无监督深度学习

Unsurpervised Deep Learning

任课教师：孔雨秋

yqkong@dlut.edu.cn

目 录 (CONTENT)

01 PART ONE

无监督深度学习

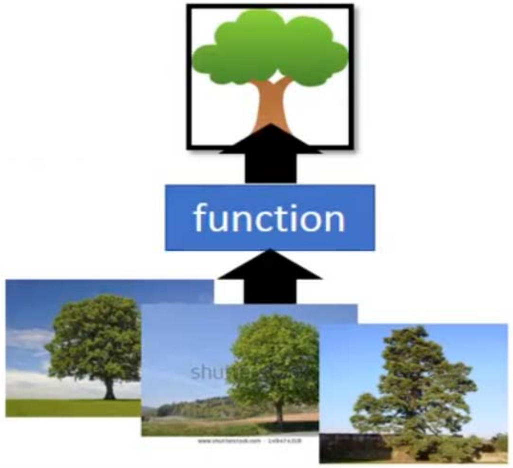
Unsupervised Deep Learning

无监督深度学习

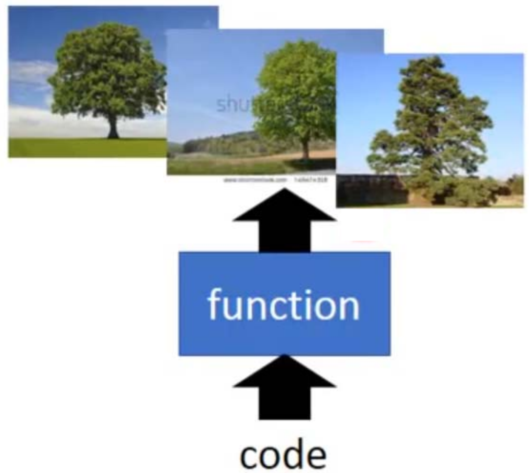
- 无监督学习(Unsupervised Learning， UL)是指从无标签的数据中学习出一些有用的模式。
  - 监督学习：建立输入-输出之间的映射关系
  - 无监督学习：发现隐藏的数据中的有价值信息，包括有效的特征、 类别、结构以及概率分布等

化繁为简
训练时只有输入数据

降维 →



无中生有
训练时只有输出数据



无监督深度学习

- 无监督学习(Unsupervised Learning， UL)是指从无标签的数据中学习出一些有用的模式。
  - 无监督特征学习：从无标签的训练数据中挖掘有效的特征或表示。用来进行降维、数据可视化或监督学习前期的数据预处理
  - 概率密度估计(Probabilistic Density Estimation): 简称密度估计，是根据一组训练样本来估计样本空间的概率密度。参数密度估计是假设数据服从某个已知概率密度函数形式的分布（比如高斯分布），然后根据训练样本去估计概率密度函数的参数。非参数密度估计是不假设数据服从某个已知分布， 只利用训练样本对密度进行估计，可以进行任意形状密度的估计。
  - 聚类是将一组样本根据一定的准则划分到不同的组(簇，cluster) 。一个比较通用的准则是组内样本的相似性要高于组间样本的相似性。

K-means算法、谱聚类等.

无监督深度学习

- 无监督学习三要素：模型、学习准则、优化算法
  - 无监督特征学习：最小重构误差
    - ✓ 对特征进行独立性、非负性或稀疏性约束等
  - 概率密度估计：最大似然估计

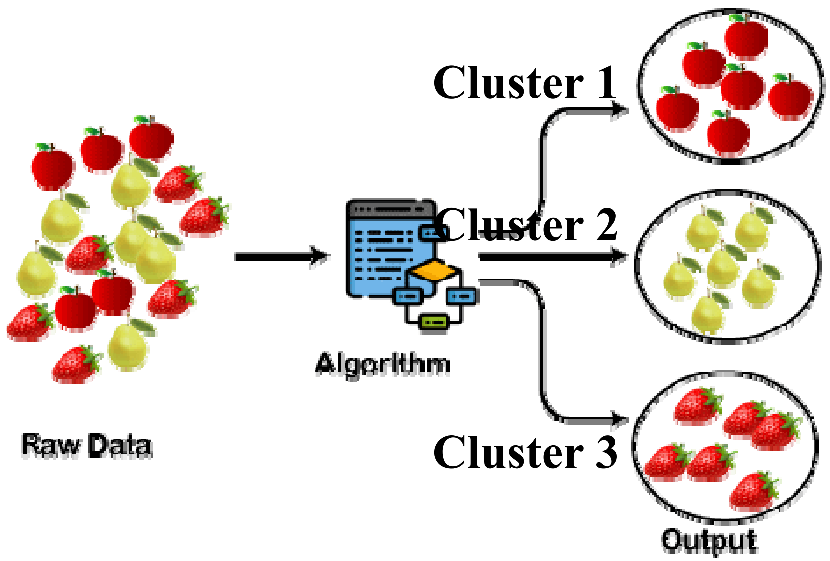
02 PART TWO

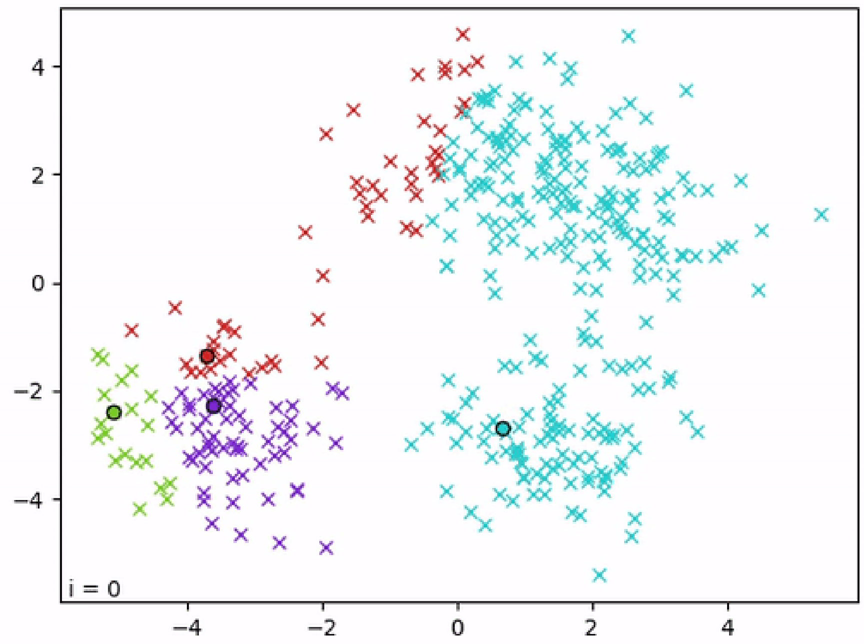
无监督特征学习

Unsupervised Feature Learning

聚类 Clustering

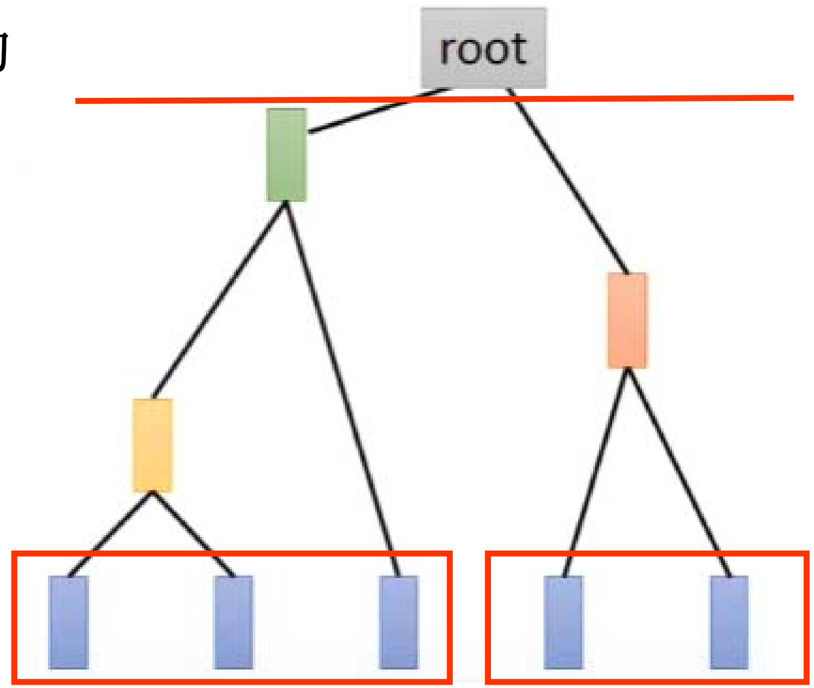
- 应该有几个簇(cluster)?
- **K-Means:** 按照经验选择簇的个数





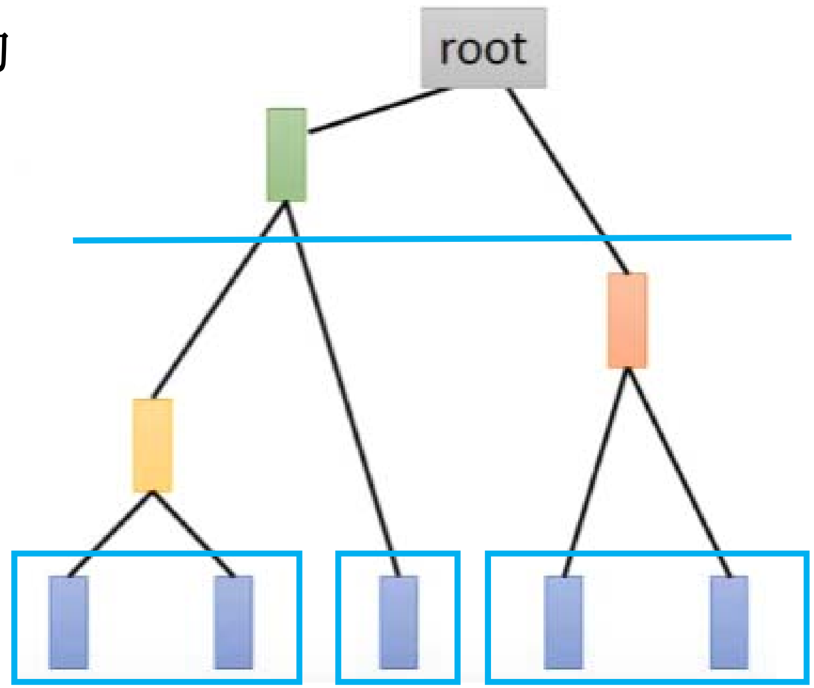
聚类 Clustering

- 应该有几个簇(cluster)?
- **Hierarchical Agglomerative Clustering (HAC)**
  - 构建一个树，对比每两个样本的相似性，并配对
  - 设置阈值，切一刀



聚类 Clustering

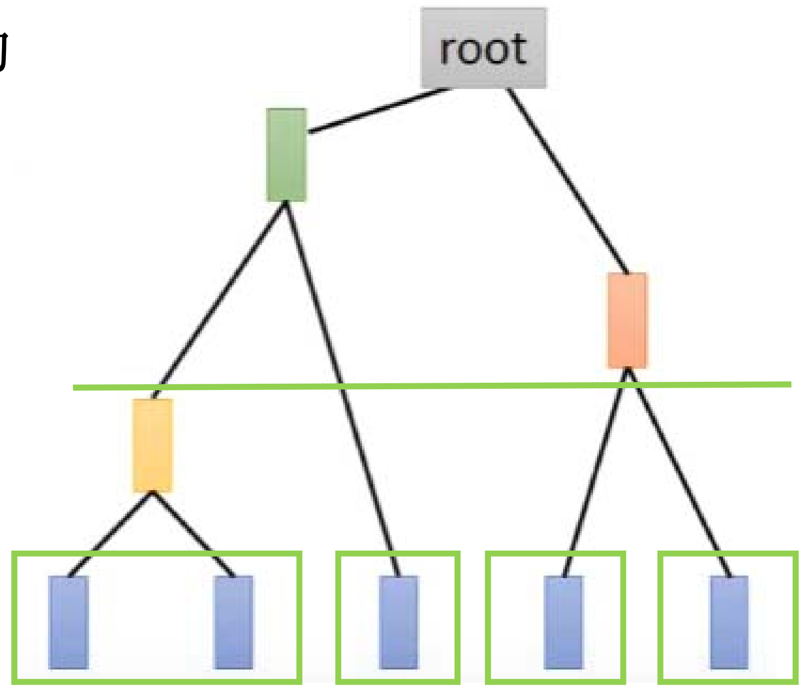
- 应该有几个簇(cluster)?
- **Hierarchical Agglomerative Clustering (HAC)**
  - 构建一个树，对比每两个样本的相似性，并配对
  - 设置阈值，切一刀



聚类 Clustering

大連理工大學 Dalian University of Technology

- 应该有几个簇(cluster)?
- **Hierarchical Agglomerative Clustering (HAC)** 无需自己设定
  - 构建一个树，对比每两个样本的相似性，并配对
  - 设置阈值，切一刀



分布式表示 Distributed Representation

大连理工大学 Dalian University of Technology

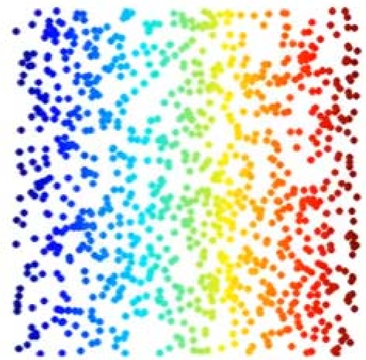
- 聚类：每个样本只能属于一个簇
- 分布式表示：降维表示

颜色	局部表示	分布式表示
琥珀色	$[1, 0, 0, 0]^T$	$[1.00, 0.75, 0.00]^T$
天蓝色	$[0, 1, 0, 0]^T$	$[0.00, 0.5, 1.00]^T$
中国红	$[0, 0, 1, 0]^T$	$[0.67, 0.22, 0.12]^T$
咖啡色	$[0, 0, 0, 1]^T$	$[0.44, 0.31\ 0.22]^T$

降维表示 Dimension Reduction

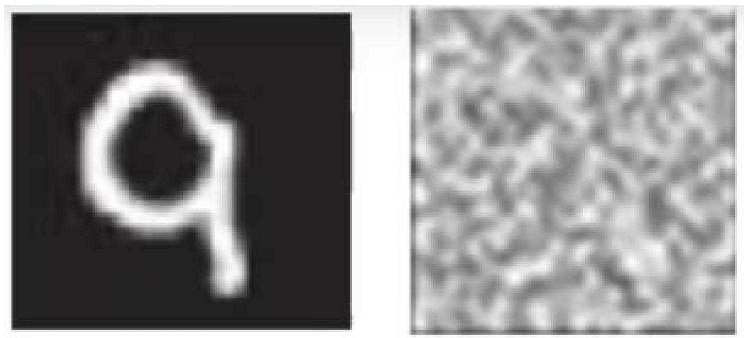
- 聚类：每个样本只能属于一个簇
- 分布式表示：降维表示
  - □ **3-D**空间 → **2-D**空间

3-D



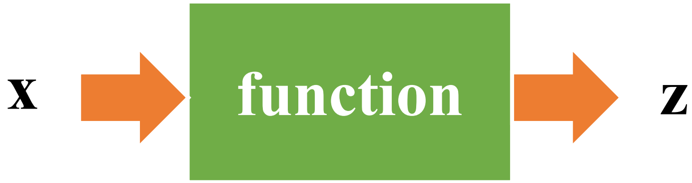
2-D

  - □ **MNIST**图像特征表示，**28*28**维特征通常没有任何意义 → 降维提取有用特征



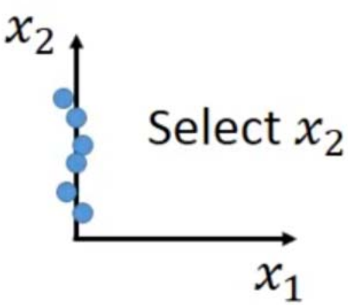
降维表示 Dimension Reduction

■ 降维



Z的维数小于x的维数

□ 特征选择



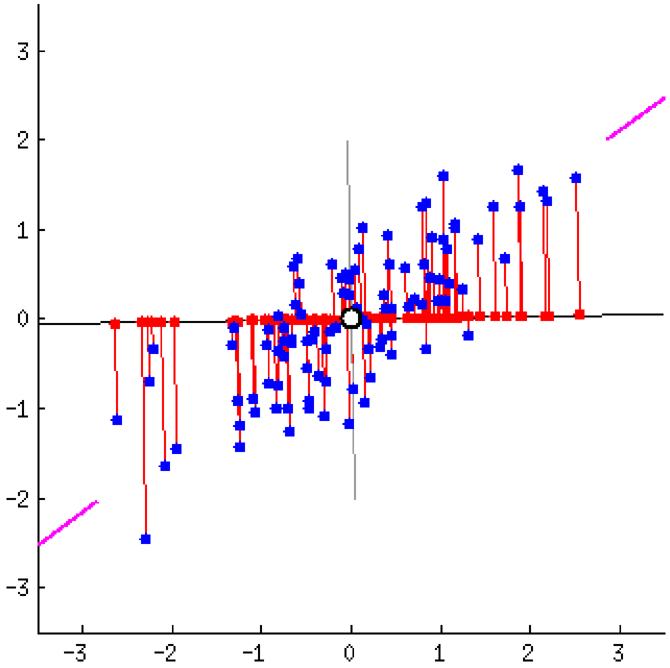
□ 主成分分析(Principle Component Analysis, PCA)

$$\mathbf{z} = \mathbf{W}\mathbf{x}$$

主成分分析

- 主成分分析(**Principal Component Analysis, PCA**):是一种最常用的数据降维方法， 使得在转换后的空间中数据的方差最大
  - 最大化方差
  - 最小化重构误差

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{x}$$ ← 线性降维



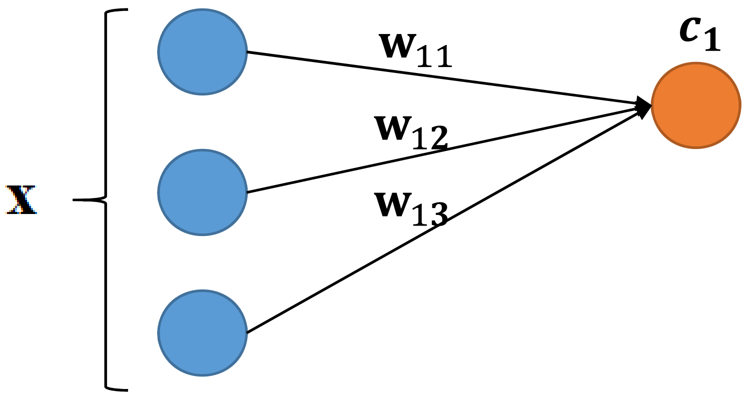
主成分分析

- **PCA VS. NN**
  - 最小化重构误差：如果$\{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_K\}$是主成分

重构结果：$\hat{\mathbf{x}} = \sum_{k=1}^{K} c_k \mathbf{w}_k$ $\xleftrightarrow{\text{最小化重建误差}}$ $\mathbf{x}$

$\Rightarrow$ $c_k = \mathbf{w}_k^T \cdot \mathbf{x}$

K=2：

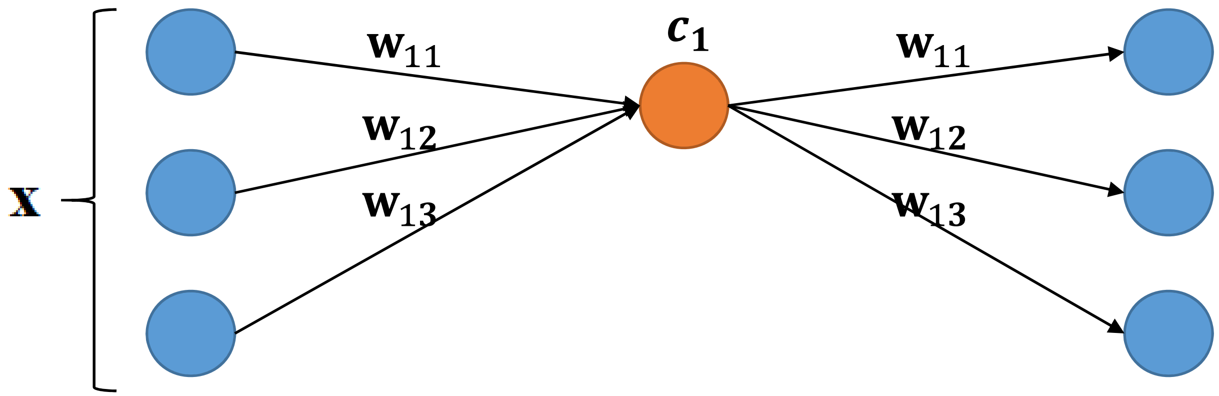


- **PCA VS. NN**
  - 最小化重构误差：如果$\{w_1, w_2, \cdots, w_K\}$是主成分

重构结果： $\hat{\mathbf{x}} = \sum_{k=1}^{K} c_k \mathbf{w}_k$ 最小化重建误差 $\mathbf{x}$

$$\Rightarrow c_k = \mathbf{w}_k^T \cdot \mathbf{x}$$

K=2：



■ PCA VS. NN

- □ 最小化重构误差：如果$\{w_1, w_2, \cdots, w_K\}$是主成分

重构结果：$\hat{\mathbf{x}} = \sum_{k=1}^{K} c_k \mathbf{w}_k$ ⟷ 最小化重建误差 ⟷ $\mathbf{x}$

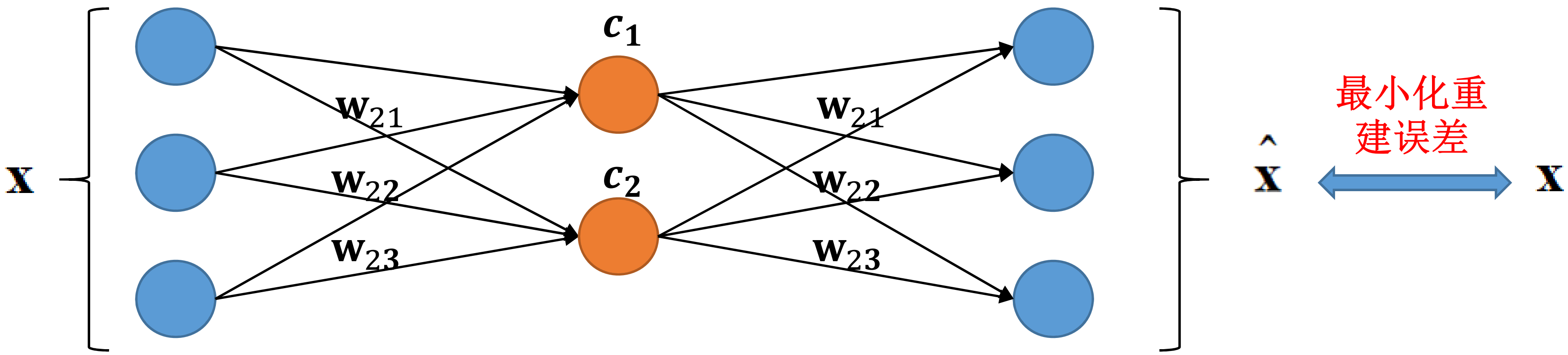
⇨ $c_k = \mathbf{w}_k^T \cdot \mathbf{x}$

