VAE

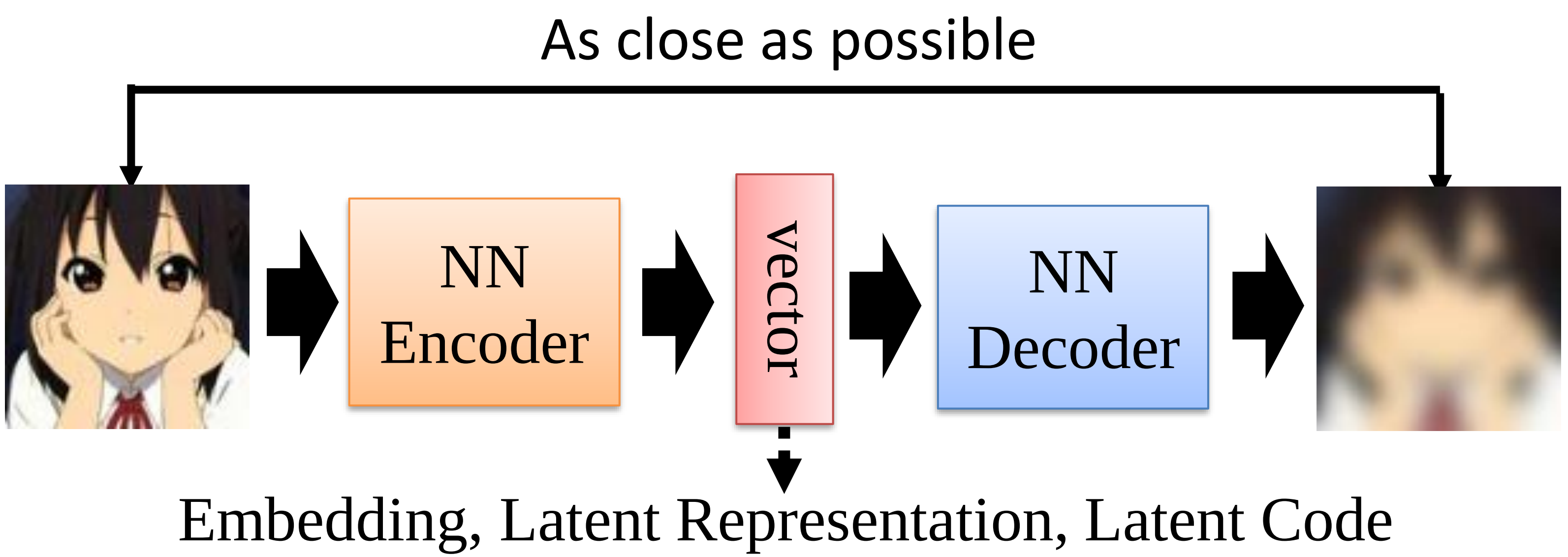
赵洲

浙江大学计算机学院

VAE模型家族

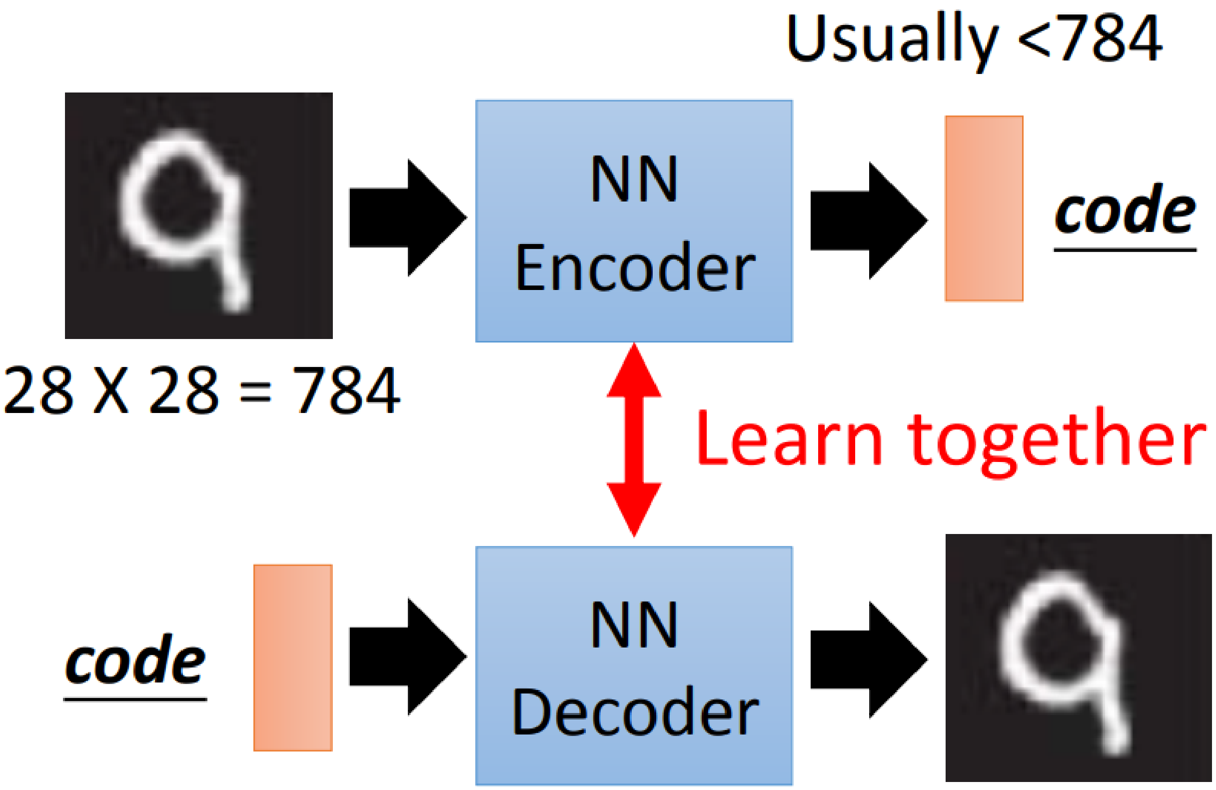
- AE
- VAE

Auto-encoder



- More than minimizing reconstruction error
- More interpretable embedding

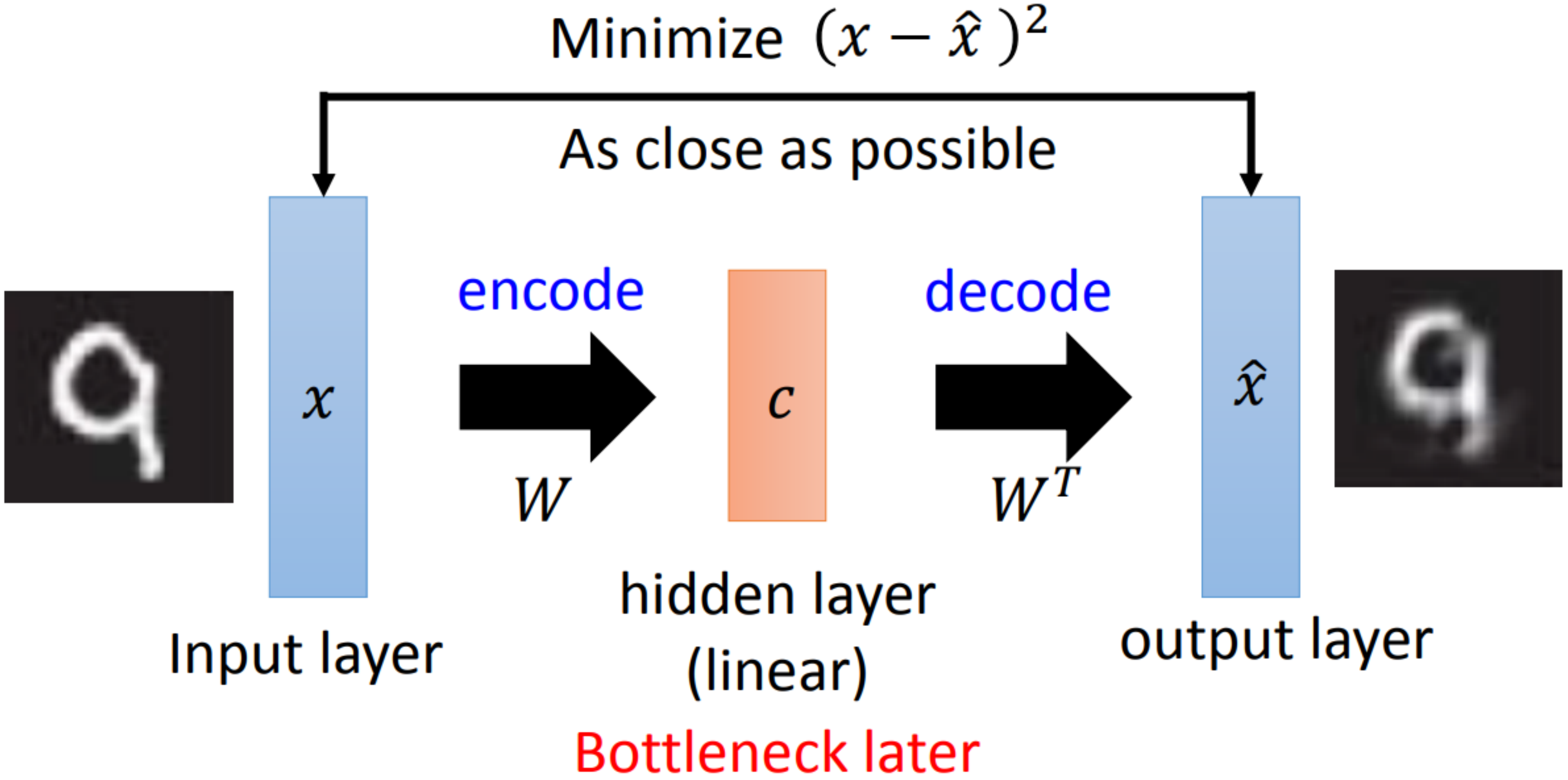
Auto-encoder (MINIST)



