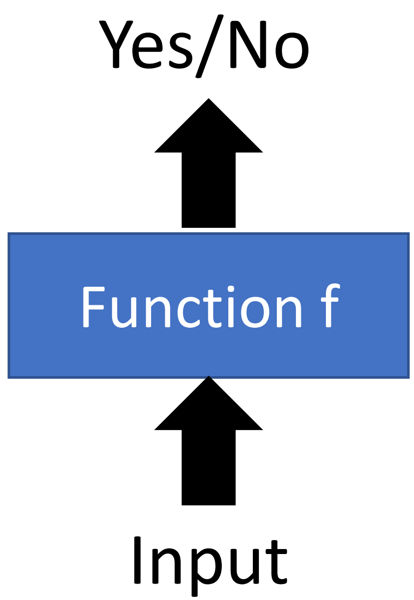
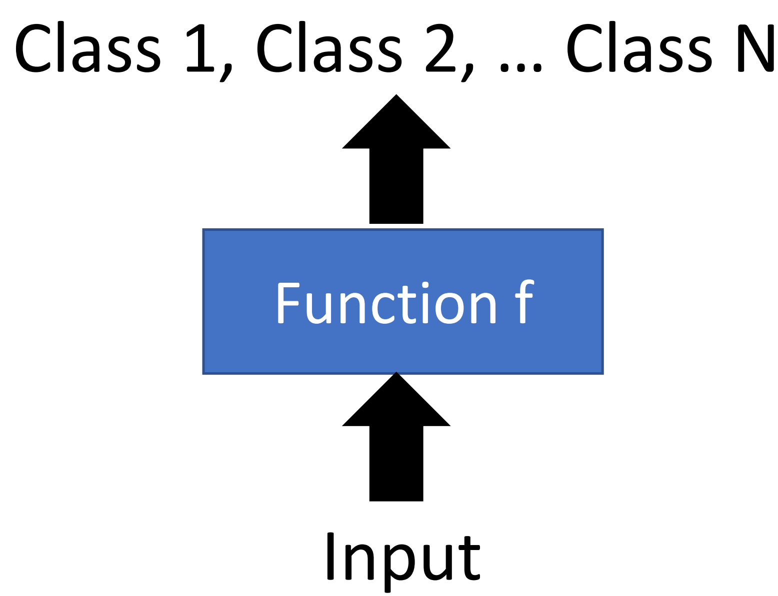
AI Alchemy:
Encoder, Generator, and put them together

李宏毅
Hung-yi Lee

Machine Learning
≈ Looking for a Function

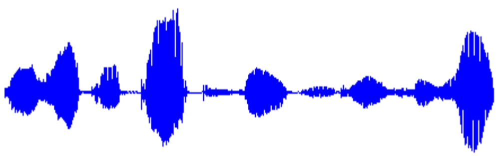
- Binary Classification (是非題)
- Multi-class Classification (選擇題)





Machine Learning
≈ Looking for a Function

- Structured input/output



→ f → "大家好，歡迎大家來修機器學習"

Speech Recognition

→ f → "近半選民挺脫歐公投 ……"

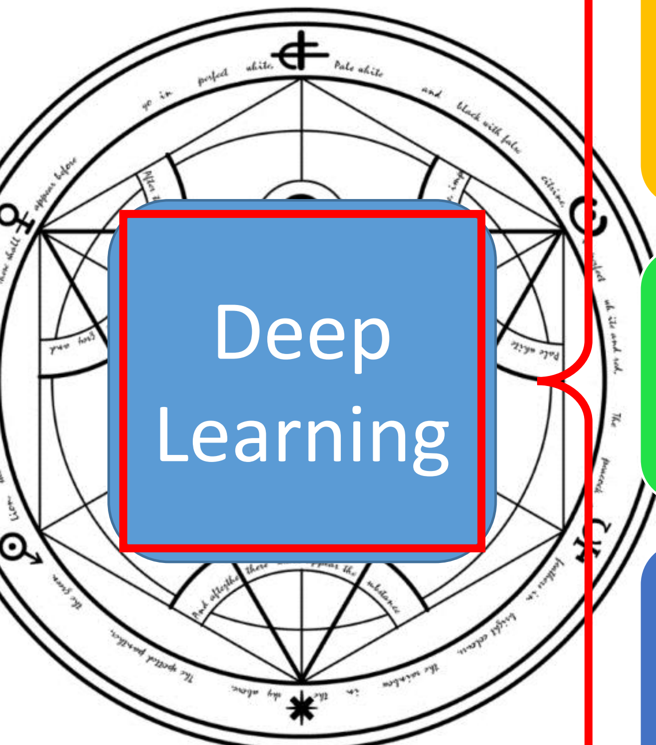
(title, summary)

Summarization

"girl with red hair and red eyes" → f →

http://yumekui.pixnet.net/album/photo/13915541-%E8%B6%85%E5%A4%A7%EF%BC%8E%E7%B7%B4%E6%88%90%E9%99%A3

Outline



Auto-encoder

理解事物的法則後，加以分解

Deep Generative Model

構築

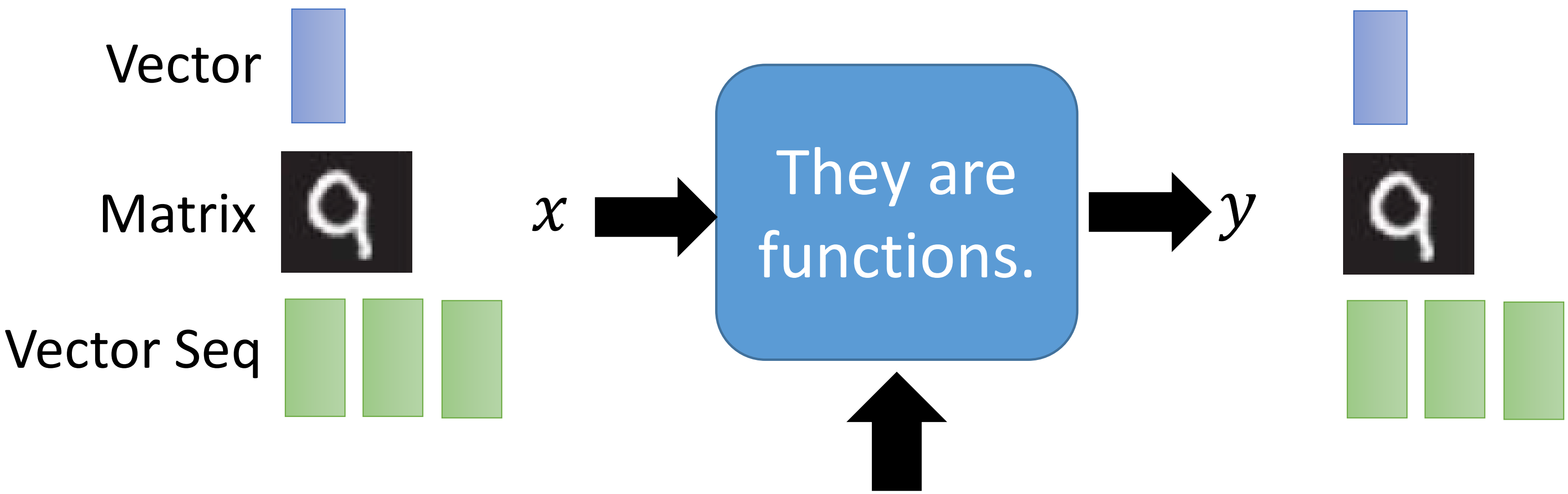
Conditional Generation

理解、分解 → 再構築

Deep Learning in One Slide

Many kinds of networks:

- Fully connected feedforward network (MLP)
- Convolutional neural network (CNN)
- Recurrent neural network (RNN)

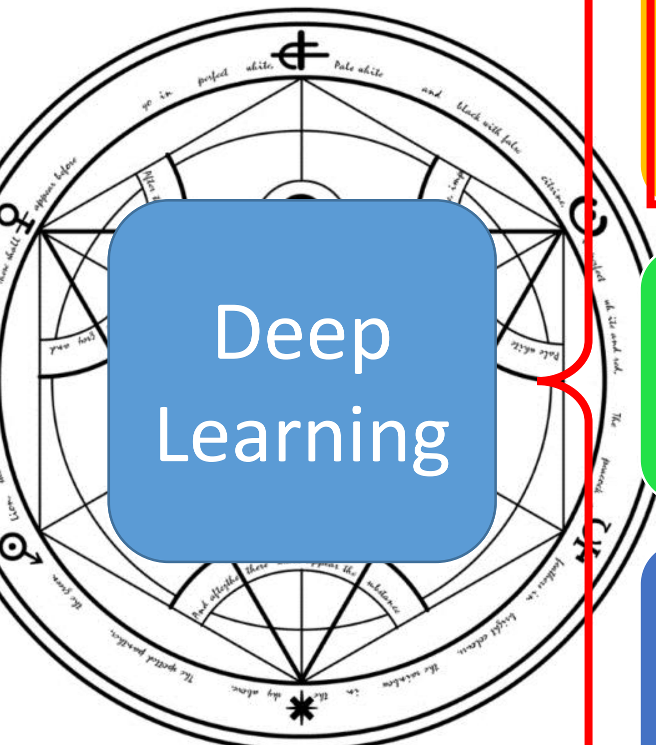


How to find the function?

Given the example inputs/outputs as training data: $\{(x_1,y_1),(x_2,y_2), \ldots\ldots, (x_{1000},y_{1000})\}$

http://yumekui.pixnet.net/album/photo/13915541-%E8%B6%85%E5%A4%A7%EF%BC%8E%E7%B7%B4%E6%88%90%E9%99%A3

Outline



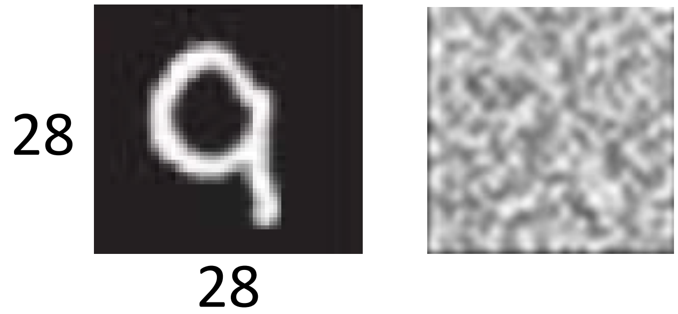
Deep Learning

Auto-encoder

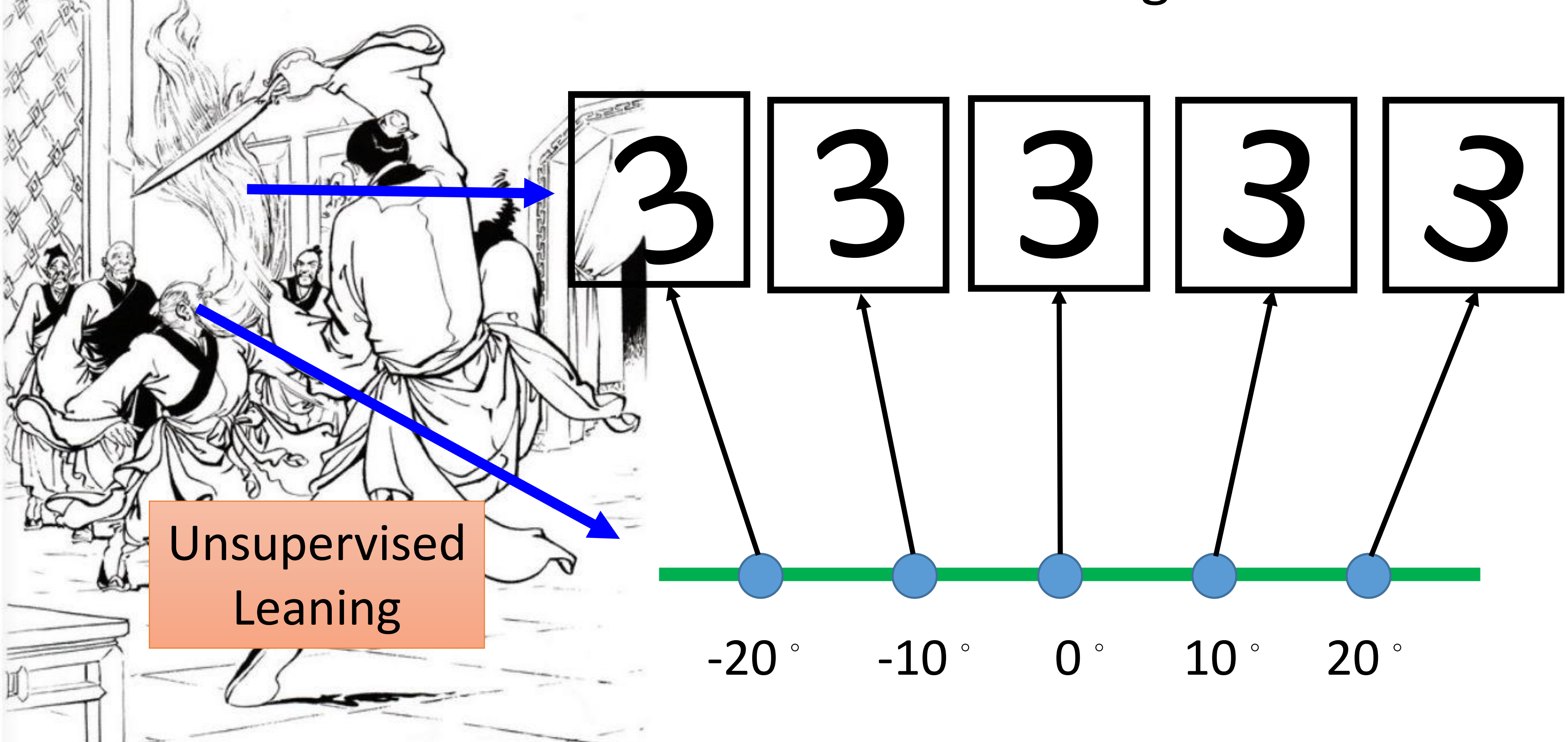
Deep Generative Model

Conditional Generation

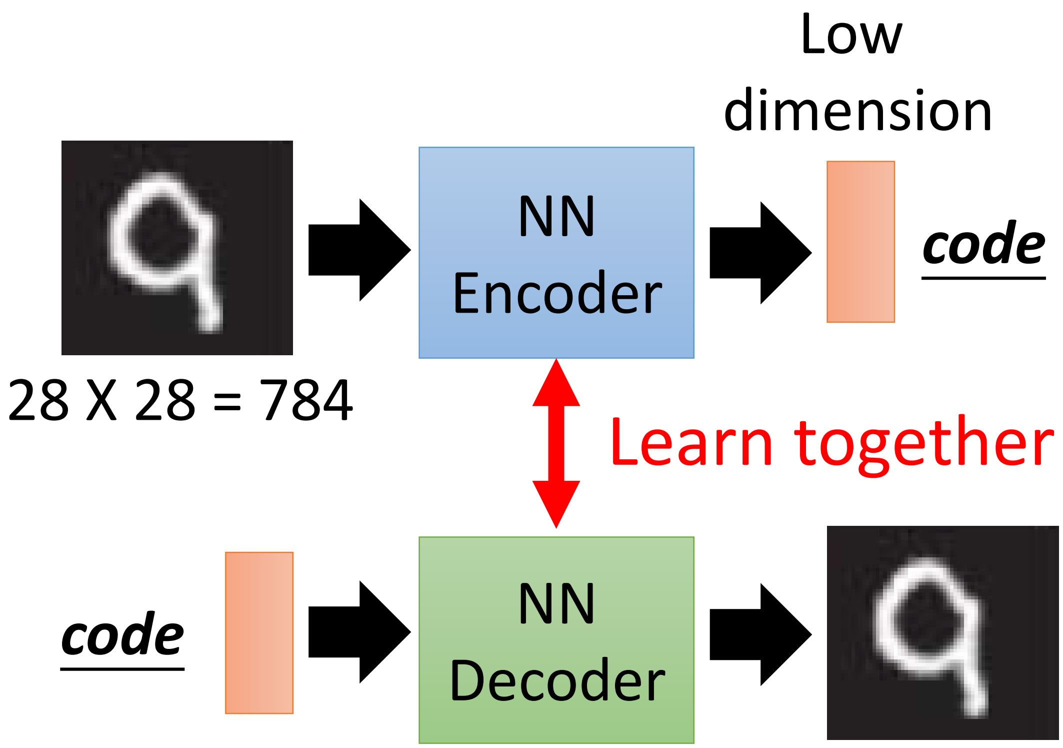
化繁為簡



- A digit can be represented as a 28 x 28 dim vector
  - Most 28 x 28 dim vectors are not digits



Auto-encoder



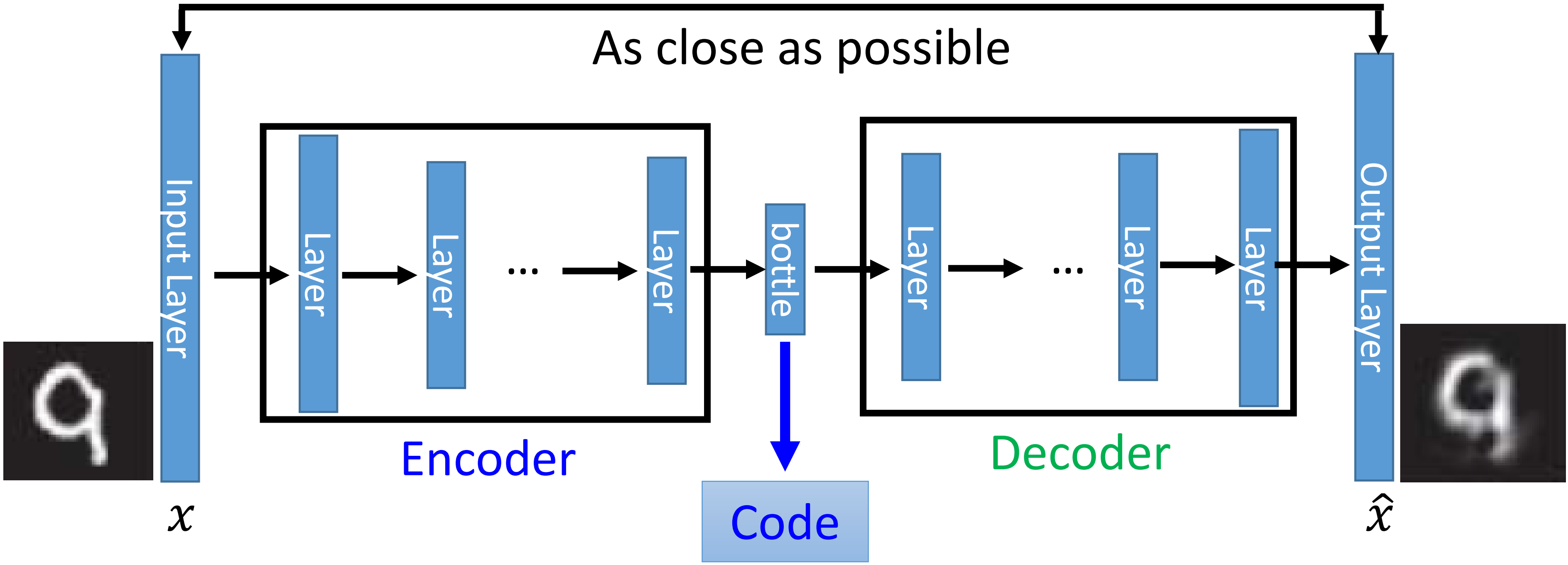
Compact representation of the input object

