



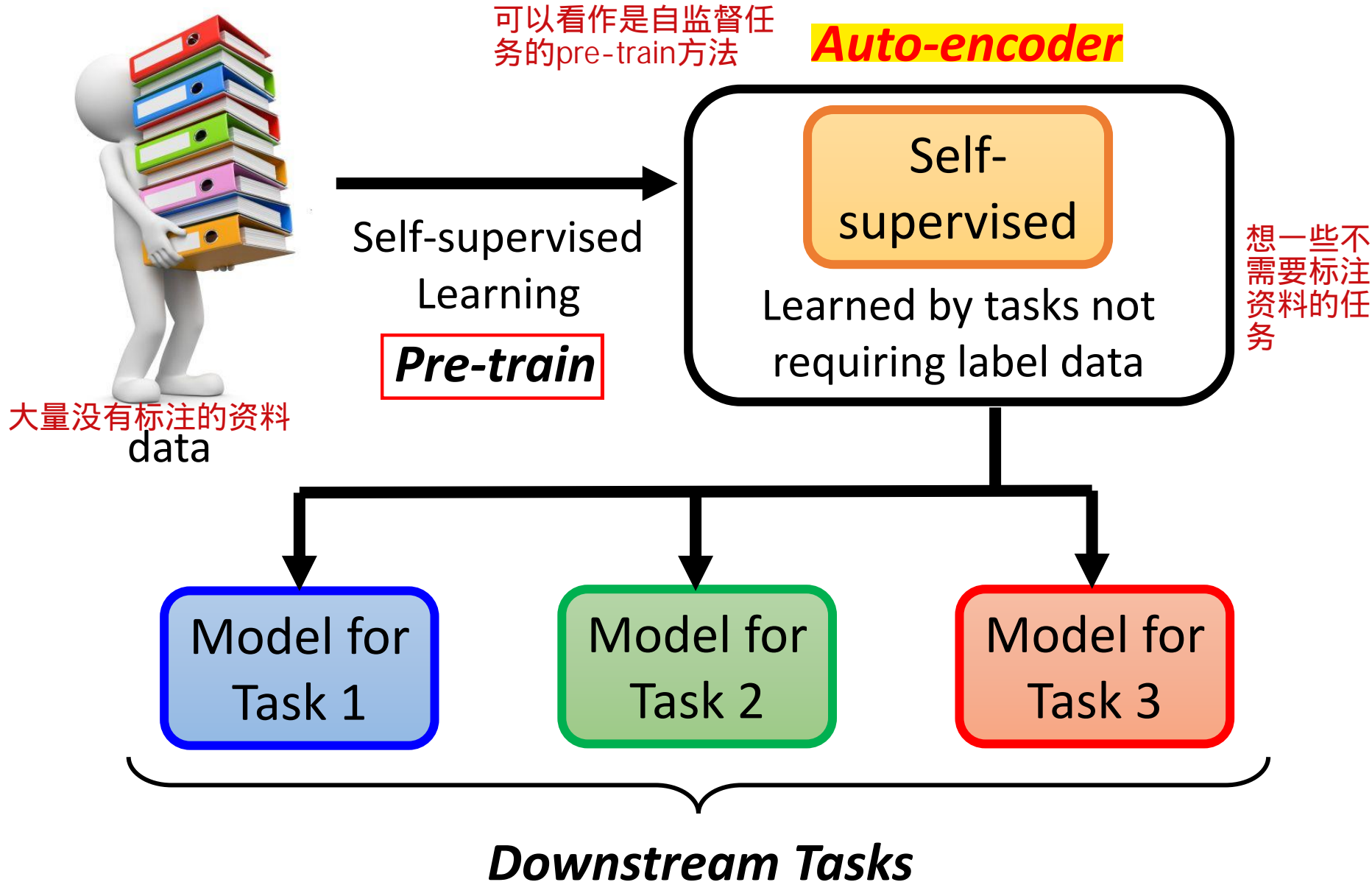
P53- Auto-encoder (上)

2021年5月1日
自编码器 (1) - 基本概念

AUTO-ENCODER

Hung-yi Lee 李宏毅

Self-supervised Learning Framework



Outline

Basic Idea of Auto-encoder

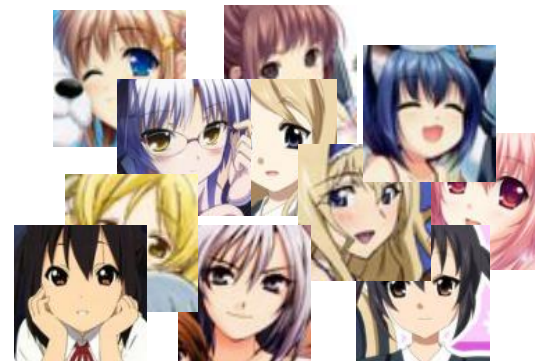
Feature Disentanglement

Discrete Latent Representation

More Applications

Auto-encoder

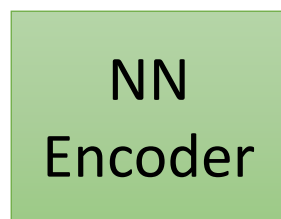
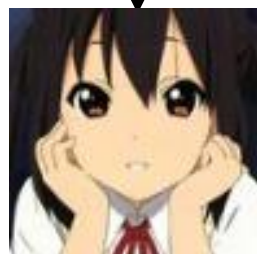
Unlabeled
Images



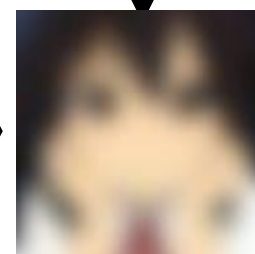
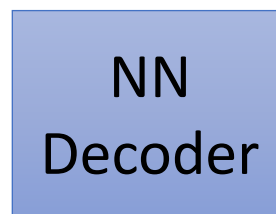
Sounds familiar? We have seen the same idea in Cycle GAN. 😊

训练目标 As close as possible (**reconstruction**)

high dim



像GAN里的generator



old feature

因为是低维，也被叫做
bottleneck (瓶颈)

low dim (bottleneck)

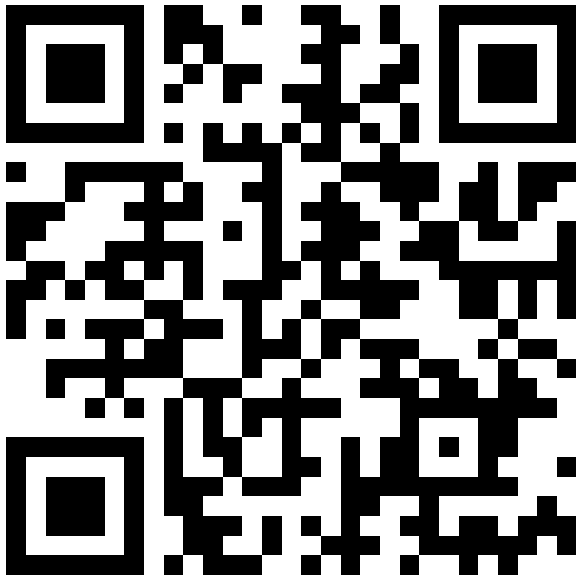
Dimension reduction

Embedding, Representation, Code (三个名字都可以)

New feature for downstream tasks

More Dimension Reduction

(not based on deep learning)



https://youtu.be/iwh5o_M4BNU

PCA P59



<https://youtu.be/GBUEjkpoxXc>

t-SNE P60

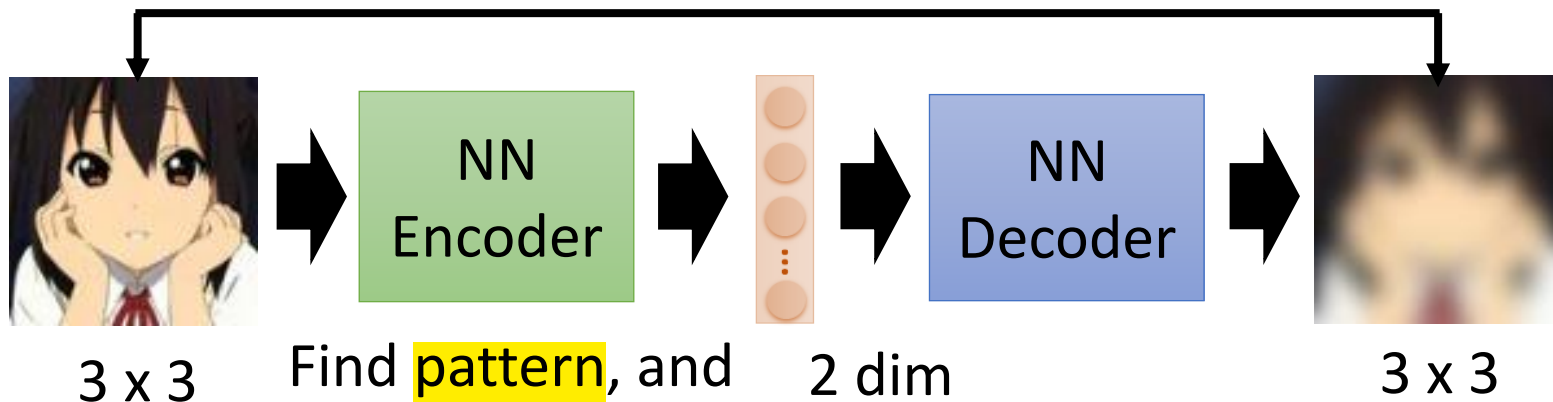
Why Auto-encoder?



《神鵰俠侶》

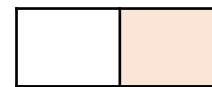
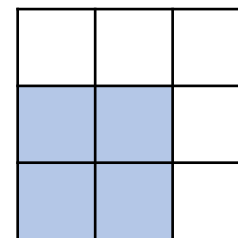
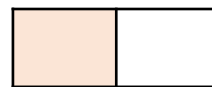
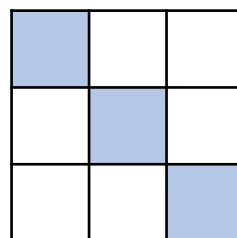
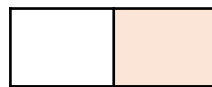
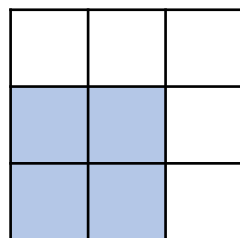
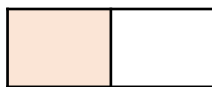
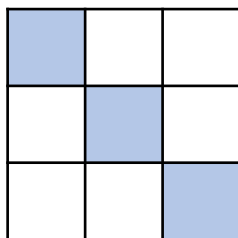
Why Auto-encoder?

As close as possible (reconstruction)



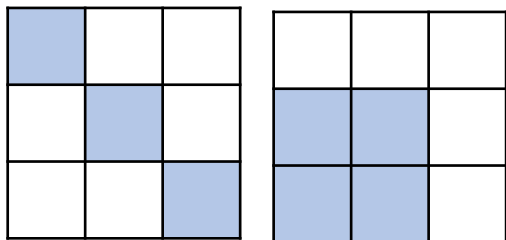
then compress

图片的变化有限 能找到pattern 可以用低纬度表示



Why Auto-encoder?