Auto-encoder（自编码器）

⇧

PCA VS. 有一个隐层的NN

K=2：

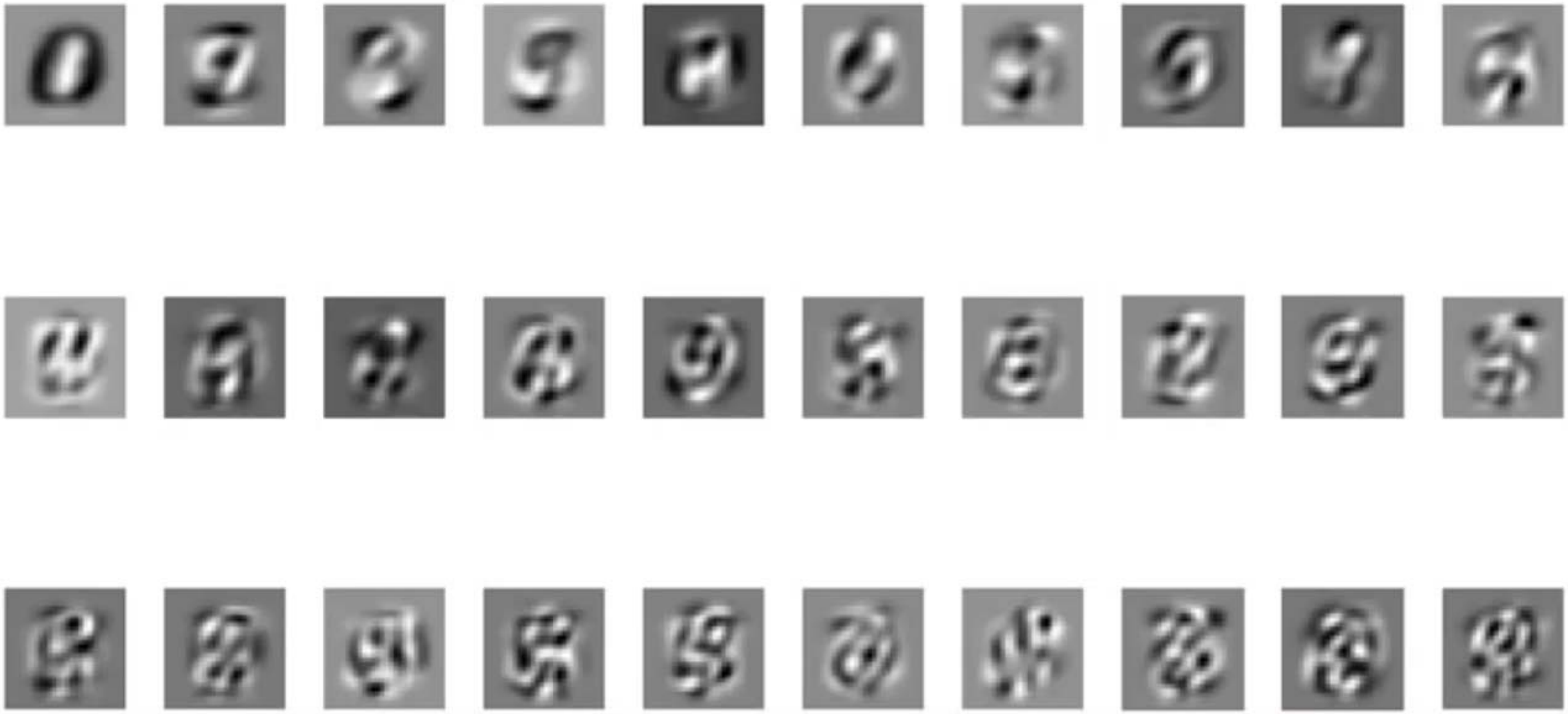


主成分分析

- **PCA on MNIST**
  - **30个主成分可视化**

$$= a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_K\mathbf{w}_K$$

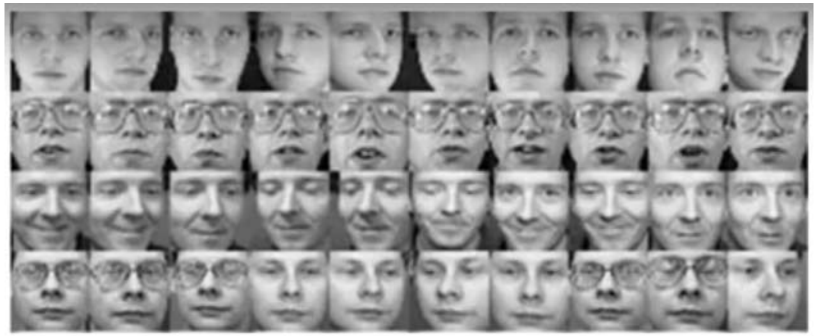
Eigen Digits



主成分分析

- PCA on MNIST
  - 30个主成分可视化

Eigen Faces



主成分分析

$$= a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_K\mathbf{w}_K$$

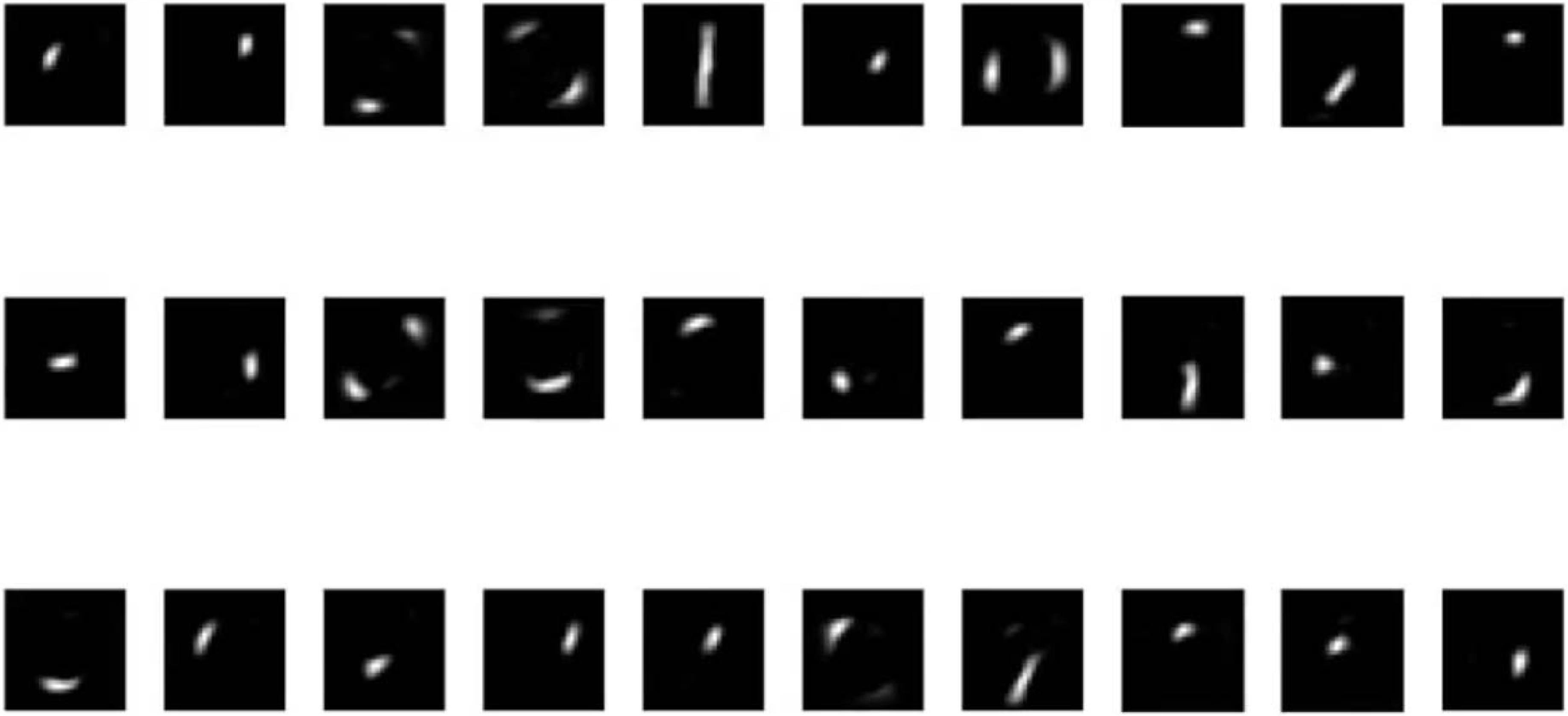
可正可负

- **PCA：** 可加/减主成分
- **Non-Negative Matrix Factorization (NMF)**
  - $\boldsymbol{a}_1, \boldsymbol{a}_2 \cdots$是非负的
  - $\mathbf{w}_1, \mathbf{w}_2 \cdots$所有元素是非负的

Algorithms for non-negative matrix factorization. NIPS 2001

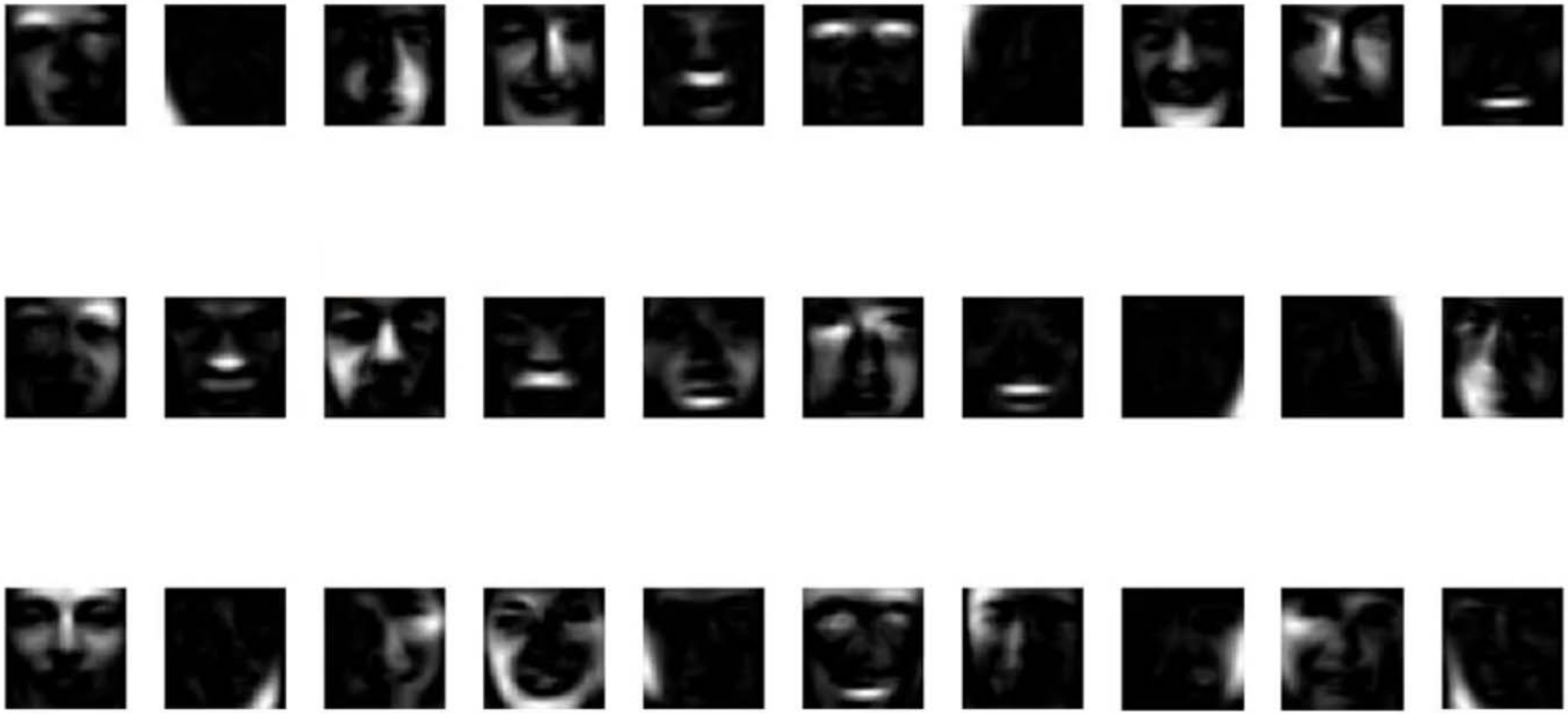
主成分分析

- NMF on MNIST



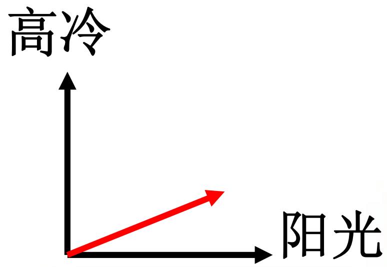
主成分分析

■ **NMF on Face**



矩阵因子分解

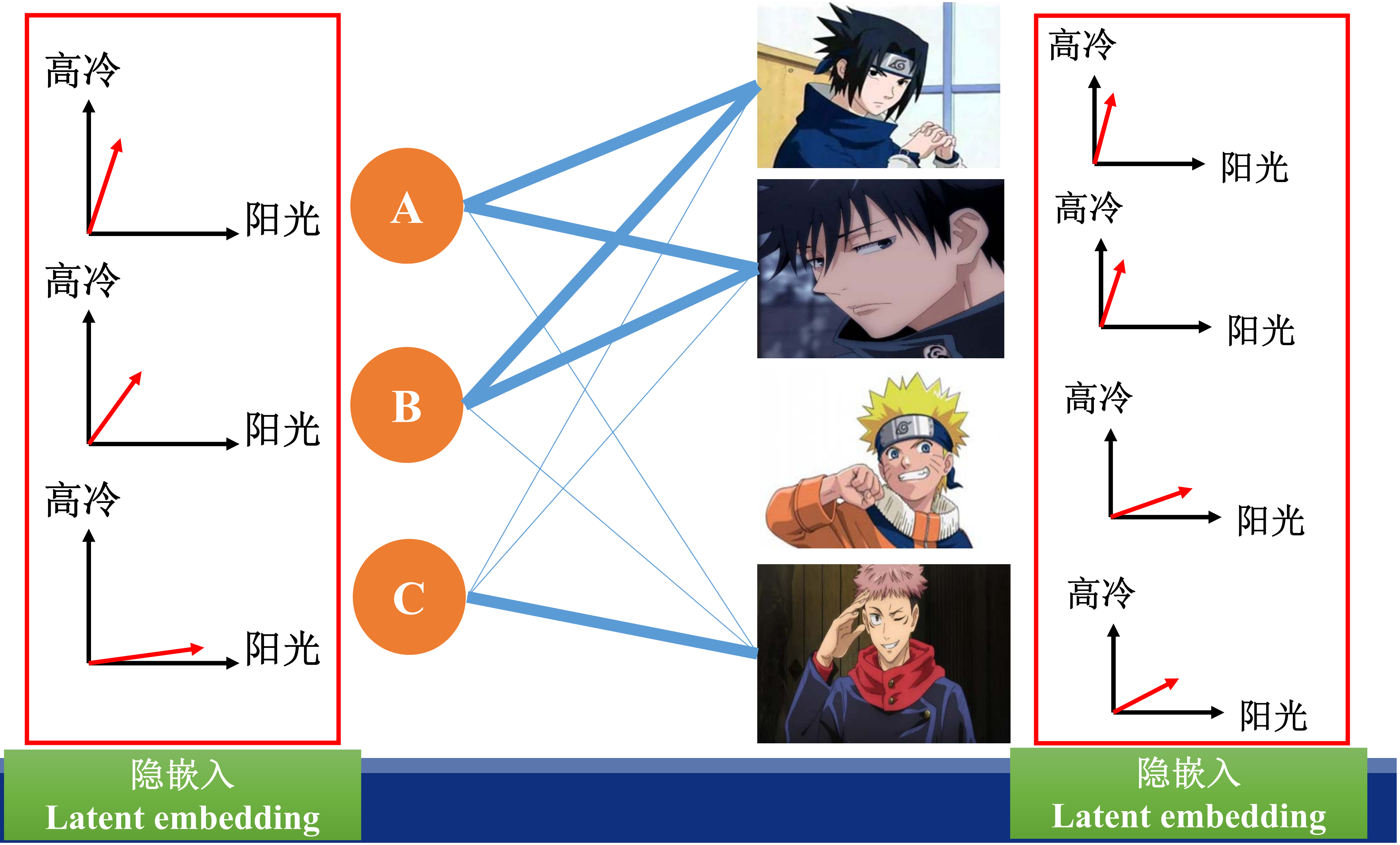
- 矩阵因子分解(**Matrix Factorization**)：隐式嵌入模型，常用于推荐系统



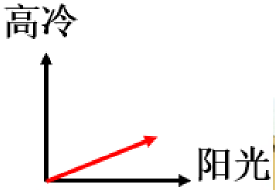
				
A	5	3	0	1
B	4	3	0	1
C	1	1	0	5
D	1	0	4	4
E	0	1	5	4

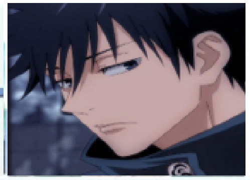
矩阵因子分解

■ 矩阵因子分解(Matrix Factorization)：隐式嵌入模型，常用于推荐系统



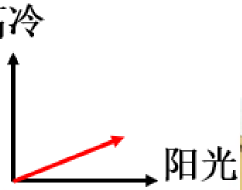
矩阵因子



		r_1	r_2	r_3	r_4
					
r_A	A	5	3	0	1
r_B	B	4	3	0	1
r_C	C	1	[illegible]	[illegible]	5
r_D	D	1	0	4	4
r_E	E	0	1	5	4

反馈矩阵 X

矩阵因子分



		$\mathbf{r}_1$	$\mathbf{r}_2$	$\mathbf{r}_3$	$\mathbf{r}_4$
$\mathbf{r}_A$	A	5	3	0	1
$\mathbf{r}_B$	B	4	3	0	1
$\mathbf{r}_C$	C	1	[illegible]	[illegible]	5
$\mathbf{r}_D$	D	1	0	4	4
$\mathbf{r}_E$	E	0	1	5	4

反馈矩阵 X

$$\mathbf{r}_A \cdot \mathbf{r}_1 \approx 5$$

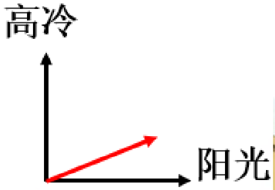
$$\mathbf{r}_B \cdot \mathbf{r}_1 \approx 4$$

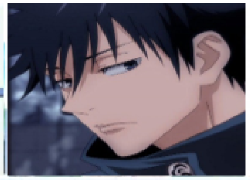
$$\mathbf{r}_C \cdot \mathbf{r}_1 \approx 1$$

$$\vdots$$

$$\begin{bmatrix} n_{A1} & n_{A2} & \\ n_{B1} & n_{B2} & \\ & & \end{bmatrix} \approx \mathrm{M} \begin{array}{c} \mathrm{K} \\ \begin{bmatrix} \mathbf{r}_A \\ \mathbf{r}_B \\ \vdots \end{bmatrix} \end{array} \times \mathrm{K} \begin{array}{c} \mathrm{N} \\ \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots \end{bmatrix} \end{array}$$

矩阵因子



	$\mathbf{r}_1$ 	$\mathbf{r}_2$ 	$\mathbf{r}_3$ 	$\mathbf{r}_4$ 
$\mathbf{r}_A$ A	5	3	0	1
$\mathbf{r}_B$ B	4	3	0	1
$\mathbf{r}_C$ C	1			5
$\mathbf{r}_D$ D	1	0	4	4
$\mathbf{r}_E$ E	0	1	5	4

反馈矩阵 X

$\mathbf{r}_A \cdot \mathbf{r}_1 \approx 5$

$\mathbf{r}_B \cdot \mathbf{r}_1 \approx 4$

$\mathbf{r}_C \cdot \mathbf{r}_1 \approx 1$

$\vdots$

n_{A1} n_{A2}

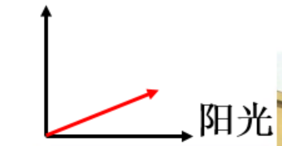
n_{B1} n_{B2}

反馈矩阵 X $\approx$ M [K: $\mathbf{r}_A$, $\mathbf{r}_B$] x K [N: $\mathbf{r}_1$, $\mathbf{r}_2$]

最小化重构误差

SVD分解求解

矩阵因子



		$\mathbf{r}_1$ 	$\mathbf{r}_2$ 	$\mathbf{r}_3$ 	$\mathbf{r}_4$ 
$\mathbf{r}_A$	A	5	3	?	1
$\mathbf{r}_B$	B	4	3	?	1
$\mathbf{r}_C$	C	1	1	?	5
$\mathbf{r}_D$	D	1	?	4	4
$\mathbf{r}_E$	E	?	1	5	4

- 反馈矩阵中有数据缺失，无法进行**SVD**
- 可忽略缺失部分，优化损失函数，实现特征嵌入

$$\mathbf{r}_A \cdot \mathbf{r}_1 \approx 5$$

$$\mathbf{r}_B \cdot \mathbf{r}_1 \approx 4$$

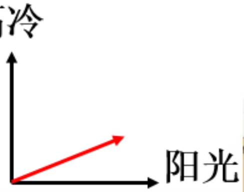
$$\mathbf{r}_C \cdot \mathbf{r}_1 \approx 1$$

$$\vdots$$

$$\min L = \sum_{(i,j)} \left(\mathbf{r}_i \cdot \mathbf{r}_j - n_{ij}\right)^2$$

SGD求解

矩阵因子



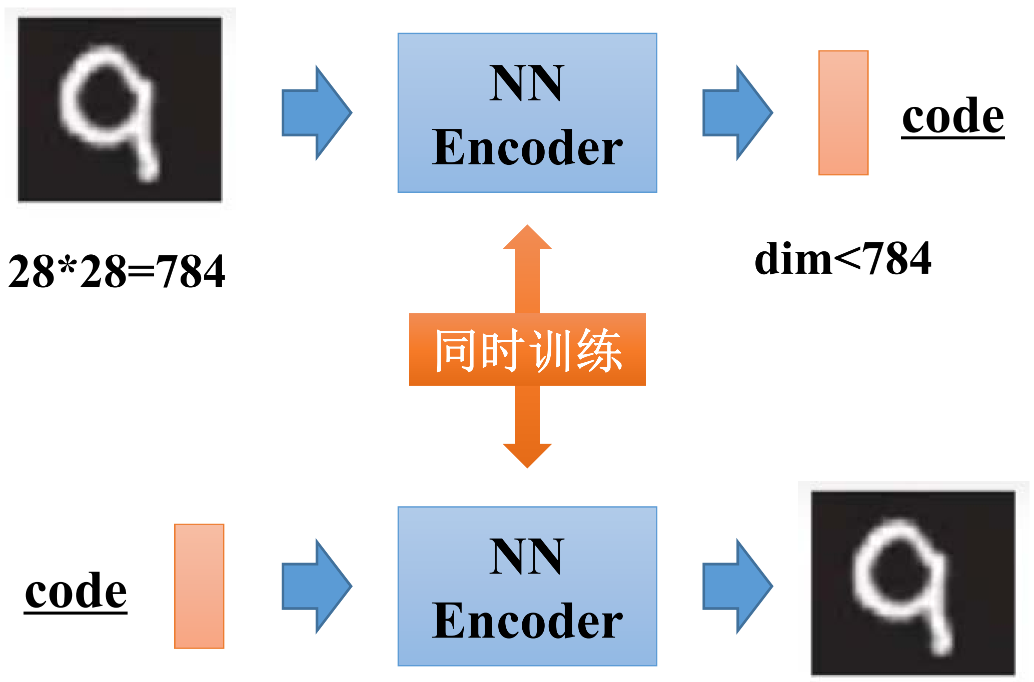
		r_1	r_2	r_3	r_4
r_A	A	5 n_{A1}	3	-0.4	1
r_B	B	4	3	-0.3	1
r_C	C	1	1	2.2	5
r_D	D	1	0.6	4	4
r_E	E	0.1	1	5	4

