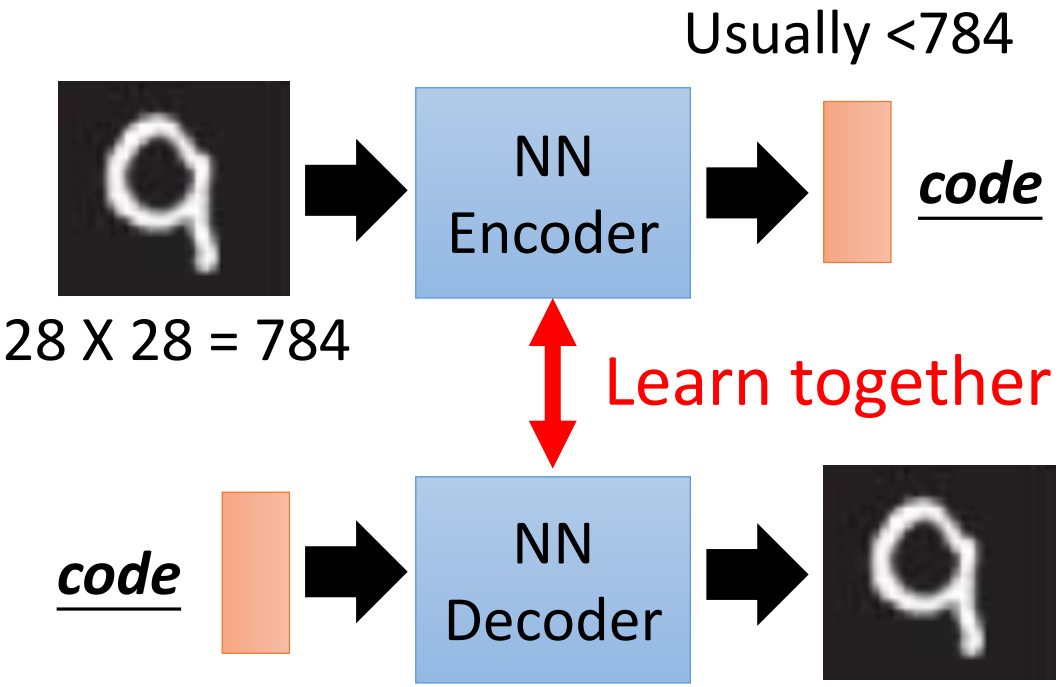
Unsupervised Learning:
Deep Auto-encoder

Unsupervised Learning

“We expect unsupervised learning to become far more important in the longer term. Human and animal learning is largely unsupervised: we discover the structure of the world by observing it, not by being told the name of every object.”
– LeCun, Bengio, Hinton, Nature 2015

As I've said in previous statements: most of human and animal learning is unsupervised learning. If intelligence was a cake, unsupervised learning would be the cake, supervised learning would be the icing on the cake, and reinforcement learning would be the cherry on the cake. We know how to make the icing and the cherry, but we don't know how to make the cake.
- Yann LeCun, March 14, 2016 (Facebook)

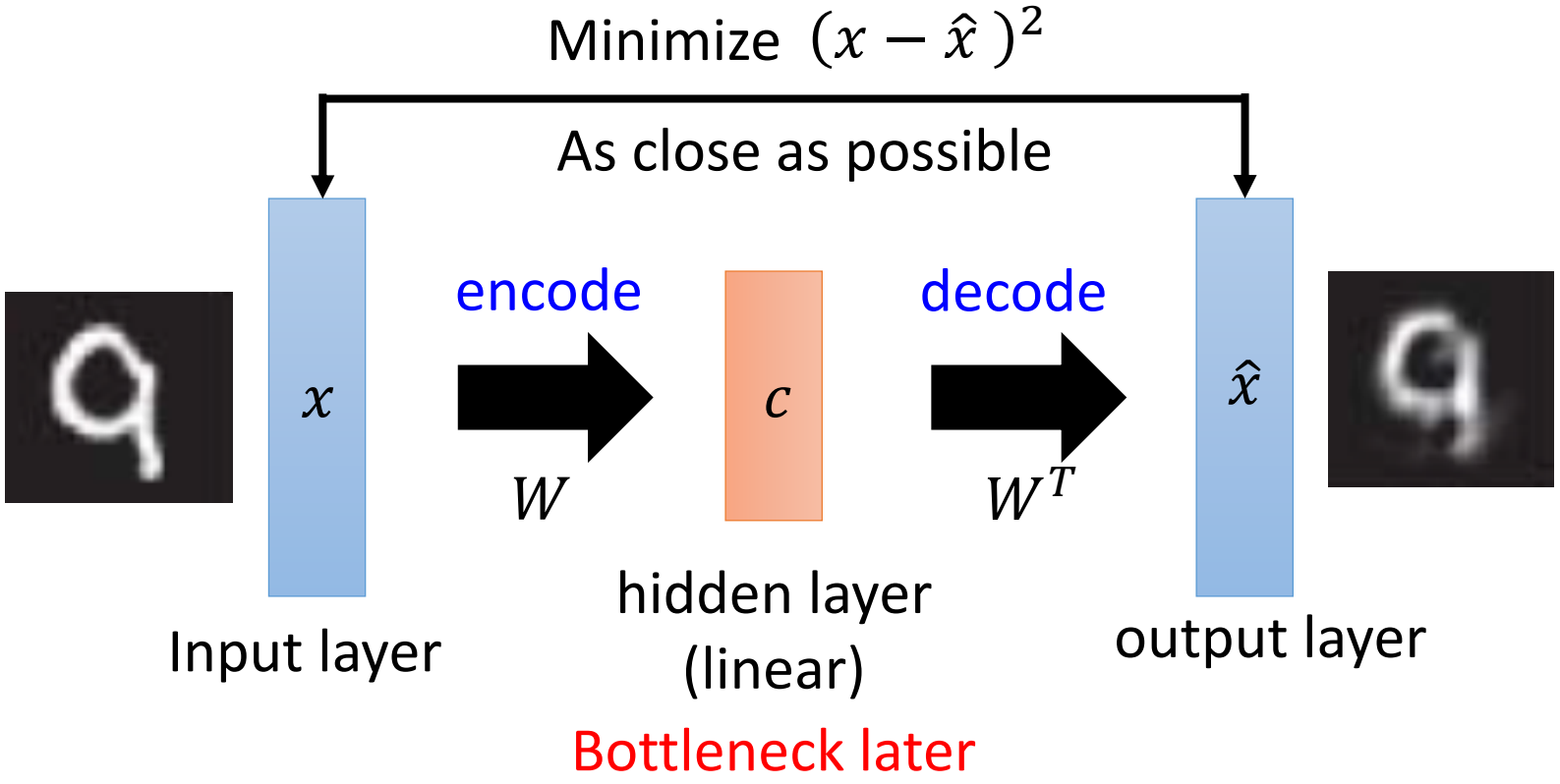
Auto-encoder



Compact representation of the input object

Can reconstruct the original object

Recap: PCA

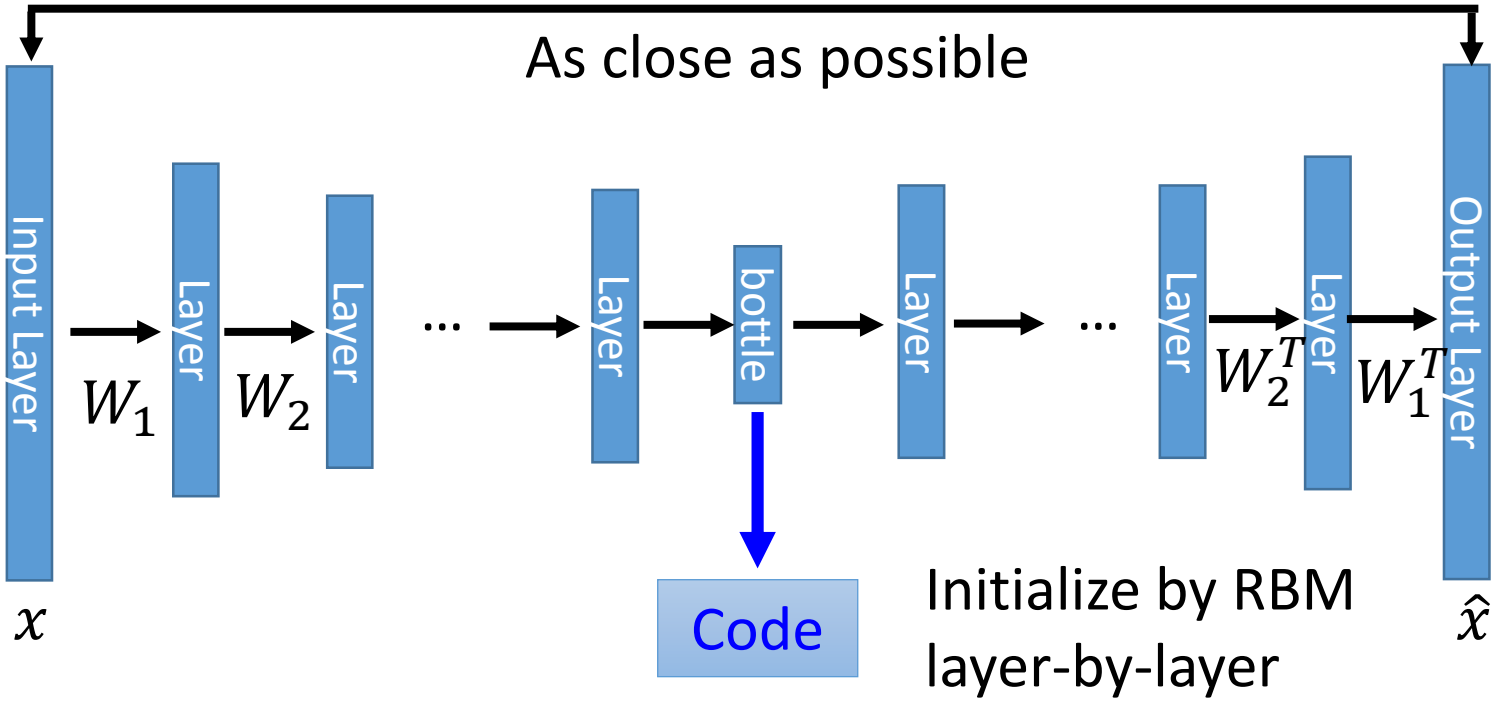


Output of the hidden layer is the code

Deep Auto-encoder

Symmetric is not necessary.

- Of course, the auto-encoder can be deep



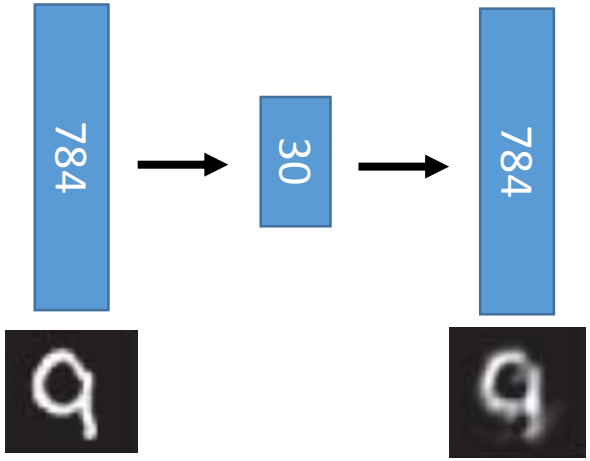
Reference: Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." *Science* 313.5786 (2006): 504-507

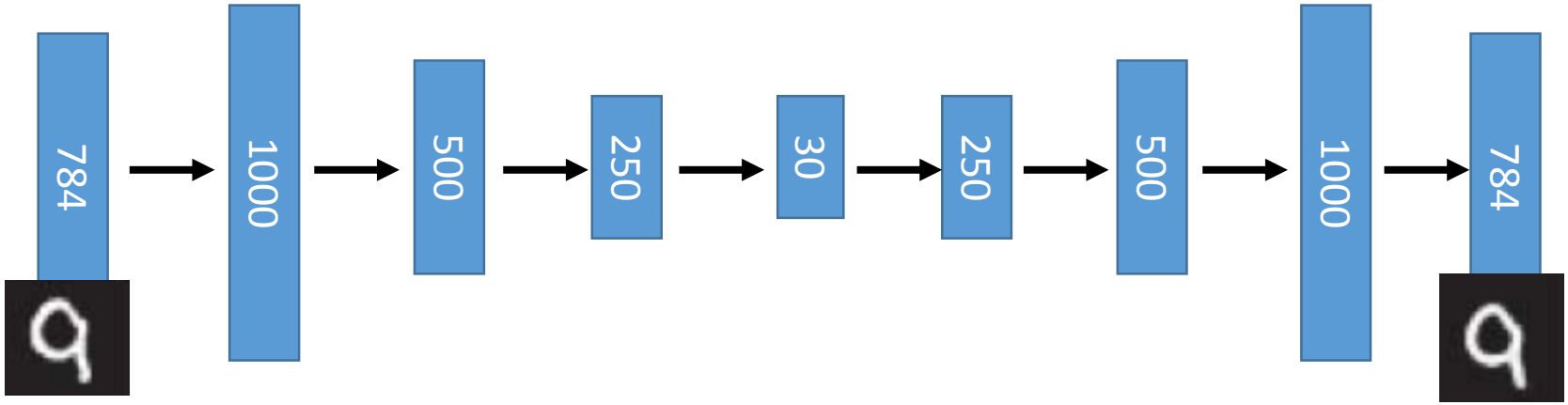
Deep Auto-encoder

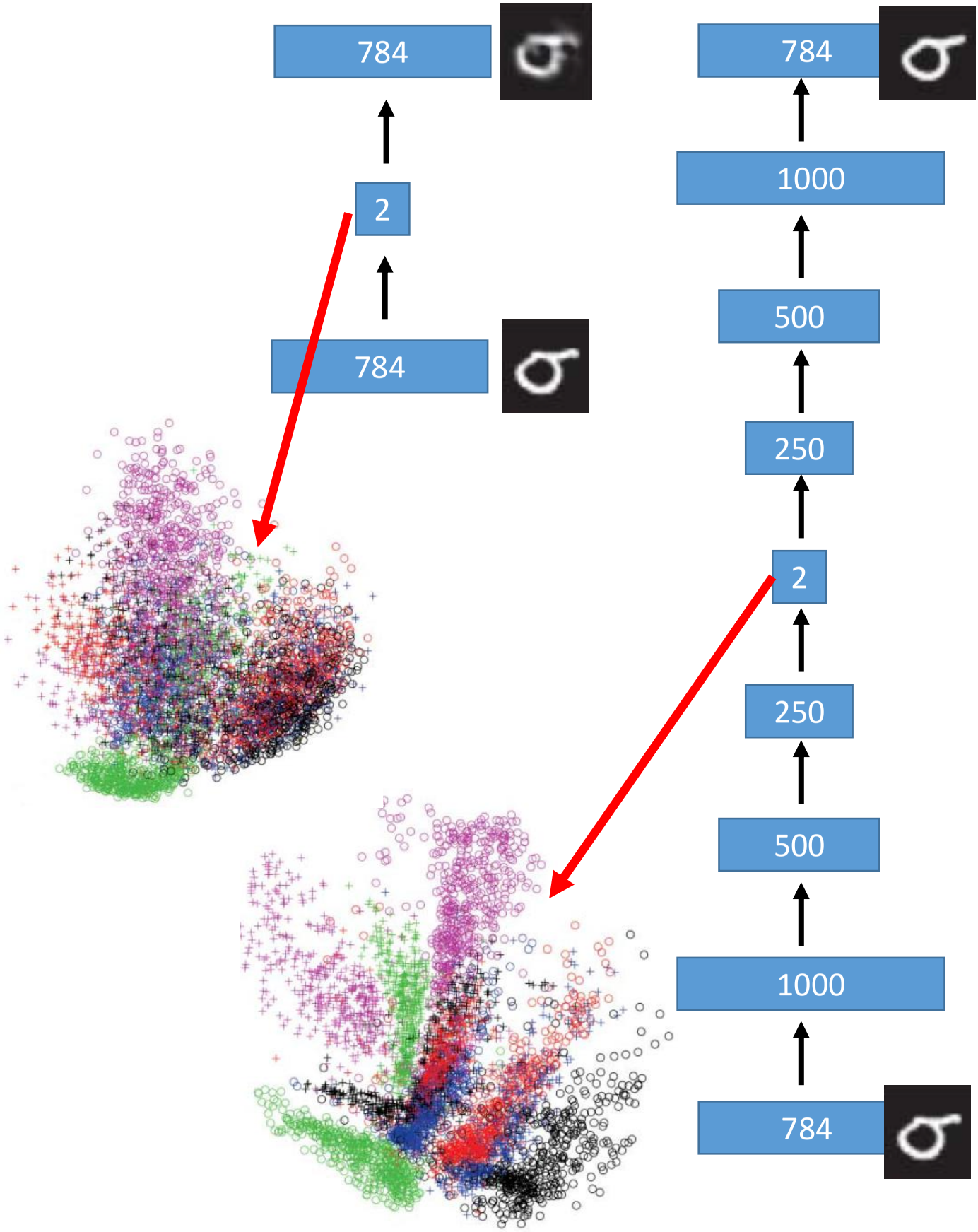
Original Image

PCA

Deep Auto-encoder





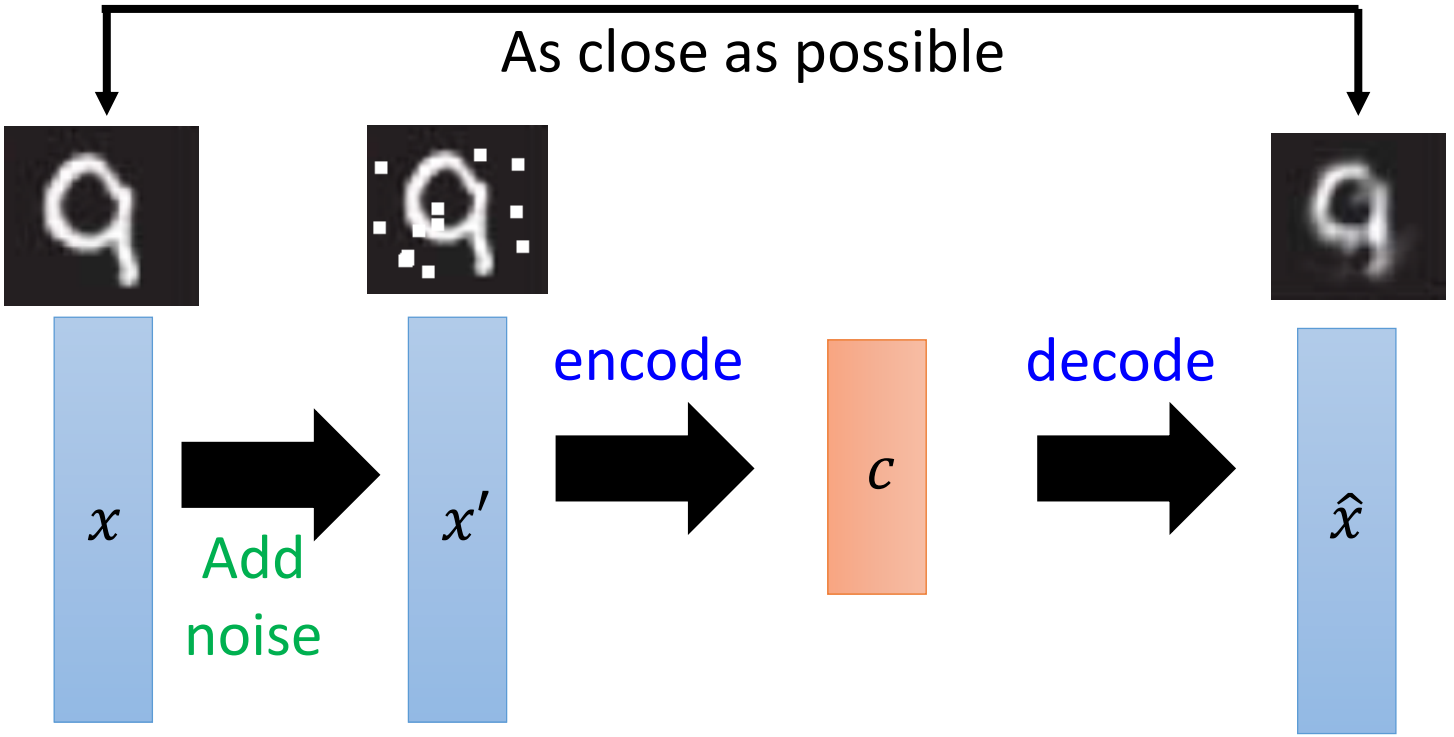

784
2
784
784
1000
500
250
2
250
500
1000
784

More: Contractive auto-encoder

Auto-encoder

Ref: Rifai, Salah, et al. "Contractive auto-encoders: Explicit invariance during feature extraction." *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011.