目标物体
紧凑表示

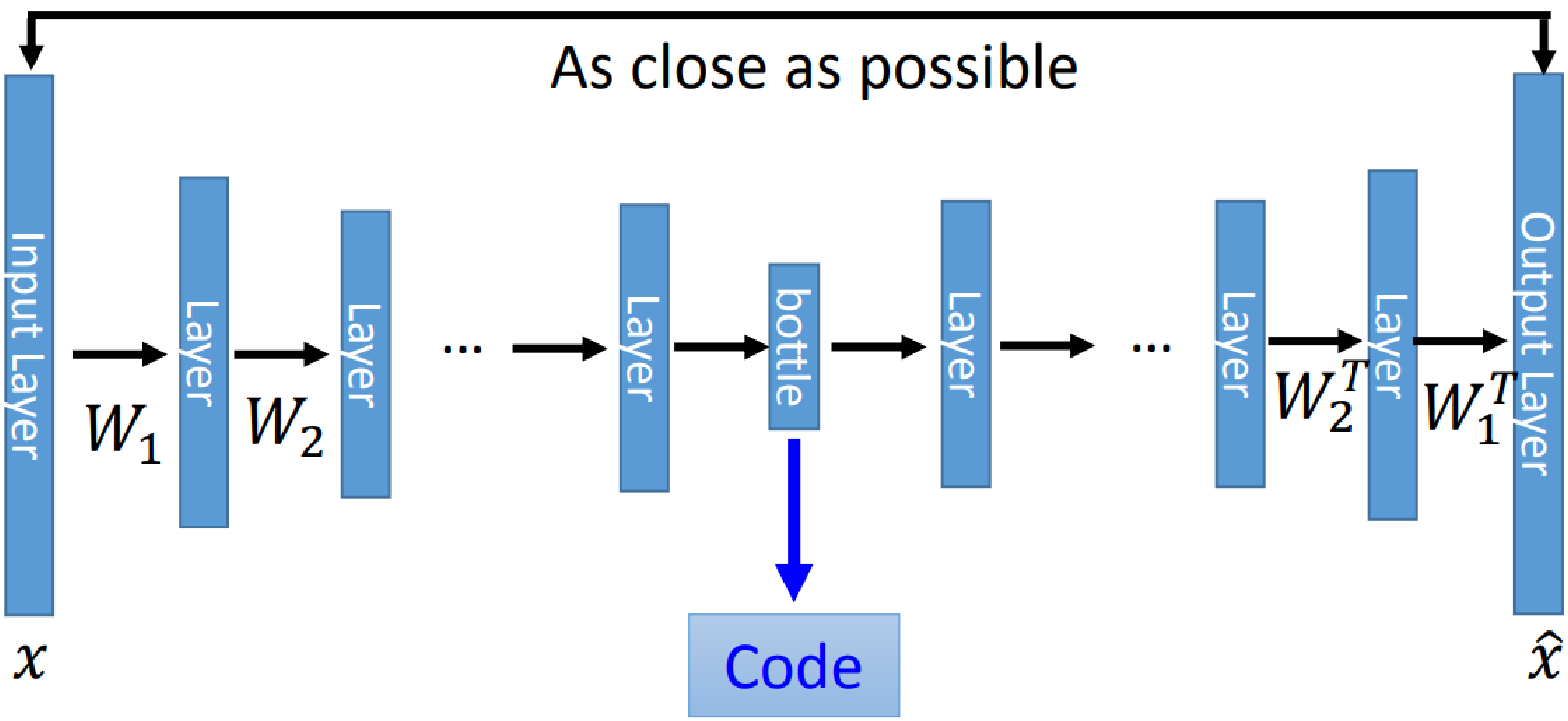
目标物体
数据重构

与PCA等降维方法联系

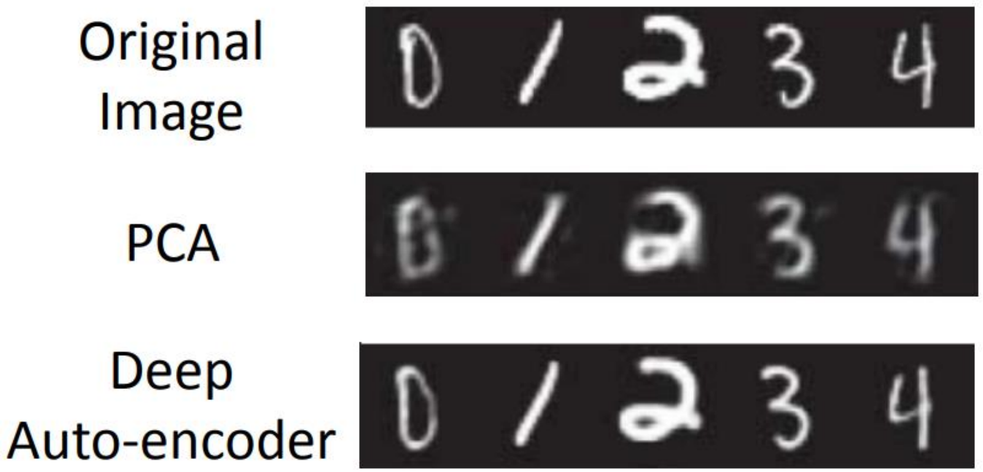


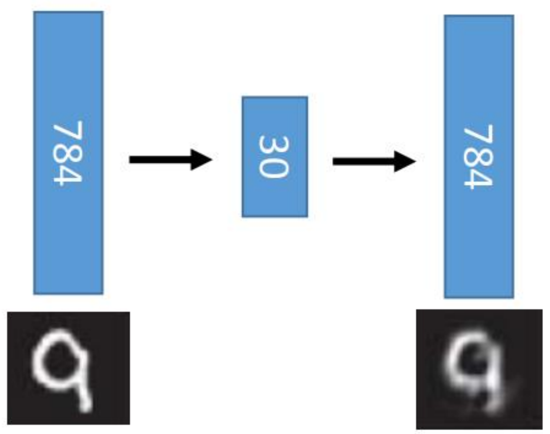
线性层的输出是目标物体的编码

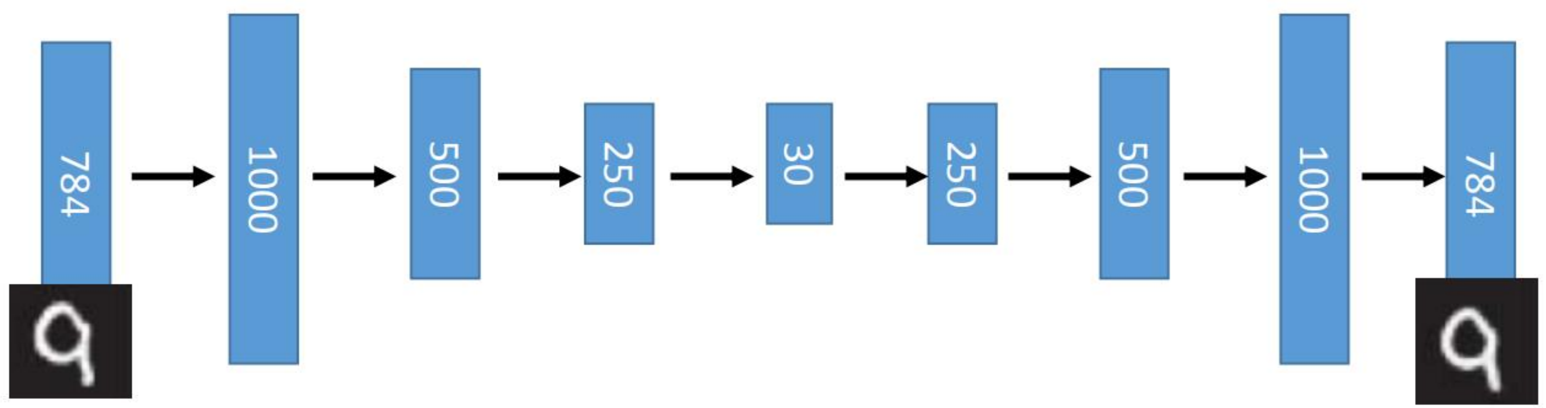