- 假设隐嵌入特征维数为2

A	0.2	2.1
B	0.2	1.8
C	1.3	0.7
D	1.9	0.2
E	2.2	0.0

1	0.0	2.2
2	0.1	1.5
3	1.9	-0.3
4	2.2	0.5

自编码 Auto-encoder

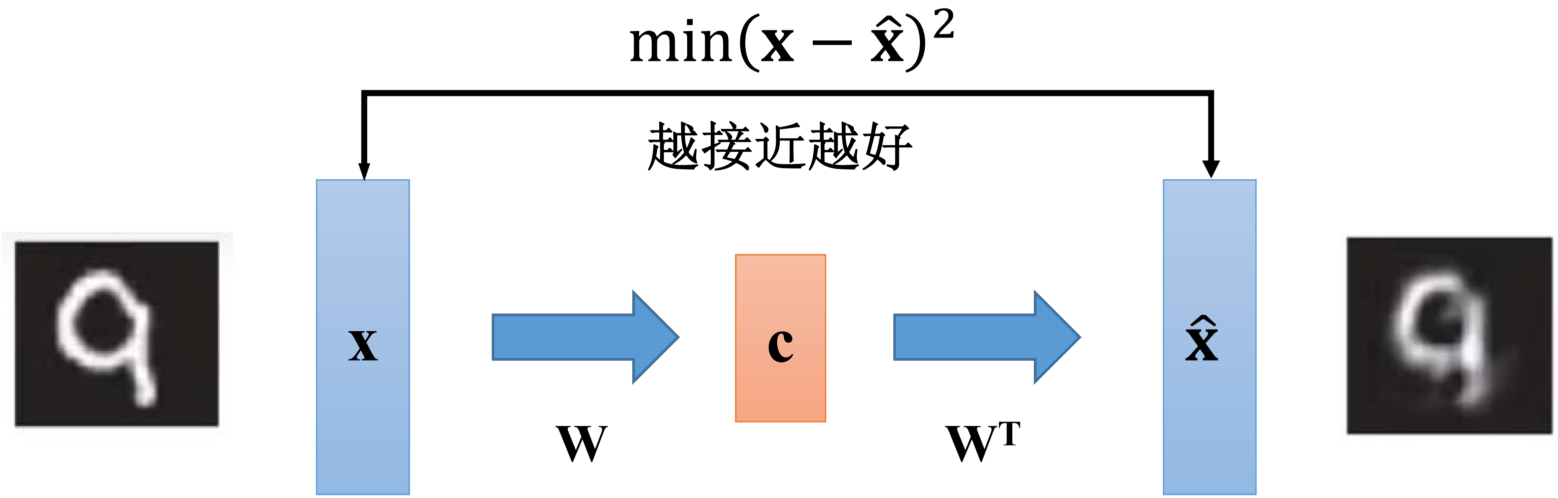


输入图像的降维表示

重建回原始图像

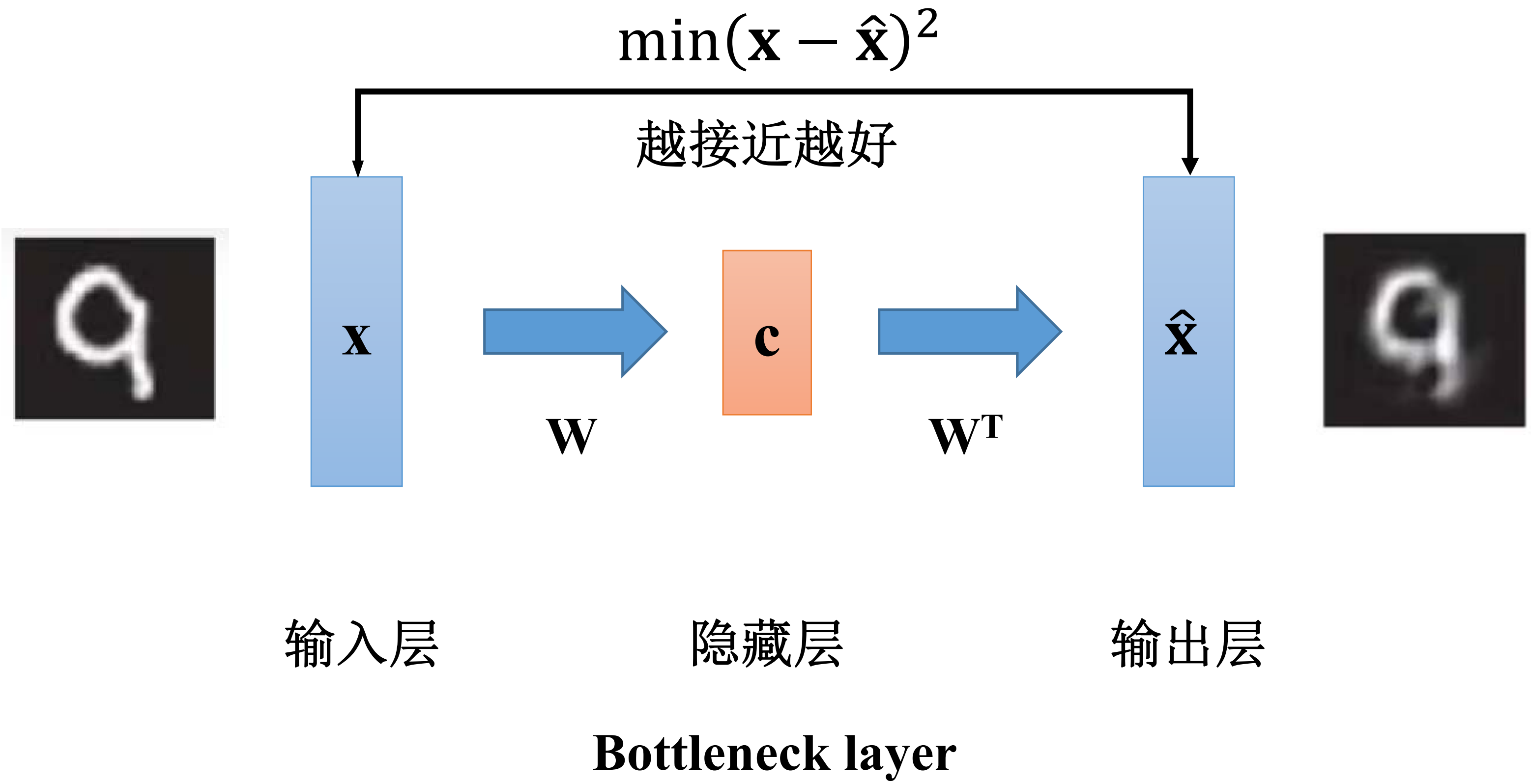
自编码 Auto-encoder

■ **PCA**



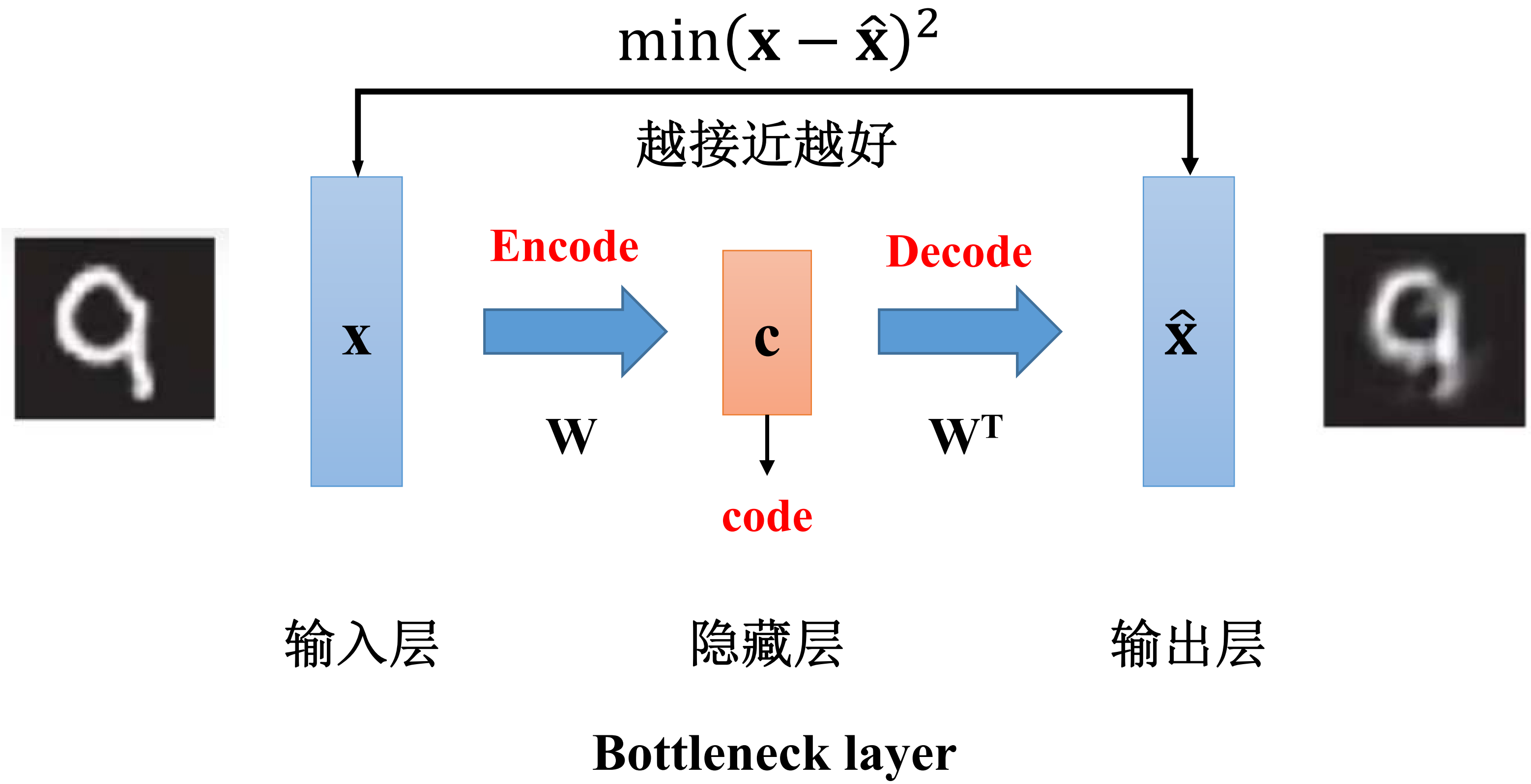
自编码 Auto-encoder

- **PCA→NN**



自编码 Auto-encoder

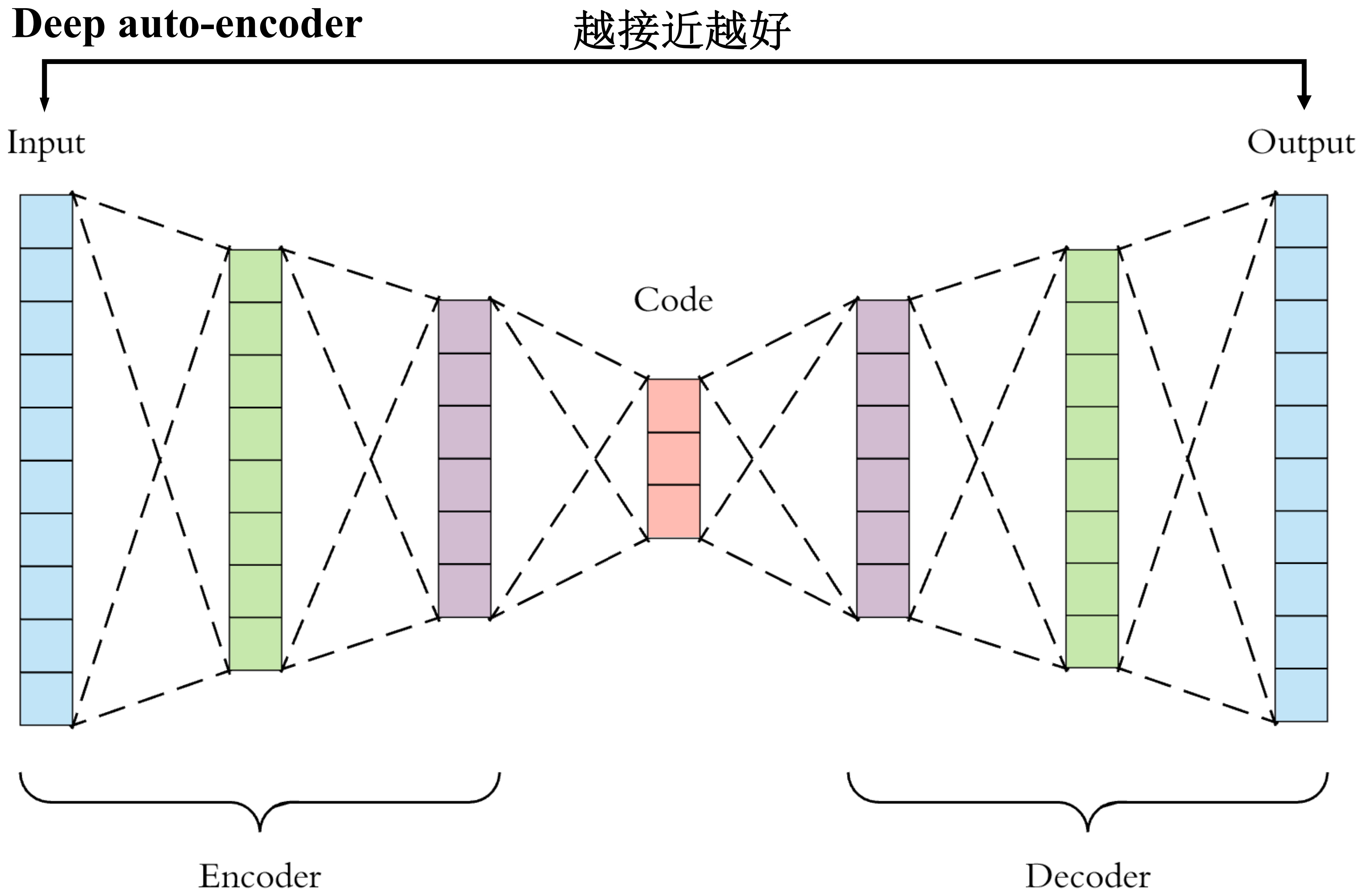
- **PCA→NN**



隐藏层的输出即是降维特征表示

自编码 Auto-encoder

- **Deep auto-encoder**



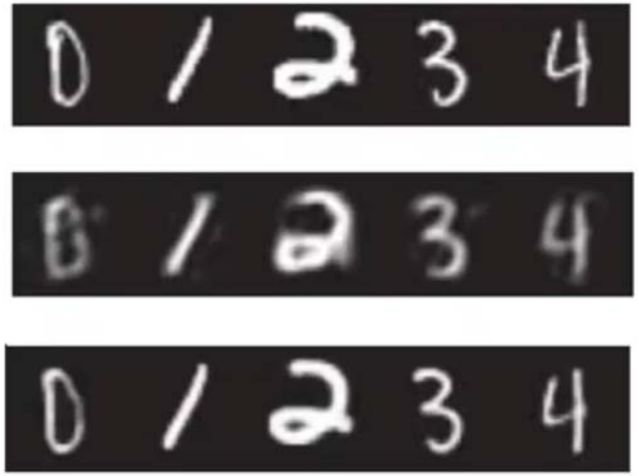
自编码 Auto-encoder

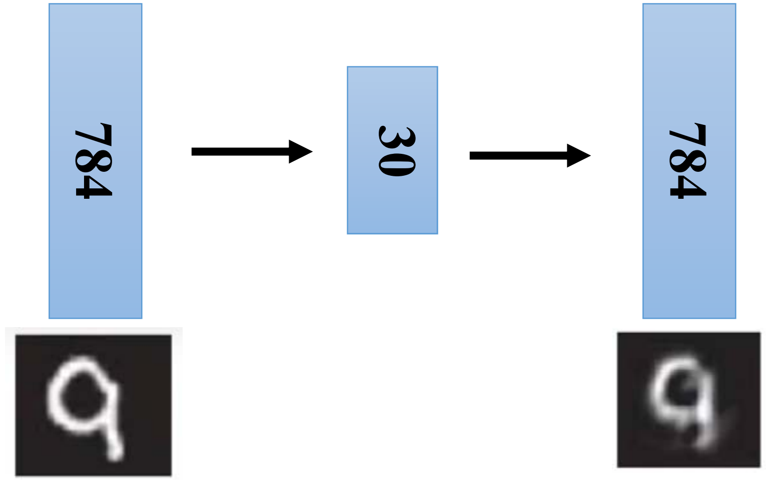
■ **Deep auto-encoder**

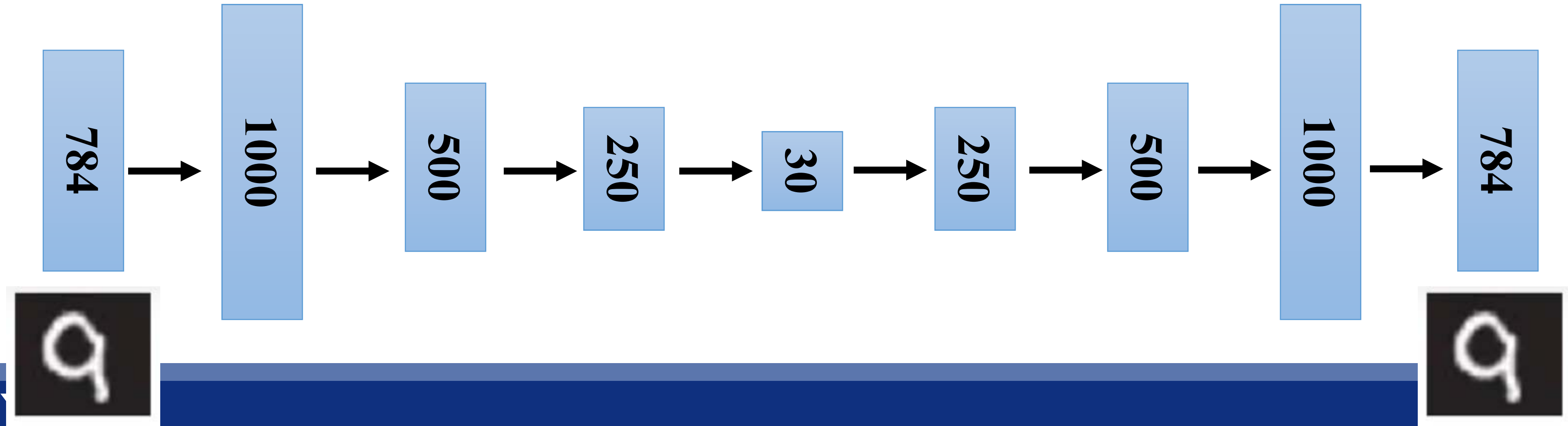
输入图像

PCA

Deep Auto-encoder



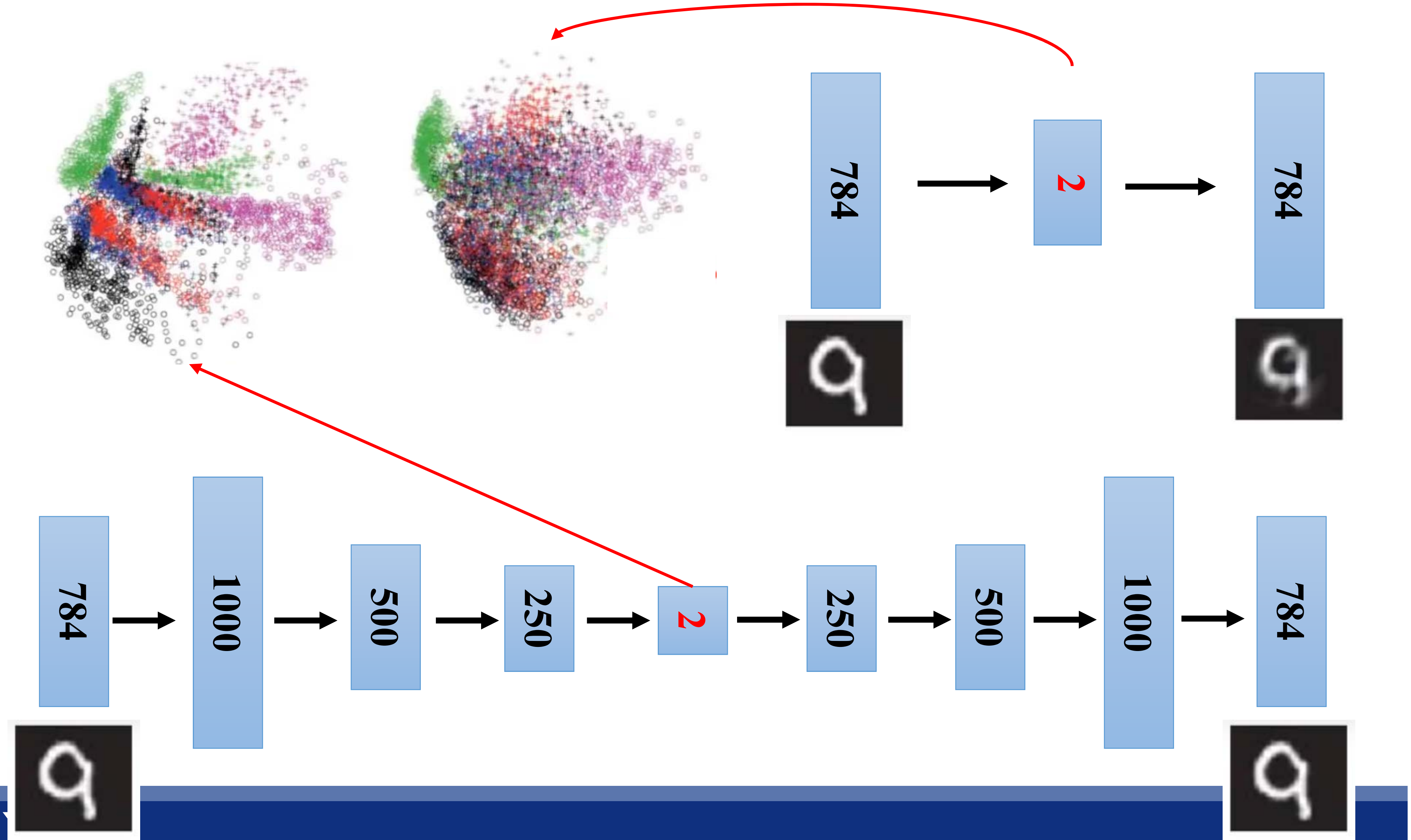




自编码 Auto-encoder

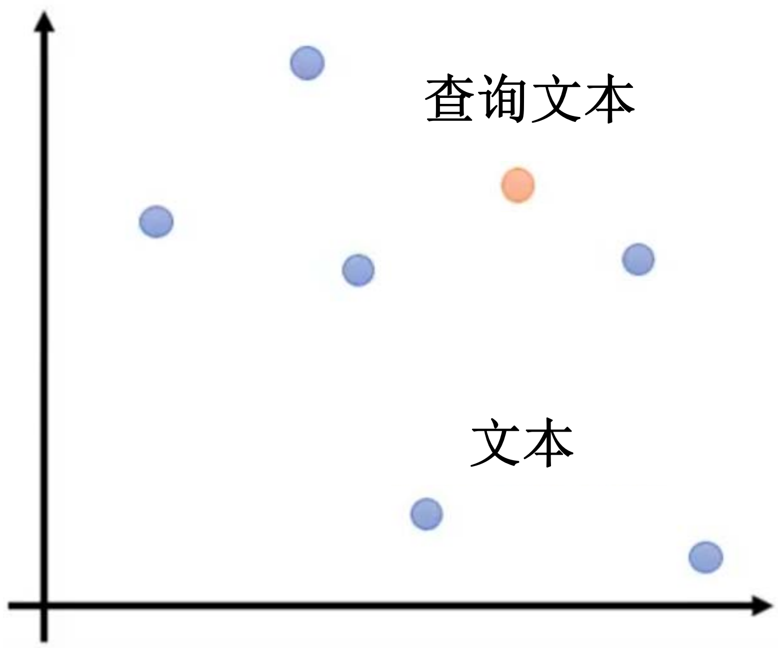
大连理工大学 Dalian University of Technology

■ **Deep auto-encoder**



自编码 Auto-encoder

- **Deep auto-encoder –文本检索**
  - 文本 → 编码表示

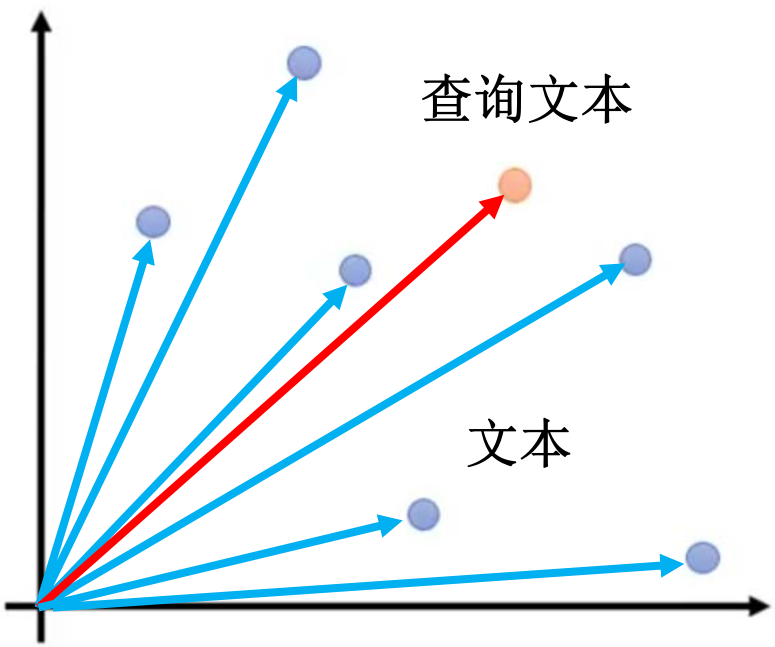


自编码 Auto-encoder

- **Deep auto-encoder –文本检索**
  - 文本 → 编码表示

Vector Space Model

向量内积/余弦相似性

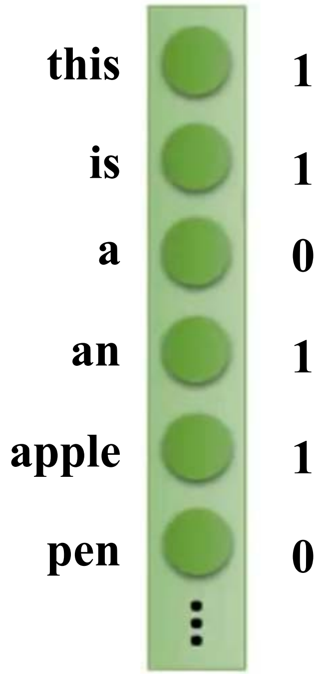


Bag-of-word

“This is an apple”

