深度生成模型
Deep Generative Model

- http://speech.ee.ntu.edu.tw/~tlkagk/courses_MLDS18.html

  李宏毅 台湾大学

目 录 (CONTENT)

大连理工大学 Dalian University Of Technology

PART THREE

变分自编码器

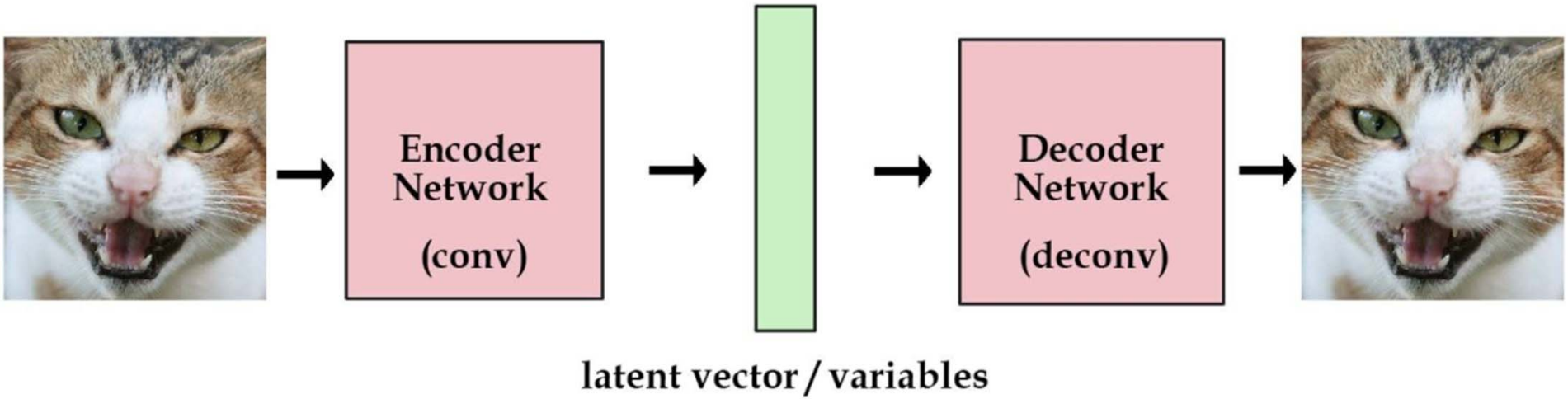
Variational AutoEncoder

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 回顾：自编码器（Autoencoder）

- 编码器 (Encoder) 将输入数据映射为一个隐含变量
- 解码器 (Decoder) 从隐含变量中重构出原始输入数据

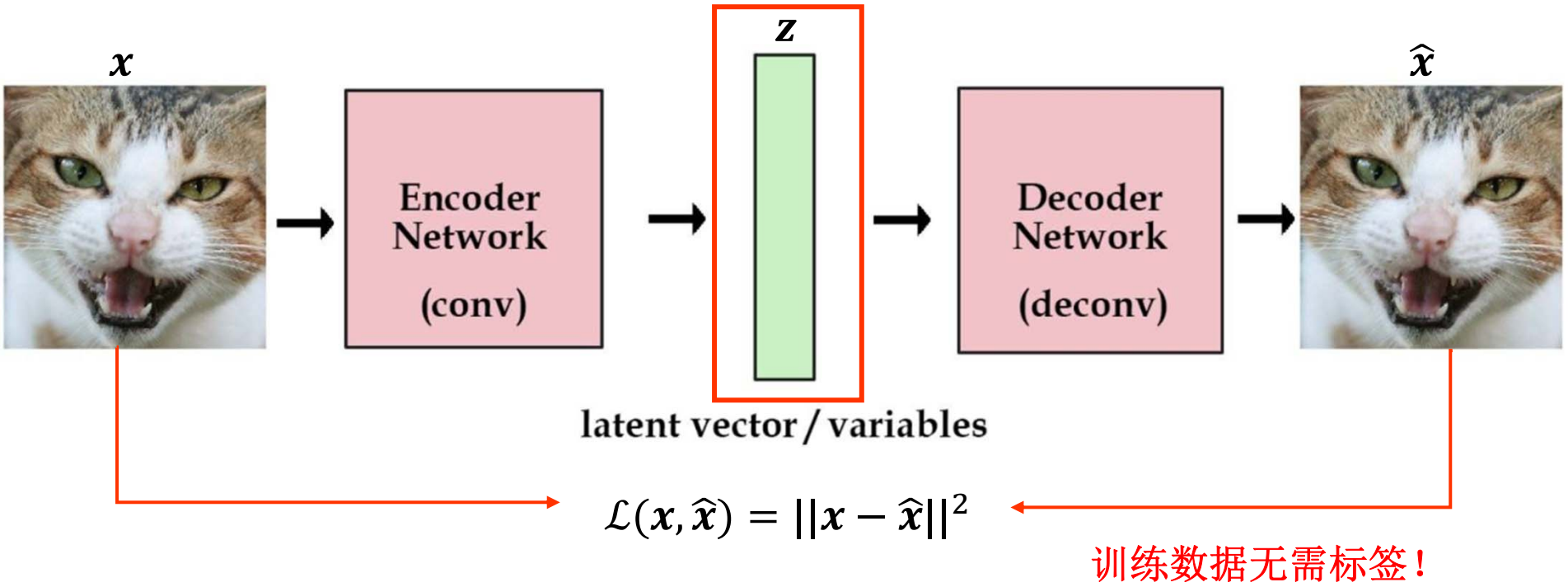


变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 回顾：自编码器（Autoencoder）

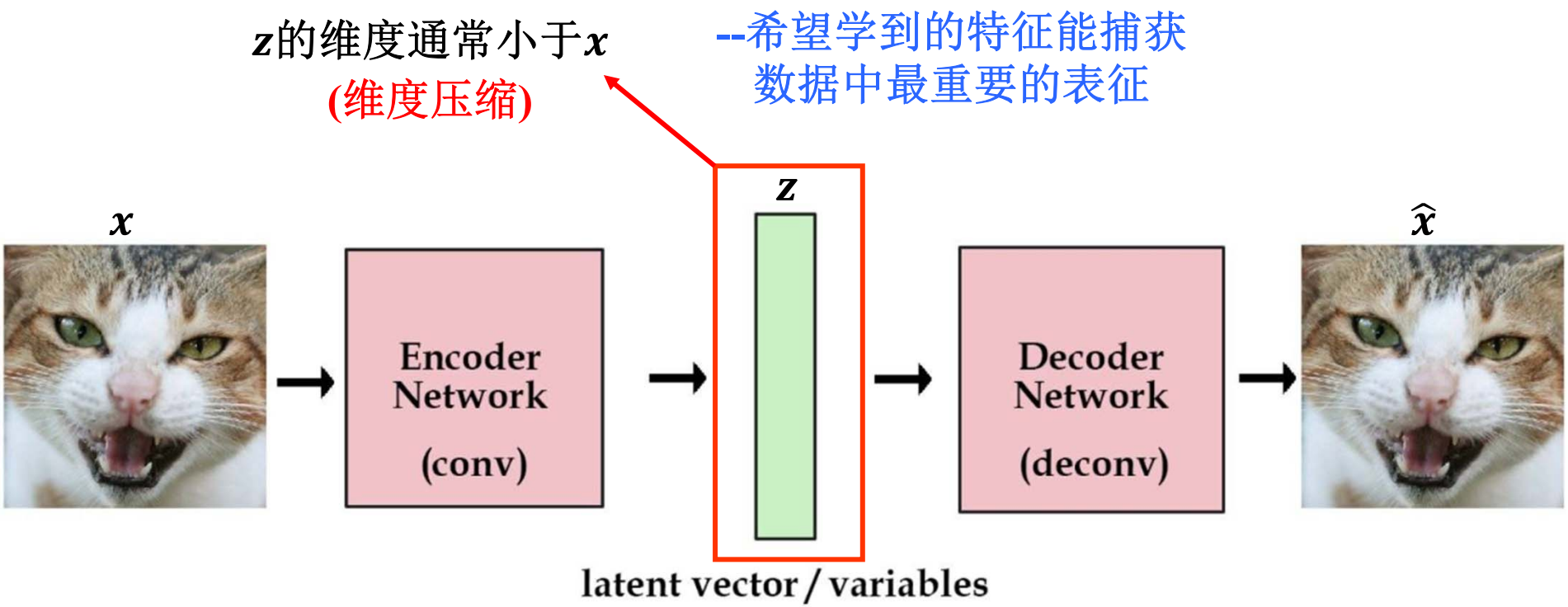
- 最小化重构误差，使网络学会捕获对数据的隐含变量



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 回顾： 自编码器（**Autoencoder**）



自编码是一种数据压缩方式

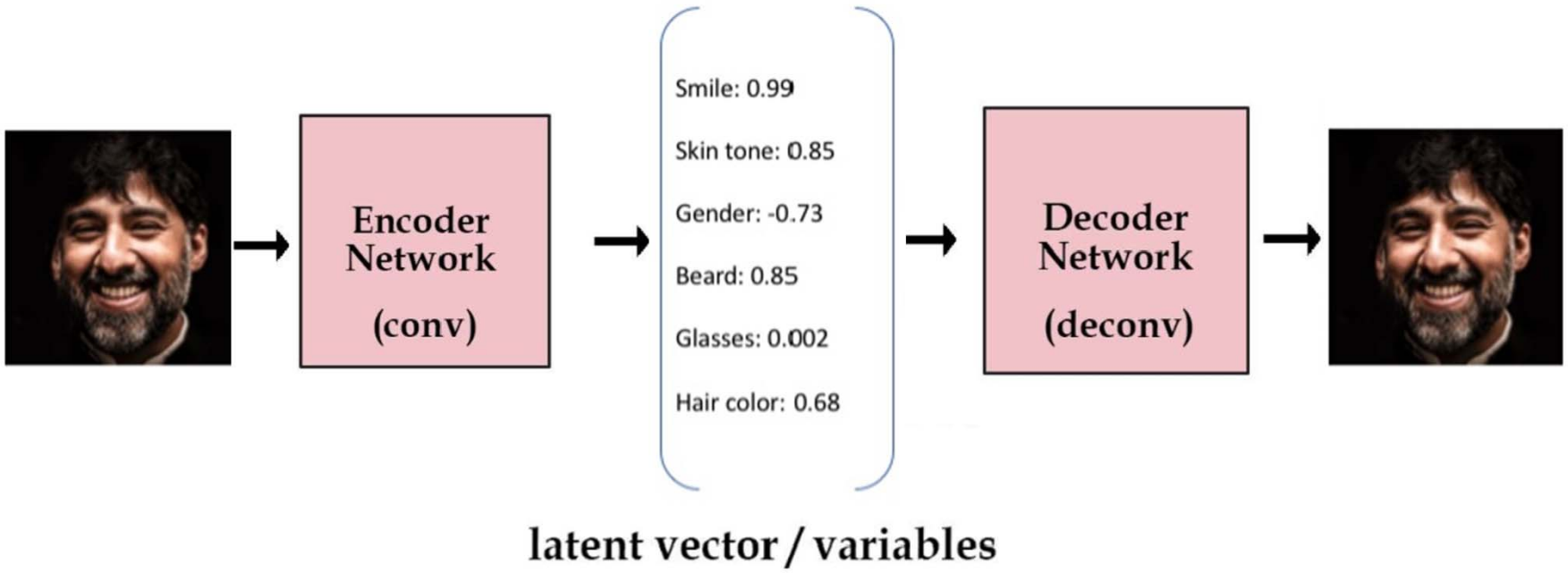
Autoencoding = **Auto**matically **encoding** data