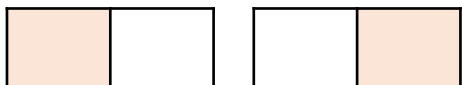
看起来很复杂



化繁为简

NN
Encoder

但是变化有限

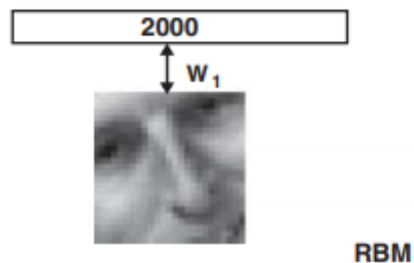
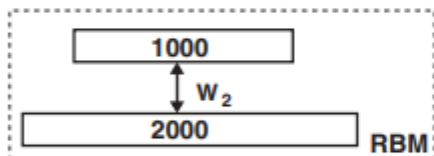
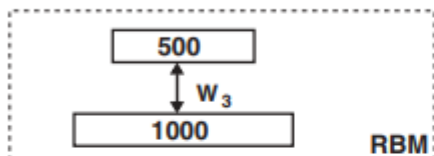
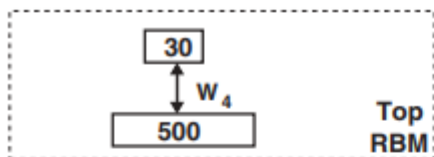


《神鵰俠侶》

Auto-encoder is not a new idea

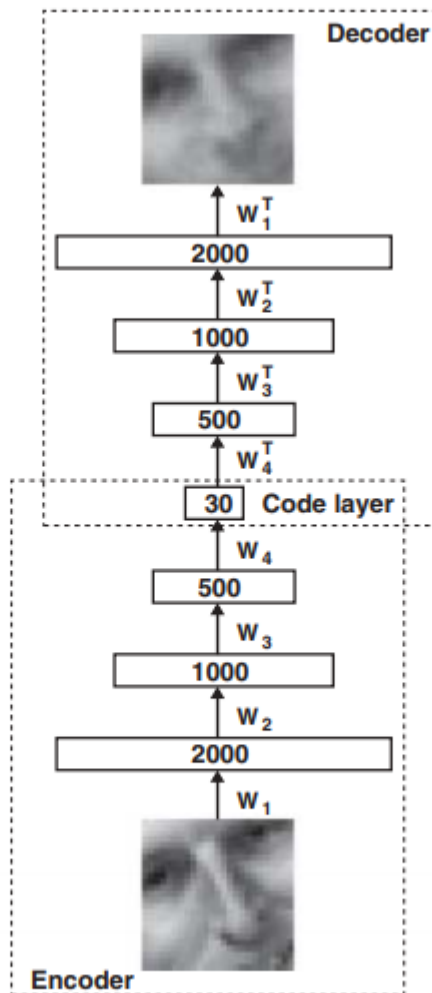
Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." *Science* 313.5786 (2006): 504-507

分开训练



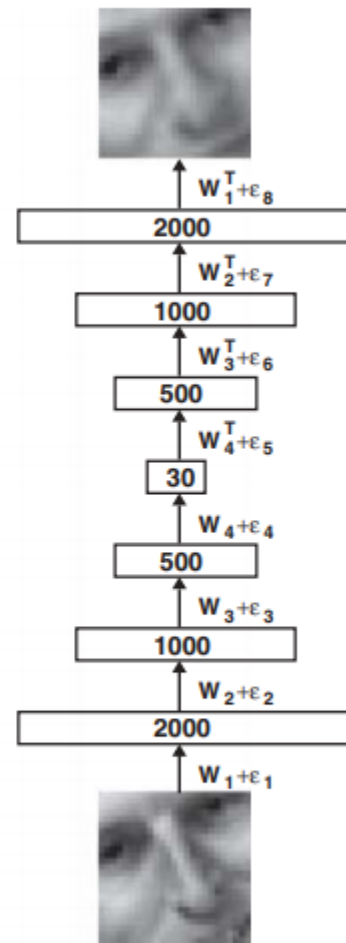
Pretraining

pre-train的pre-train



Unrolling

当时觉得要对称

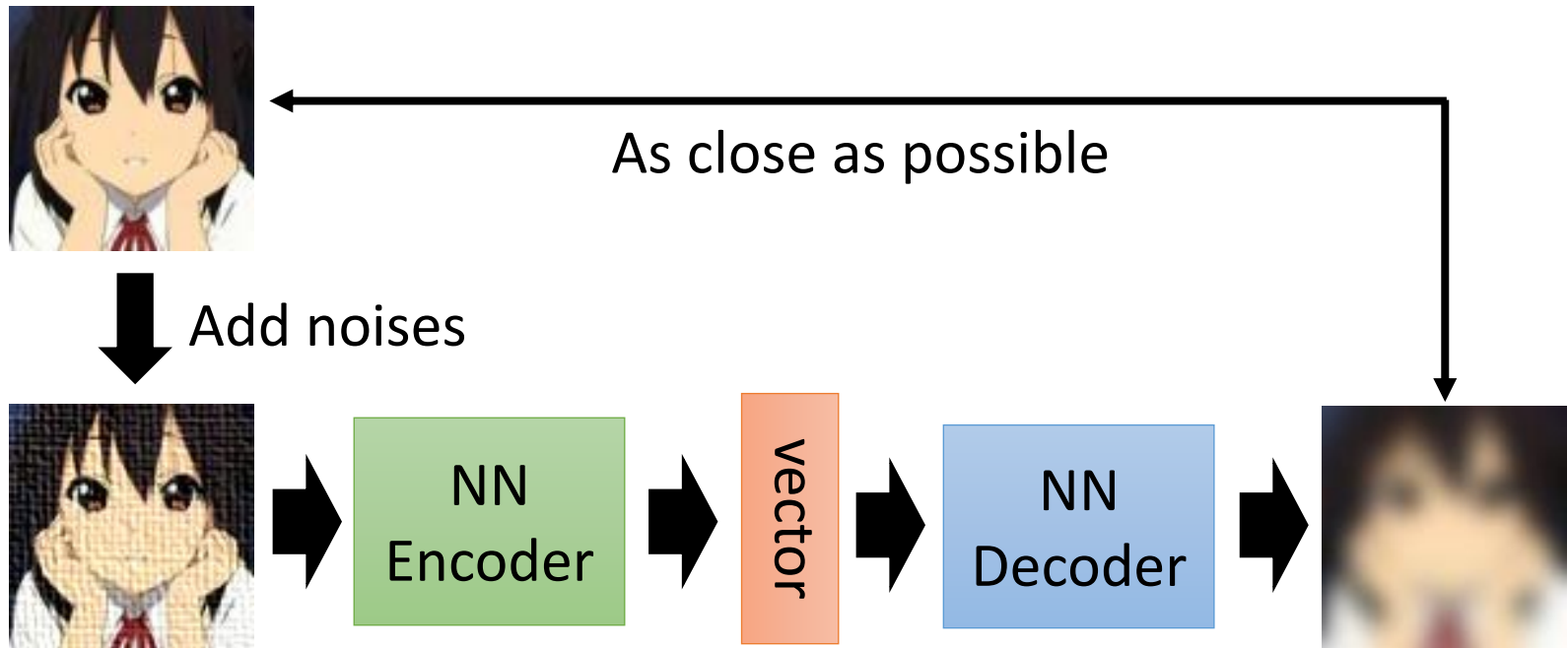


Fine-tuning

微调pretrain的模型

RBM : 受限玻尔茨曼机

De-noising Auto-encoder

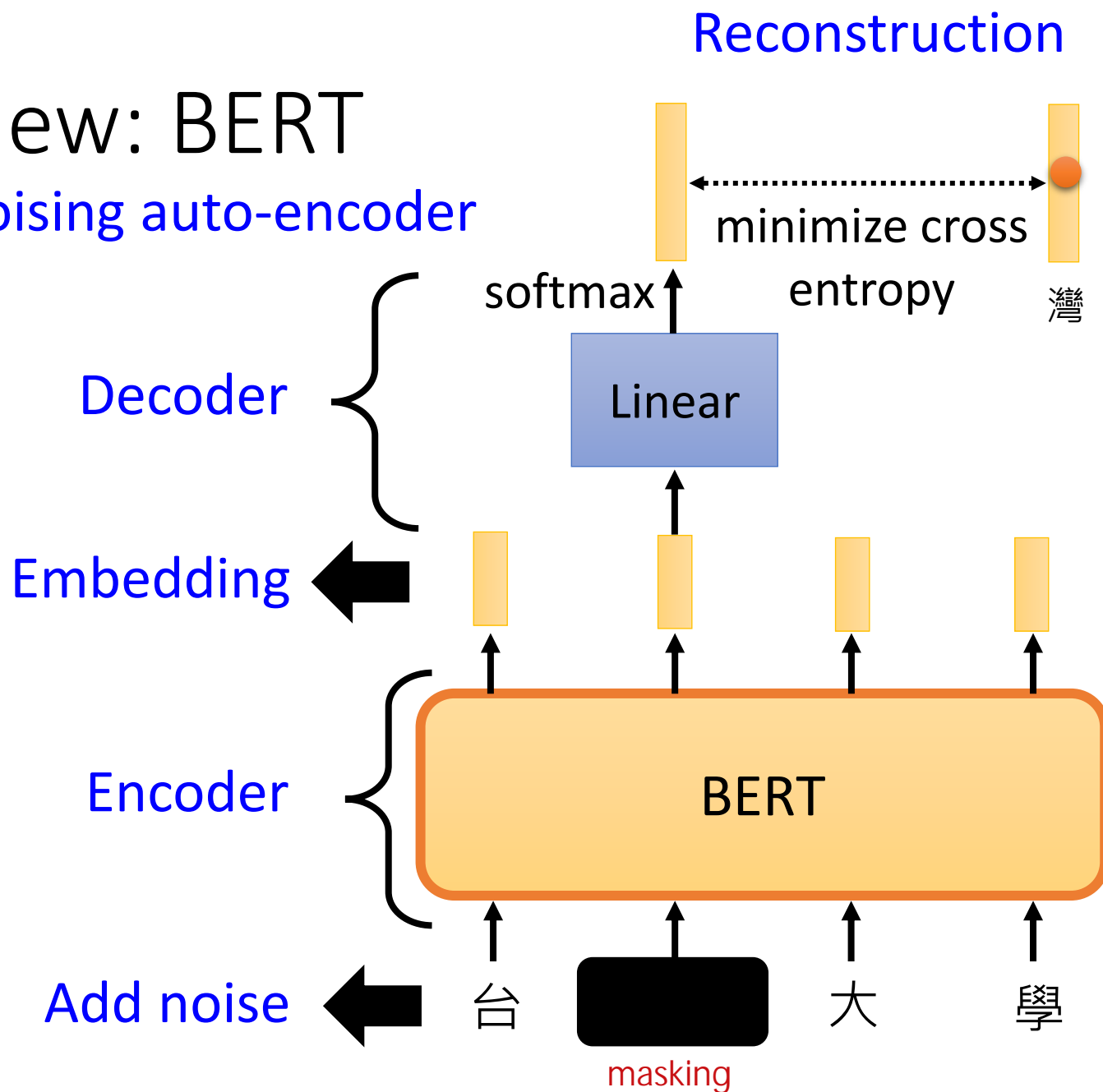


The idea sounds familiar? 😊 Bert

Vincent, Pascal, et al. "Extracting and composing robust features with denoising autoencoders." *ICML*, 2008.