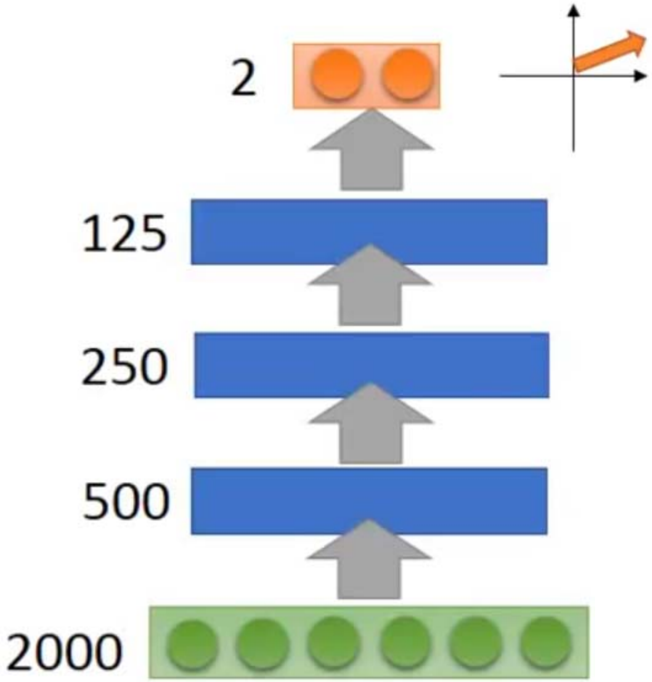
无法考虑词汇间

的语义相关性



自编码 Auto-encoder

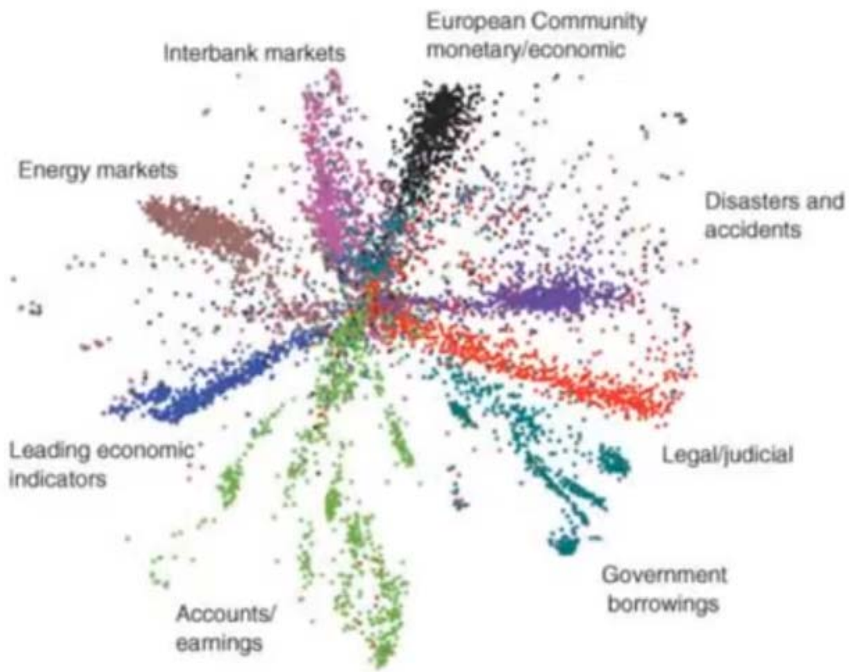
- **Deep auto-encoder –文本检索**
  - 文本 → 编码表示



Bag-of-word

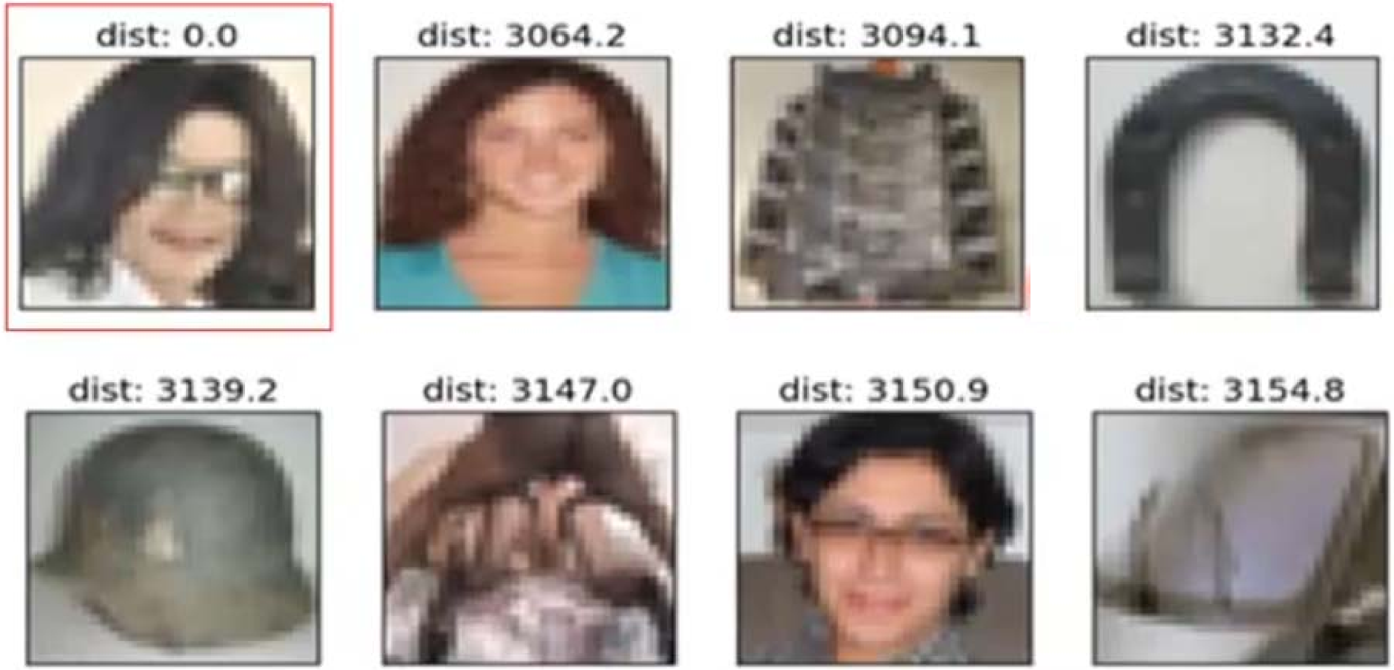
文本/查询词汇

训练准则：语义相近的文本有相似的编码



自编码 Auto-encoder

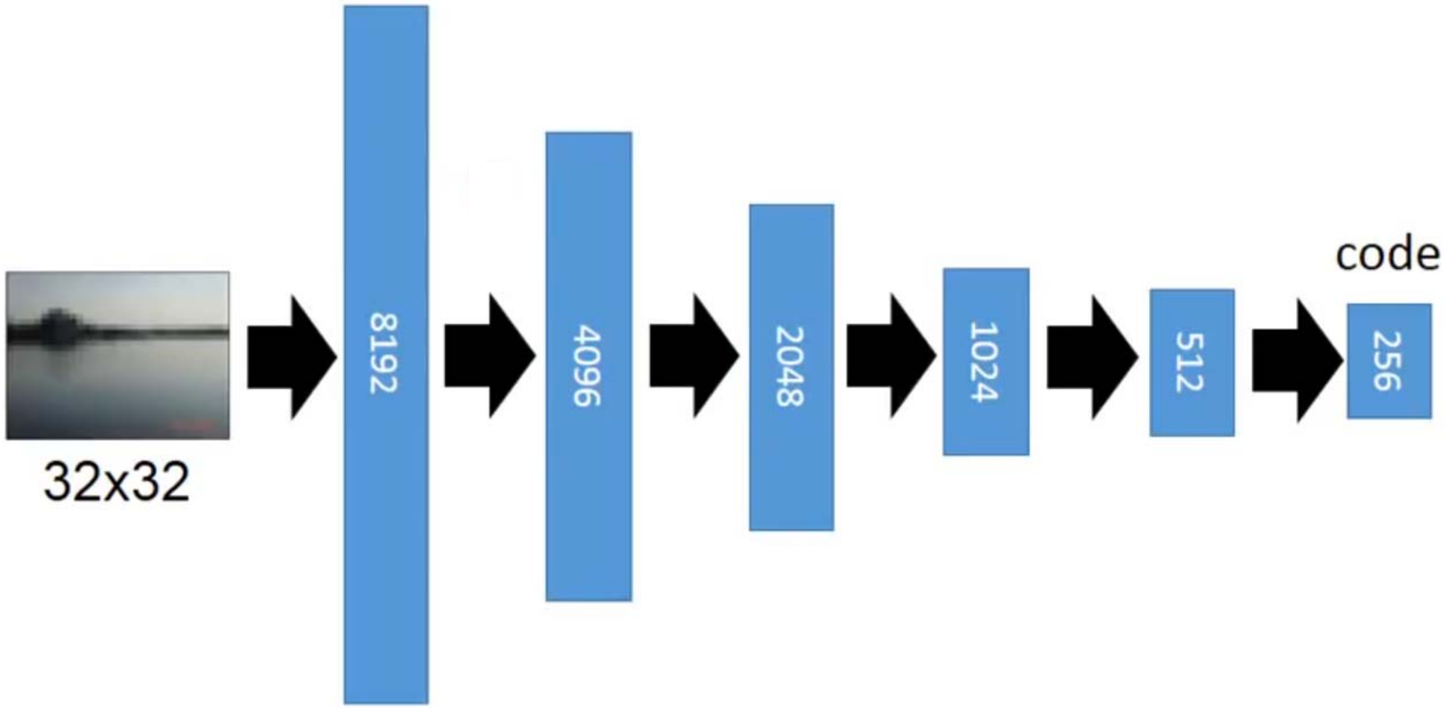
- **Deep auto-encoder –图像检索**
  - 在欧式空间中做图像检索



Using very deep antoencoders for content-based image retrieval. ESANN2011

自编码 Auto-encoder

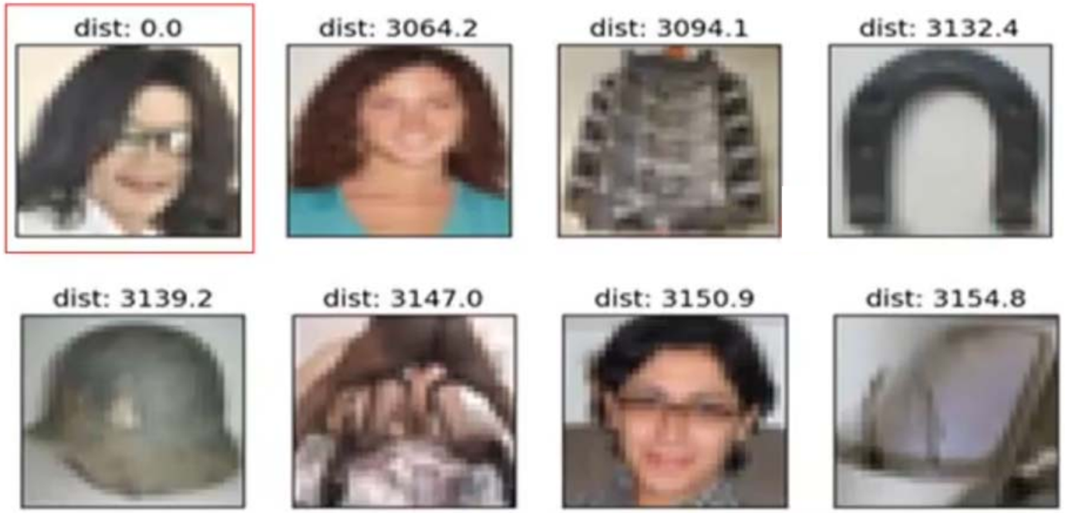
- **Deep auto-encoder –图像检索**
  - 在隐式嵌入空间中做图像检索

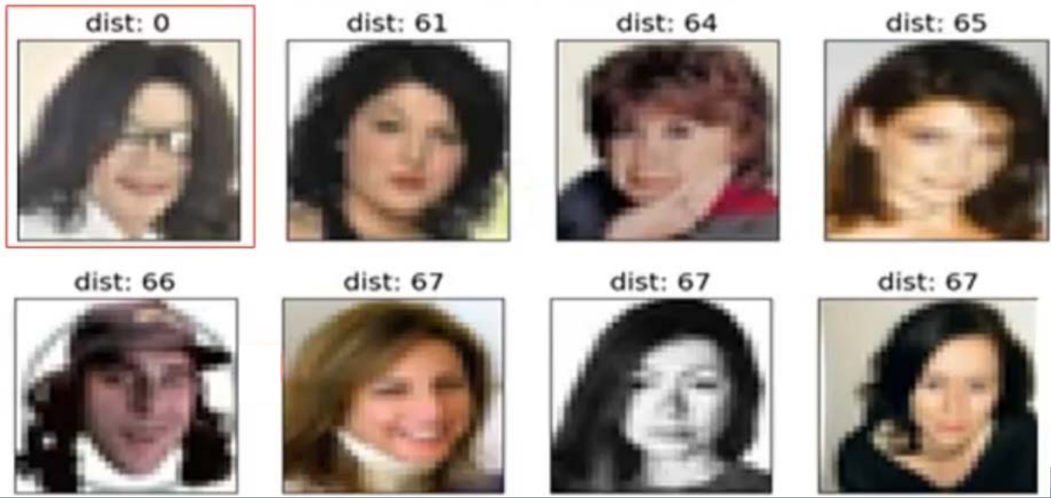


Using very deep antoencoders for content-based image retrieval. ESANN2011

自编码 Auto-encoder

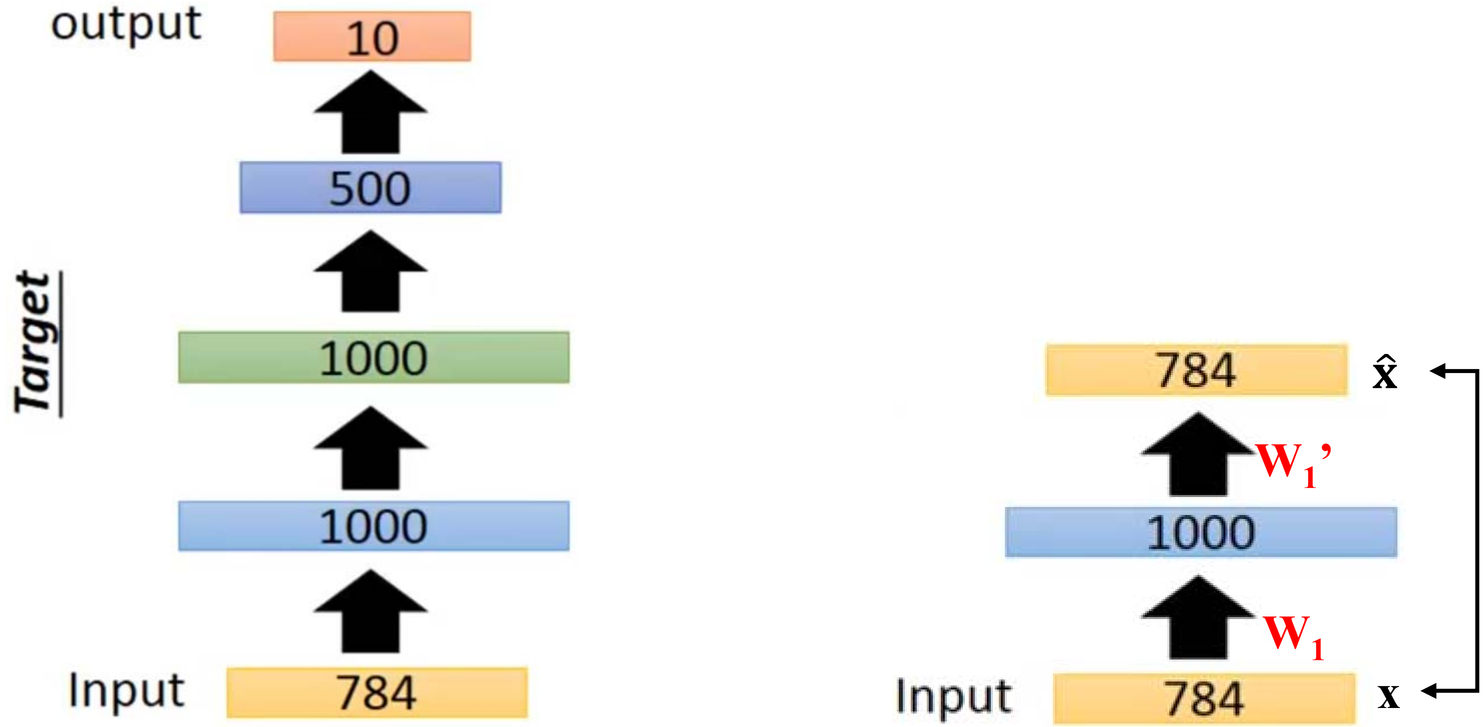
- **Deep auto-encoder –图像检索**
  - 在欧式空间中做图像检索
  - 在隐式嵌入空间中做图像检索