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 回顾：自编码器（Autoencoder）

- 隐含变量可表征输入数据的某些结构或属性

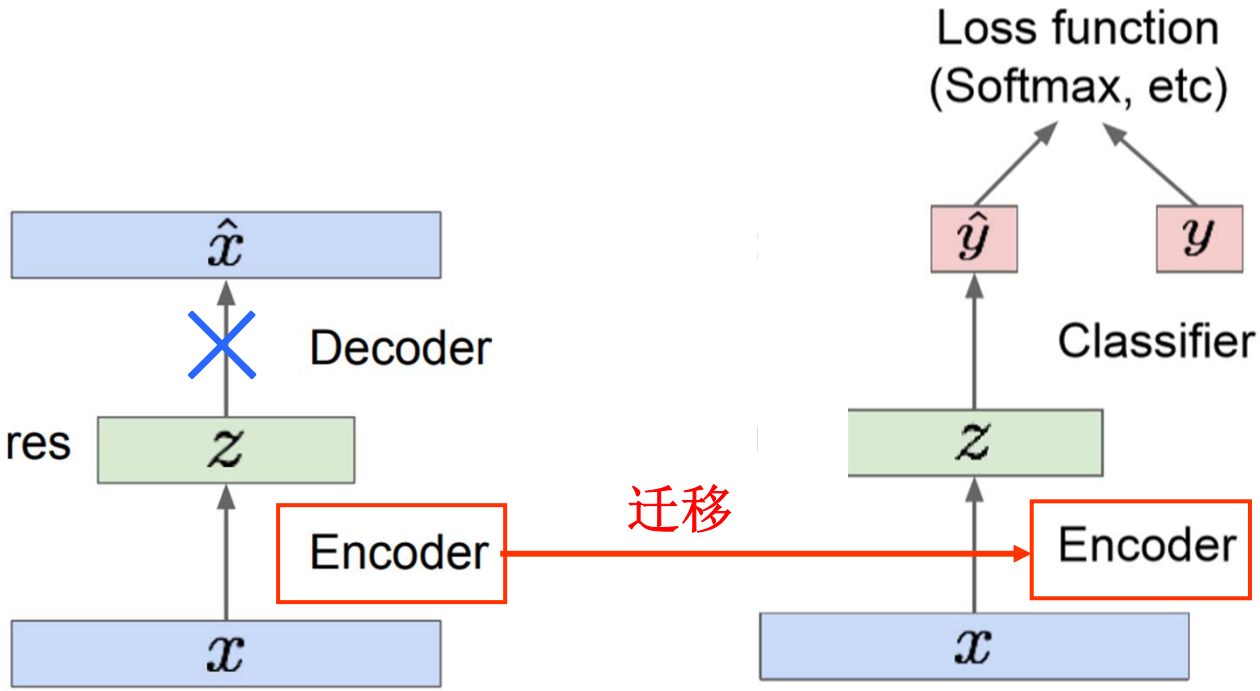


变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 回顾：自编码器（Autoencoder）

- 训练完成后， Encoder能够将训练数据映射至一个有效地隐含空间
- 可抛开Decoder，将学好的Encoder迁移至其他任务中



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

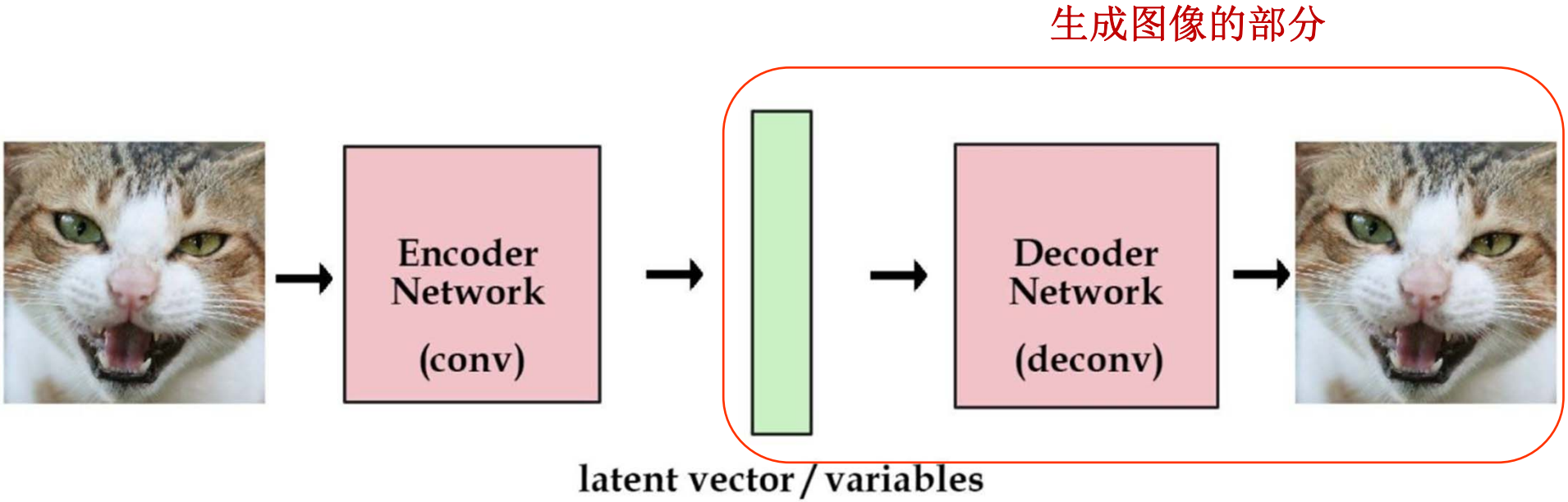
- 自编码器是否能生成新的图像？

AutoEncoder

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

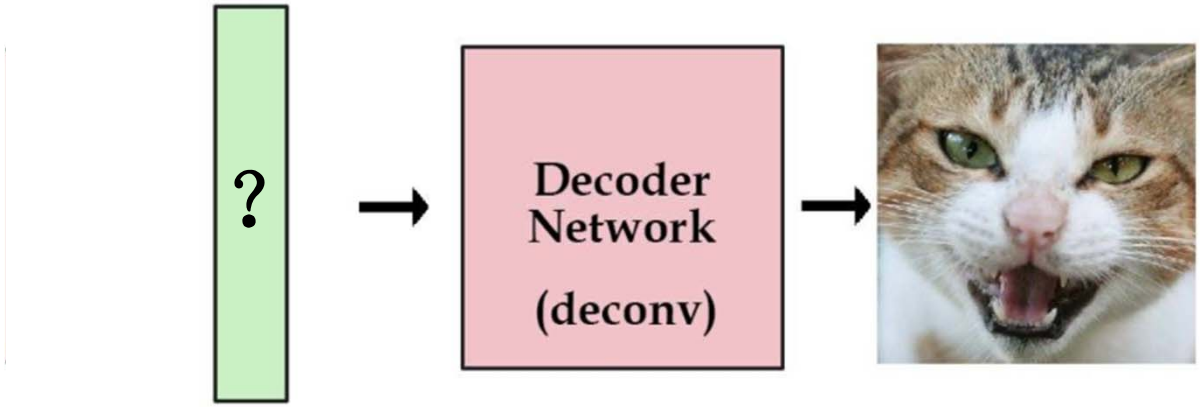
- 自编码器是否能生成新的图像？



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 自编码器是否能生成新的图像？
  - 测试阶段
    - 不知道如何创建需要输入的隐变量



生成图像的部分

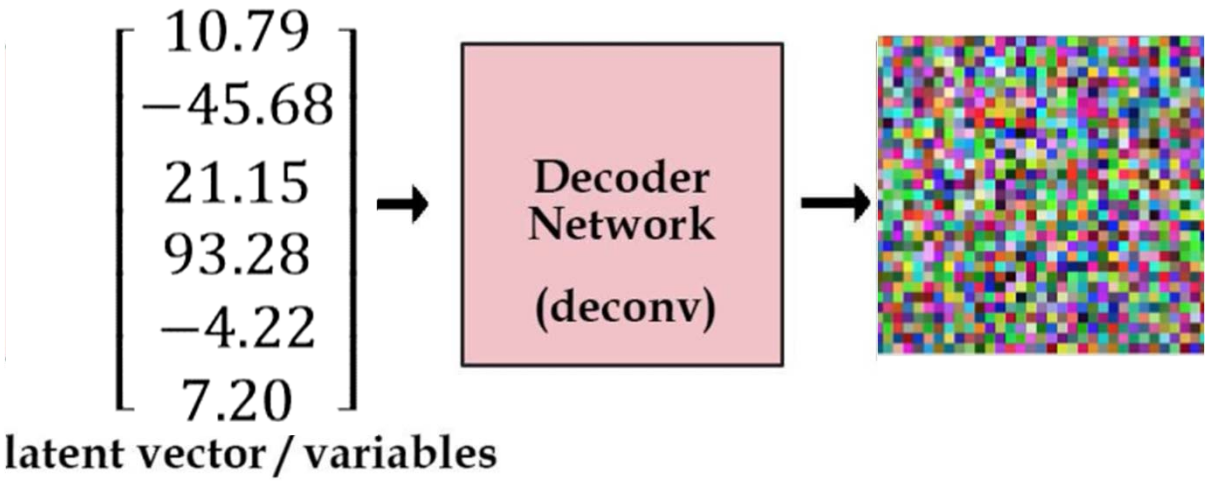
变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 自编码器是否能生成新的图像？