Review: BERT

A de-noising auto-encoder



Outline

2020年5月1日
自编码器 (2)

Basic Idea of Auto-encoder

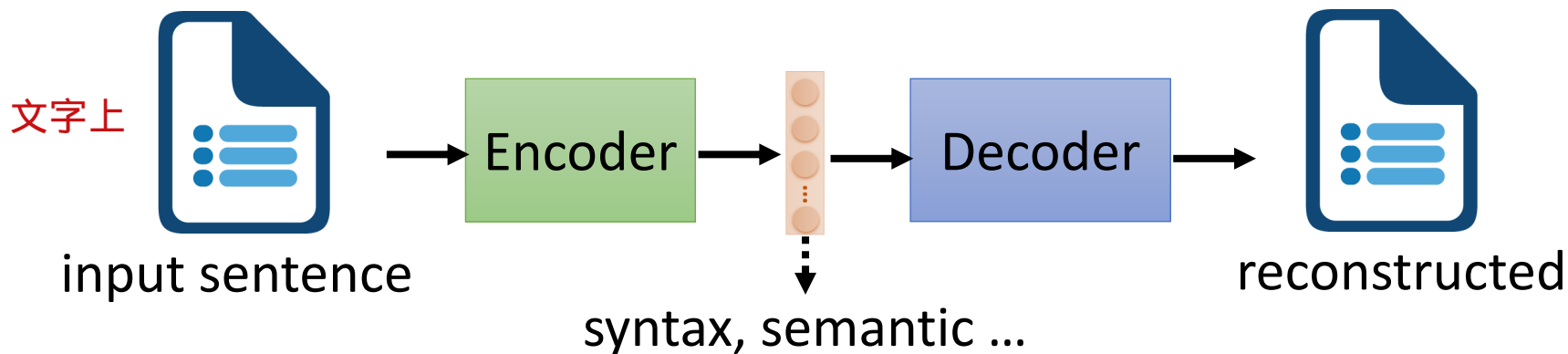
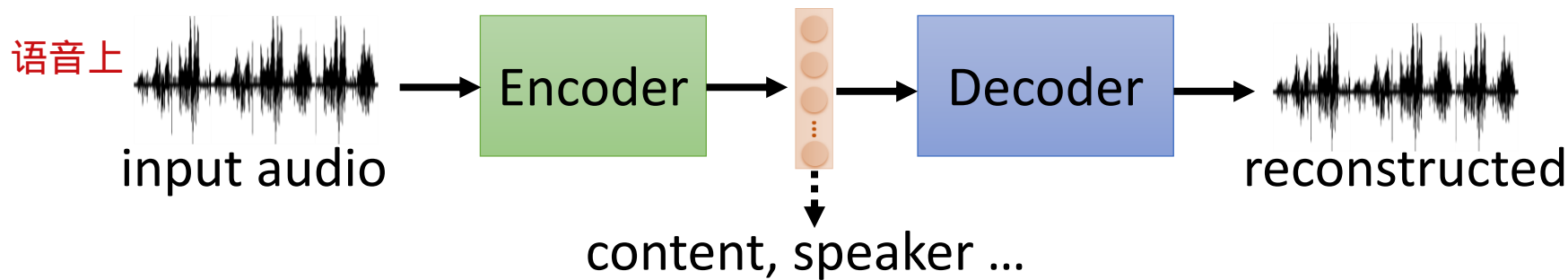
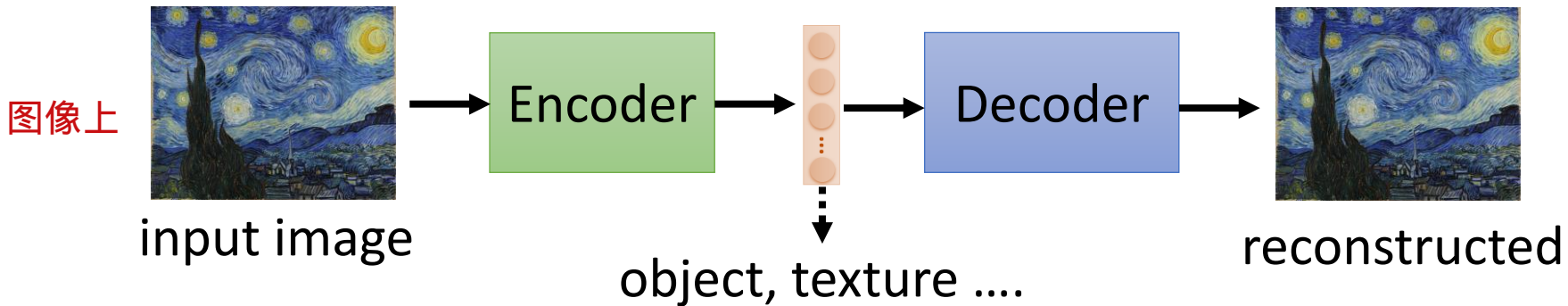
Feature Disentanglement

Discrete Latent Representation

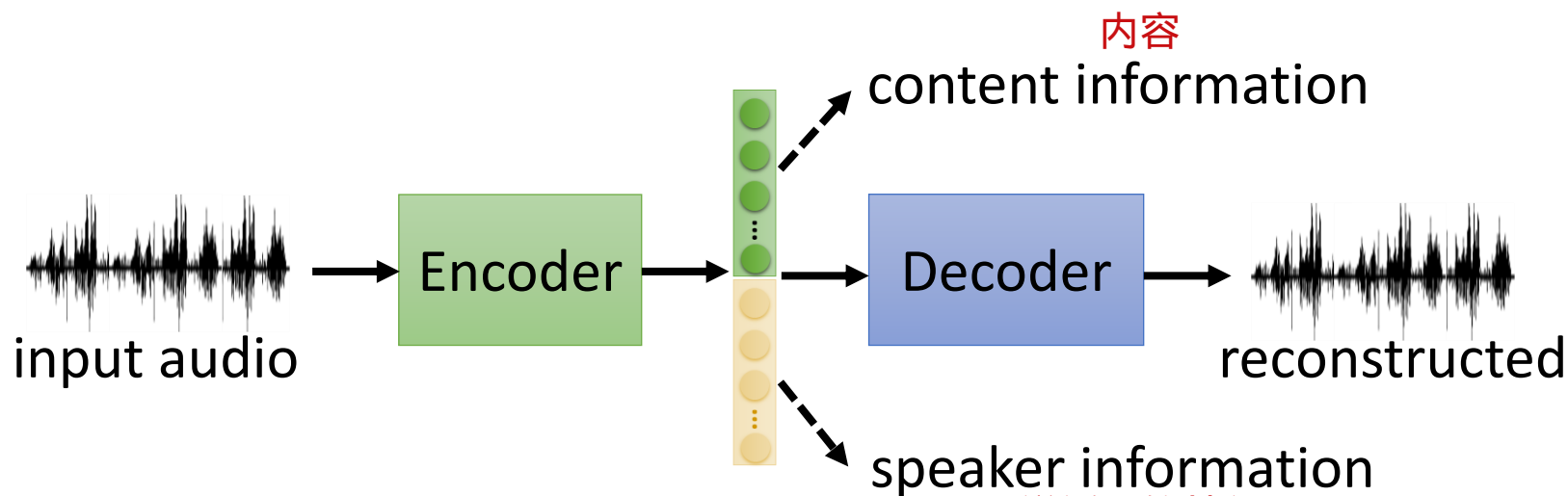
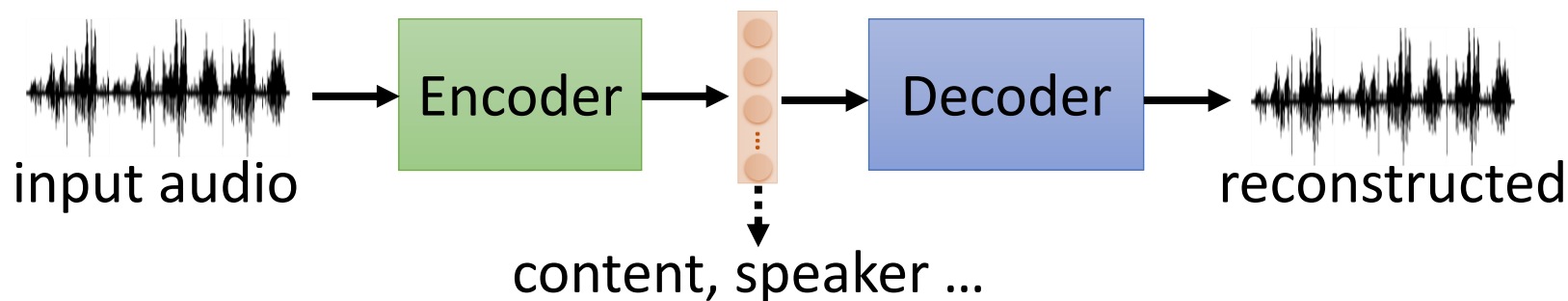
More Applications