- De-noising auto-encoder

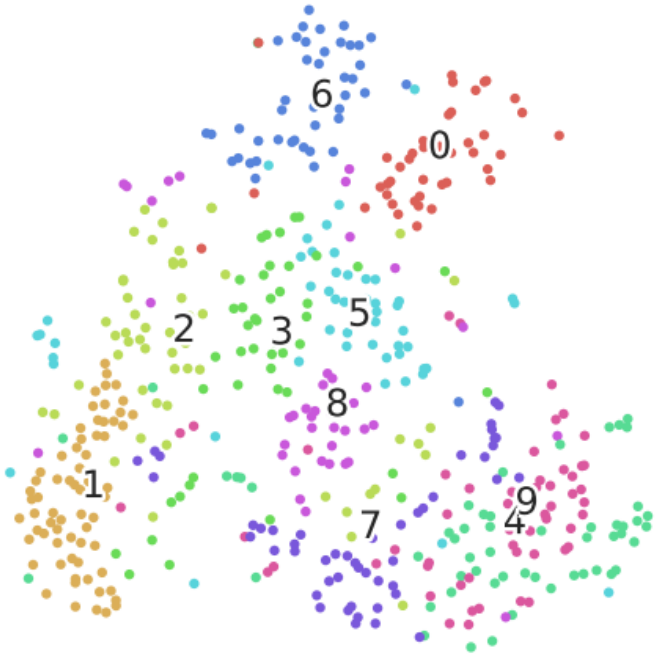


Vincent, Pascal, et al. "Extracting and composing robust features with denoising autoencoders." *ICML,* 2008.

Deep Auto-encoder - Example

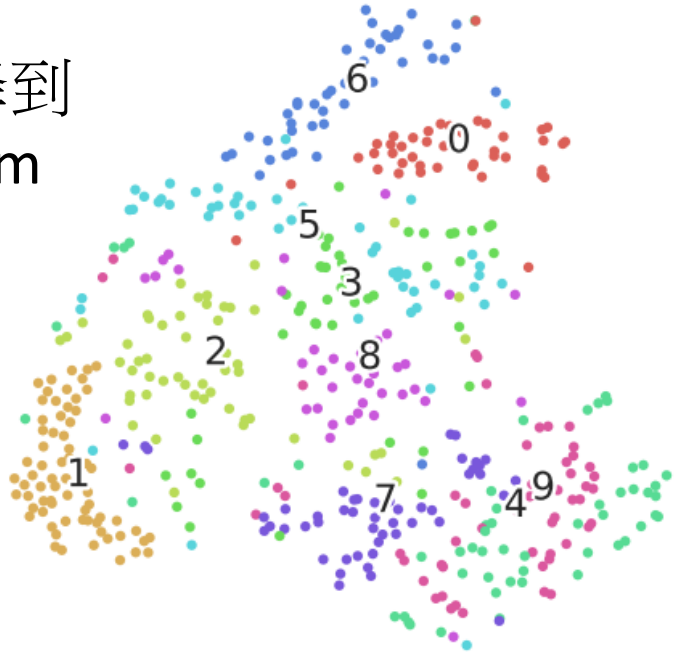
NN Encoder

c



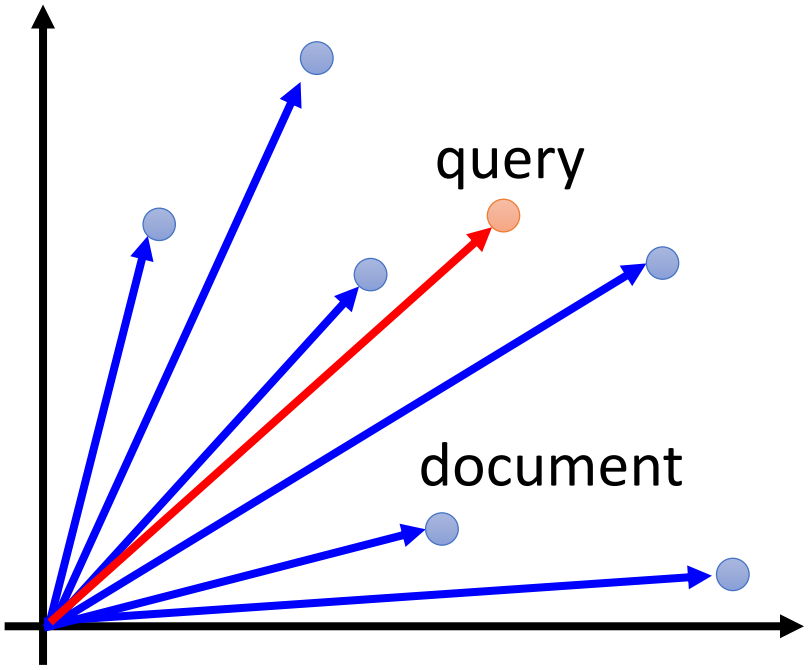
Pixel -> tSNE

PCA 降到 32-dim



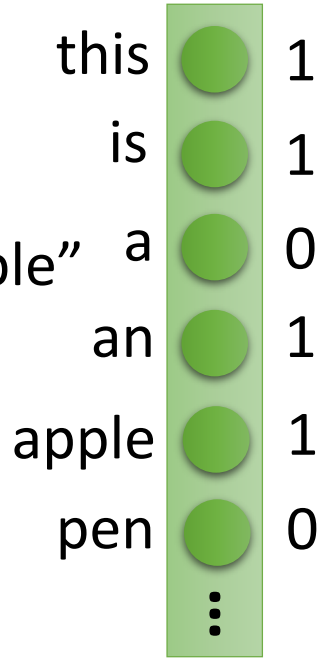
Auto-encoder – Text Retrieval

Vector Space Model



Bag-of-word

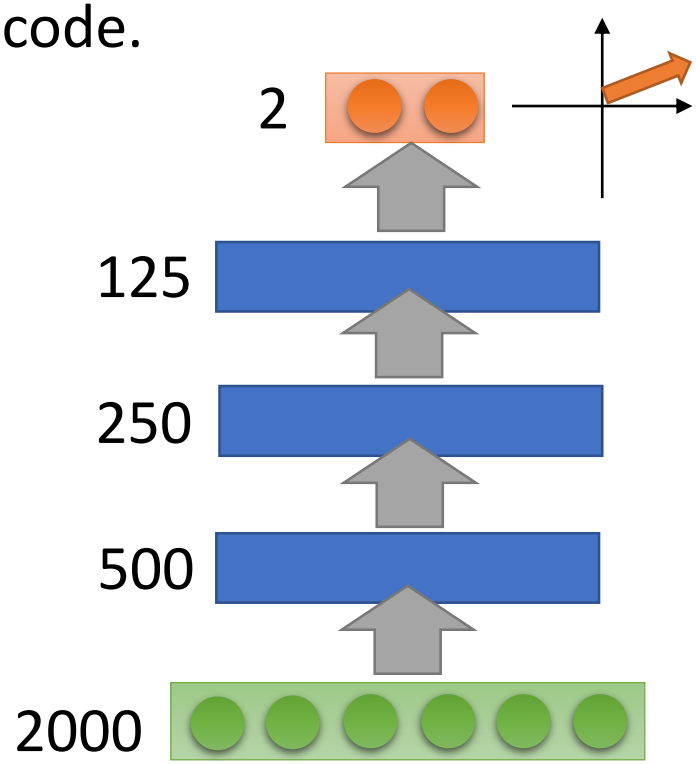
word string:
"This is an apple"



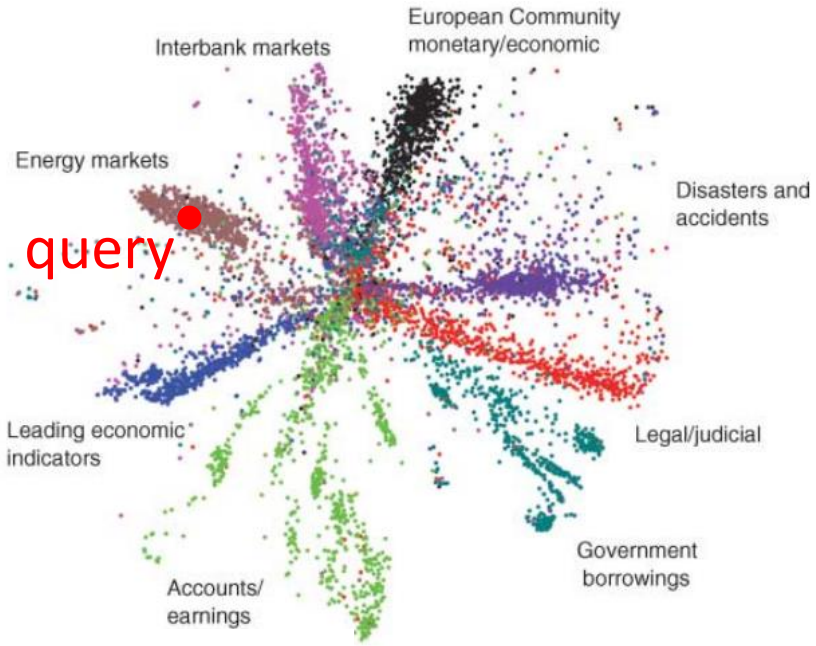
Semantics are not considered.

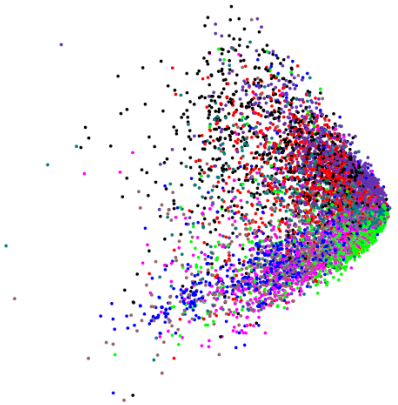
Auto-encoder – Text Retrieval

The documents talking about the same thing will have close code.



Bag-of-word
(document or query)





LSA: project documents to 2 latent topics

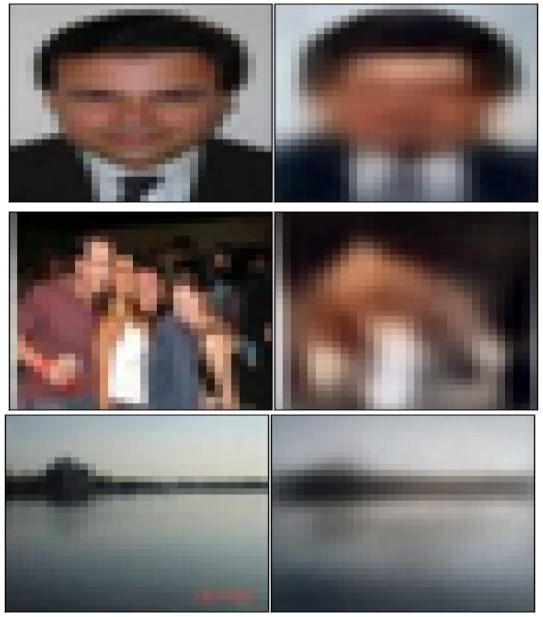
Auto-encoder – Similar Image Search

Retrieved using Euclidean distance in pixel intensity space

(Images from Hinton's slides on Coursera)

Reference: Krizhevsky, Alex, and Geoffrey E. Hinton. "Using very deep autoencoders for content-based image retrieval." *ESANN*. 2011.

Auto-encoder – Similar Image Search



32x32

8192

4096

2048

1024

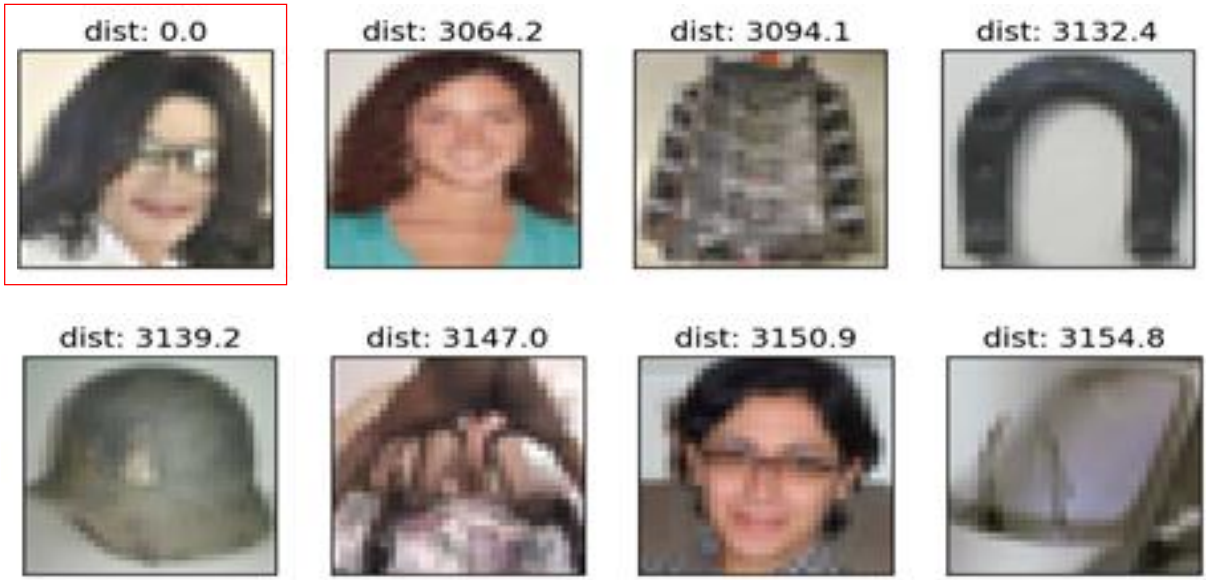
512

code

256

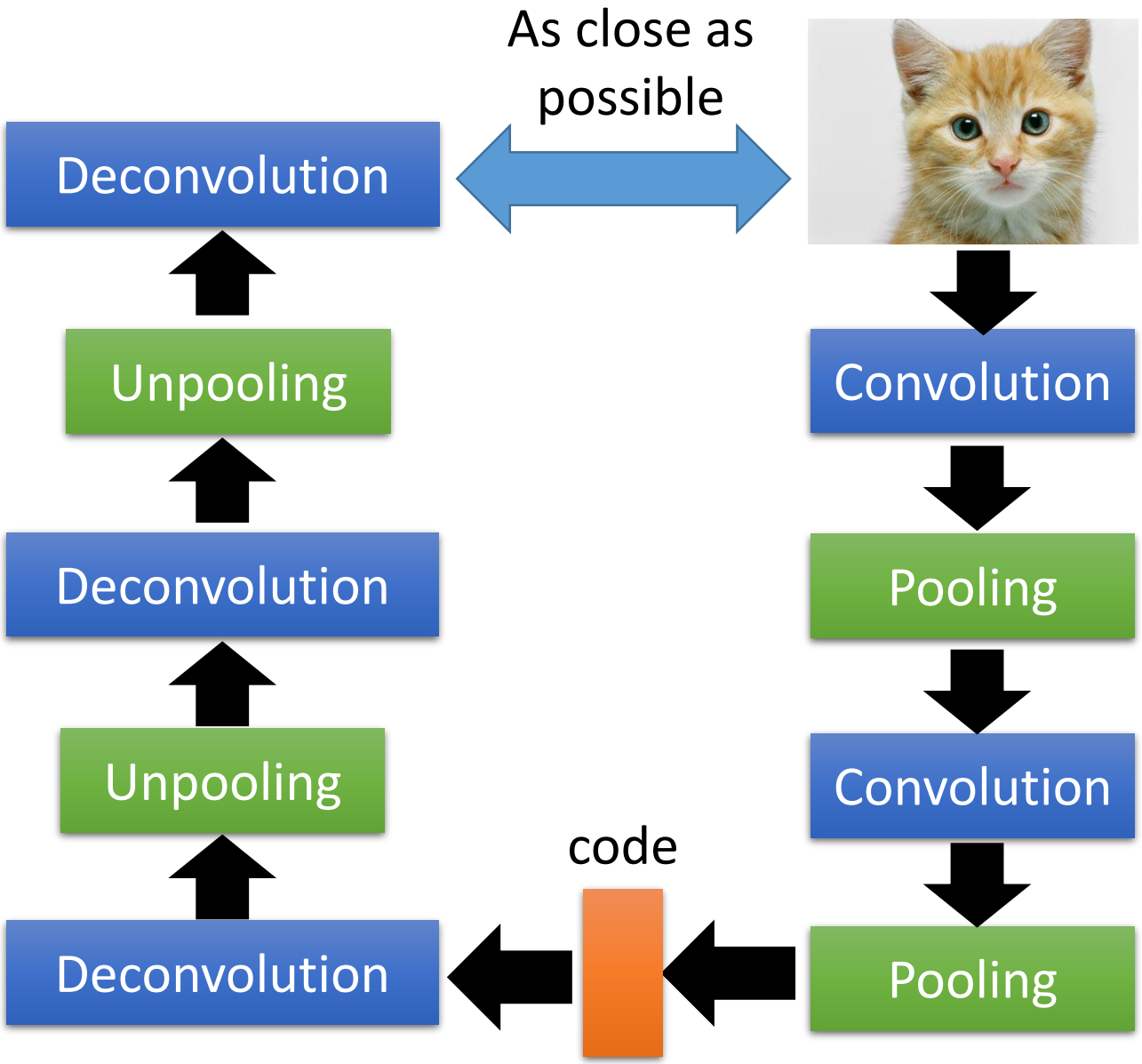
(crawl millions of images from the Internet)