深度Auto-encoder



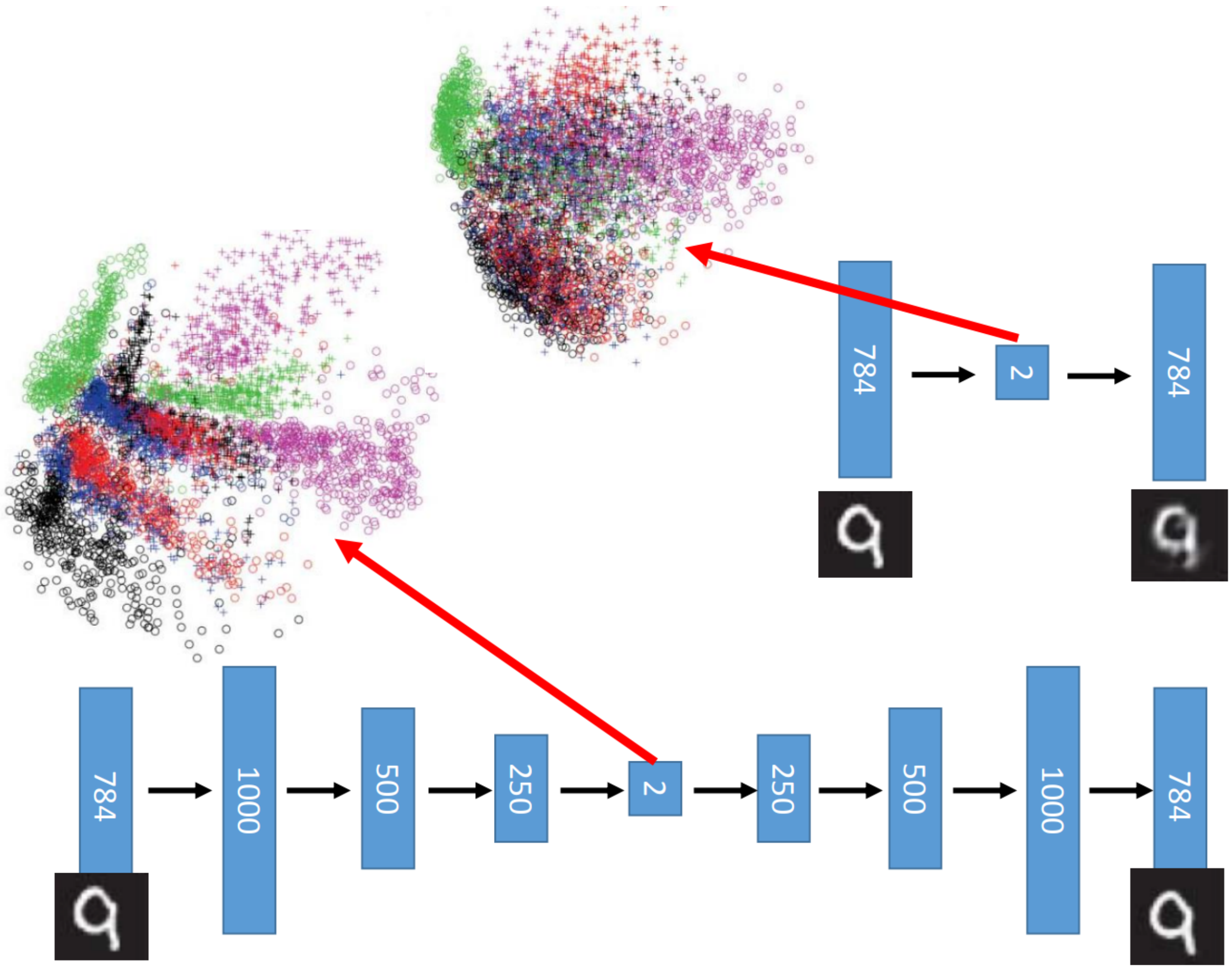
与PCA等降维方法对比



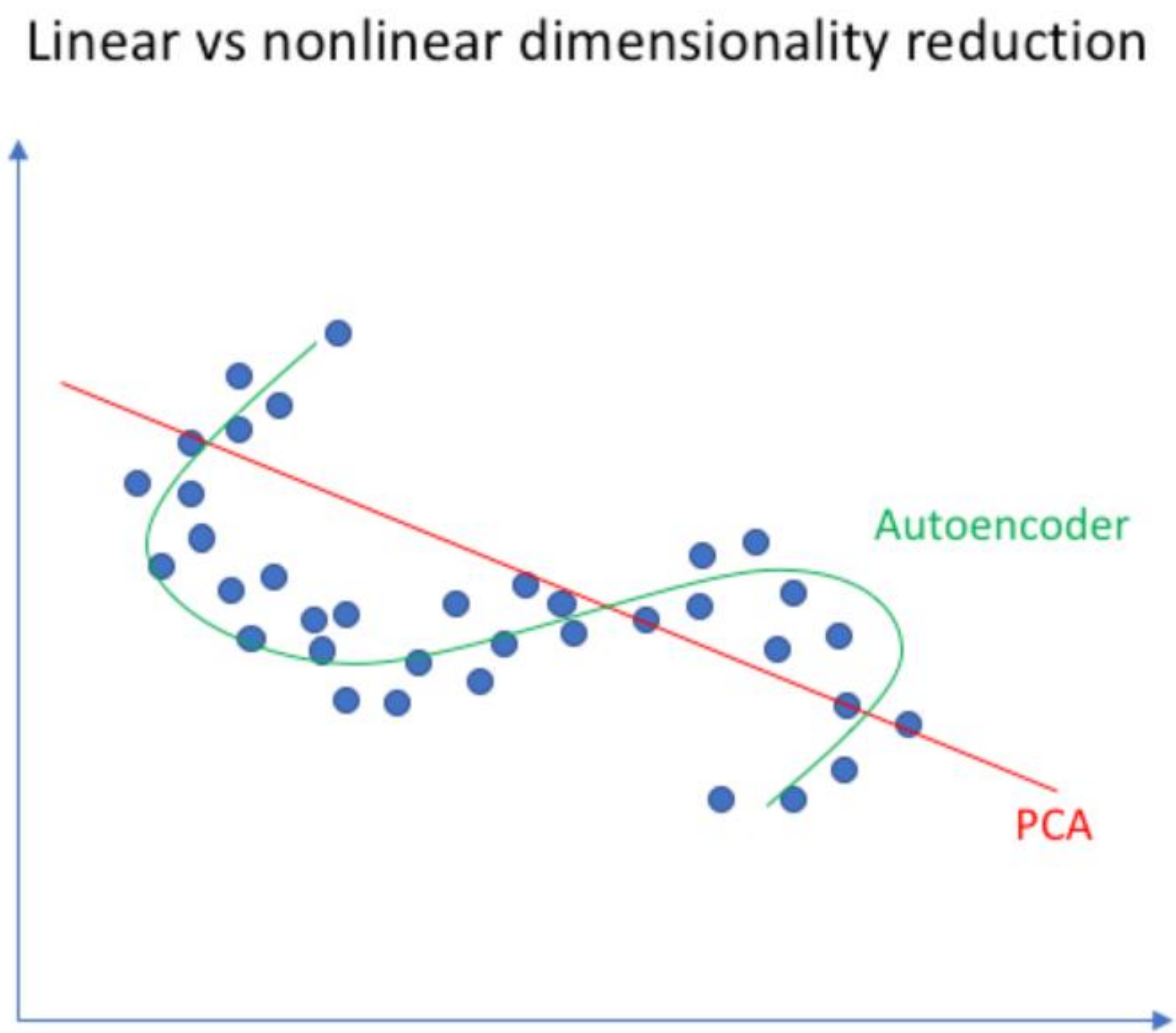




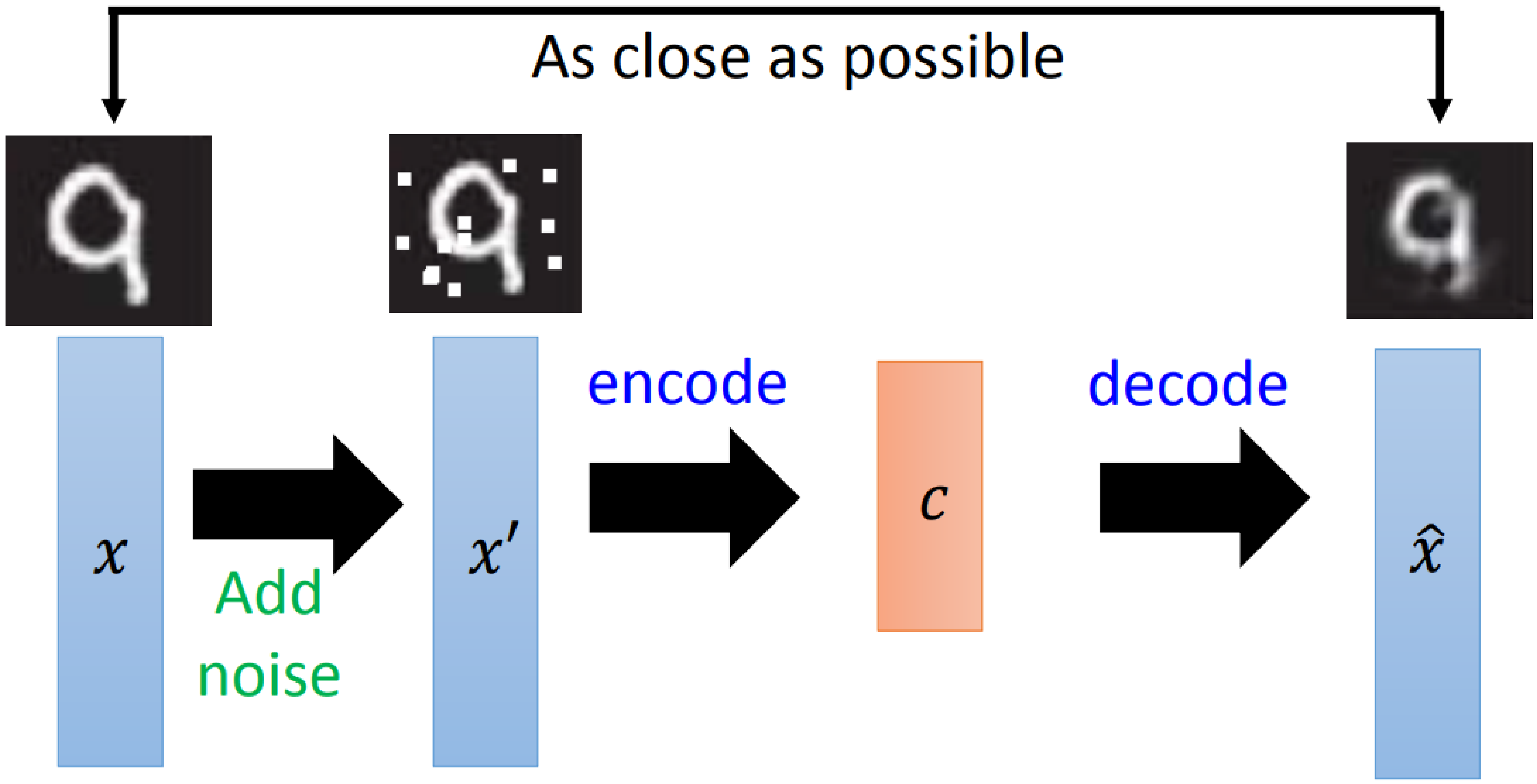
编码可视化



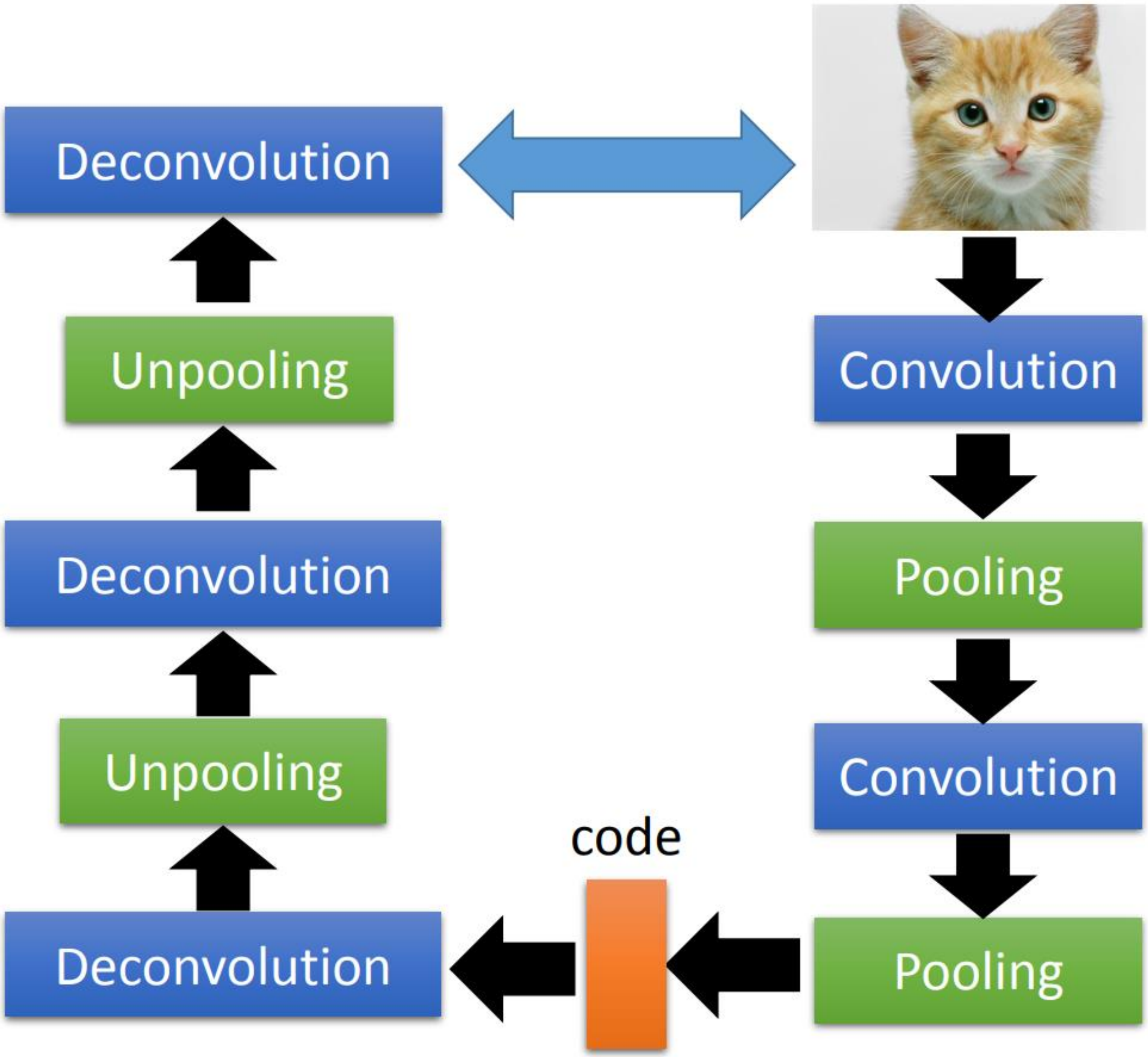