什么是Disentanglement

Representation includes information of different aspects



Feature Disentangle



<https://arxiv.org/abs/1904.05742>

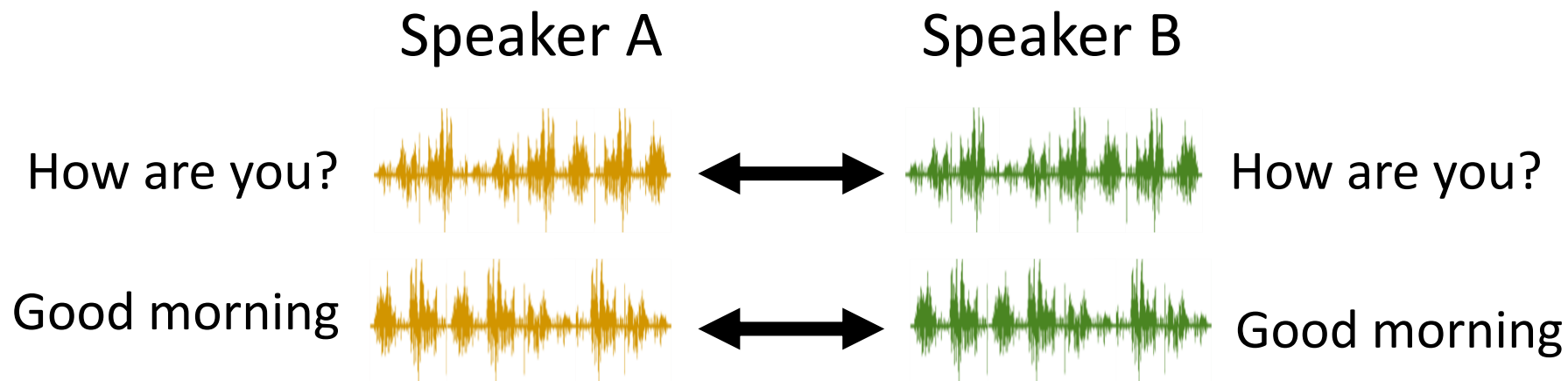
<https://arxiv.org/abs/1804.02812>

<https://arxiv.org/abs/1905.05879>



Application: Voice Conversion

In the past

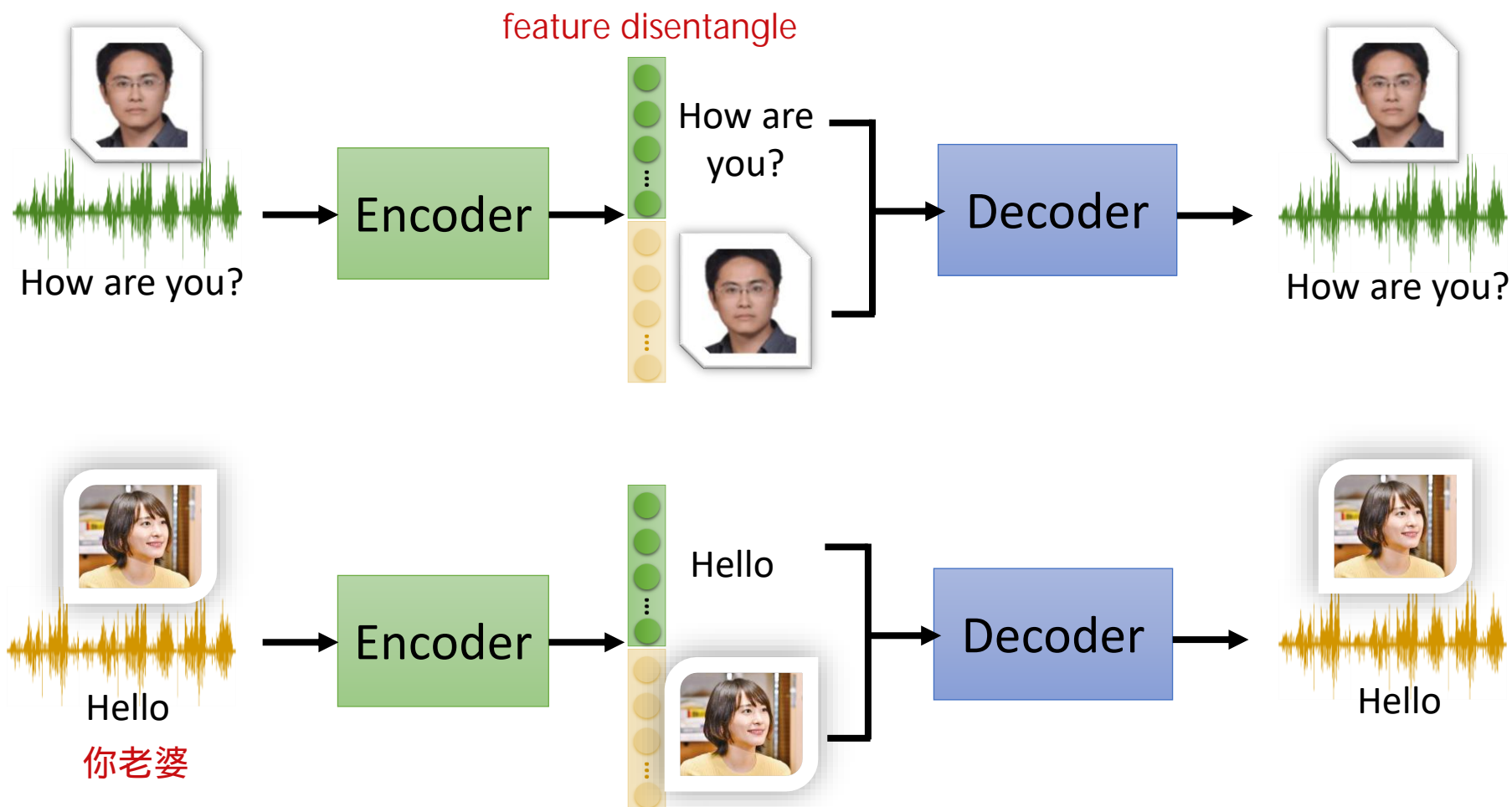


Today

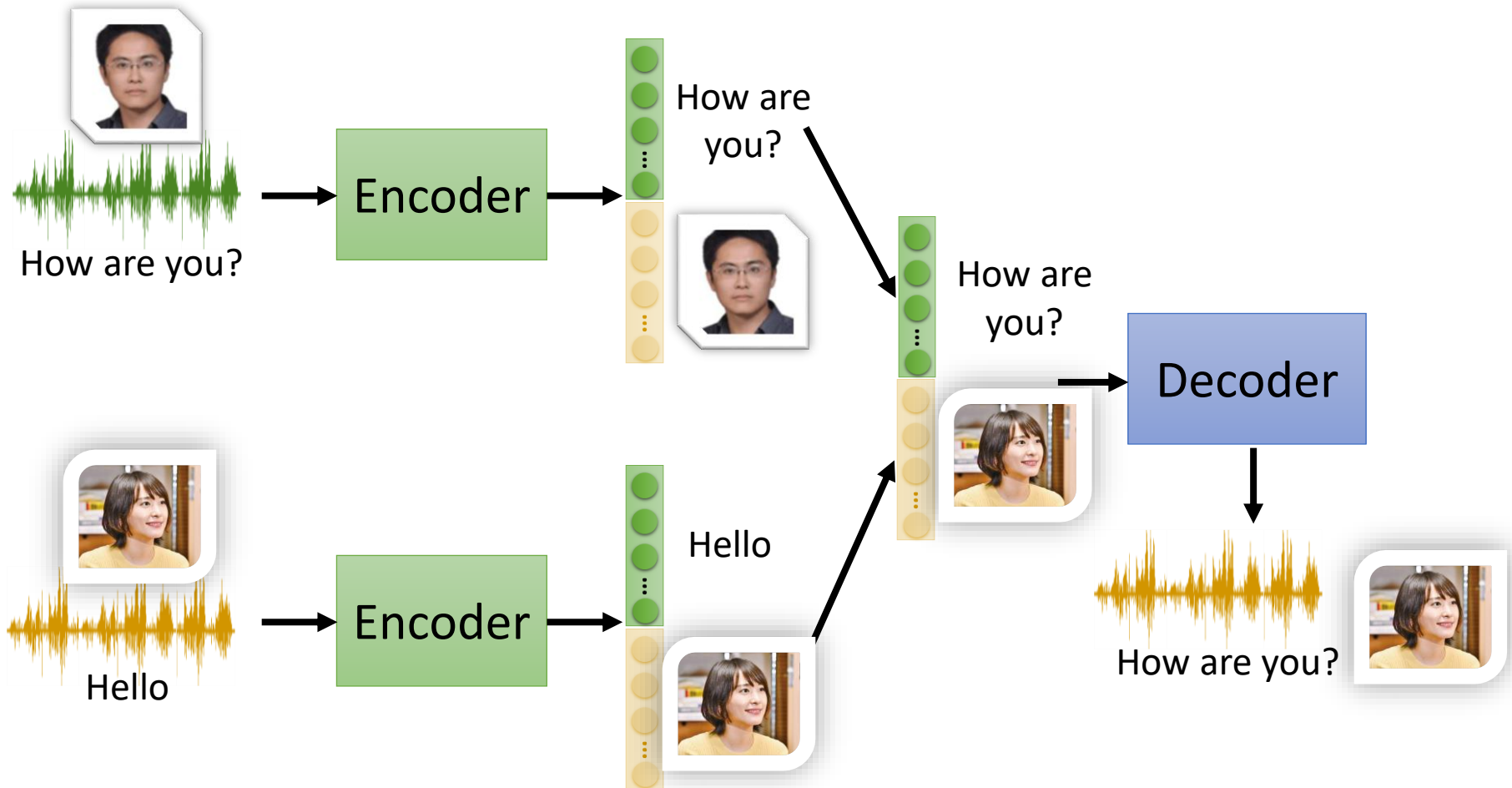


Speakers A and B are talking about completely different things.

