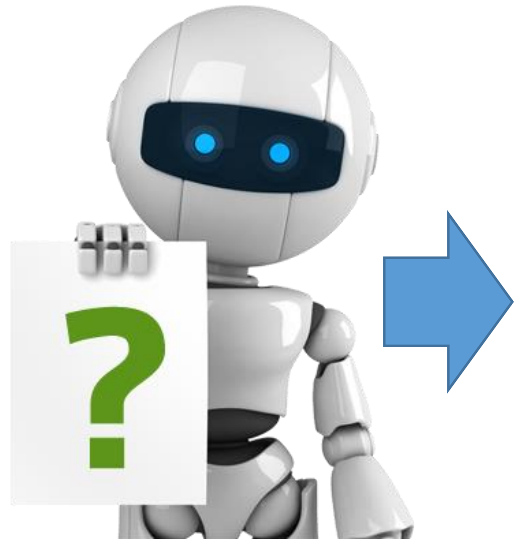
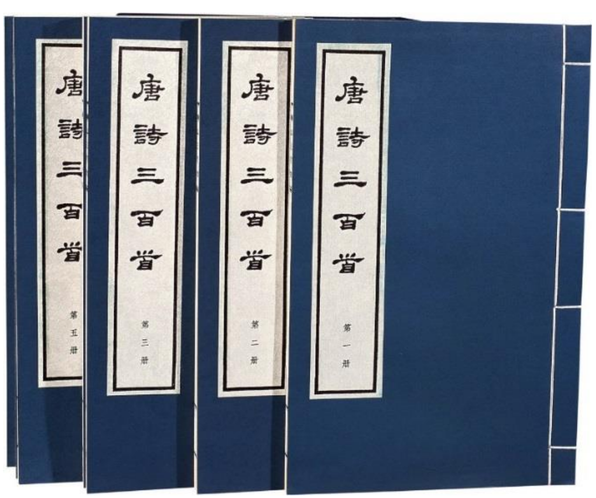
Unsupervised Learning: Generation

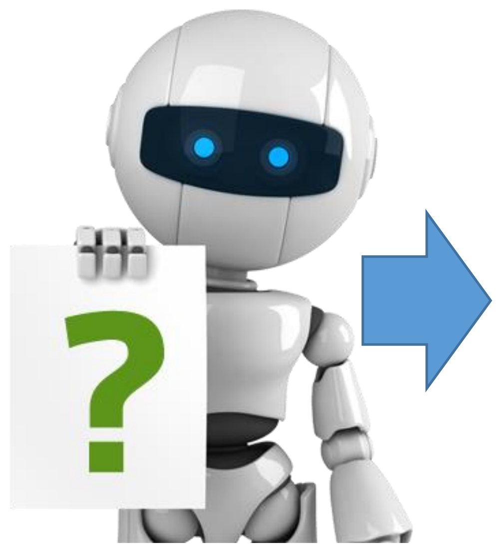
http://www.rb139.com/index.php?s=/Lot/44547

Creation



Drawing?

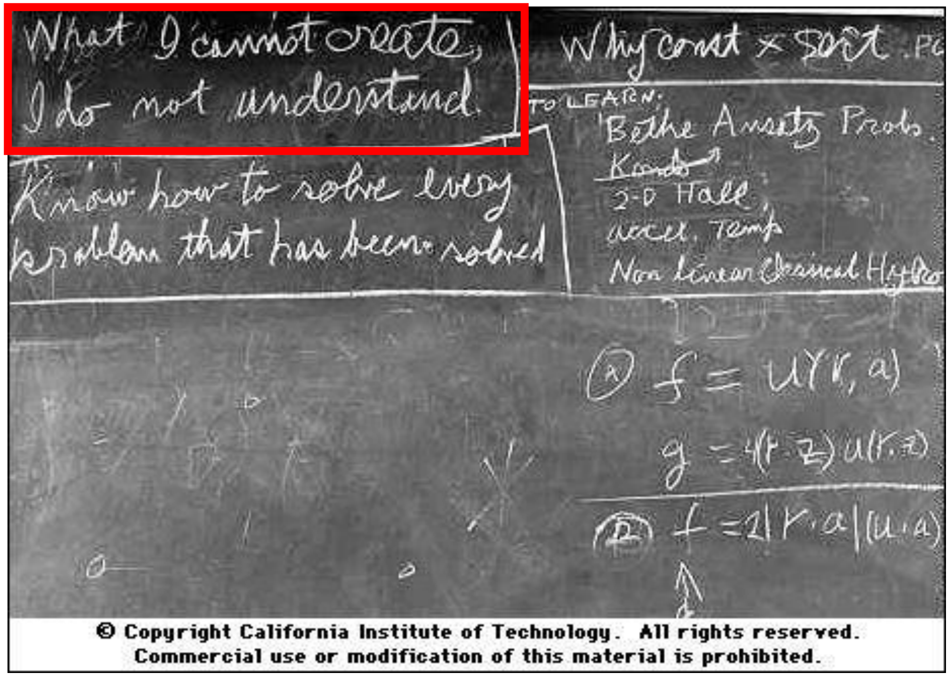




Writing Poems?

Creation

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

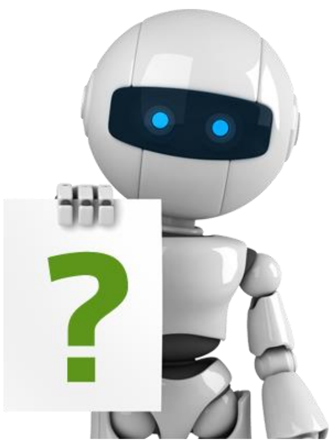


What I cannot create,
I do not understand.

Richard Feynman

https://www.quora.com/What-did-Richard-Feynman-mean-when-he-said-What-I-cannot-create-I-do-not-understand

Creation



Now

v.s.

In the future

Machine draws a cat

http://www.wikihow.com/Draw-a-Cat-Face

Generative Models

Component-by-component

Autoencoder

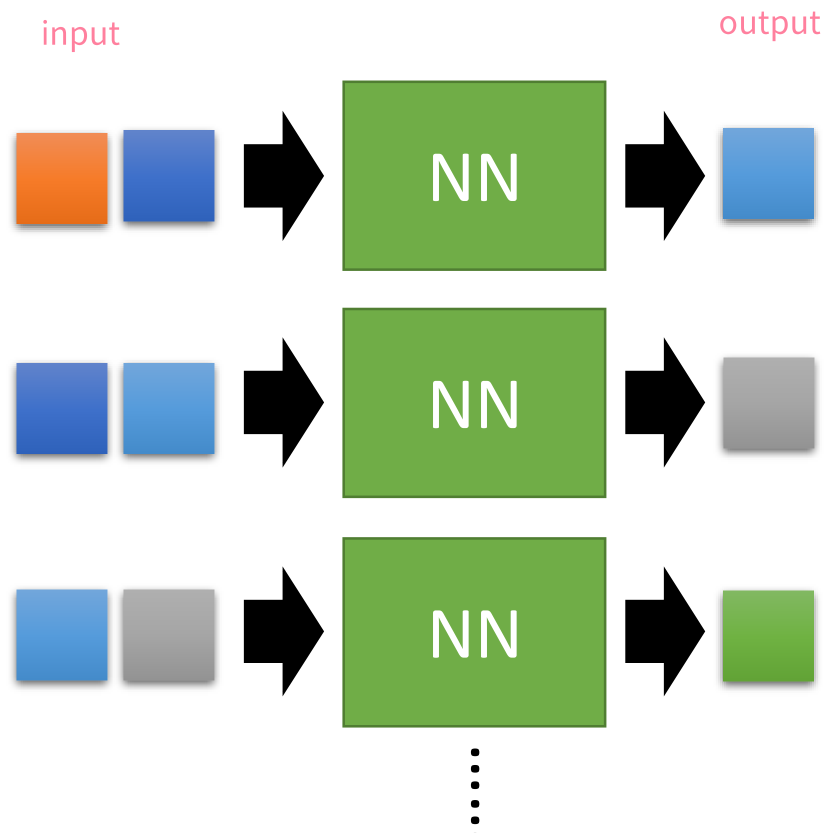
Generative Adversarial Network (GAN)

PixelRNN

Component-by-component

- Image generation

E.g. 3 x 3 images

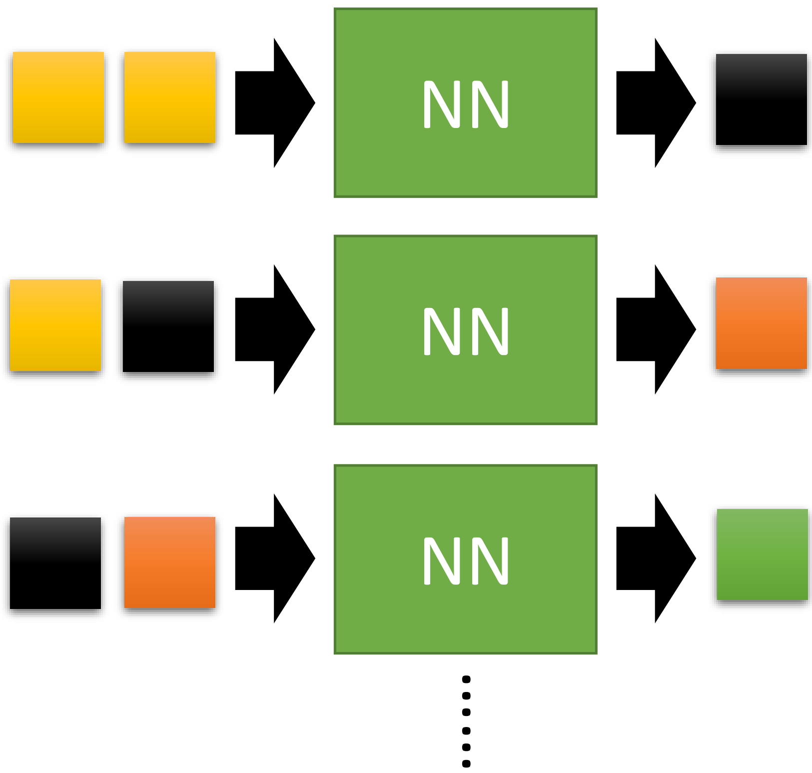


Can be trained just with a large collection of images without any annotation

Component-by-component

- Image generation

E.g. 3 x 3 images



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

Practicing Generation Models: Pokémon Creation

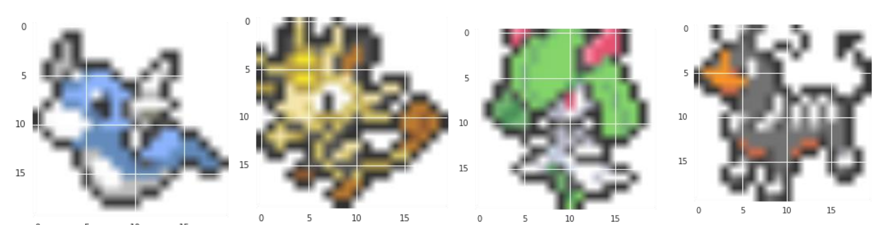
- Small images of 792 Pokémon's
  - Can machine learn to create new Pokémons?

