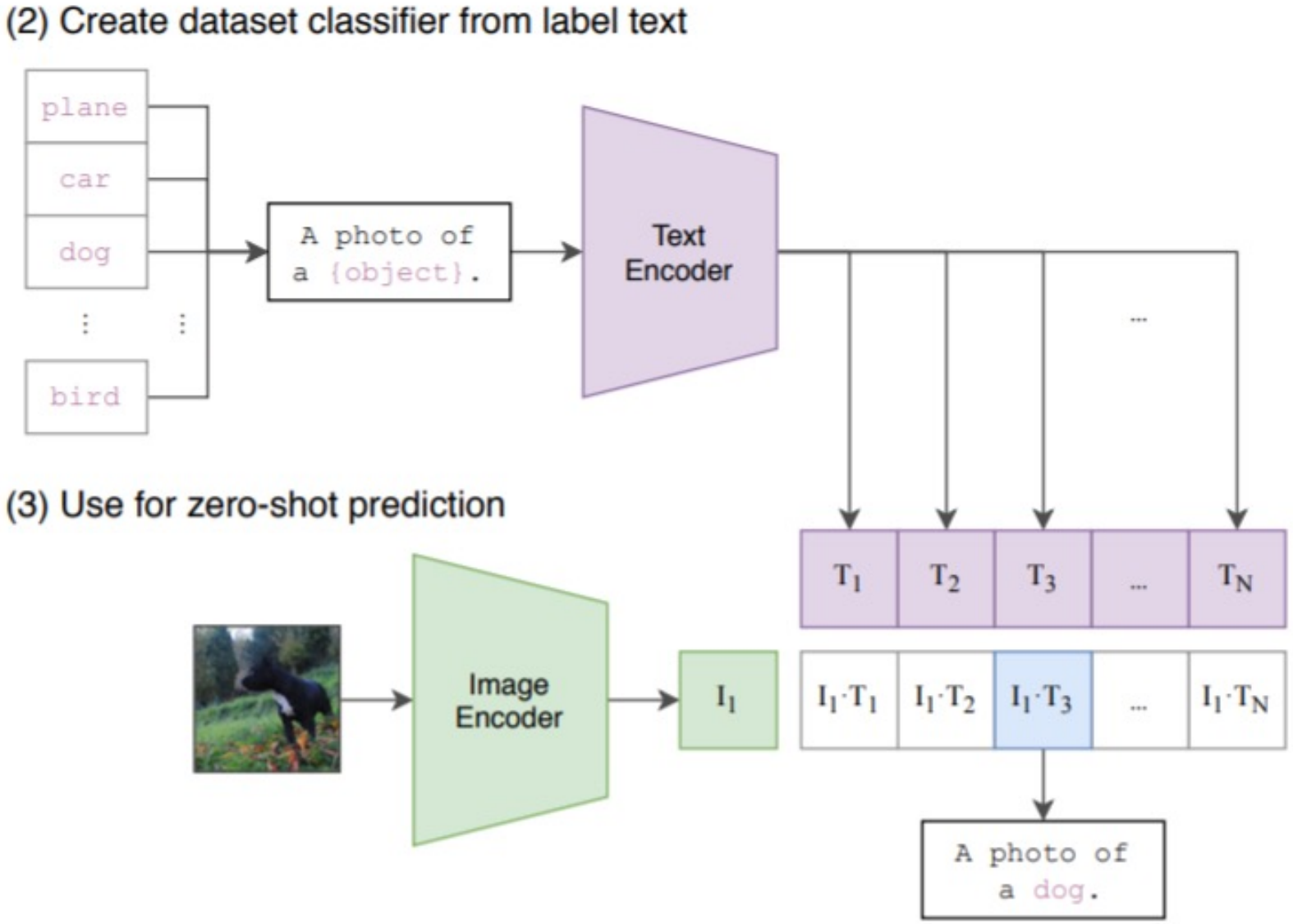
Transformer



CLIP Contrastive Language-Image Pre-Training



CLIP Contrastive Language-Image Pre-Training



LDM -Latent Diffusion Models -潜在扩散模型