➢ 测试阶段

- 假设输入一组随机的向量，我们很大程度上将得到一个噪声图像

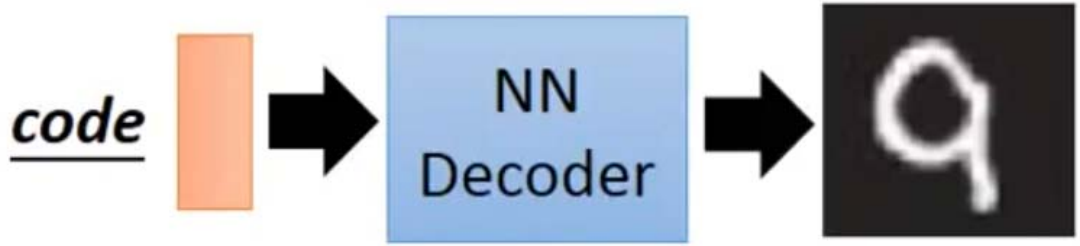


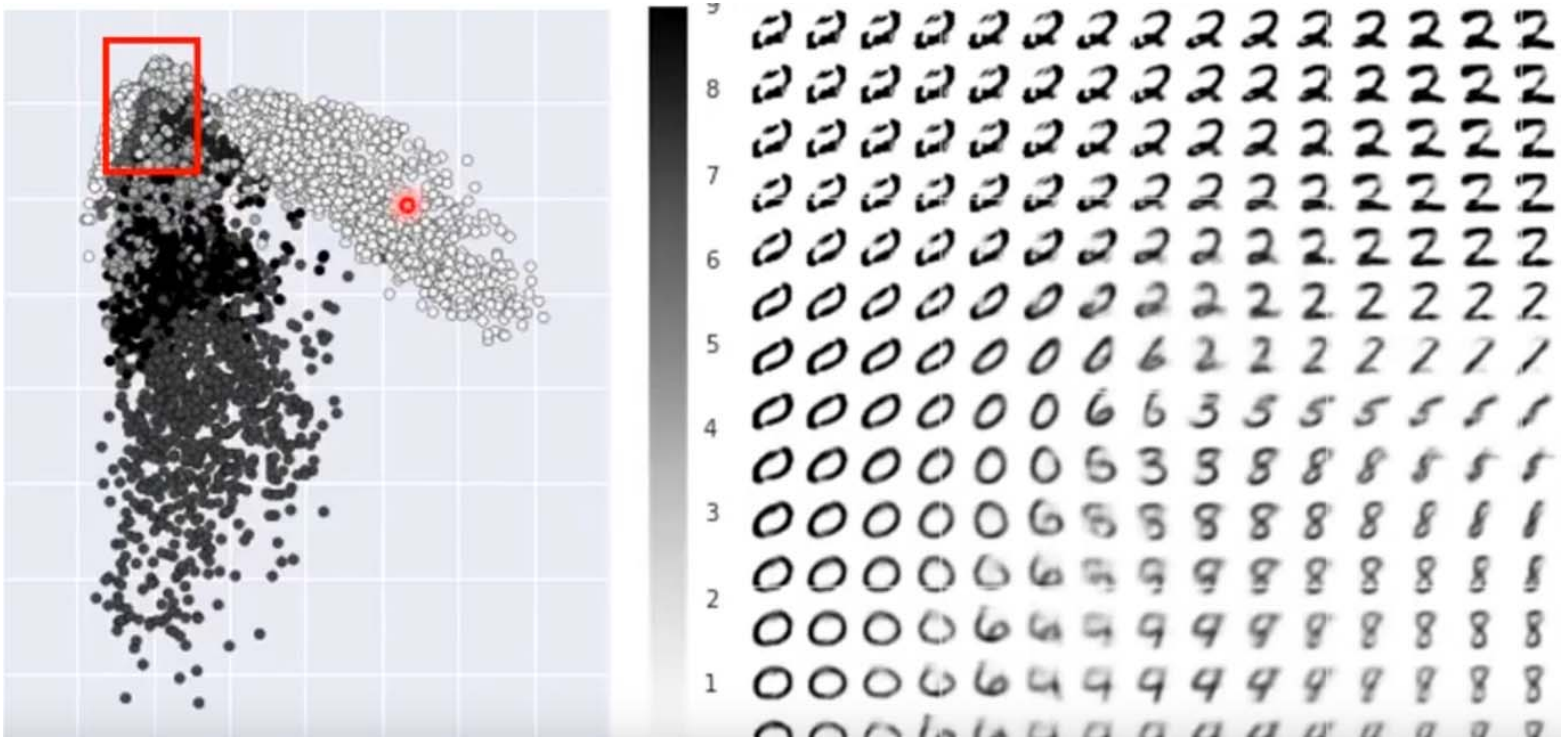
生成图像的部分

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

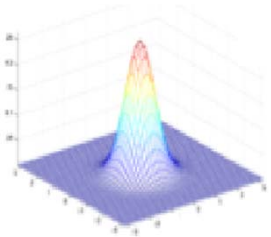
- 自编码器是否能生成新的图像？
  - 测试阶段



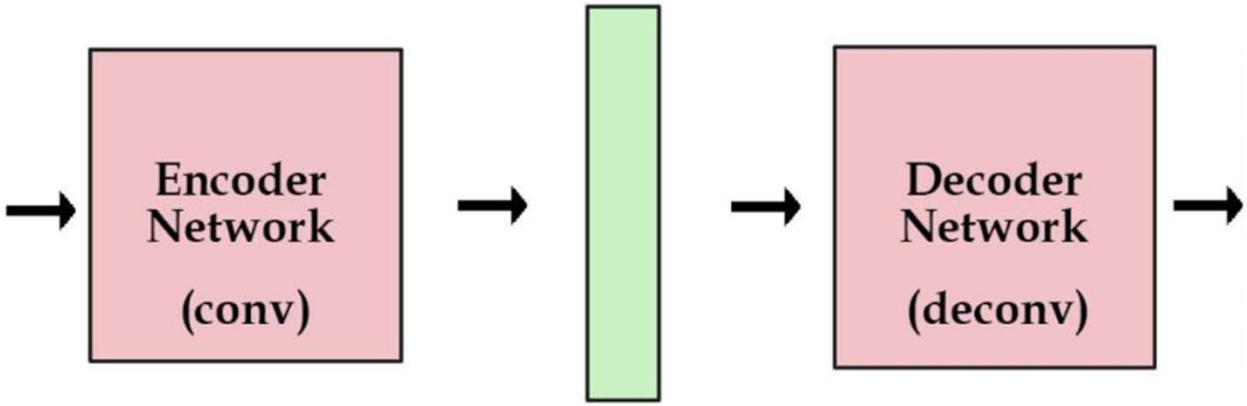


变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

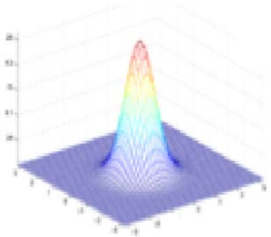


- 自编码器是否能生成新的图像？
  - 解决方案—变分自编码器
    - 对Encoder增加约束，强制其产生的隐含变量基本服从某种分布



变分自编码器（VAE）

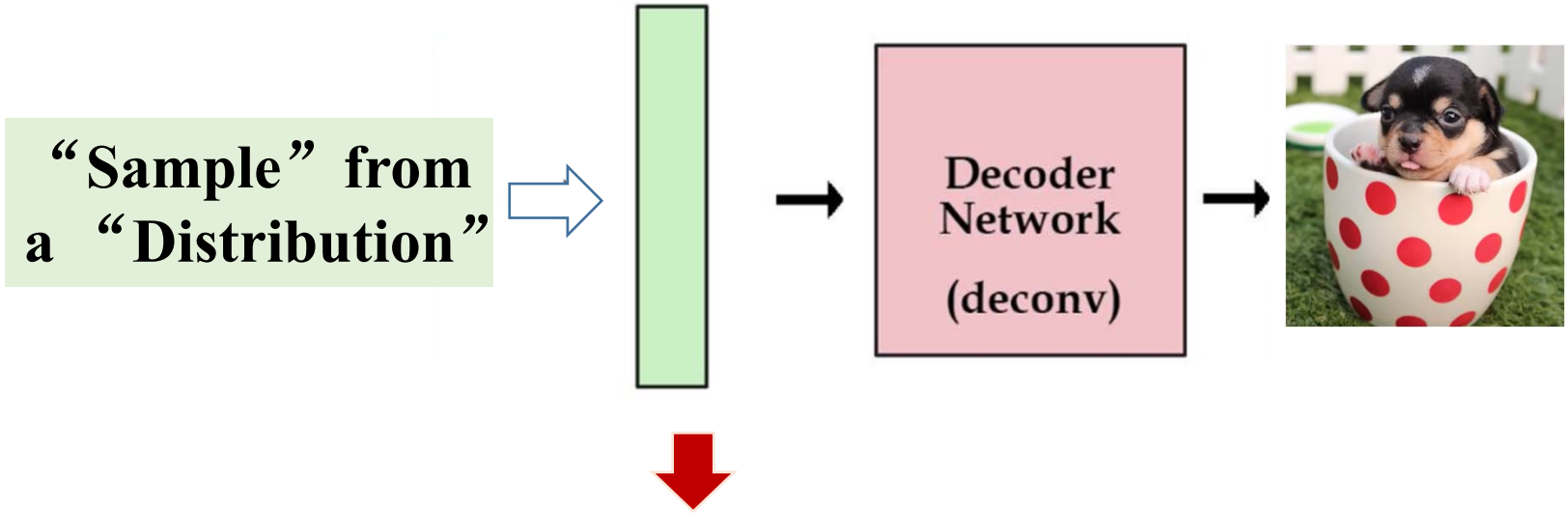
大连理工大学 Dalian University Of Technology



■ 自编码器是否能生成新的图像？

➢ 解决方案—变分自编码器

- 测试时，从该分布中采样作为隐含量，并传递到解码器，生成新图像



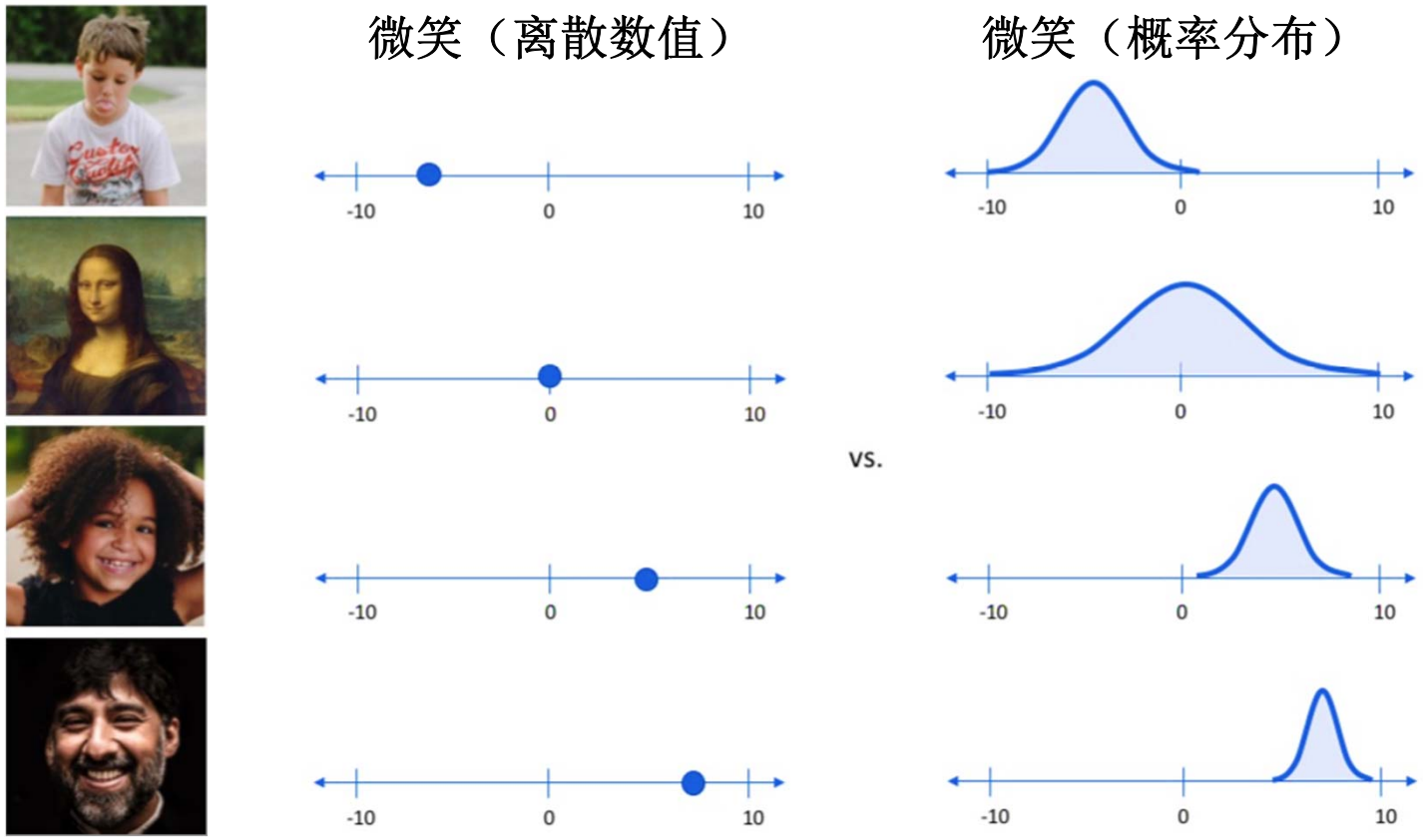
变分自编码器

Variational Autoencoder, AVE

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

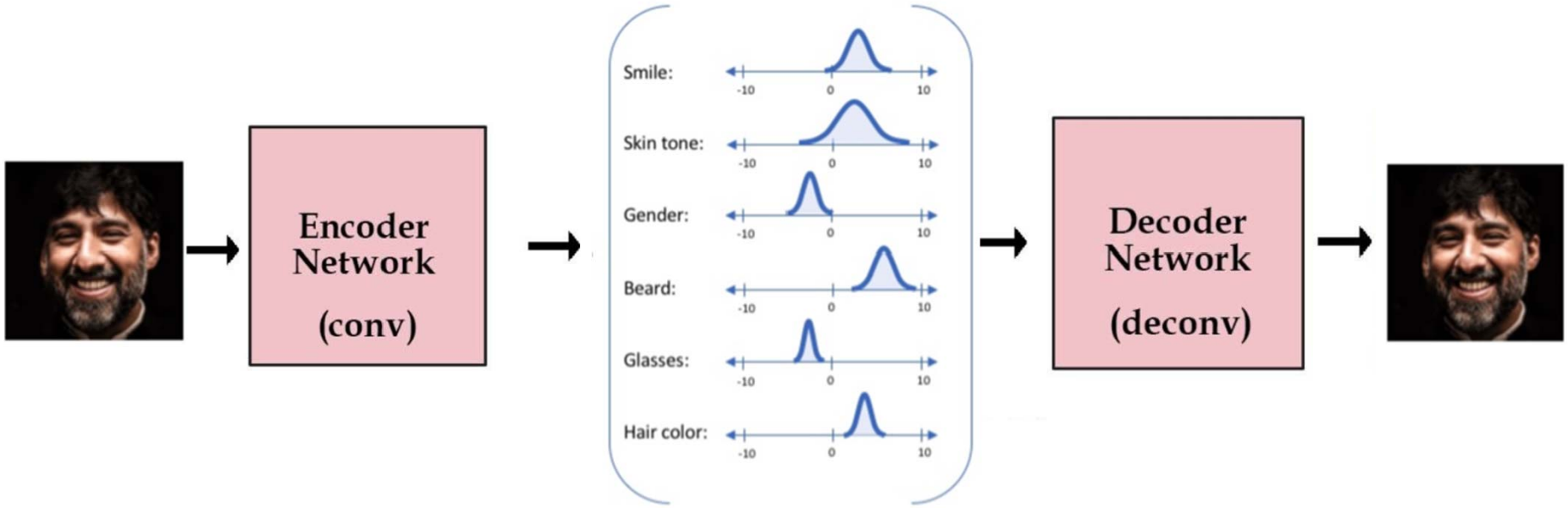
- 隐含变量的每个元素由原先的“单值” =>“取值的概率分布”



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 生成图像（即对隐含变量解码）时，从每个隐含分布中随机采样，生成一个向量作为解码器模型的输入