<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



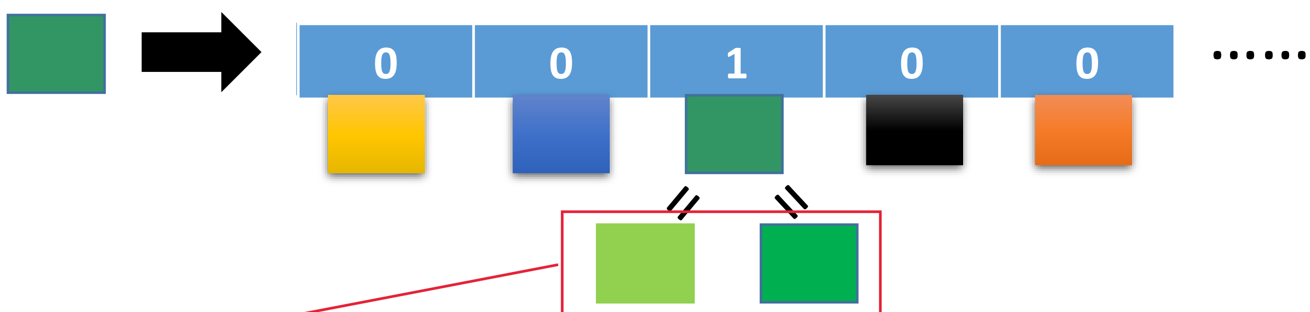
Practicing Generation Models: Pokémon Creation

- Tips (?)
  - ➢ Each pixel is represented by 3 numbers (corresponding to RGB)

R=50, G=150, B=100

  - ➢ Each pixel is represented by a 1-of-N encoding feature

類似 dummy coding



Clustering the similar color ➡ 167 colors in total

Practicing Generation Models: Pokémon Creation

- Original image (40 x 40): http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Pokemon_creation/image.rar
- Pixels (20 x 20): http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Pokemon_creation/pixel_color.txt
  - Each line corresponds to an image, and each number corresponds to a pixel
    - http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2016/Pokemon_creation/colormap.txt

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 19 41 34 0 0 19 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 44 74 44 51 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 21 80 80 81 0 0 0 0 0 0 0 0
0 0 0 0 0 1 2 3 18 35 22 0 5 2 0 0 0 0 0 0
93 94 93 93 85 95 38 96 97 98 99 99 67 99 9
0 0 0 0 0 0 1 106 106 106 106 106 61 107 0
```

……

0
1
2