Retrieved using Euclidean distance in pixel intensity space

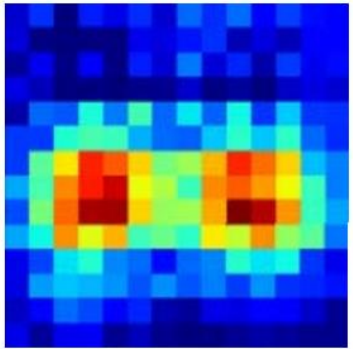


retrieved using 256 codes

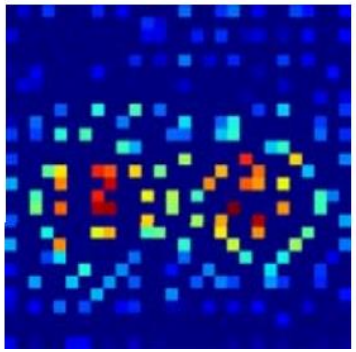
Auto-encoder for CNN



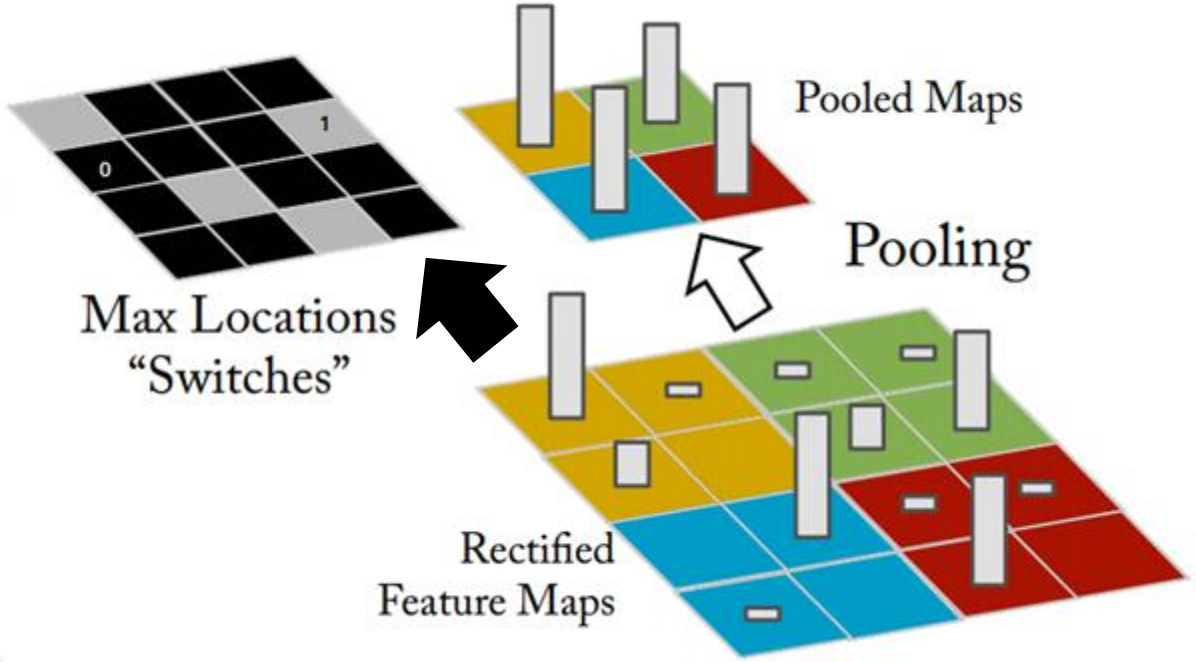
CNN -Unpooling



14 x 14



28 x 28

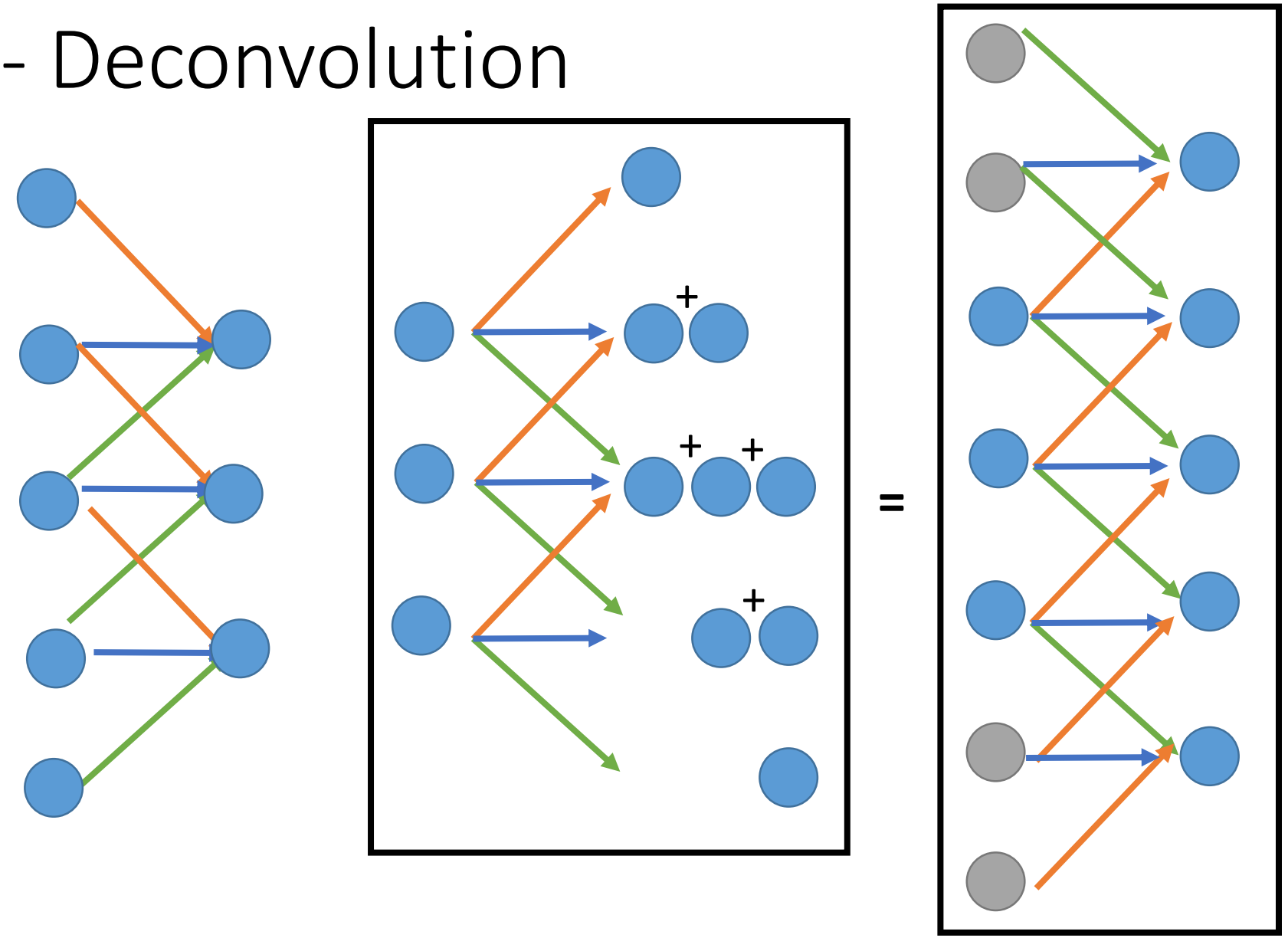


Alternative: simply repeat the values

Source of image : https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/image_segmentation.html

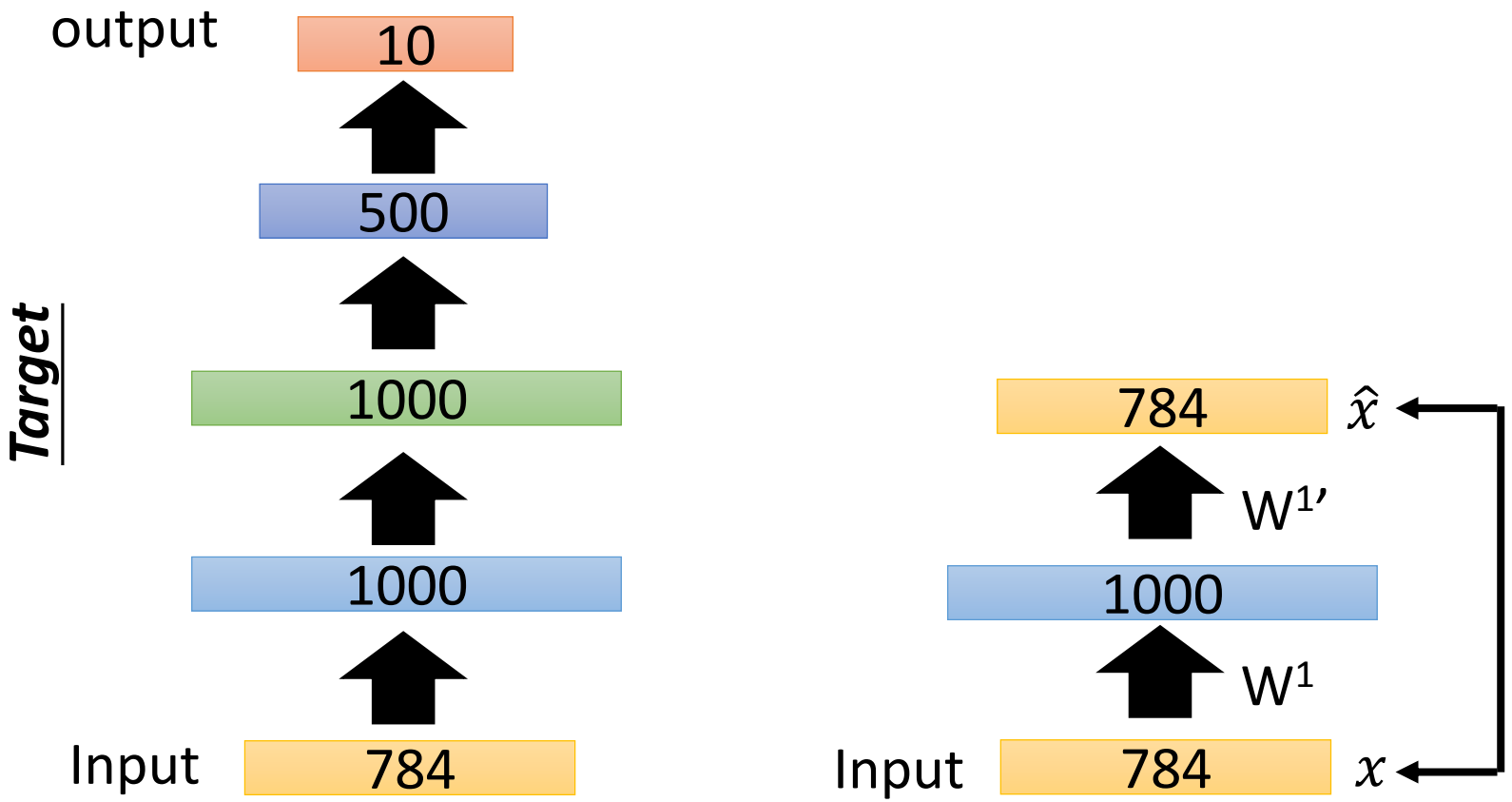
CNN
- Deconvolution

Actually, deconvolution is convolution.