PCA的非线性泛化


Linear vs nonlinear dimensionality reduction
Autoencoder
PCA

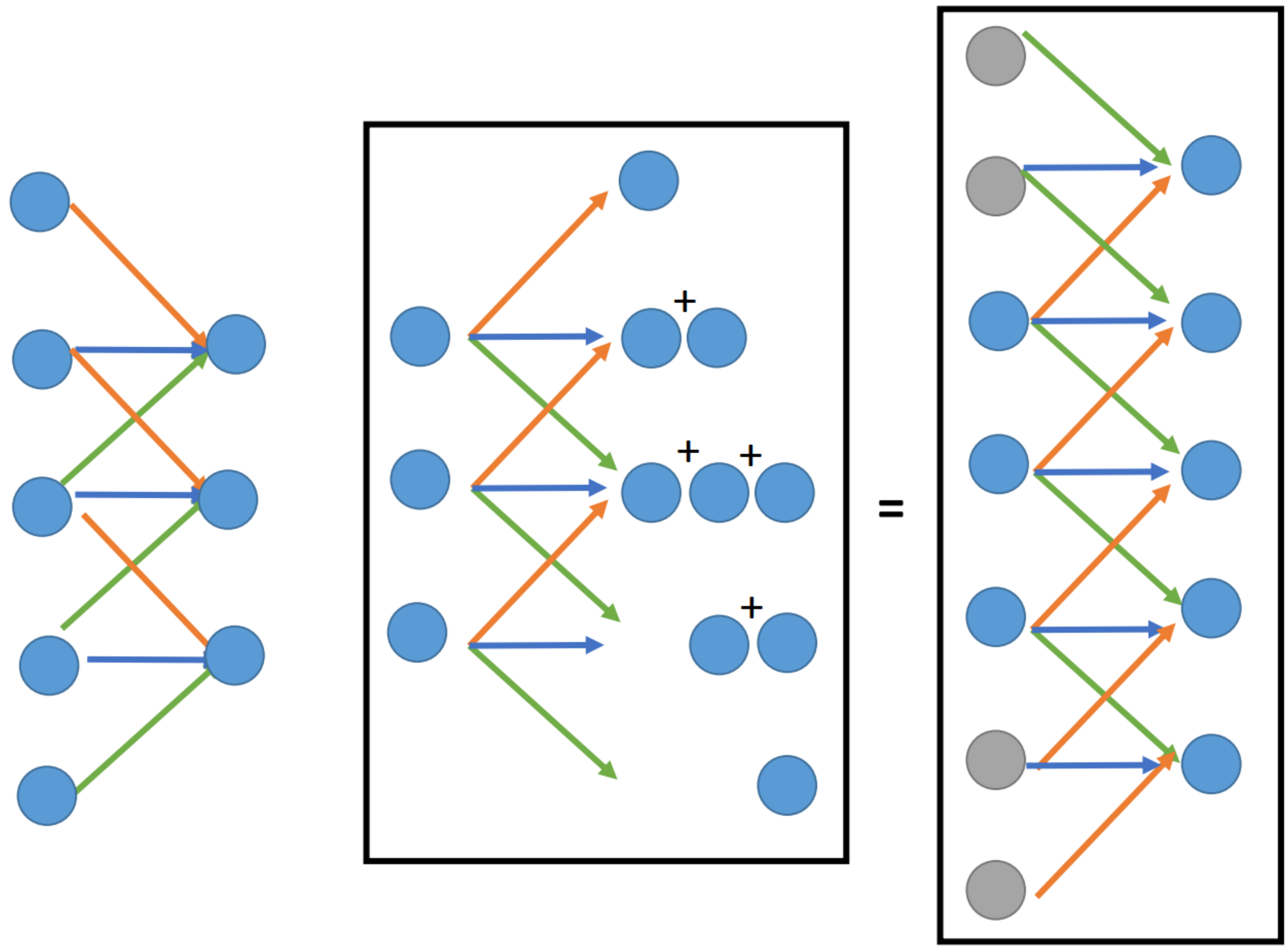
去噪Auto-encoder



基于CNN的Auto-encoder

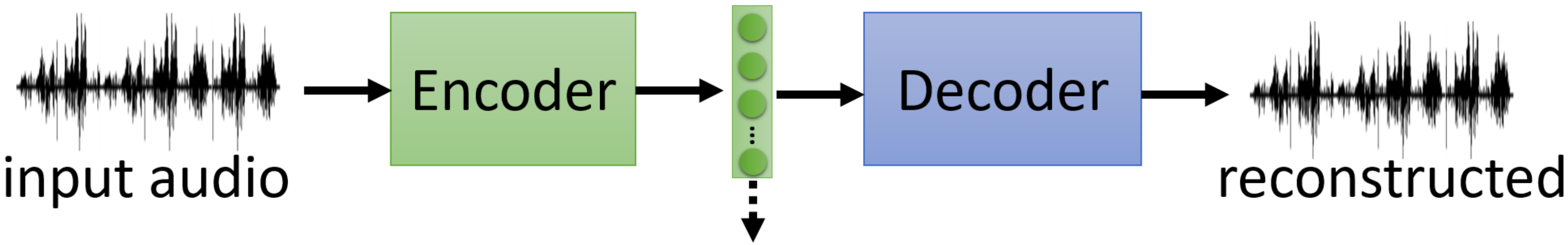


Deconvolution