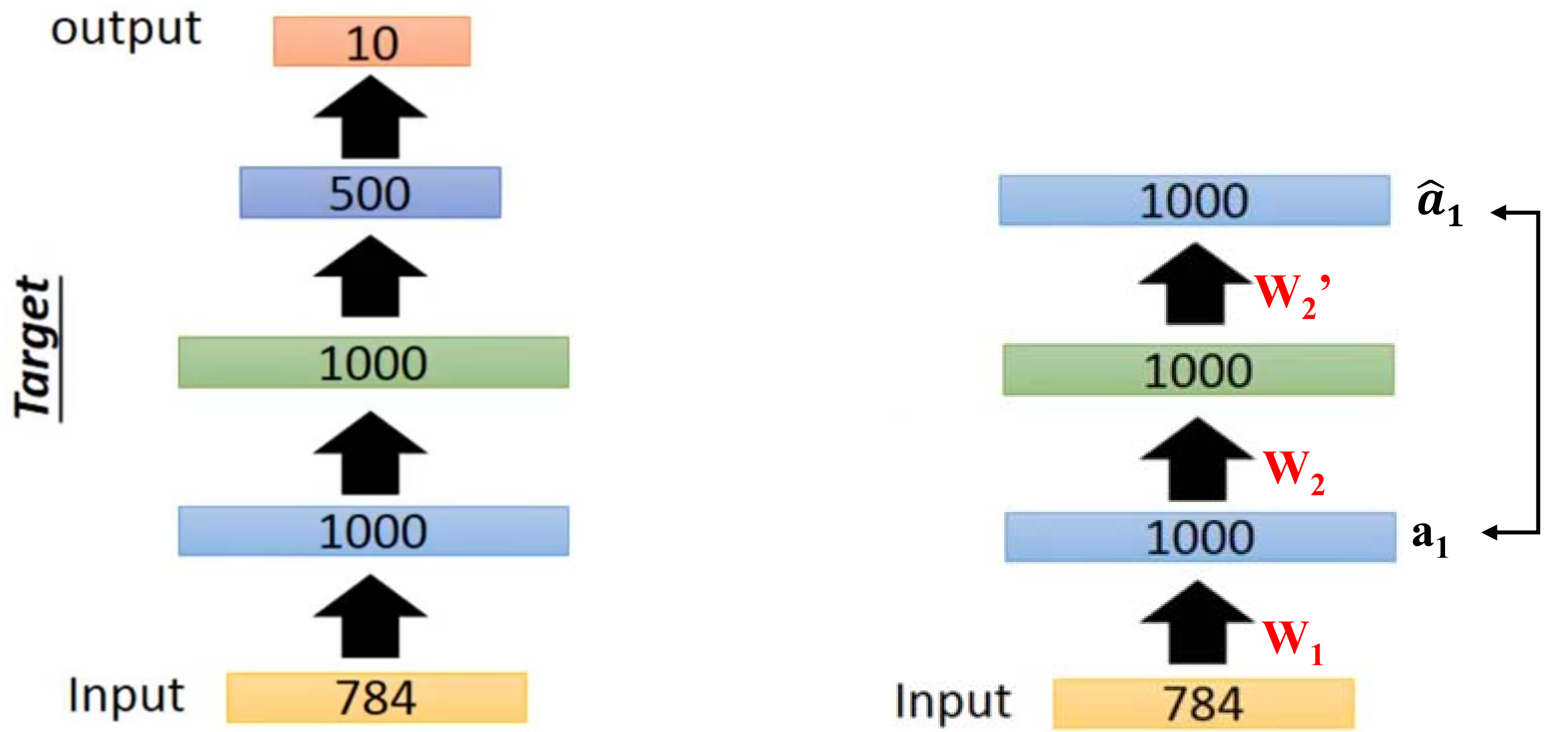
自编码 Auto-encoder

- **Deep auto-encoder –预训练DNN**
  - 用**auto-encoder学习W_1**



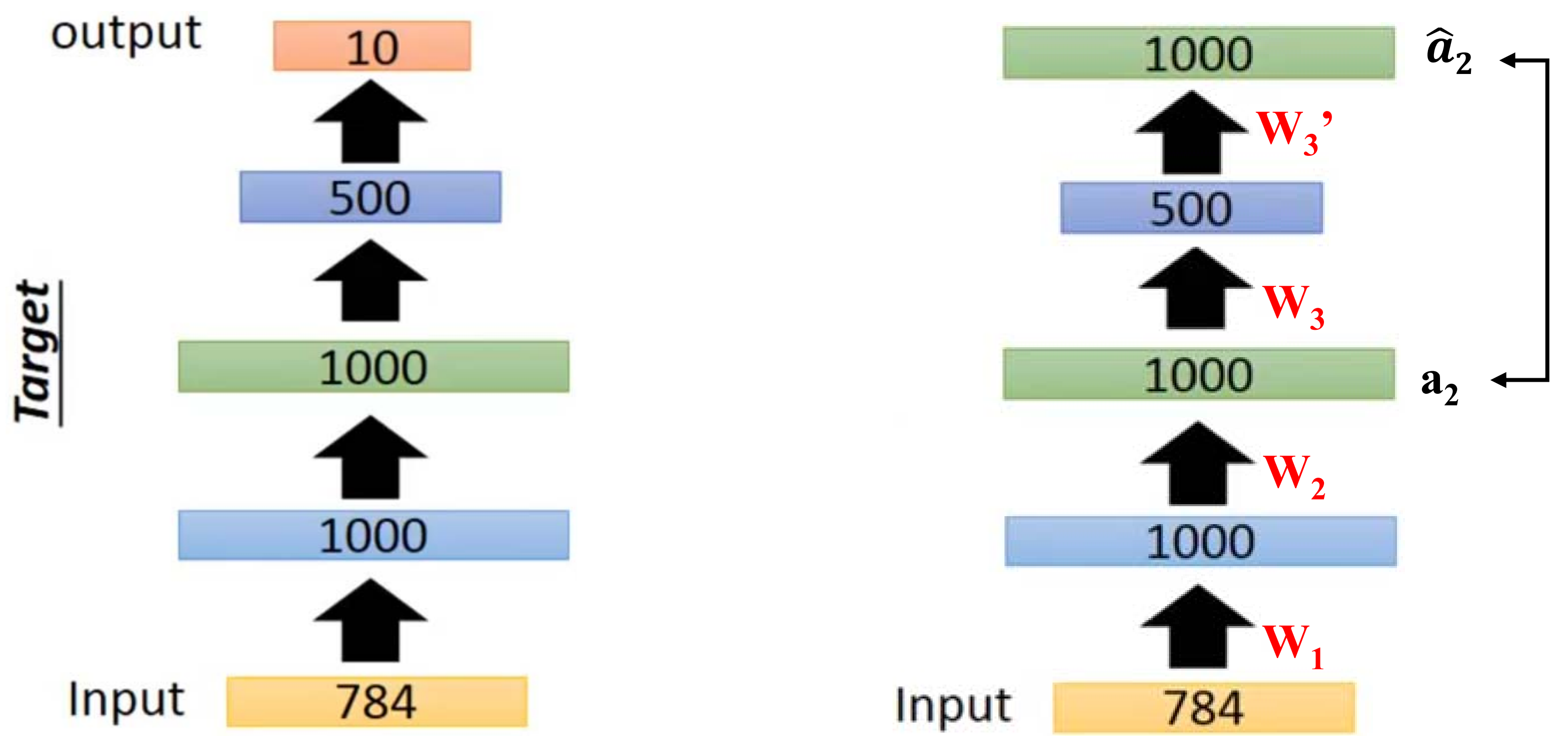
自编码 Auto-encoder

- **Deep auto-encoder –预训练DNN**
  - 固定W_1，用auto-encoder学习W_2



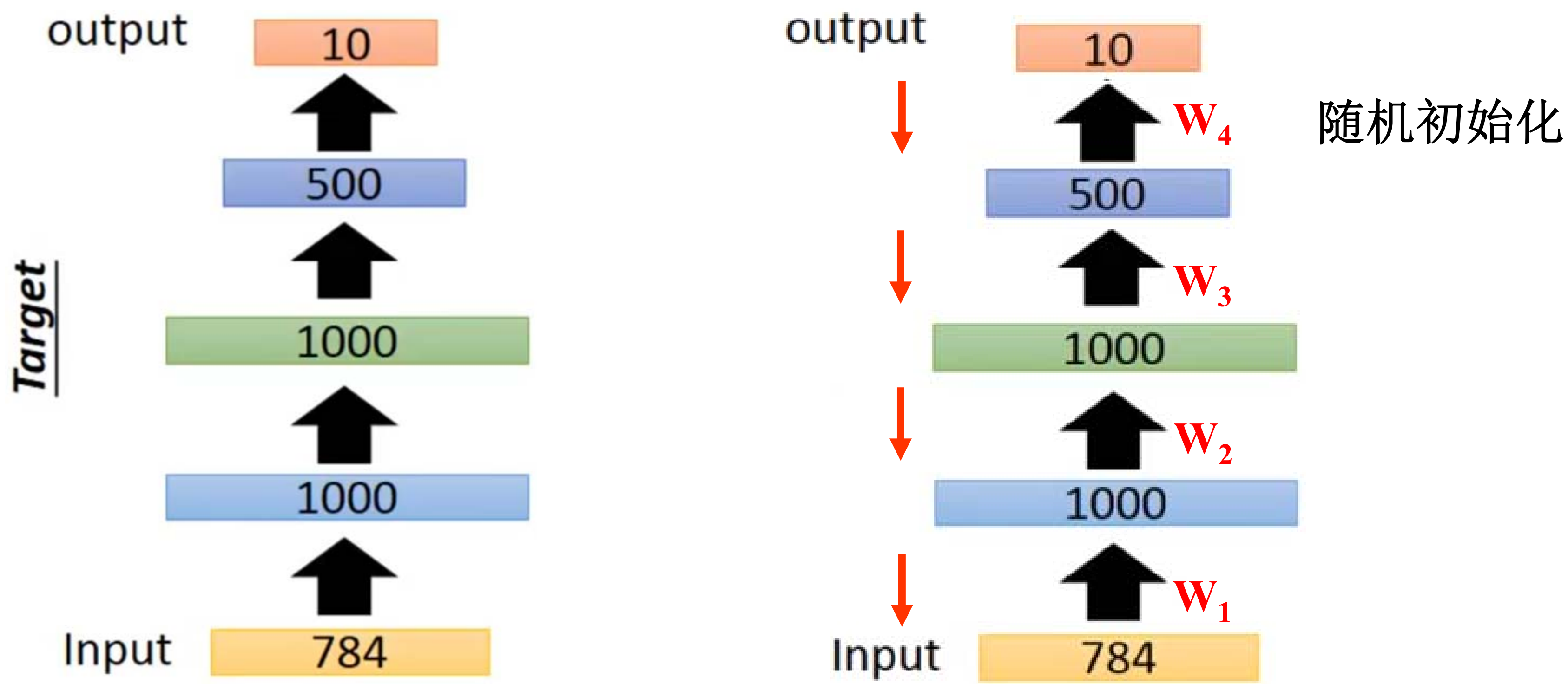
自编码 Auto-encoder

- **Deep auto-encoder –预训练DNN**
  - 固定W_1、W_2，用auto-encoder学习W_3



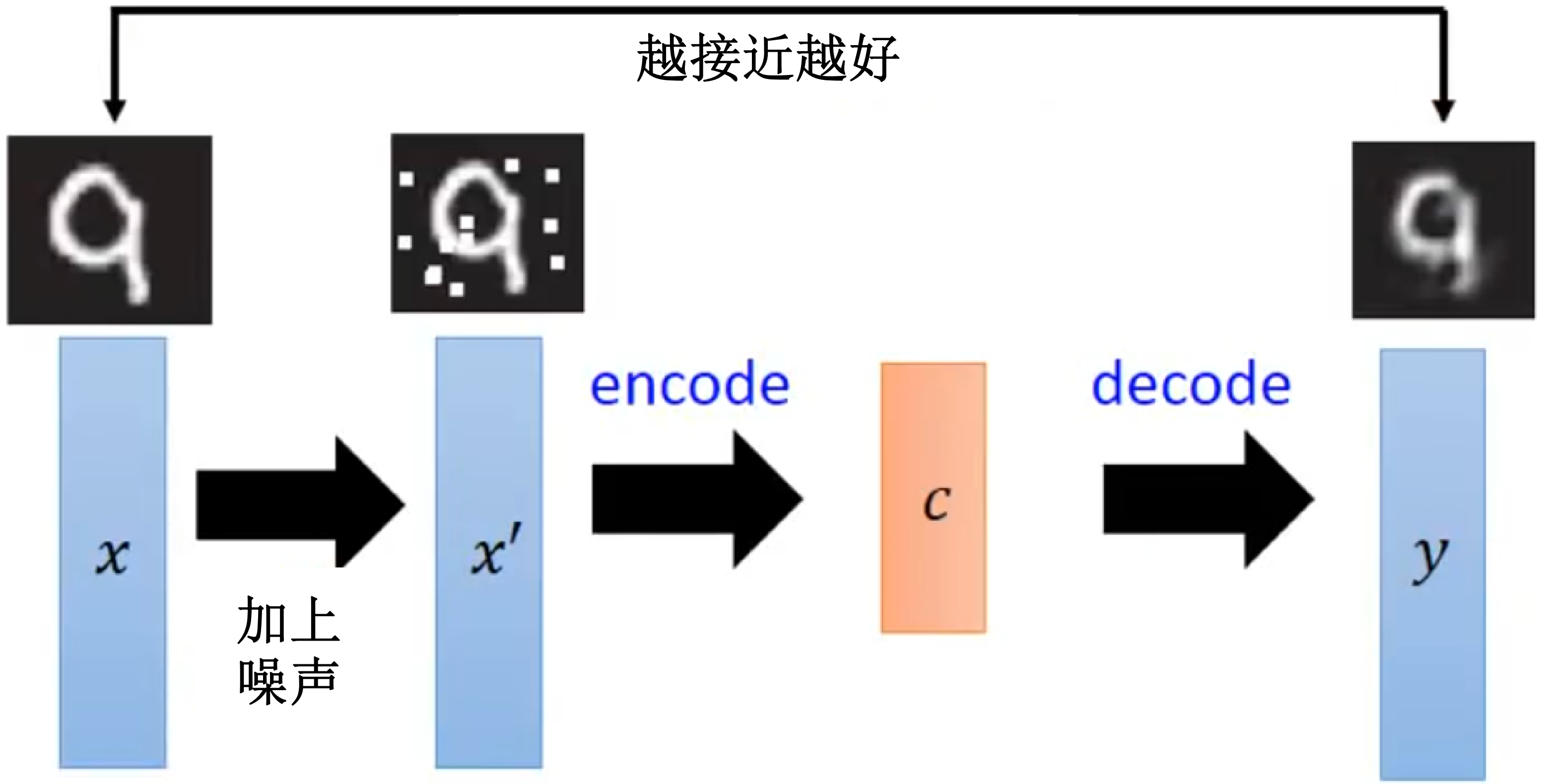
自编码 Auto-encoder

- **Deep auto-encoder –预训练DNN**
  - 固定W_1、W_2、W_3，随机初始化W_4
  - 反向传播微调参数



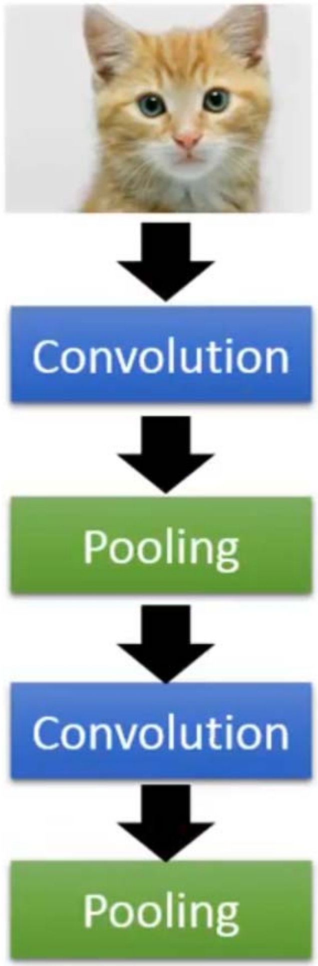
自编码 Auto-encoder

- **Deep auto-encoder** –去噪自编码
  - 学习特征嵌入
  - 学习过滤噪声



自编码 Auto-encoder

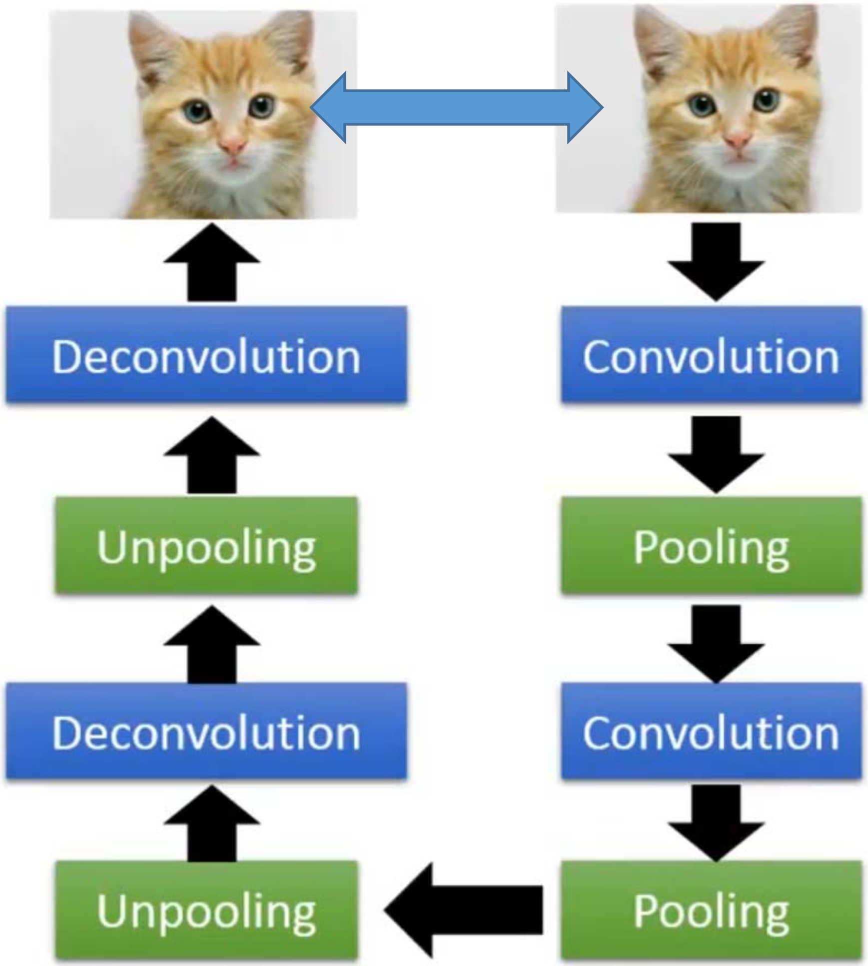
- **Auto-encoder for CNN**



自编码 Auto-encoder

- **Auto-encoder for CNN**

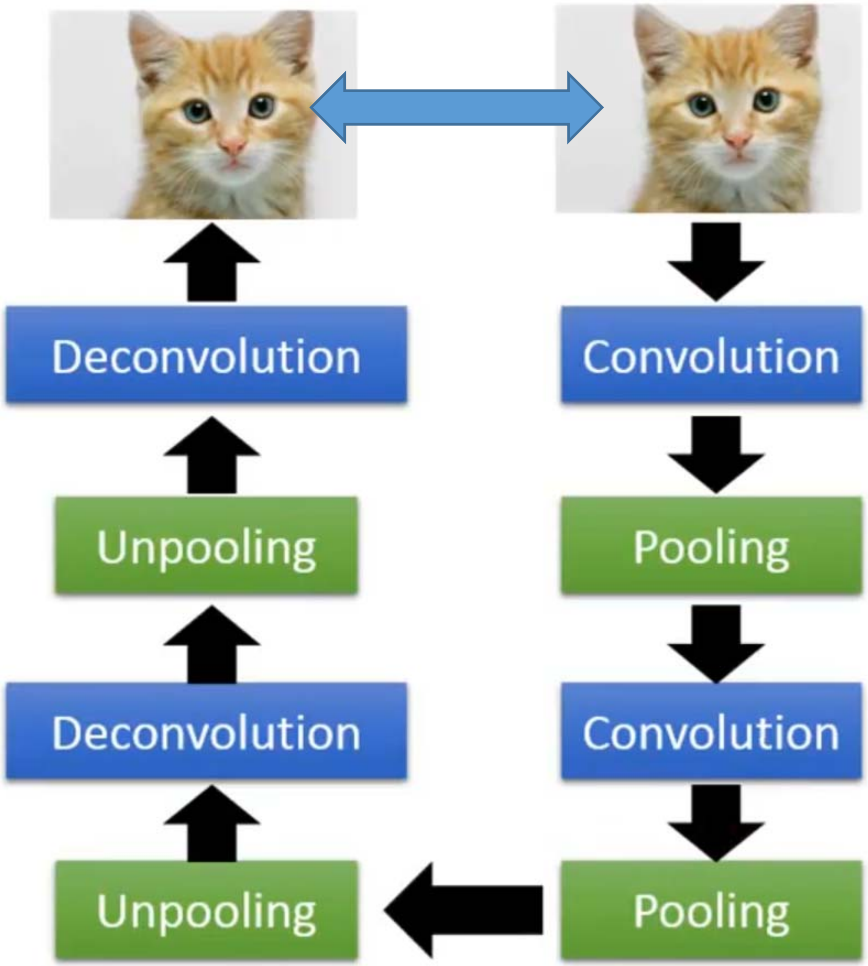
越接近越好



自编码 Auto-encoder

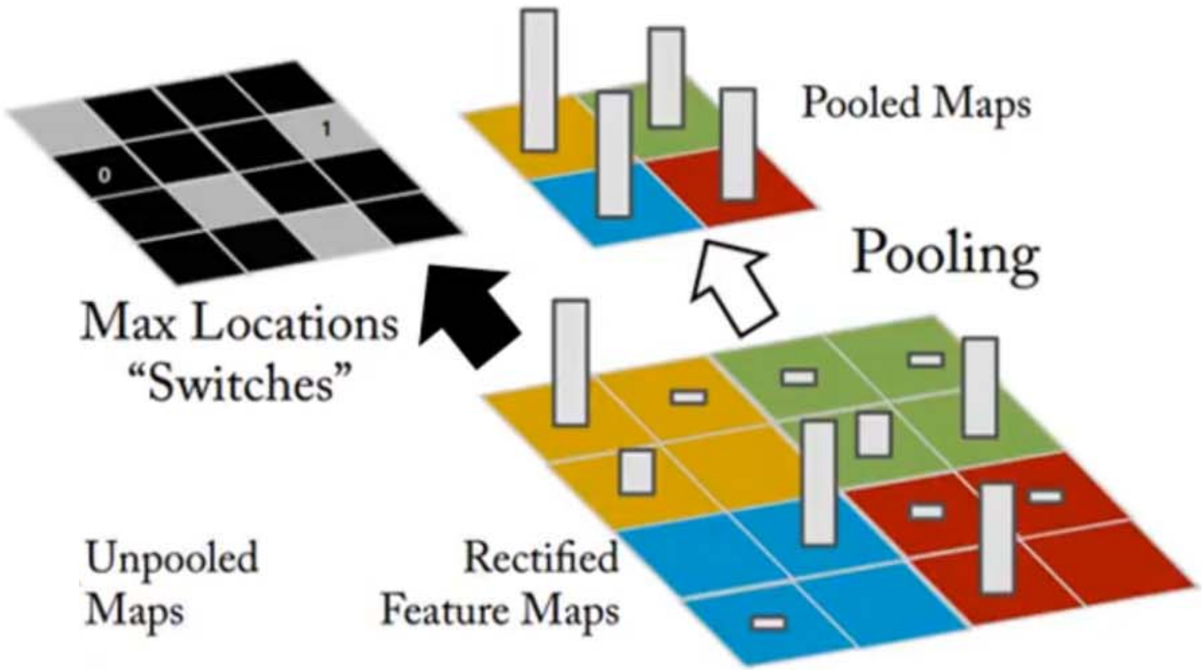
- **Auto-encoder for CNN**

越接近越好

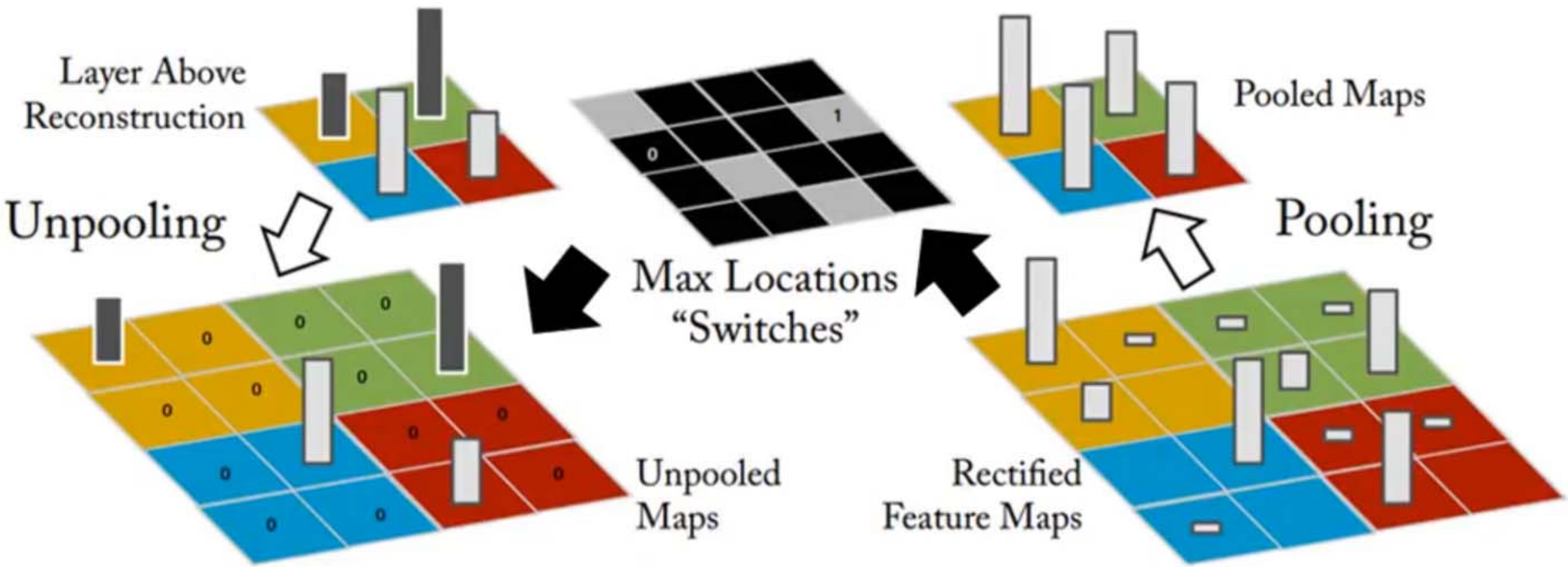


自编码 Auto-encoder

- **Auto-encoder for CNN**
  - **Unpooling**

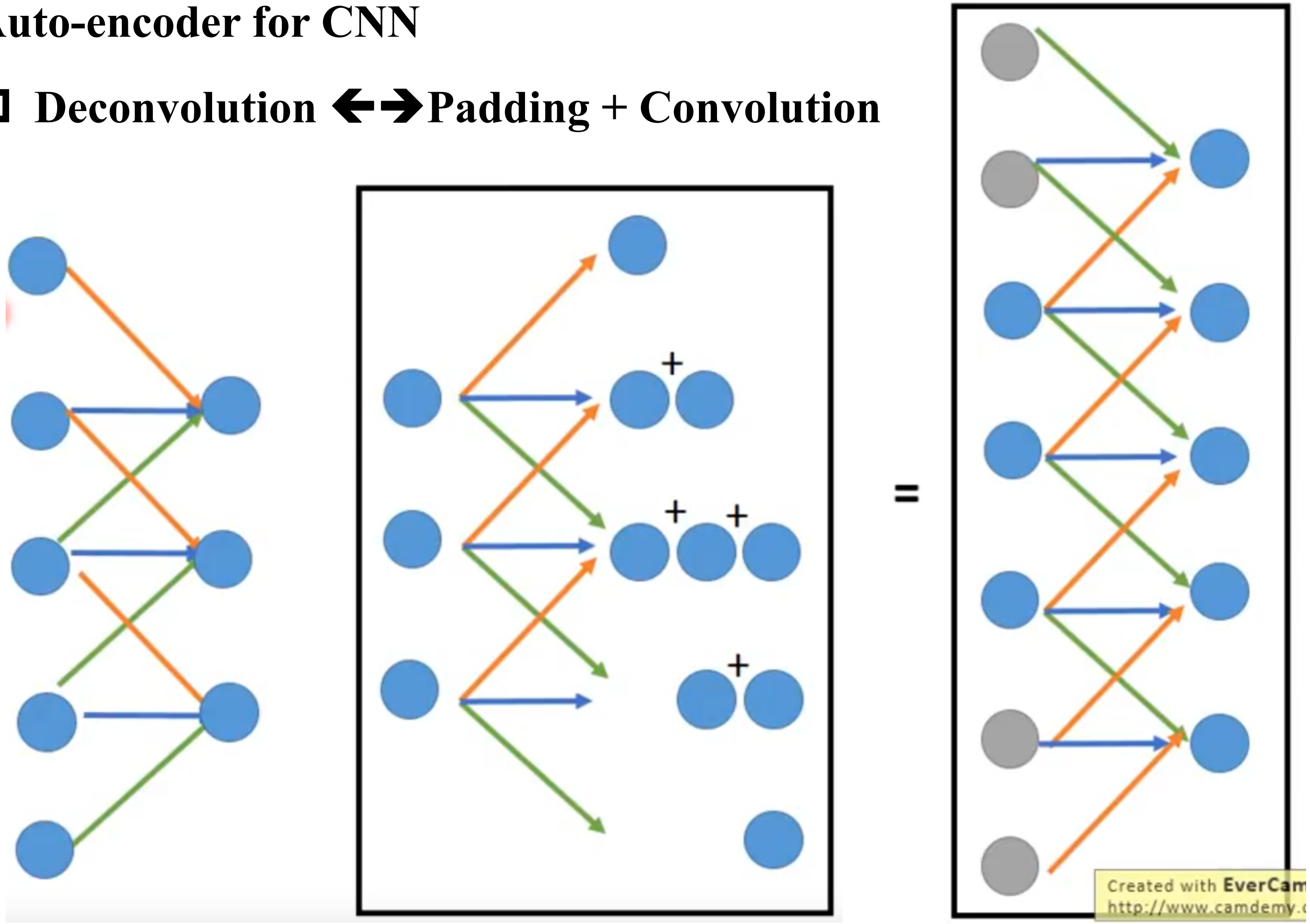


- **Auto-encoder for CNN**
  - **Unpooling**



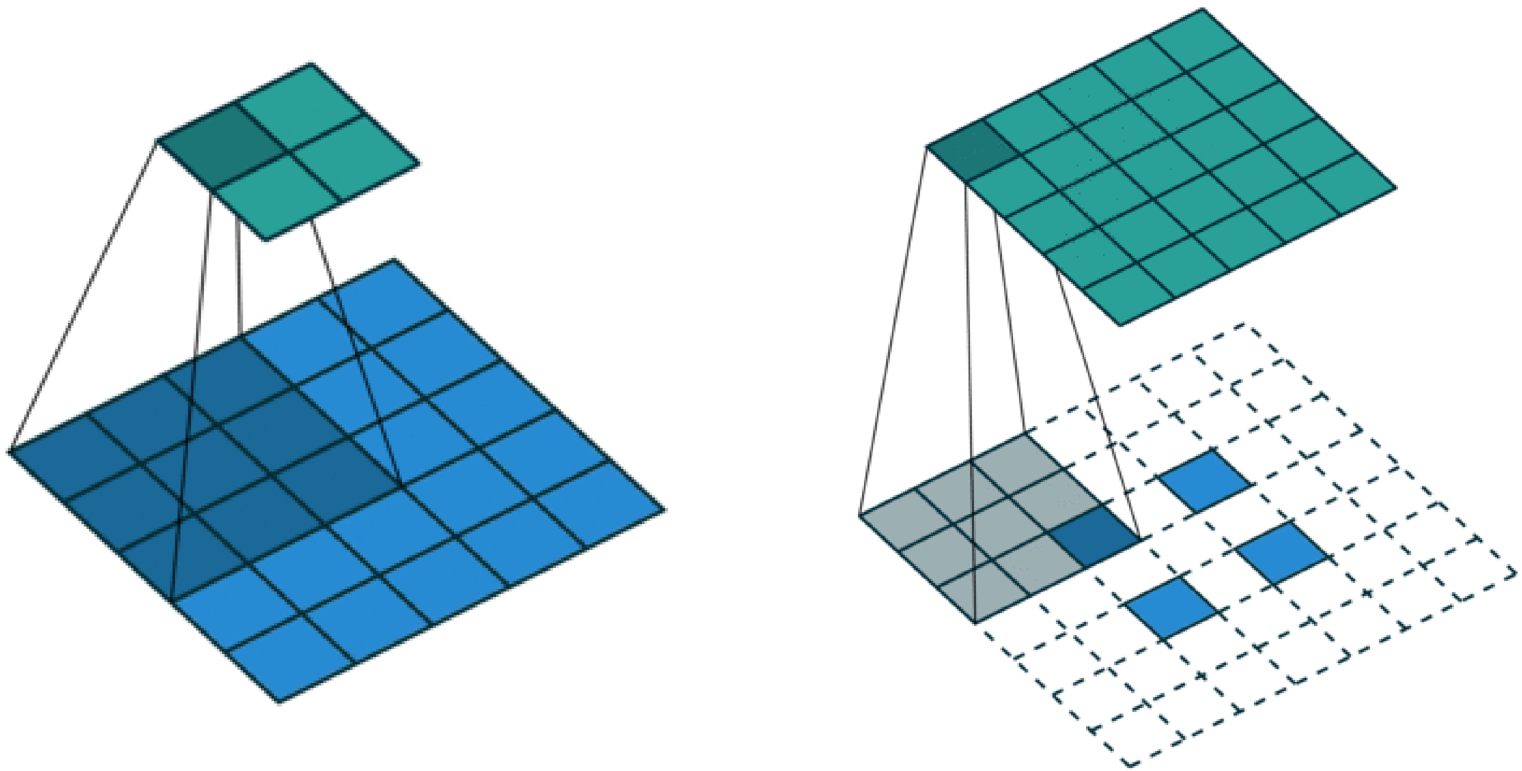
自编码 Auto-encoder

- **Auto-encoder for CNN**
  - **Deconvolution ←→ Padding + Convolution**



自编码 Auto-encoder

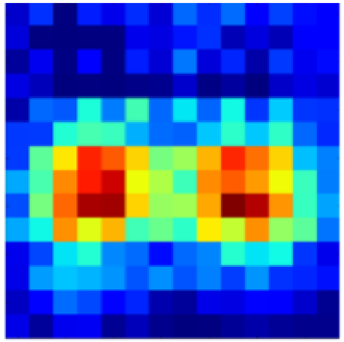
- **Auto-encoder for CNN**
  - **Deconvolution ←→Padding + Convolution**