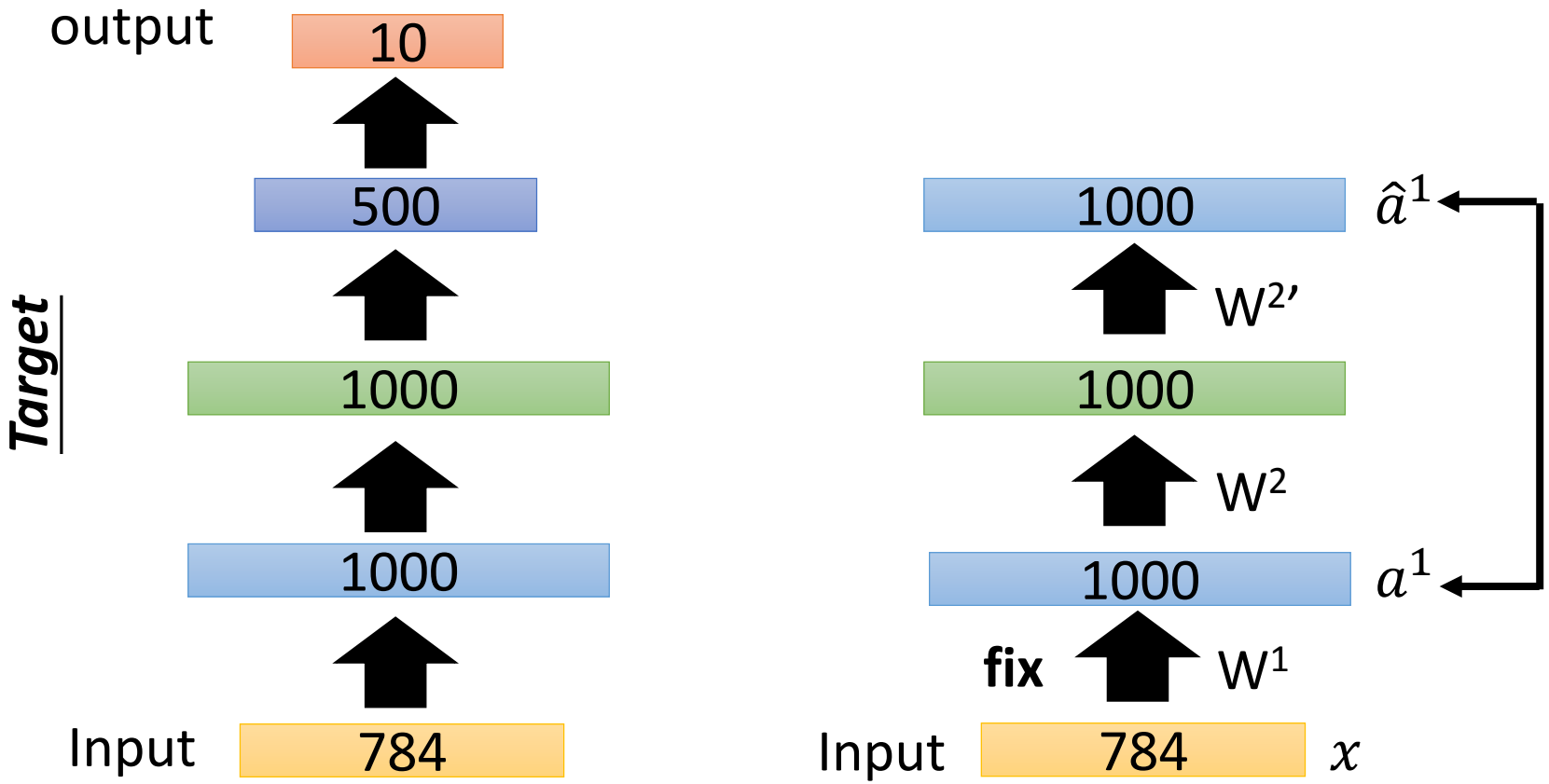
Auto-encoder – Pre-training DNN

- Greedy Layer-wise Pre-training *again*



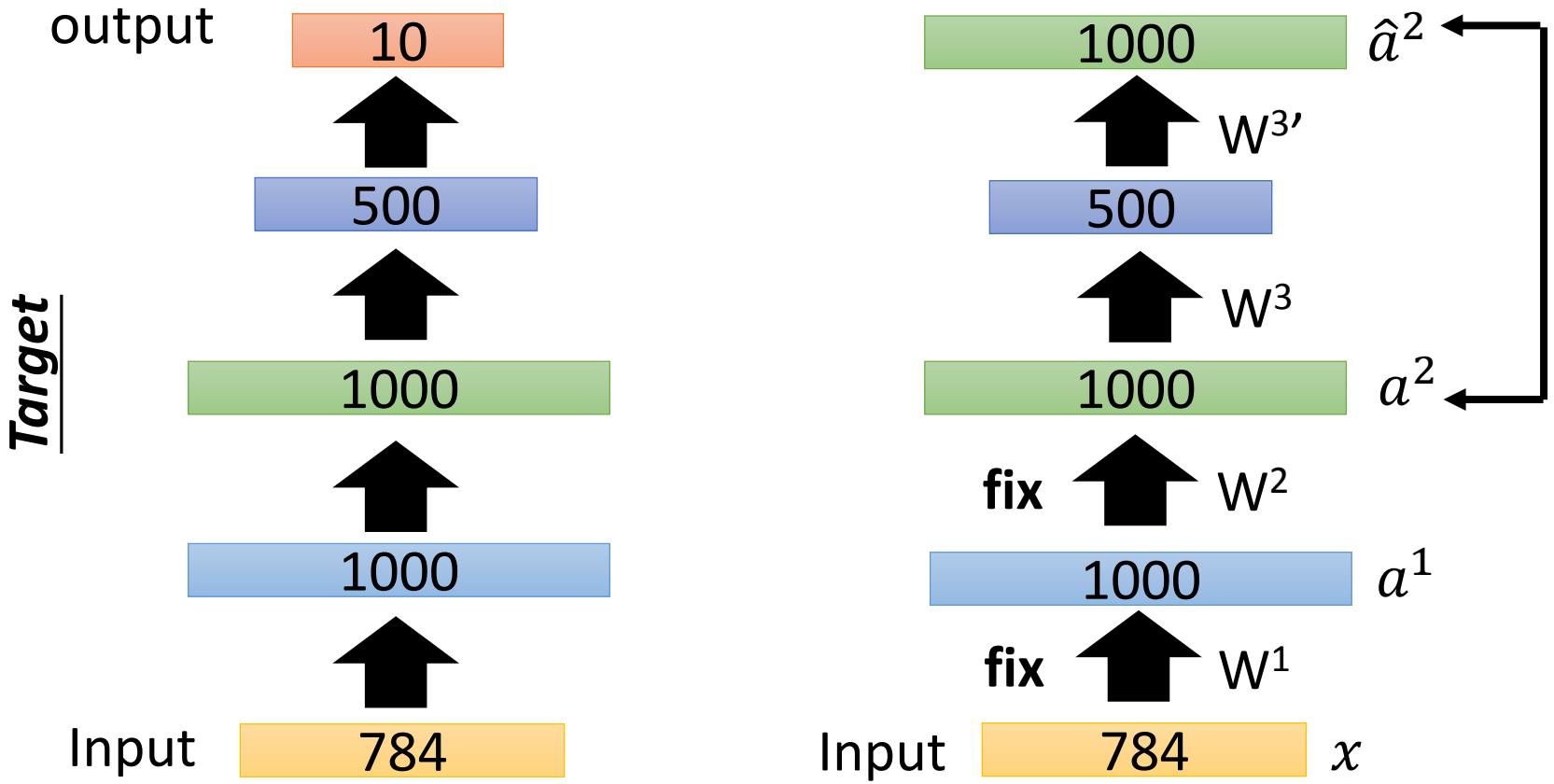
Auto-encoder – Pre-training DNN

- Greedy Layer-wise Pre-training *again*



Auto-encoder – Pre-training DNN

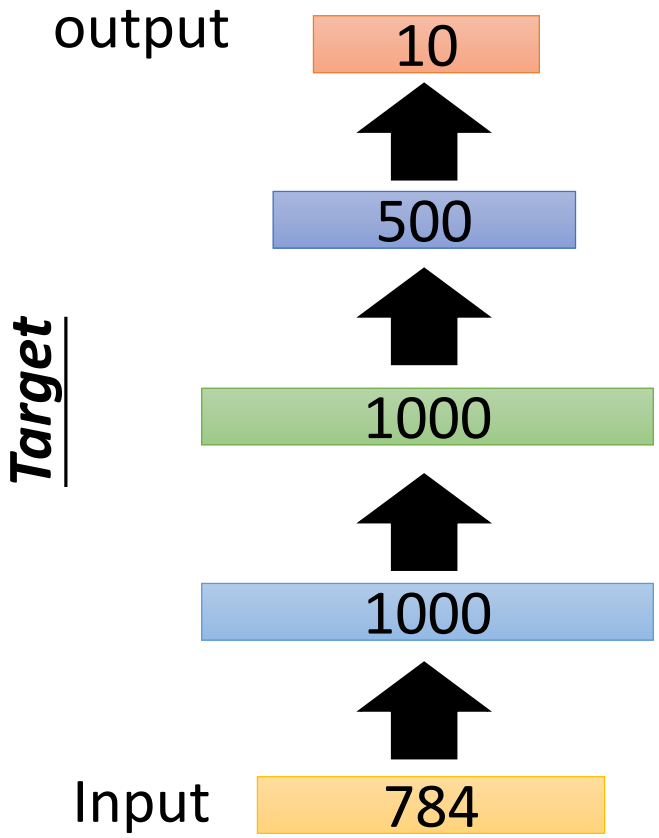
- Greedy Layer-wise Pre-training *again*

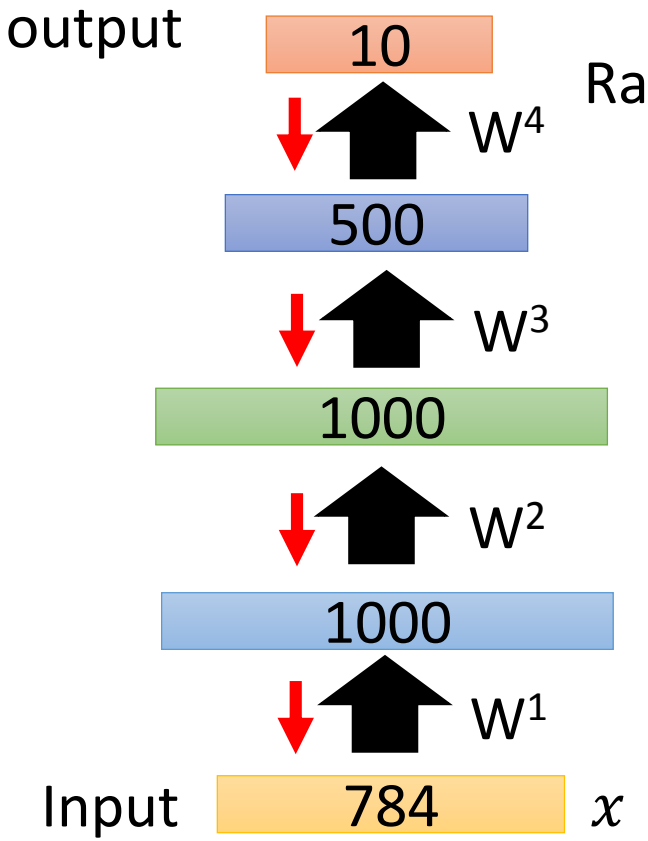


Auto-encoder – Pre-training DNN

- Greedy Layer-wise Pre-training *again*

Find-tune by backpropagation





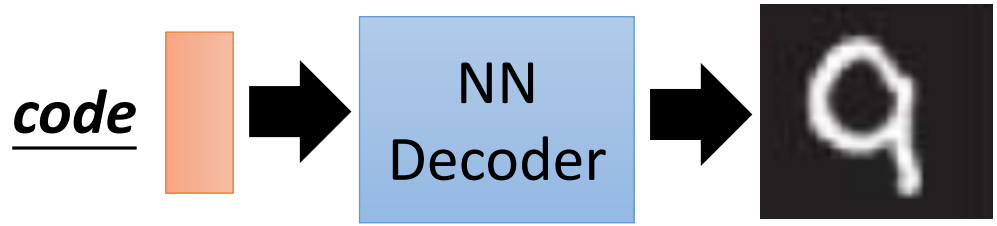
Learning More
- Restricted Boltzmann Machine

- Neural networks [5.1] : Restricted Boltzmann machine – definition
  - https://www.youtube.com/watch?v=p4Vh_zMw-HQ&index=36&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH
- Neural networks [5.2] : Restricted Boltzmann machine – inference
  - https://www.youtube.com/watch?v=lekCh_i32iE&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH&index=37
- Neural networks [5.3] : Restricted Boltzmann machine - free energy
  - https://www.youtube.com/watch?v=e0Ts_7Y6hZU&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH&index=38

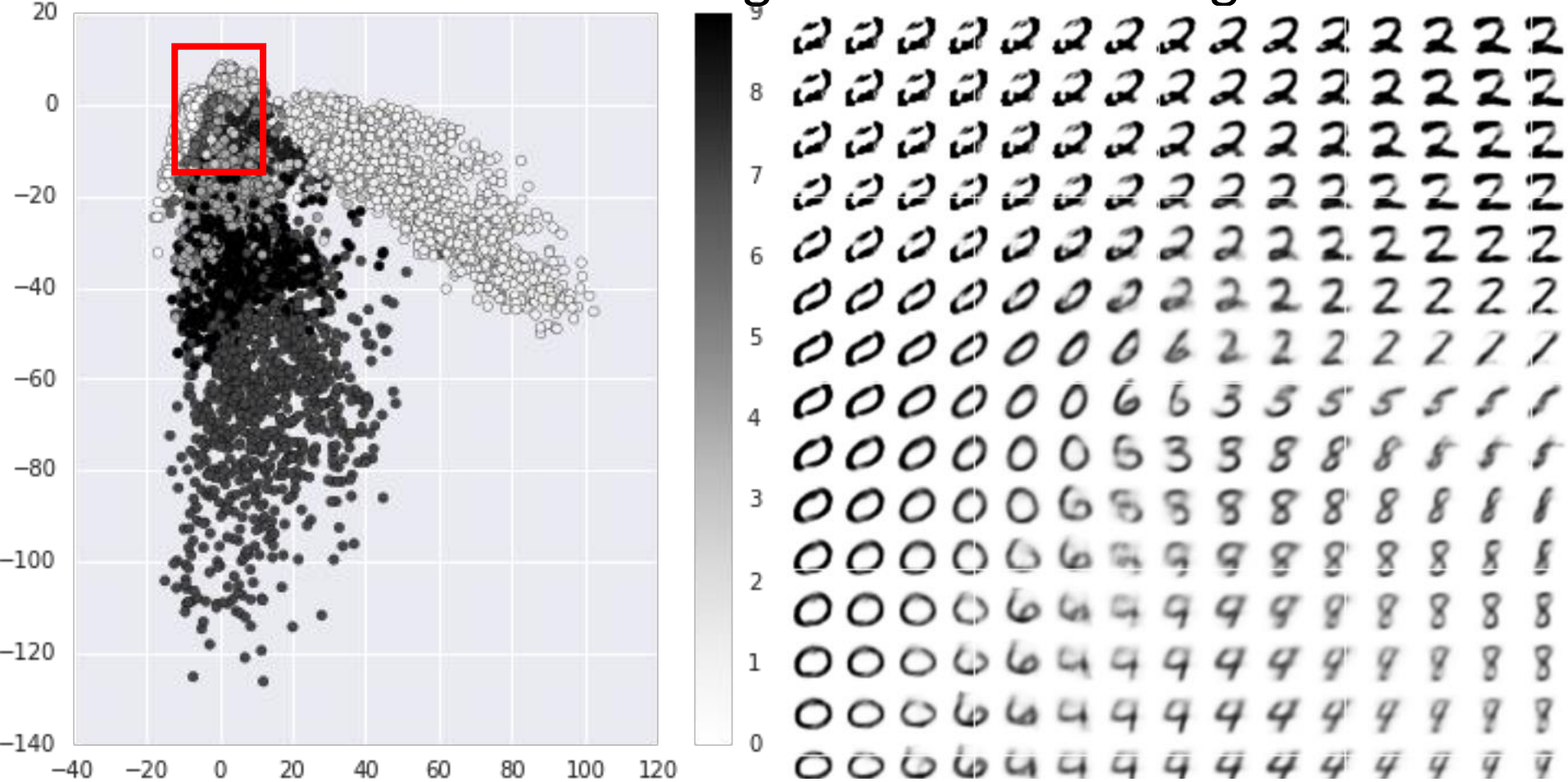
Learning More
- Deep Belief Network

- Neural networks [7.7] : Deep learning - deep belief network
  - https://www.youtube.com/watch?v=vkb6AWYXZ5I&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH&index=57
- Neural networks [7.8] : Deep learning - variational bound
  - https://www.youtube.com/watch?v=pStDscJh2Wo&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH&index=58
- Neural networks [7.9] : Deep learning - DBN pre-training
  - https://www.youtube.com/watch?v=35MUlYCColk&list=PL6Xpj9I5qXYEcOhn7TqghAJ6NAPrNmUBH&index=59

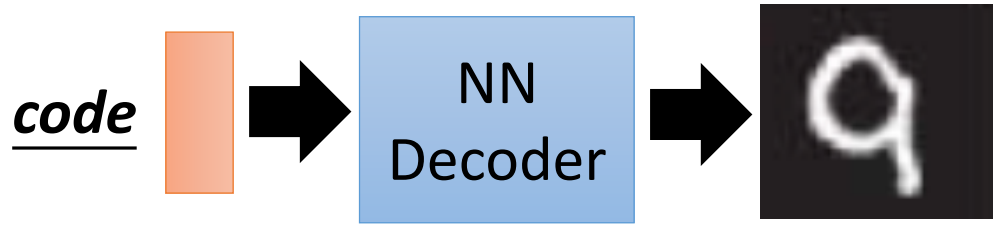
Next .....



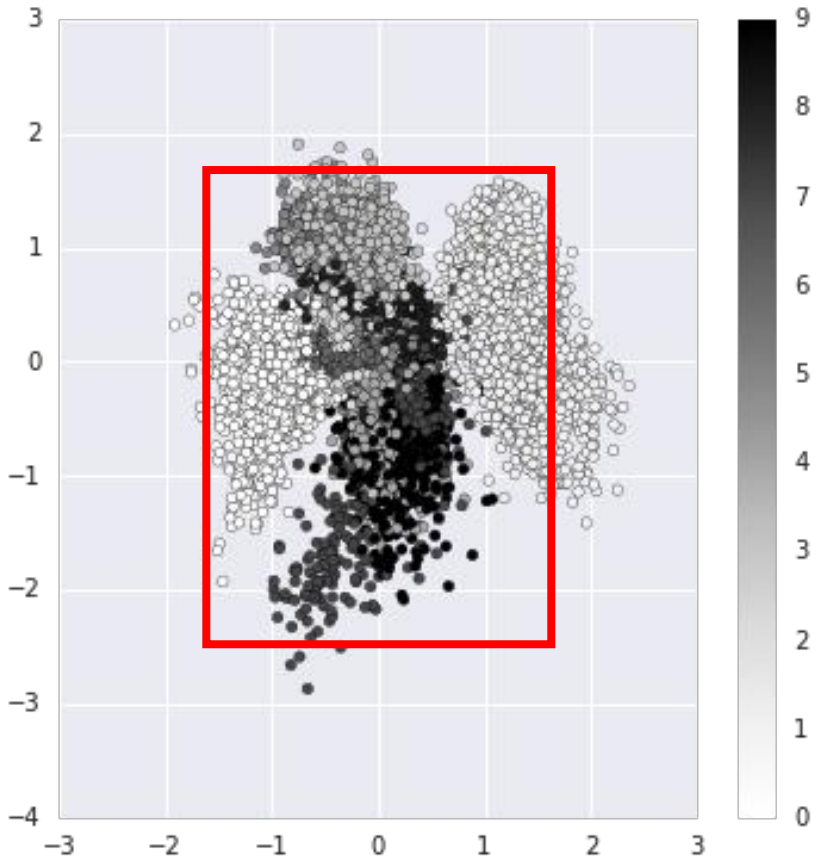
- Can we use decoder to generate something?

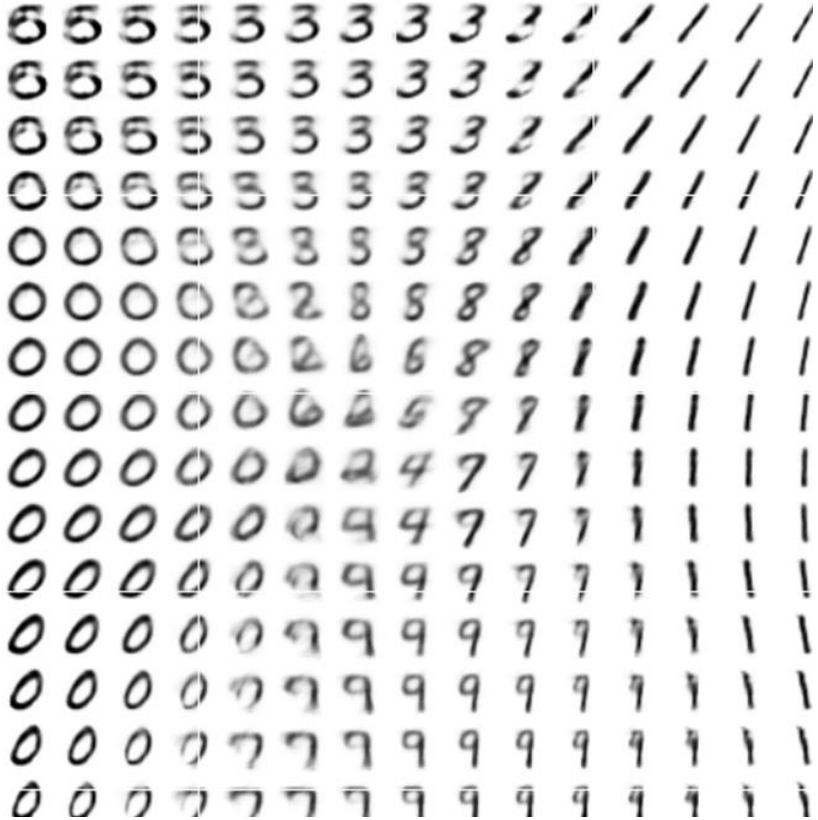


Next .....



- Can we use decoder to generate something?