```
255 255 255
53 53 53
49 49 49
186 186 186
51 51 51
54 54 54
187 187 187
83 83 83
50 51 52
251 251 251
52 52 52
```

⋮

- Following experiment: 1-layer LSTM, 512 cells

Real Pokémon

Never seen by machine!

Cover 50%

It is difficult to evaluate generation.

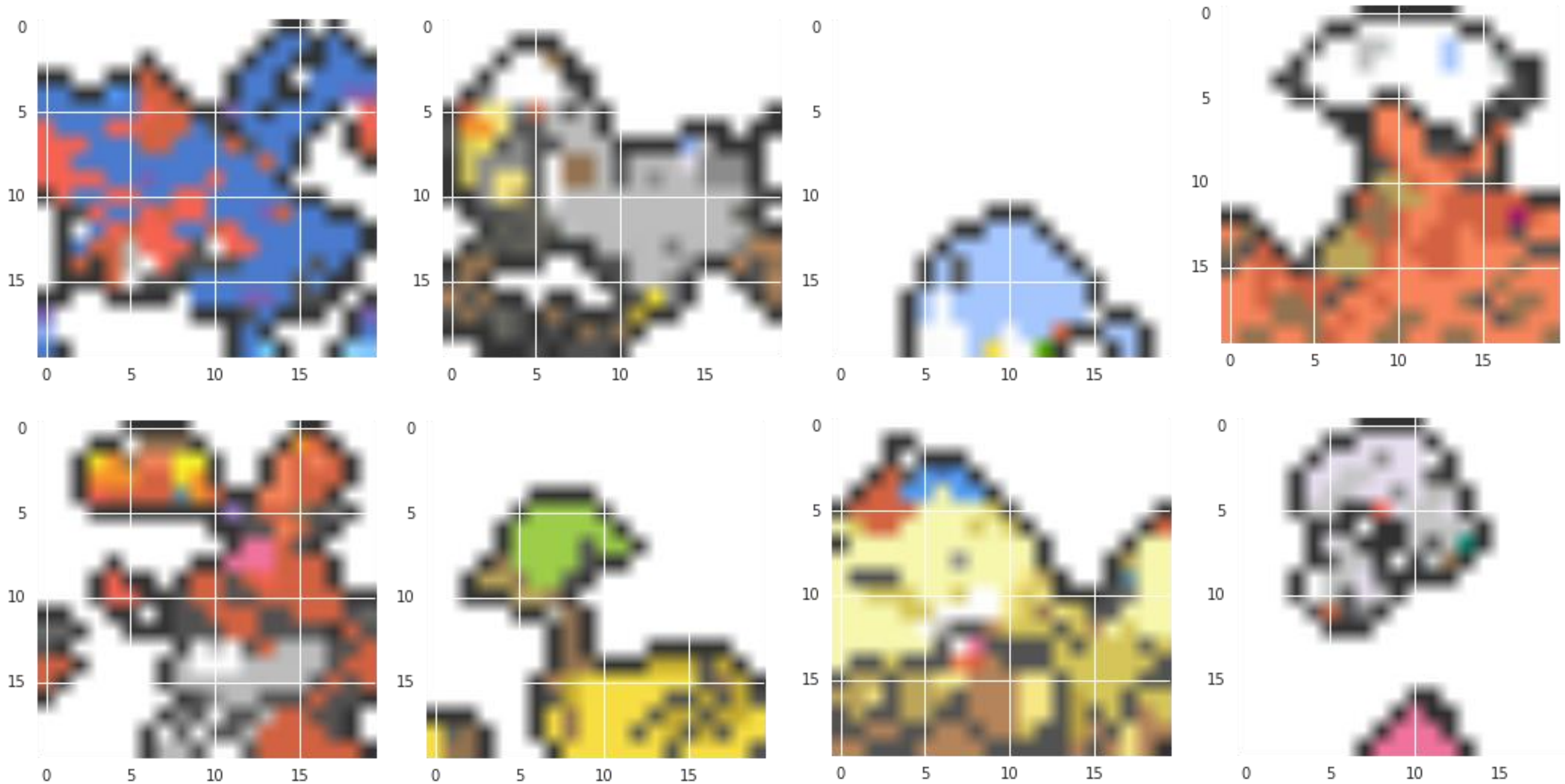
畫的跟原本不一樣，不代表畫錯

Cover 75%

Pokémon Creation

Drawing from scratch

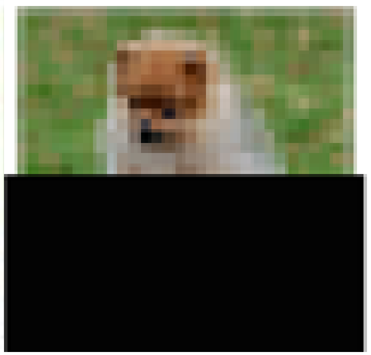
Need some randomness



PixelRNN

Ref: Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu, Pixel Recurrent Neural Networks, arXiv preprint, 2016

To create an image, generating a pixel each time

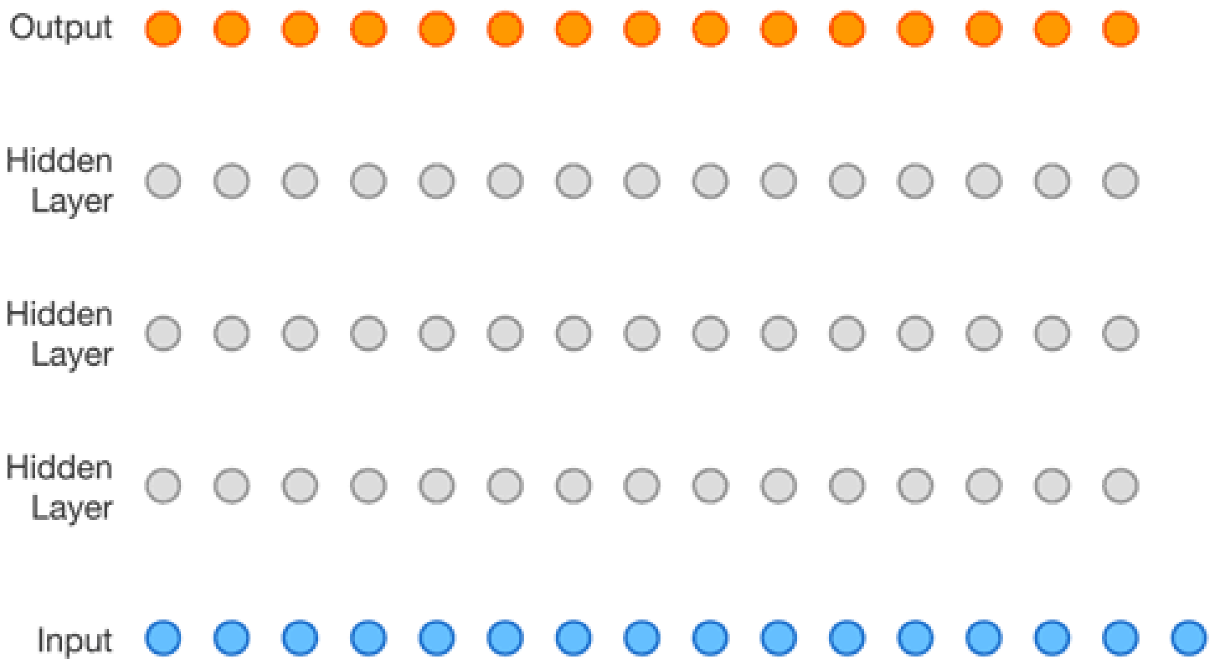


Real World

遮住下半身，機器畫出下面的圖

More than images ……

合出一段聲音



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

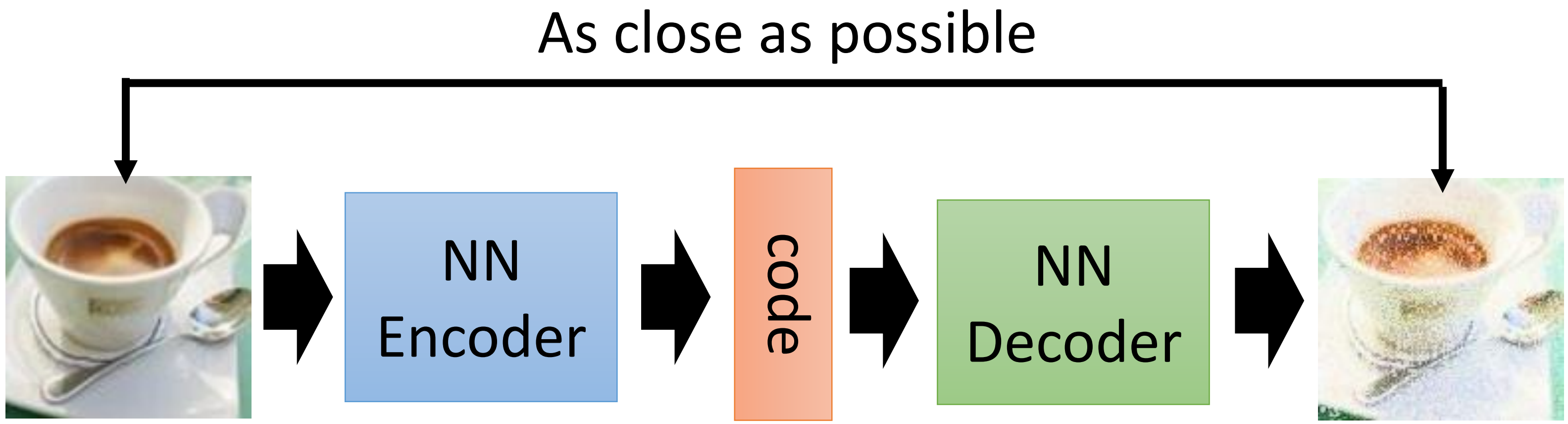
Generative Models

Component-by-component

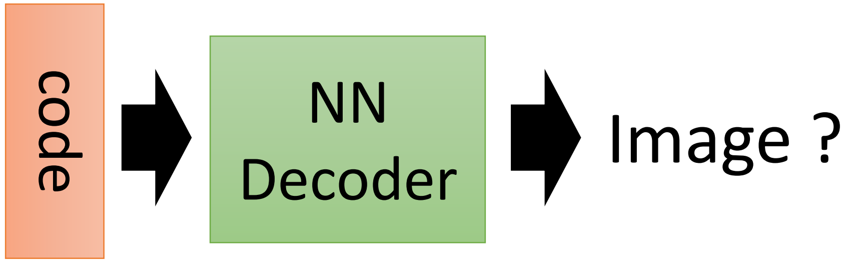
Autoencoder

Generative Adversarial Network (GAN)