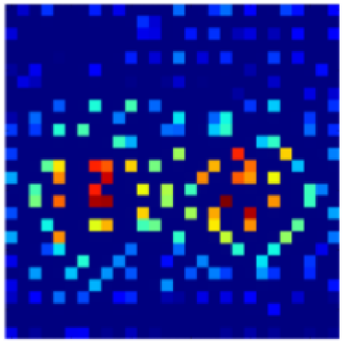
自编码 Auto-encoder

- **Auto-encoder for CNN**
  - **Deconvolution VS. Unpooling**

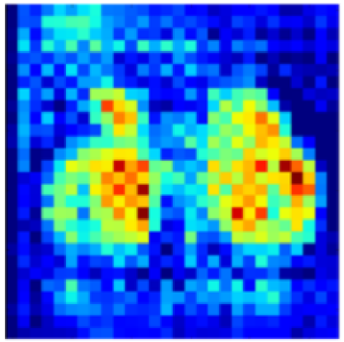
输入图像



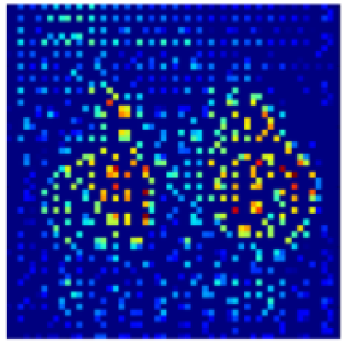
14*14反卷积



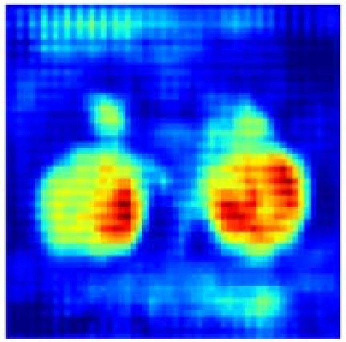
28*28 Unpooling



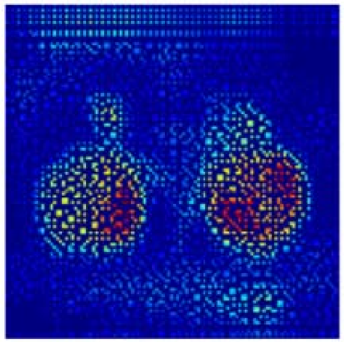
28*28反卷积



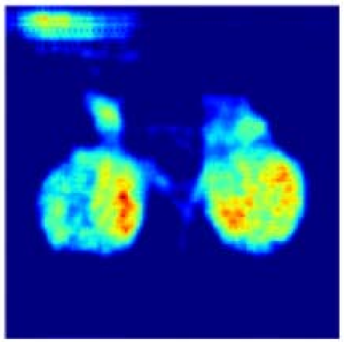
56*56 Unpooling



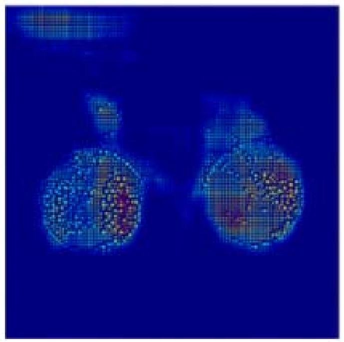
56*56 反卷积



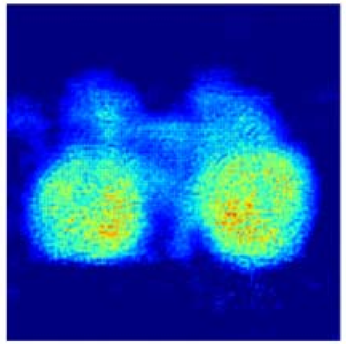
112*112 Unpooling



112*112 反卷积



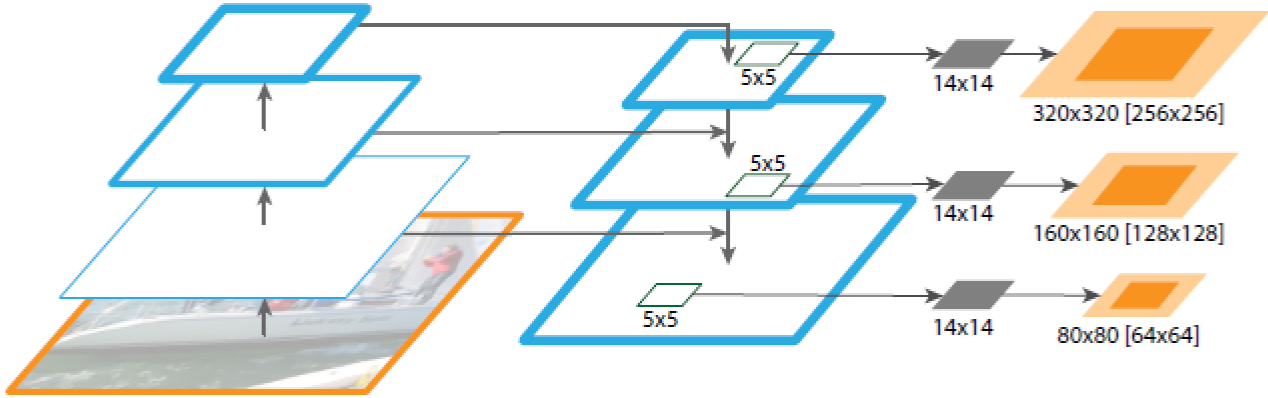
224*224 Unpooling



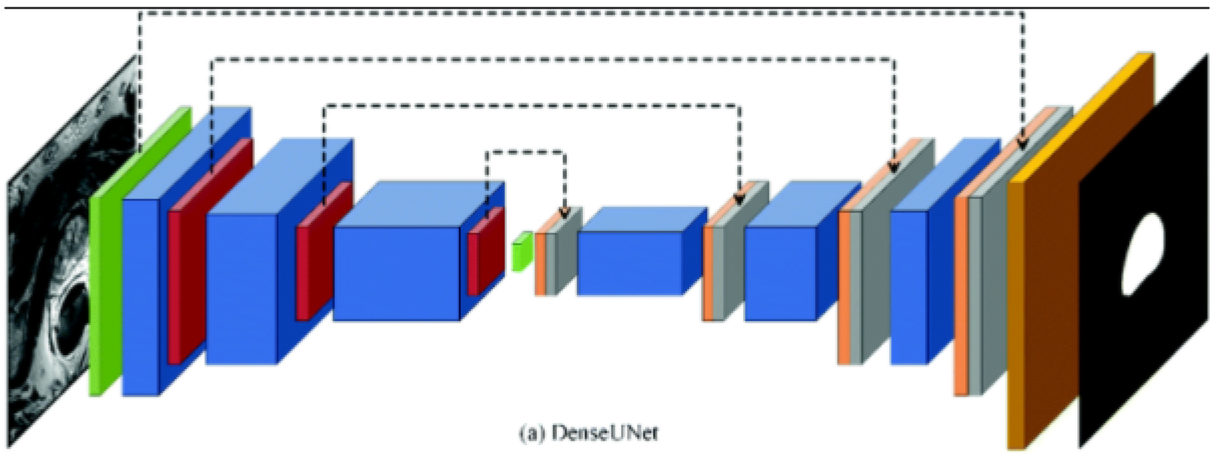
224*224 反卷积

自编码 Auto-encoder

- **FPN**

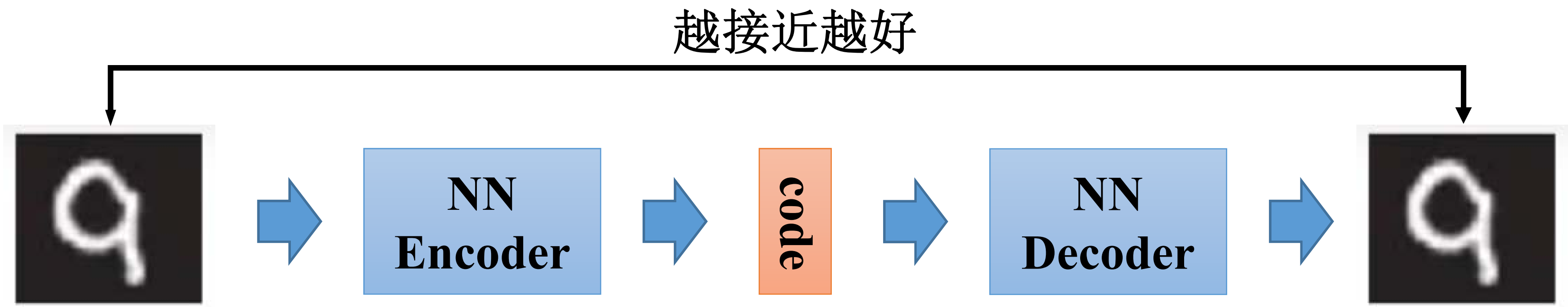


- **U-NET**



(a) DenseUNet

深入理解 Auto-encoder

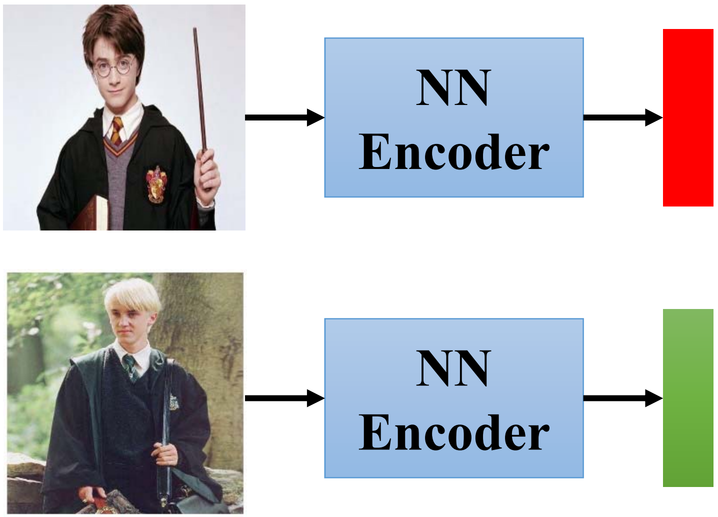


特征嵌入、隐式表示、隐式编码

- 一定要最小化重建误差吗？
- 如何学好嵌入特征？
  - 好的嵌入特征应有效表示输入数据

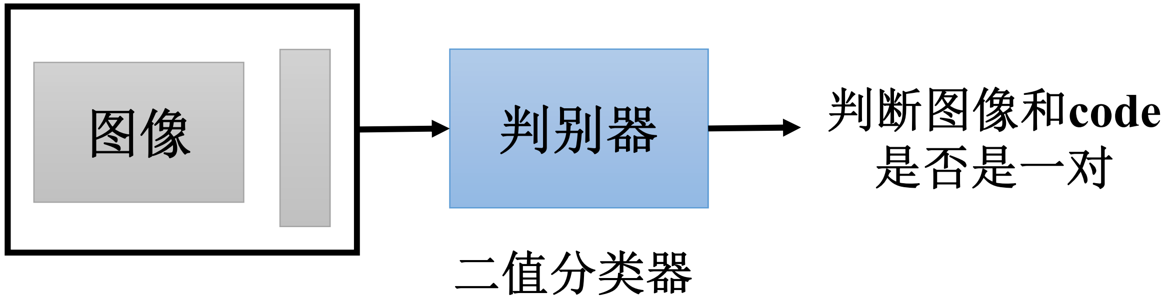
深入理解 Auto-encoder

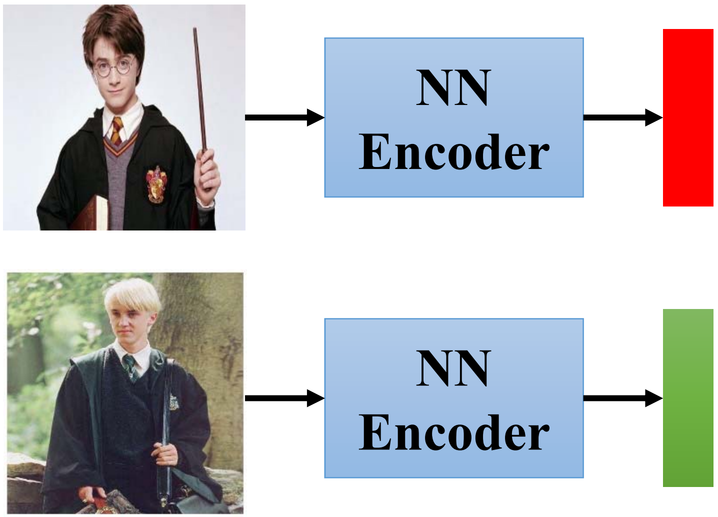
■ 如何评价encoder的好坏



深入理解 Auto-encoder

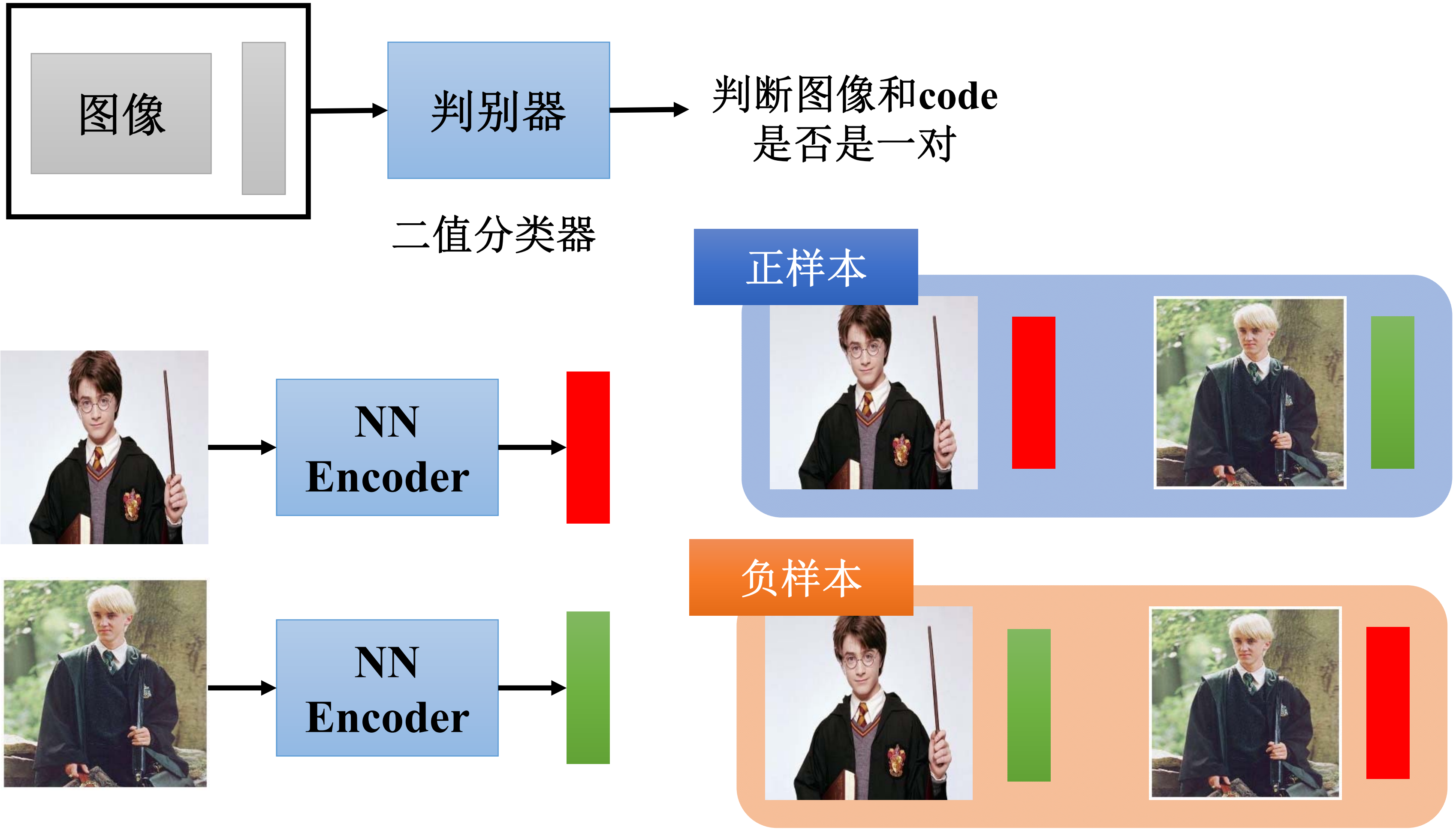
- 如何评价encoder的好坏





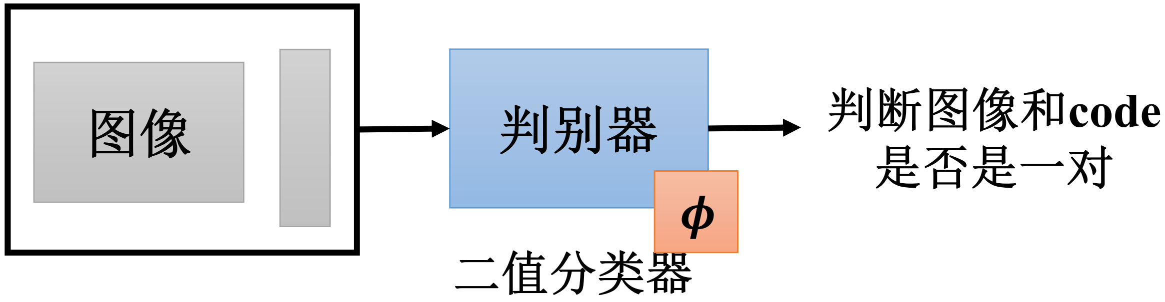
深入理解 Auto-encoder

■ 如何评价encoder的好坏



深入理解 Auto-encoder

■ 如何评价encoder的好坏