Can reconstruct the original object

Deep Auto-encoder

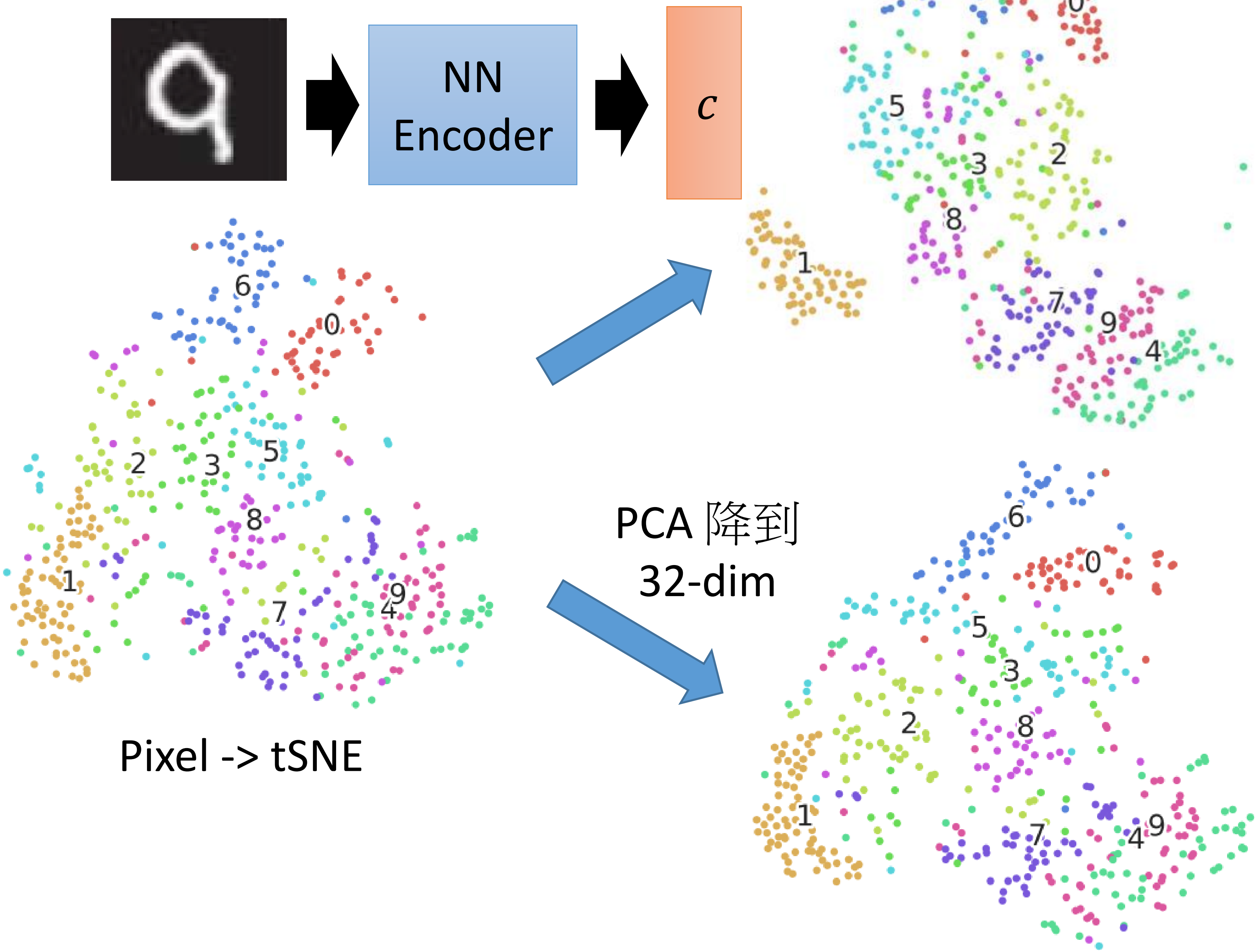
Unsupervised Leaning

- NN encoder + NN decoder = a deep network



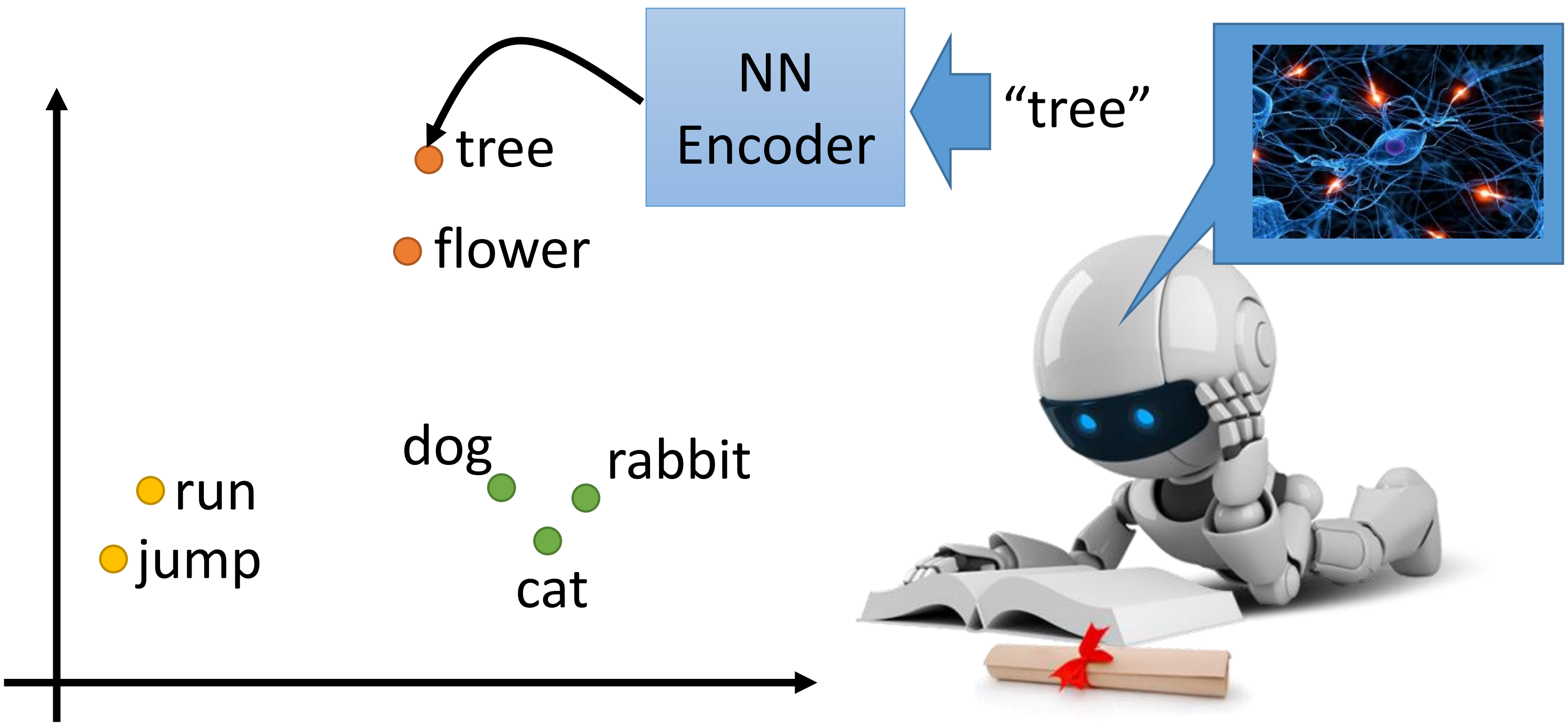
Reference: Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." *Science* 313.5786 (2006): 504-507

Deep Auto-encoder - Example



Word Embedding

- Machine learn the meaning of words from reading a lot of documents without supervision



Word Embedding

To learn more
https://www.youtube.com/watch?v=X7PH3NuYW0Q

- Machine learn the meaning of words from reading a lot of documents without supervision
- A word can be understood by its context

蔡英文、馬英九 are something very similar

You shall know a word by the company it keeps

馬英九 520宣誓就職

蔡英文 520宣誓就職

Word Embedding

- Characteristics

$$V(Germany) \approx V(Berlin) - V(Rome) + V(Italy)$$

$$V(hotter) - V(hot) \approx V(bigger) - V(big)$$

$$V(Rome) - V(Italy) \approx V(Berlin) - V(Germany)$$

$$V(king) - V(queen) \approx V(uncle) - V(aunt)$$

- Solving analogies

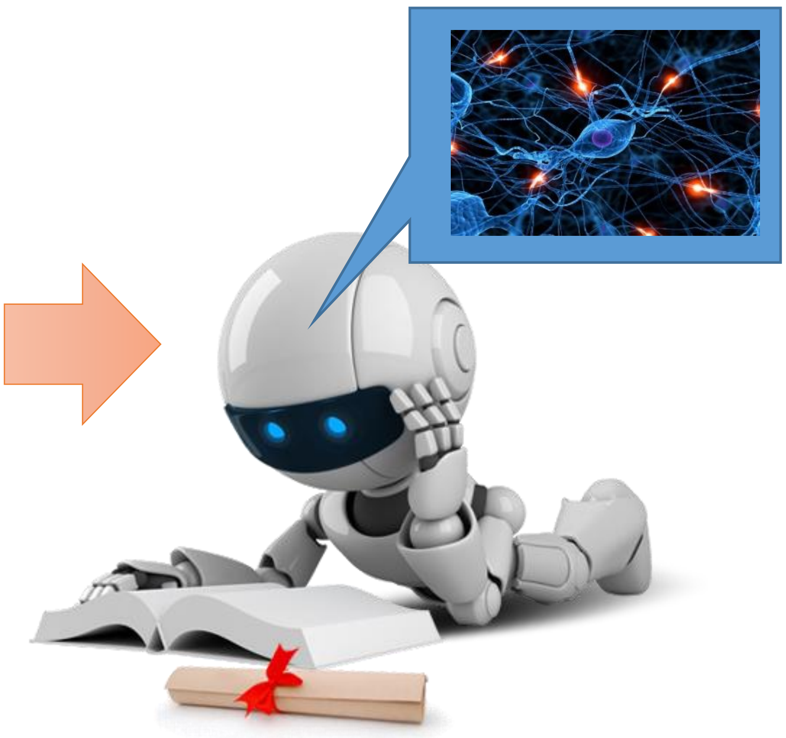
Rome : Italy = Berlin : ?

Compute $\underline{V(Berlin) - V(Rome) + V(Italy)}$

Find the word w with the closest V(w)

Word Embedding - Demo

- Machine learn the meaning of words from reading a lot of documents without supervision



Word Embedding - Demo

- Model used in demo is provided by 陳仰德
  - Part of the project done by 陳仰德、林資偉
  - TA: 劉元銘
  - Training data is from PTT (collected by 葉青峰)

Audio Word to Vector

Machine does not have any prior knowledge

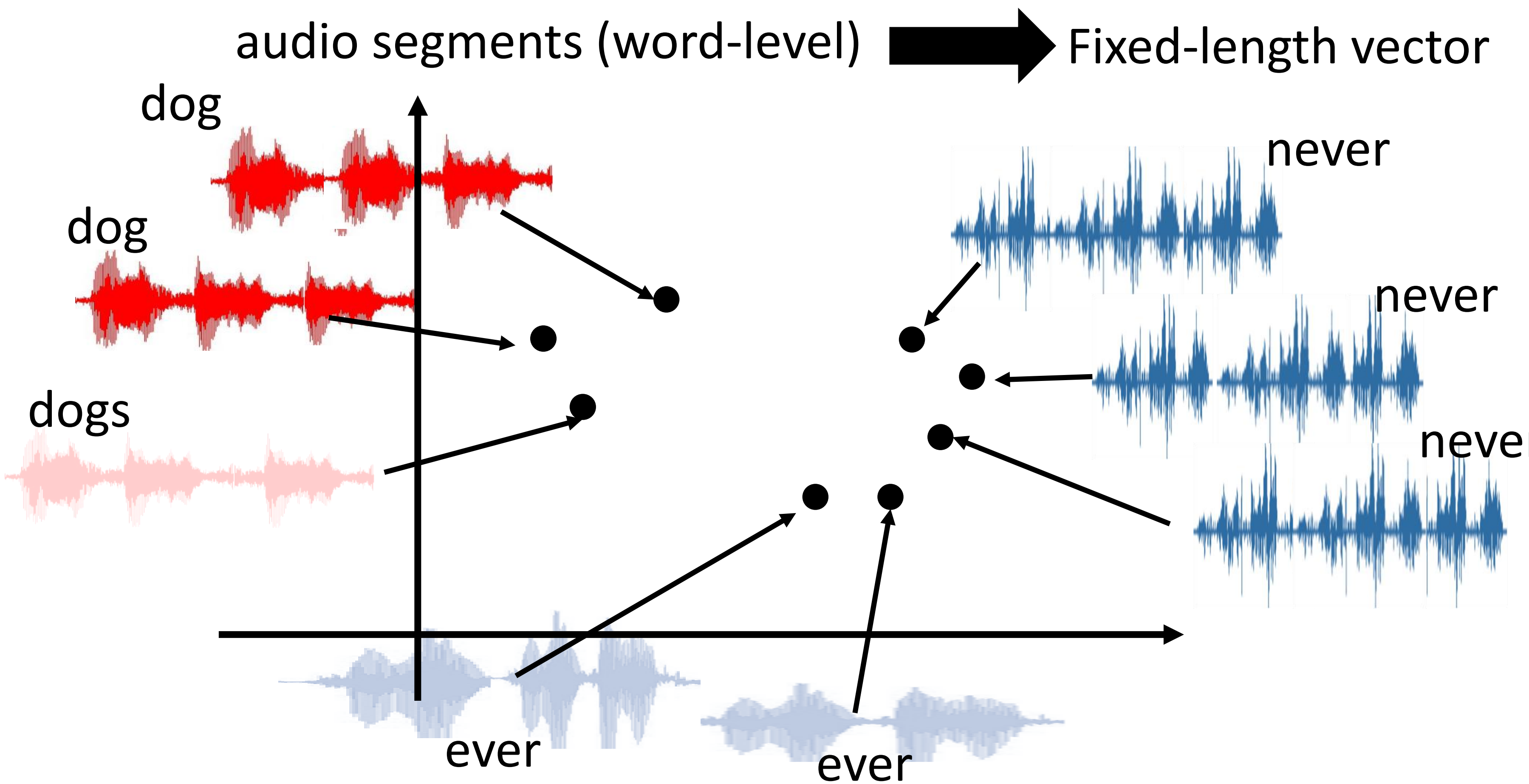
Machine listens to lots of audio book

Like an infant

[Chung, Interspeech 16)

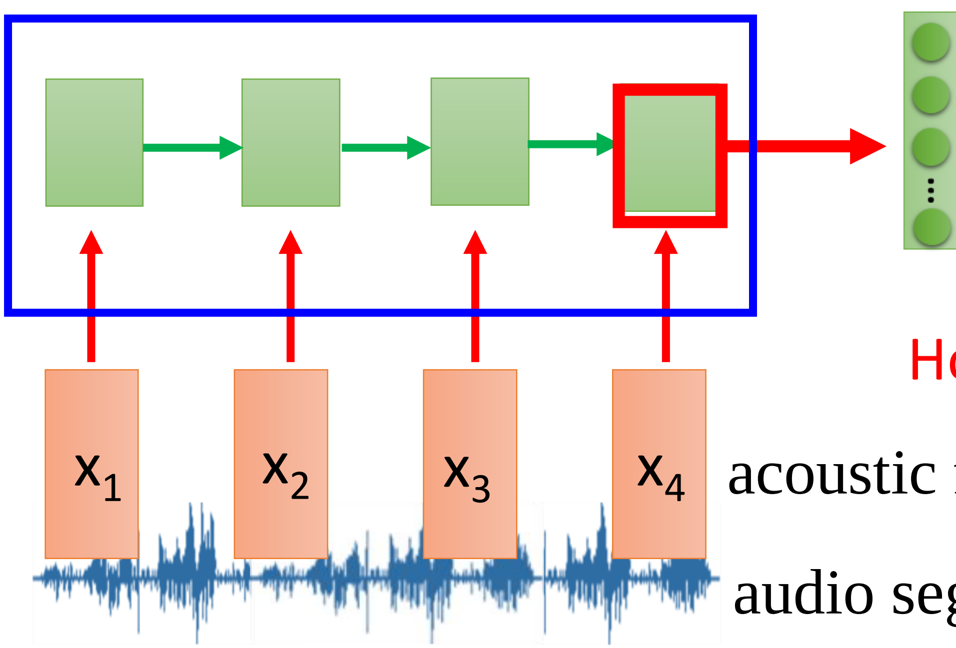
Audio Word to Vector

- Dimension reduction for a sequence with variable length



Sequence-to-sequence Auto-encoder

RNN Encoder



The vector we want

Can represent the whole audio segment

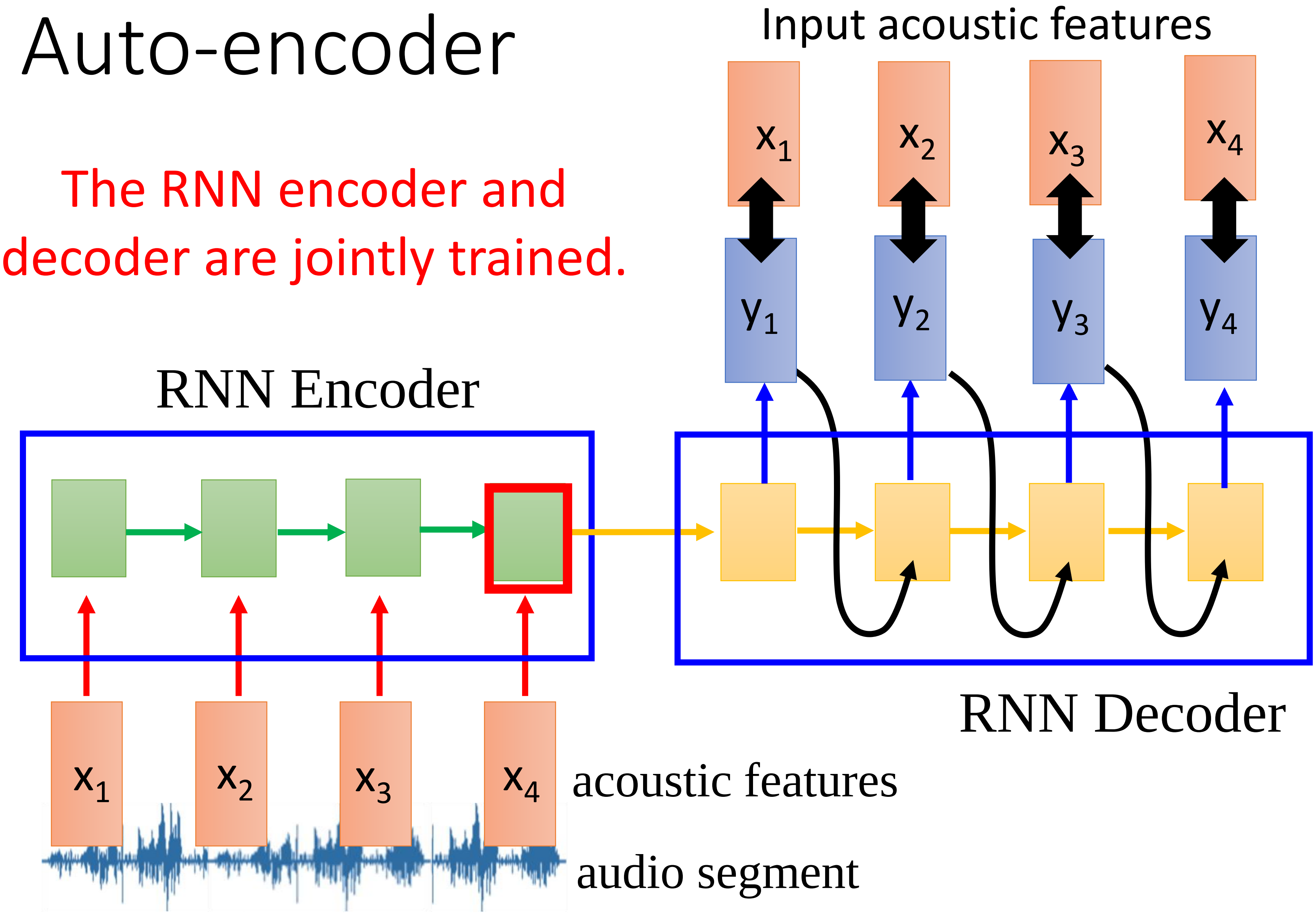
How to train RNN Encoder?

acoustic features

audio segment

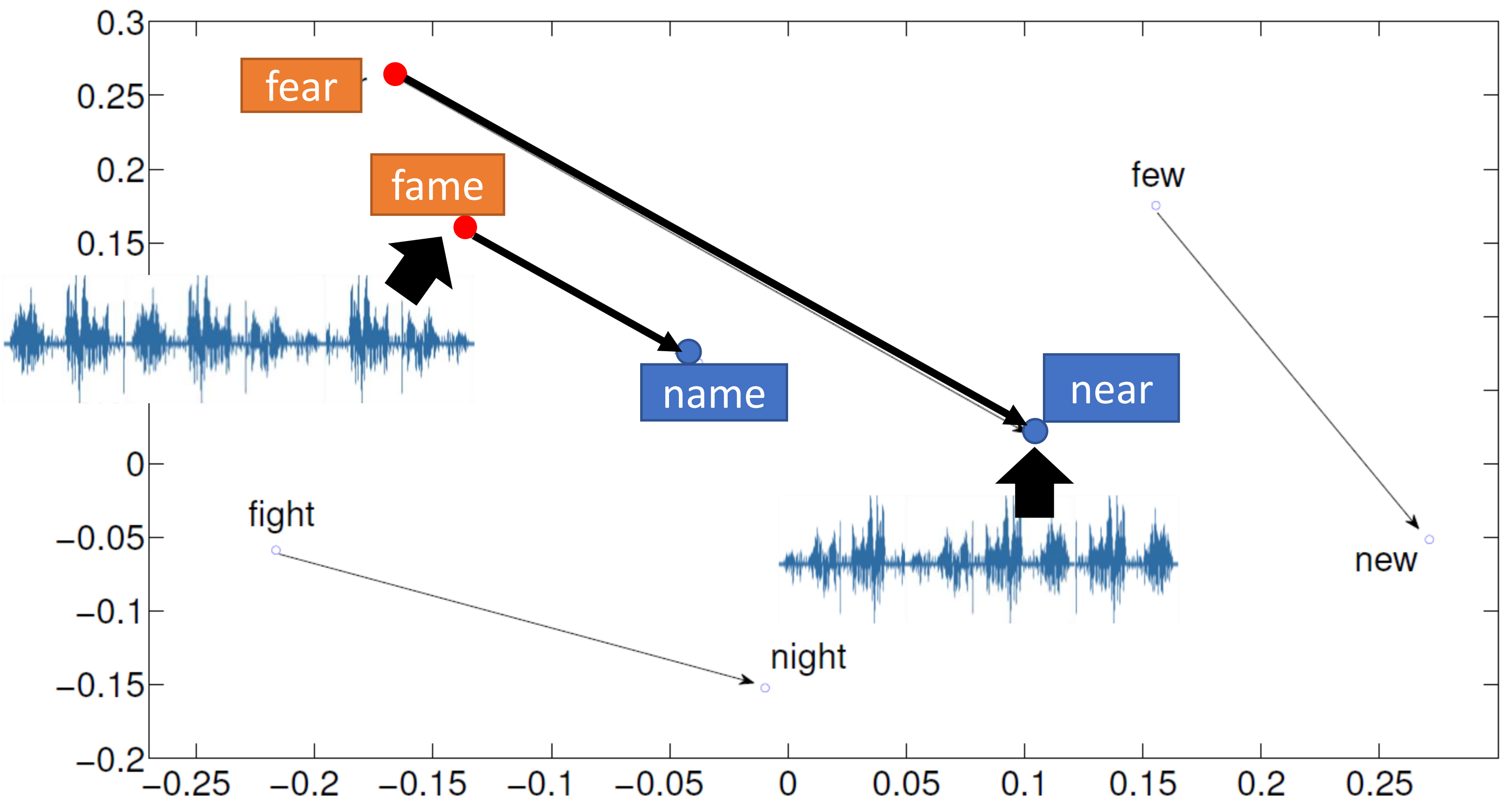
Sequence-to-sequence Auto-encoder

The RNN encoder and decoder are jointly trained.