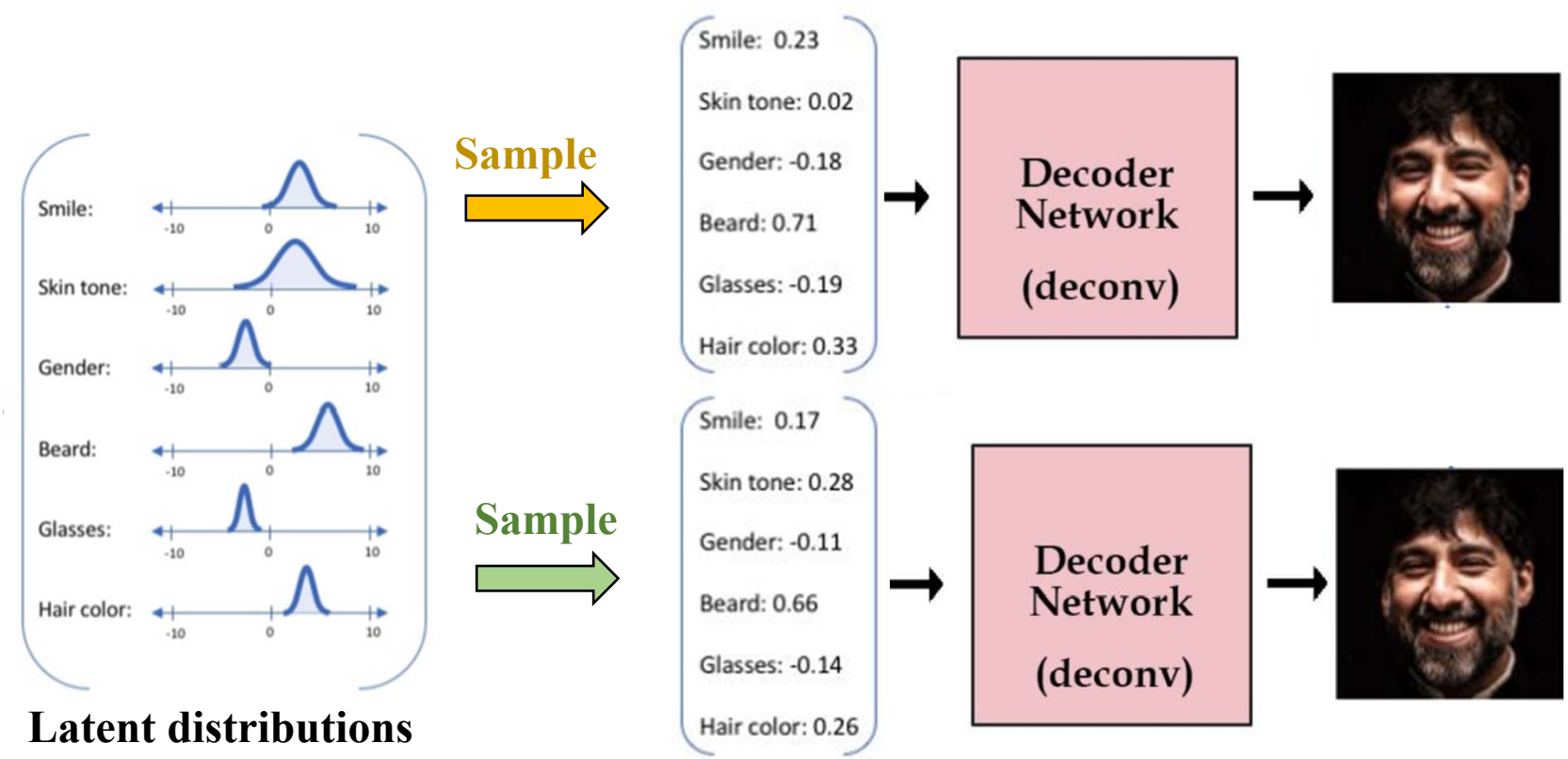


Latent distributions

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

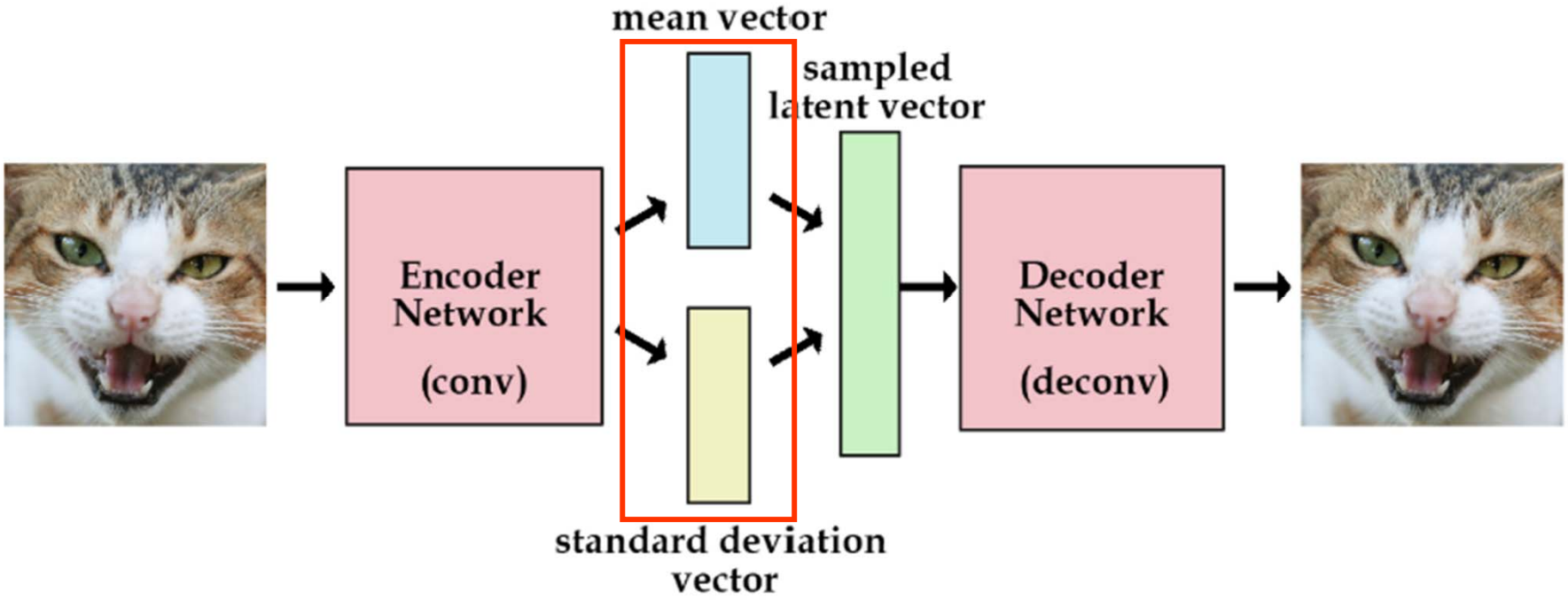
- 在连续平滑的隐含分布空间随机采样，彼此相邻的值具有类似重构



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

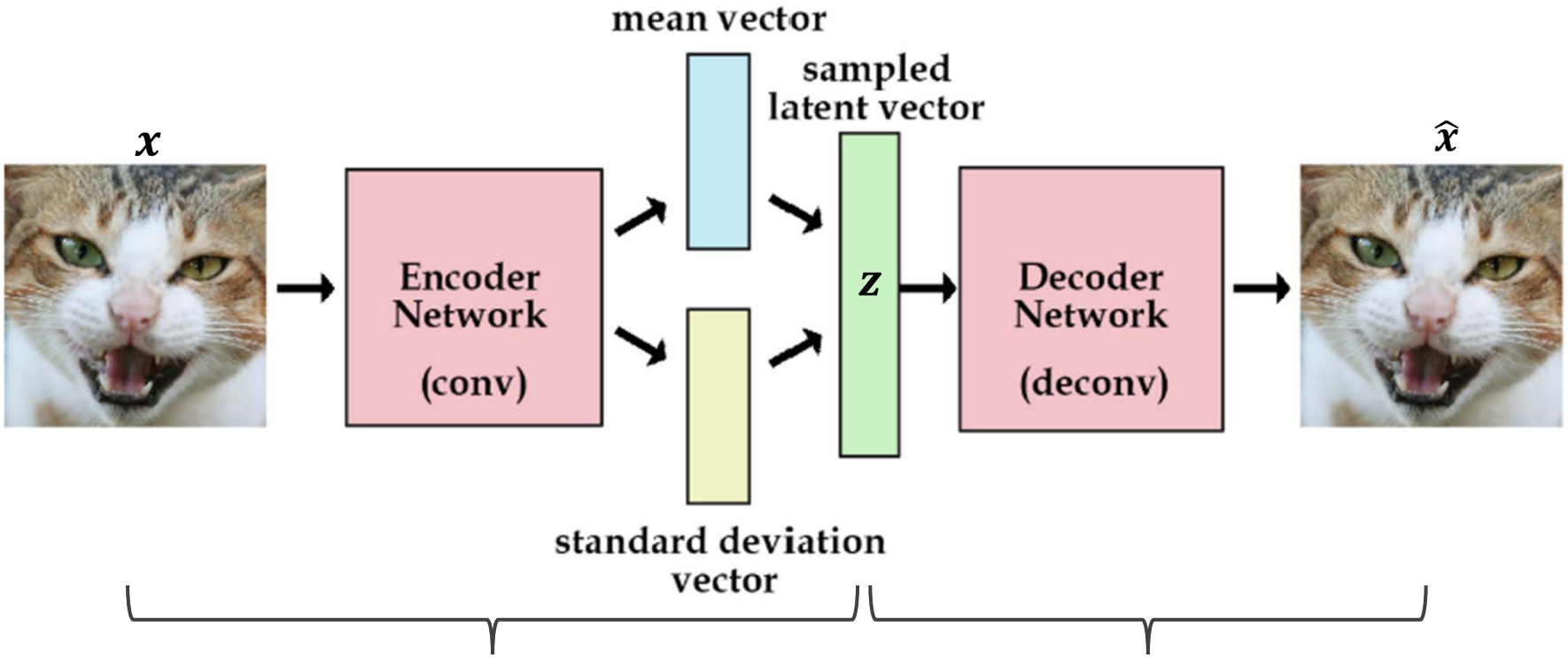
- 通常假设采样的隐含变量服从正态分布 $\mathcal{N}(\mu, \sigma^2)$
- 从均值分布和标准差分布中随机抽取，计算隐含变量



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- **VAE是一种概率形式的自编码器**



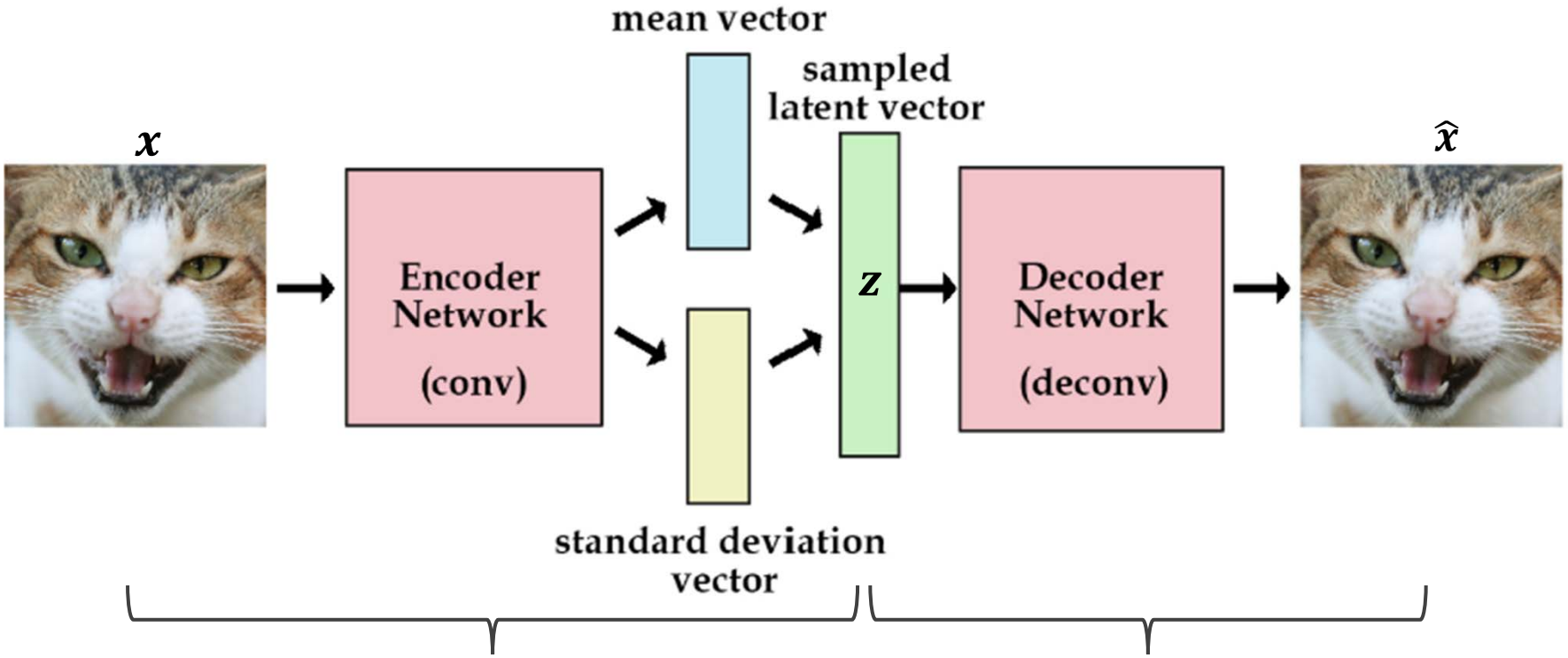
z 的后验分布的推断过程，即$p_\phi(z|x)$

$\hat{x}$的条件分布生成过程，即 $q_\theta(\hat{x}|z)$

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 模型优化



z 的后验分布的推断过程，即$p_\phi(z|x)$

$\hat{x}$的条件分布生成过程，即 $q_\theta(\hat{x}|z)$

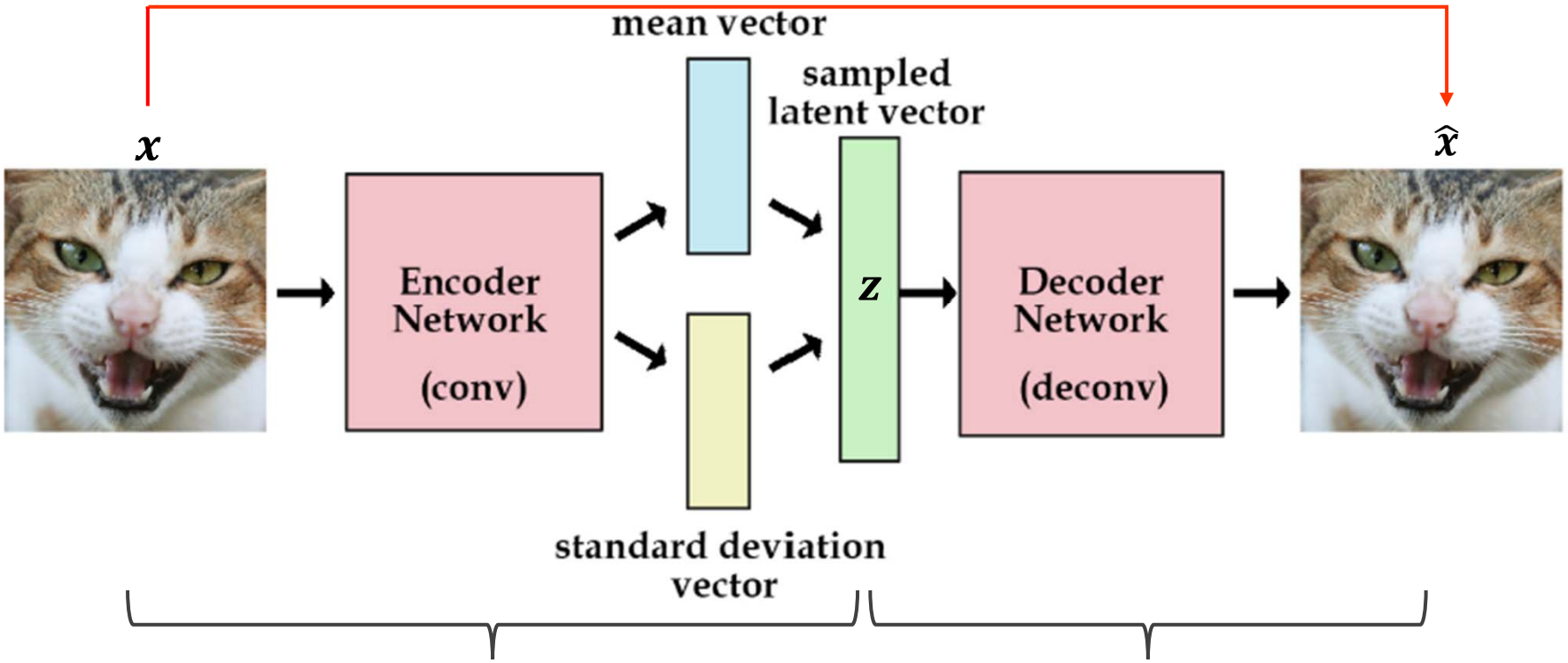
$\mathcal{L}(\phi, \theta)$ = 重构误差+ 约束项

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 模型优化

$e.g.\ ||x-\hat{x}||^2$



z 的后验分布的推断过程，即$p_\phi(z|x)$

$\hat{x}$的条件分布生成过程，即 $q_\theta(\hat{x}|z)$

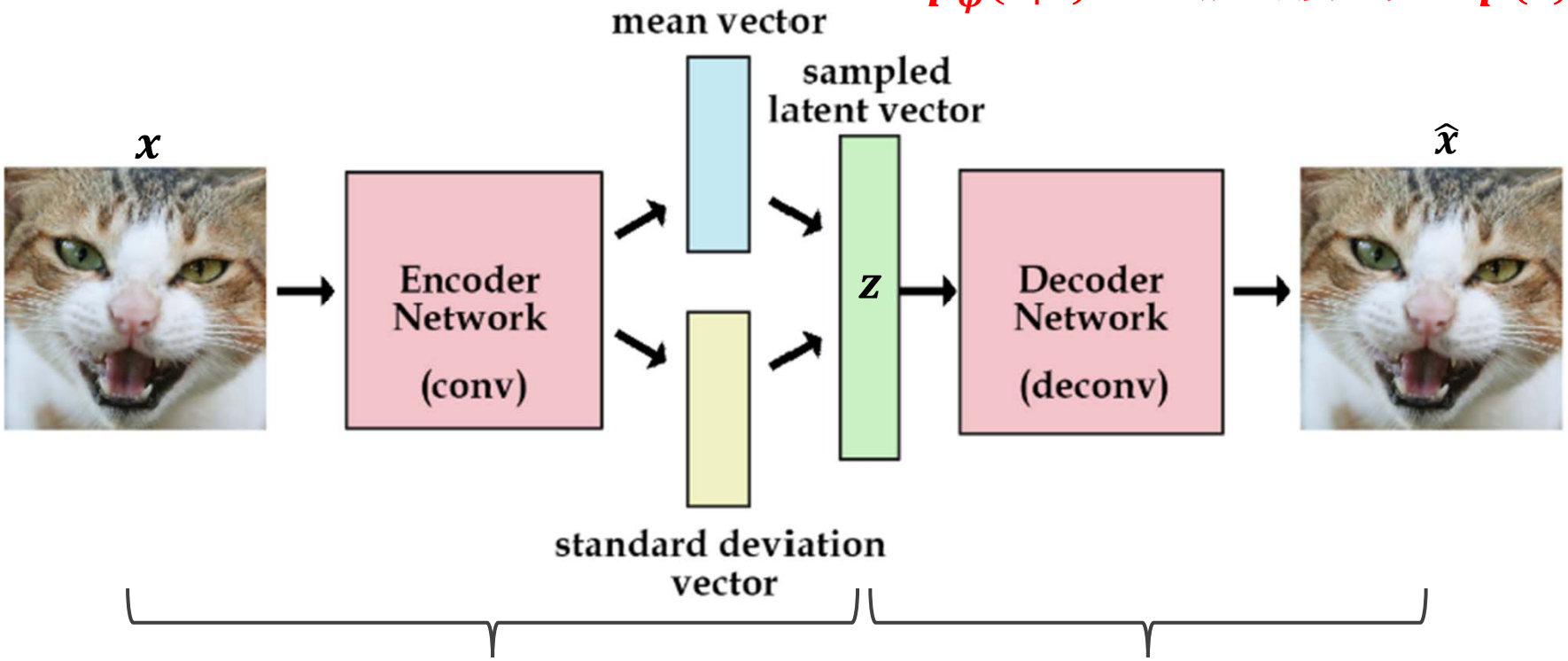
$\mathcal{L}(\phi,\theta)$ = 重构误差+ 约束项

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 模型优化

$p_\phi(z|x)$ 尽可能的接近先验$p(z)$



z 的后验分布的推断过程，即$p_\phi(z|x)$

$\hat{x}$的条件分布生成过程，即 $q_\theta(\hat{x}|z)$

$\mathcal{L}(\phi, \theta)$ = 重构误差+约束项

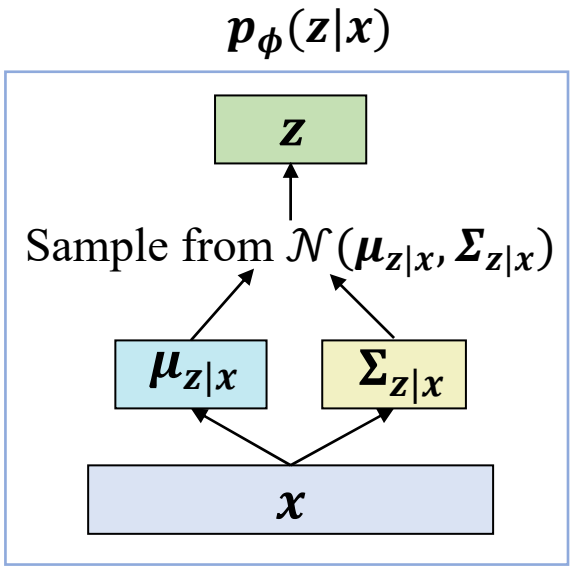
变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 模型优化