- ☐ 训练函数ϕ，最小化交叉熵损失L_D

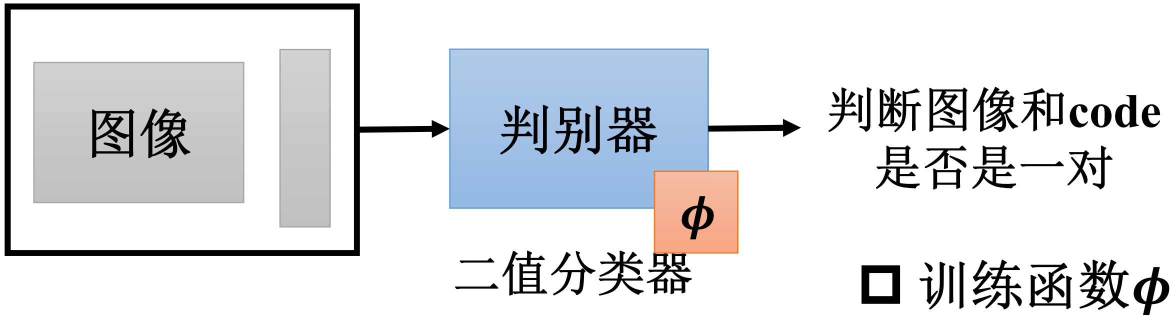
$$L_D^* = min_\phi L_D$$

- ☐ L_D^*小，即嵌入特征具有代表性，encoder是好的

- ☐ 训练encoder：最小化重建误差→最小化L_D

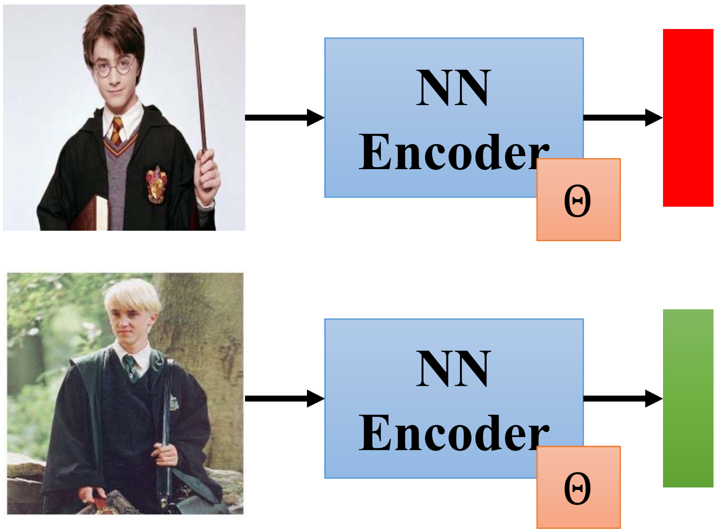
深入理解 Auto-encoder

■ 如何评价encoder的好坏



- □ 训练函数ϕ，最小化交叉熵损失L_D

$$L_D^* = min_\phi L_D$$



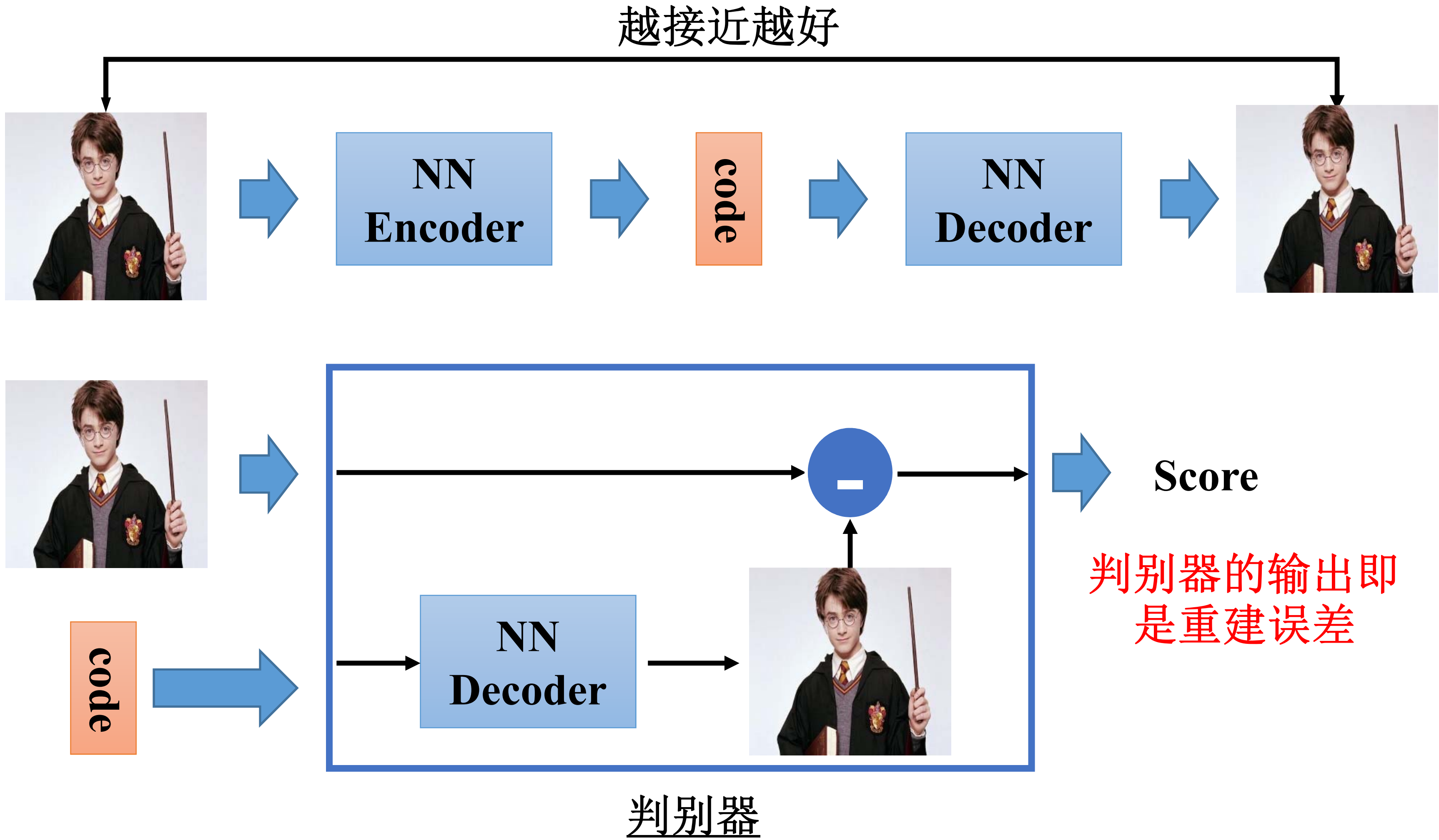
- □ 训练encoder：最小化L_D

$$\theta^* = \arg\min_\theta L_D^*$$

$$= \arg\min_\theta \min_\phi L_D$$

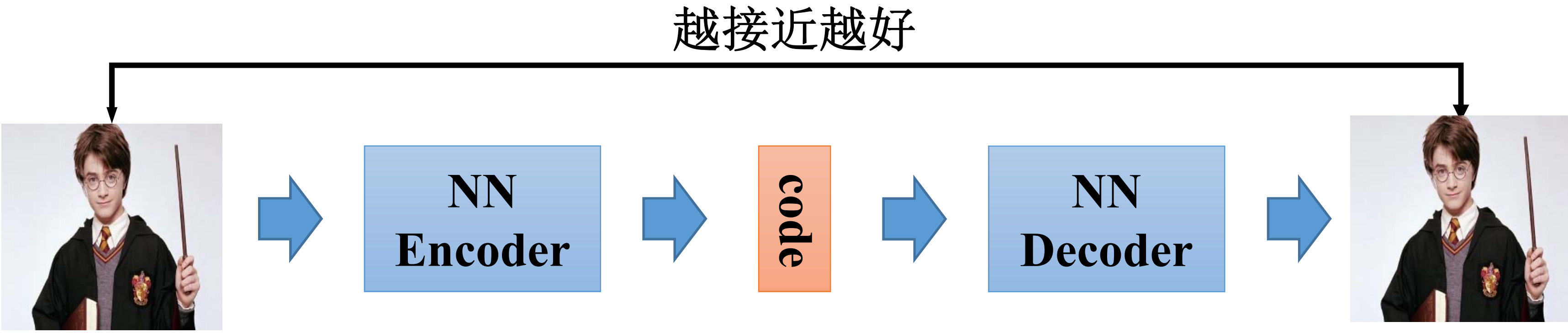
深入理解 Auto-encoder

- Auto-encoder是上述过程的特例



深入理解 Auto-encoder

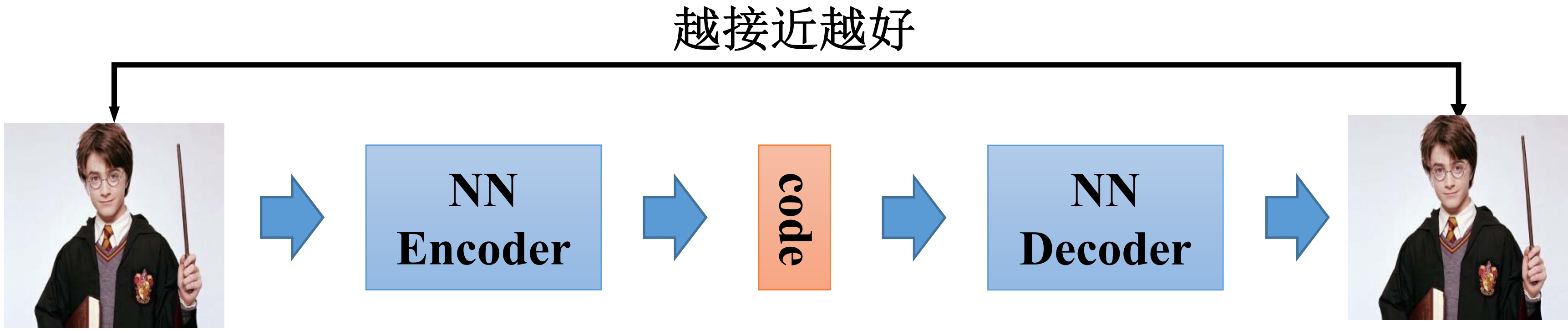
■ **Auto-encoder是上述过程的特例**



- ☐ 重建误差←→判别器输出的数值
- ☐ **Auto-encoder**训练的时候没有用负样本

深入理解 Auto-encoder

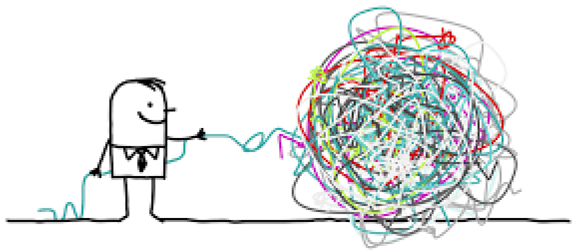
- Auto-encoder是上述过程的特例

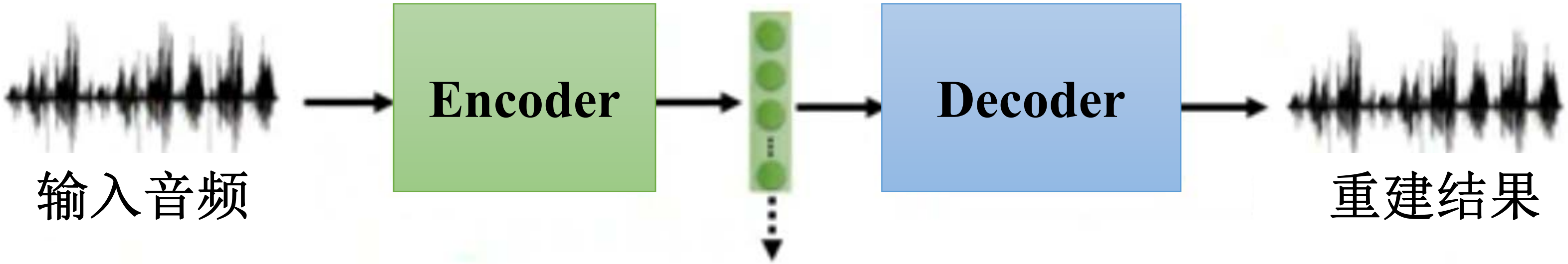


- 重建误差←→判别器输出的数值
- Auto-encoder训练的时候没有用负样本

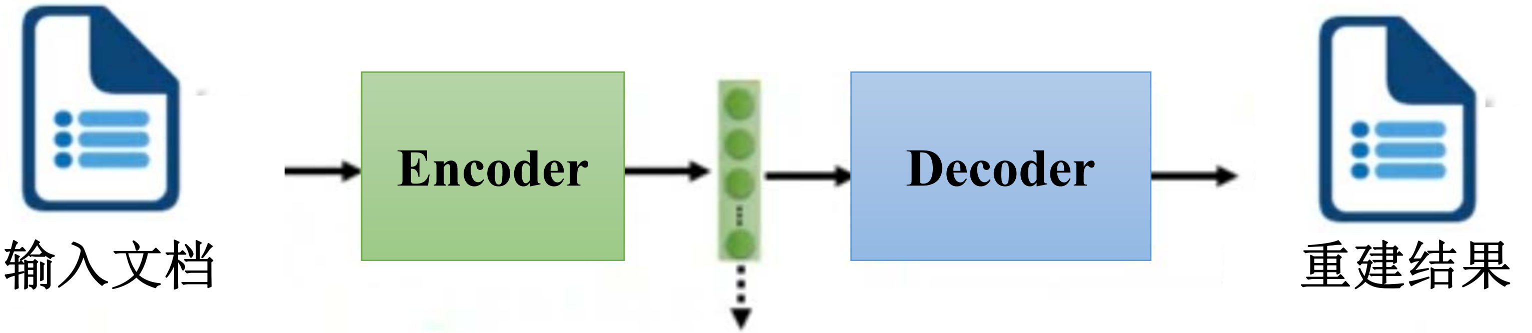
深入理解 Auto-encoder

- 让嵌入特征具有可解释性
- **Feature disentangle**
  - 输入数据包含多个方面的信息





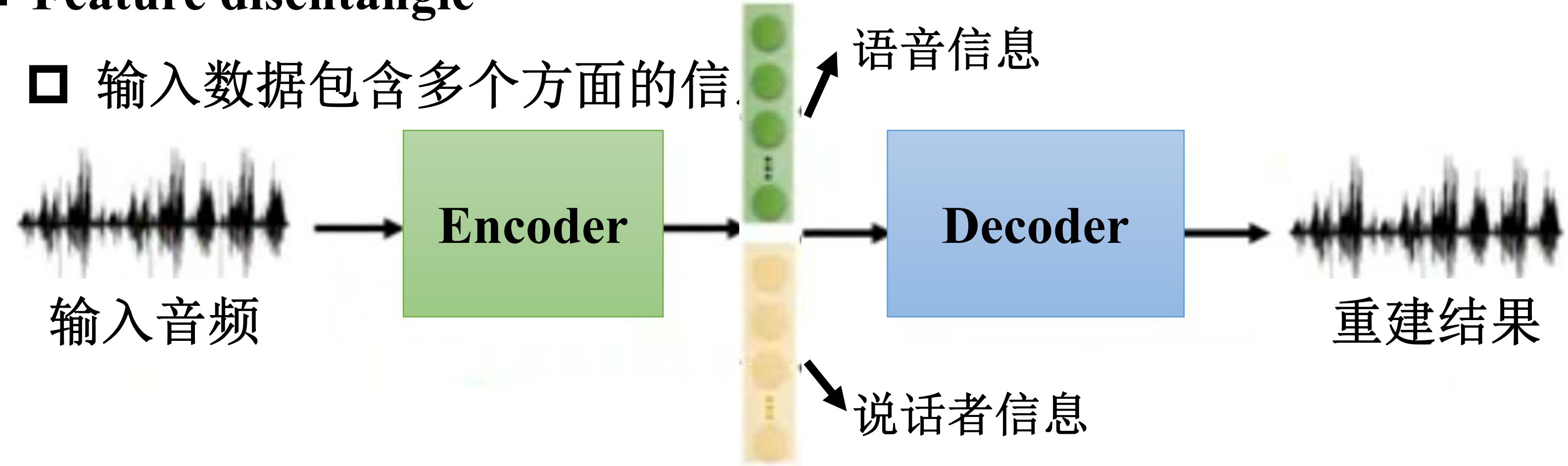
包括语音信息、说话者信息等

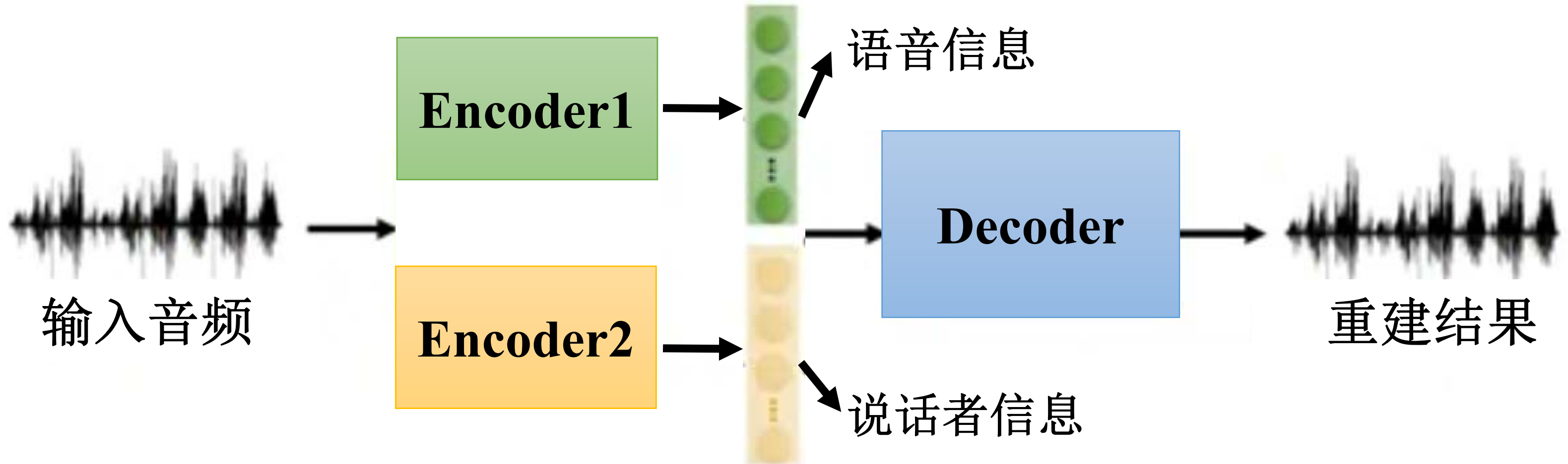


包括语法信息、语义信息等

深入理解 Auto-encoder

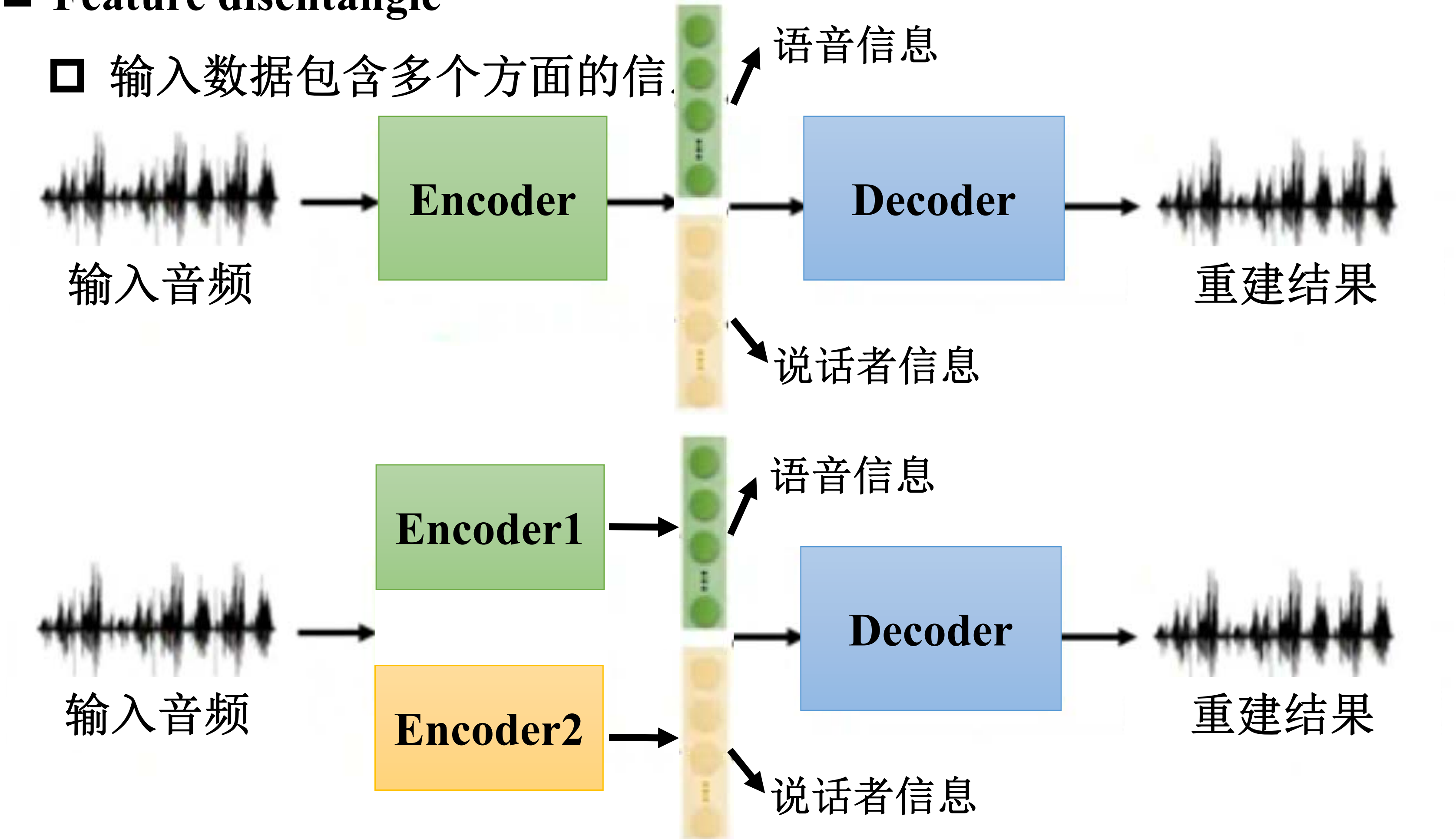
- 让嵌入特征具有可解释性
- **Feature disentangle**
  - 输入数据包含多个方面的信