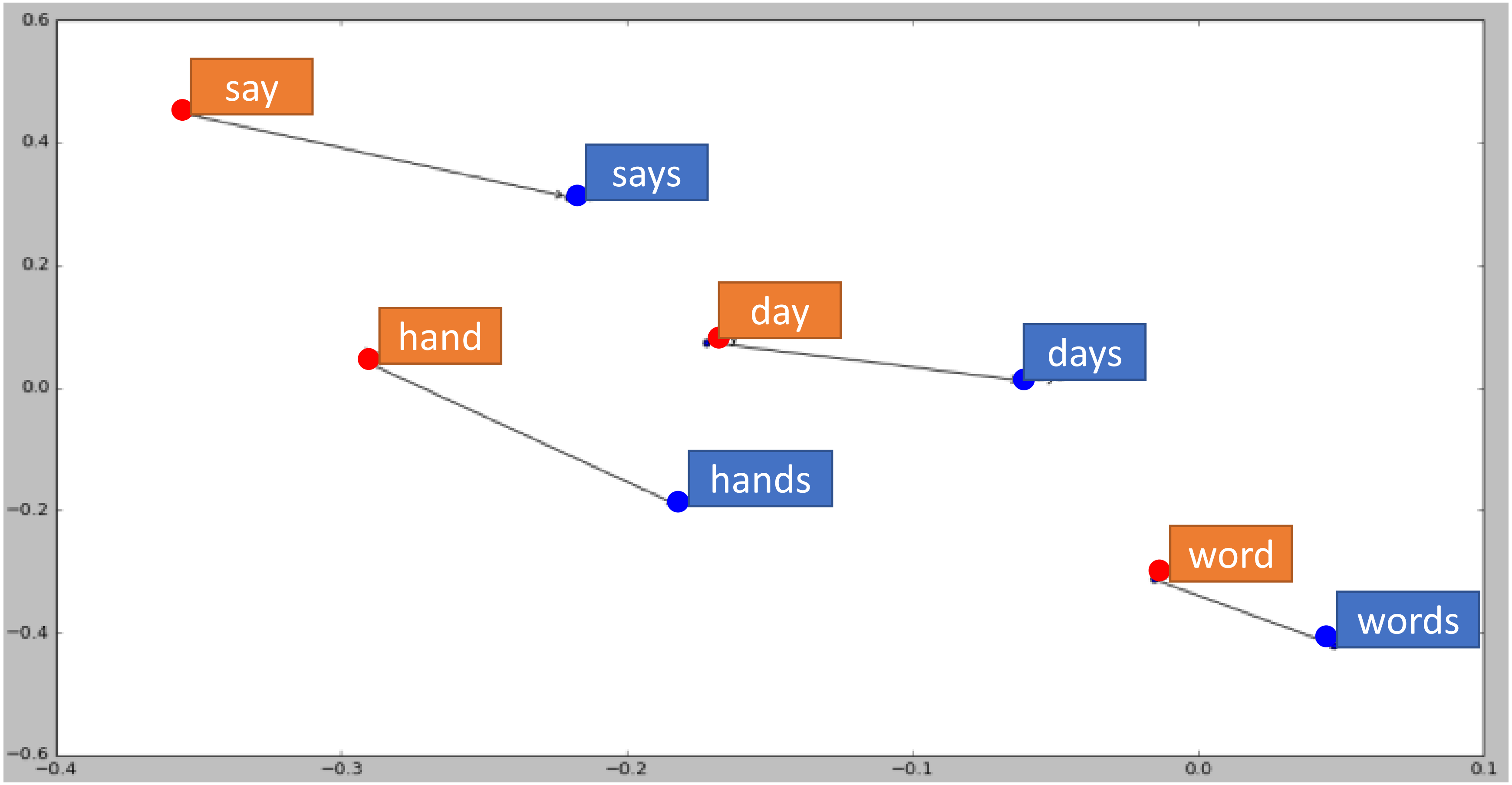
Sequence-to-sequence Auto-encoder

- Visualizing embedding vectors of the words

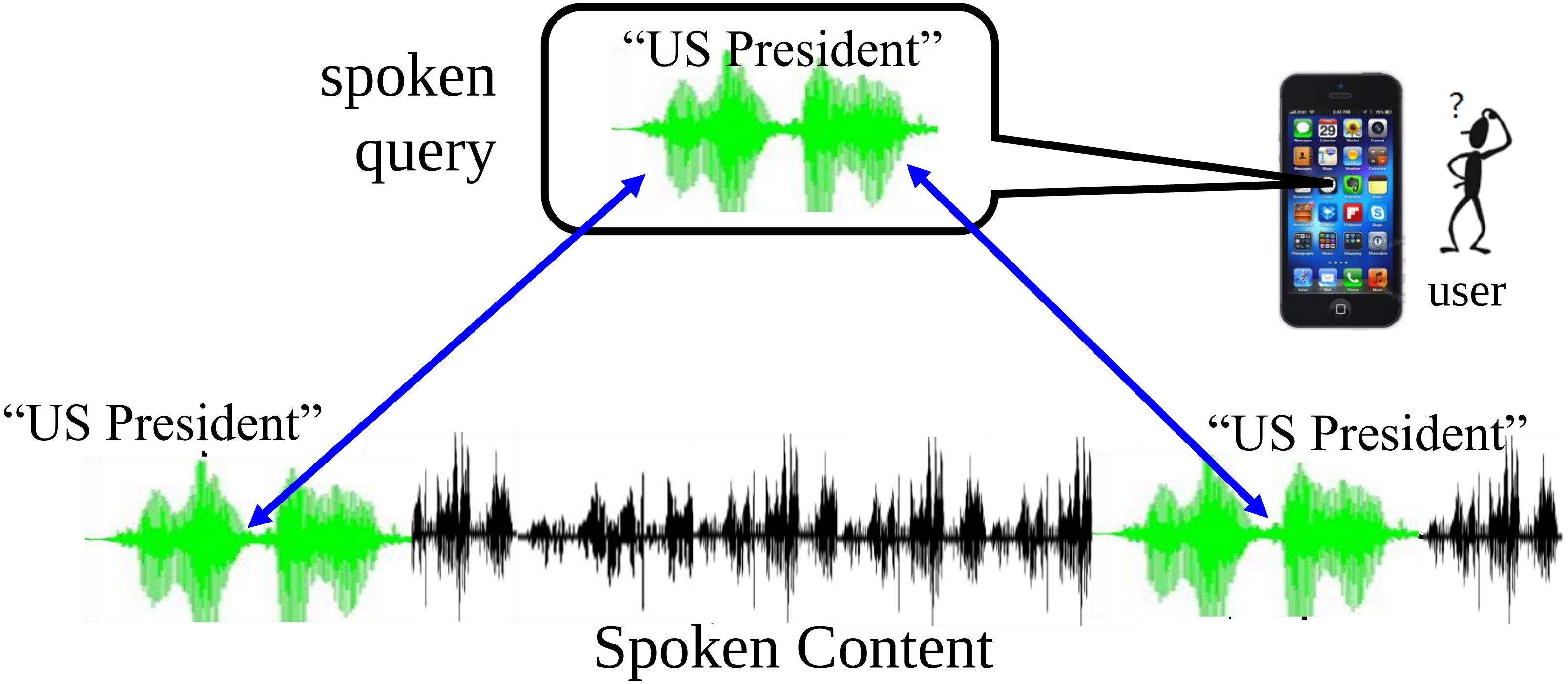


Sequence-to-sequence Auto-encoder

- Visualizing embedding vectors of the words



Audio Word to Vector –Application

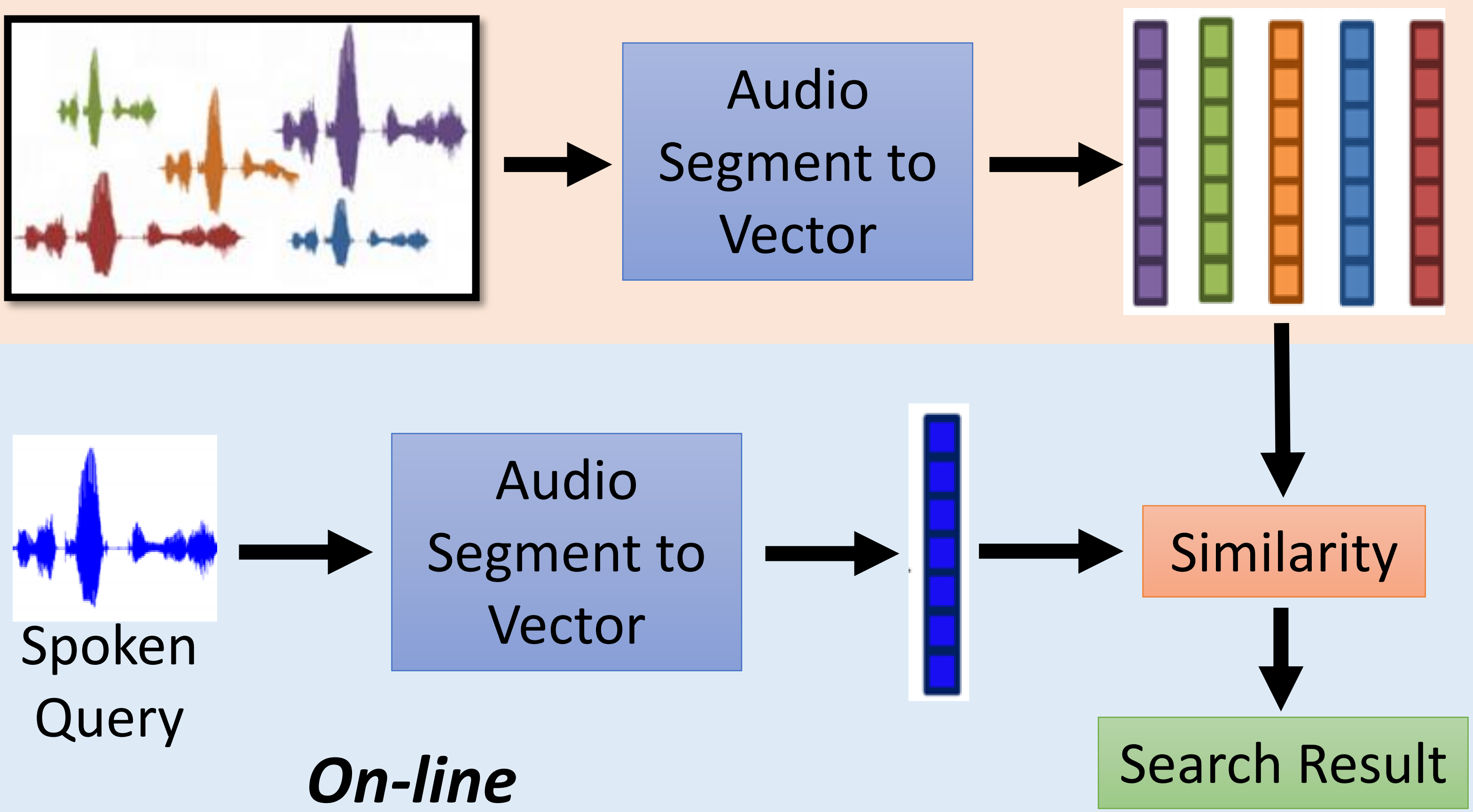


Compute similarity between spoken queries and audio files on acoustic level, and find the query term

Audio Word to Vector –Application

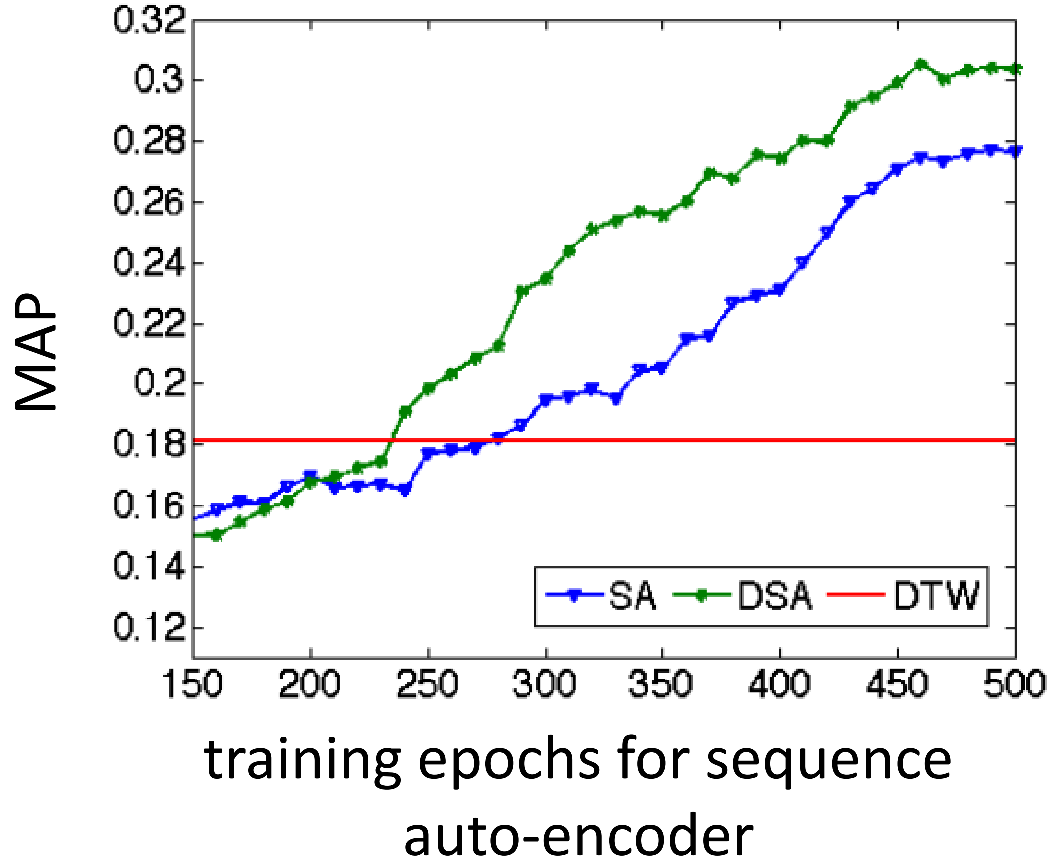
Audio archive divided into variable-length audio segments

Off-line



Audio Word to Vector –Application

- Query-by-Example Spoken Term Detection



SA: sequence auto-encoder

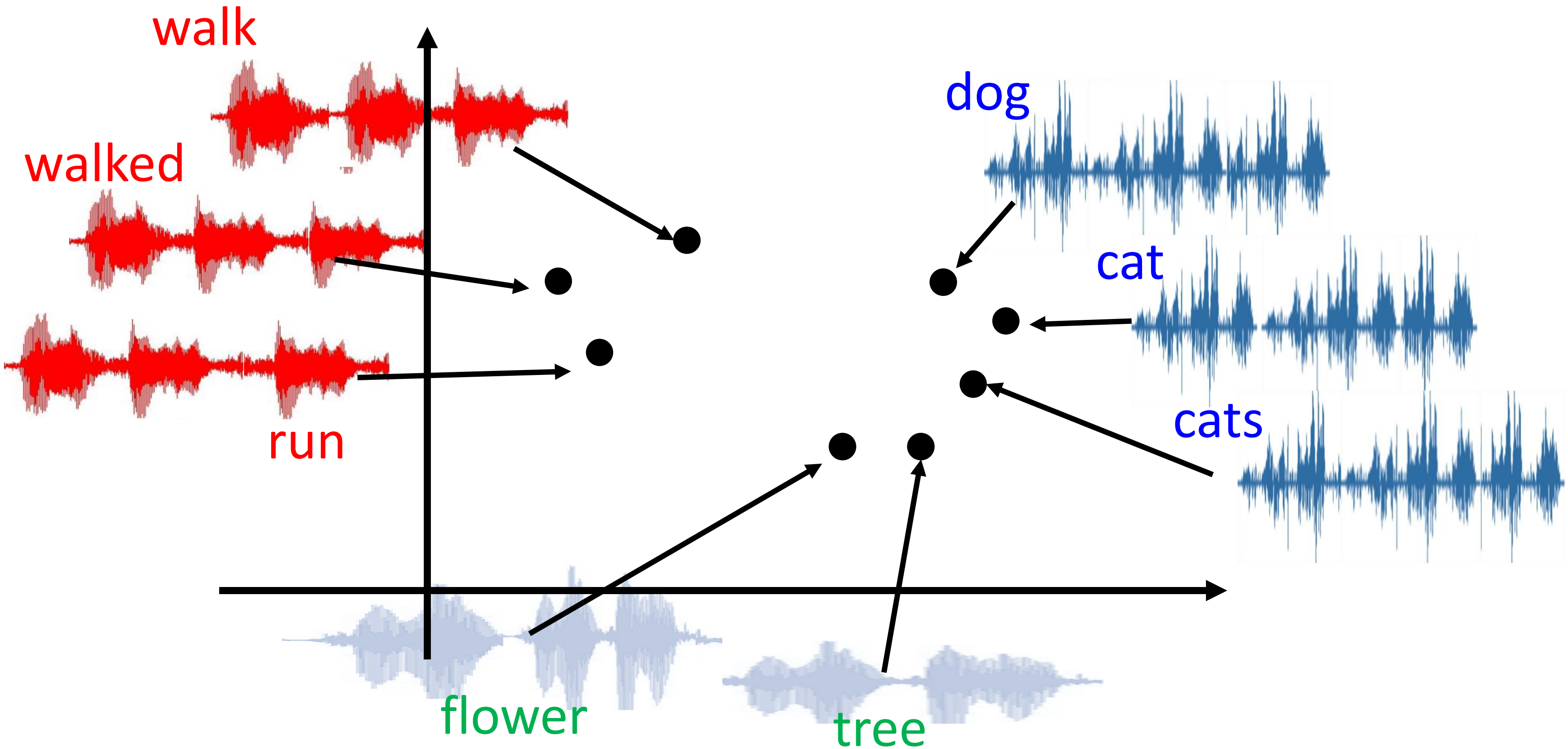
DSA: de-noising sequence auto-encoder

Input: clean speech + noise

output: clean speech

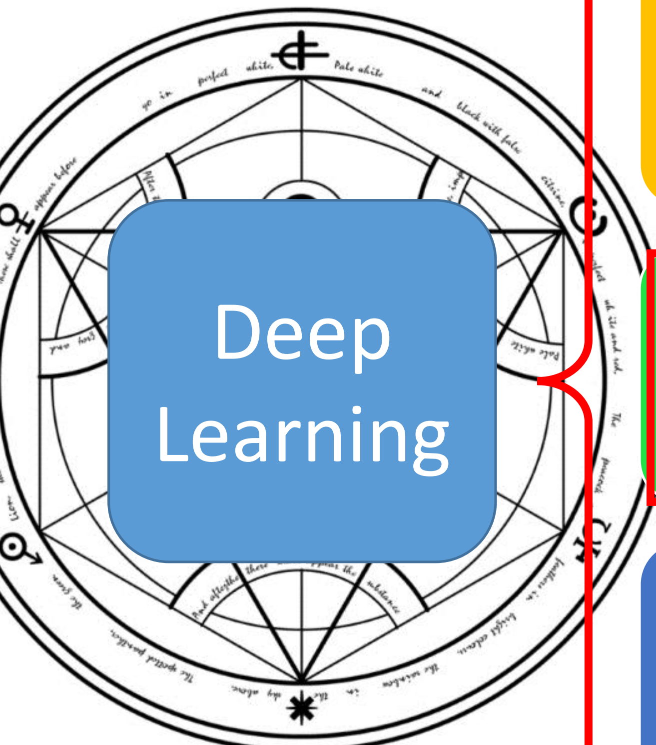
Next Step ……

- Can we include semantics?



http://yumekui.pixnet.net/album/photo/13915541-%E8%B6%85%E5%A4%A7%EF%BC%8E%E7%B7%B4%E6%88%90%E9%99%A3

Outline



Deep Learning

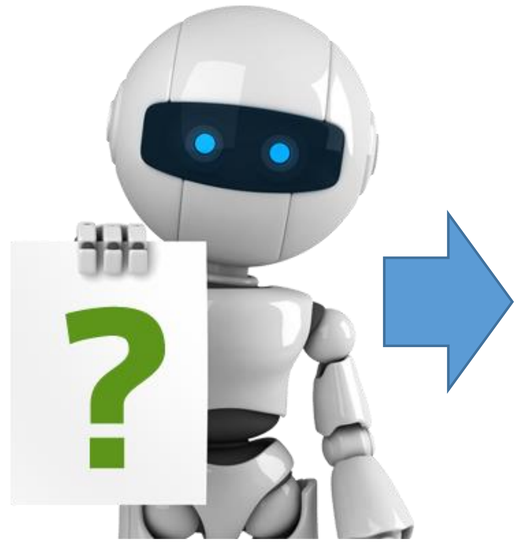
Auto-encoder

Deep Generative Model

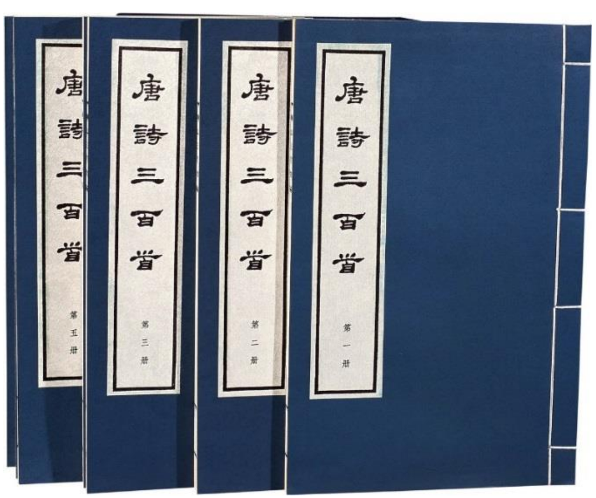
Conditional Generation

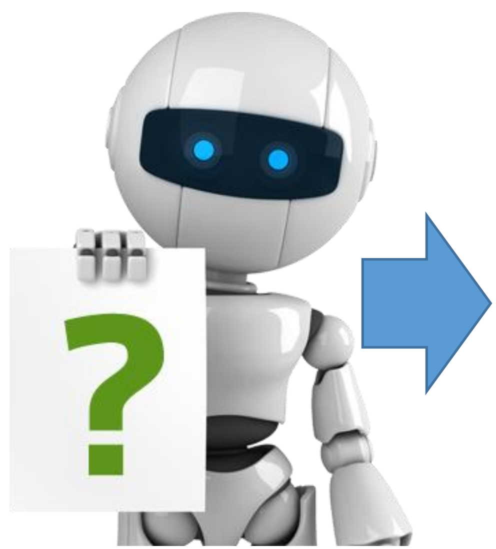
http://www.rb139.com/index.php?s=/Lot/44547

Creation



Drawing?