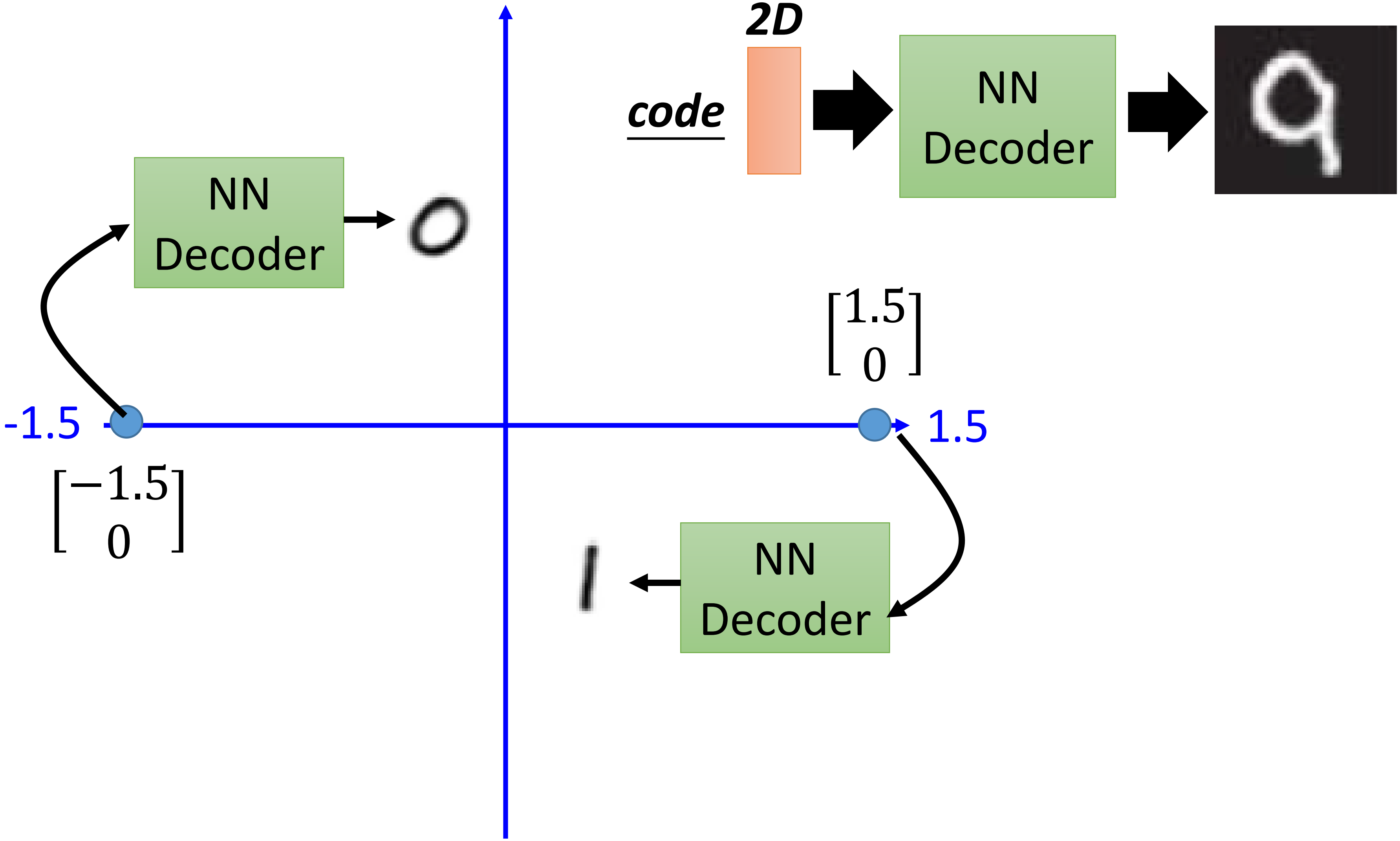
Auto-encoder



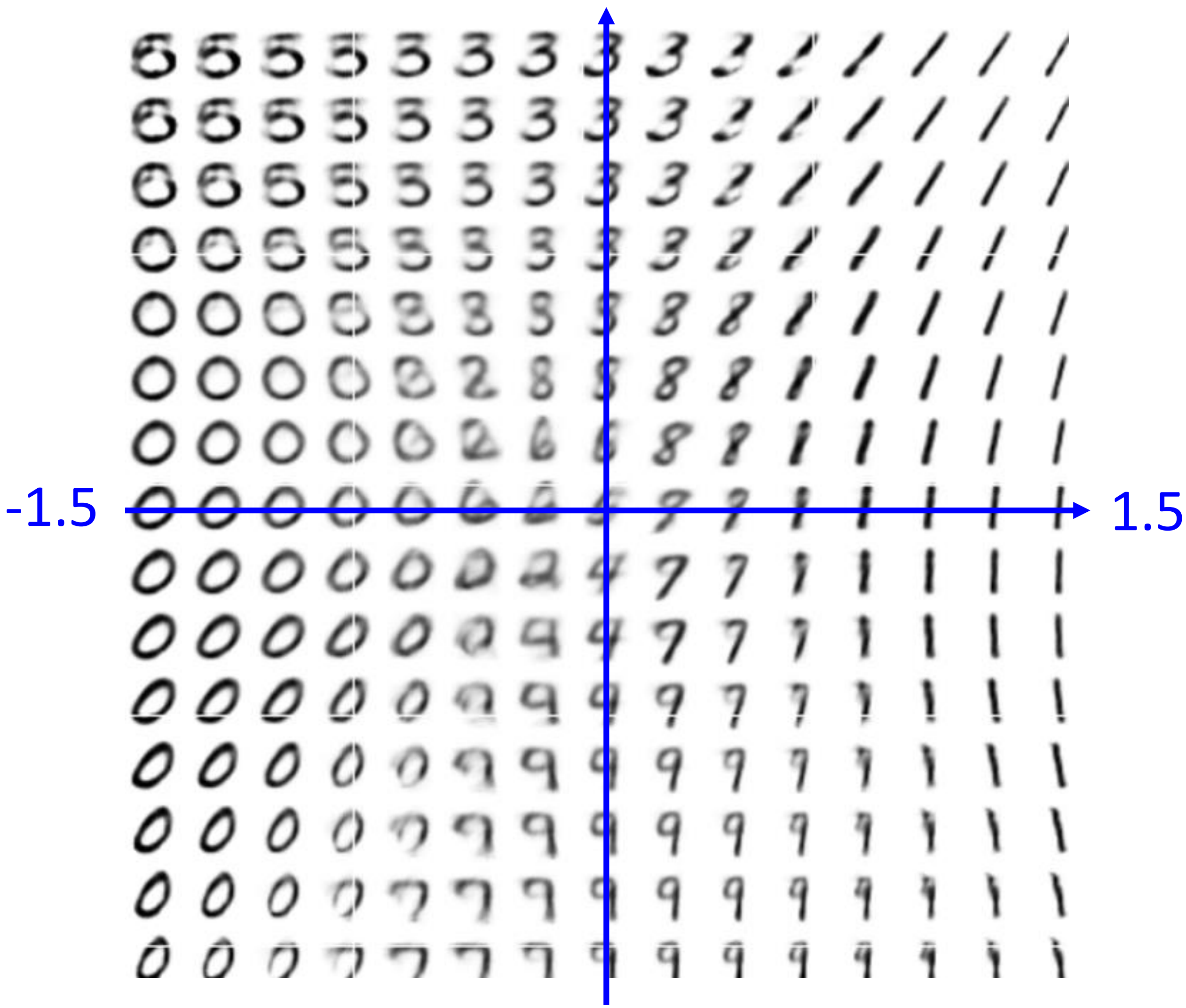
Randomly generate
a vector as code



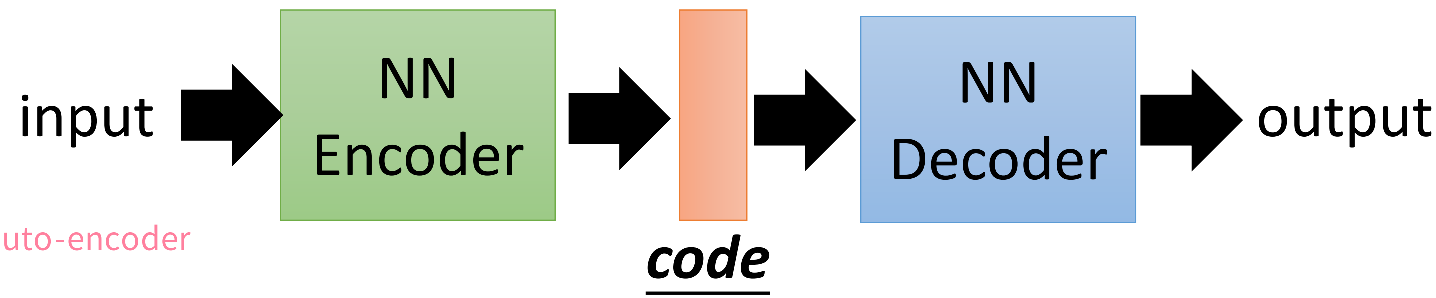
Review: Auto-encoder



Review: Auto-encoder

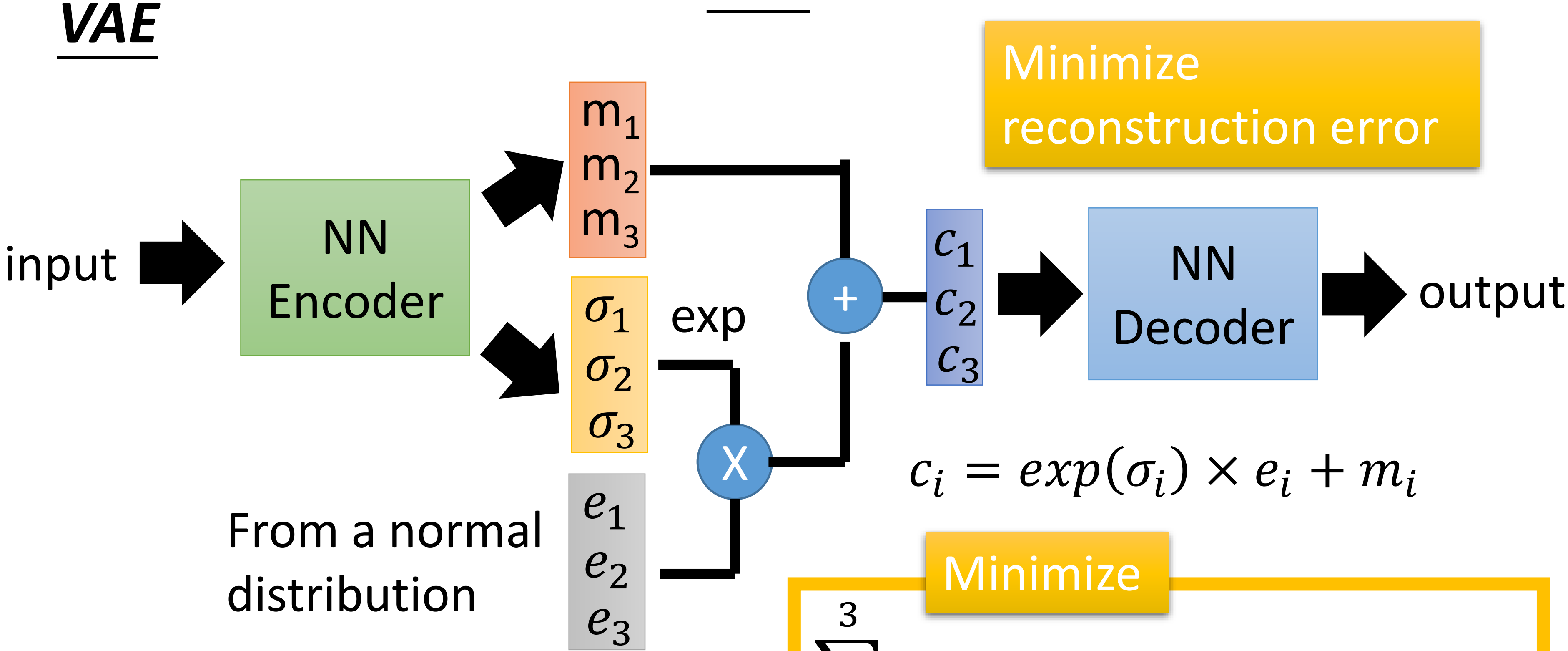


Auto-encoder



variant auto-encoder

VAE



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

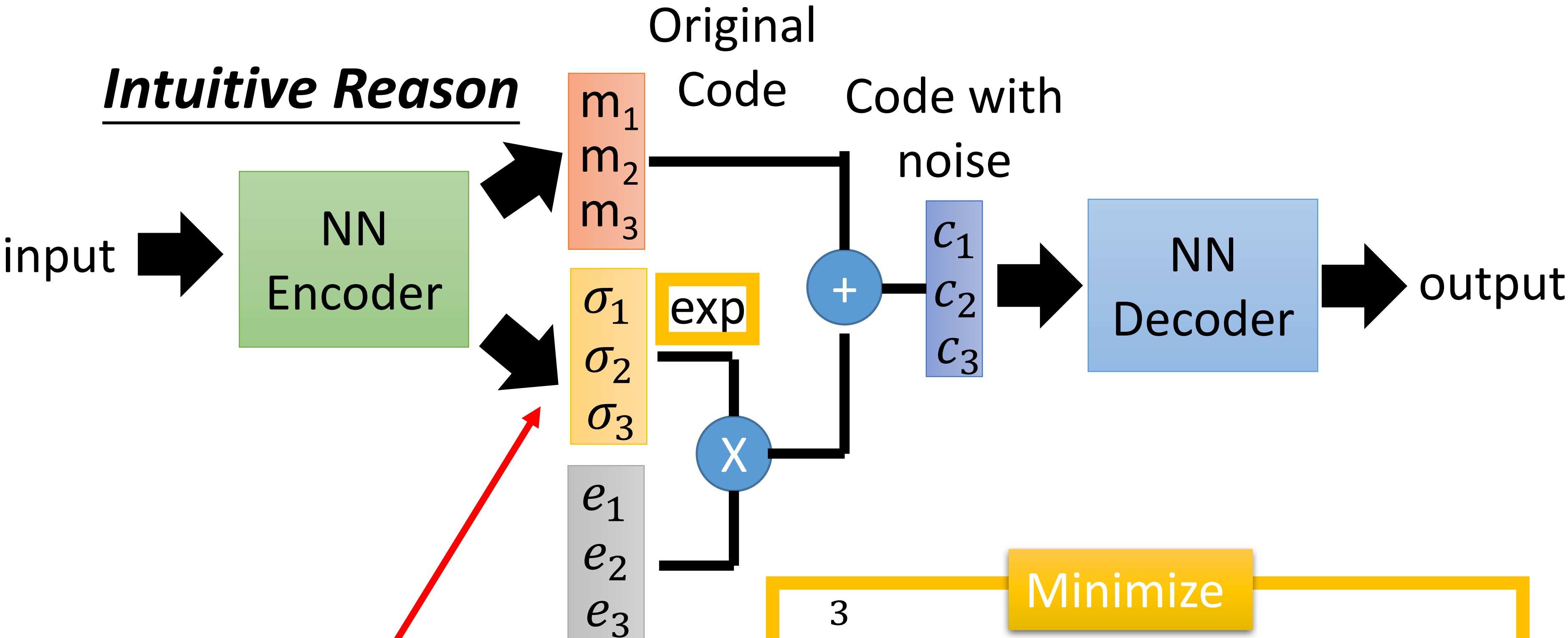
Minimize

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

Why VAE?

What will happen if we only minimize reconstruction error?

Intuitive Reason



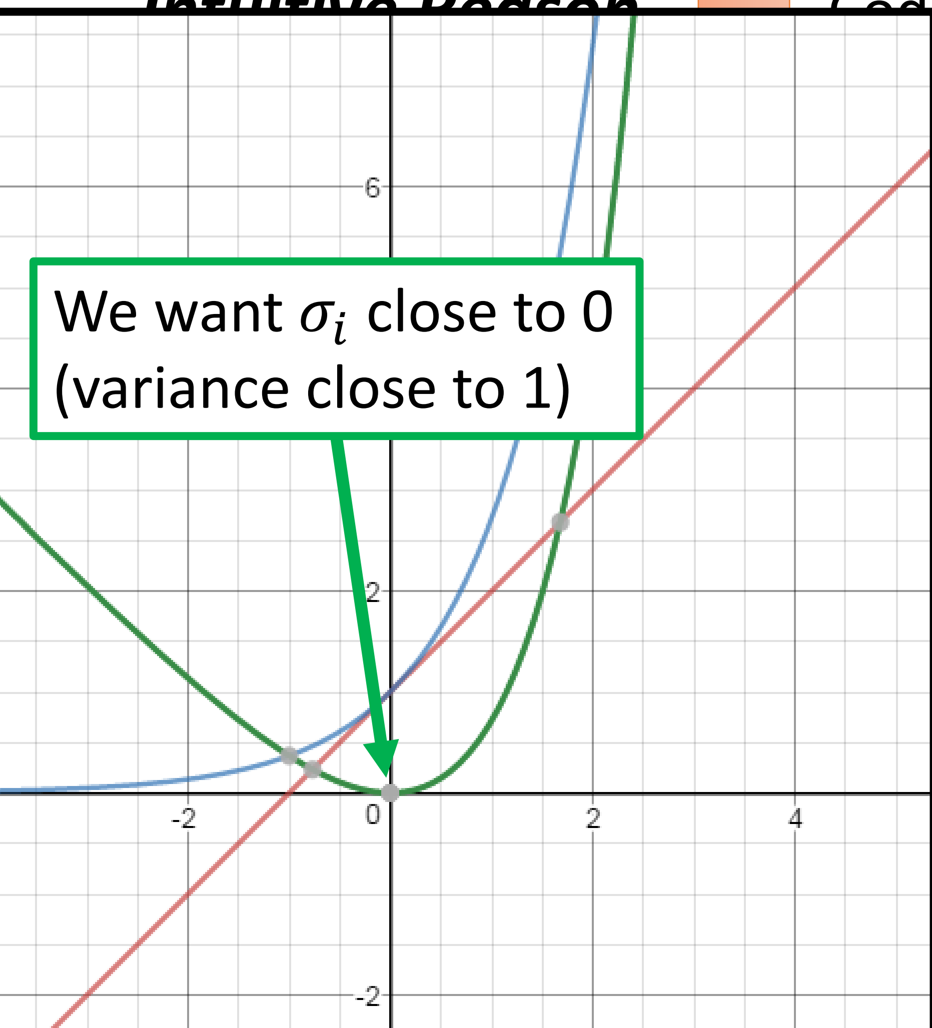
The variance of noise is automatically learned

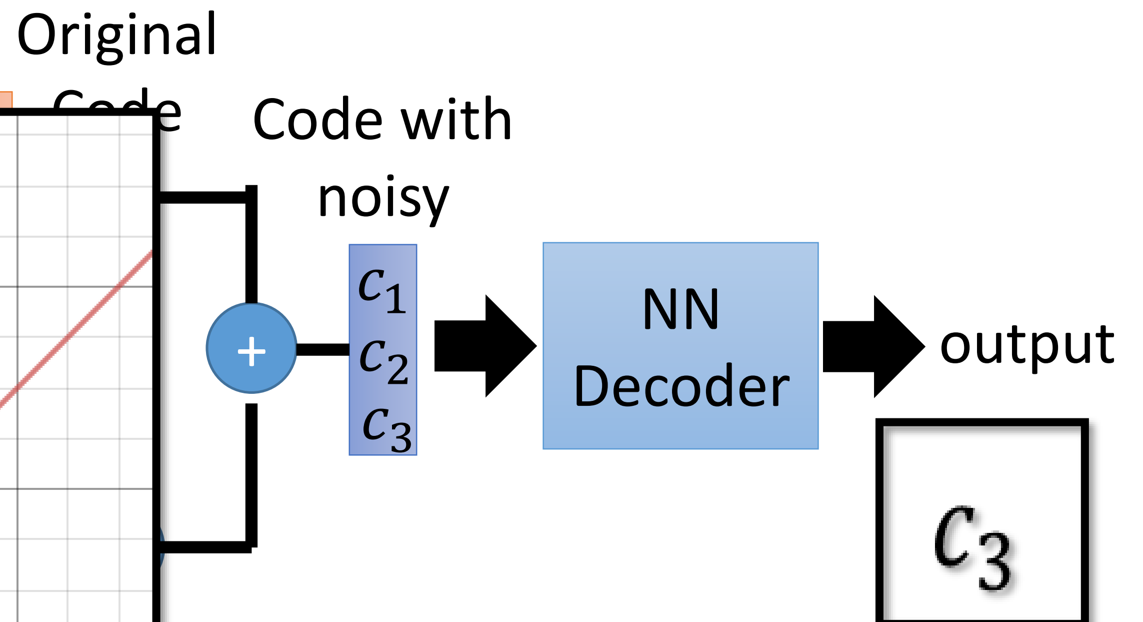
Minimize

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

Why VAE?

What will happen if we only minimize reconstruction error?





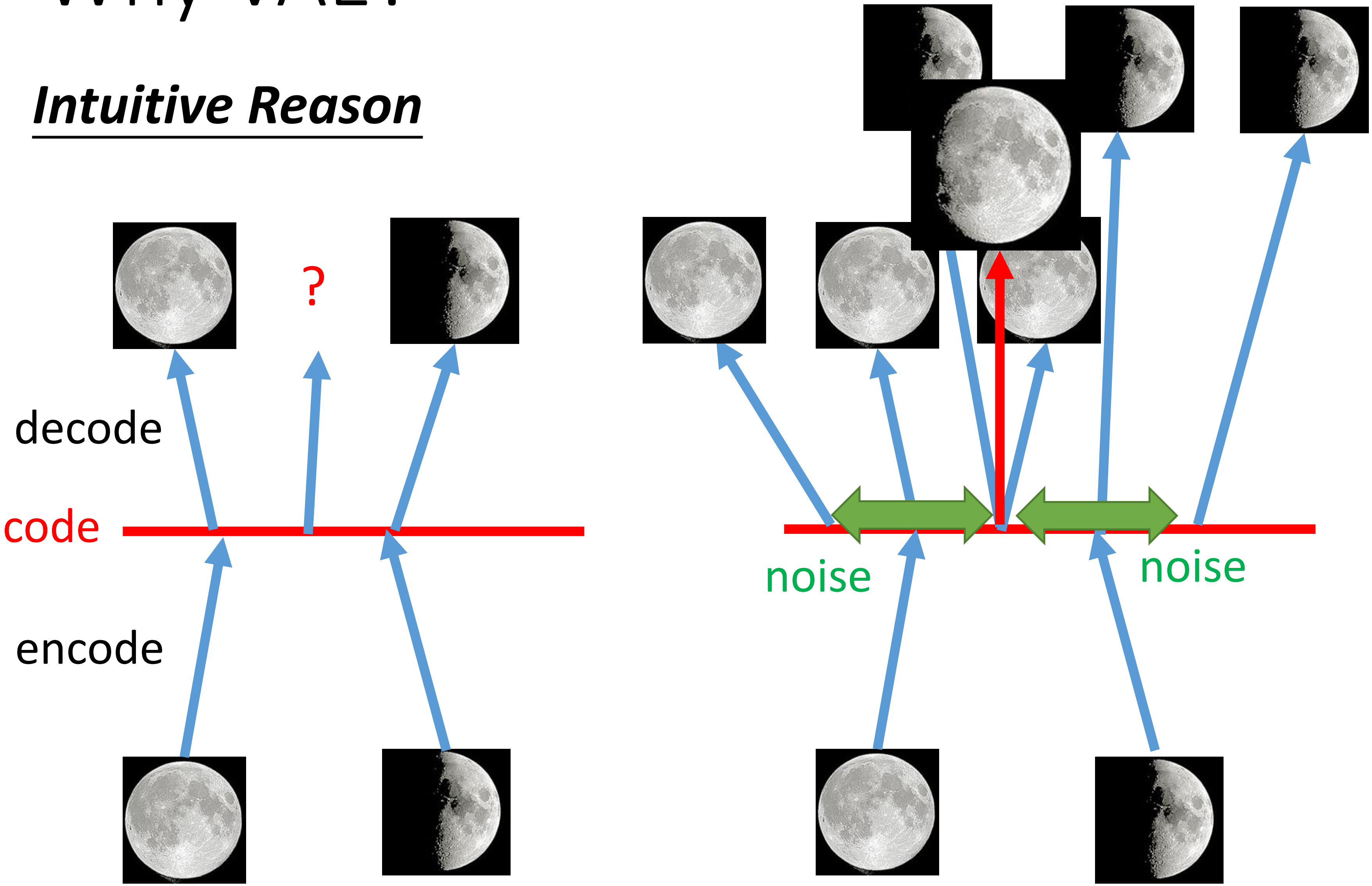
Minimize

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