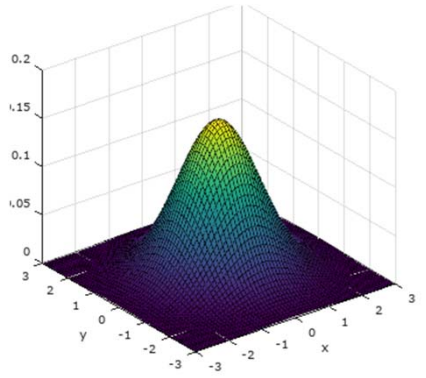
➢ 约束项

- 目标：$p_\phi(z|x)$ 尽可能的接近先验 $p(z)$



Encoder

通常假设$p(z)\sim\mathcal{N}(0,1)$



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

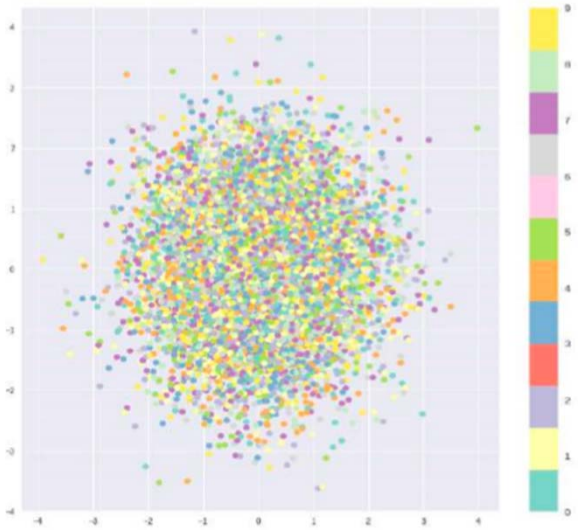
■ 模型优化

➢ 正则项

$\boldsymbol{p}(\boldsymbol{z})\sim\mathcal{N}(0,1)$

- 目标：$\boldsymbol{p_\phi}(\boldsymbol{z}|\boldsymbol{x})$ 尽可能的接近先验$\boldsymbol{p}(\boldsymbol{z})$
- 采用KL散度，即：

$$D_{KL}\left(\boldsymbol{p_\phi}(\boldsymbol{z}|\boldsymbol{x})\parallel\boldsymbol{p}(\boldsymbol{z})\right)$$
$$=D_{KL}\left(\mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{z}|\boldsymbol{x}},\boldsymbol{\Sigma}_{\boldsymbol{z}|\boldsymbol{x}}\right)\parallel\mathcal{N}(0,1)\right)$$
$$=-\frac{1}{2}\sum_{j}^{K}(\boldsymbol{\Sigma}_{\boldsymbol{z}|\boldsymbol{x}_j}+\boldsymbol{\mu}_{\boldsymbol{z}|\boldsymbol{x}_j}^2-1-log\boldsymbol{\Sigma}_{\boldsymbol{z}|\boldsymbol{x}_j})$$



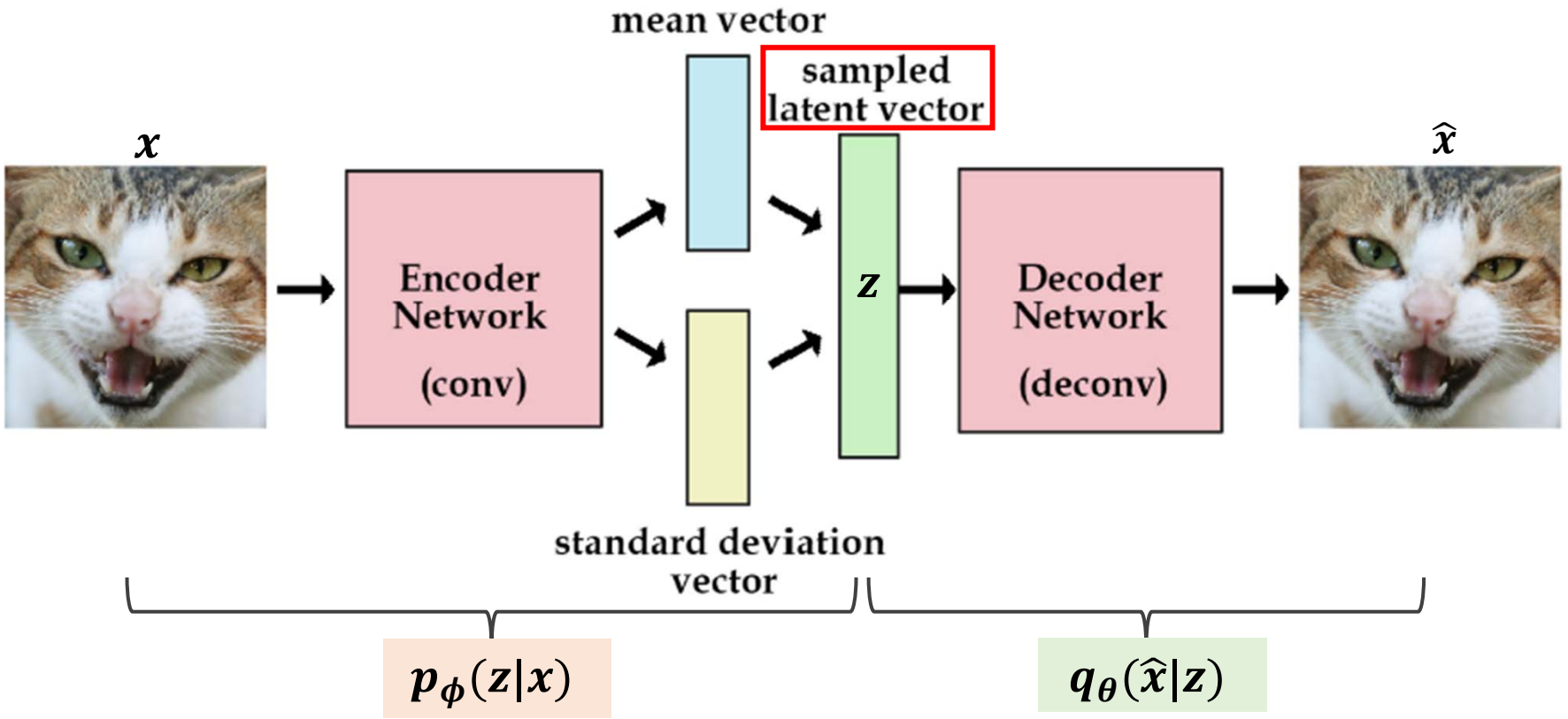
- 鼓励Encoder的输出分布$\boldsymbol{p_\phi}(\boldsymbol{z}|\boldsymbol{x})$均衡地分布在隐变量的先验空间

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 模型优化

问题：**sampling layer无法后向传递梯度!**



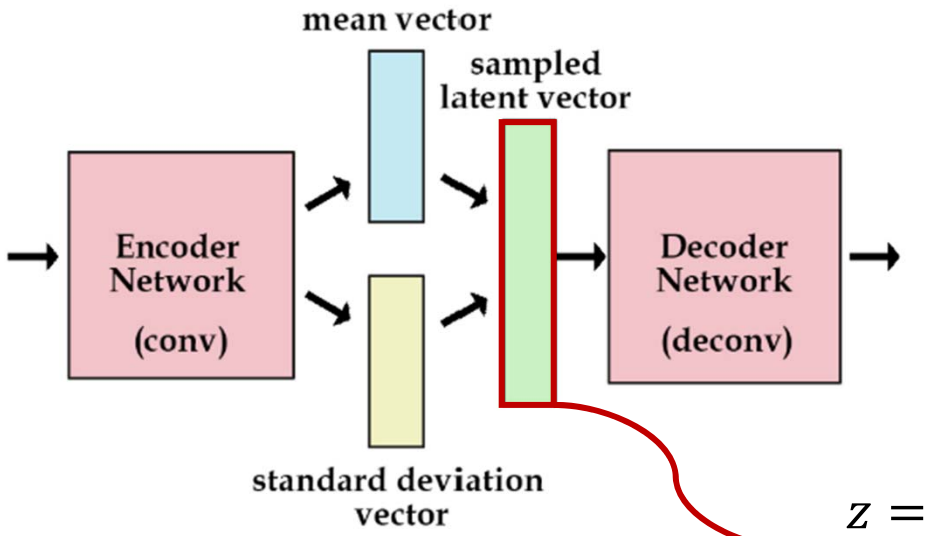
$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta})$ =重构误差 + 正则项

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 模型优化
  - 问题：**Sampling layer无法后向传递梯度!**

解决方案：再参数化 （**Reparameterization Trick** ）



$$z = \mu + \Sigma \odot \varepsilon$$

其中，$\varepsilon \sim \mathcal{N}(0,1)$

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

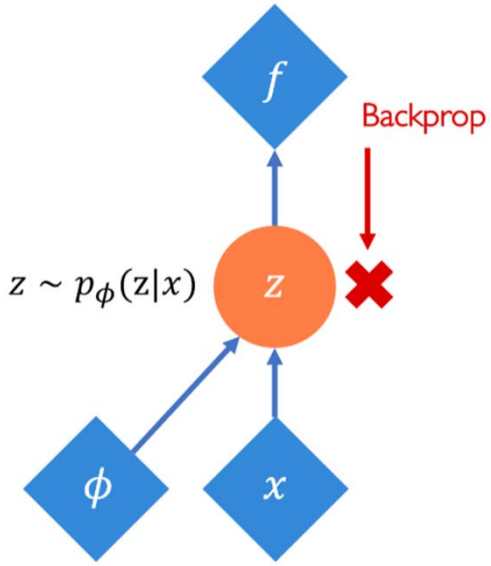
■ 模型优化

➢ 问题：**Sampling layer无法后向传递梯度!**

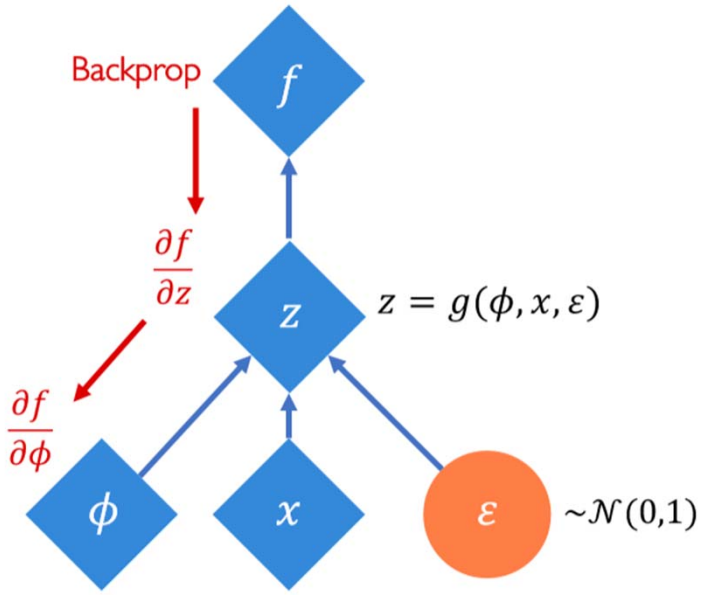
解决方案：对**Sampling layer**再参数化 （**Reparameterization Trick** ）