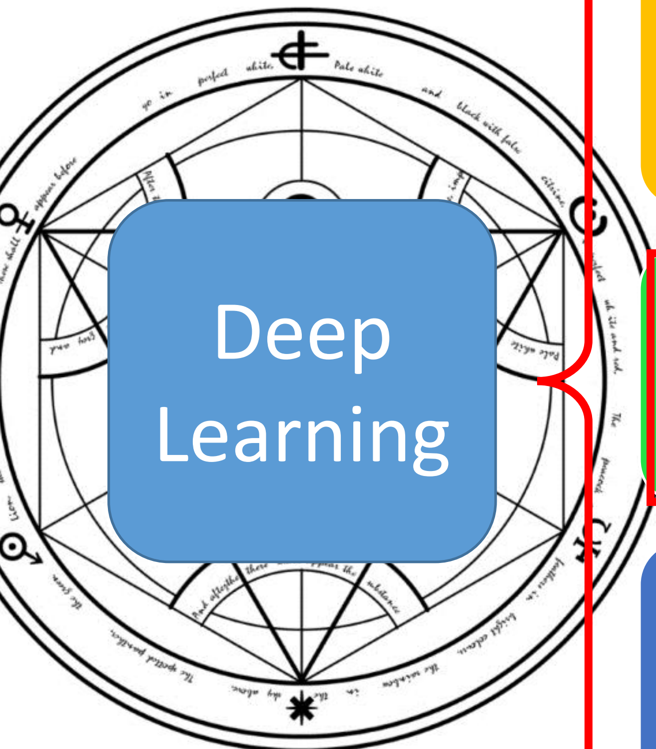




Writing Poems?

http://yumekui.pixnet.net/album/photo/13915541-%E8%B6%85%E5%A4%A7%EF%BC%8E%E7%B7%B4%E6%88%90%E9%99%A3

Outline



Deep Learning

Auto-encoder

Deep Generative Model

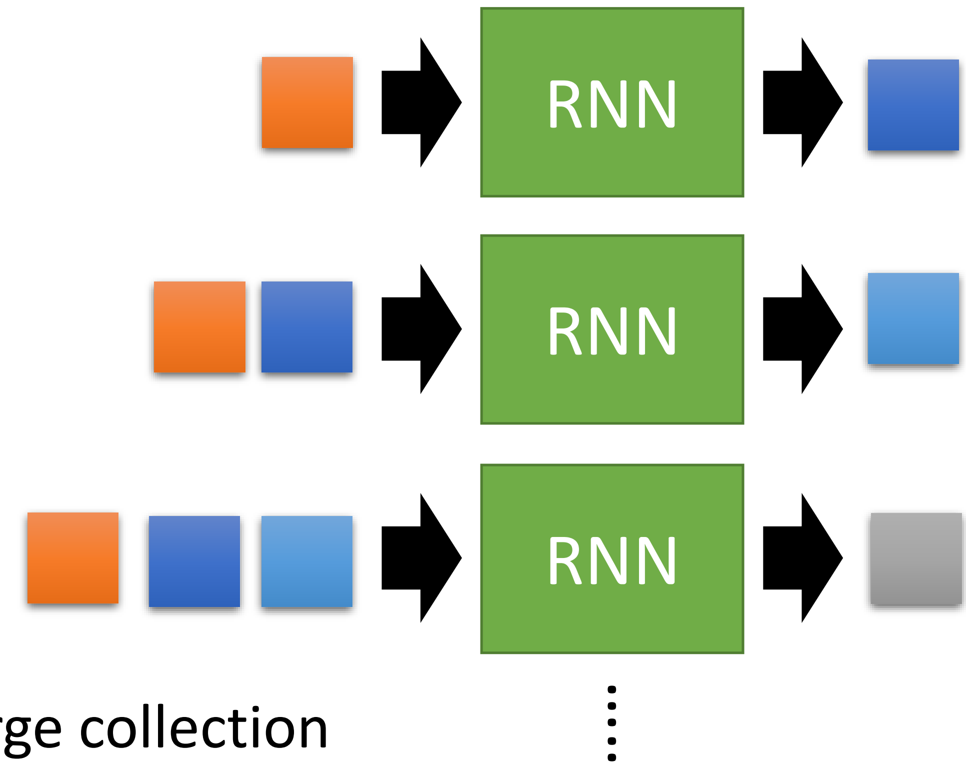
Component-wised, VAE, GAN

Conditional Generation

Component-by-component

- Images are composed of pixels
- To create an image, generating a pixel each time

E.g. 3 x 3 images



Can be trained just with a large collection of images without any annotation

Component-by-component
- 寶可夢鍊成

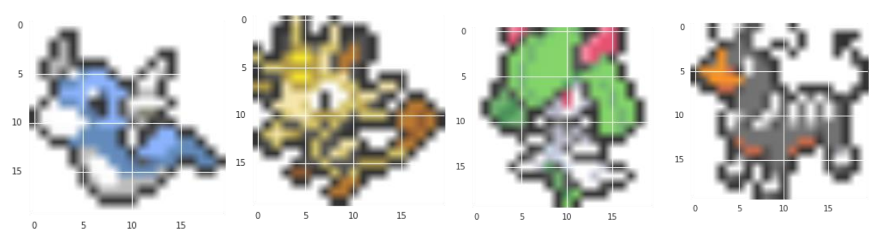
- Small images of 792 Pokémon's
  - Can machine learn to create new Pokémons?

<u>Don't catch them! Create them!</u>

- Source of image: http://bulbapedia.bulbagarden.net/wiki/List_of_Pok%C3%A9mon_by_base_stats_(Generation_VI)

Original image is 40 x 40

Making them into 20 x 20



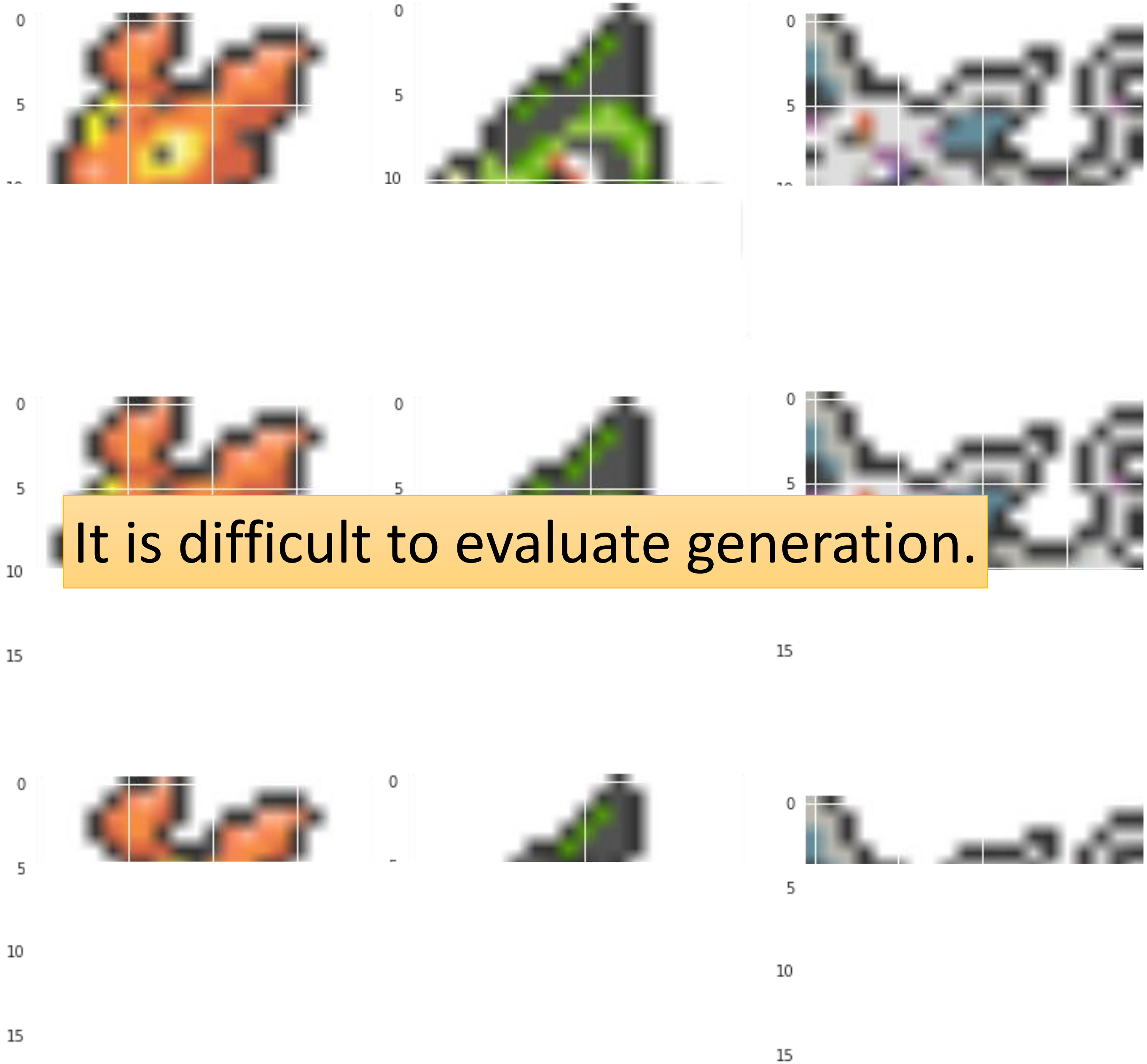
- Using 1-layer RNN with 512 LSTM cells

Real Pokémon

Never seen by machine!

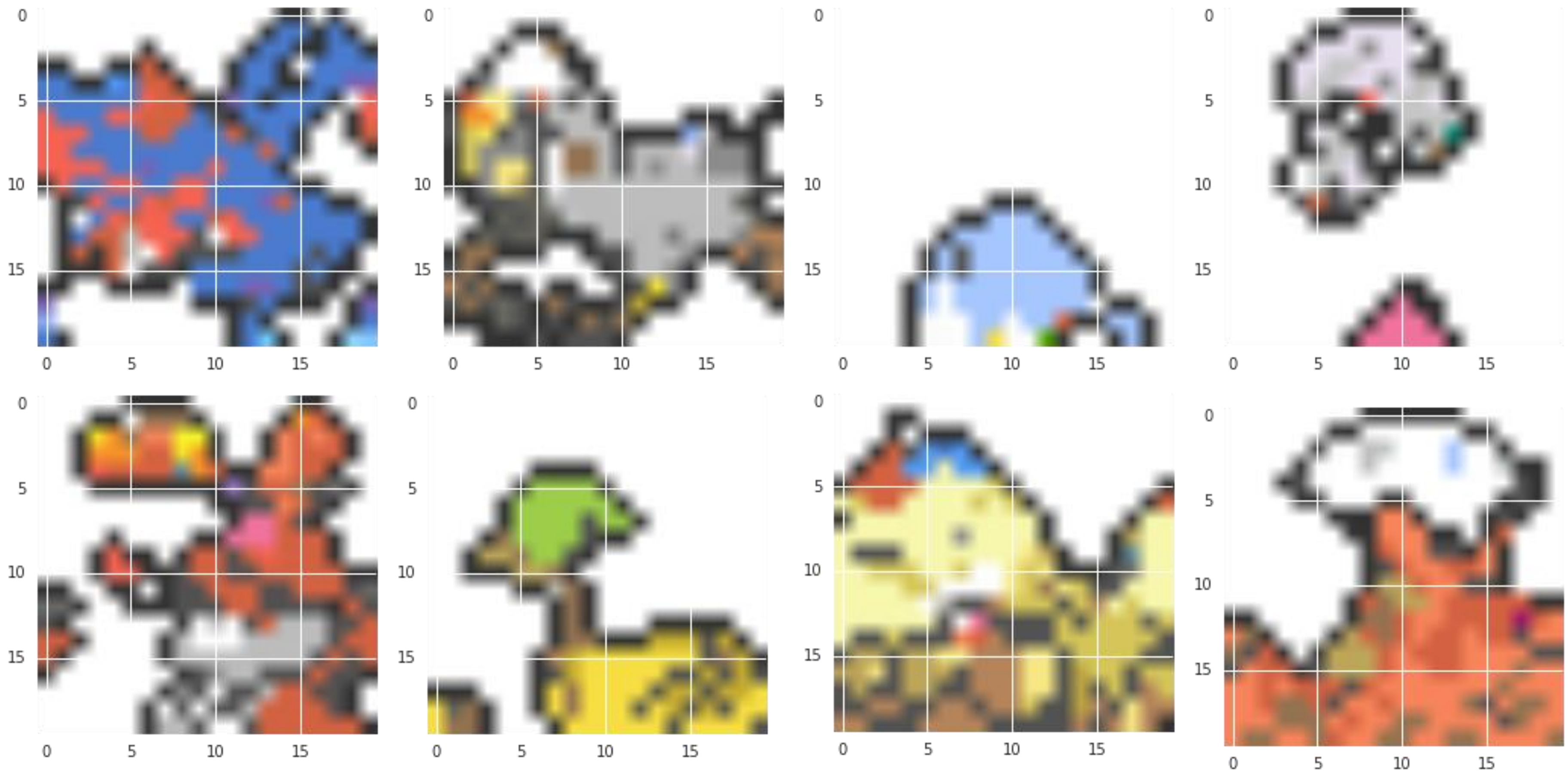
Cover 50%

Cover 75%



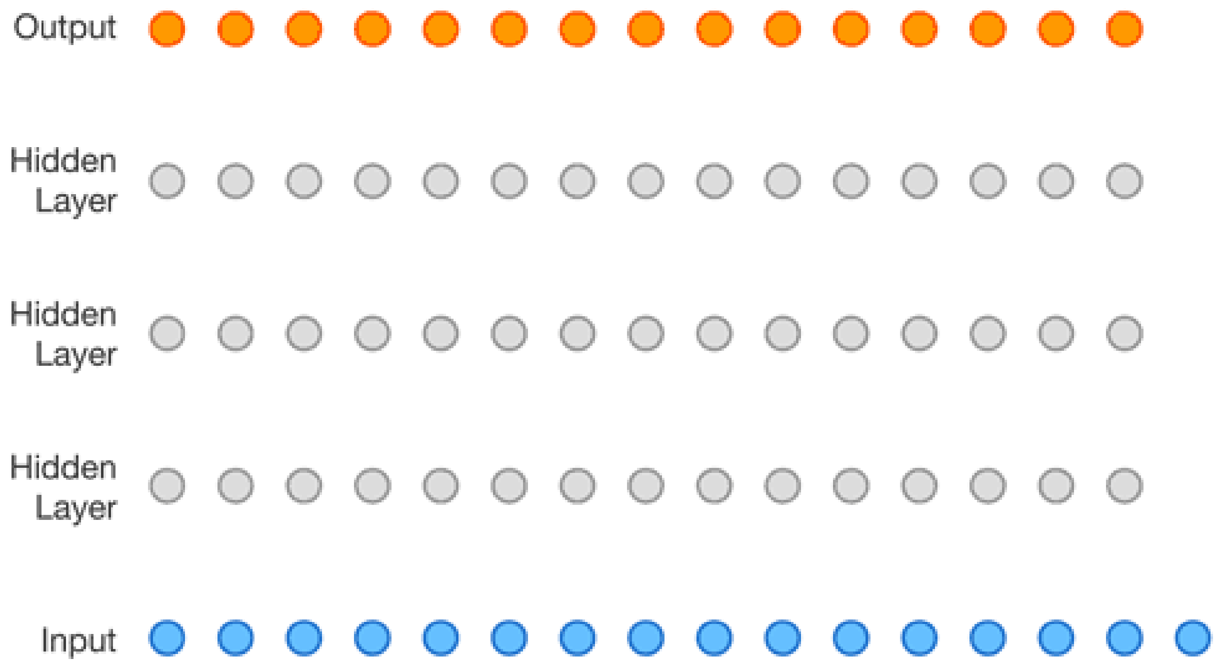
It is difficult to evaluate generation.

Component-by-component
- 寶可夢鍊成



Drawing from scratch　　Need some randomness

Component-by-component



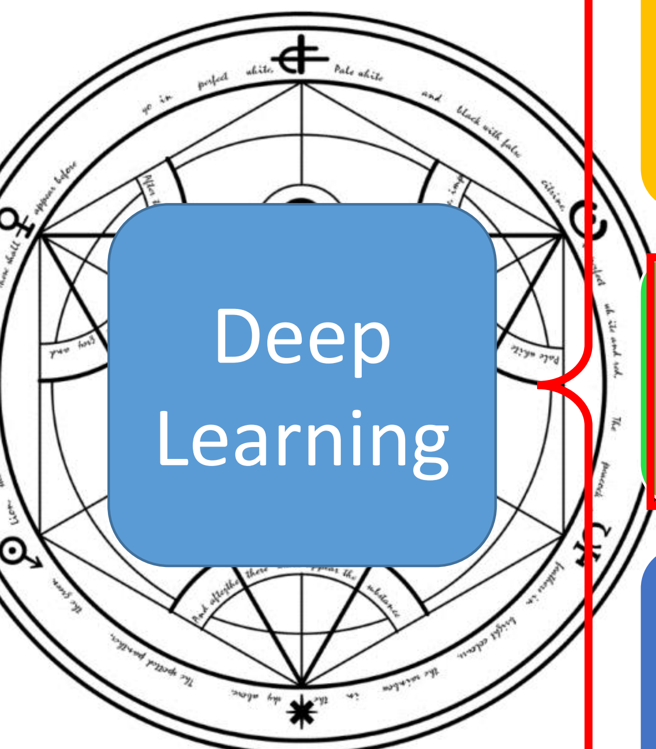
Audio: Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, arXiv preprint, 2016

Video: Nal Kalchbrenner, Aaron van den Oord, Karen Simonyan, Ivo Danihelka, Oriol Vinyals, Alex Graves, Koray Kavukcuoglu, Video Pixel Networks , arXiv preprint, 2016

http://yumekui.pixnet.net/album/photo/13915541-%E8%B6%85%E5%A4%A7%EF%BC%8E%E7%B7%B4%E6%88%90%E9%99%A3