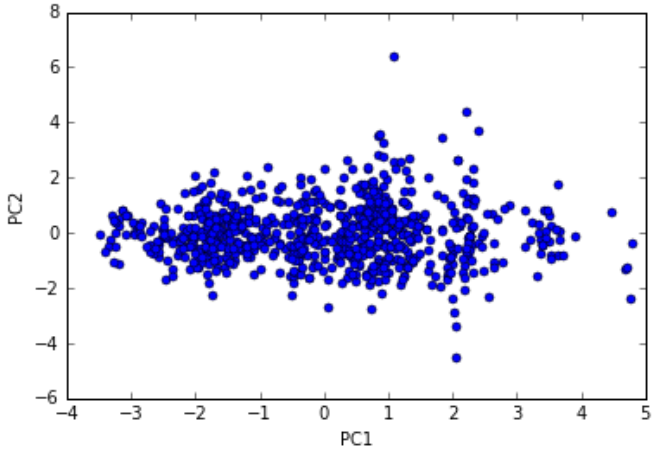




Appendix

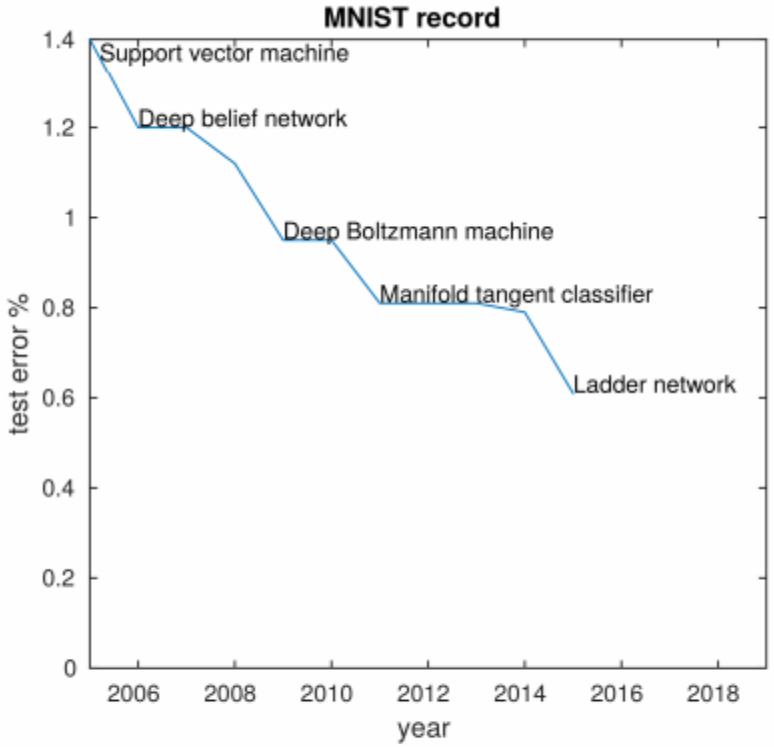
Pokémon

- http://140.112.21.35:2880/~tlkagk/pokemon/pca.html
- http://140.112.21.35:2880/~tlkagk/pokemon/auto.html
- The code is modified from
  - http://jkunst.com/r/pokemon-visualize-em-all/



Add: Ladder Network

- http://rinuboney.github.io/2016/01/19/ladder-network.html
- https://mycourses.aalto.fi/pluginfile.php/146701/mod_resource/content/1/08%20semisup%20ladder.pdf
- https://arxiv.org/abs/1507.02672

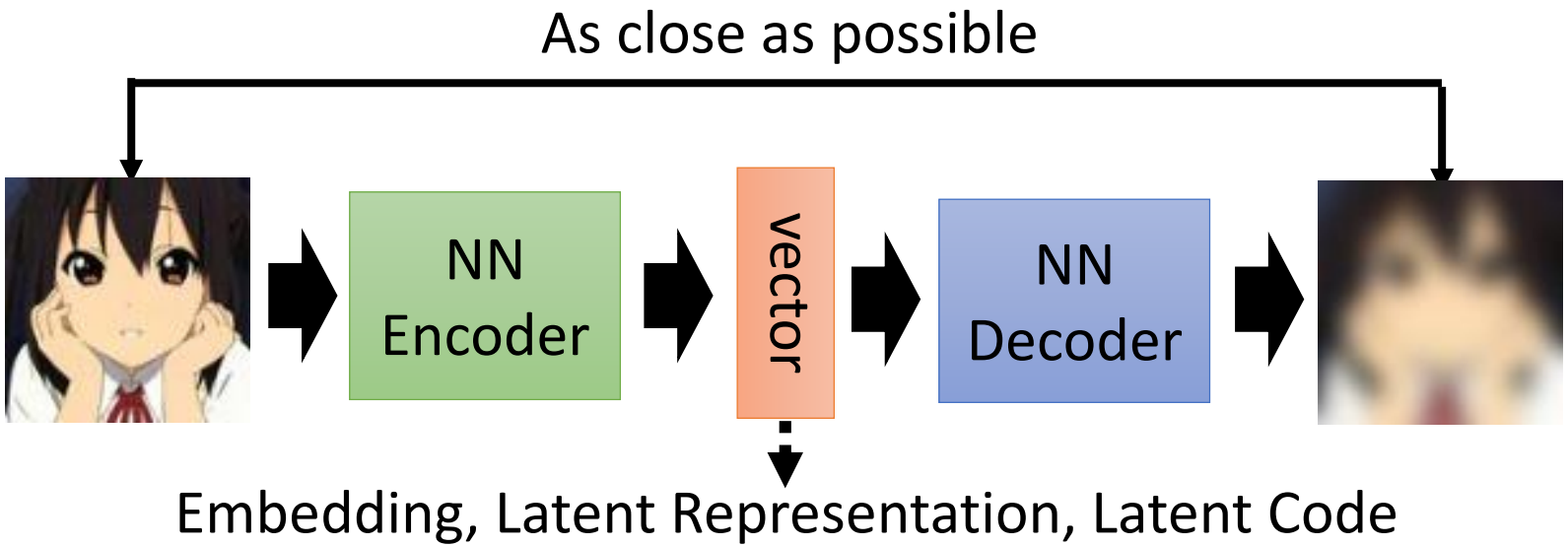


Yearly progress in permutation-invariant MNIST.

A. Rasmus, H. Valpola, M. Honkala, M. Berglund, and T. Raiko.

Semi-Supervised Learning with Ladder Network. To appear in NIPS 2015.

Auto-encoder



- More than minimizing reconstruction error
- More interpretable embedding

What is good embedding?

- An embedding should represent the object.

是一對

不是一對

Beyond Reconstruction

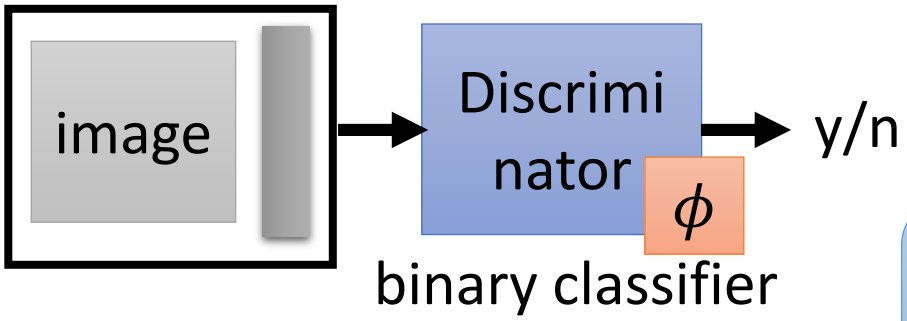
How to evaluate an encoder?

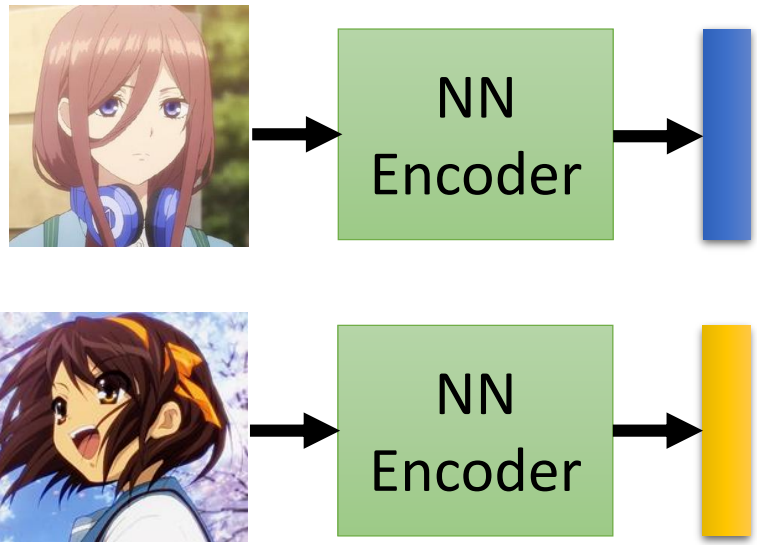
loss of the classification task is L_D

Train ϕ to minimize L_D

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

Small L_D^* ➡ The embeddings are representative.





Beyond Reconstruction

How to evaluate an encoder?

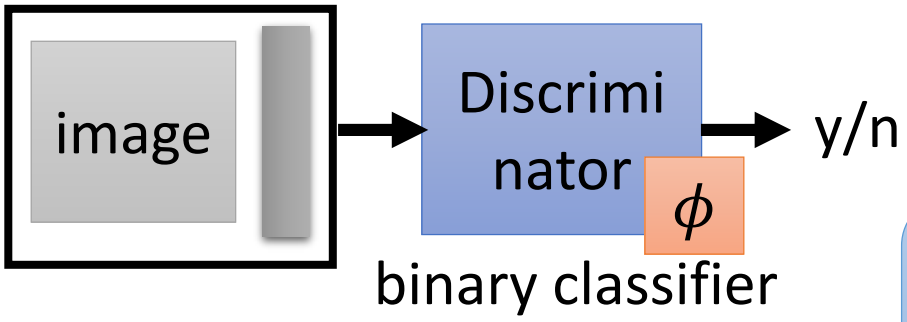
loss of the classification task is L_D

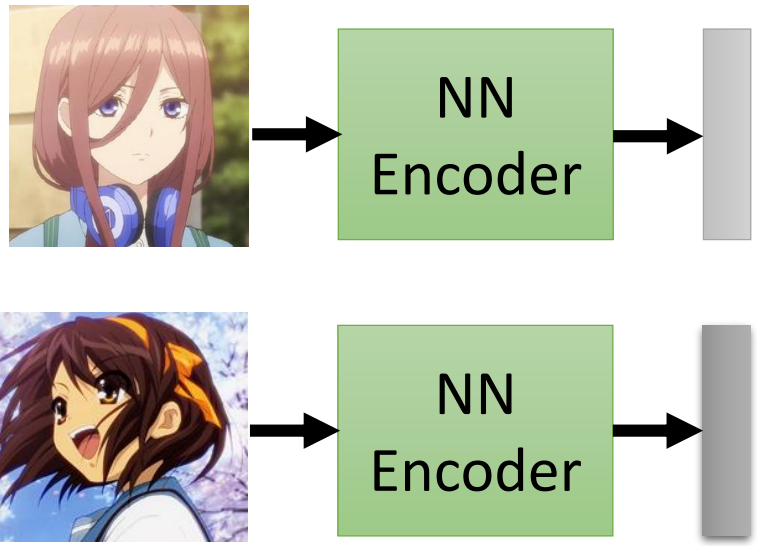
Train ϕ to minimize L_D

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

Small L_D^* ➡ The embeddings are representative.

Large L_D^* ➡ Not representative

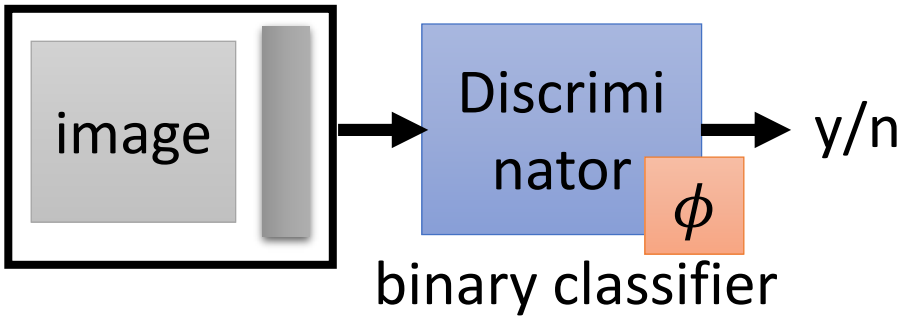




Beyond Reconstruction

How to evaluate an encoder?

loss of the classification task is L_D



Train ϕ to minimize L_D

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

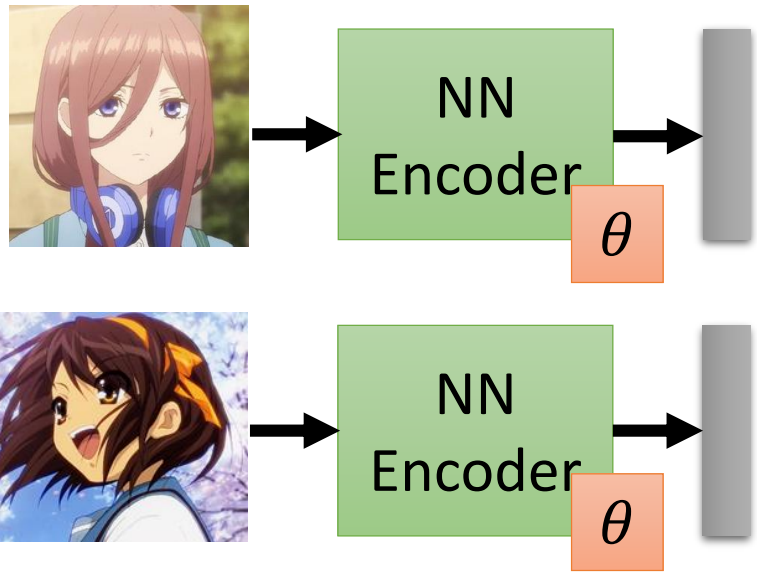
Small L_D^* ➡ The embeddings are representative.

Large L_D^* ➡ Not representative

Train θ to minimize L_D^*

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

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

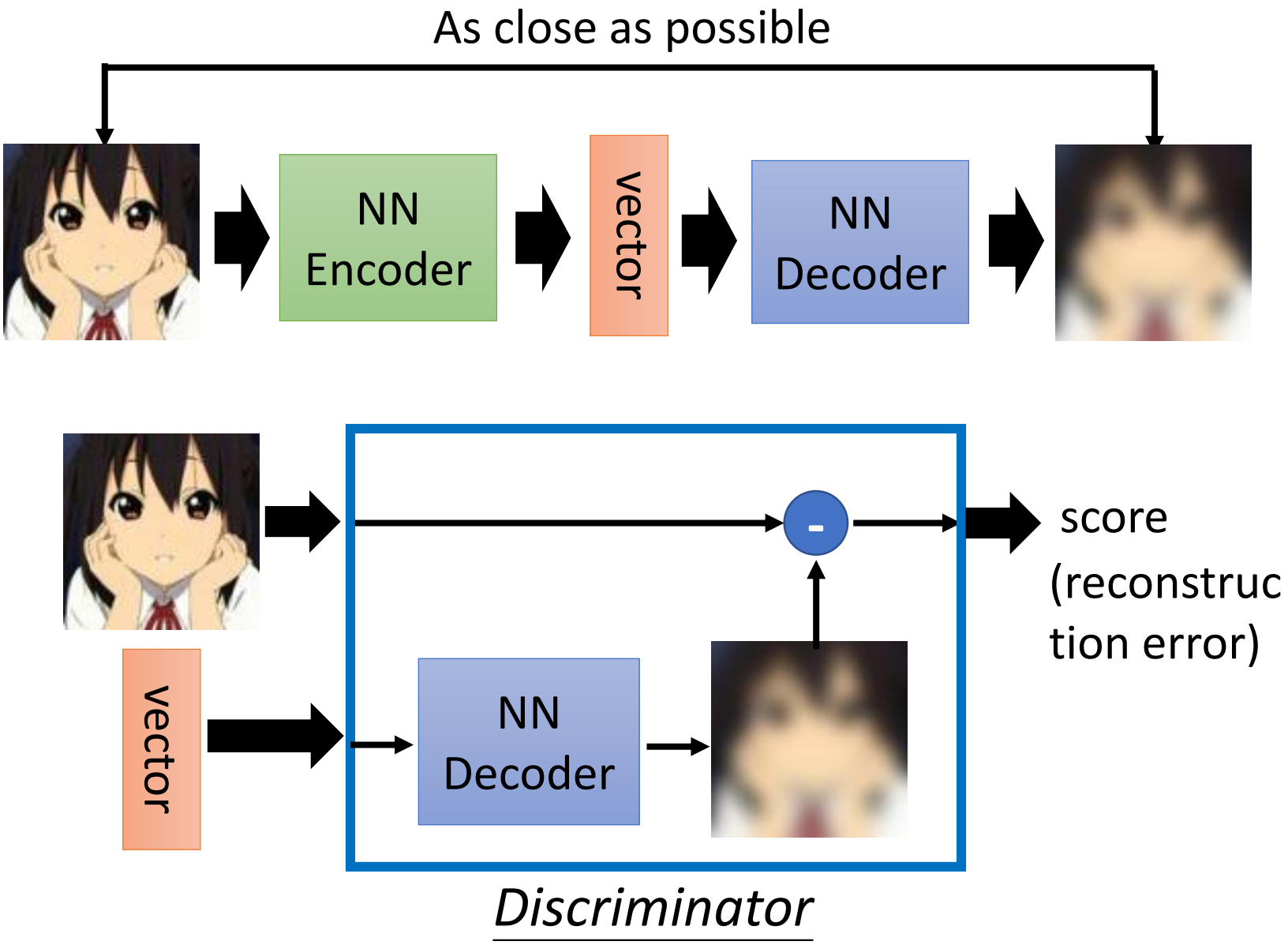


Train the encoder θ and discriminator ϕ to minimize L_D

Deep InfoMax (DIM)

(c.f. training encoder and decoder to minimize reconstruction error)

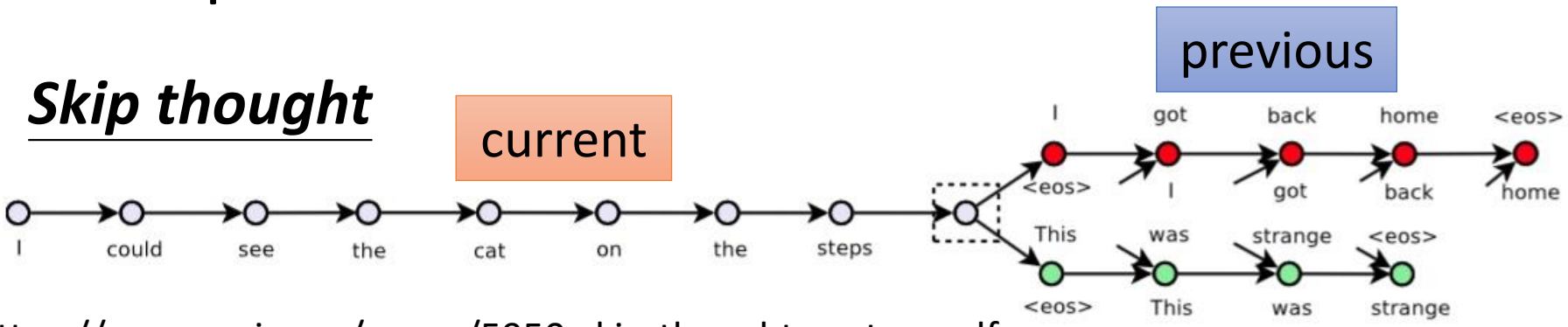
Typical auto-encoder is a special case



Sequential Data

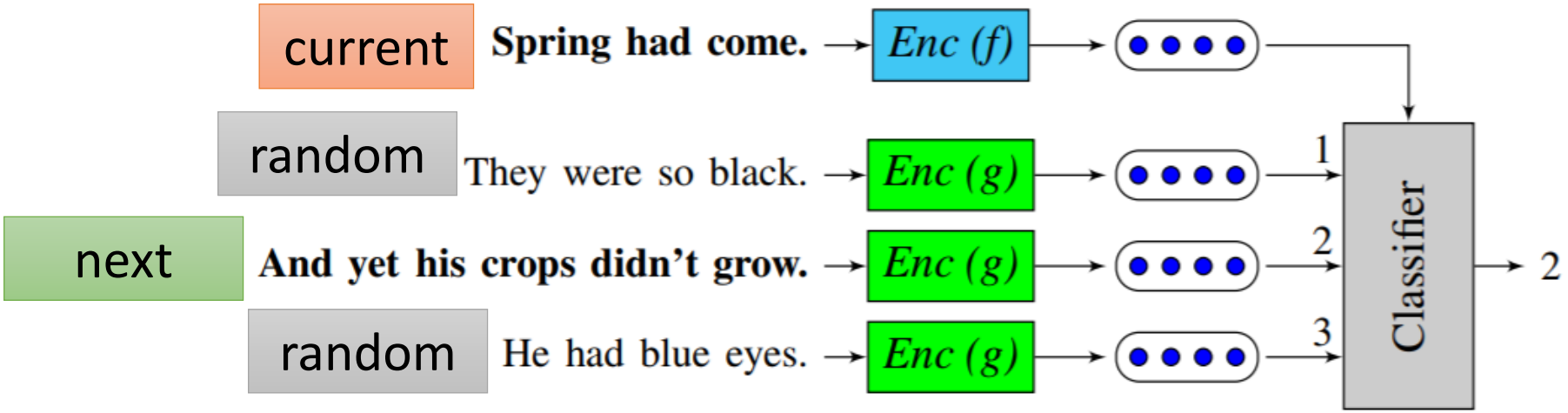
A document is a sequence of sentences.