确定节点

随机节点



原始形式



再参数化形式

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

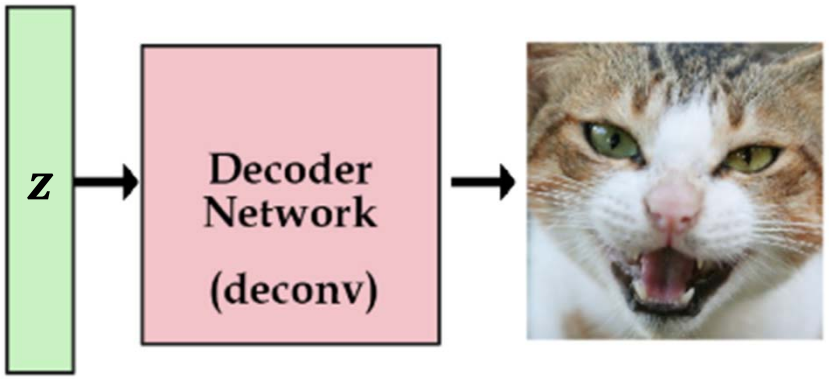
- 模型优化

```python
import tensorflow as tf

def gaussian_MLP_encoder(x, n_hidden, n_output, keep_prob):

    ...

    gaussian_params = tf.matmul(h1, wo) + bo
    # The mean parameter is unconstrained
    mean = gaussian_params[:, :n_output]
    # The standard deviation must be positive:
    stddev = 1e-6 + tf.nn.softplus(gaussian_params[:, n_output:])

    return mean, stddev

def autoencoder(x_hat, x, dim_img, dim_z, n_hidden, keep_prob):

    # encoding
    mu, sigma = gaussian_MLP_encoder(x_hat, n_hidden, dim_z, keep_prob)
    # sampling by re-parameterization technique
    z = mu + sigma * tf.random_normal(tf.shape(mu), 0, 1, dtype=tf.float32)
    # decoding
    y = bernoulli_MLP_decoder(z, n_hidden, dim_img, keep_prob)

    # loss
    marginal_likelihood = tf.reduce_sum(x * tf.log(y) + (1 - x) * tf.log(1 - y), 1)
    KL_divergence = 0.5 * tf.reduce_sum(tf.square(mu) + tf.square(sigma) - tf.log(1e-8 + tf.square(sigma)) - 1, 1)

    ELBO = tf.reduce_mean(marginal_likelihood) - tf.reduce_mean(KL_divergence)
    loss = -ELBO
```

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 数据生成
  - 从先验分布中随机抽取$\boldsymbol{z}$送入Decoder中，则可随机生成相应的图像

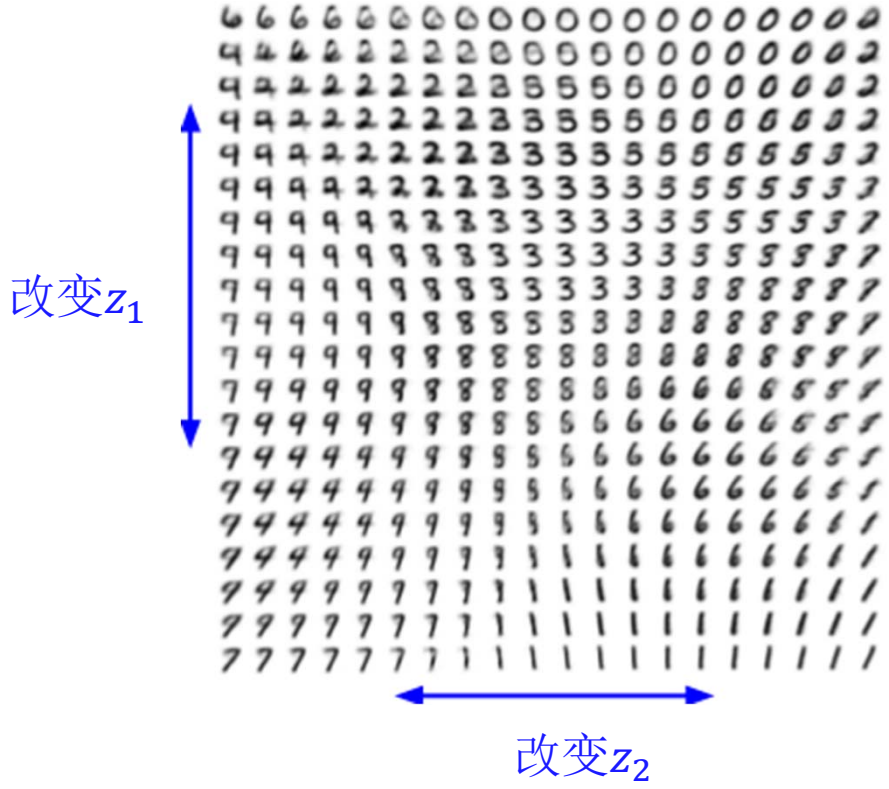


变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 数据生成

- 例：在手写数字数据集上训练，且隐含变量$\boldsymbol{z} = [z_1, z_2]^T$为二维向量



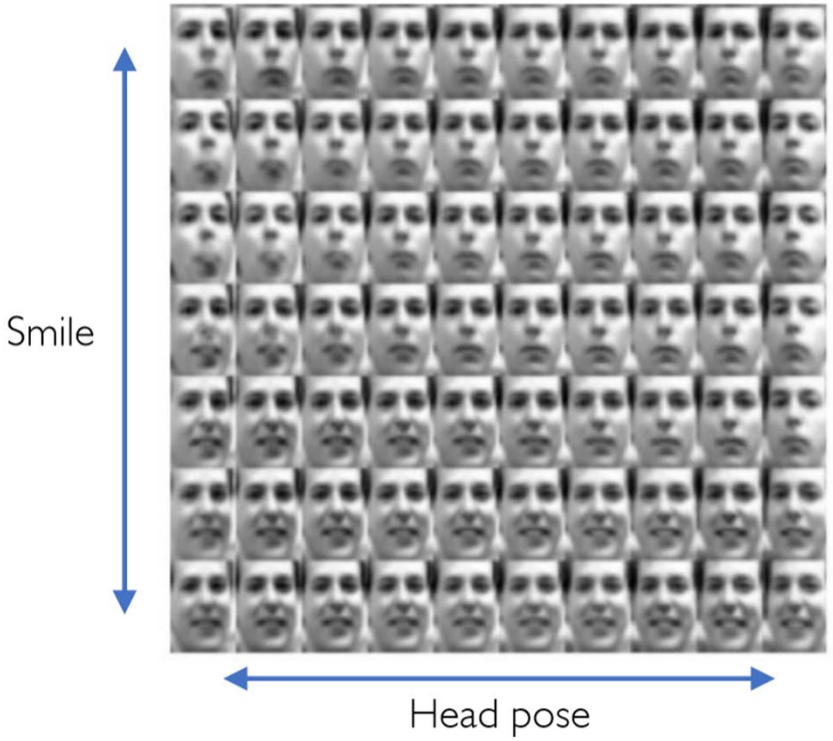
变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 数据生成

➢ Disentangled VAE

- 在隐含变量中，每个点之间彼此无关，各自学习输入数据中不同隐含特征
- 对$\boldsymbol{z}$增加约束，使每一维的隐含分布相互独立



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 数据生成

➢ Disentangled AVE

- 在隐含变量中，每个点之间彼此无关，各自学习输入数据中不同隐含特征
- 对$\boldsymbol{z}$增加约束，使每一维的隐含分布相互独立

对隐含变量的某一维度进行平滑扰动，观察相应所生成的数据，来理解该维度隐变量的物理含义

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 小结

- 自编码器 + 概率分布估计 => **生成模型**，可生成新的数据
- 隐含变量是对可观测数据的一种**压缩表征**
- 通过重构的方式进行**无监督学习**，降低对数据标注的要求
- 隐含变量的各维度对应着**可解释的物理含义**

Variational AutoEncoder

生成对抗网络（GAN）

大连理工大学 Dalian University Of Technology

■ 小结

- 关注于数据生成问题，不涉及对数据分布的显式建模问题
- 借助博弈思想：通过“两方博弈”，从训练数据分布中学习生成新数据

- 优点与不足：

  --优：能生成非常生动清晰的样本

  --缺：训练过程不易稳定，需要更多的训练技巧

- 活跃研究领域：

  --寻求更好的损失函数，更稳定的训练策略 (Wasserstein GAN, LSGAN等)

  --广泛的应用 (图像转换、动漫生成、语音识别、去雾等)

04

PART FOUR

小结和资源

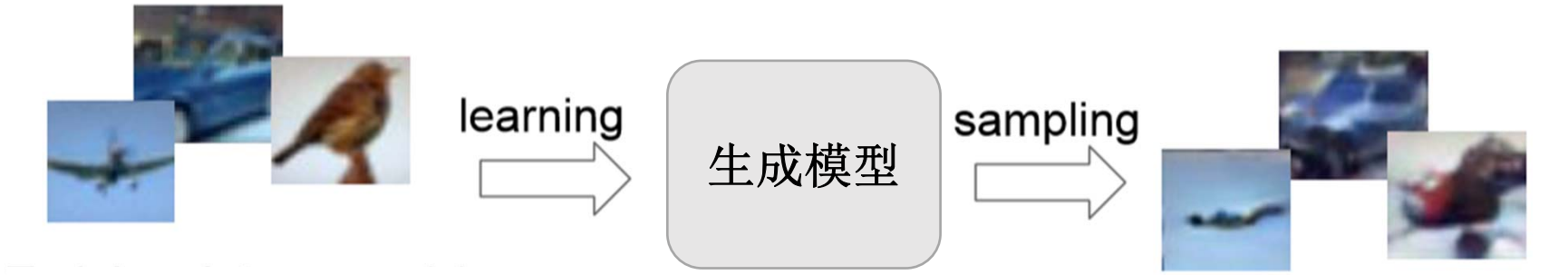
Take Home Message and Resource

小结（Take Home Message）

大连理工大学 Dalian University Of Technology

■ 生成网络

- 希望模型生成的新数据与训练数据的真实分布尽可能地相似



训练数据 $\sim p_{\text{data}}(x)$

生成样本 $\sim p_{\text{model}}(x)$

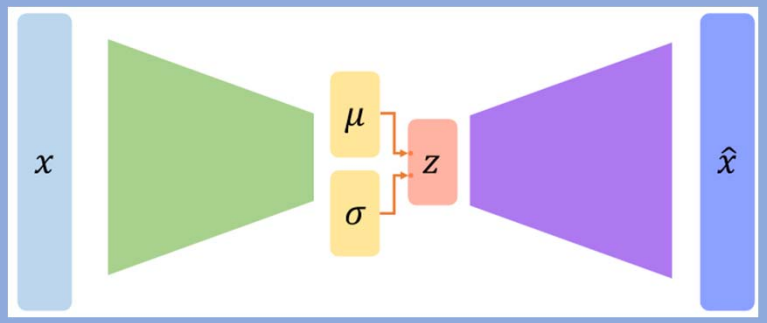
$p_{\text{model}} \approx p_{data}$

小结（Take Home Message）

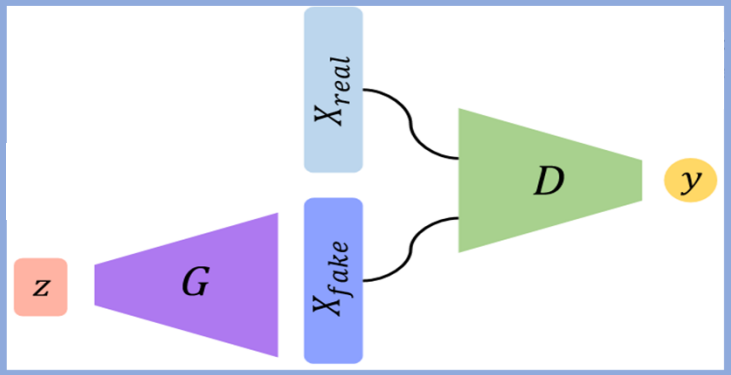
大连理工大学 Dalian University Of Technology

- 两种生成模型

变分自编码器（Variational Autoencoder, AVE）



生成对抗网络（Generative Adversarial Network, GAN）



小结（Take Home Message）

大连理工大学 Dalian University Of Technology

■ 变分自编码器（VAE）

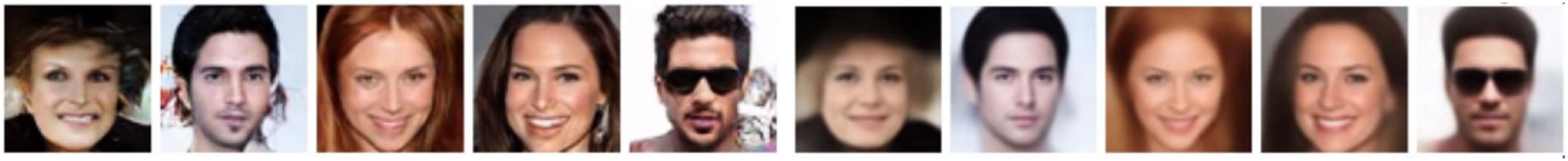
普通的自编码器 （Normal AutoEncoder）	变分自编码器 （Variational AutoEncoder）
• 作用 -- 学习输入数据的隐含表征 • 只能重构数据，无法生成新数据 • 优化： -- min 重构误差	• 作用 -- 学习生成新的可观测数据 • 优化： -- min (重构误差+隐变量的约束项) -- 隐变量从正态分布中随机抽取

小结（Take Home Message）

大连理工大学 Dalian University Of Technology

■ 变分自编码器（VAE）

生成对抗网络 （Generative Adversarial Network）	变分自编码器 （Variational AutoEncoder）
• 隐式密度估计 • 网络结构： -- Generator + Discriminator • 如何学习： -- Minmax Game 交替训练	• 显式密度估计 • 网络结构： -- Encoder + Decoder • 如何学习： -- min (重构误差+隐变量的约束项)



GAN生成图像的质量更清晰！

资源参考（Resource and Reference）

大连理工大学 Dalian University Of Technology

➢ Paper

[1] Karras et al., A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR2019.

[2] Goodfellow, *et al*., Generative adversarial nets. NIPS2014.

[3] Ledig *et al*., Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. CVPR2017.

[4] Isola *et al*., Image-to-Image Translation with Conditional Adversarial Networks. CVPR2017.

[5] C Doersch, Tutorial on Variational Autoencoders. arXiv abs/1606.05908.

[6] Higgins *et al*., β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. ICLR2017.

[7] Radford et al., Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, ICLR 2016.