Application: Voice Conversion



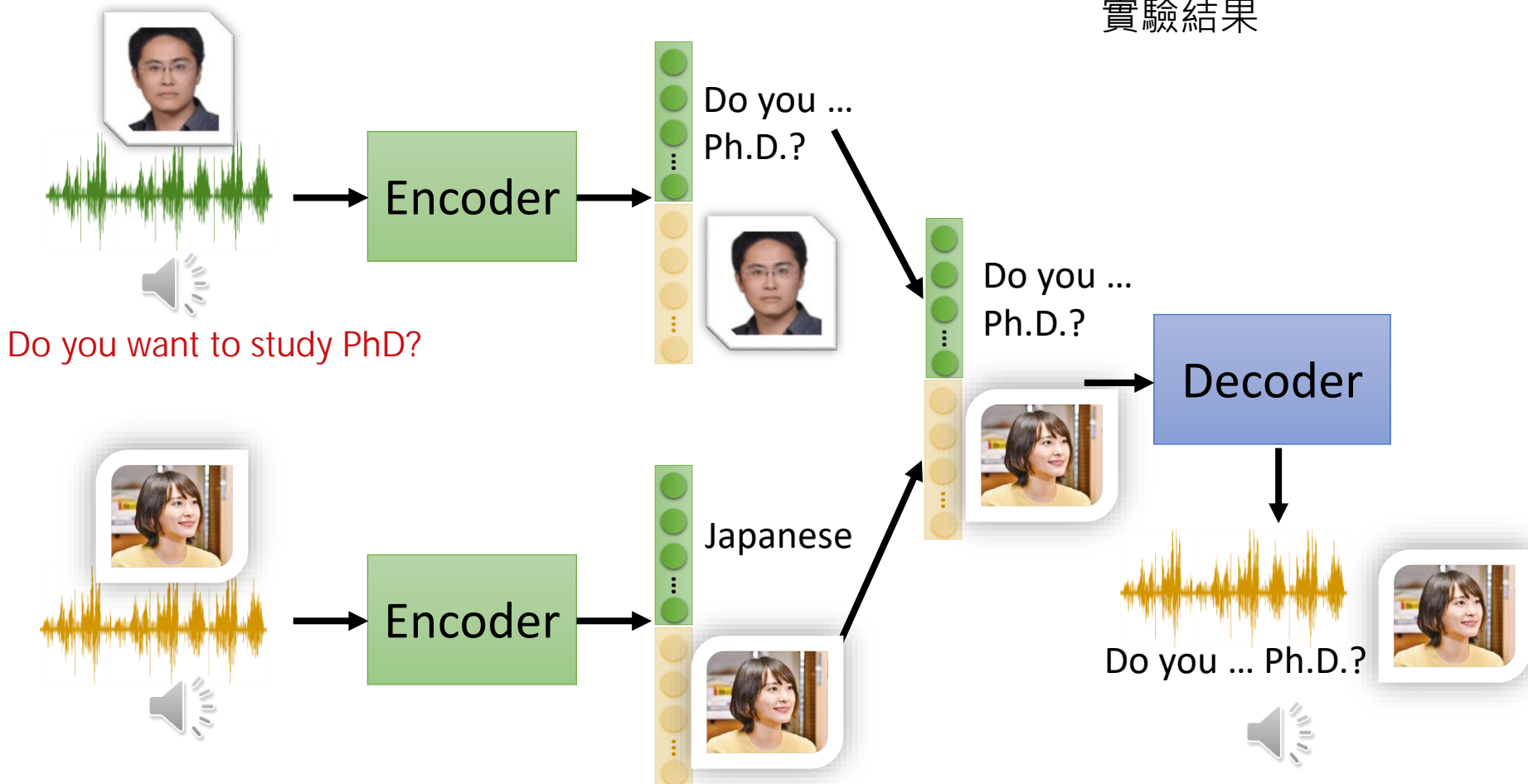
Application: Voice Conversion



实际操作

Application: Voice Conversion

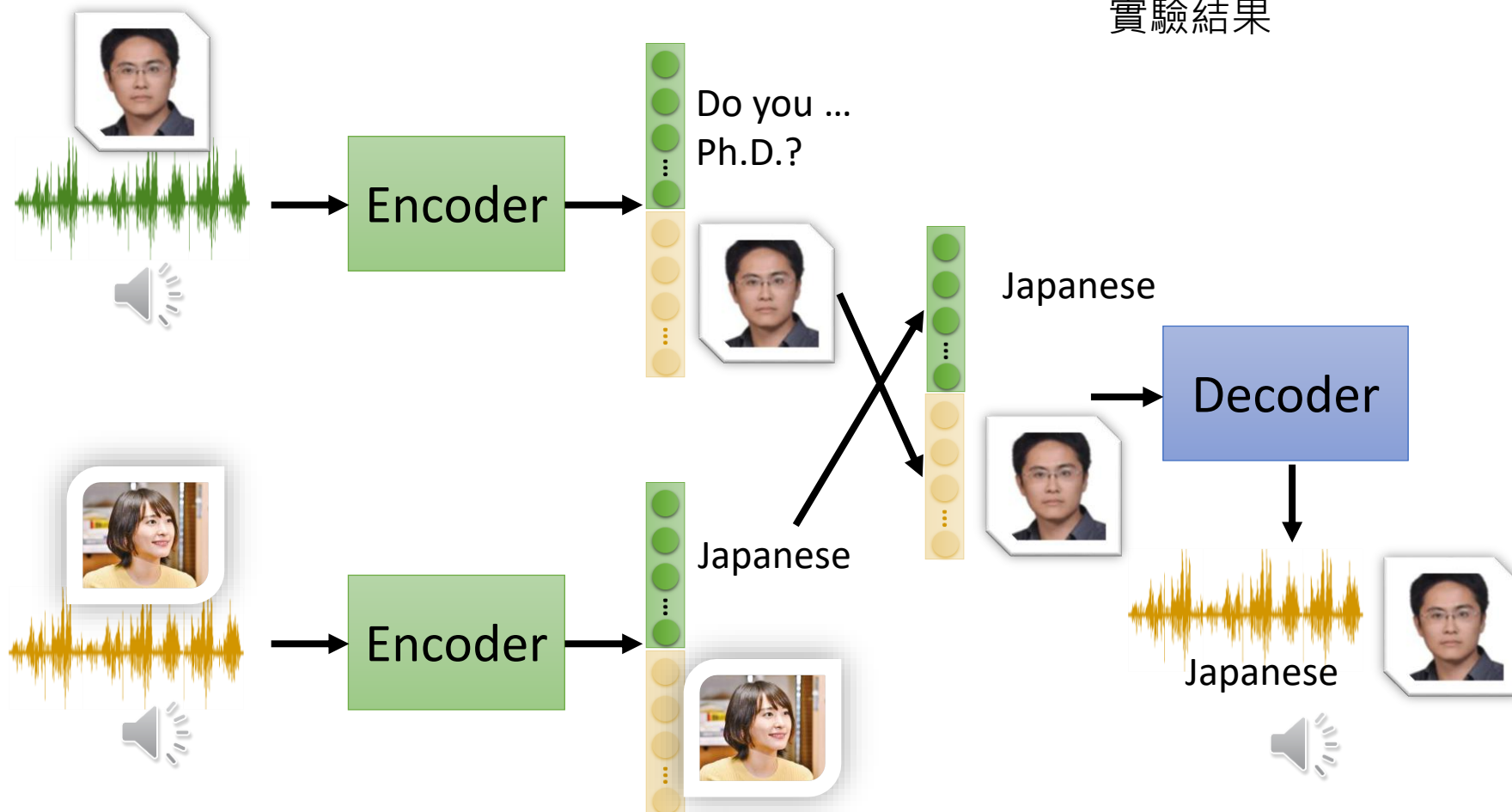
感謝解正平同學提供
實驗結果



李宏毅：你老婆在叫你读博班！

Application: Voice Conversion

感謝解正平同學提供
實驗結果



Outline

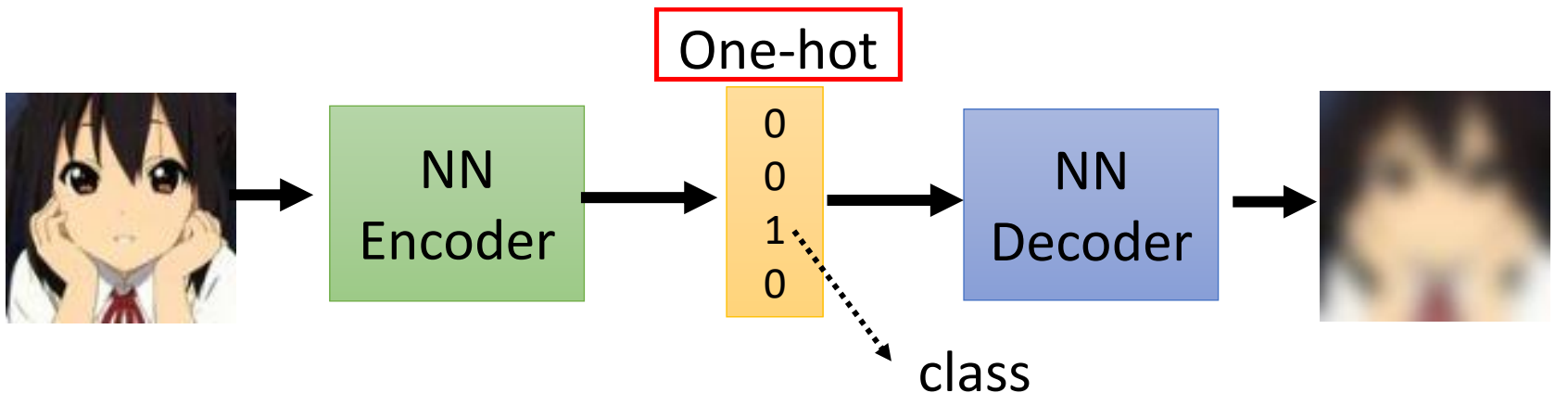
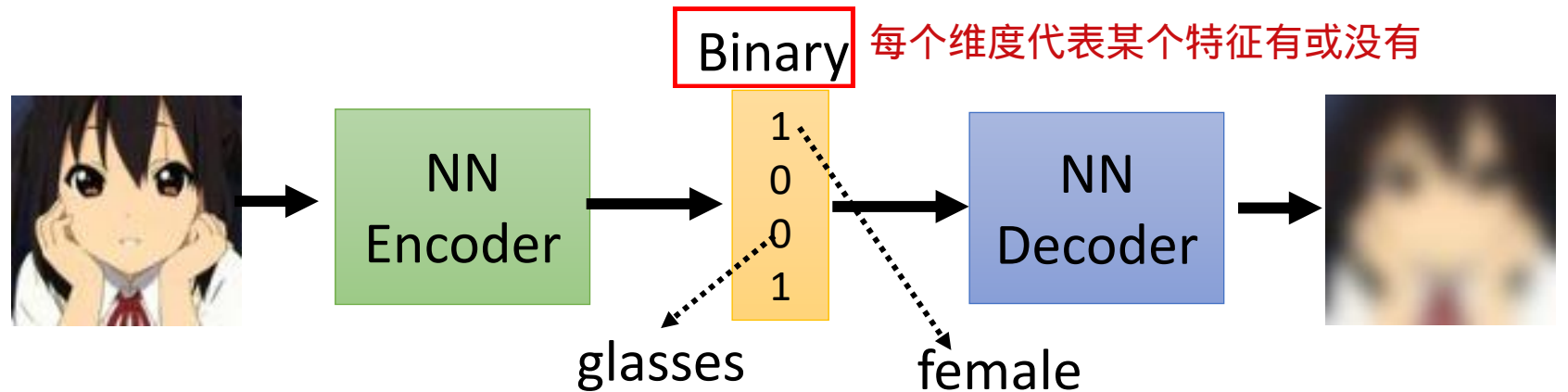
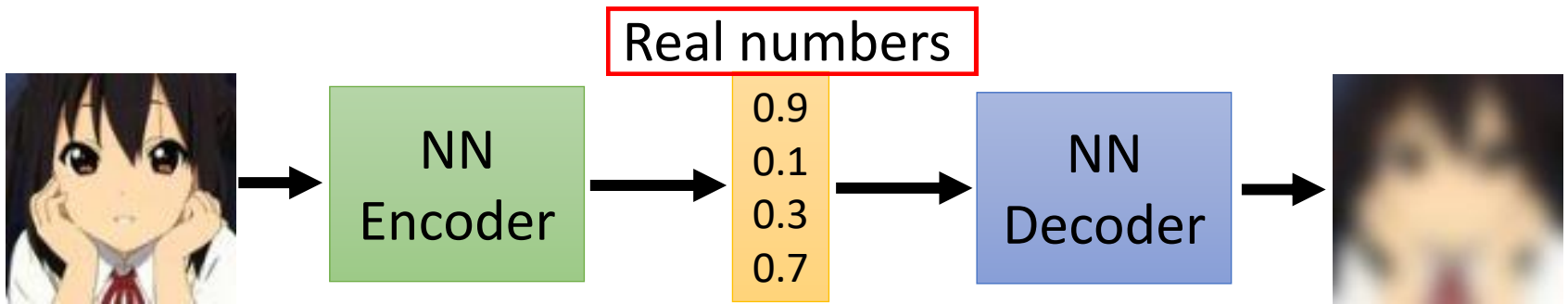
Basic Idea of Auto-encoder