L2 regularization

Why VAE?

Intuitive Reason



Warning of Math

Gaussian Mixture Model

How to sample?

$$m \sim P(m) \text{ (multinomial)}$$

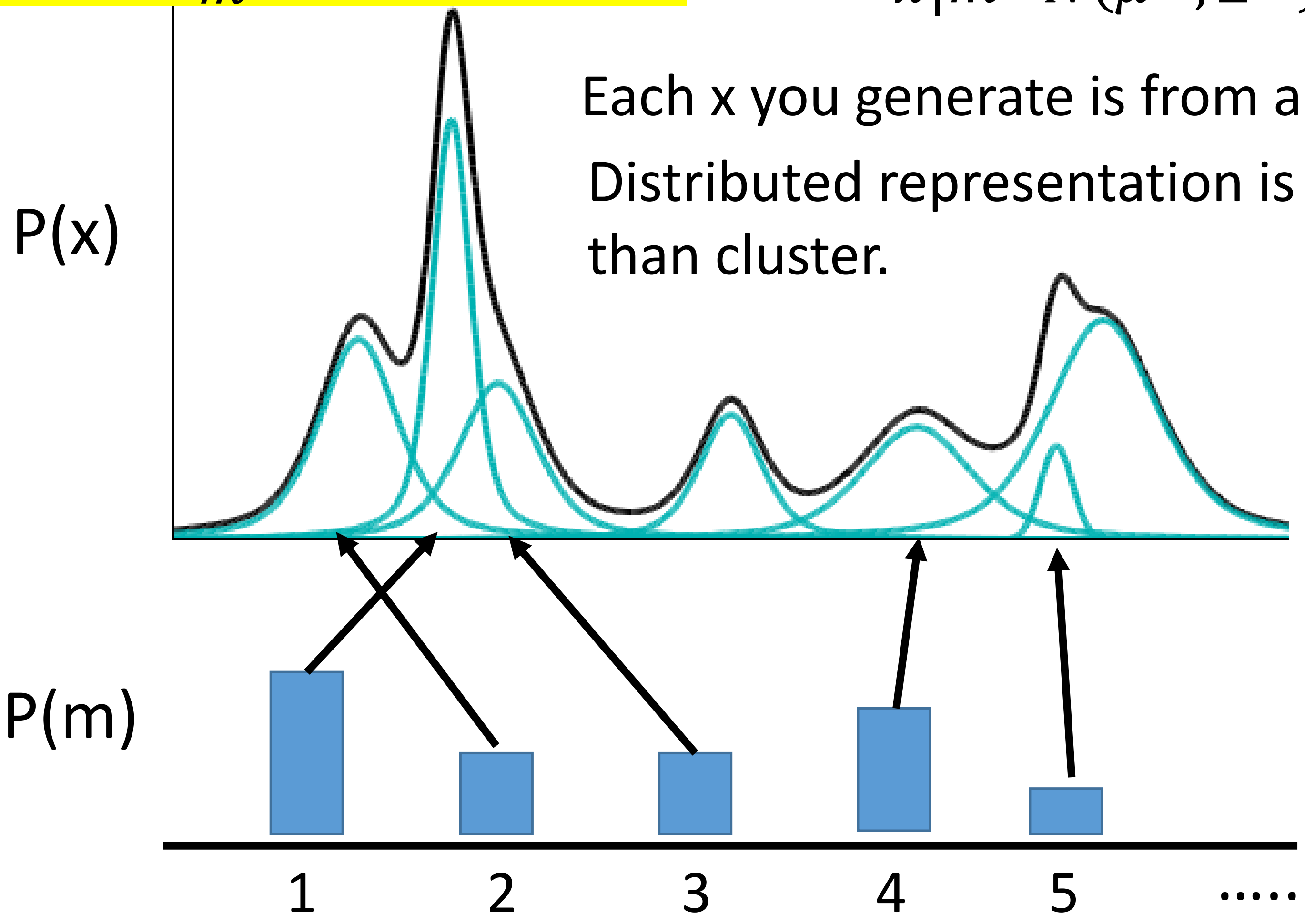
m is an integer

$$x|m \sim N(\mu^m, \Sigma^m)$$

$$P(x) = \sum_m P(m)P(x|m)$$

Each x you generate is from a mixture

Distributed representation is better than cluster.



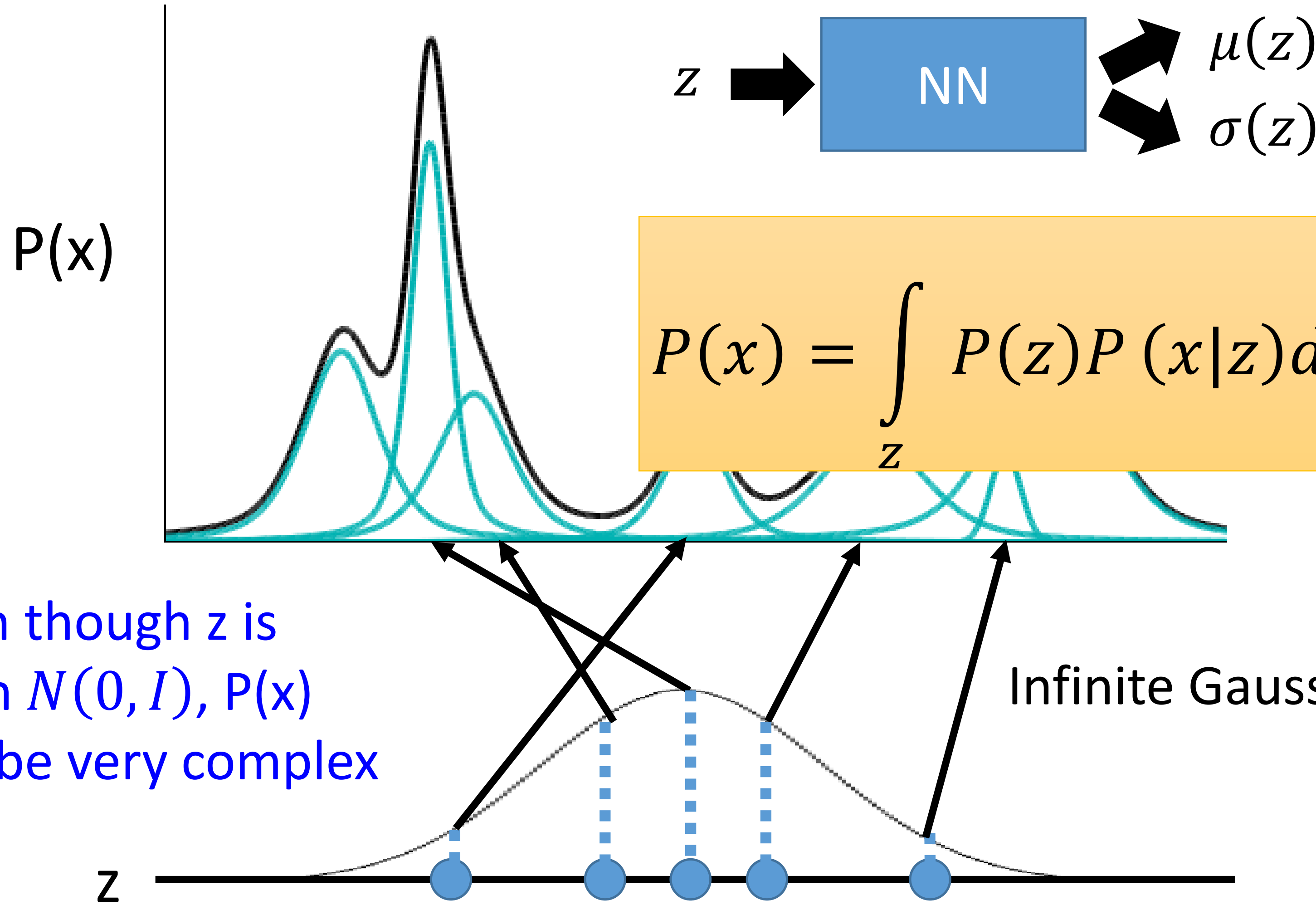
VAE

$z \sim N(0, I)$

z is a vector from normal distribution

$x|z \sim N\big(\mu(z), \sigma(z)\big)$

Each dimension of z represents an attribute



$$P(x) = \int_z P(z) P(x|z) dz$$

Infinite Gaussian

Even though z is from $N(0, I)$, P(x) can be very complex

Maximizing Likelihood

$$P(x) = \int_z P(z)P(x|z)dz$$

P(z) is normal distribution

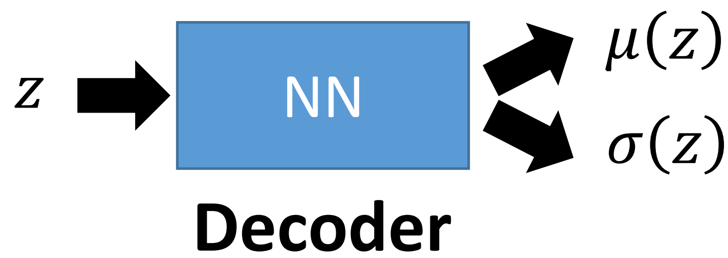
$x|z \sim N\big(\mu(z), \sigma(z)\big)$

$\mu(z), \sigma(z)$ is going to be estimated

$$L = \sum_x logP(x)$$

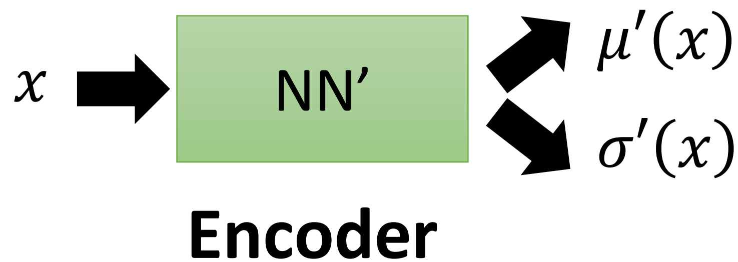