➢ Other useful links ZOO

https://github.com/hindupuravinash/the-gan-zoo **GAN ZOO**

https://github.com/soumith/ganhacks **Tips for Training GANs**

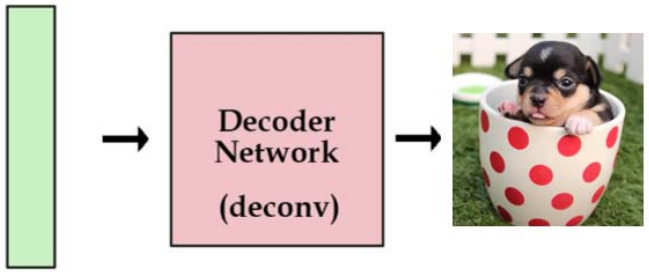
VAE

大连理工大学 Dalian University Of Technology

- **VAE优化密度函数（似然函数）**

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

  - 通过附加的隐变量 z 对密度函数进行建模
  - 全概率公式，对所有可能的 z 求期望
  - 最大化该似然函数



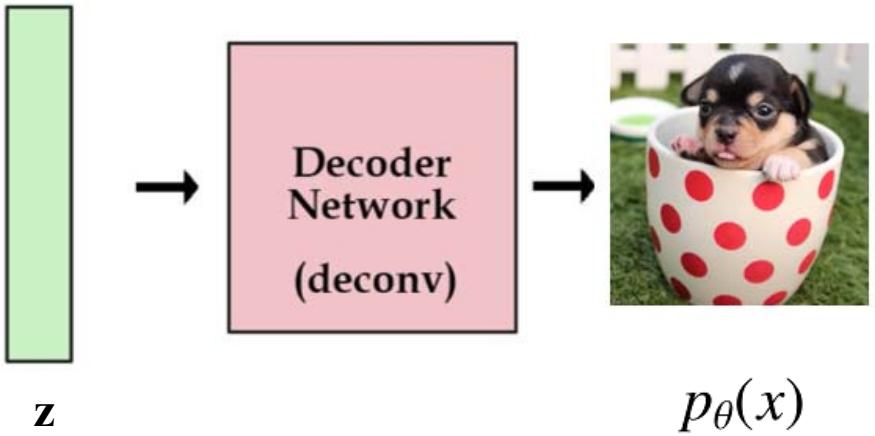
VAE

大连理工大学 Dalian University Of Technology

- **VAE优化密度函数（似然函数）**

$$p_\theta(x) = \int p_\theta(z) p_\theta(x|z) dz$$

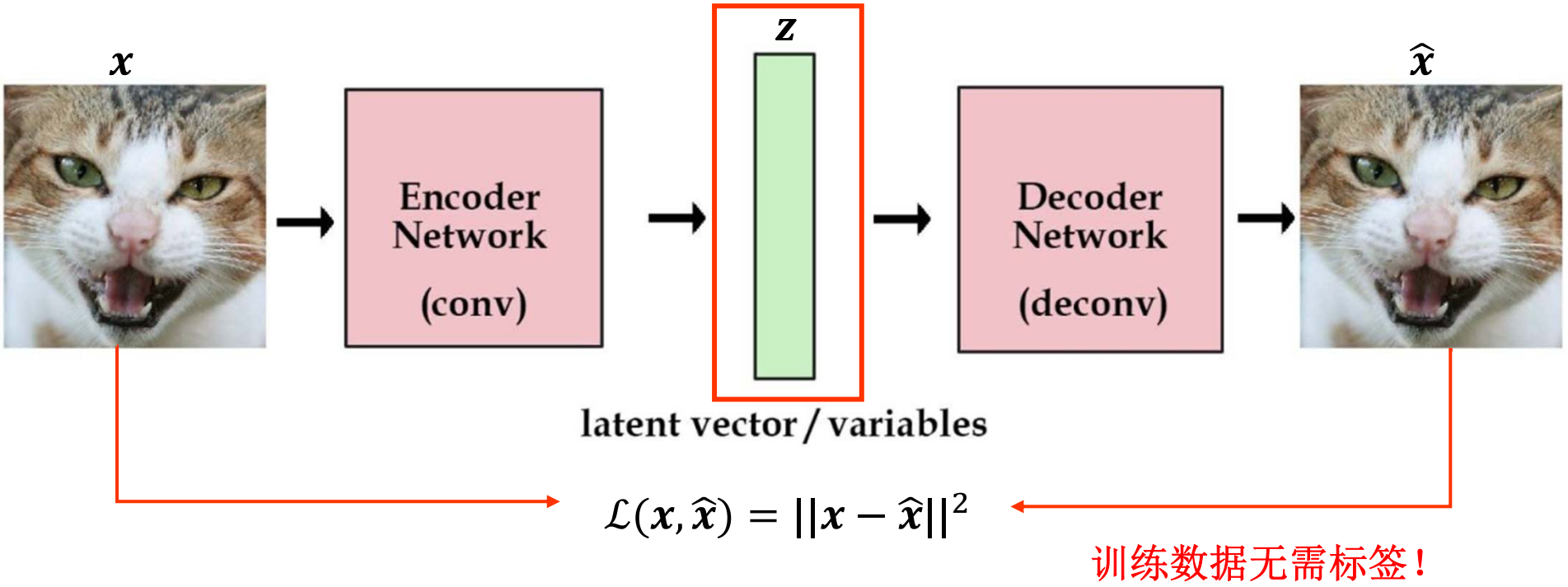
  - 通过附加的隐变量 **z** 对密度函数进行建模
  - 全概率公式，对所有可能的 **z** 求期望
  - 最大化该似然函数



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

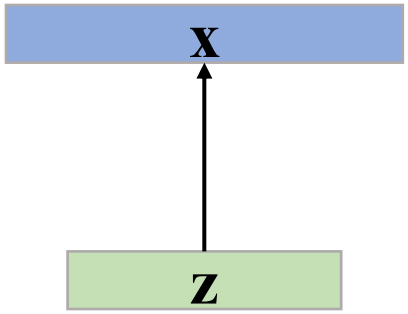
- 自编码器：能够重建数据



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 变分自编码器：在自编码器中引入随机因子获得的一种模型，从该模型中采样从而生成新数据
- 训练数据 $\{x^{(i)}\}_{i=1}^{N}$ 由隐变量 **z** 生成



直观上讲：z的每个维度是训练数据中的某个属性的变化程度，如脸的朝向、笑容程度、头发多少等

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

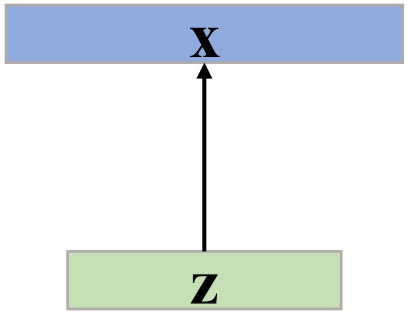
- 变分自编码器：在自编码器中引入随机因子获得的一种模型，从该模型中采样从而生成新数据
- 训练数据 $\{x^{(i)}\}_{i=1}^{N}$ 由隐变量 **z** 生成

从真实条件分布中采样

$p_{\theta^*}(x|z^{(i)})$

从真实先验中采样

$p_{\theta^*}(z)$



直观上讲：**z**的每个维度是训练数据中的某个属性的变化程度，如脸的朝向、笑容程度、头发多少等

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

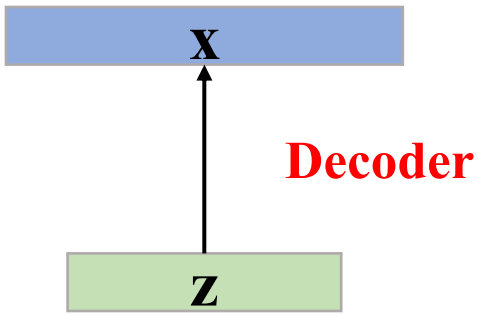
- 目标：估计该生成模型中的参数 θ^*
- 选择简单的先验 $p(z)$ 即可，例如高斯分布
- 估计 $p(x|z)$ 很复杂，用NN对其进行建模

从真实条件分布中采样

$p_{\theta^*}(x|z^{(i)})$

从真实先验中采样

$p_{\theta^*}(z)$



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

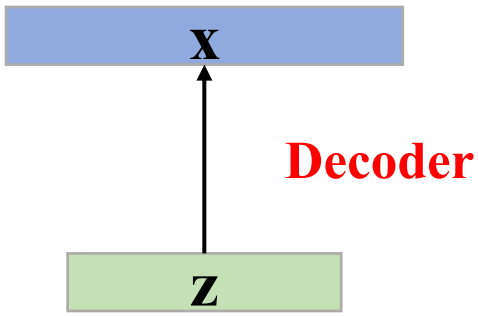
- 问题：该积分似然函数的最大化很难求解

从真实条件分布中采样

$p_{\theta^*}(x|z^{(i)})$

从真实先验中采样

$p_{\theta^*}(z)$



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z) p_\theta(x|z) dz$$

  - 高斯函数

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z) p_\theta(x|z) dz$$

  - 高斯函数
  - 神经网络

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

  - 高斯函数
  - 神经网络
  - 此积分难算

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

  - 高斯函数
  - 神经网络
  - 此积分难算
  - 导致后验概率也无法计算

$$p_\theta(z|x) = p_\theta(x|z)p_\theta(z)/p_\theta(x)$$

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

  - 高斯函数
  - 神经网络
  - 此积分难算
  - 导致后验概率也无法计算

$$p_\theta(z|x) = p_\theta(x|z)p_\theta(z)/p_\theta(x)$$

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

□ 解决方案

- 用**decoder**建模 $p_\theta(x|z)$
- 用另外一个网络 $q_\phi(z|x)$ **(encoder network)**建模 $p_\theta(z|x)$

$$p_\theta(z|x) = p_\theta(x|z)p_\theta(z)/p_\theta(x)$$

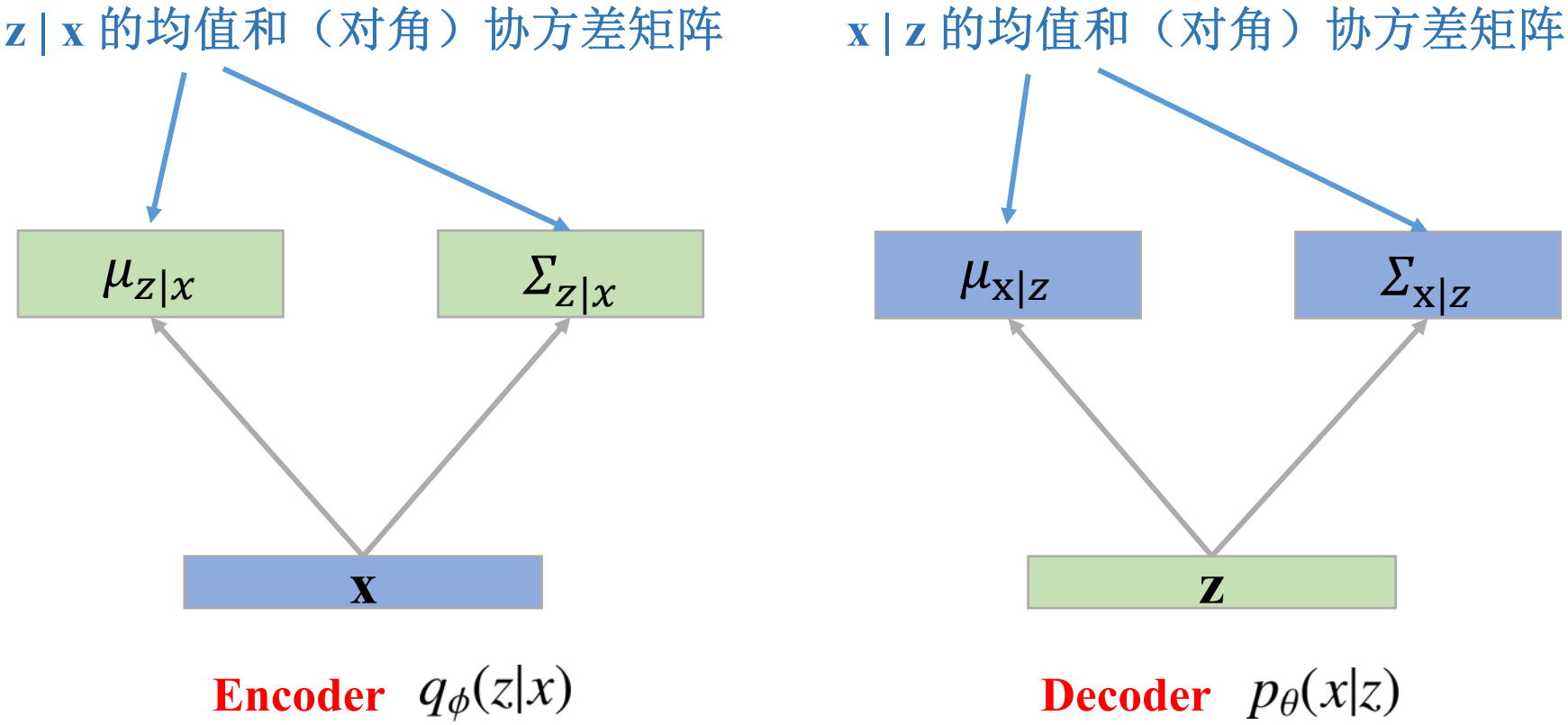
Encoder　　Decoder

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ **VAE网络：建模数据的生成概率，encoder和decoder也要是概率性模型**

即将随机因子引入网络

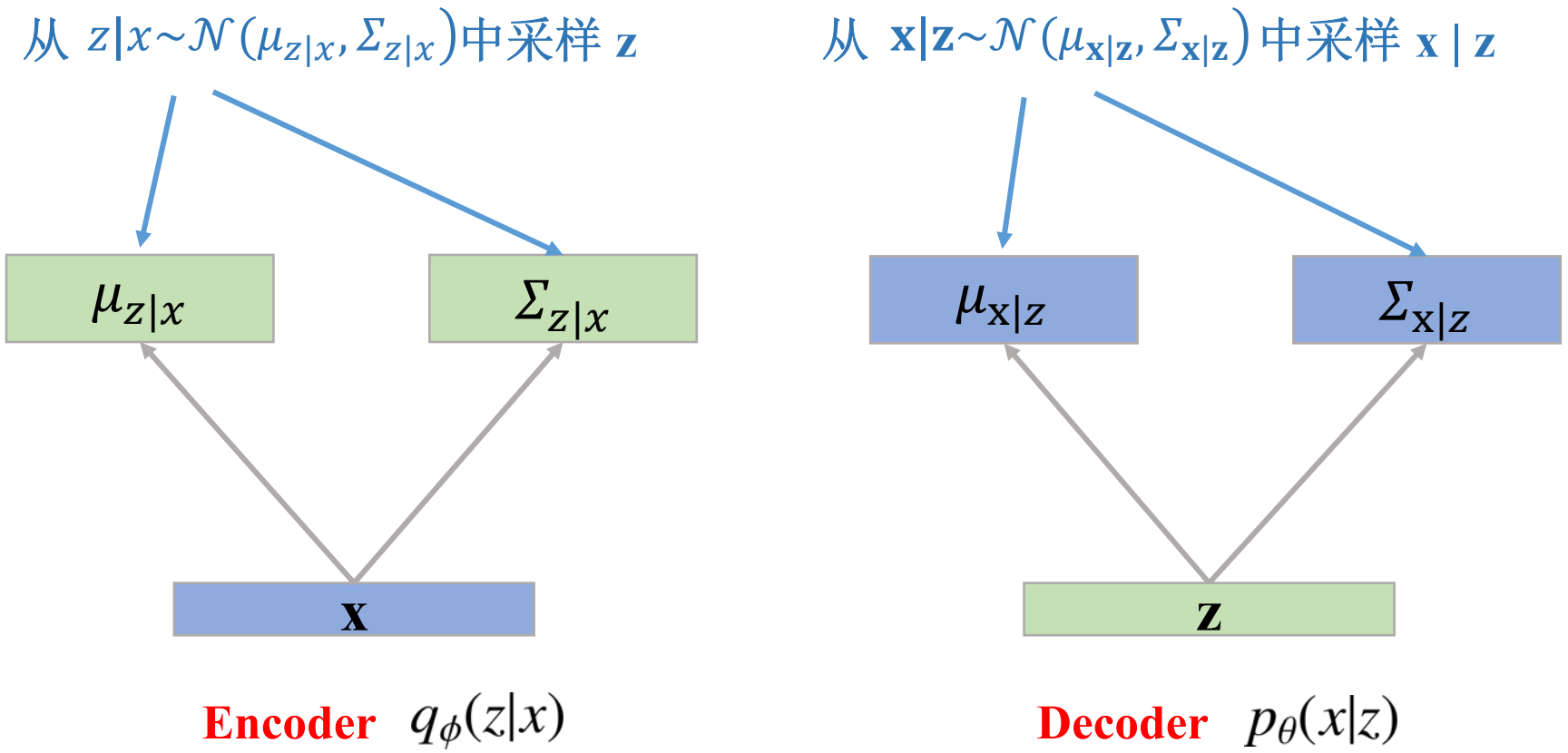


变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- **VAE网络：建模数据的生成概率，encoder和decoder也要是概率性模型**