Feature Disentanglement

Discrete Latent Representation

More Applications

Discrete Representation

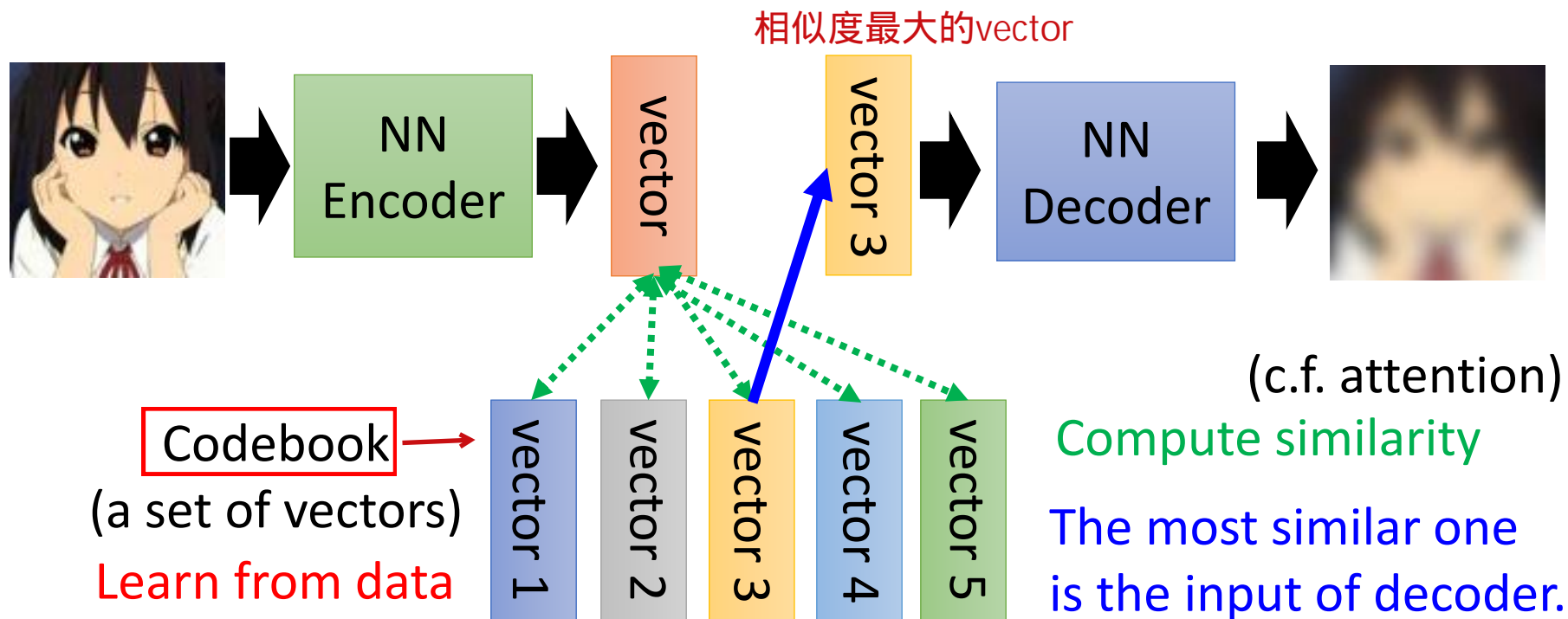


Discrete Representation

最知名的Discrete Representation

<https://arxiv.org/abs/1711.00937>

- Vector Quantized Variational Auto-encoder (**VQVAE**)



For speech, the codebook represents phonetic information

<https://arxiv.org/pdf/1901.08810.pdf>

Text as Representation

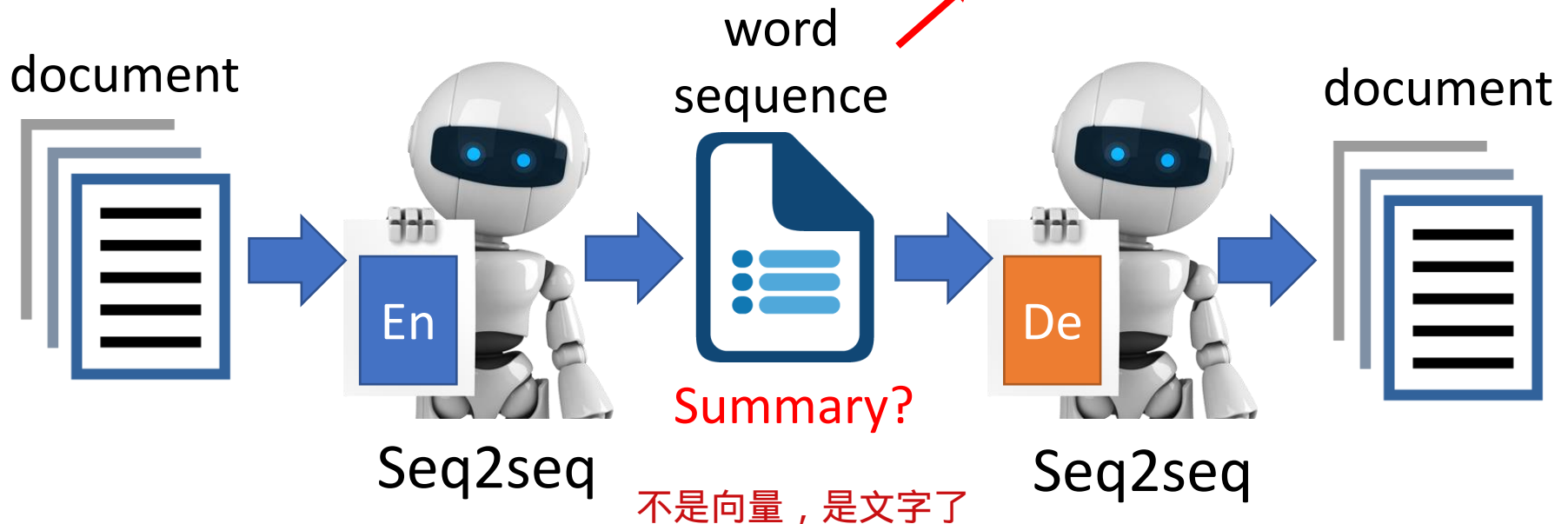
Only need a lot
of documents to
train the model



长seq 短seq 长seq
seq2seq2seq
auto-encoder

Unsupervised Summarization

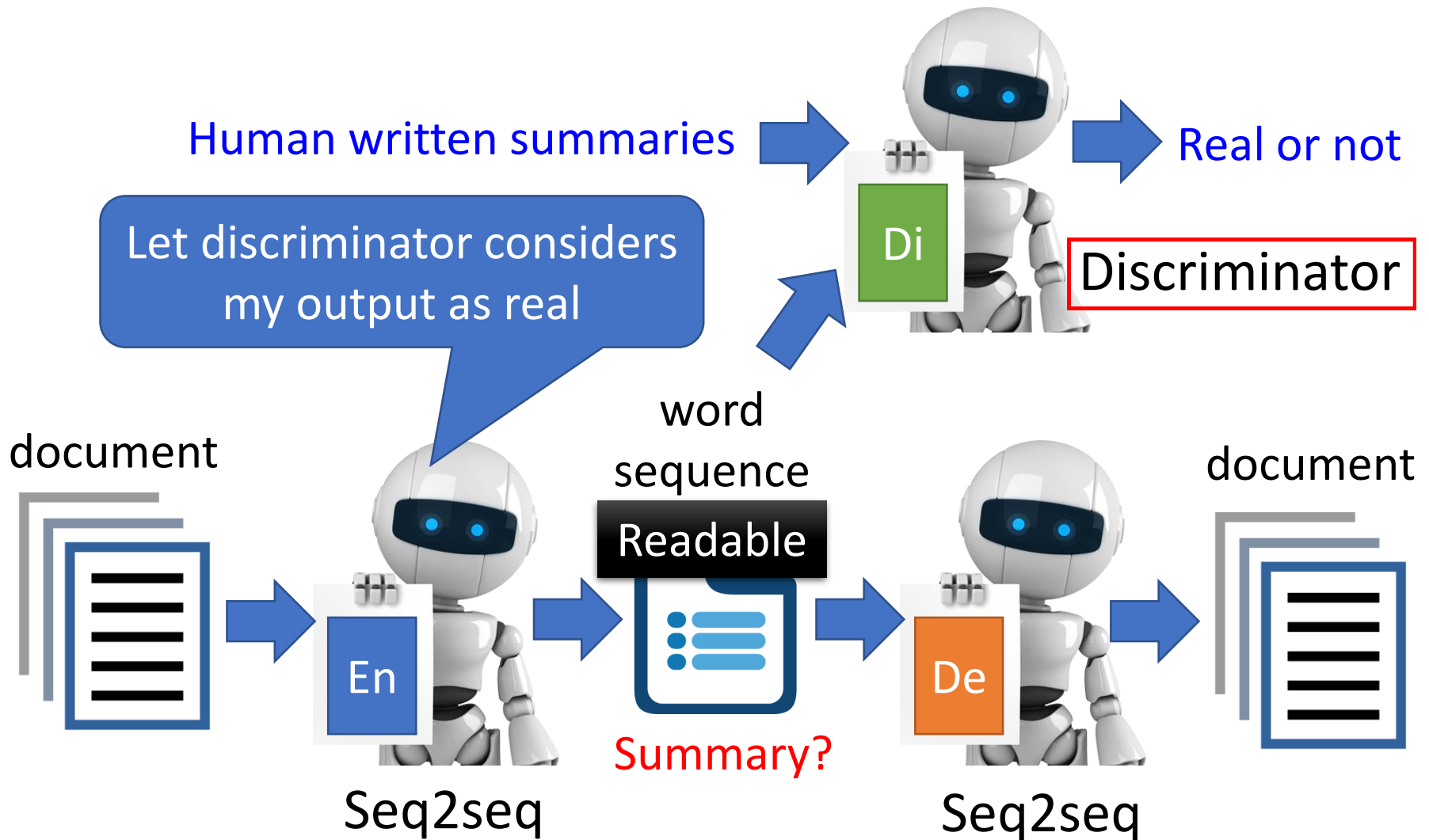
人看不懂，decoder看得懂
not readable ...