Outline

VAE = Variational Auto-Encoder



Deep Learning

Auto-encoder

+

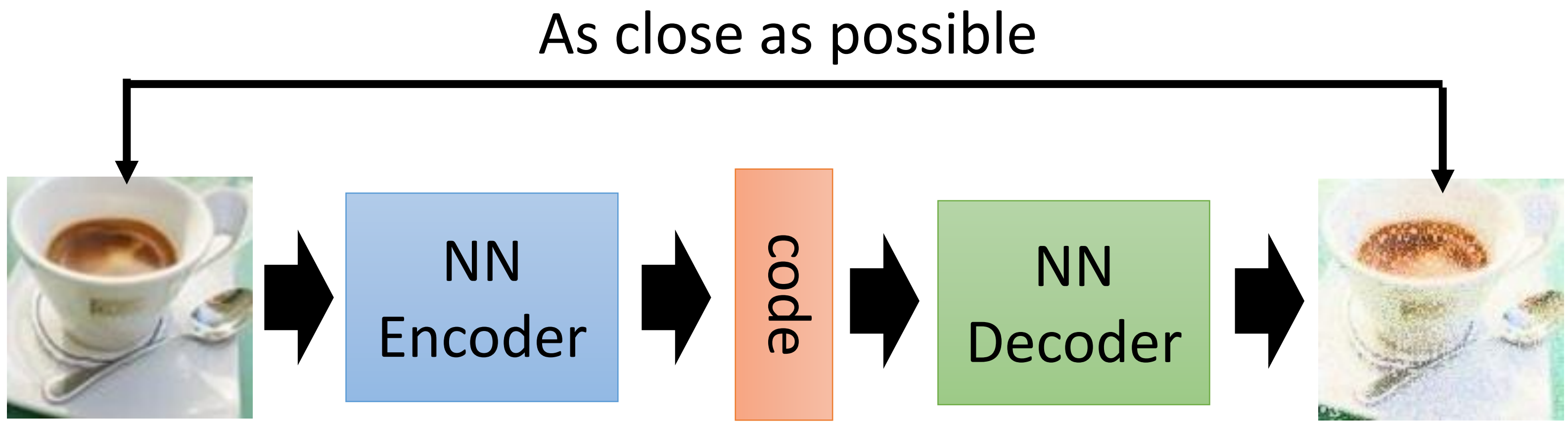
Deep Generative Model

=

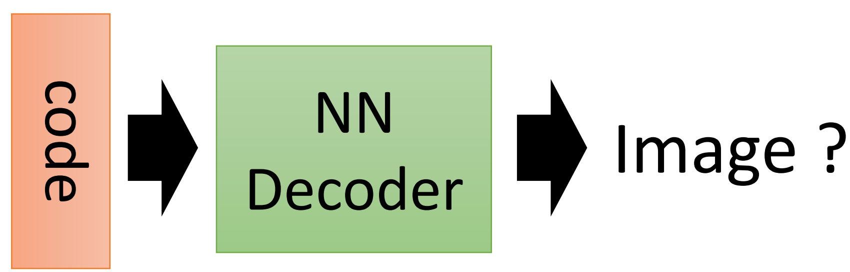
Conditional Generation

Component-wised, VAE, GAN

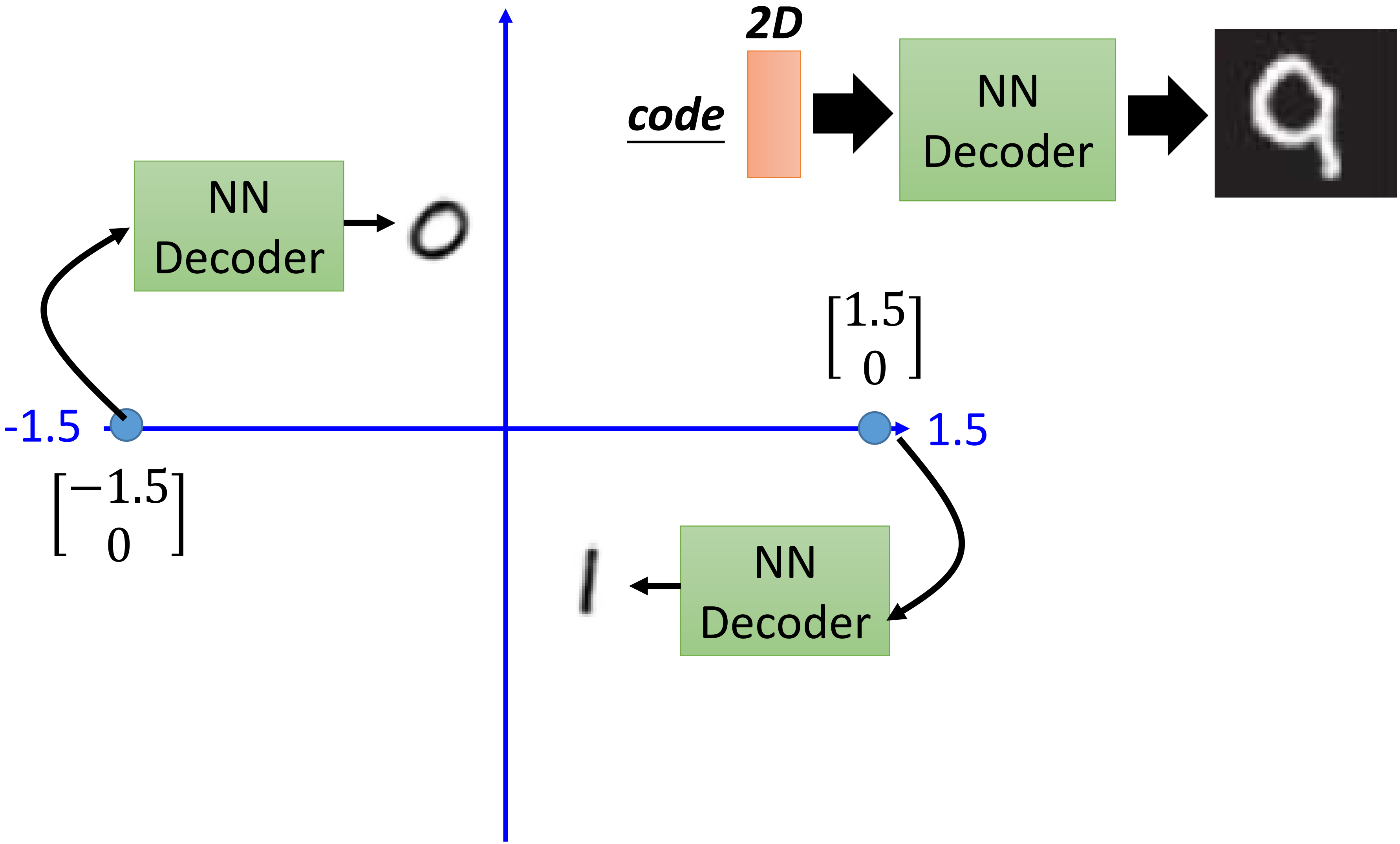
Remember Auto-encoder?



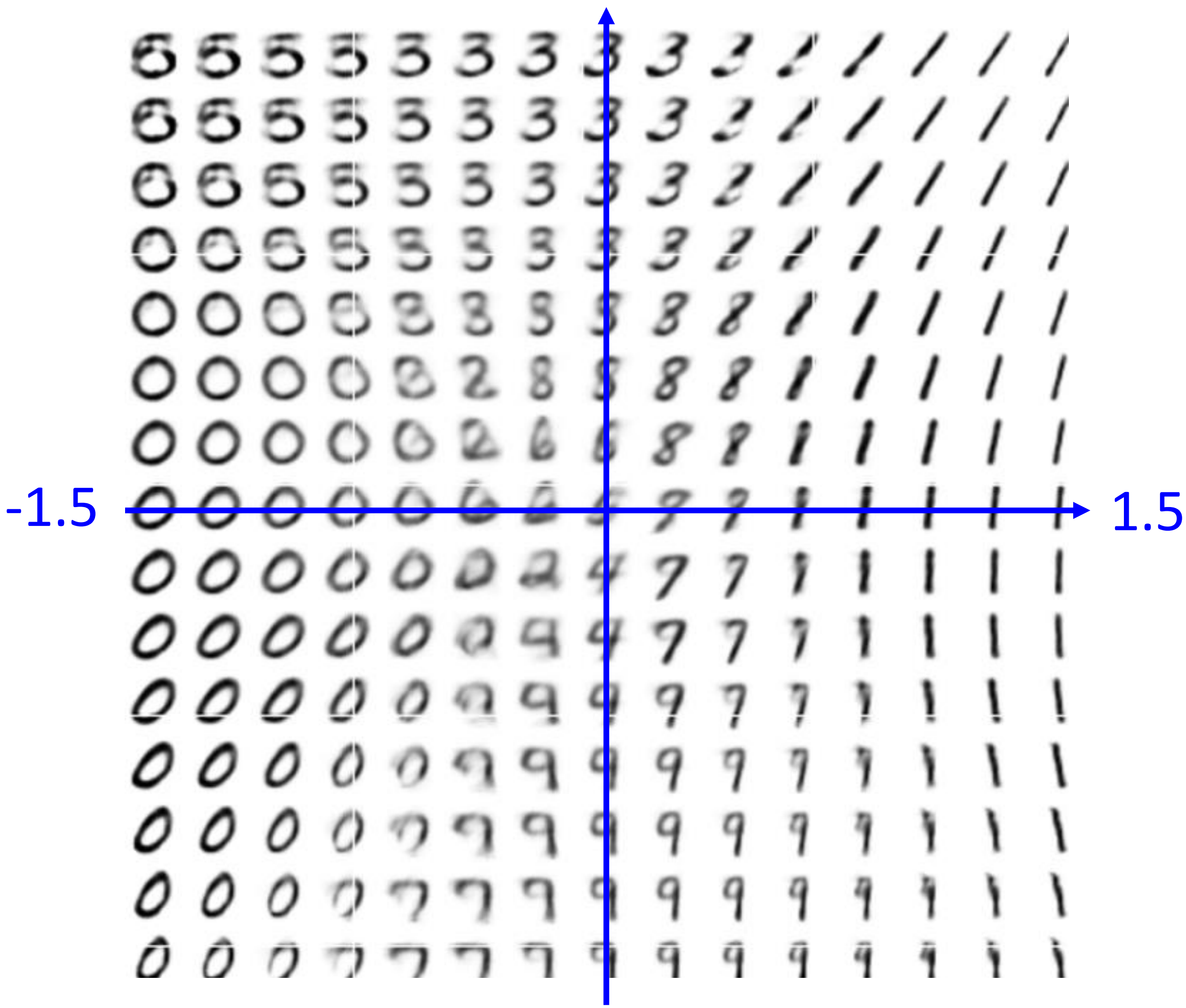
Randomly generate
a vector as code



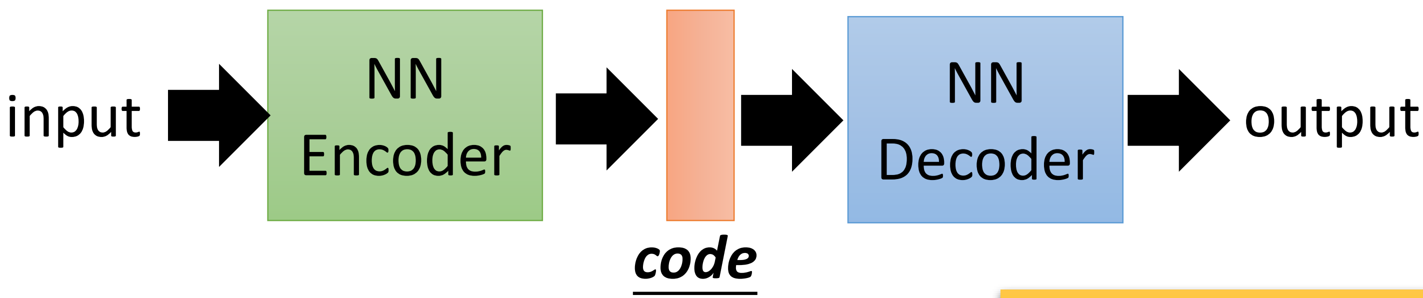
Remember Auto-encoder?



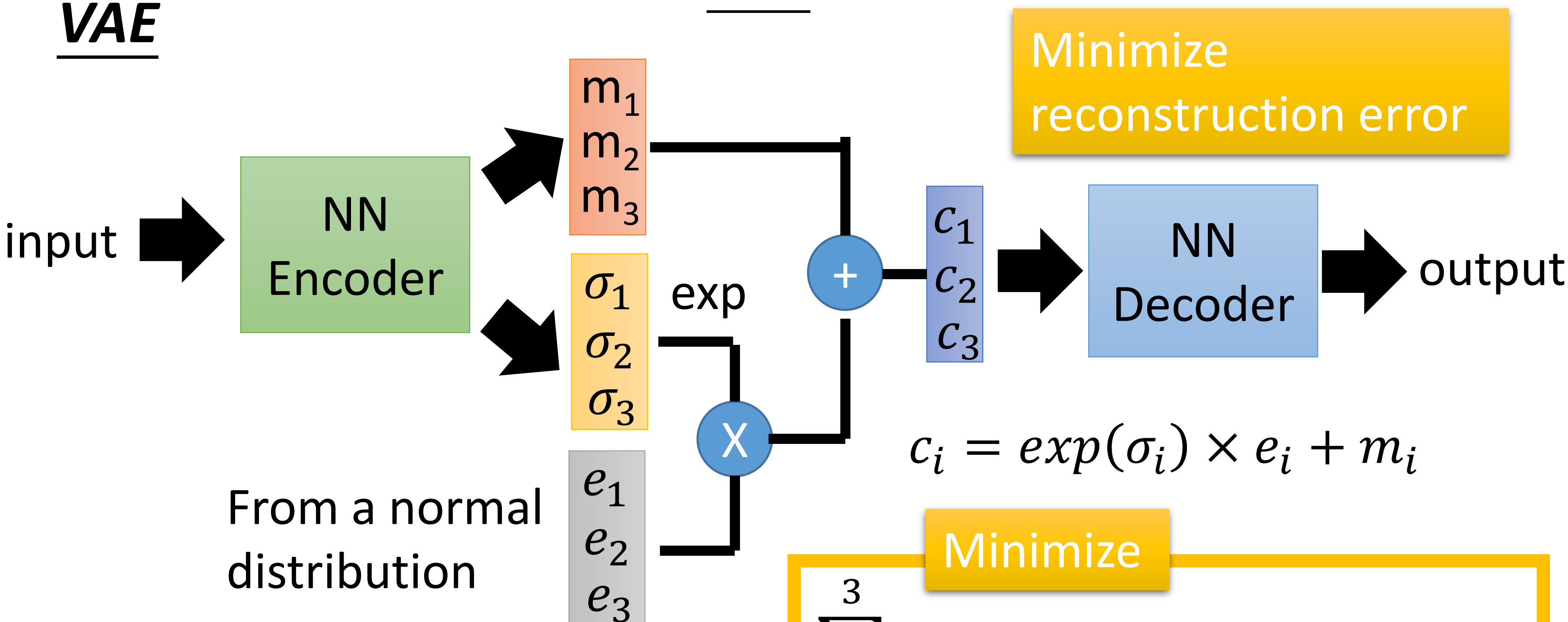
Remember Auto-encoder?



Auto-encoder



VAE



$$c_i = exp(\sigma_i) \times e_i + m_i$$

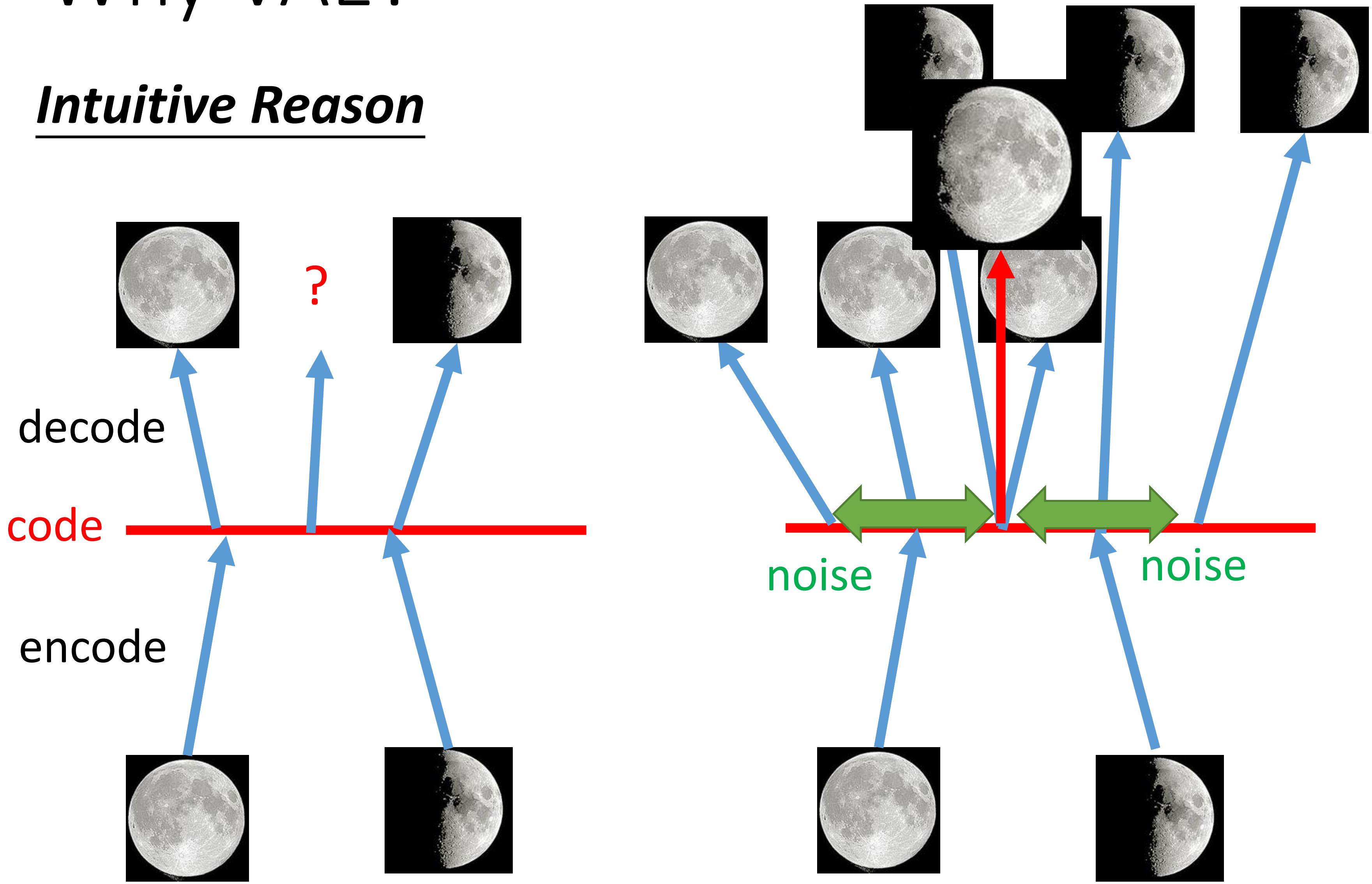
Minimize

$$\sum_{i=1}^{3}\left(exp(\sigma_i) - (1+\sigma_i) + (m_i)^2\right)$$

Auto-Encoding Variational Bayes,
https://arxiv.org/abs/1312.6114

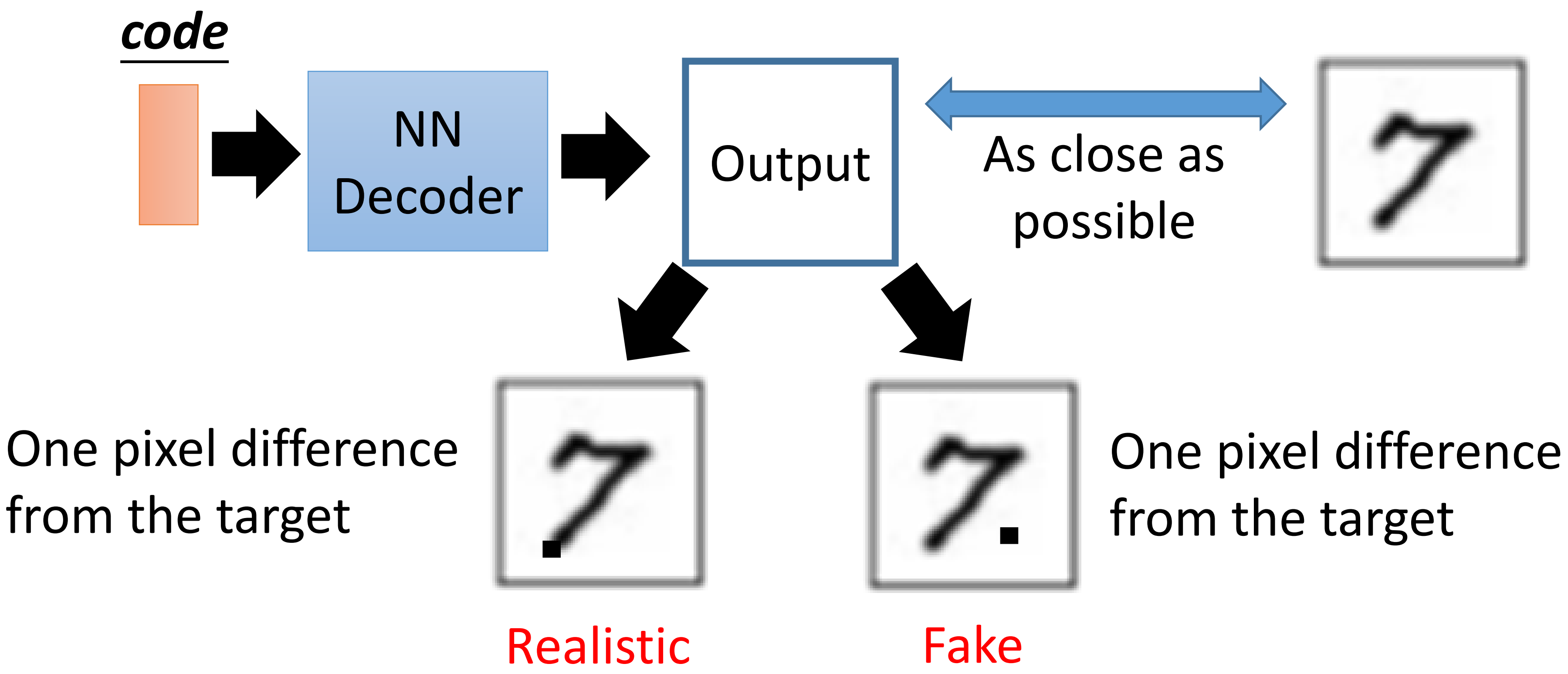
Why VAE?

Intuitive Reason



Problems of VAE

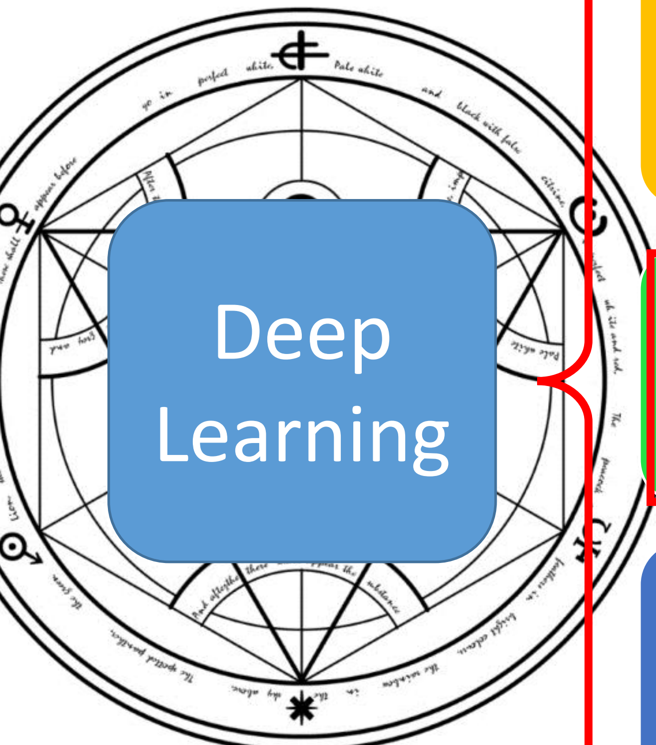
- It does not really try to simulate real images



http://yumekui.pixnet.net/album/photo/13915541-%E8%B6%85%E5%A4%A7%EF%BC%8E%E7%B7%B4%E6%88%90%E9%99%A3

Outline

GAN =
Generative Adversarial Network



Auto-encoder

+

Deep Generative Model

Component-wised,
VAE, $\boxed{\text{GAN}}$

=

Conditional Generation

Yann LeCun’s comment

What are some recent and potentially upcoming breakthroughs in deep learning?