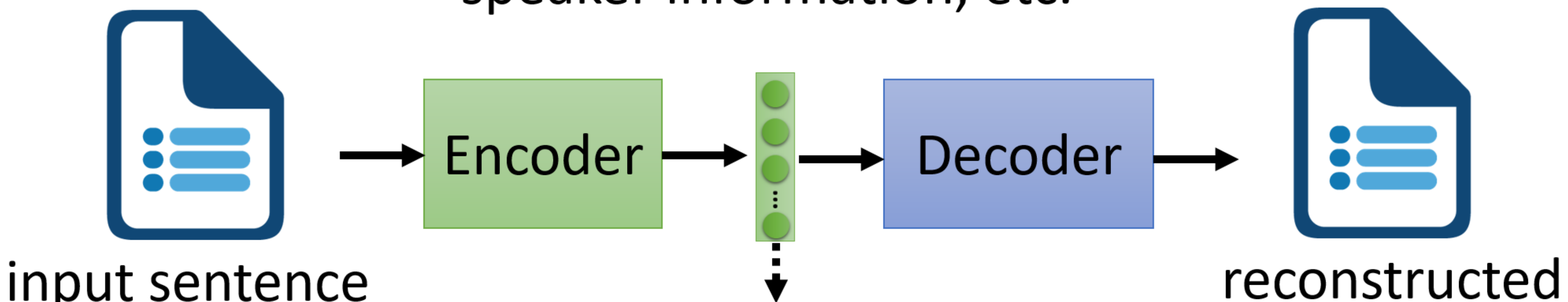


Auto-encoder (特征解耦)

An object contains multiple aspect information

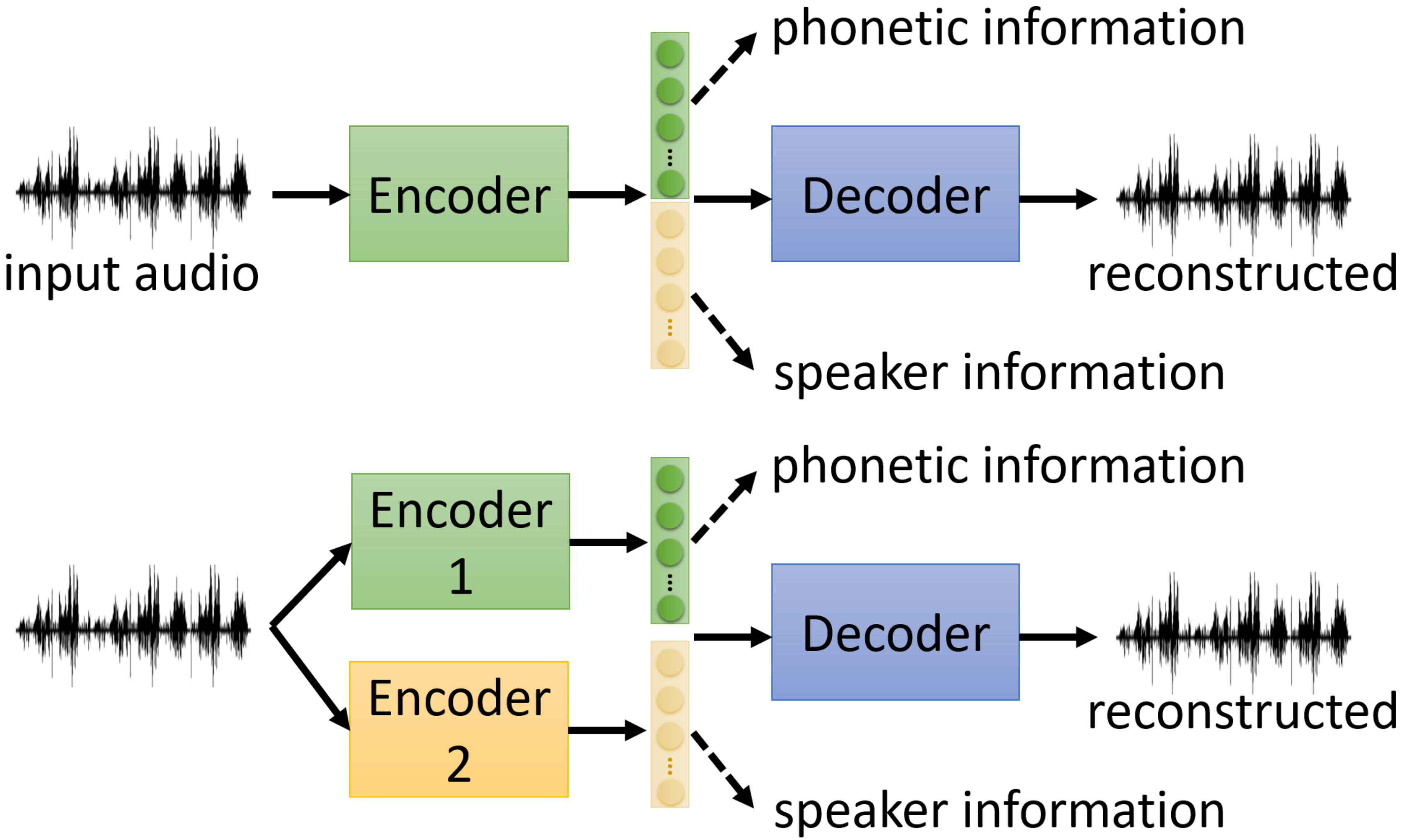


Include phonetic information,
speaker information, etc.