  即将随机因子引入网络



变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

$$\log p_\theta(x^{(i)}) = \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)})]$$

$$= \mathbf{E}_z[\log \frac{p_\theta(x^{(i)} \mid z)p_\theta(z)}{p_\theta(z \mid x^{(i)})}]$$

贝叶斯公式

$$= \mathbf{E}_z[\log \frac{p_\theta(x^{(i)} \mid z)p_\theta(z)}{p_\theta(z \mid x^{(i)})} \frac{q_\theta(z \mid x^{(i)})}{q_\theta(z \mid x^{(i)})}]$$

$$= \mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - \mathbf{E}_z[\log \frac{q_\theta(z \mid x^{(i)})}{p_\theta(z)}] + \mathbf{E}_z[\log \frac{q_\theta(z \mid x^{(i)})}{p_\theta(x^{(i)} \mid z)}]$$

$$= \mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z)) + D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z \mid x^{(i)}))$$

Encoder的输出

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

$$\log p_\theta(x^{(i)}) = \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)})]$$

$$= \mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z)) + D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z \mid x^{(i)}))$$

  - $p_\theta(x^{(i)} \mid z)$ 由**Decoder**给出，通过采样可计算该项

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

$$\log p_\theta(x^{(i)}) = \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)})]$$

$$=\mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z)) + D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z \mid x^{(i)}))$$

  - $p_\theta(x^{(i)} \mid z)$ 由**Decoder**给出，通过采样可计算该项
  - **Encoder的输出（高斯函数）和 z 的先验间的KL散度，有闭形式解**

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

$$\log p_\theta(x^{(i)}) = \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)})]$$

$$= \mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z)) + D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z \mid x^{(i)}))$$

  - $p_\theta(x^{(i)} \mid z)$ 由**Decoder**给出，通过采样可计算该项
  - **Encoder**的输出（高斯函数）和 **z** 的先验间的**KL**散度，有闭形式解
  - $p_\theta(z \mid x^{(i)})$ 无法计算，此项不可解，但是此项一定大于零（由于**KL**散度的定义）

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

$$\log p_\theta(x^{(i)}) = \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)})]$$

$$= \underbrace{\mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - D_{KL}(q_\phi(z \mid x^{(i)}) \,\|\, p_\theta(z))}_{\mathcal{L}(x^{(i)},\theta,\phi)} + \underbrace{D_{KL}(q_\phi(z \mid x^{(i)}) \,\|\, p_\theta(z \mid x^{(i)}))}_{\geq 0}$$

似然函数的下界，可计算梯度并进行优化

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

$$\log p_\theta(x^{(i)}) = \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)})]$$

$$= \underbrace{\mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z))}_{\mathcal{L}(x^{(i)},\theta,\phi)} + \underbrace{D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z \mid x^{(i)}))}_{\geq 0}$$

训练模型：最优化似然函数的下界

$$\theta^*, \phi^* = \arg\max_{\theta,\phi} \sum_{i=1}^{N} \mathcal{L}(x^{(i)}, \theta, \phi)$$

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

- 训练模型：最大化似然函数，从而得到最优参数 θ^*

$$p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz$$

$$\log p_\theta(x^{(i)}) = \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}[\log p_\theta(x^{(i)})]$$

$$= \mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z)) + D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z \mid x^{(i)}))$$

重建输入数据

使隐变量z的后验概率与z的先验概率接近

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

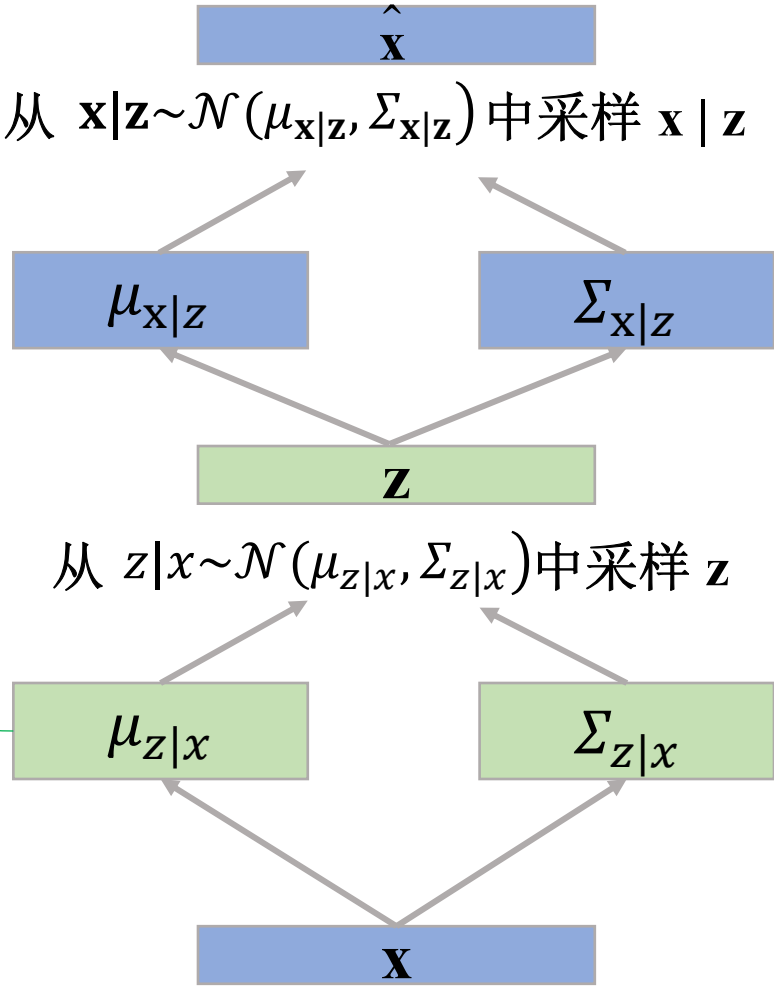
■ 训练模型：最大化似然函数的下界

最大化重建输入数据的似然函数

$$\mathbf{E}_z[\log p_\theta(x^{(i)} \mid z)] - D_{KL}(q_\phi(z \mid x^{(i)}) \| p_\theta(z))$$

$$\mathcal{L}(x^{(i)}, \theta, \phi)$$

使隐变量z的后验概率
与z的先验概率接近



Encoder $q_\phi(z|x)$

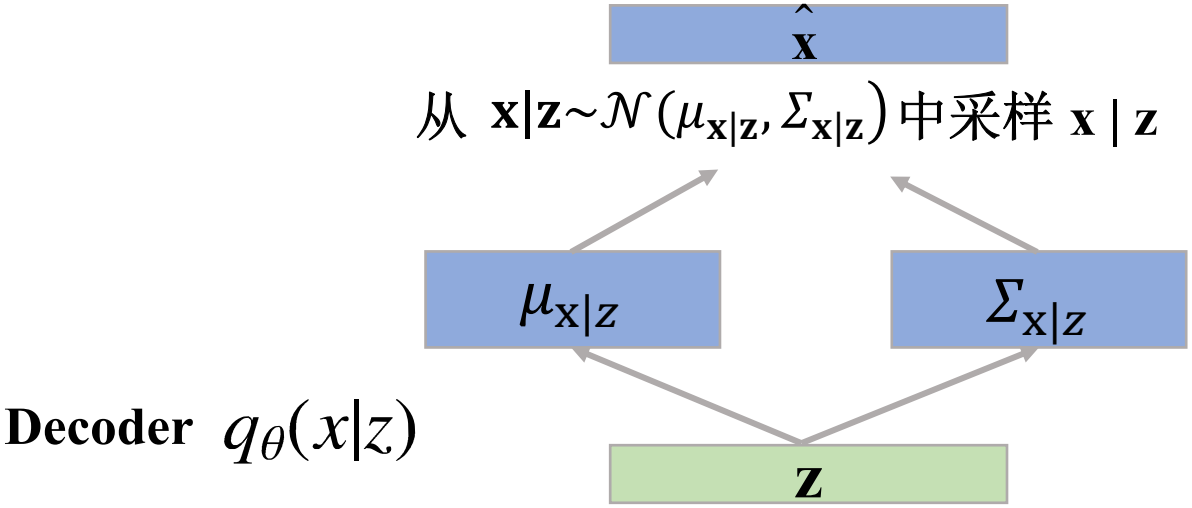
输入数据

变分自编码器（VAE）

大连理工大学 Dalian University Of Technology

■ 训练模型：最大化似然函数的下界

- ❑ 用Decoder生成图像
- ❑ 从z的先验分布中采样



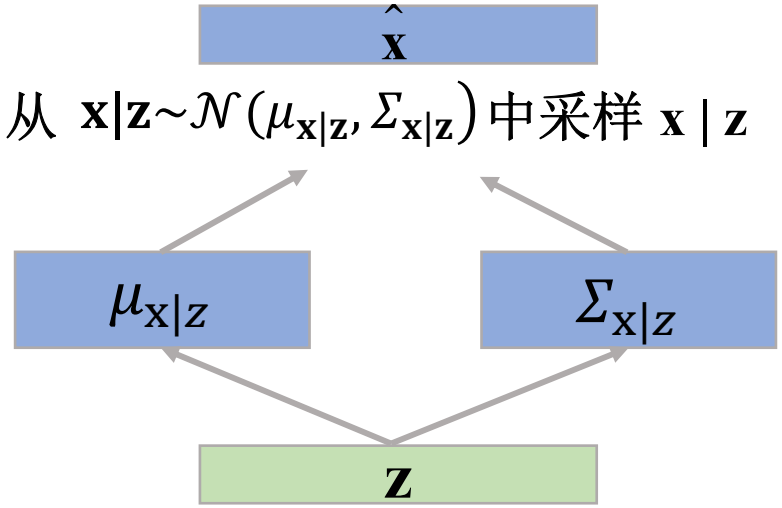
从 $z\sim\mathcal{N}(0,I)$ 中采样 $\mathbf{z}$

变分自编码器（VAE）

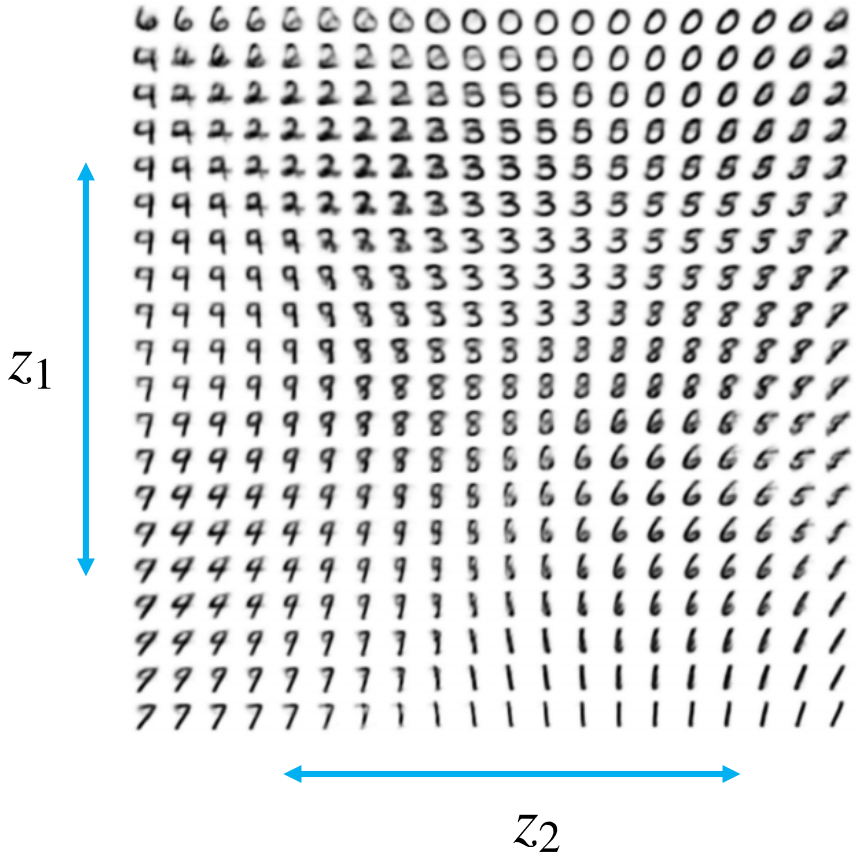
大连理工大学 Dalian University Of Technology

■ 训练模型：最大化似然函数的下界

- ❑ 用Decoder生成图像
- ❑ 从z的先验分布中采样



从 $z\sim\mathcal{N}(0,I)$ 中采样 $\mathbf{z}$



变分自编码器（VAE）

大连理工大学
Dalian University Of Technology

- **GAN**

变分自编码器（VAE）

大连理工大学
Dalian University Of Technology

- **GAN**