Yann LeCun, Director of AI Research at Facebook and Professor at NYU
Written Jul 29 · Upvoted by Joaquin Quiñonero Candela, Director Applied Machine Learning at Facebook and Nikhil Garg, I lead a team of Quora engineers working on ML/NLP problems

……

The most important one, in my opinion, is adversarial training (also called GAN for Generative Adversarial Networks). This is an idea that was originally proposed by Ian Goodfellow when he was a student with Yoshua Bengio at the University of Montreal (he since moved to Google Brain and recently to OpenAI).

This, and the variations that are now being proposed is the most interesting idea in the last 10 years in ML, in my opinion.

https://www.quora.com/What-are-some-recent-and-potentially-upcoming-breakthroughs-in-deep-learning

Evolution

http://peellden.pixnet.net/blog/post/40406899-2013-%E7%AC%AC%E5%9B%9B%E5%AD%A3%EF%BC%8C%E5%86%AC%E8%9D%B6%E5%AF%82%E5%AF%A5

Brown

Kallima inachus

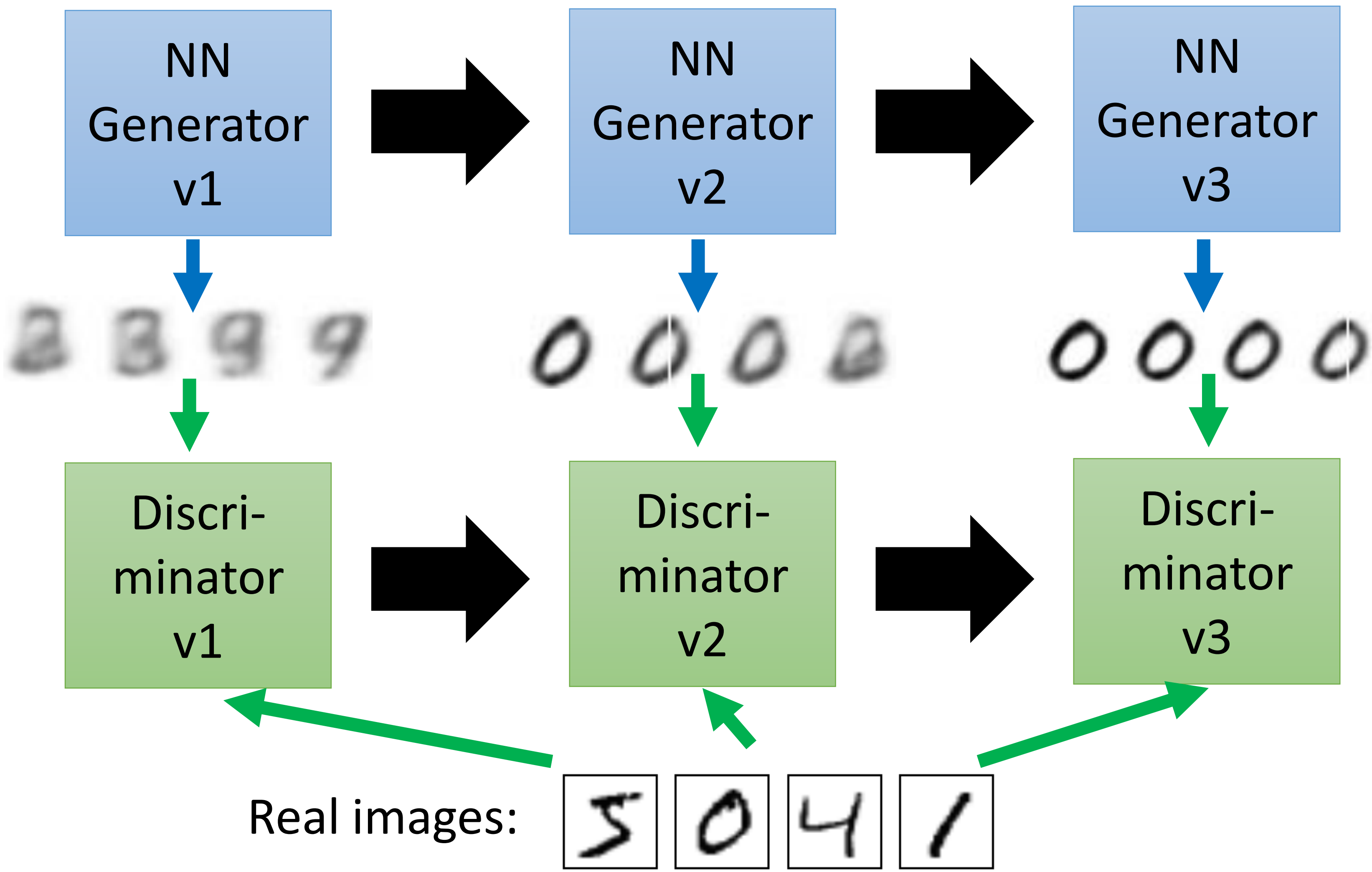
veins

Butterflies are not brown

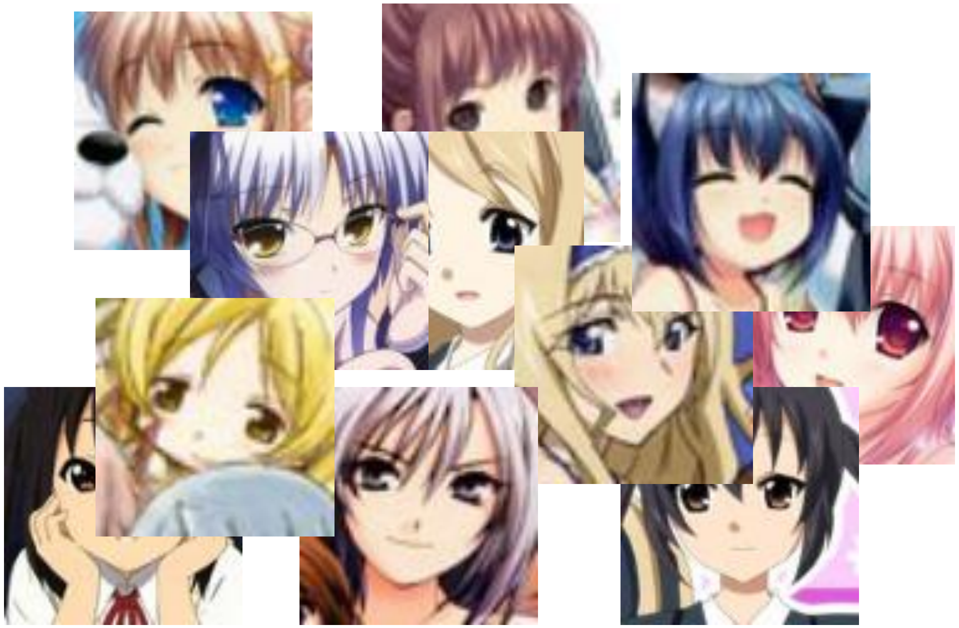
Butterflies do not have veins

....

The evolution of generation



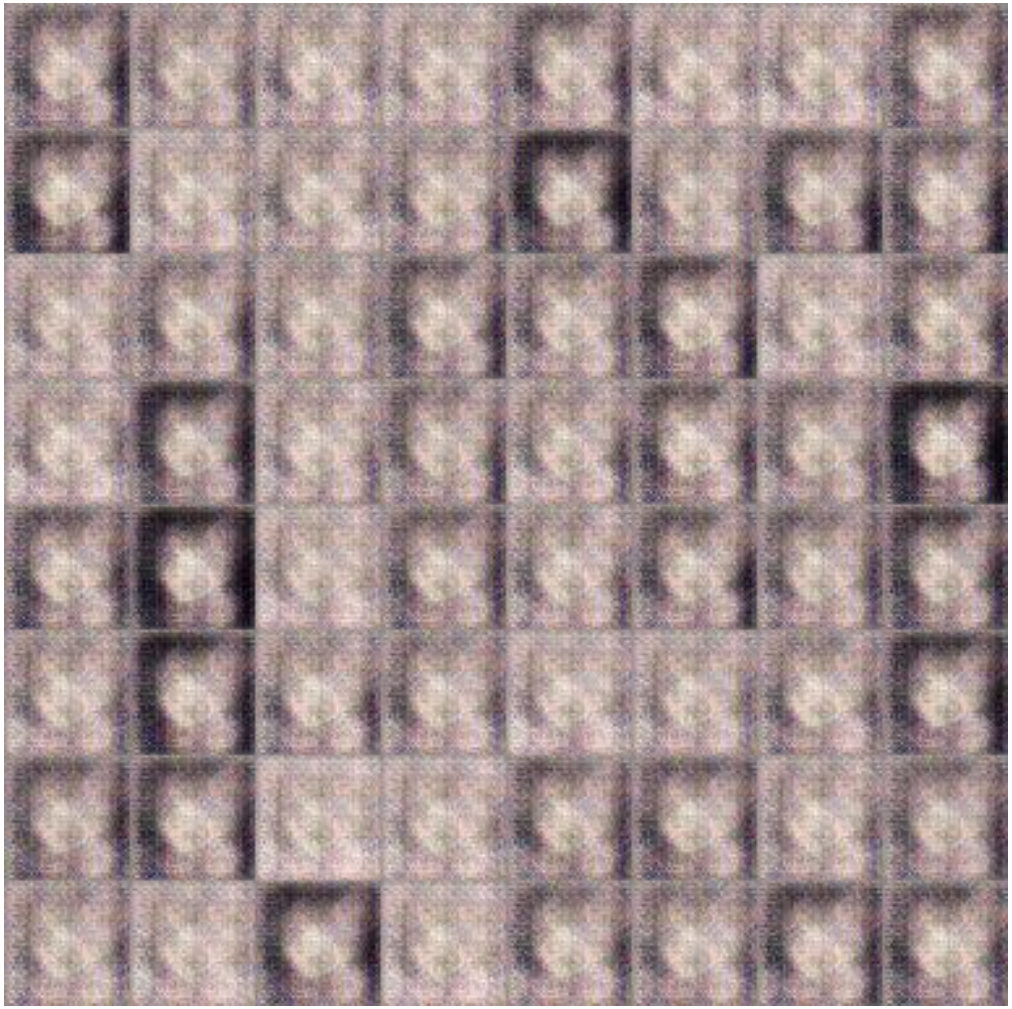
GAN – 二次元人物頭像鍊成



Source of images: https://zhuanlan.zhihu.com/p/24767059

DCGAN: https://github.com/carpedm20/DCGAN-tensorflow

GAN – 二次元人物頭像鍊成



100 rounds

GAN – 二次元人物頭像鍊成

1000 rounds

GAN – 二次元人物頭像鍊成

2000 rounds

GAN – 二次元人物頭像鍊成

5000 rounds

GAN – 二次元人物頭像鍊成

10,000 rounds

GAN – 二次元人物頭像鍊成

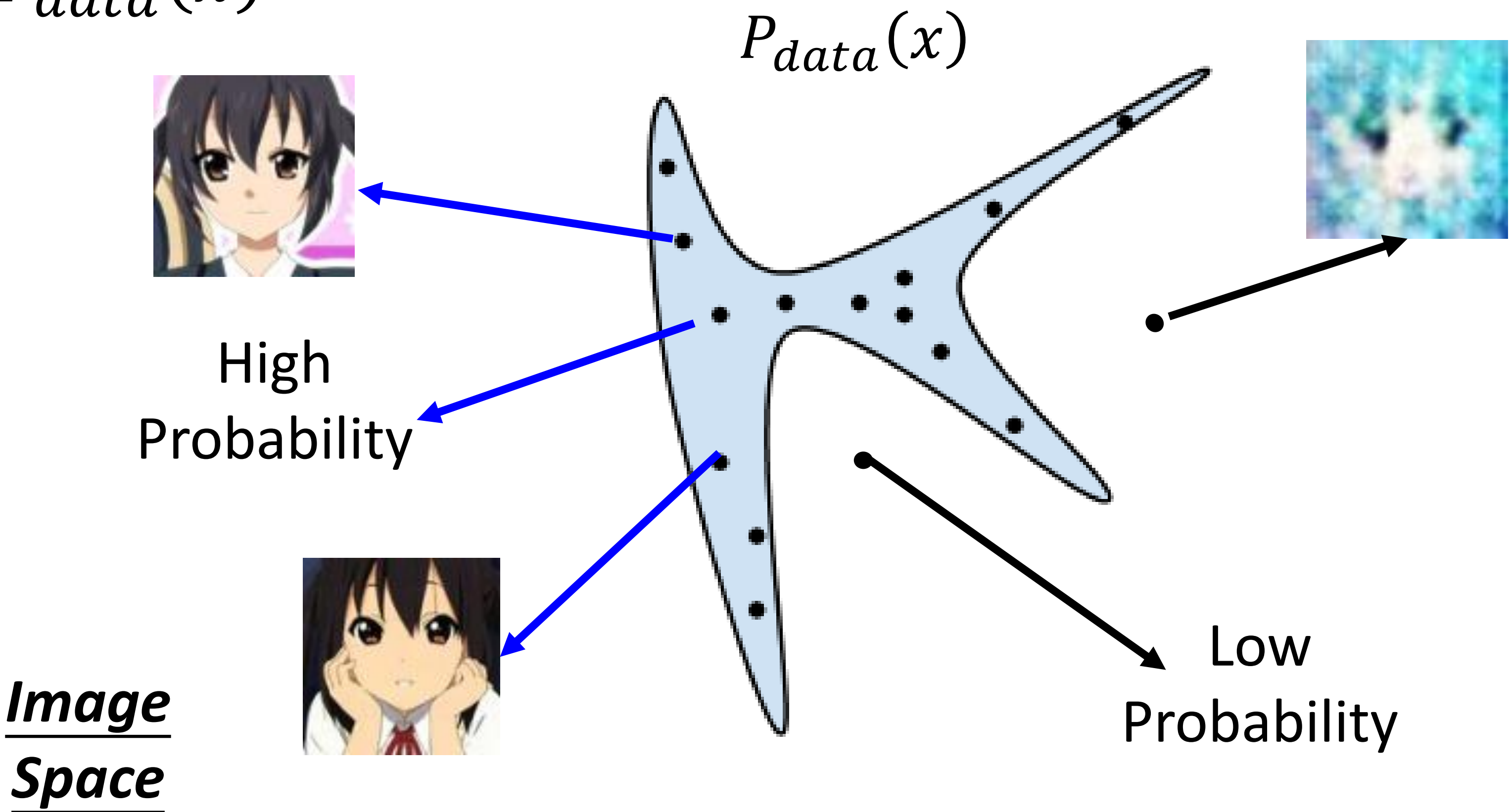
20,000 rounds

GAN – 二次元人物頭像鍊成

50,000 rounds

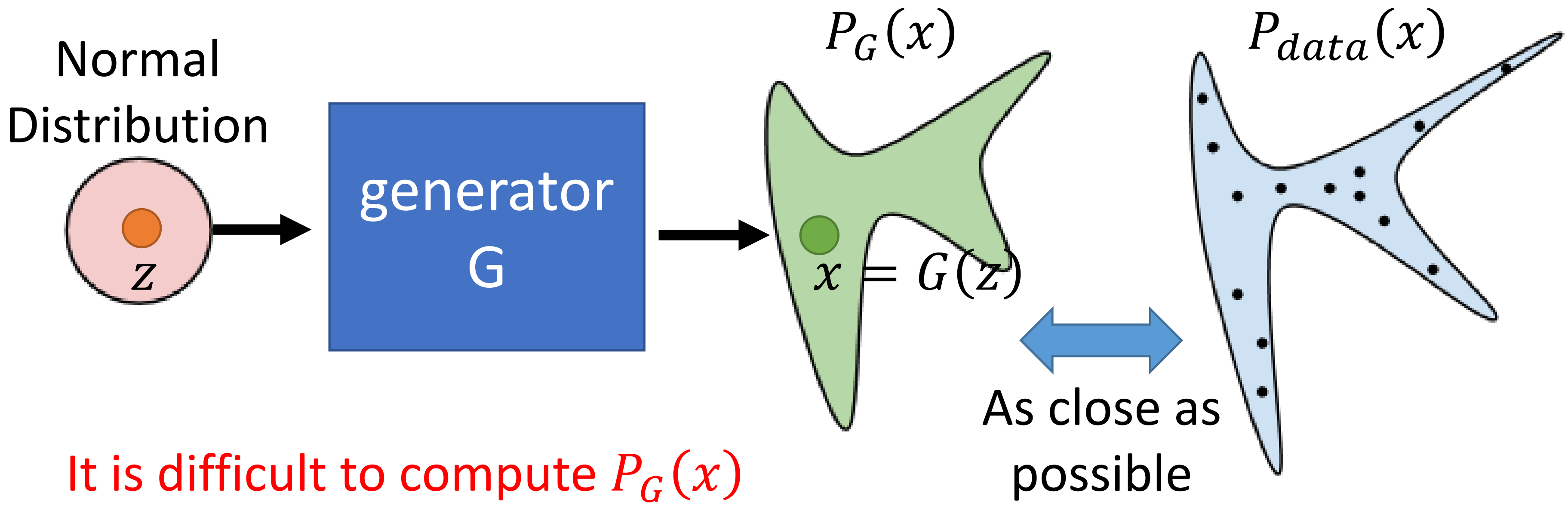
Basic Idea of GAN

- The data we want to generate has a distribution $P_{data}(x)$



Basic Idea of GAN

- A generator G is a network. The network defines a probability distribution.

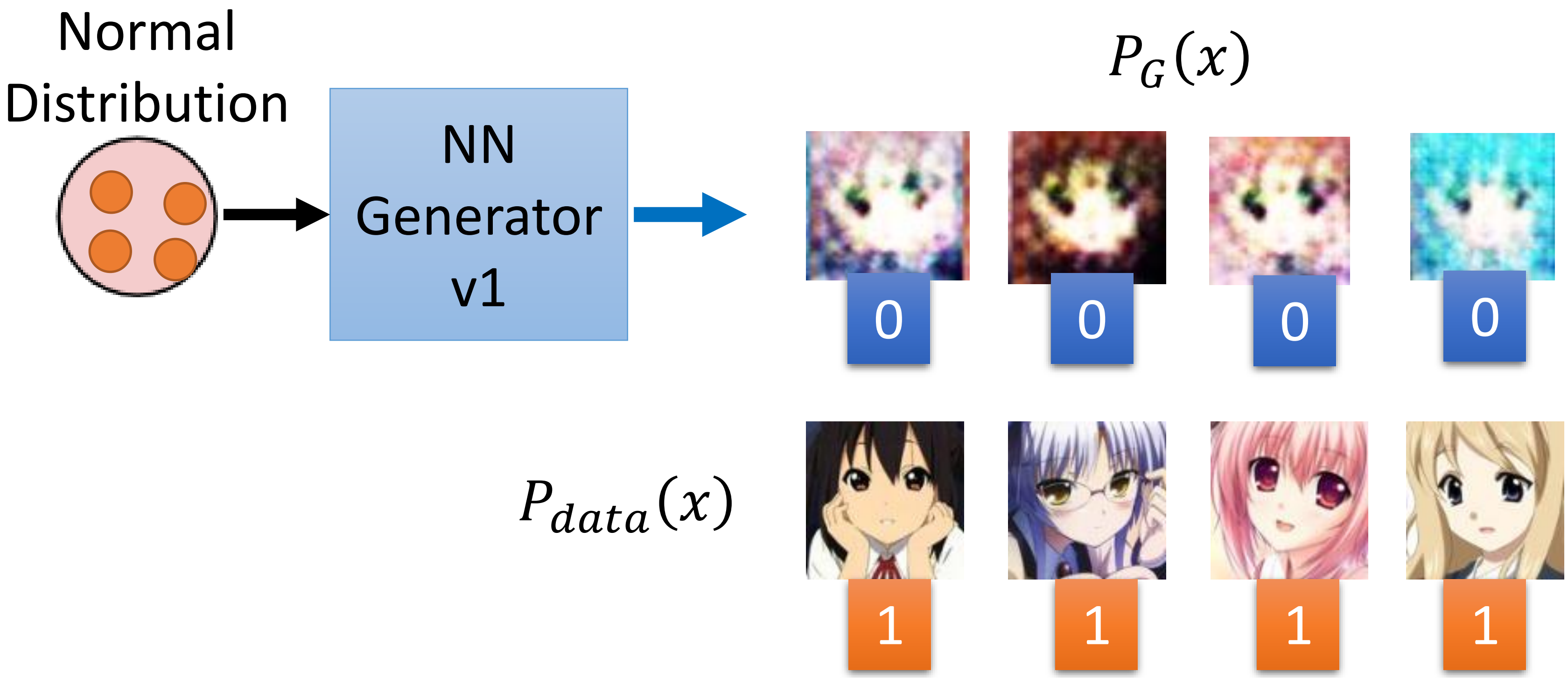


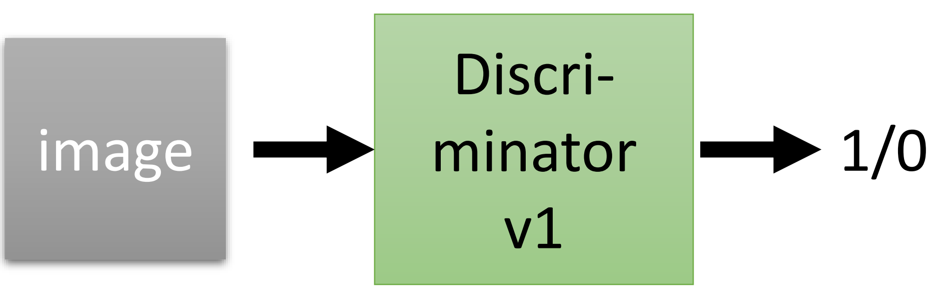
It is difficult to compute $P_G(x)$

We do not know what the distribution looks like.

https://blog.openai.com/generative-models/

Basic Idea of GAN





It can be proofed that the ***loss the discriminator*** related to ***JS divergence***.