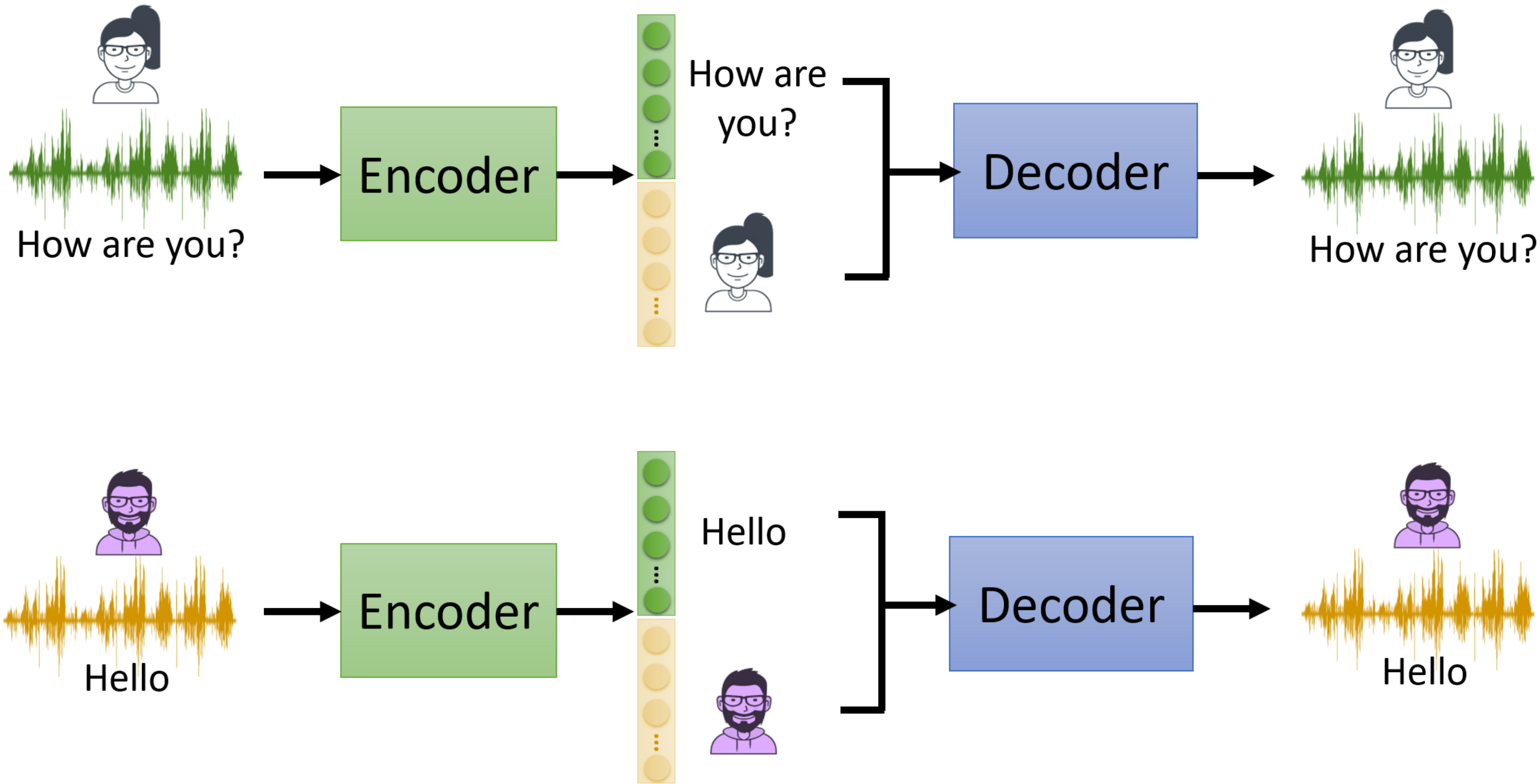


Include syntactic information,
semantic information, etc.

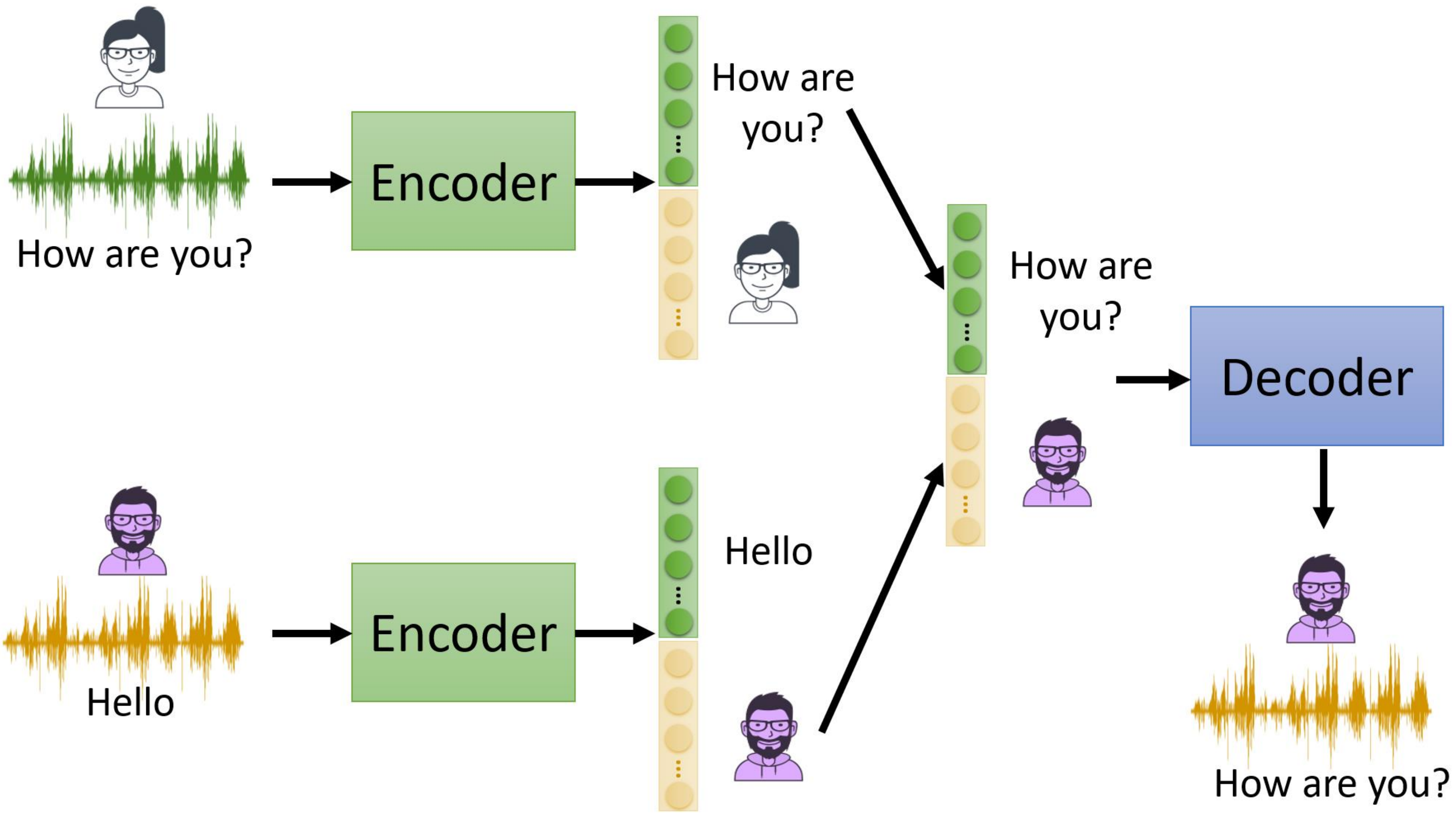
特征解耦


phonetic information
input audio
Encoder
Decoder
reconstructed
speaker information
phonetic information
Encoder 1
Encoder 2
Decoder
reconstructed
speaker information

语音转换


How are you?
Encoder
How are
you?
Decoder
How are you?
Hello
Encoder
Hello
Decoder
Hello

语音转换



How are you?
Encoder
How are
you?
Hello
Encoder
Hello
How are
you?
Decoder
How are you?

例子

Do you want to study a PhD?

Go away!

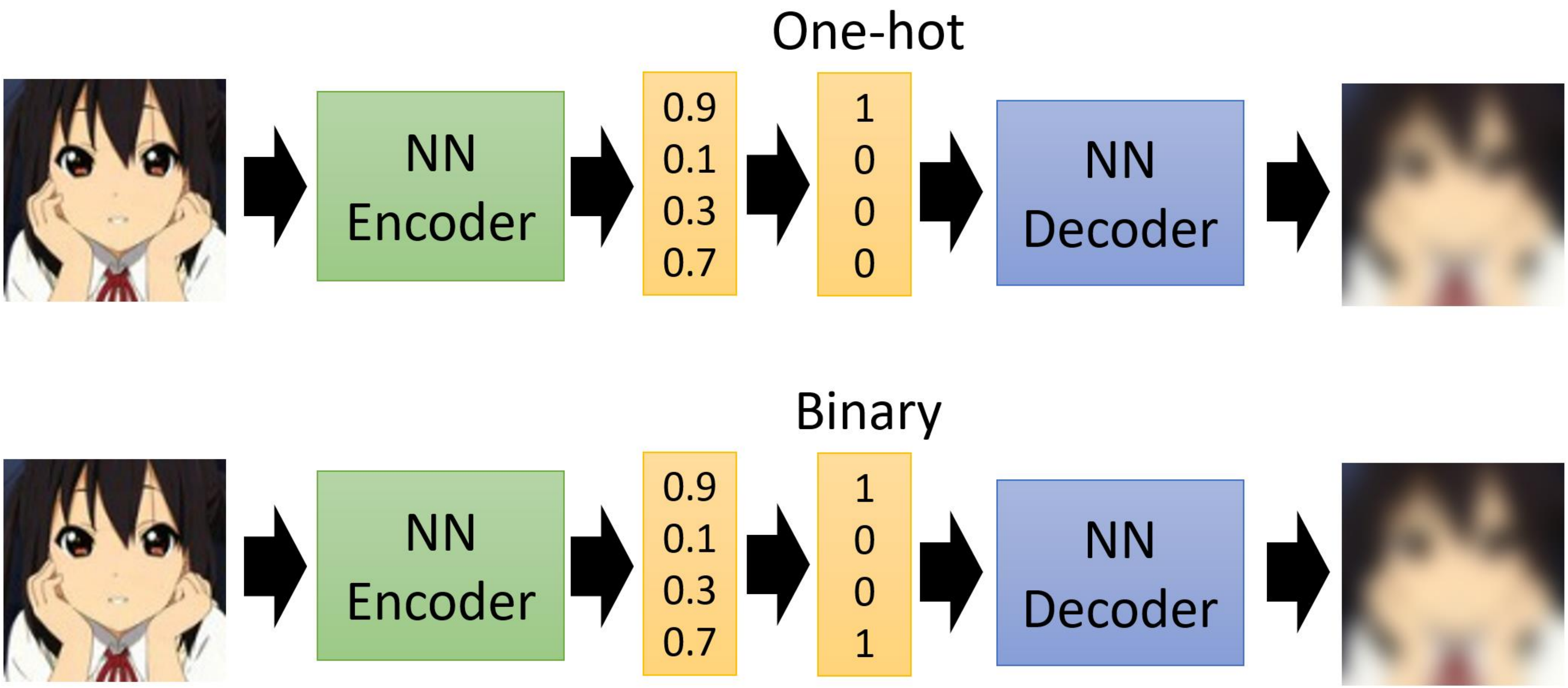
Student

新垣結衣
(Aragaki Yui)