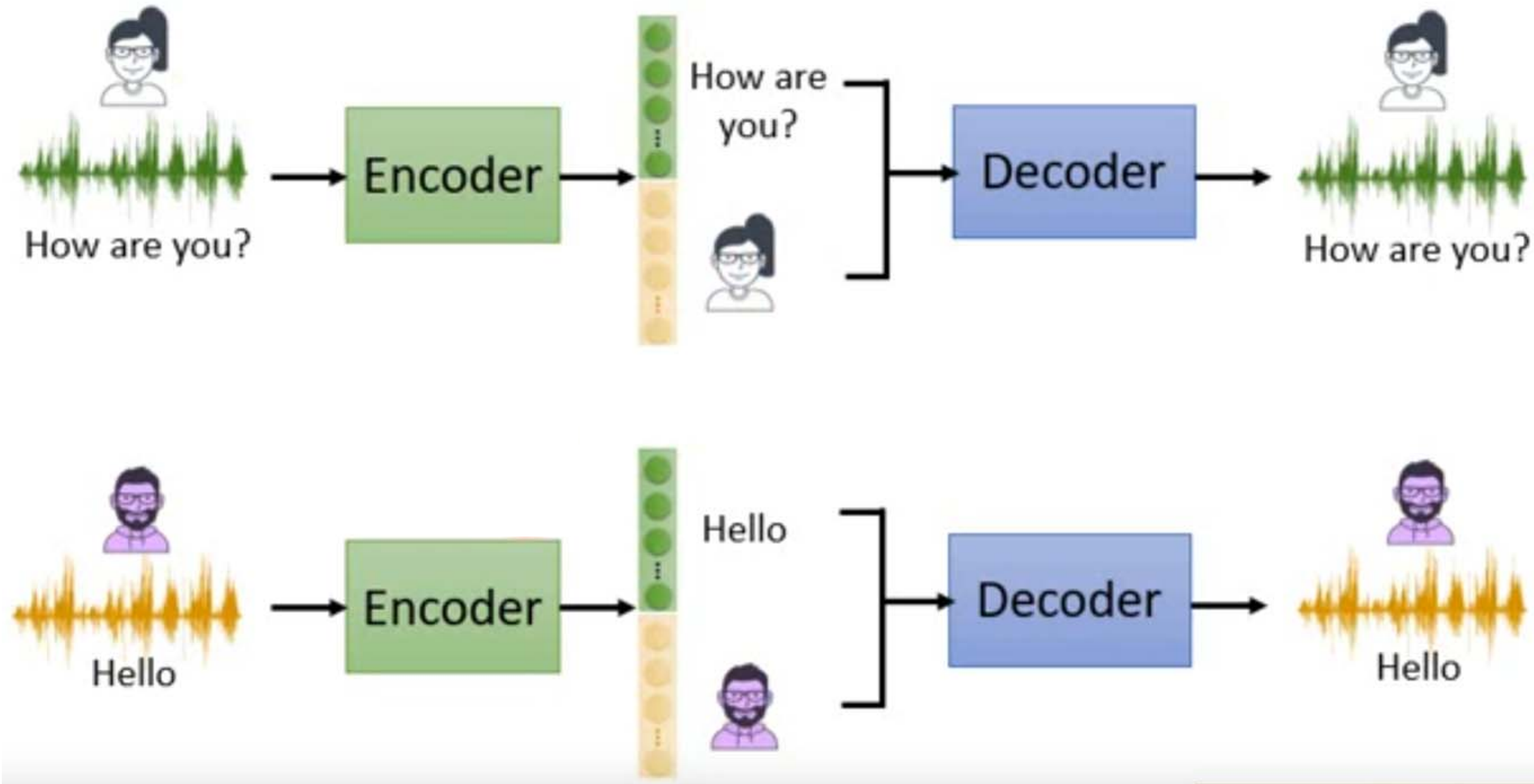
深入理解 Auto-encoder

- 让嵌入特征具有可解释性
- **Feature disentangle**
  - 输入数据包含多个方面的信



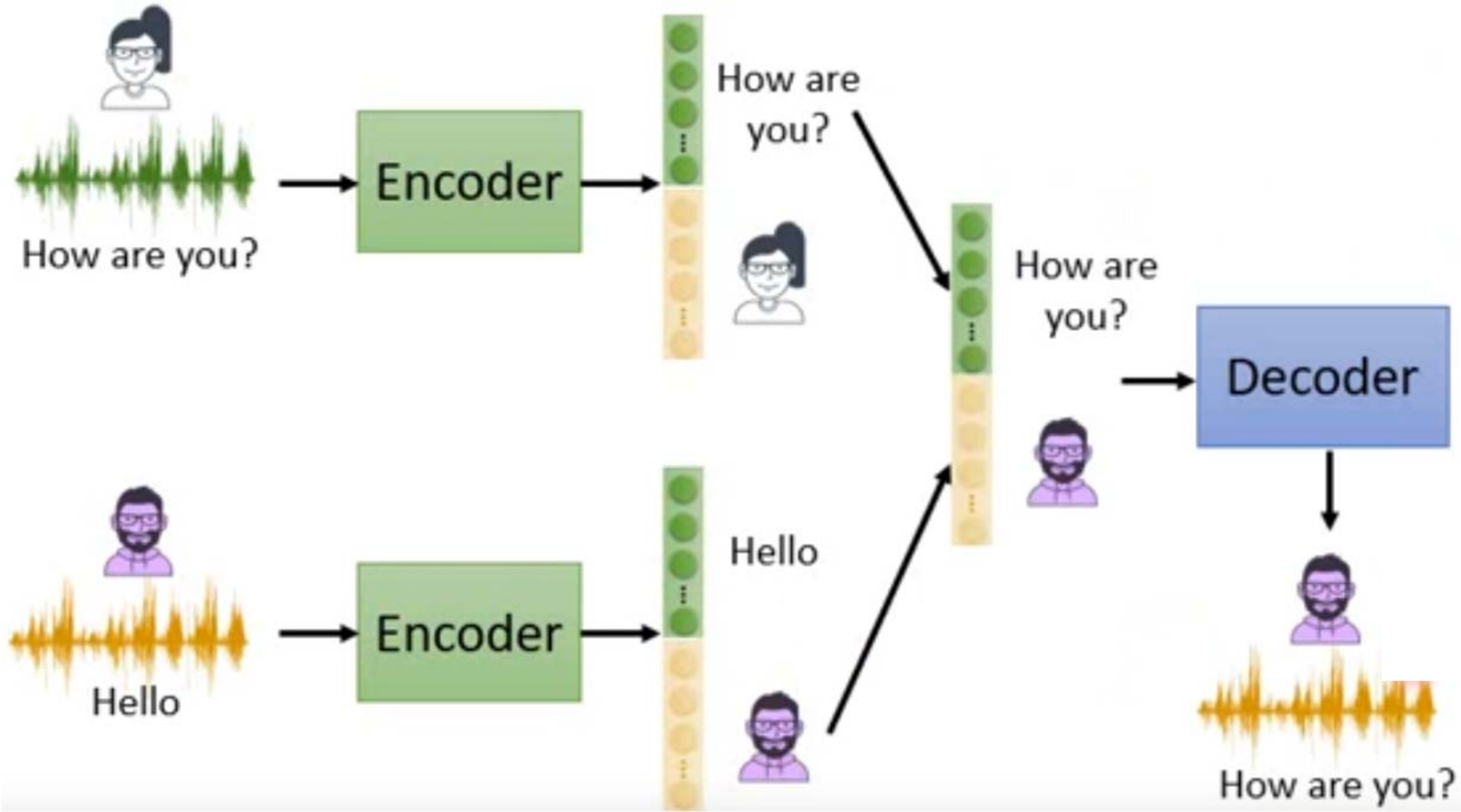
深入理解 Auto-encoder

- 让嵌入特征具有可解释性
- **Feature disentangle—语音转换**



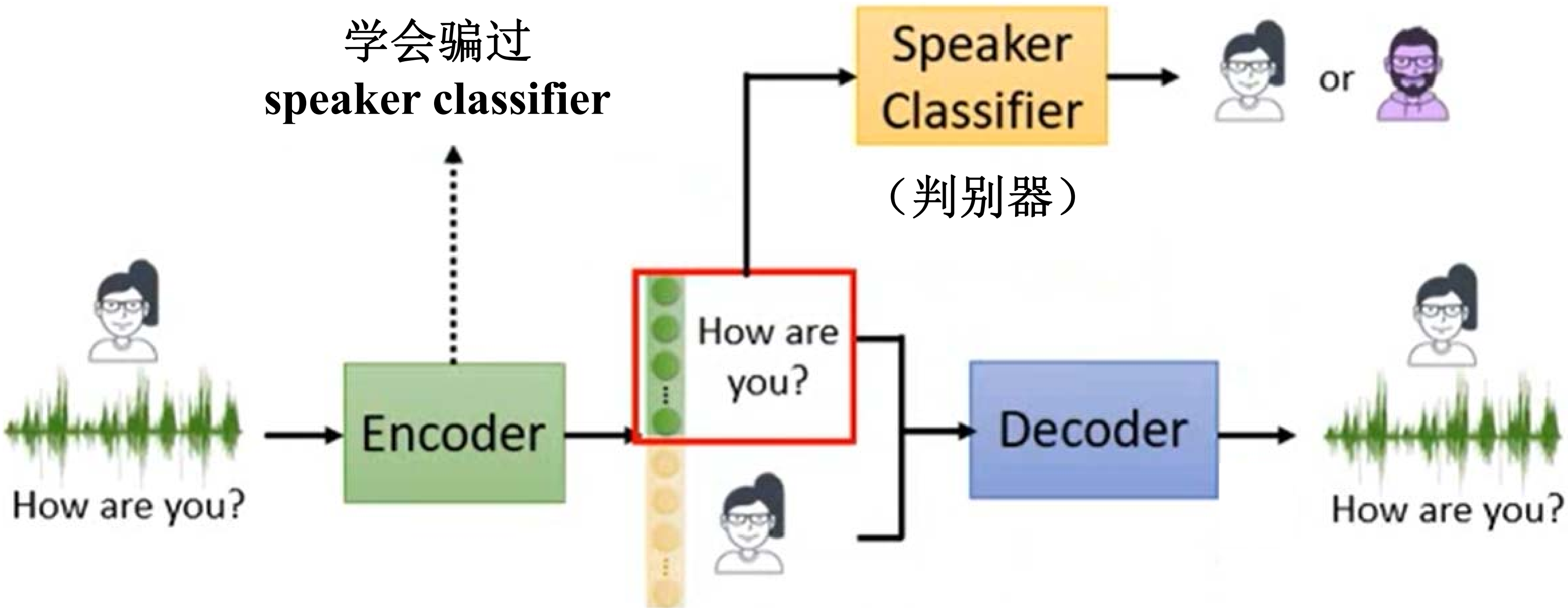
深入理解 Auto-encoder

- 让嵌入特征具有可解释性
- **Feature disentangle—语音转换**



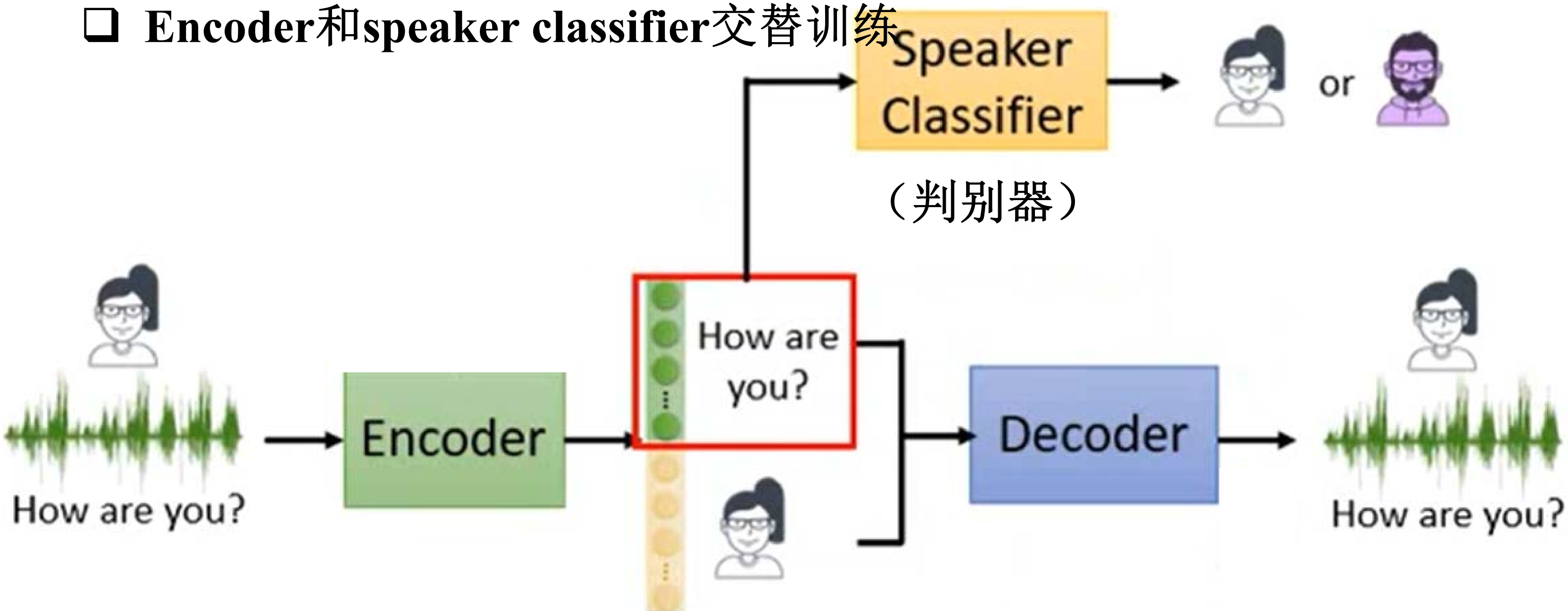
深入理解 Auto-encoder

- 让嵌入特征具有可解释性
- **Feature disentangle**—语音转换
  - **Encoder**学会骗过**speaker classifier**，使其正确率最低，说明**code**的前段特征没有语者信息



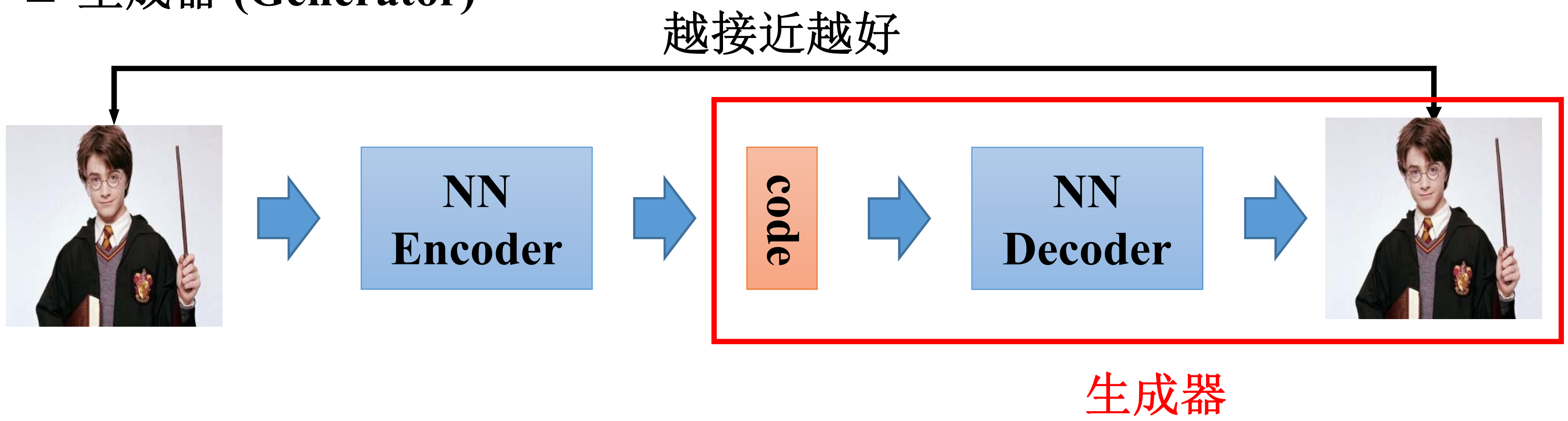
深入理解 Auto-encoder

- 让嵌入特征具有可解释性
- **Feature disentangle—语音转换**
  - ❑ **Encoder学会骗过speaker classifier，使其正确率最低，说明code的前段特征没有语者信息**
  - ❑ **Encoder和speaker classifier交替训练**



更多应用

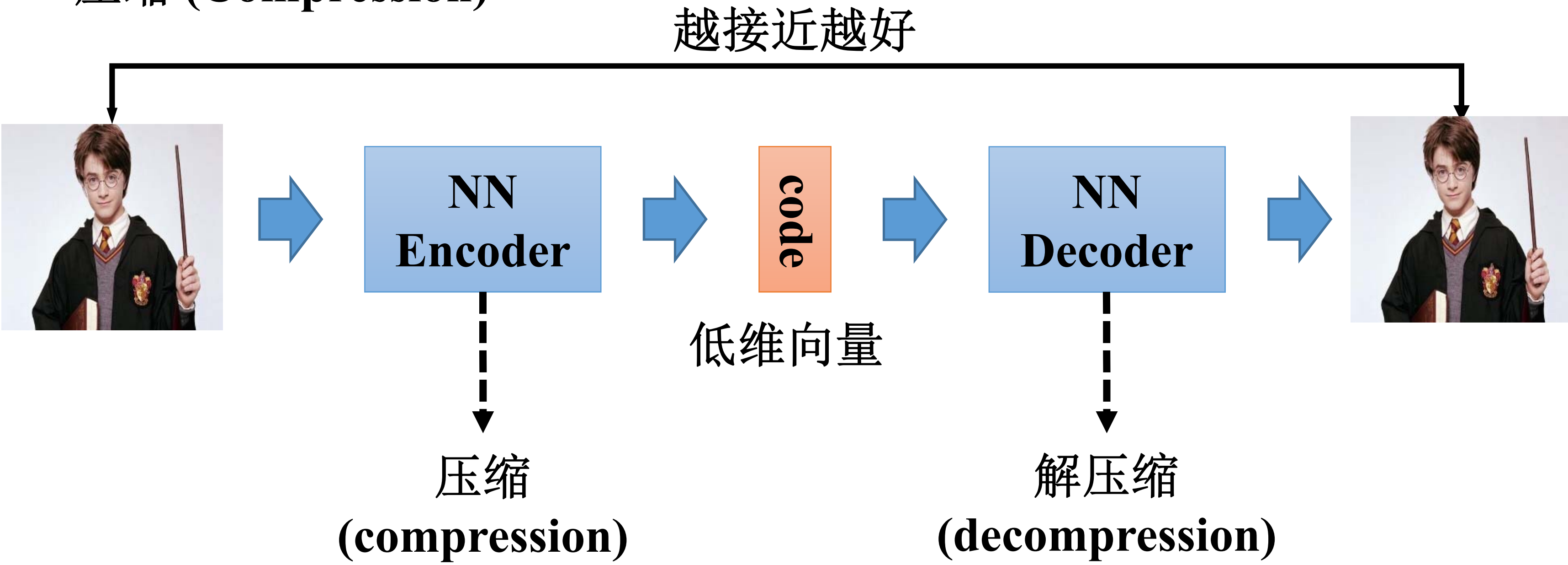
■ 生成器 (Generator)



- ☐ 生成对抗网络 (Generative Adversarial Network, GAN)
- ☐ 变分自编码器 (Variational auto-encoder, VAE)

更多应用

■ 压缩 (Compression)

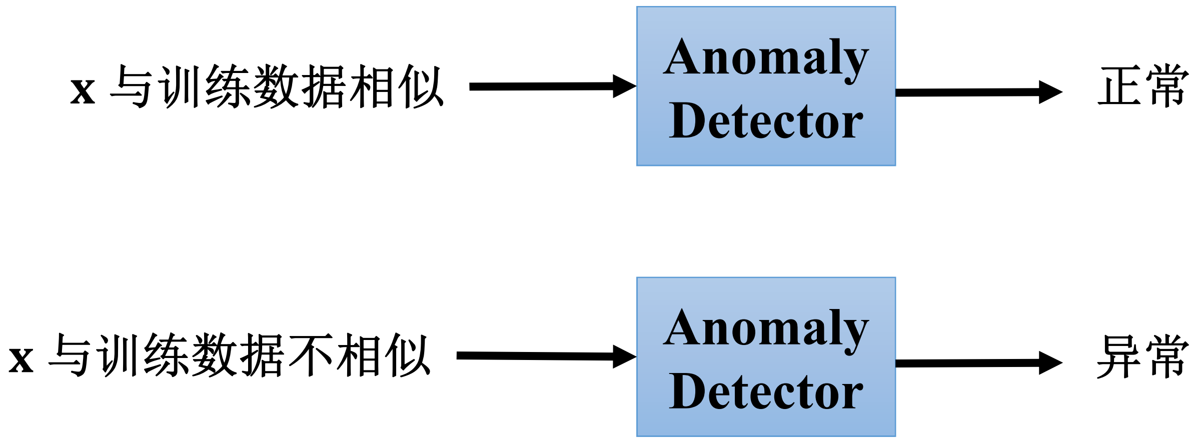


□ 用压缩结果重建的图像会失真

更多应用

■ 异常检测(Anomaly Detection)

- □ 已知训练数据 $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$
- □ 检测输入数据 **x** 是否与训练数据相似



更多应用

- 异常检测(Anomaly Detection)
  - 已知训练数据 $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$
  - 检测输入数据 **x** 是否与训练数据相似

训练数据

正常

训练数据

异常

更多应用

- 异常检测(**Anomaly Detection**)
  - 已知训练数据 $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$
  - 检测输入数据 **x** 是否与训练数据相似
  - 二值分类，但是训练数据只有一类

更多应用

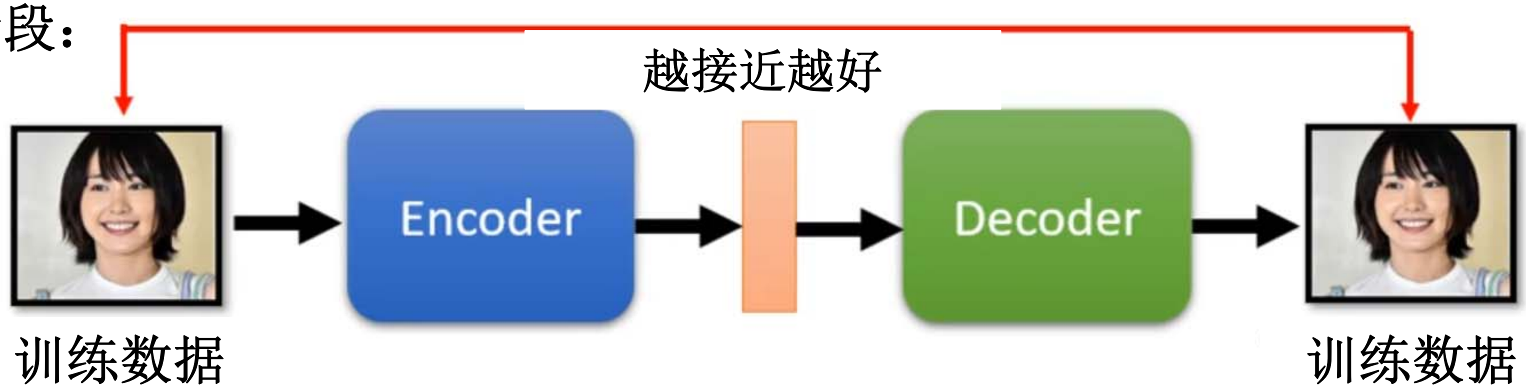
- 异常检测(**Anomaly Detection**)
  - 欺诈侦测(**Fraud Detection**)
    - 训练数据：信用卡消费记录
    - **https://www.kaggle.com/ntnu-testimon/paysim1/home**
  - 网络侵入检测(**Network Intrusion Detection**)
    - 训练数据：连接行为，**x:** 是否受到攻击
    - **http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html**
  - 癌细胞检测(**Cancer Detection**)
    - 训练数据：正常细胞，**x:**是否是癌细胞
    - **https://www.kaggle.com/uciml/breast-cancer-wisconsin-data/home**

更多应用

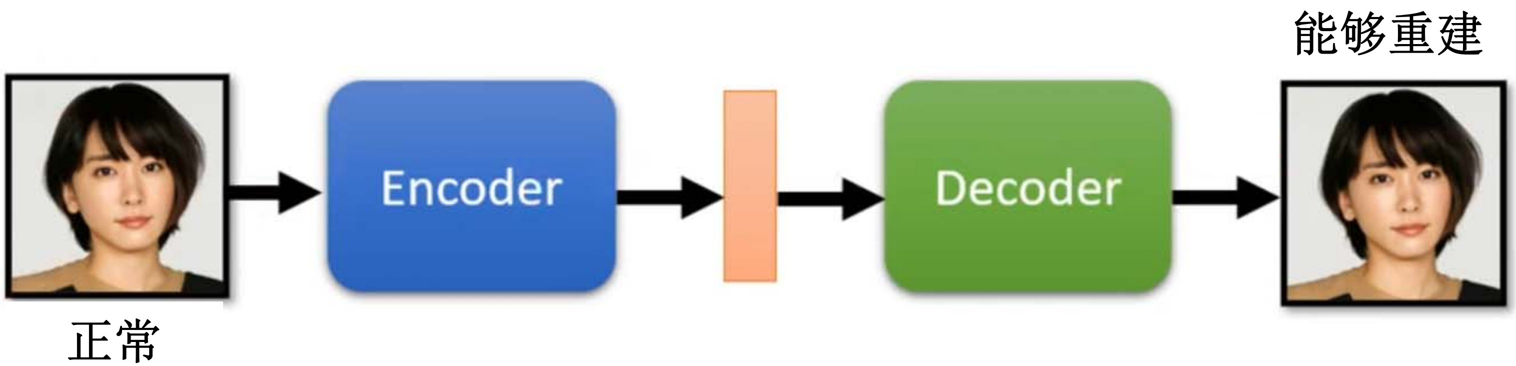
■ 异常检测(Anomaly Detection)

□ 训练auto-encoder：是否是真实的人脸

训练阶段：



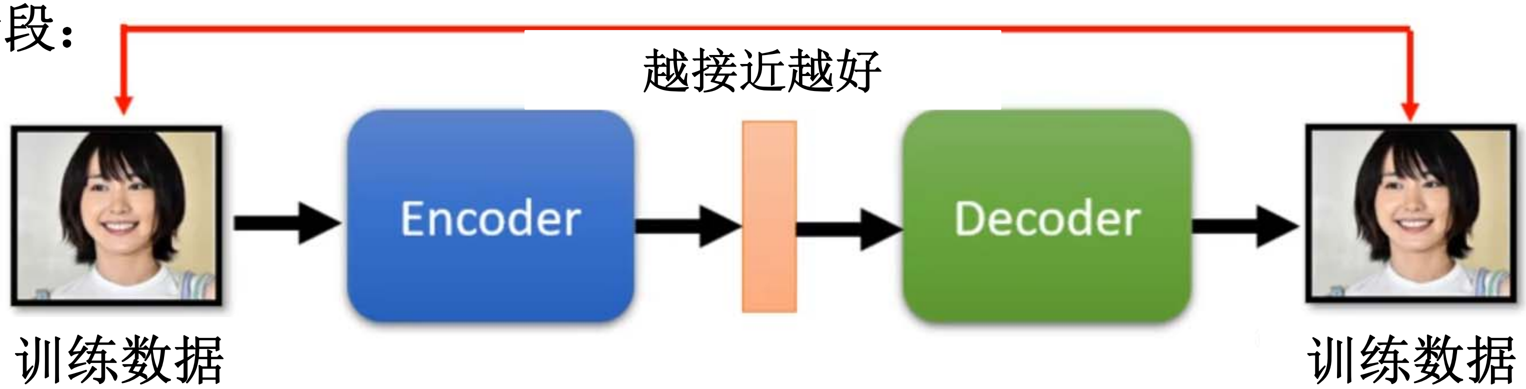
测试阶段：



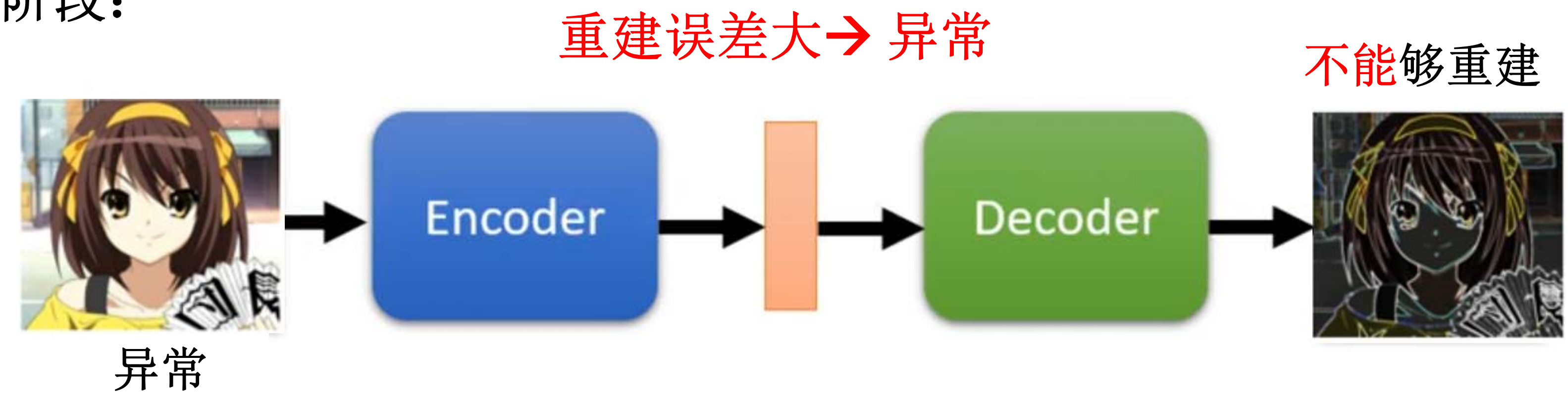
更多应用

- 异常检测(Anomaly Detection)
  - 训练auto-encoder：是否是真实的人脸

训练阶段：



测试阶段：

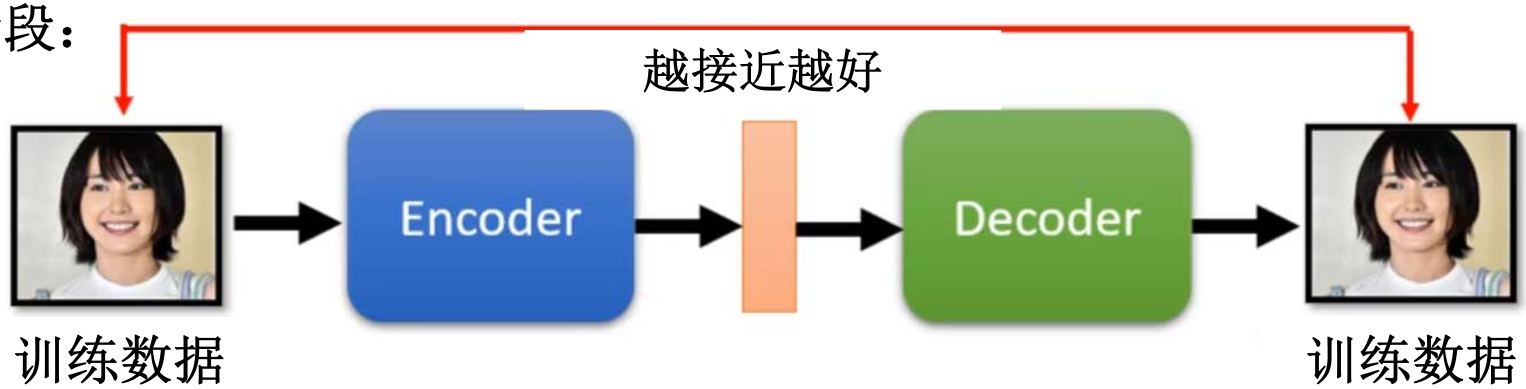


更多应用

■ 异常检测(Anomaly Detection)

□ 训练auto-encoder：是否是真实的人脸

训练阶段：



测试阶段：