Skip thought



https://papers.nips.cc/paper/5950-skip-thought-vectors.pdf

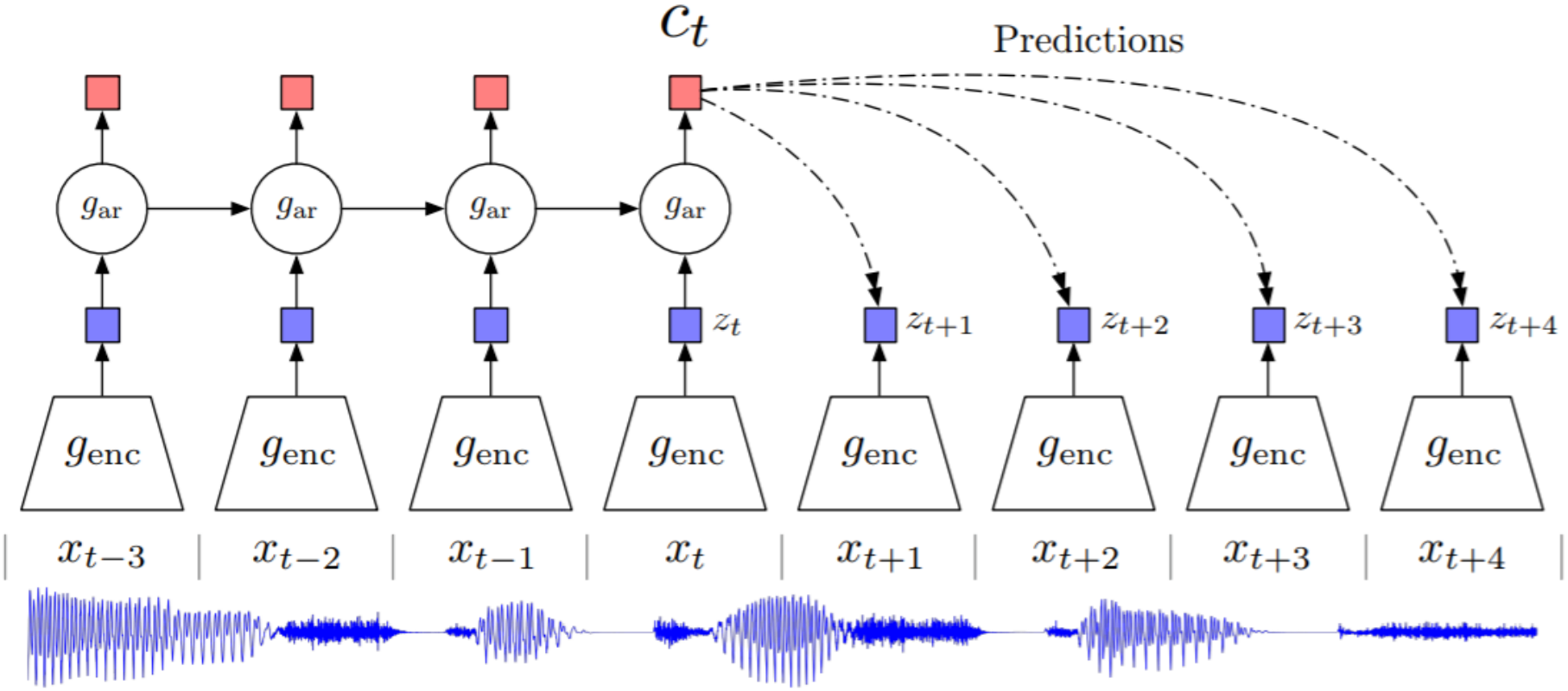
Quick thought



https://arxiv.org/pdf/1803.02893.pdf

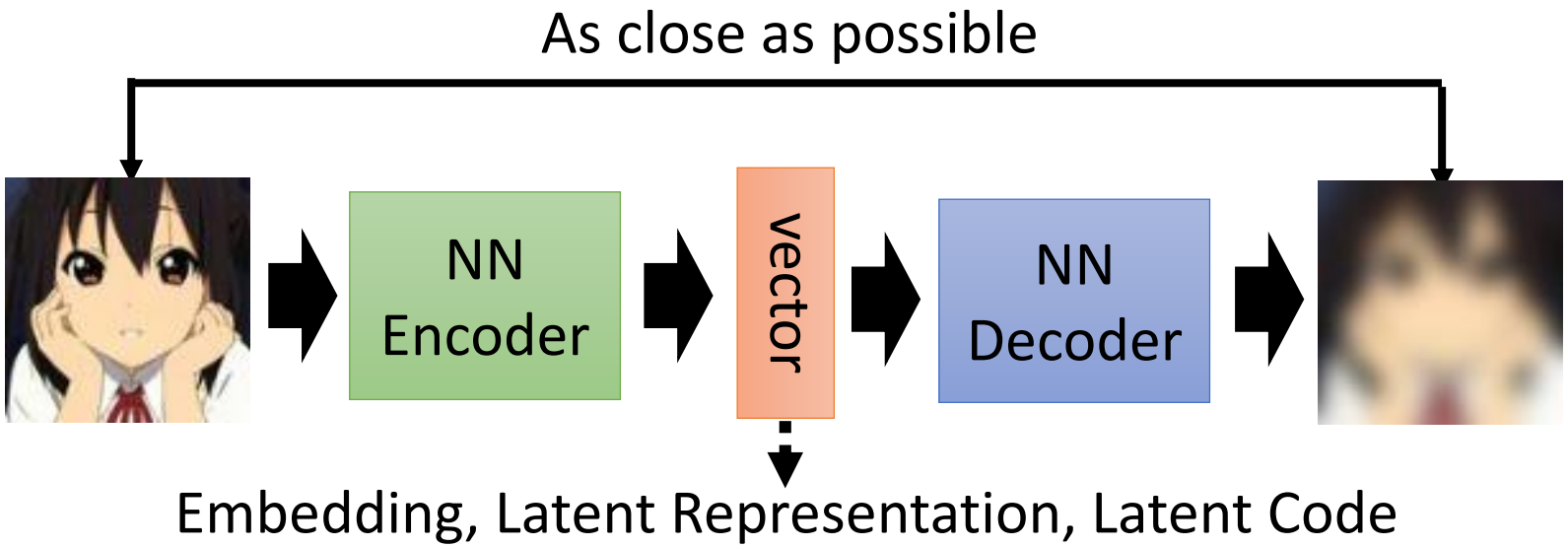
Sequential Data

- Contrastive Predictive Coding (CPC)



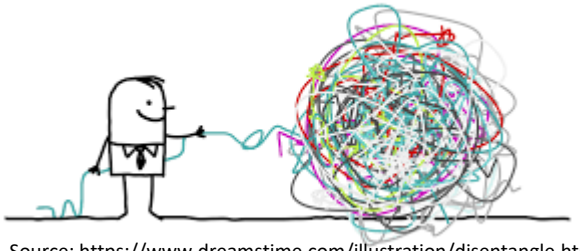
https://arxiv.org/pdf/1807.03748.pdf

Auto-encoder



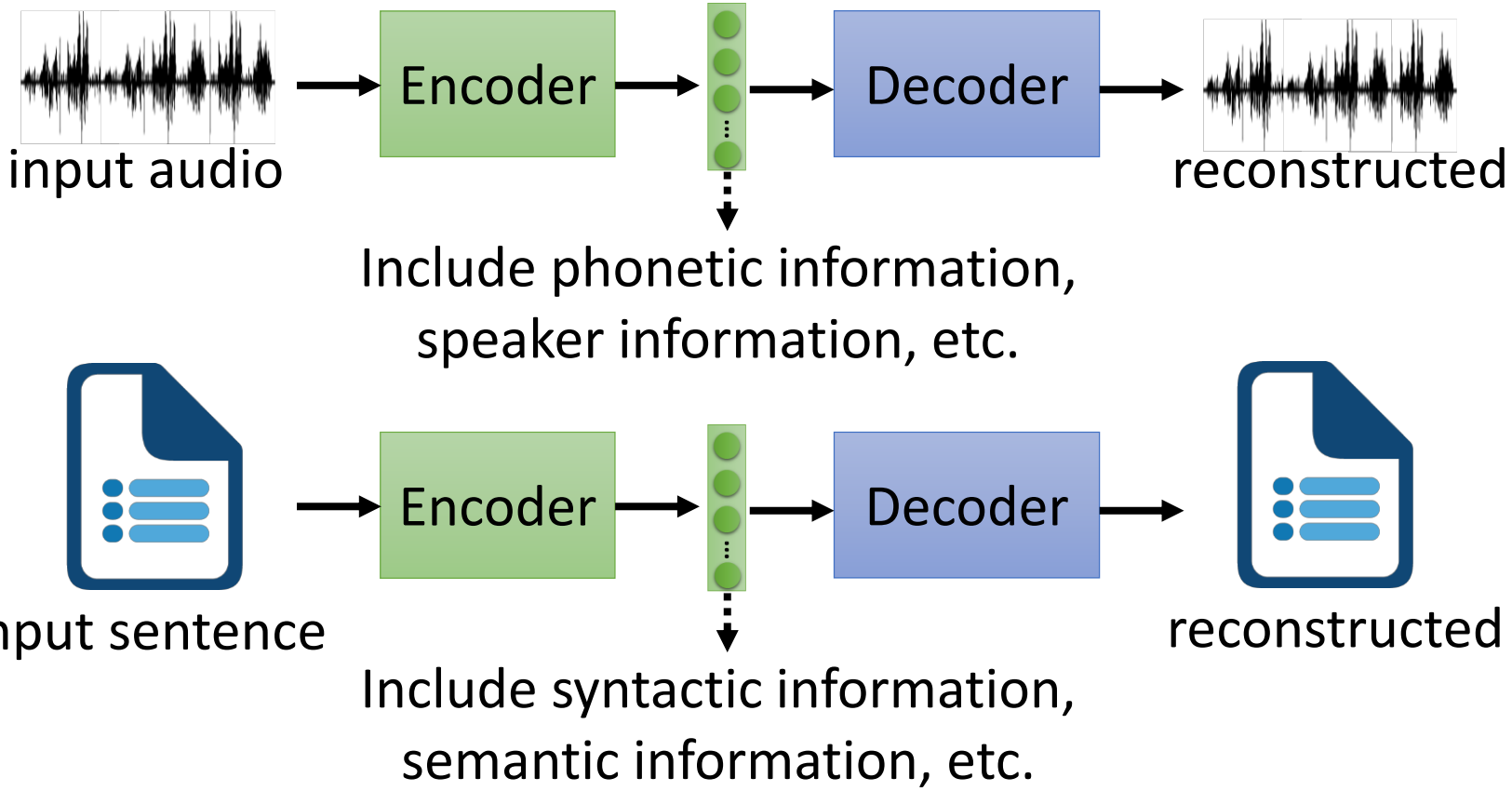
- More than minimizing reconstruction error
- More interpretable embedding

Feature Disentangle

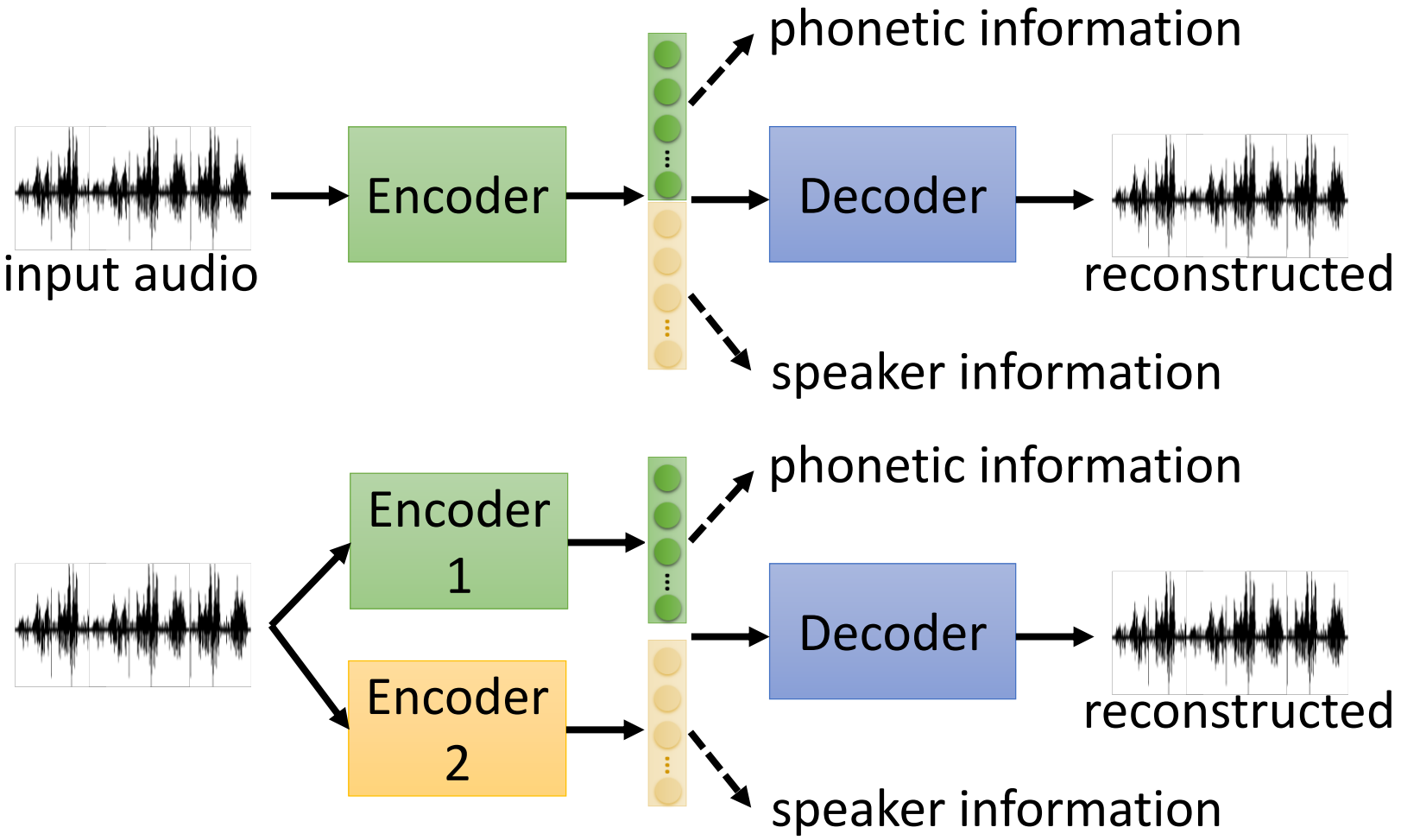


Source: https://www.dreamstime.com/illustration/disentangle.html

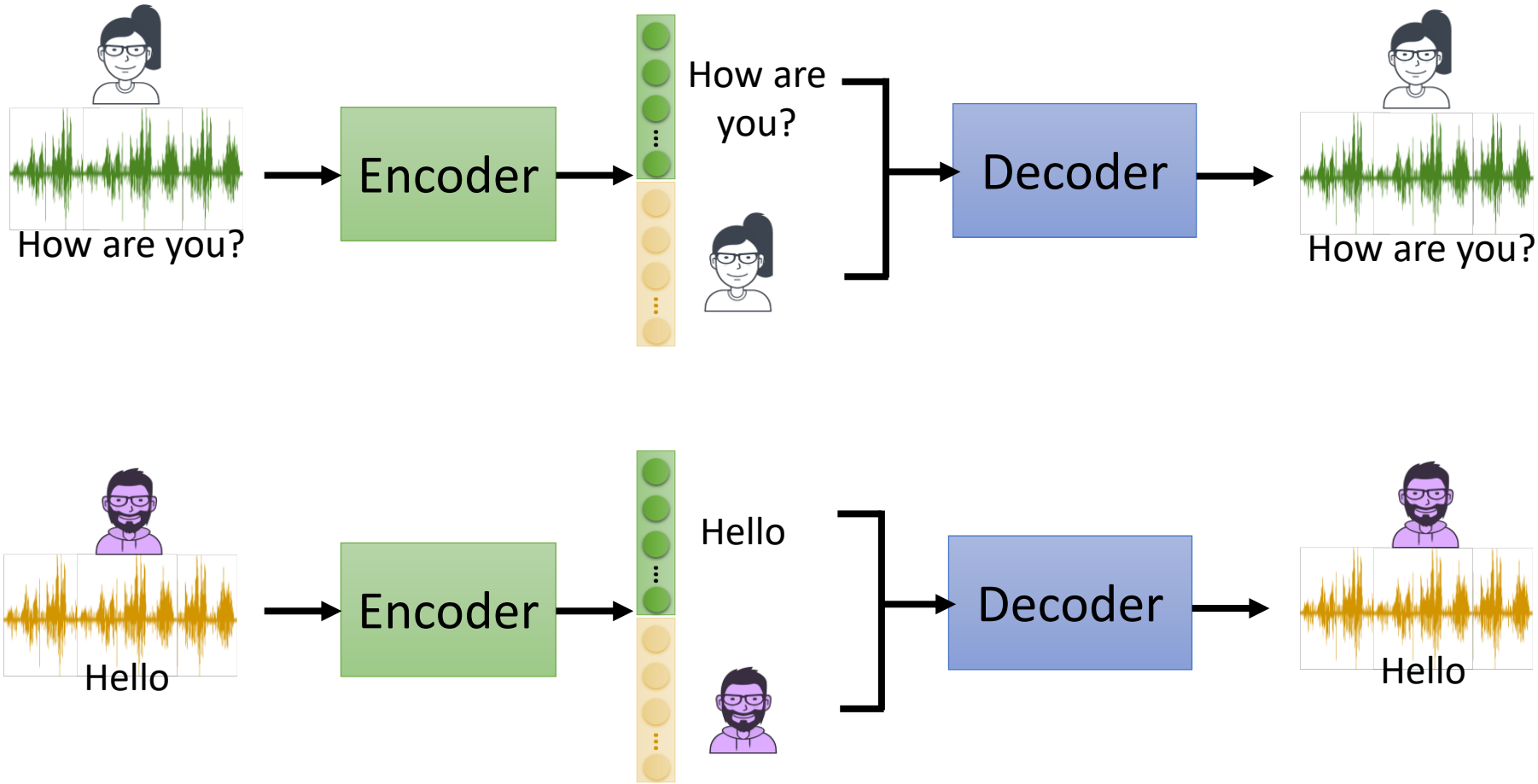
- An object contains multiple aspect information