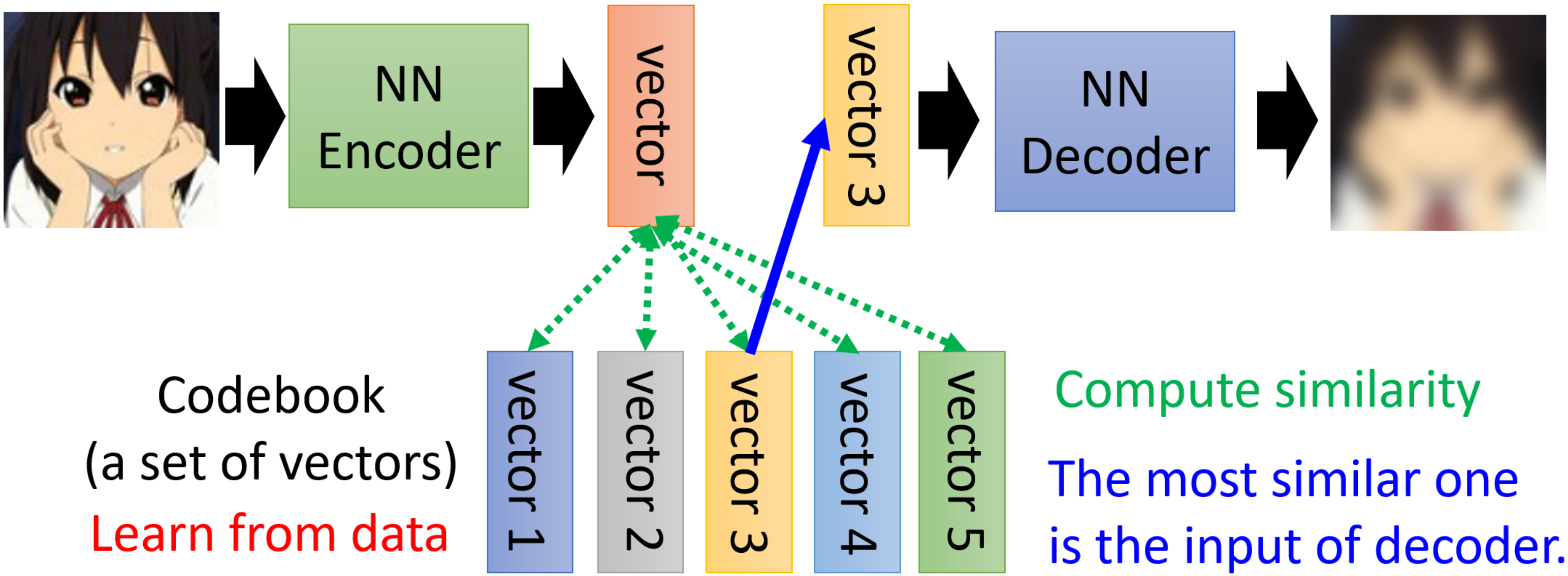
Do you want to study a PhD?

Student

离散表示（VQ-VAE）


One-hot
NN Encoder
0.9
0.1
0.3
0.7
1
0
0
0
NN Decoder
Binary
NN Encoder
0.9
0.1
0.3
0.7
1
0
0
1
NN Decoder

离散表示（VQ-VAE）

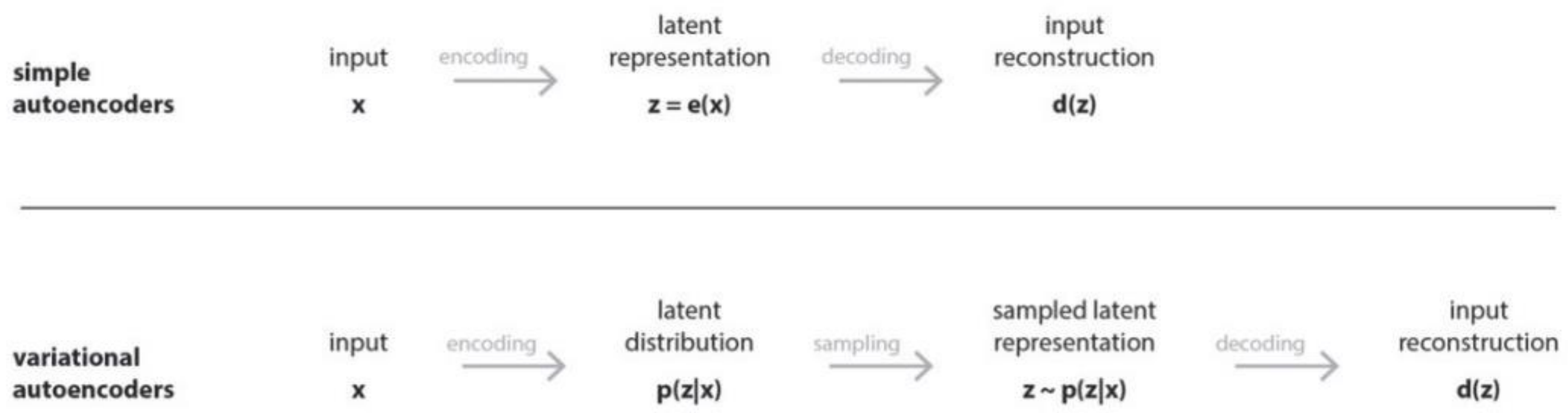

NN Encoder
vector
vector 3
NN Decoder
Codebook
(a set of vectors)
Learn from data
vector 1
vector 2
vector 3
vector 4
vector 5
Compute similarity
The most similar one is the input of decoder.

AE的问题

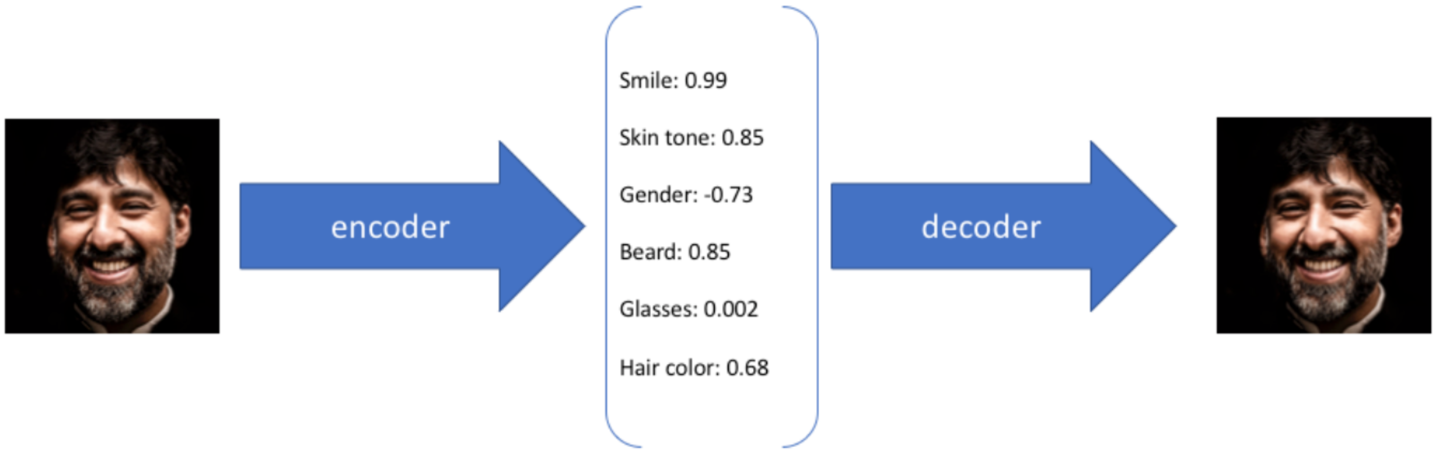
- AE通过编码器z = g(X), 将每个图片编码成向量z；它的解码器f(z)利用编码向量z来重构原始图片。
- 当AE作为合成模型，针对随机生成的编码向量z, f(z)只会生成一些没有意义的噪声。
- 原因在与AE没有对z的分布p(z)进行建模，所以不确定哪些z能够生成有用的图片（训练f(z)数据有限，f只能对极有限的z响应，而编码向量是一个太大的空间）。