Maximizing the likelihood of the observed x

Tuning the parameters to maximize likelihood L



We need another distribution q(z|x)

$$z|x \sim N\big(\mu'(x), \sigma'(x)\big)$$



Maximizing Likelihood

P(z) is normal distribution

$x|z \sim N\big(\mu(z), \sigma(z)\big)$

$\mu(z), \sigma(z)$ is going to be estimated

$$P(x) = \int_z P(z)P(x|z)dz$$

$$L = \sum_x logP(x)$$

Maximizing the likelihood of the observed x

$$logP(x) = \int_z q(z|x)logP(x)dz$$

q(z|x) can be any distribution

$$= \int_z q(z|x)log\left(\frac{P(z,x)}{P(z|x)}\right)dz = \int_z q(z|x)log\left(\frac{P(z,x)}{q(z|x)}\frac{q(z|x)}{P(z|x)}\right)dz$$

$$= \int_z q(z|x)log\left(\frac{P(z,x)}{q(z|x)}\right)dz + \underbrace{\int_z q(z|x)log\left(\frac{q(z|x)}{P(z|x)}\right)dz}_{KL\left(q(z|x)||P(z|x)\right)}$$

≥ 0

$$\geq \int_z q(z|x)log\left(\frac{P(x|z)P(z)}{q(z|x)}\right)dz$$

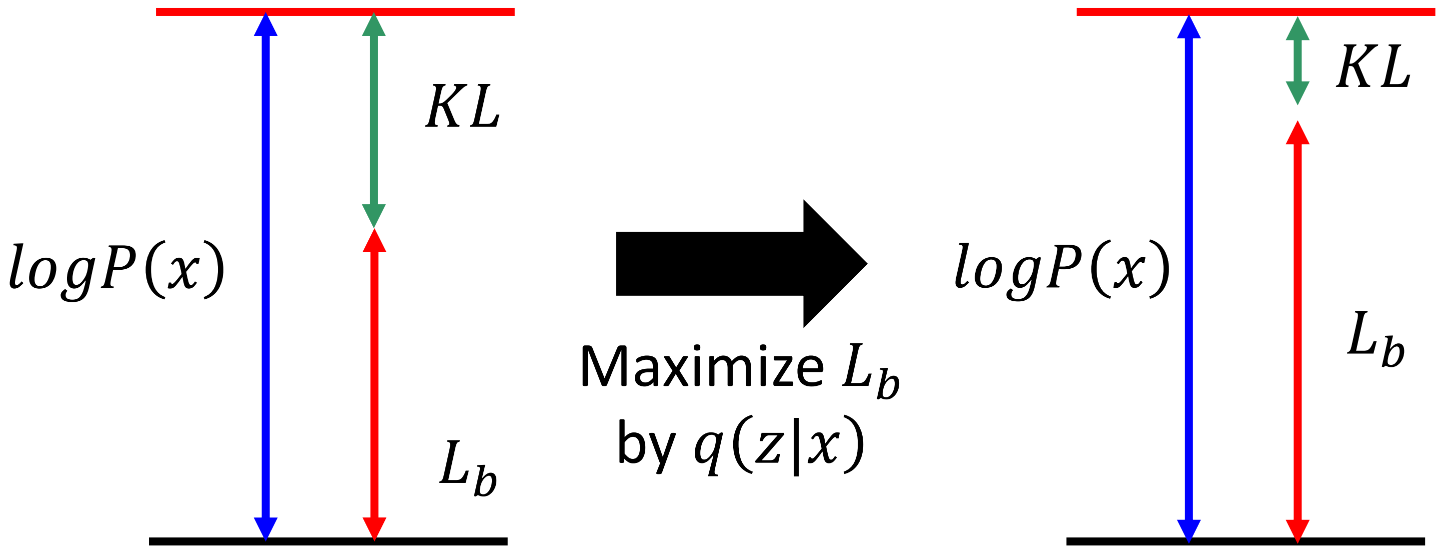
lower bound L_b

Maximizing Likelihood