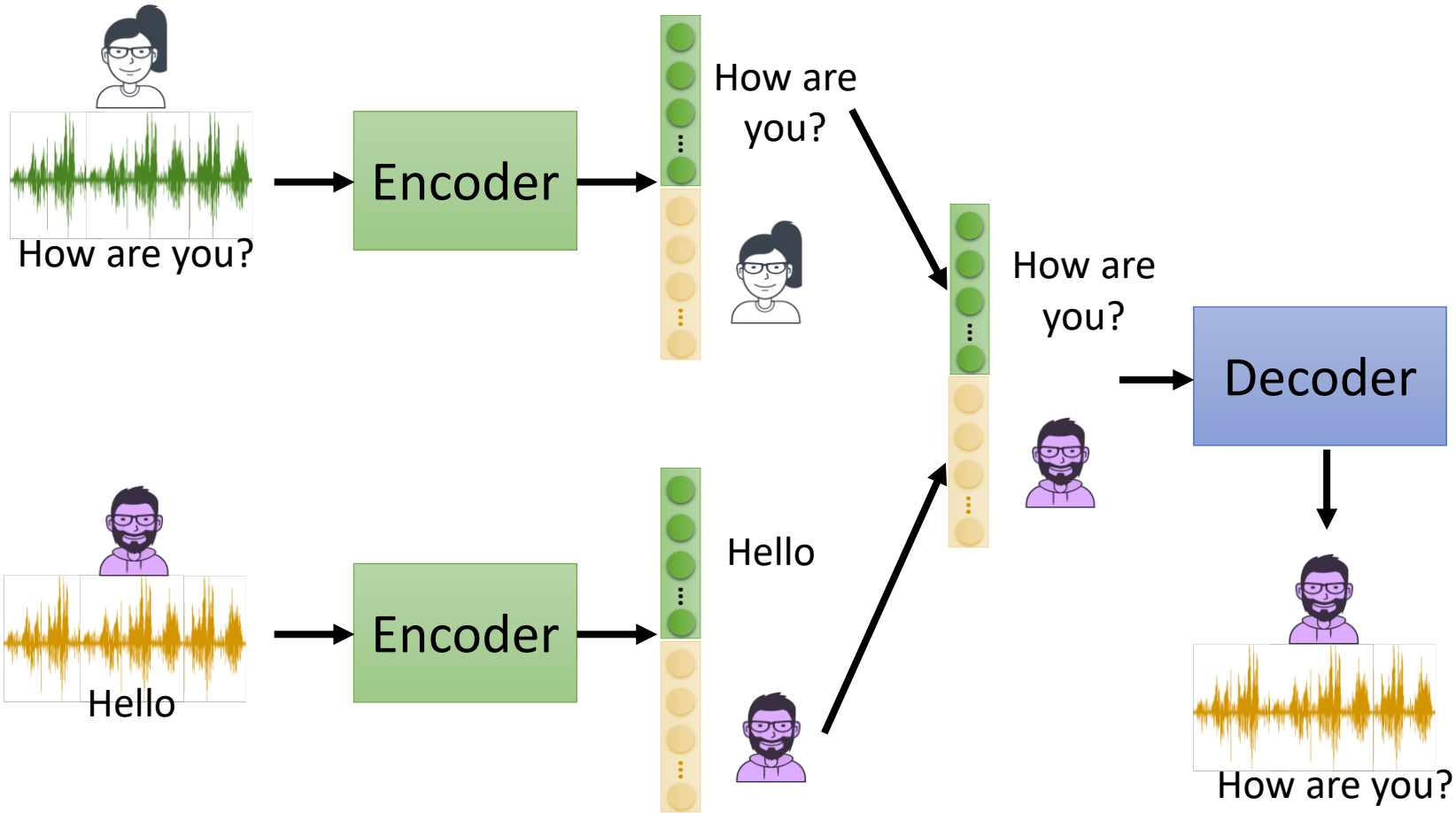
Feature Disentangle



Feature Disentangle - Voice Conversion

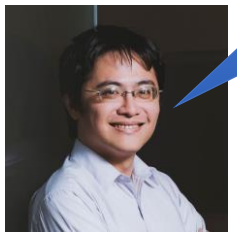


Feature Disentangle - Voice Conversion



Feature Disentangle - Voice Conversion

- The same sentence has different impact when it is said by different people.



Do you want to study a PhD?

Go away!

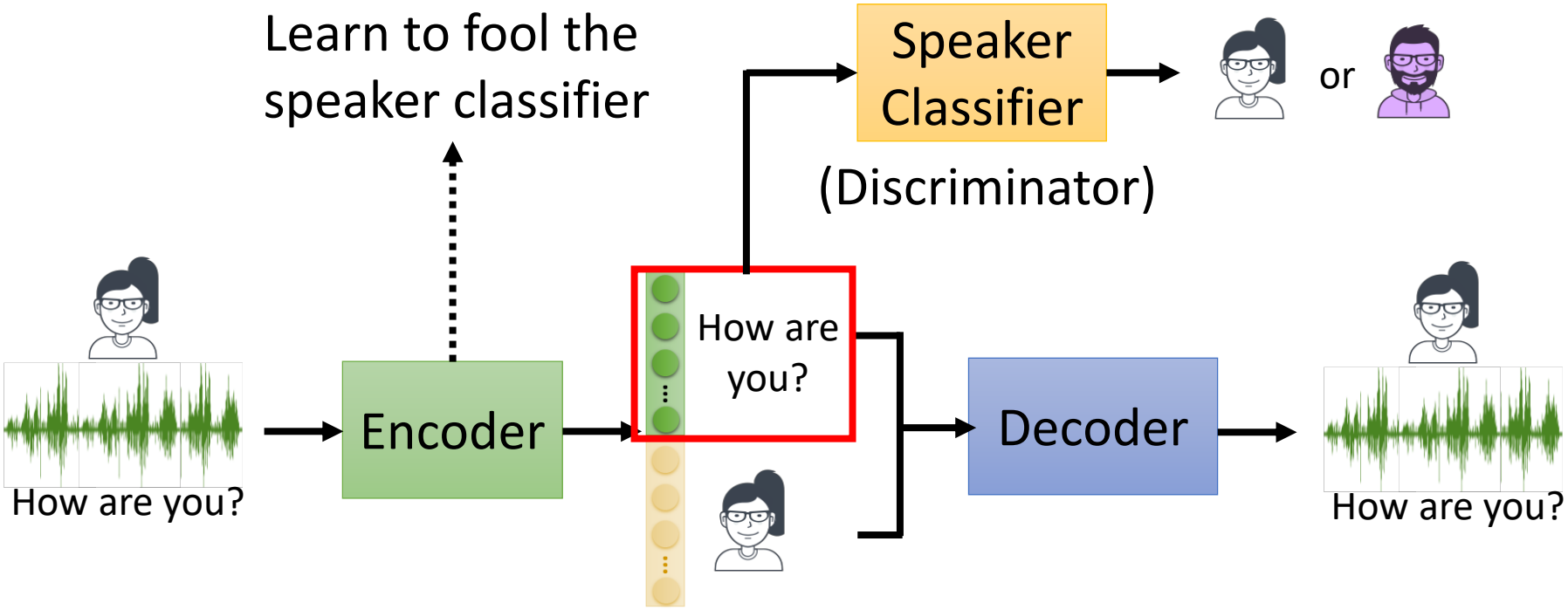
Student

新垣結衣
(Aragaki Yui)

Do you want to study a PhD?

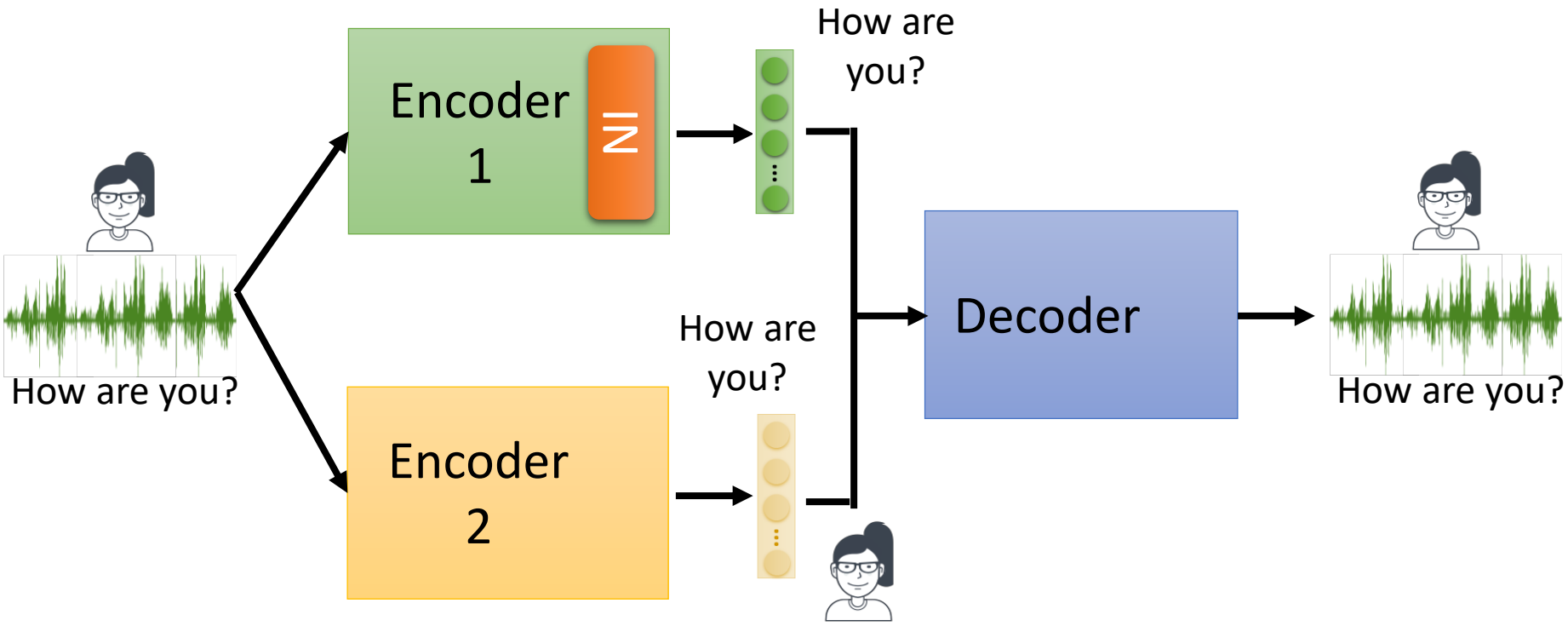
Student

Feature Disentangle - Adversarial Training



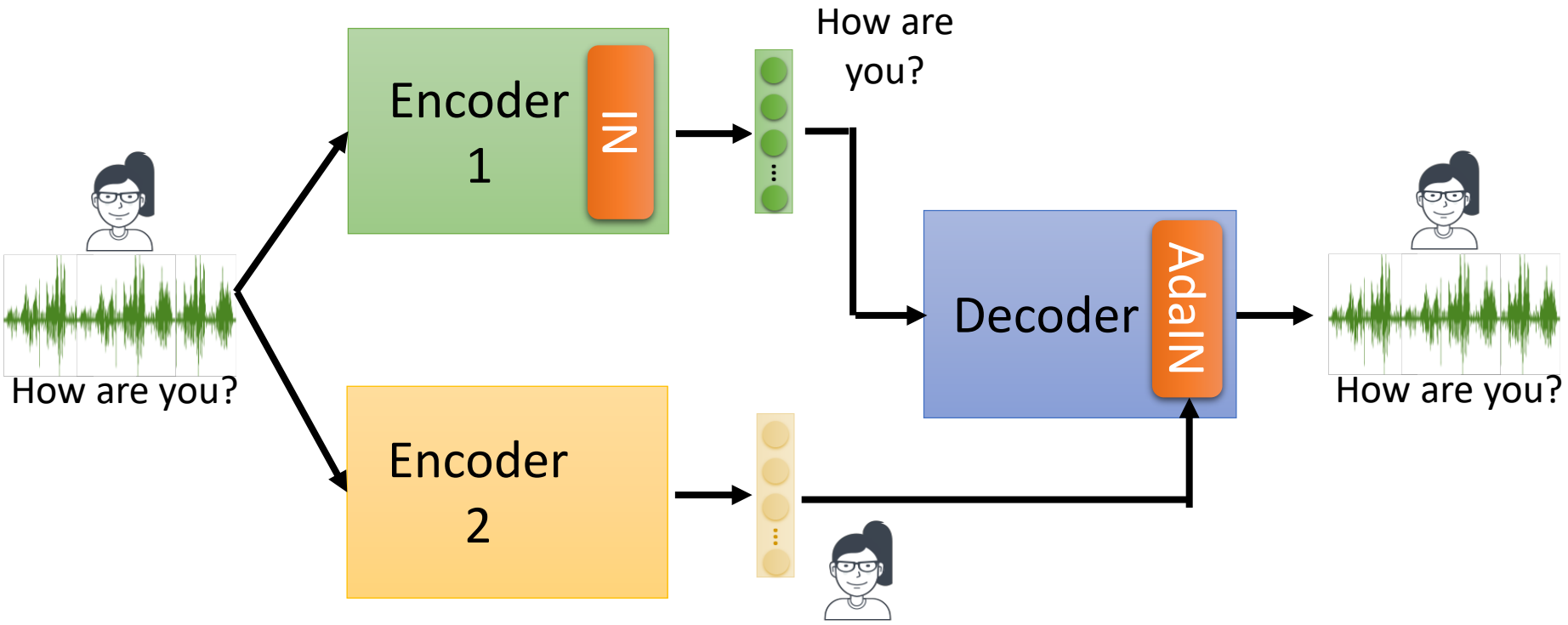
Speaker classifier and encoder are learned iteratively

Feature Disentangle
- Designed Network Architecture



IN = instance normalization (remove global information)

Feature Disentangle
- Designed Network Architecture



IN = instance normalization (remove global information)

AdaIN = adaptive instance normalization
(only influence global information)

Feature Disentangle - Adversarial Training

Target Speaker 

Source Speaker

(Never seen during training!)

Source to Target

 Me 

 Me 

 Me 

 Me 

Thanks Ju-chieh Chou for providing the results.