从AE到VAE

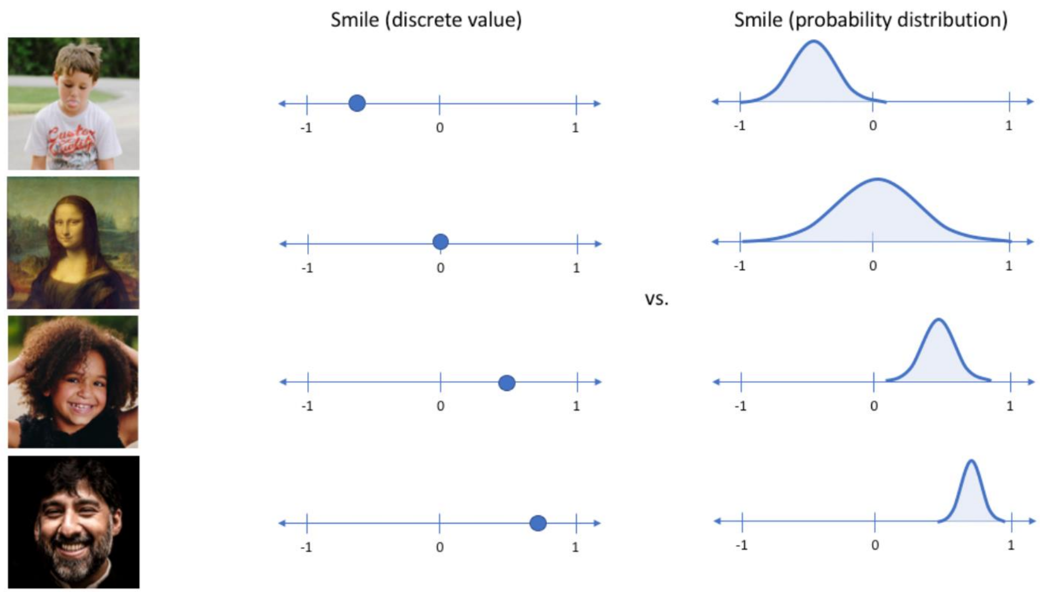
- 在AE的基础上，显性对z的分布p(z)进行建模，使得自编码器成为一个合格的生成模型（VAE）。



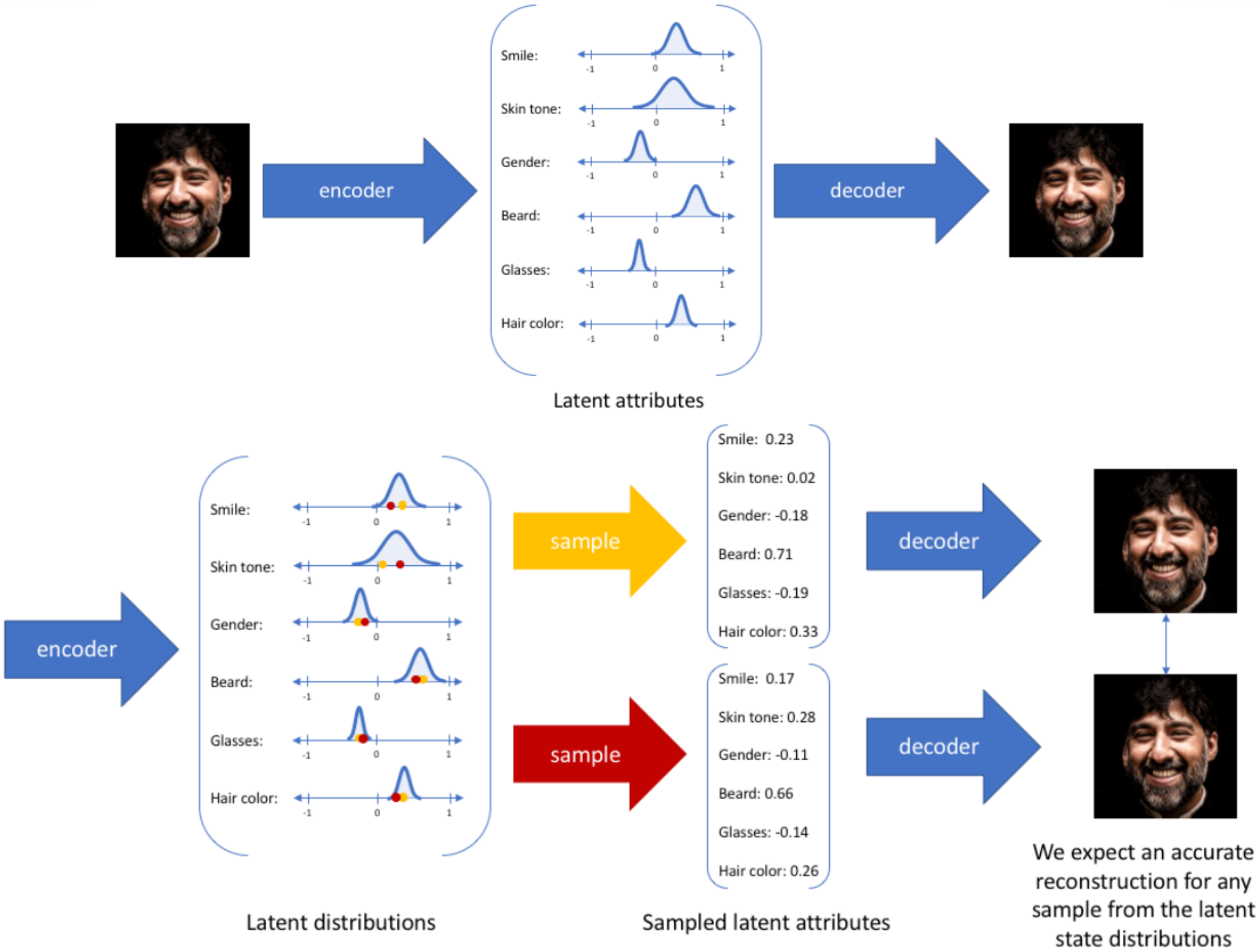
例子解释



Latent attributes

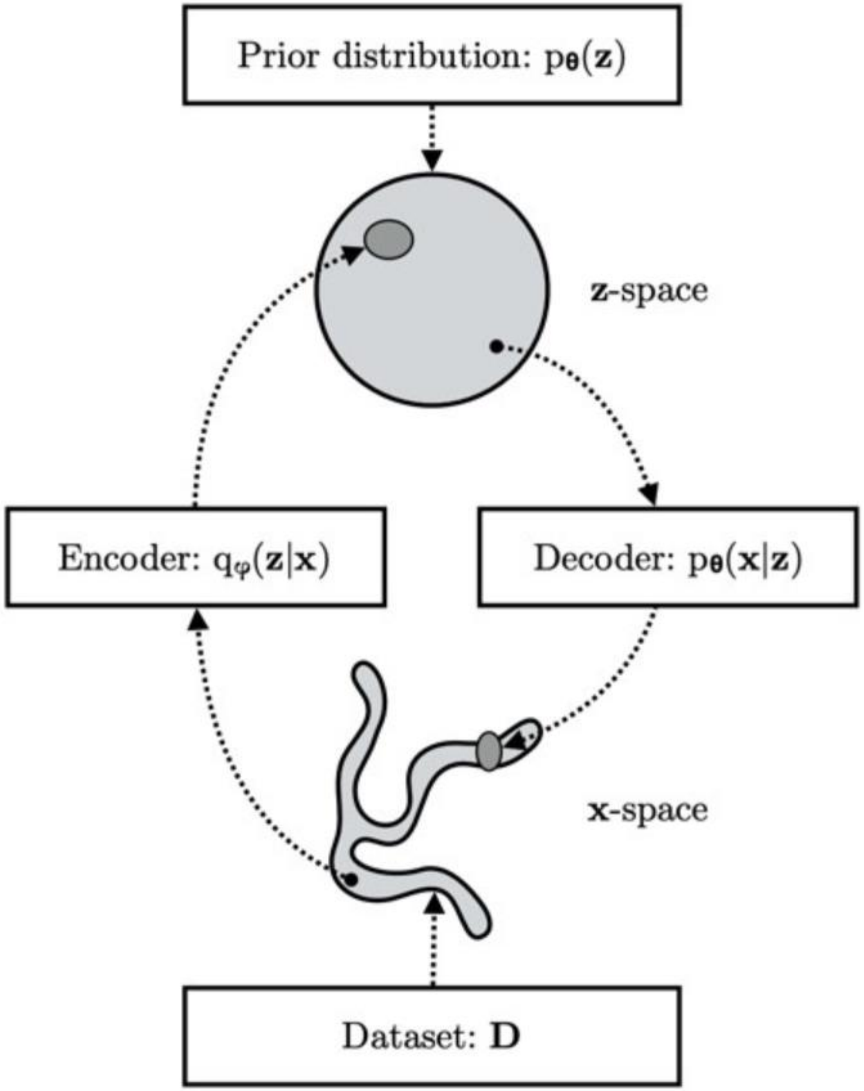


例子解释



Smile:
Skin tone:
Gender:
Beard:
Glasses:
Hair color:
encoder
decoder
Latent attributes
encoder
Smile:
Skin tone:
Gender:
Beard:
Glasses:
Hair color:
Latent distributions
sample
sample
Smile: 0.23
Skin tone: 0.02
Gender: -0.18
Beard: 0.71
Glasses: -0.19
Hair color: 0.33
Smile: 0.17
Skin tone: 0.28
Gender: -0.11
Beard: 0.66
Glasses: -0.14
Hair color: 0.26
Sampled latent attributes
decoder
decoder
We expect an accurate reconstruction for any sample from the latent state distributions

VAE的编码和解码过程



优化目标

$$\mathcal{L}_{VAE} = -\lambda D(q_\phi(z) \| p(z)) + \mathbb{E}_{p_{data}(x)} \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)]$$