Basic Idea of GAN

- ***Next step***:
  - Updating the parameters of generator
  - To minimize the JS divergence

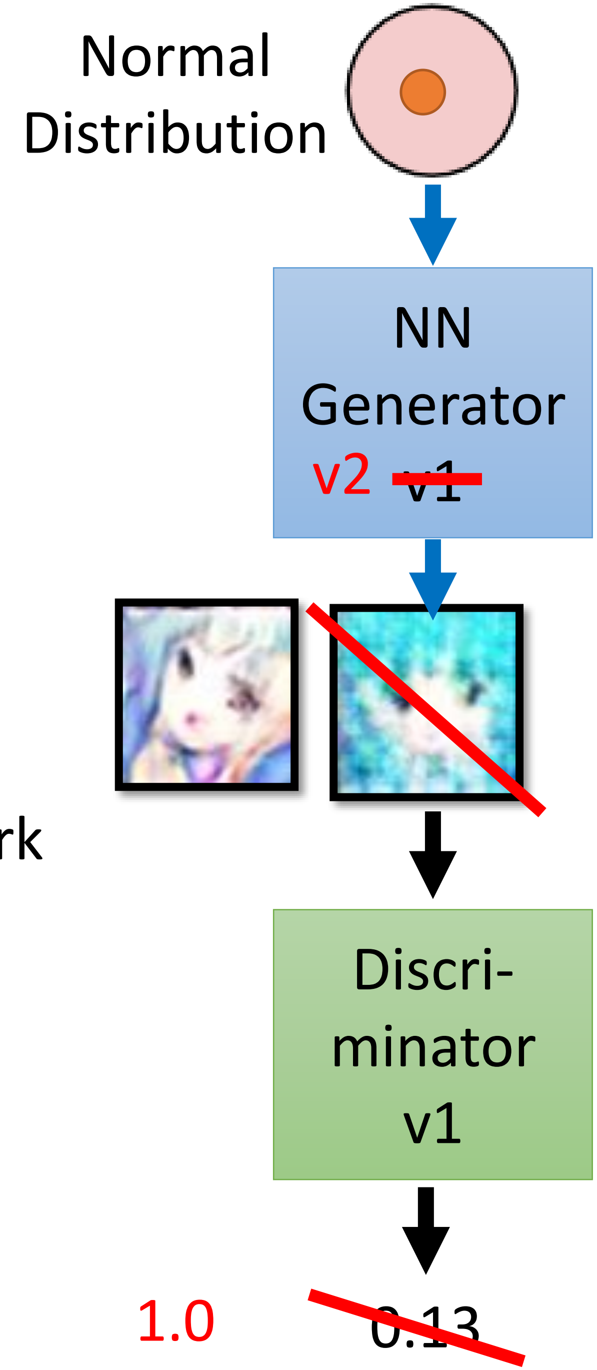
➡ The output be classified as "real" (as close to 1 as possible)

Generator + Discriminator = a network

Using gradient descent to update the parameters in the generator, but fix the discriminator

Original GAN is hard to train …….

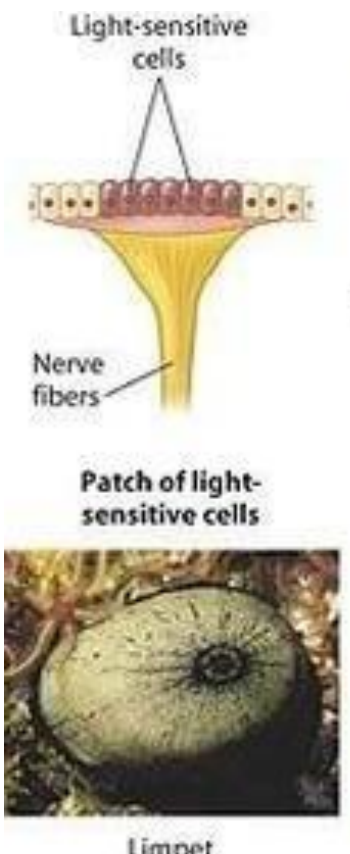
➡ ***W-GAN***



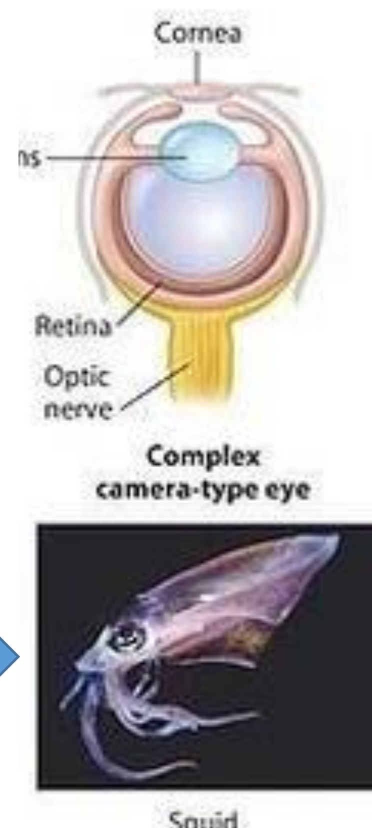
http://www.guokr.com/post/773890/

Why GAN is hard to train?

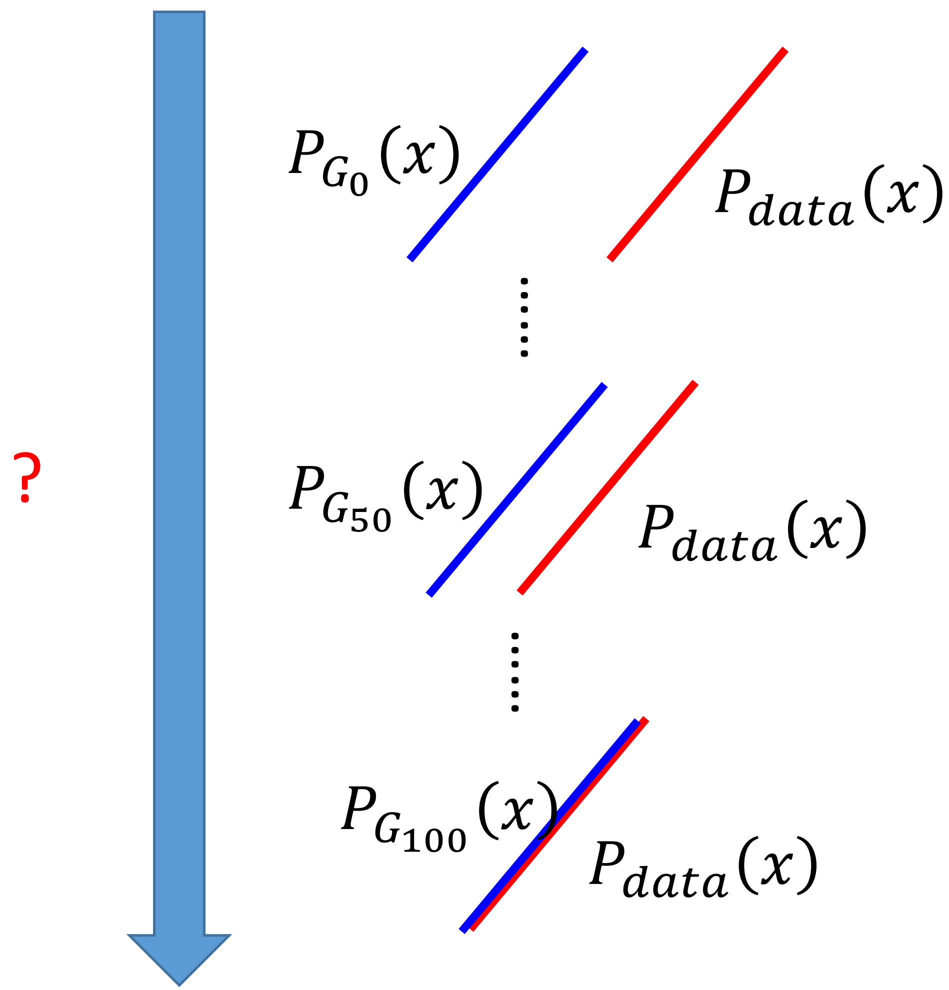
回到演化的比喻 ……



Better



Why GAN is hard to train?



$JS\left(P_{G_1}||P_{data}\right) = log2$

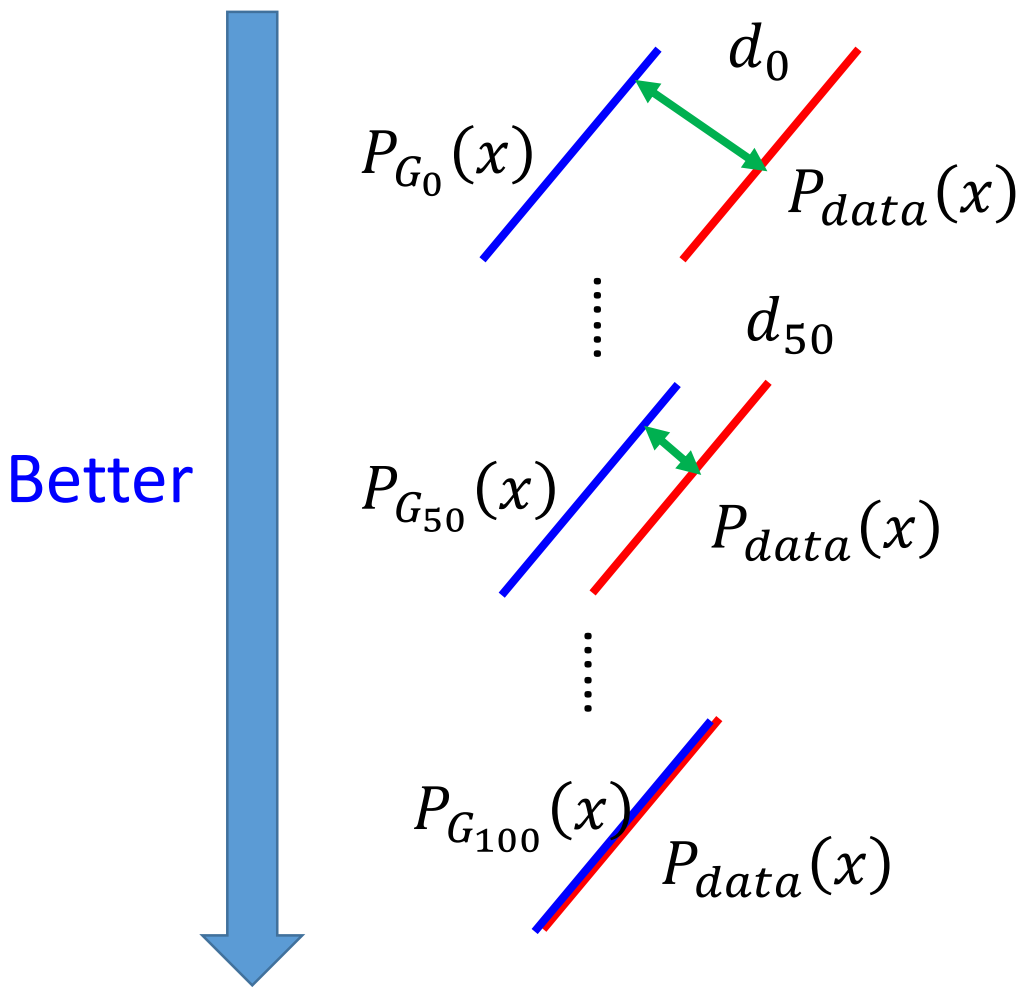
$JS\left(P_{G_2}||P_{data}\right) = log2$

Not really better ……

$JS\left(P_{G_2}||P_{data}\right) = 0$

WGAN

Using Wasserstein distance instead of JS divergence

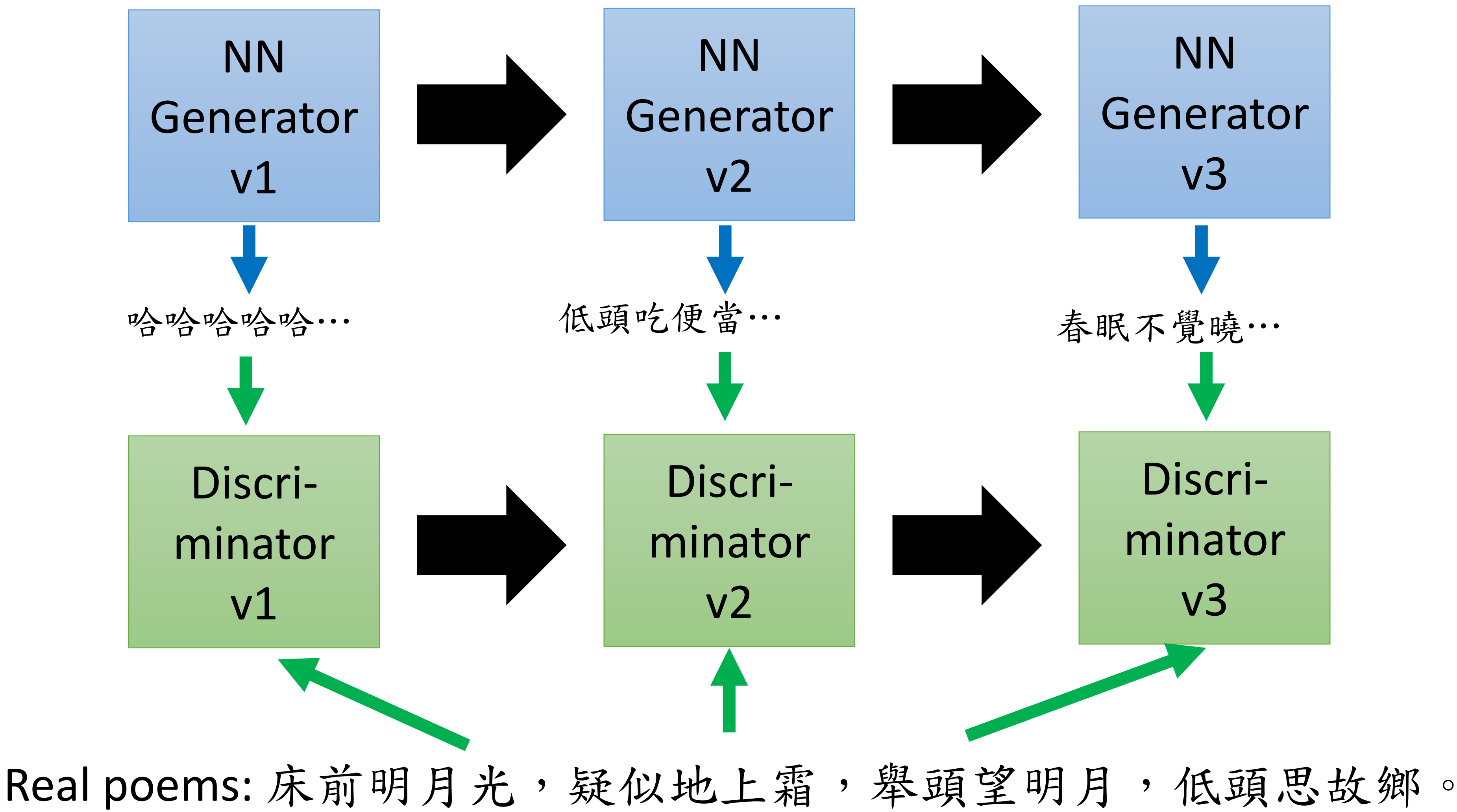


$$W\left(P_{G_1}||P_{data}\right) = d_0$$

$$W\left(P_{G_2}||P_{data}\right) = d_{50}$$

$$W\left(P_{G_2}||P_{data}\right) = 0$$

WGAN – 唐詩鍊成


NN Generator v1
NN Generator v2
NN Generator v3
哈哈哈哈哈…
低頭吃便當…
春眠不覺曉…
Discri-minator v1
Discri-minator v2
Discri-minator v3
Real poems: 床前明月光，疑似地上霜，舉頭望明月，低頭思故鄉。

由 李仲翊 同學提供實驗結果

WGAN – 唐詩鍊成

Random generated

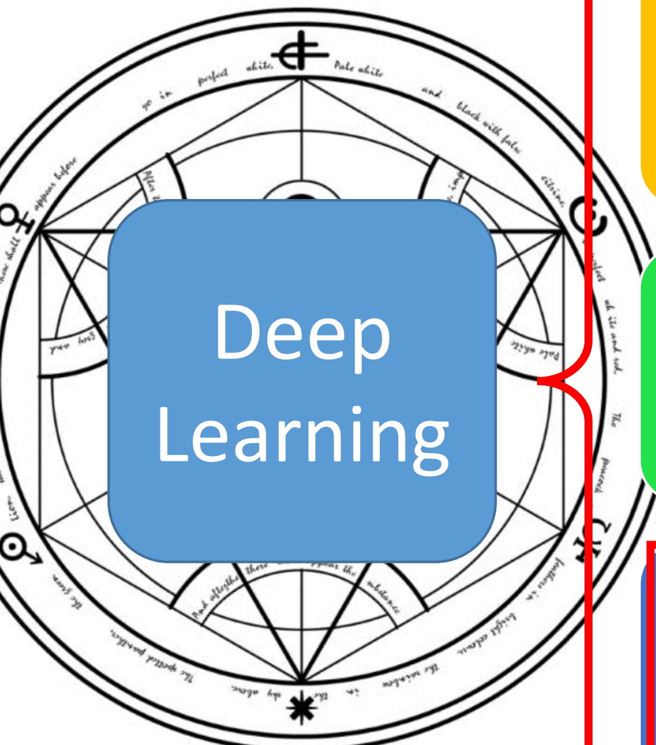
- 升雲白遲丹齋取，此酒新巷市入頭。黃道故海歸中後，不驚入得韻子門。
- 摅口容章蕃翎翎，邦貸無遊隔將毬。外蕭曾臺遶出畧，此計推上呂天夢。
- 新來寶伎泉，手雪泓臺蓑。曾子花路魏，不謀散薦船。
- 功持牧度機邈爭，不躚官嬉牧涼散。不迎白旅今掩冬，盡蘸金祇可停。
- 玉十洪沄爭春風，溪子風佛挺橫鞋。盤盤稅焰先花齋，誰過飄鶴一丞幢。
- 海人依野庇，為阻例沉迴。座花不佐樹，弟蘭十名儂。
- 入維當興日世瀕，不評皺。頭醉空其杯，駸園凋送頭。
- 鉢笙動春枝，寶絫潔長知。官爲窰爛去，絆粒薛一靜。
- 吾涼腕不楚，縱先待旅知。楚人縱酒待，一蔓飄聖猜。
- 折幕故癘應韻子，徑頭霜瓊老徑徑。尚錯春鏘熊悽梅，去吹依能九將香。
- 通可矯目鷄須浄，丹迤挈花一抵嫖。外子當目中前醒，迎日幽筆鈎弧前。
- 庭愛四樹人庭好，無衣服仍繡秋州。更怯風流欲鳩雲，帛陽舊據畆婷儻。
- 日小輪壯杜，浮夕到來態。外日三落去，仙林停得動。

So many GANs
…… Just name a few

Modifying the Optimization of GAN	Different Structure from the Original GAN
fGAN	Conditional GAN
WGAN	Semi-supervised GAN
Least-square GAN	InfoGAN
Loss Sensitive GAN	BiGAN
Energy-based GAN	Cycle GAN
Boundary-seeking GAN	Disco GAN
Unroll GAN	VAE-GAN
……	……

http://yumekui.pixnet.net/album/photo/13915541-%E8%B6%85%E5%A4%A7%EF%BC%8E%E7%B7%B4%E6%88%90%E9%99%A3

Outline



Deep Learning

Auto-encoder

Deep Generative Model

Conditional Generation

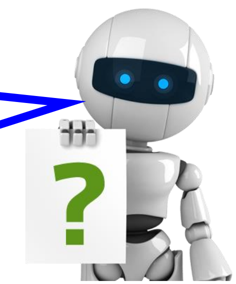
Conditional Generation

- ***We don't want to simply generate some random stuff.***
- Generate sentences based on conditions:

<u>Caption Generation</u>

Given condition:

"A dog is running."