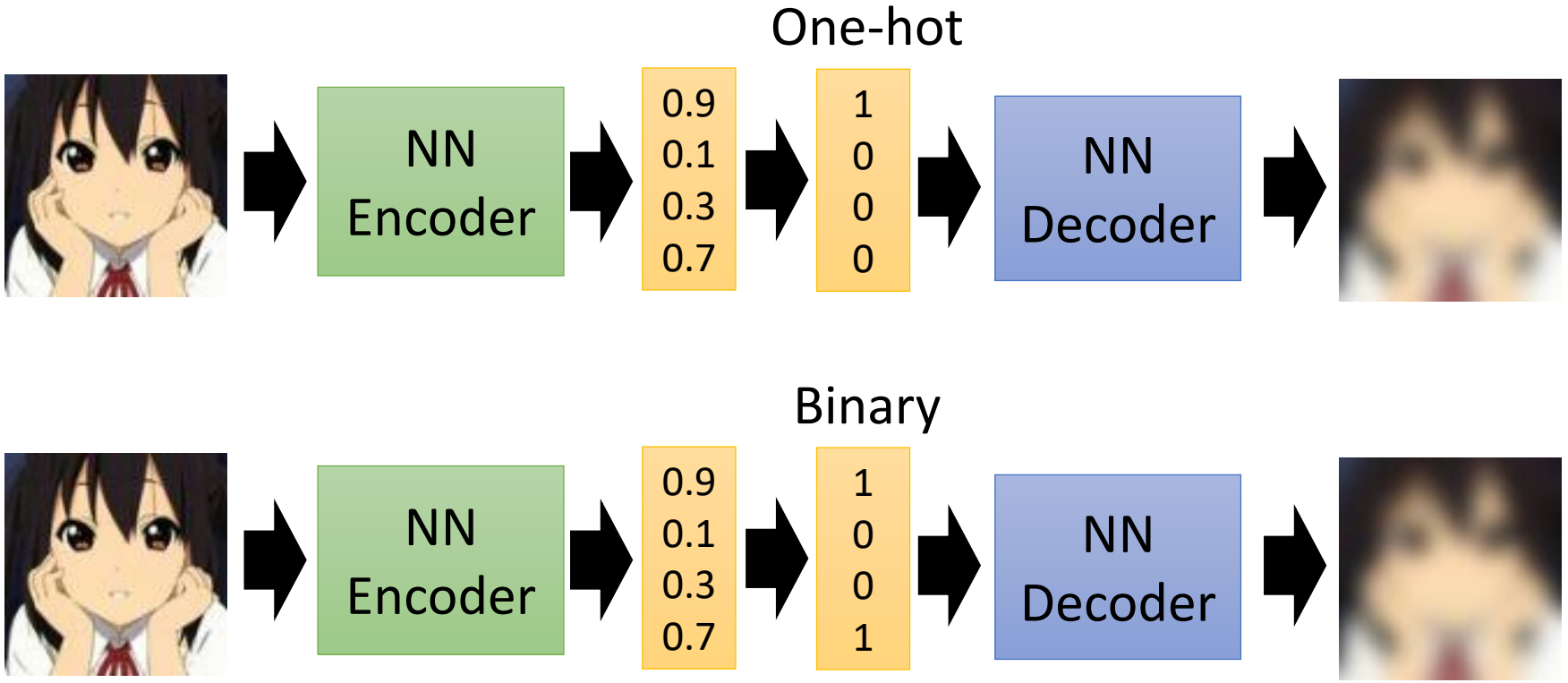
https://jjery2243542.github.io/voice_conversion_demo/

Discrete Representation

- Easier to interpret or clustering

non differentiable

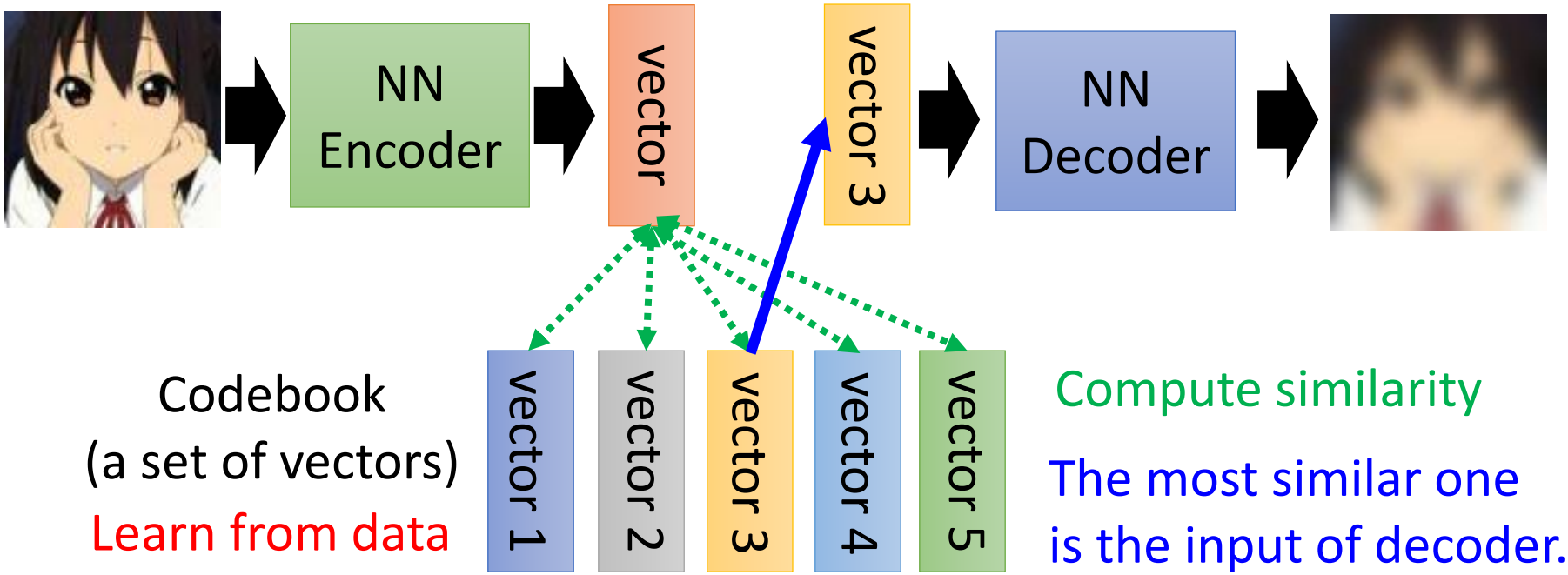
https://arxiv.org/pdf/1611.01144.pdf



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

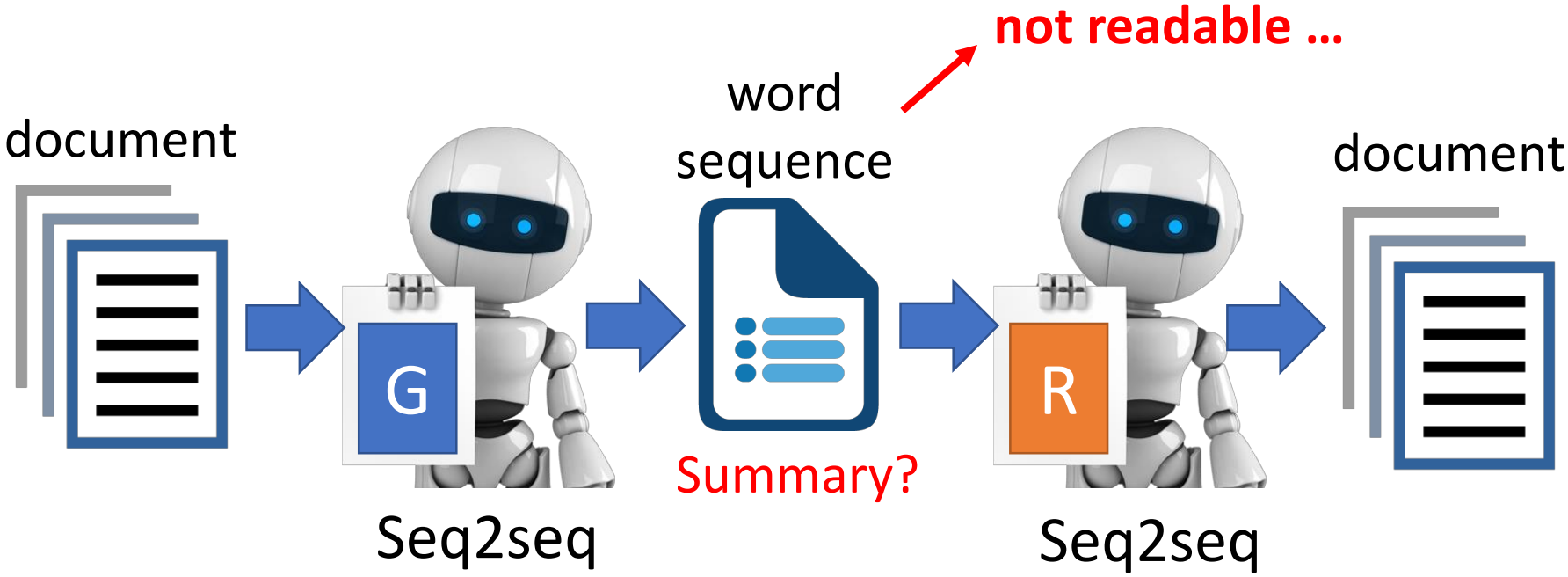
Sequence as Embedding

https://arxiv.org/abs/1810.02851

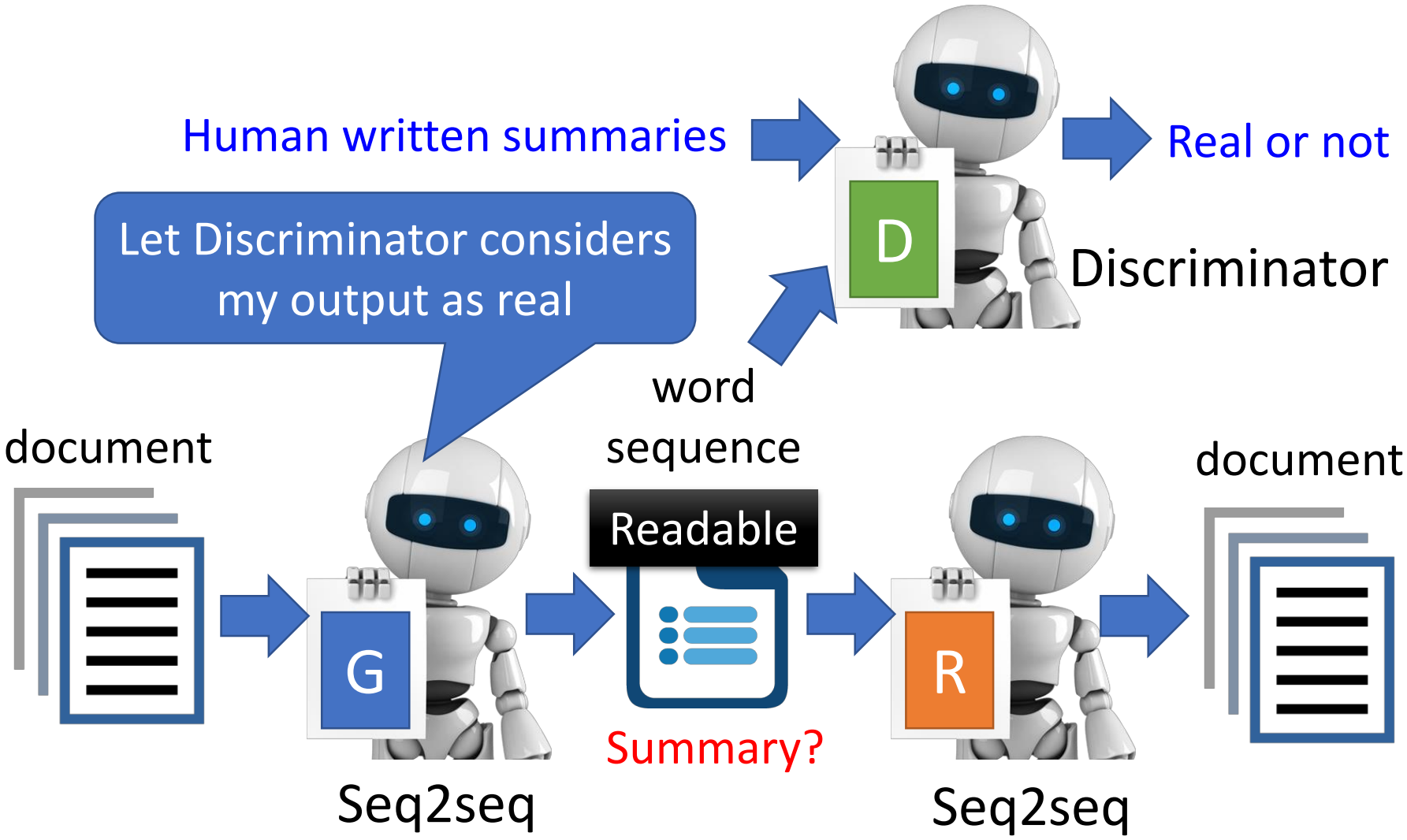
Only need a lot of documents to train the model

This is a ***seq2seq2seq auto-encoder***.

Using a sequence of words as latent representation.



Sequence as Embedding



感謝 王耀賢 同學提供實驗結果

Sequence as Embedding

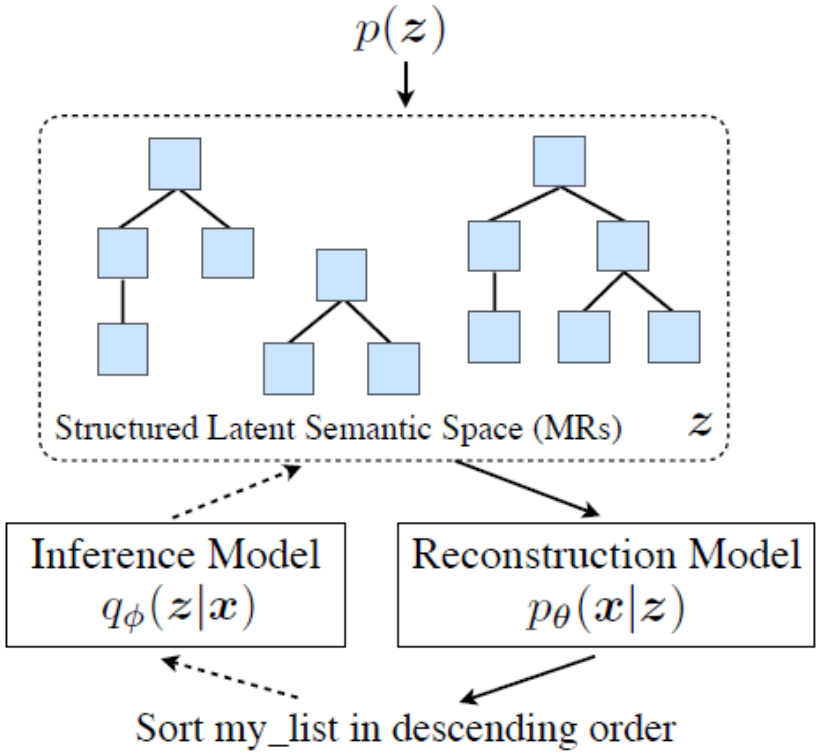
- ***Document***:澳大利亞今天與13個國家簽署了反興奮劑雙邊協議,旨在加強體育競賽之外的藥品檢查並共享研究成果 ……
- ***Summary***:
  - Human:澳大利亞與13國簽署反興奮劑協議
  - Unsupervised:澳大利亞加強體育競賽之外的藥品檢查
- ***Document***:中華民國奧林匹克委員會今天接到一九九二年冬季奧運會邀請函,由於主席張豐緒目前正在中南美洲進行友好訪問,因此尚未決定是否派隊赴賽 ……
- ***Summary***:
  - Human:一九九二年冬季奧運會函邀我參加
  - Unsupervised:奧委會接獲冬季奧運會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

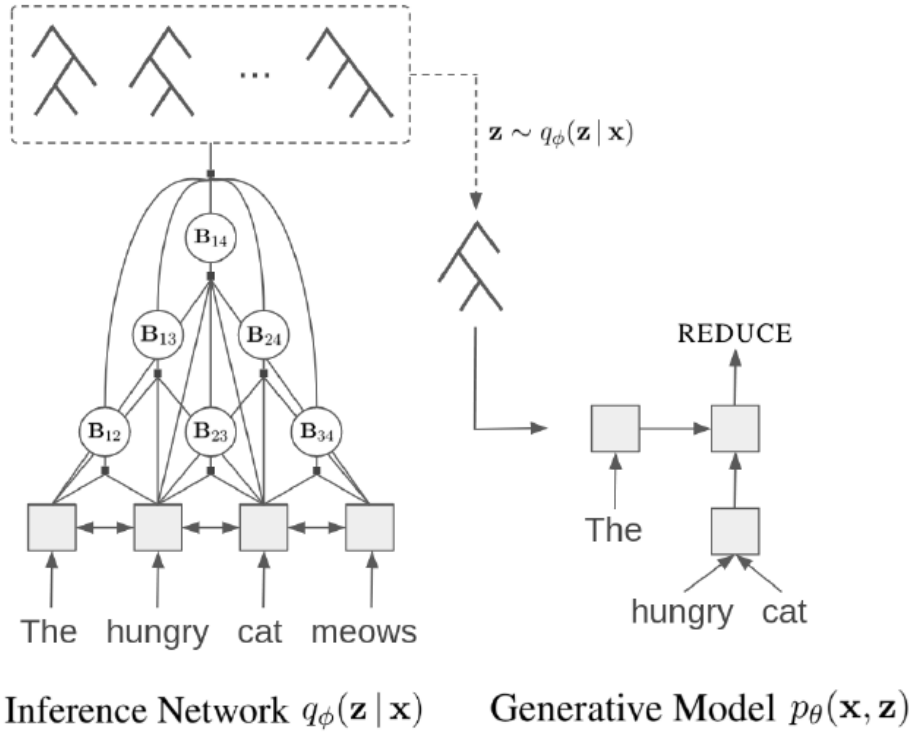
感謝 王耀賢 同學提供實驗結果

Sequence as Embedding

- ***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
- ***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

Tree as Embedding

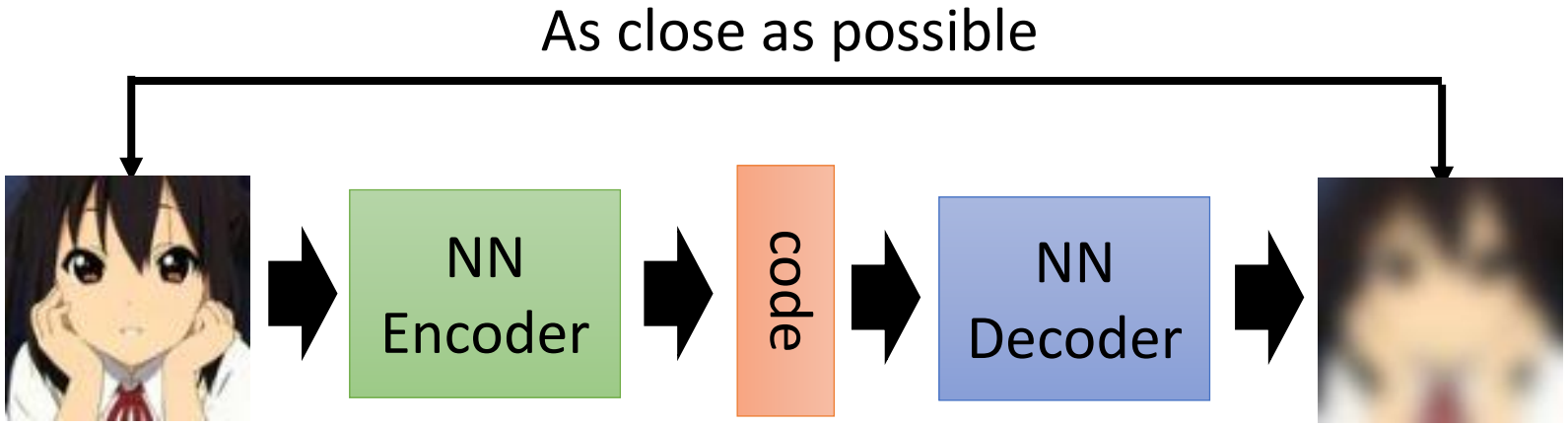




https://arxiv.org/abs/1806.07832

https://arxiv.org/abs/1904.03746

Concluding Remarks



- More than minimizing reconstruction error
  - Using Discriminator
  - Sequential Data
- More interpretable embedding
  - Feature Disentangle
  - Discrete and Structured