This is cycle GAN 😊

Text as Representation

RL硬train



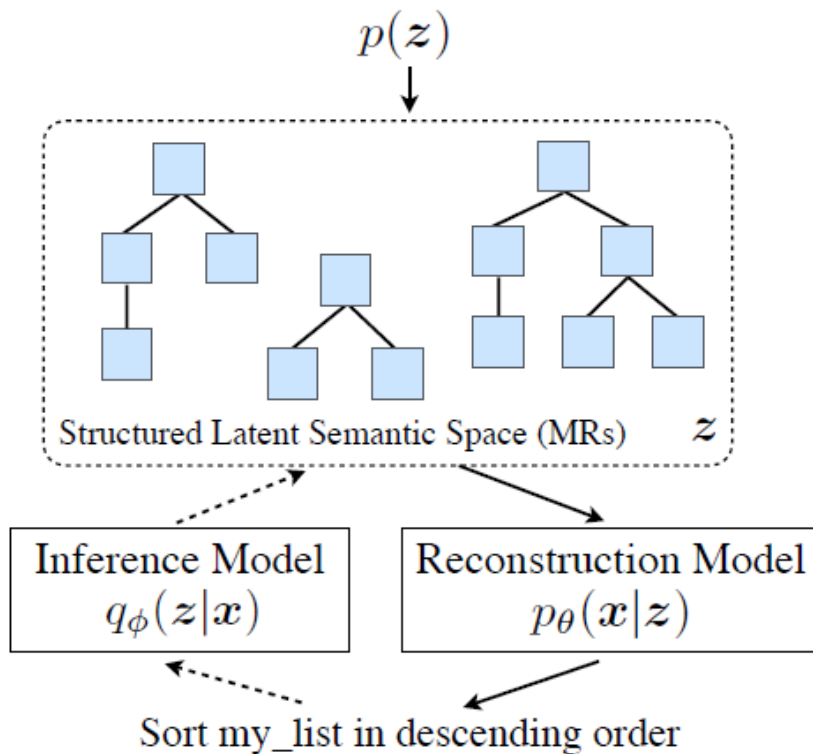
Text as Representation

- **Document**: 澳大利亞今天與13個國家簽署了反興奮劑雙邊協議,旨在加強體育競賽之外的藥品檢查並共享研究成果
- **Summary**:
 - **Human**: 澳大利亞與13國簽署反興奮劑協議
 - **Unsupervised**: 澳大利亞加強體育競賽之外的藥品檢查
- **Document**: 中華民國奧林匹克委員會今天接到一九九二年冬季奧運會邀請函,由於主席張豐緒目前正在中南美洲進行友好訪問,因此尚未決定是否派隊赴賽
- **Summary**:
 - **Human**: 一九九二年冬季奧運會函邀我參加
 - **Unsupervised**: 奧委會接獲冬季奧運會邀請函

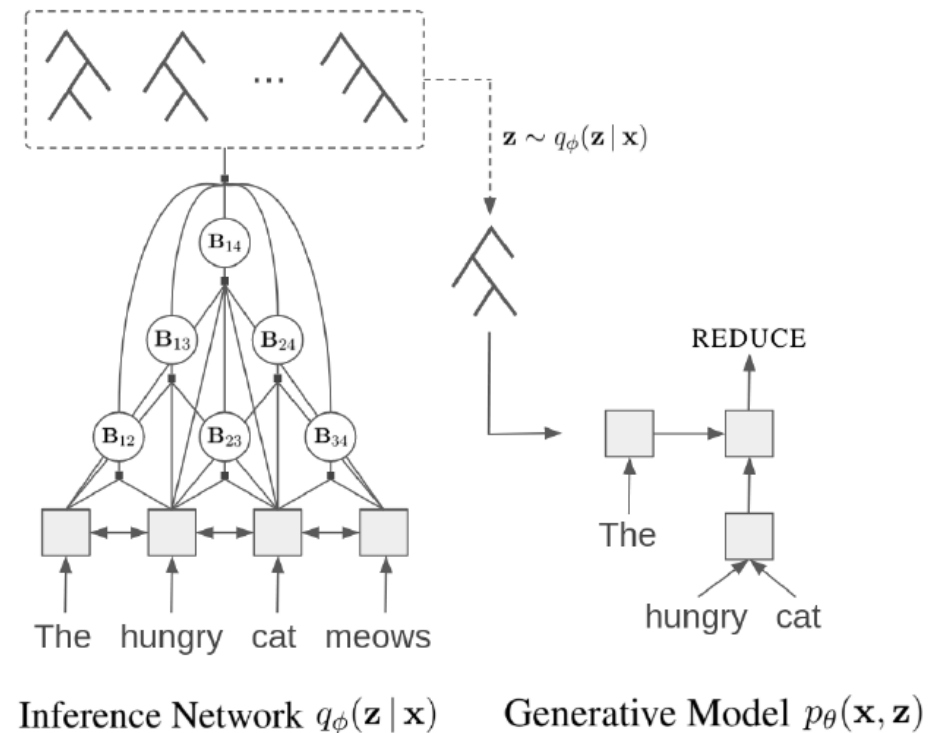
Text as Representation

- **Document**: 據此間媒體27日報道, 印度尼西亞蘇門答臘島的兩個省近日來連降暴雨, 洪水泛濫導致塌方, 到26日為止至少已有60人喪生, 100多人失蹤
- **Summary**:
 - **Human**: 印尼水災造成60人死亡
 - **Unsupervised**: 印尼門洪水泛濫導致塌雨 错误例子
- **Document**: 安徽省合肥市最近為領導幹部下基層做了新規定: 一律輕車簡從, 不準搞迎來送往、不準搞層層陪同
- **Summary**:
 - **Human**: 合肥規定領導幹部下基層活動從簡
 - **Unsupervised**: 合肥領導幹部下基層做搞迎來送往規定: 一律簡
莫名其妙的例子

Tree as Embedding



<https://arxiv.org/abs/1806.07832>



<https://arxiv.org/abs/1904.03746>

Outline

Basic Idea of Auto-encoder

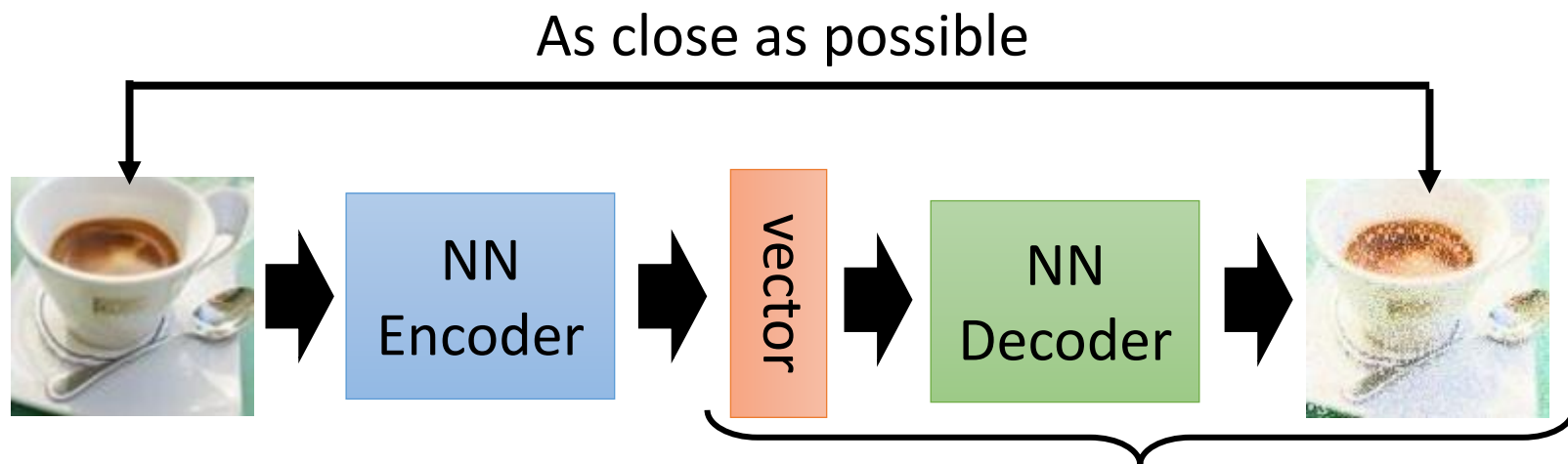
Feature Disentanglement

Discrete Latent Representation

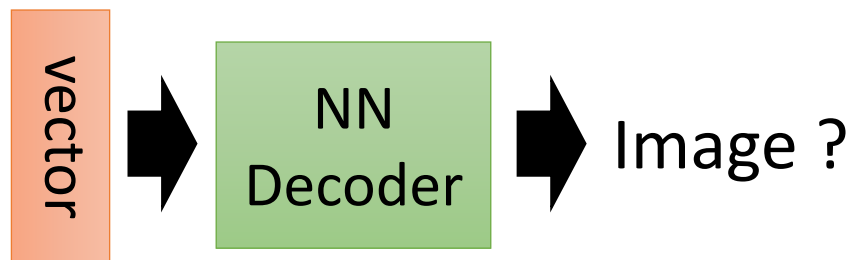
More Applications

Generator

generative模型
1. GAN
2. VAE
3.



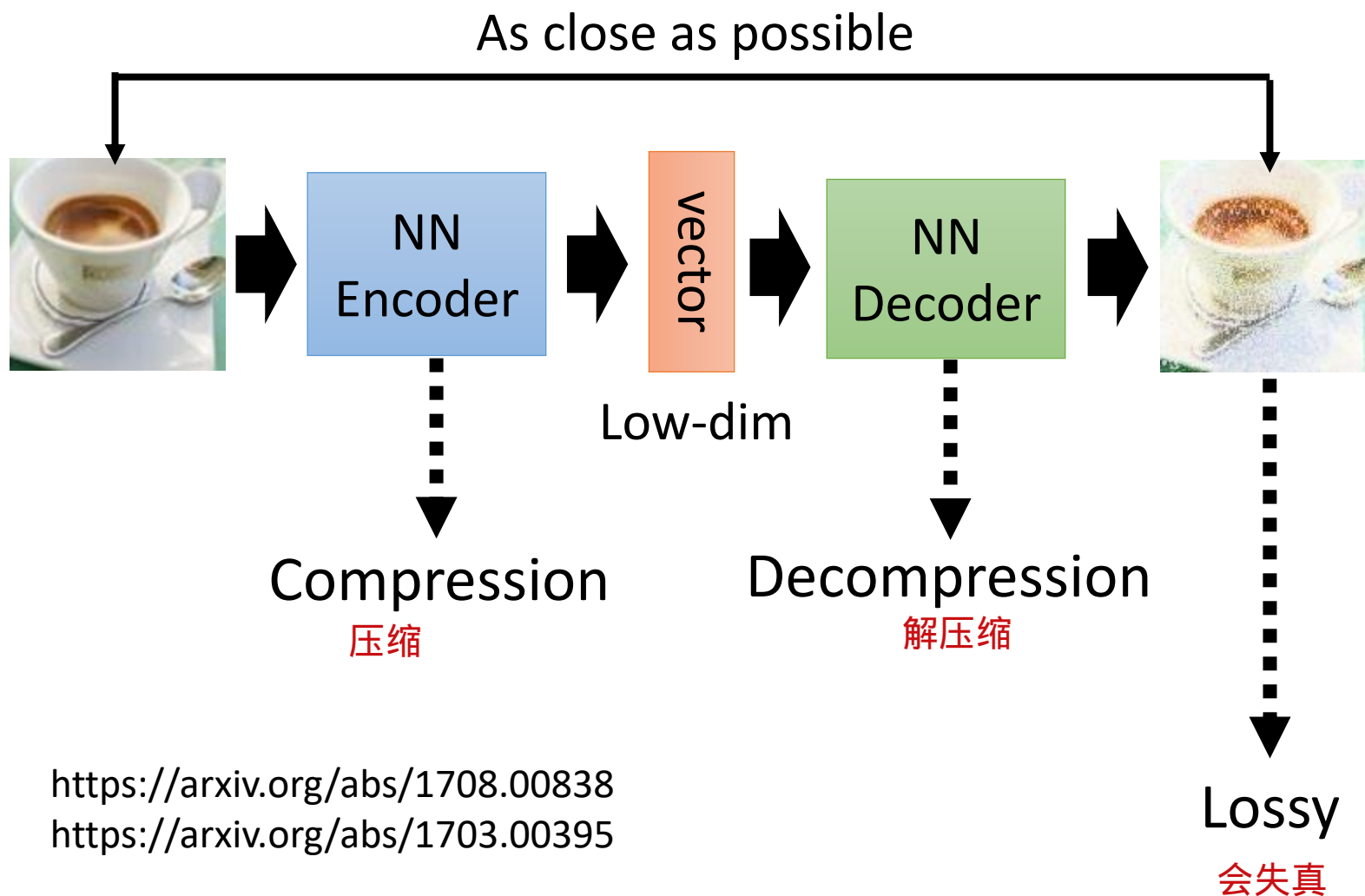
Randomly generate a vector from a distribution



With some modification, we have **variational auto-encoder (VAE)**.