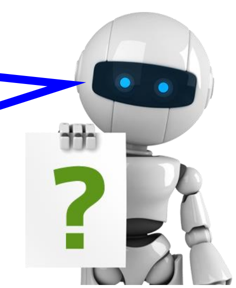


<u>Chat-bot</u>

Given condition:

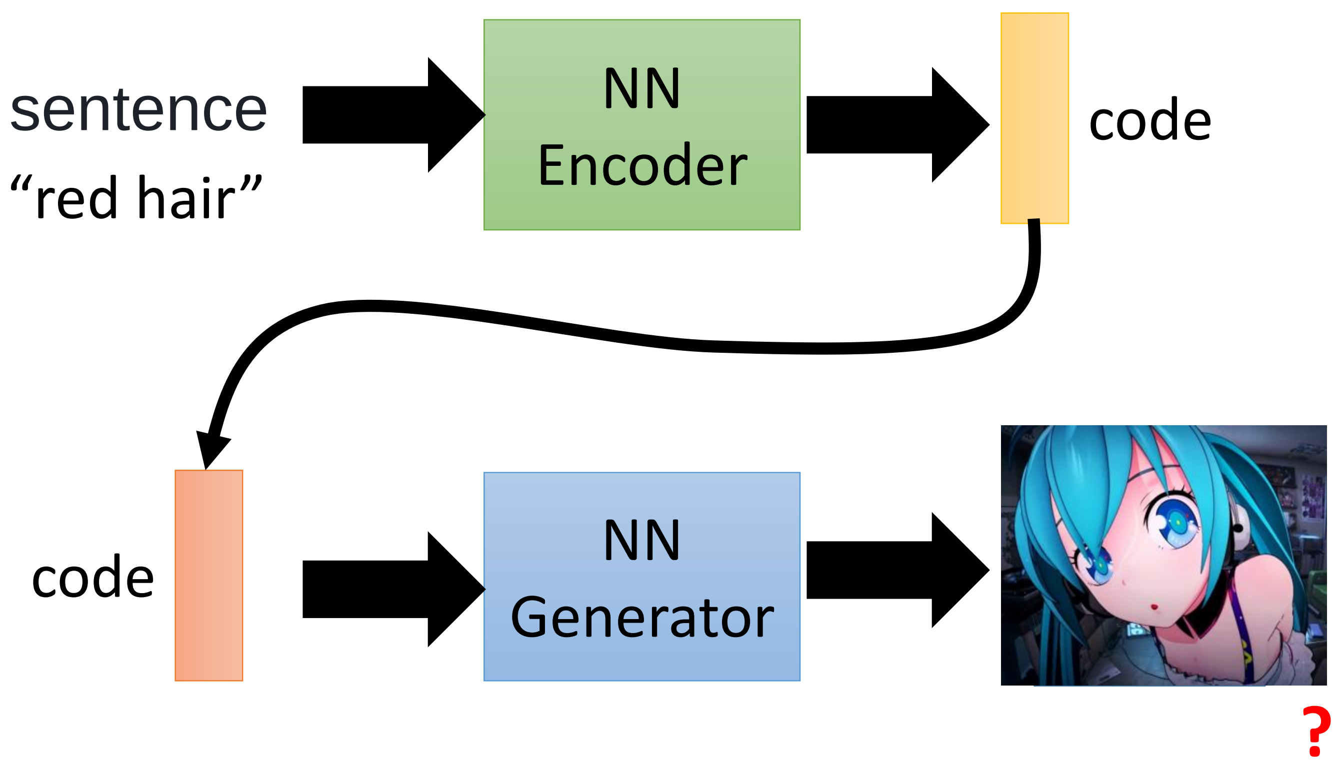
"Hello"

"Hello. Nice to see you."



Conditional Generation

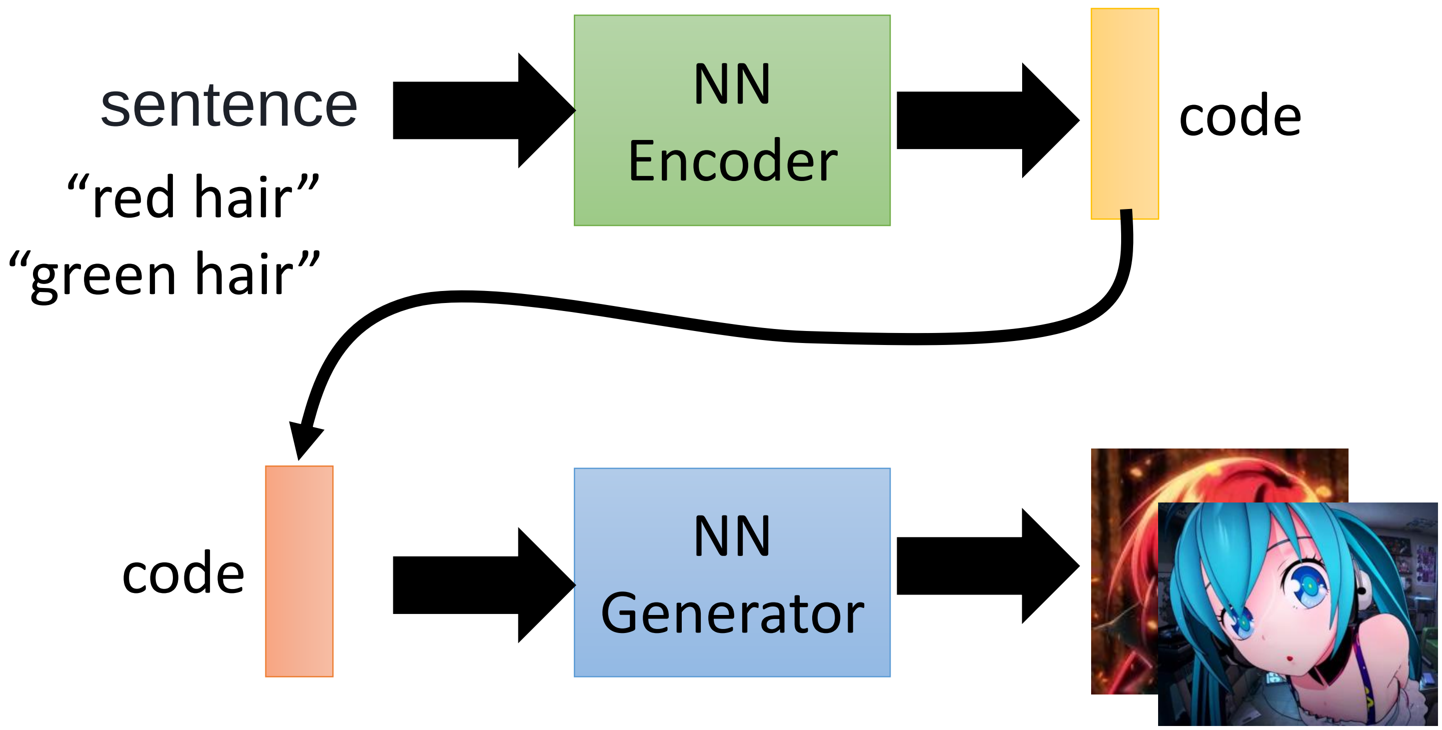
- E.g. 根據文字敘述畫出動漫人物頭像



Conditional Generation

Need some supervision

- E.g. 根據文字敘述畫出動漫人物頭像



Conditional Generation

由 曾柏翔 同學
提供實驗結果

- E.g. 根據文字敘述畫出動漫人物頭像

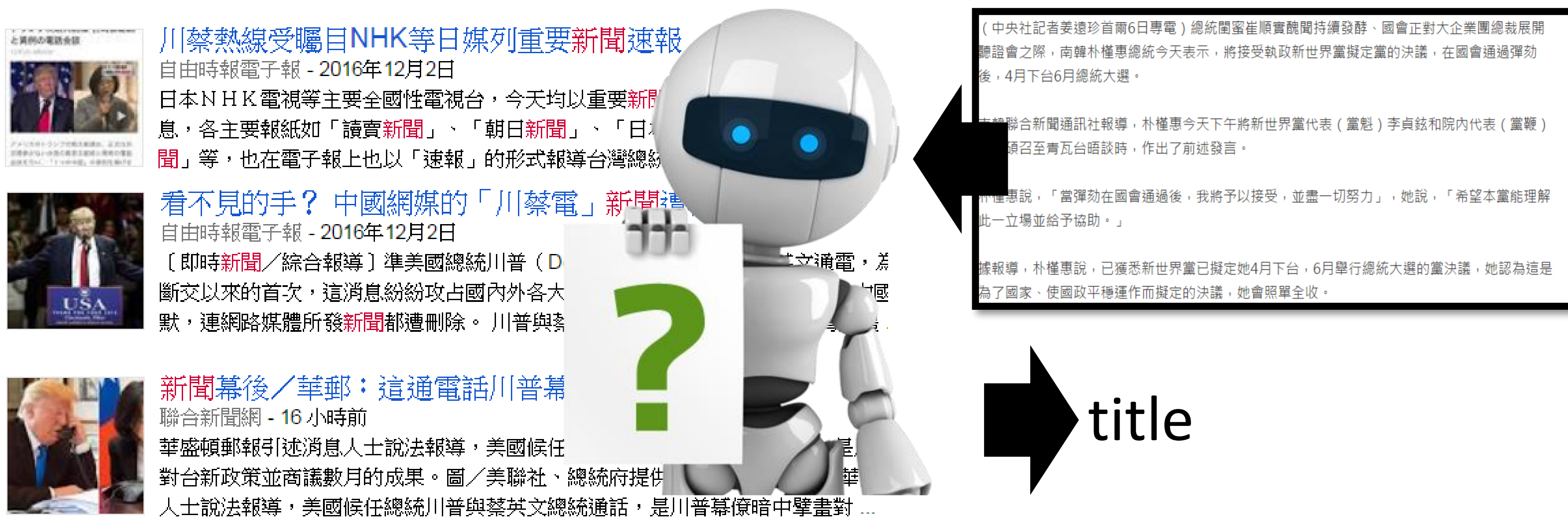
Red hair, long hair

Black hair, **blue** eyes

Blue hair, **green** eyes

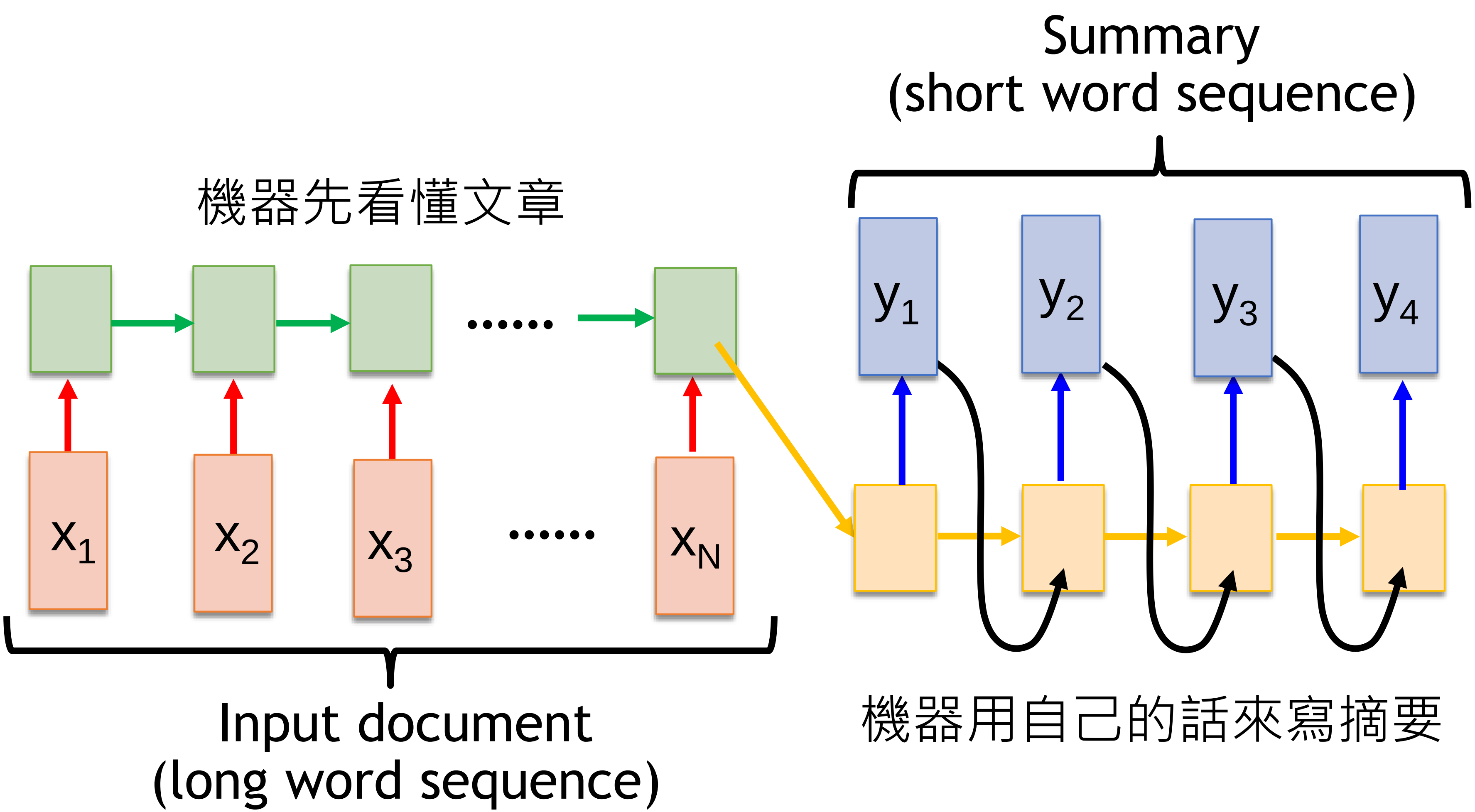
Text to Text - Summarization

- Abstractive Summary: 用自己的話寫 summary
  - Machine learns to do title generation from 2,000,000 training examples



[Yu & Lee, SLT 16]

Text to Text - Summarization



感謝 盧柏儒 提供實驗結果

Text to Text - Summarization

Document: 據印度報業托拉斯報道印度北方邦22日發生一起小公共汽車炸彈爆炸事件造成15 人死亡 3 人受傷 ……

Human: 印度汽車炸彈爆炸造成15人死亡

Machine: 印度發生汽車爆炸事件

Document: 刑事局偵四隊今天破獲一個中日跨國竊車集團，根據調查國內今年七月開放重型機車上路後 ……

Human:跨國竊車銷贓情形猖獗直得國內警方注意

Machine:刑事局破獲中國車集

Text to Text - Summarization

- Demo: 台大語音處理實驗室 葉政杰、周如杰、盧柏儒

Video to Text

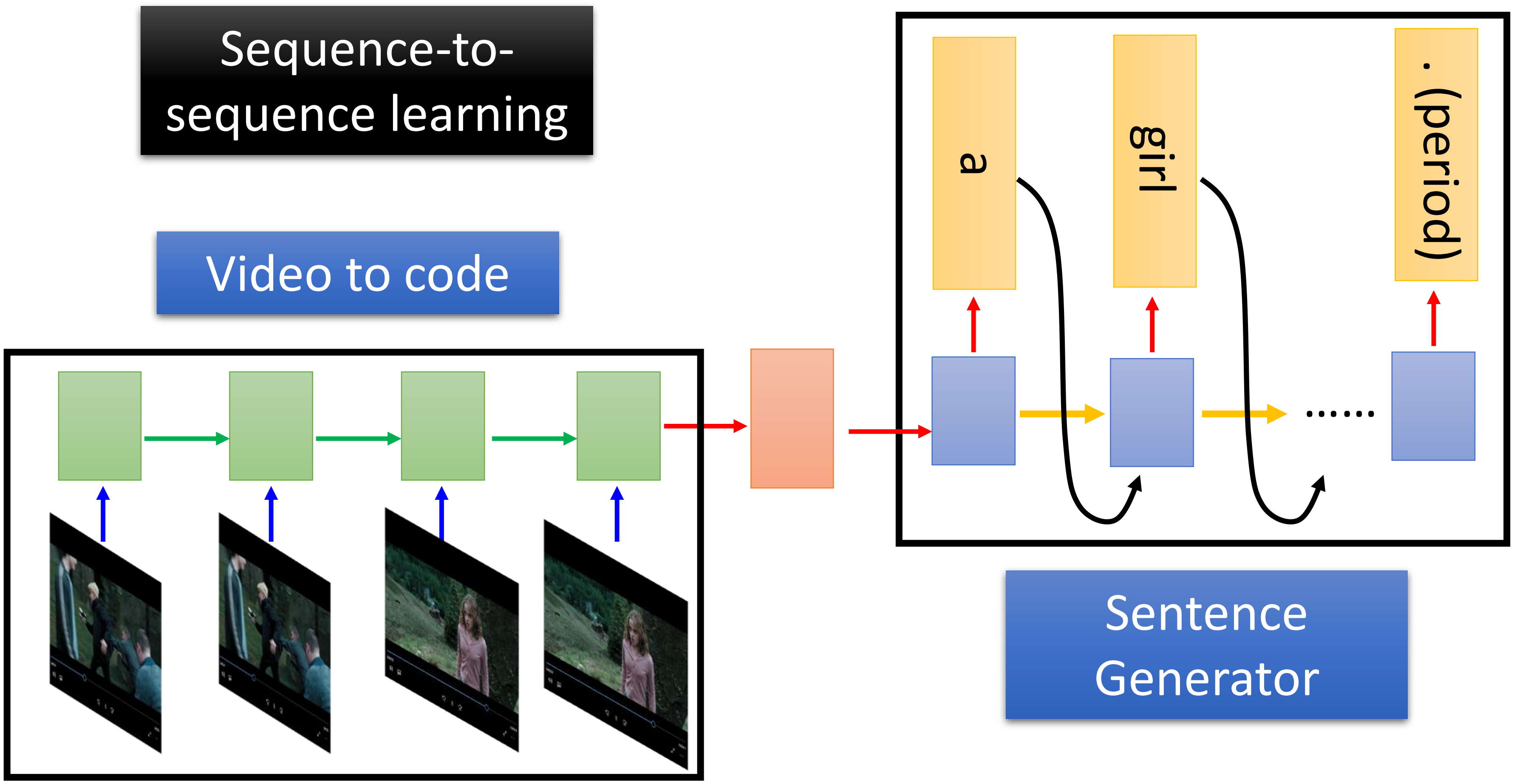
Video

A girl is running.

A group of people is knocked by a tree.

A group of people is walking in the forest.

Video to Text



Video to Text

- Can machine describe what it see from video?
- Demo: 台大語音處理實驗室 曾柏翔、吳柏瑜、盧宏宗

MTK 產學大聯盟

Image to Text

- Represent the input condition as a vector, and consider the vector as the input of RNN generator

Image Caption Generation

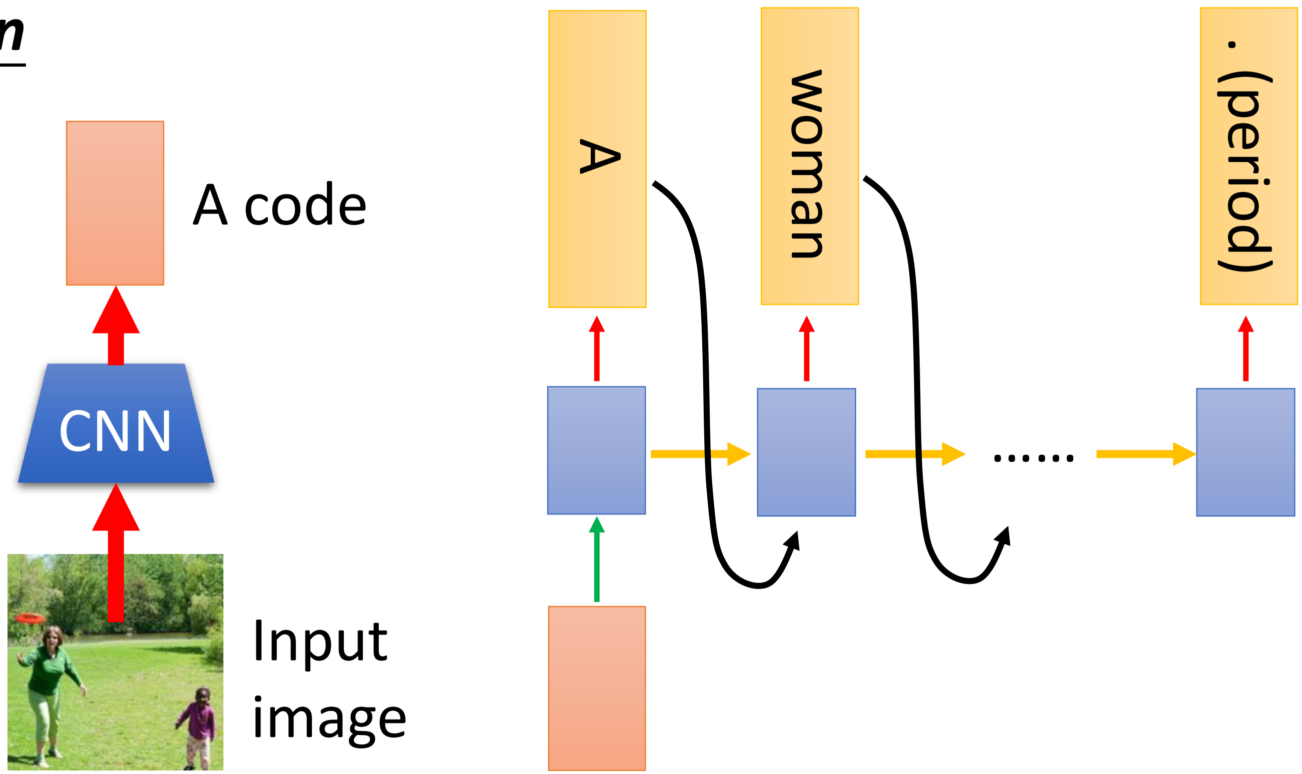


Image to Text

- Can machine describe what it see from image?
- Demo:台大電機系 大四 蘇子睿、林奕辰、徐翊祥、陳奕安

http://news.ltn.com.tw/photo/politics/breakingnews/975542_1

To Learn More …

- **Machine Learning**
  - Slides: http://speech.ee.ntu.edu.tw/~tlkagk/courses_ML16.html
  - Video: https://www.youtube.com/watch?v=fegAeph9UaA&list=PLJV_el3uVTsPy9oCRY30oBPNLCo89yu49
- **Machine Learning and Having it Deep and Structured**
  - Slides: http://speech.ee.ntu.edu.tw/~tlkagk/courses_MLDS17.html
  - Video: https://www.youtube.com/watch?v=IzHoNwlCGnE&list=PLJV_el3uVTsPMxPbjeX7PicgWbY7F8wW9

Thank you for your attention!