$$logP(x) = L_b + KL\big(q(z|x)||P(z|x)\big)$$

$$L_b = \int_z q(z|x)log\left(\frac{P(x|z)P(z)}{q(z|x)}\right)dz$$

Find $P(x|z)$ and $q(z|x)$ maximizing L_b



$q(z|x)$ will be an approximation of $p(z|x)$ in the end

Maximizing Likelihood

$$P(x) = \int_z P(z)P\left(x|z\right)dz$$

P(z) is normal distribution

$$x|z \sim N\big(\mu(z), \sigma(z)\big)$$

$\mu(z), \sigma(z)$ is going to be estimated

$$L = \sum_x logP(x)$$

Maximizing the likelihood of the observed x

$$L_b = \int_z q(z|x)log\left(\frac{P(z,x)}{q(z|x)}\right)dz = \int_z q(z|x)log\left(\frac{P(x|z)P(z)}{q(z|x)}\right)dz$$

$$= \underbrace{\int_z \boxed{q(z|x)}log\left(\frac{P(z)}{\boxed{q(z|x)}}\right)dz}_{-KL\left(q(z|x)||P(z)\right)} + \int_z q(z|x)logP(x|z)dz$$

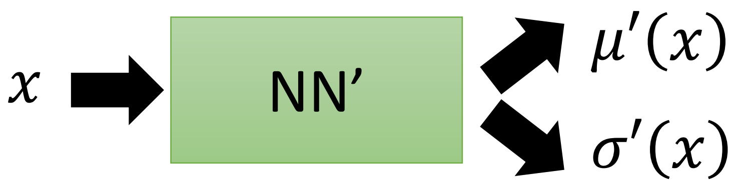
$$z|x \sim N\big(\mu'(x), \sigma'(x)\big)$$

Connection with Network

Minimize

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