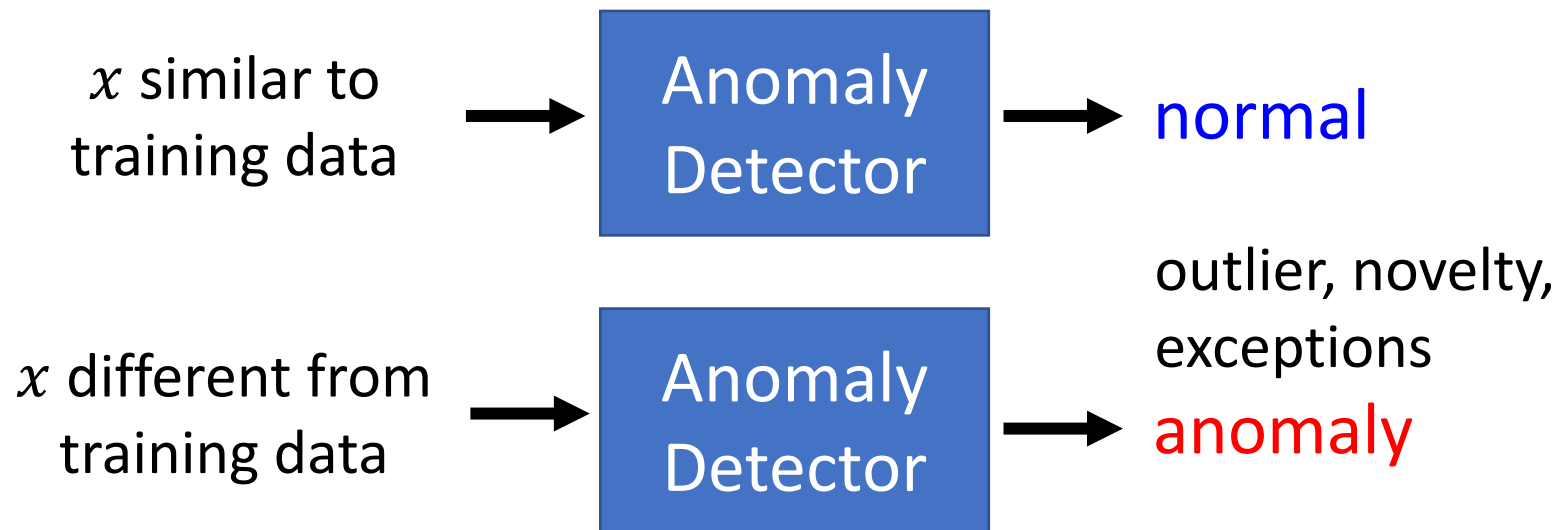
auto-encoder的decoder当作generator用

Compression



Anomaly Detection 异常检测

- Given a set of training data $\{x^1, x^2, \dots, x^N\}$
- Detecting input x is *similar* to training data or not.



Anomaly Detection

Training Data:



anomaly

Training Data:



anomaly

Training Data:



亚古兽

anomaly

宝可梦

难点：在于收资料，通常容易收到正常资料。
假设有大量的正常资料，一点点的异常资料，
所以不是一般的分类问题。

Anomaly Detection

Binary Classification?

We only have one class.

Training auto-encoder

- Fraud Detection

- Training data: credit card transactions, x : fraud or not
- Ref: <https://www.kaggle.com/ntnu-testimon/paysim1/home>
- Ref: <https://www.kaggle.com/mlg-ulb/creditcardfraud/home>

- Network Intrusion Detection

- Training data: connection, x : attack or not
- Ref: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>

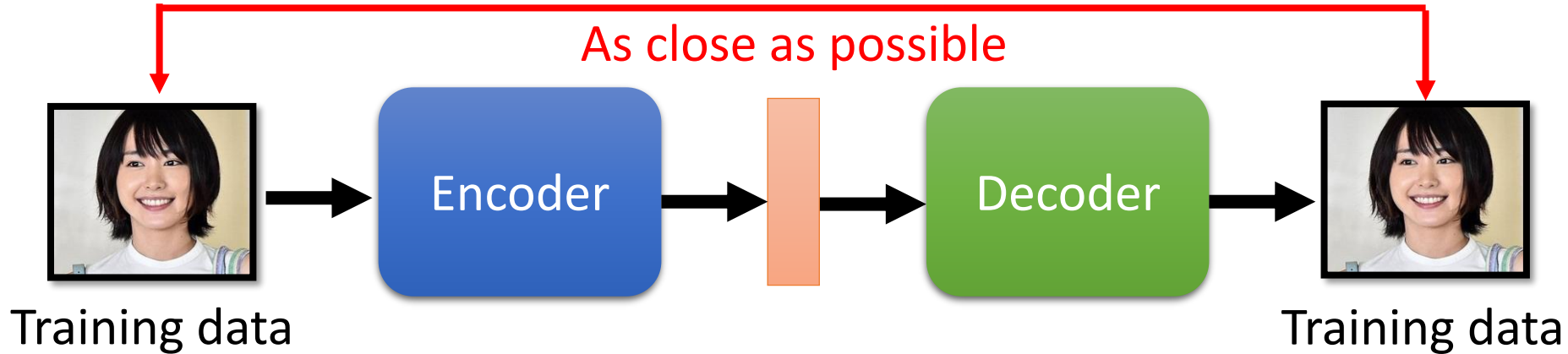
- Cancer Detection

- Training data: normal cells, x : cancer or not?
- Ref: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data/home>

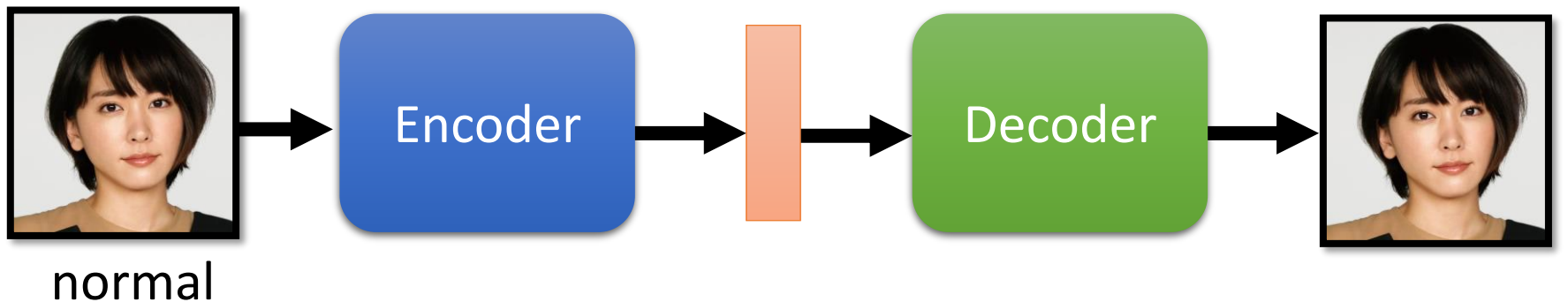
Approach: Auto-encoder

Training

Using **real human faces** to learn an autoencoder



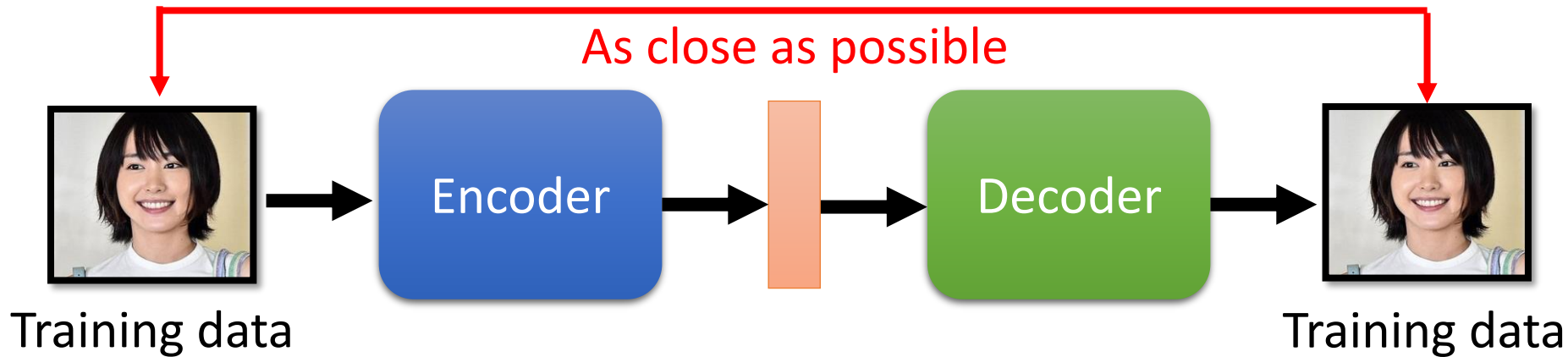
Testing



Approach: Auto-encoder

Training

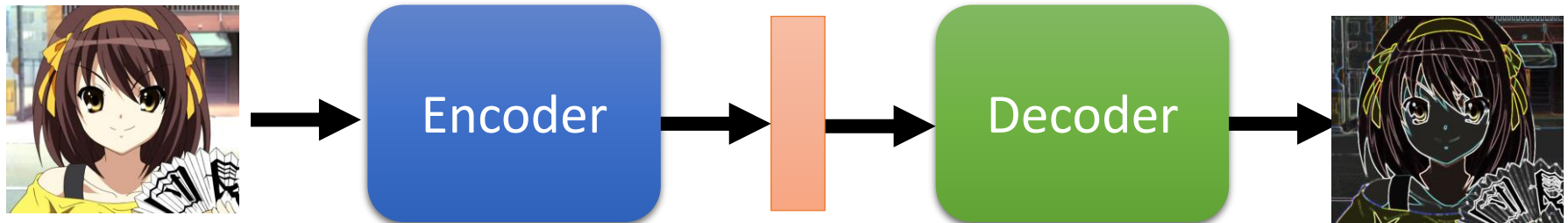
Using **real human faces** to learn an autoencoder



Testing

Large **reconstruction loss** → anomaly

cannot be reconstructed



anomaly

凉宫春日，二次元人物

More about Anomaly Detection

- Part 1: <https://youtu.be/gDp2LXGnVLQ>
- Part 2: <https://youtu.be/cYrNjLxkoXs>
- Part 3: <https://youtu.be/ueDIm2FkCnw>
- Part 4: <https://youtu.be/XwkHOUPbc0Q>
- Part 5: <https://youtu.be/Fh1xFBktRLQ>
- Part 6: <https://youtu.be/LmFWzmn2rFY>
- Part 7: <https://youtu.be/6W8FqUGYyDo>

Concluding Remarks

Basic Idea of Auto-encoder

Feature Disentanglement

Discrete Latent Representation

More Applications