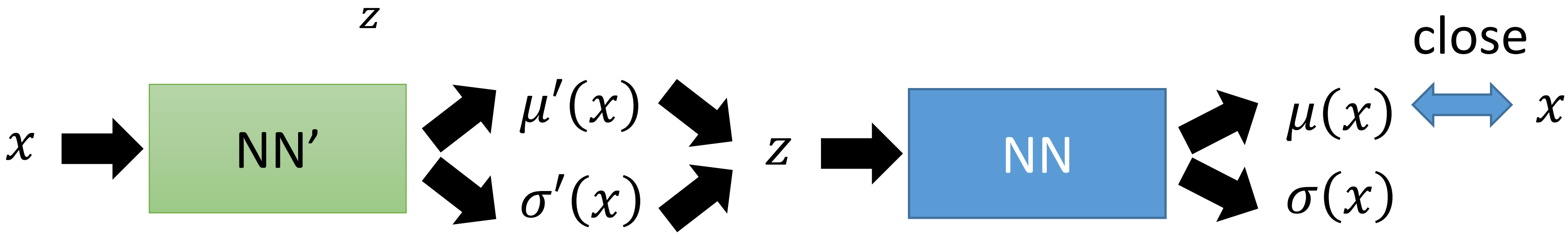
Minimizing $KL\big(q(z|x)||P(z)\big)$



(Refer to the Appendix B of the original VAE paper)

Maximizing

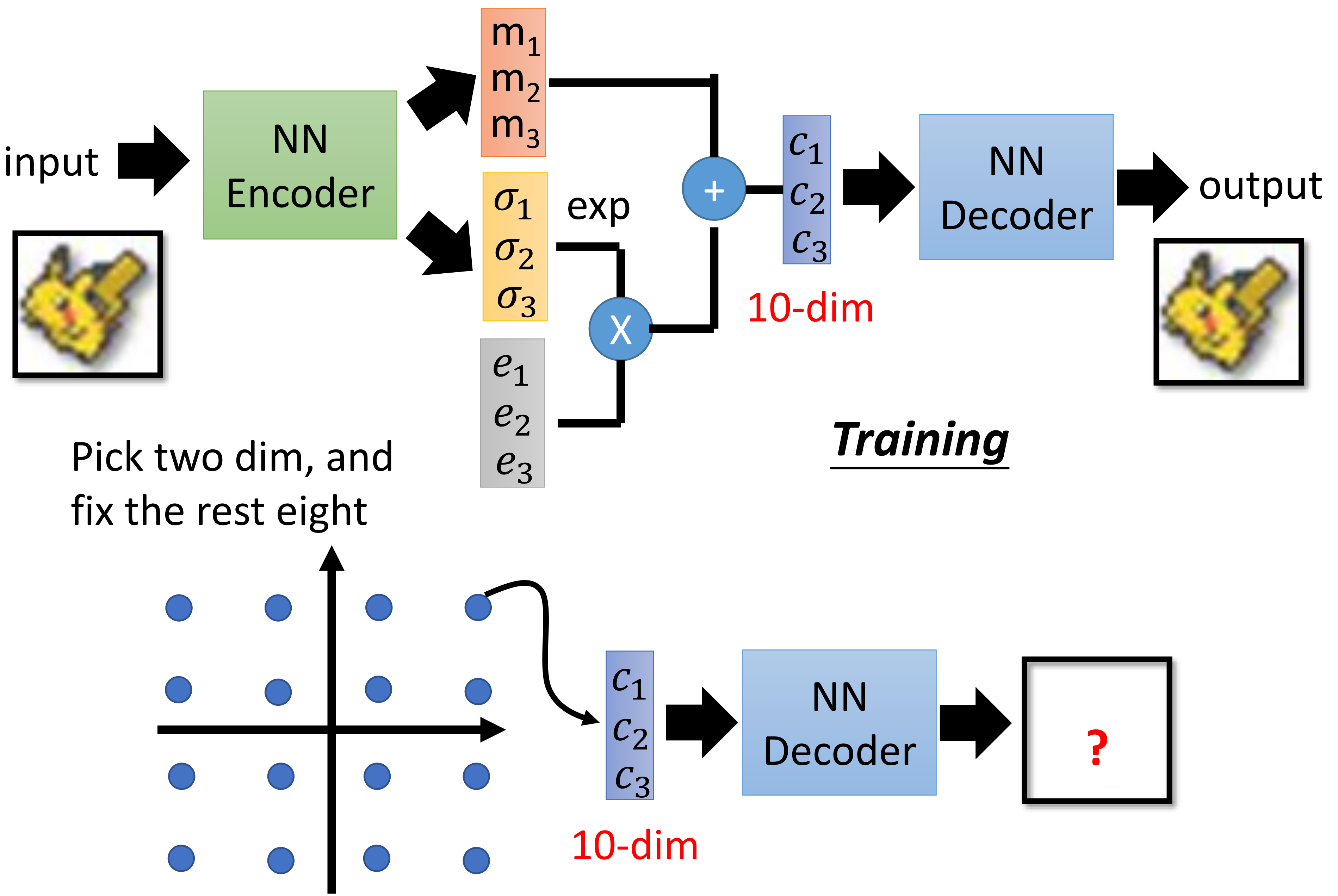
$$\int_z q(z|x) logP(x|z)dz \ = E_{q(z|x)}[logP(x|z)]$$

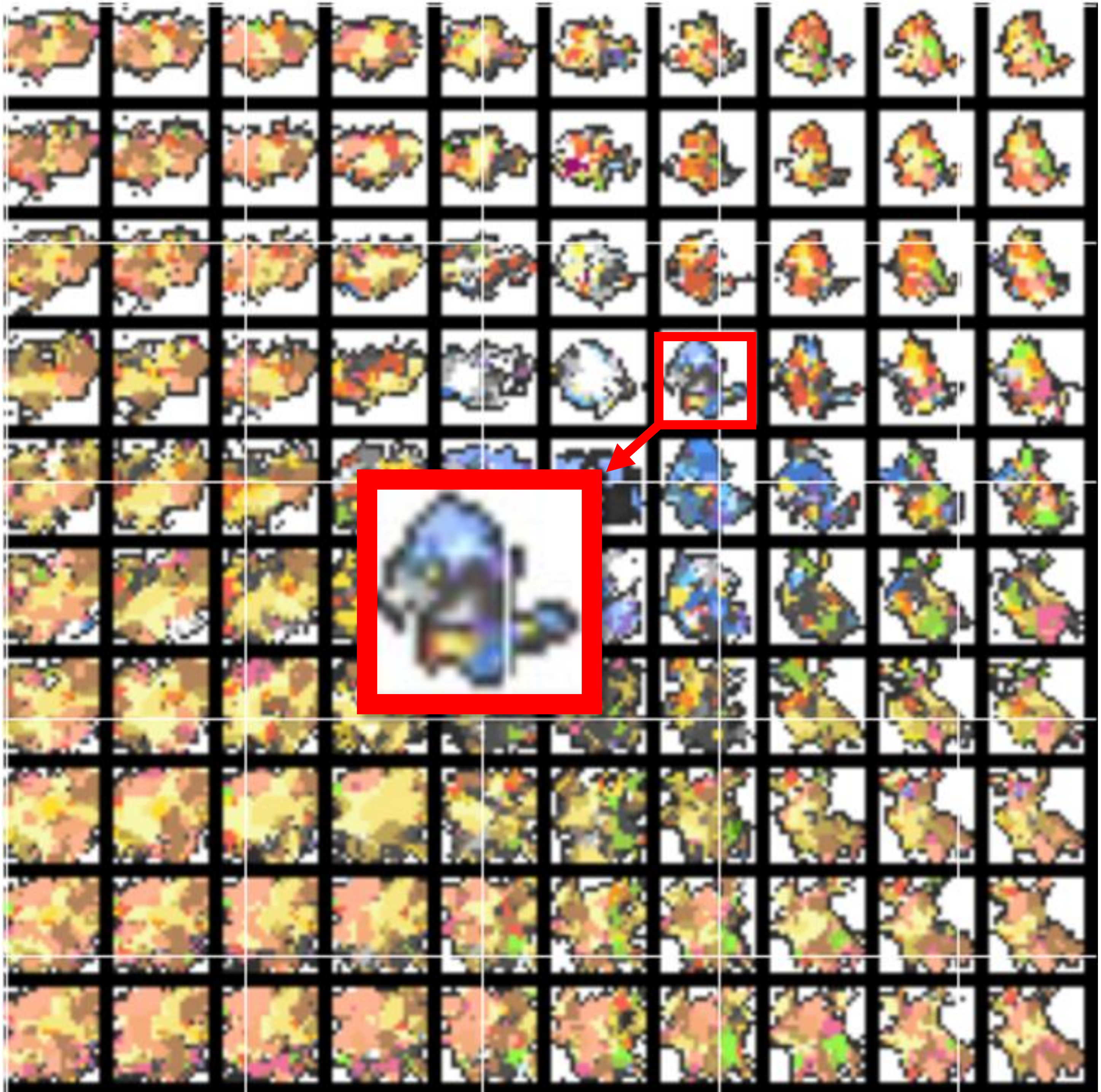


This is the auto-encoder

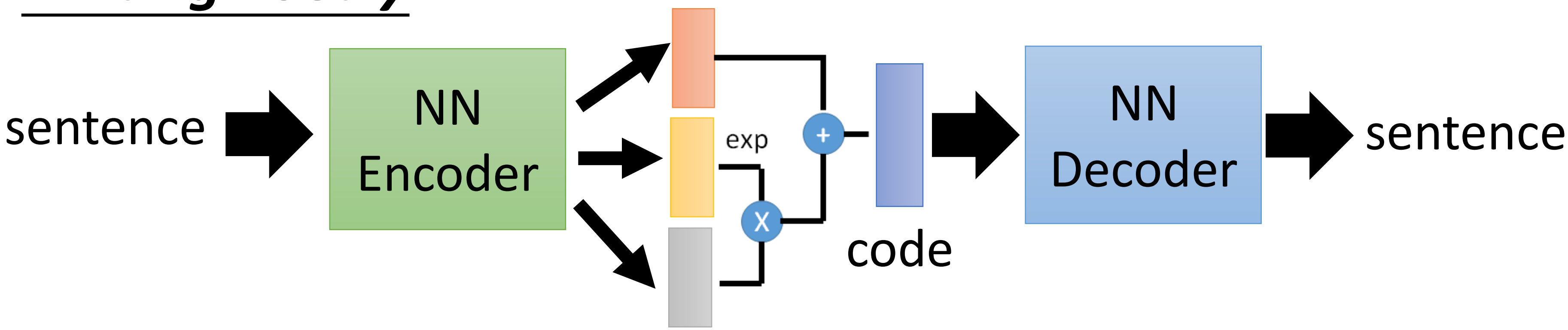
End of Warning

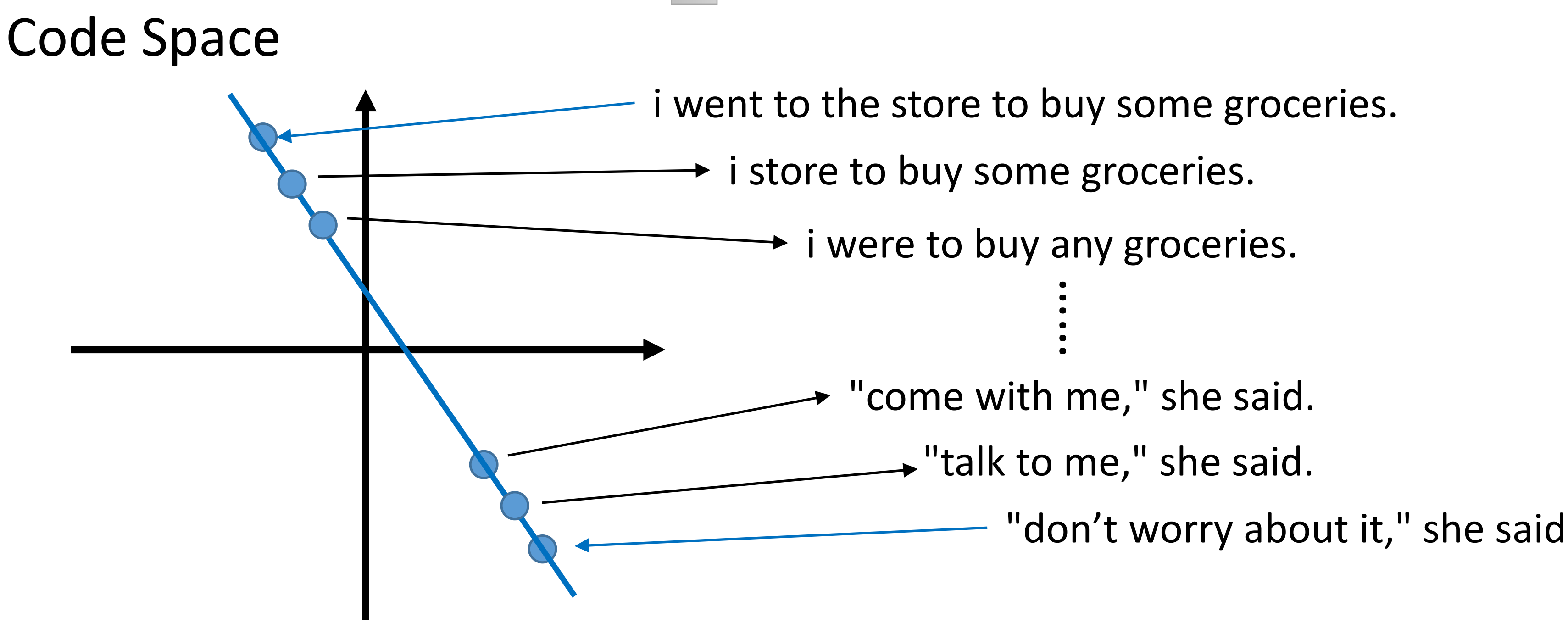
Pokémon Creation





Writing Poetry

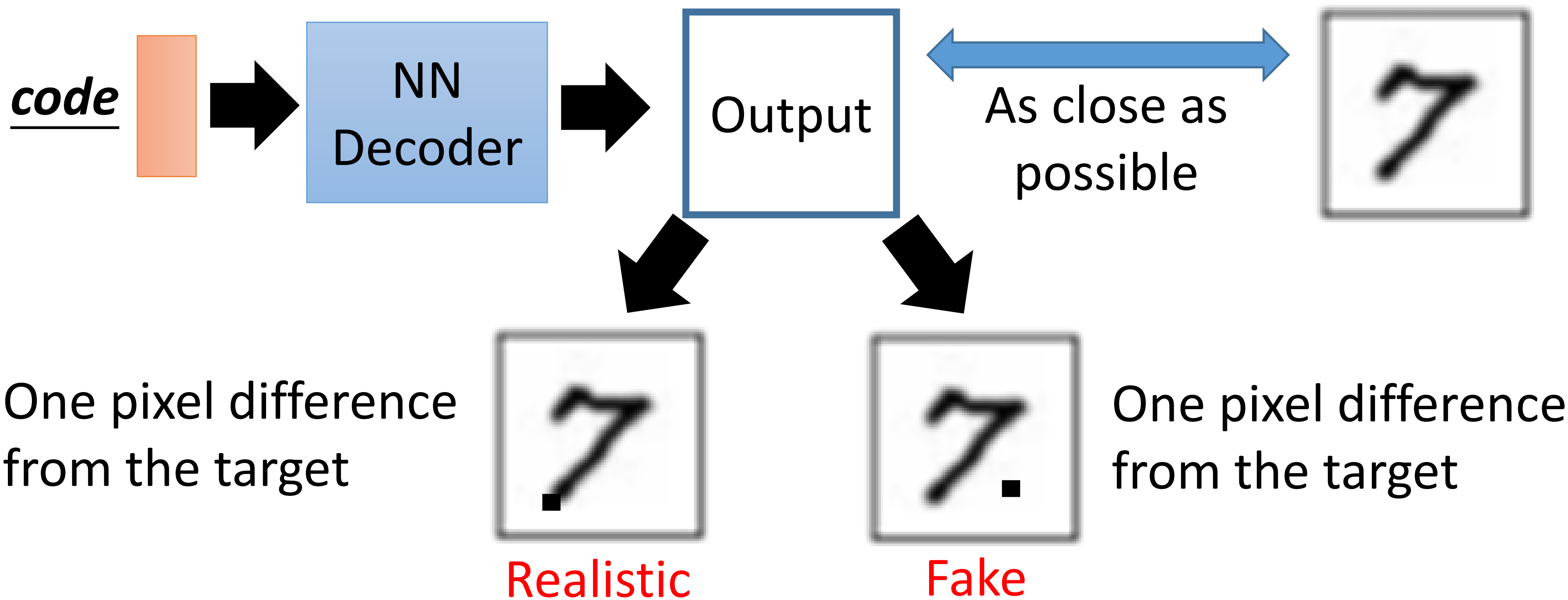




Ref: http://www.wired.co.uk/article/google-artificial-intelligence-poetry
Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, Samy Bengio, Generating Sentences from a Continuous Space, arXiv prepring, 2015

Problems of VAE

- It does not really try to simulate real images



VAE may just memorize the existing images, instead of generating new images

VAE從沒產生過新的image，只是模仿或linear combination原有的image。所以後來發展了GNN

Generative Models

Component-by-component

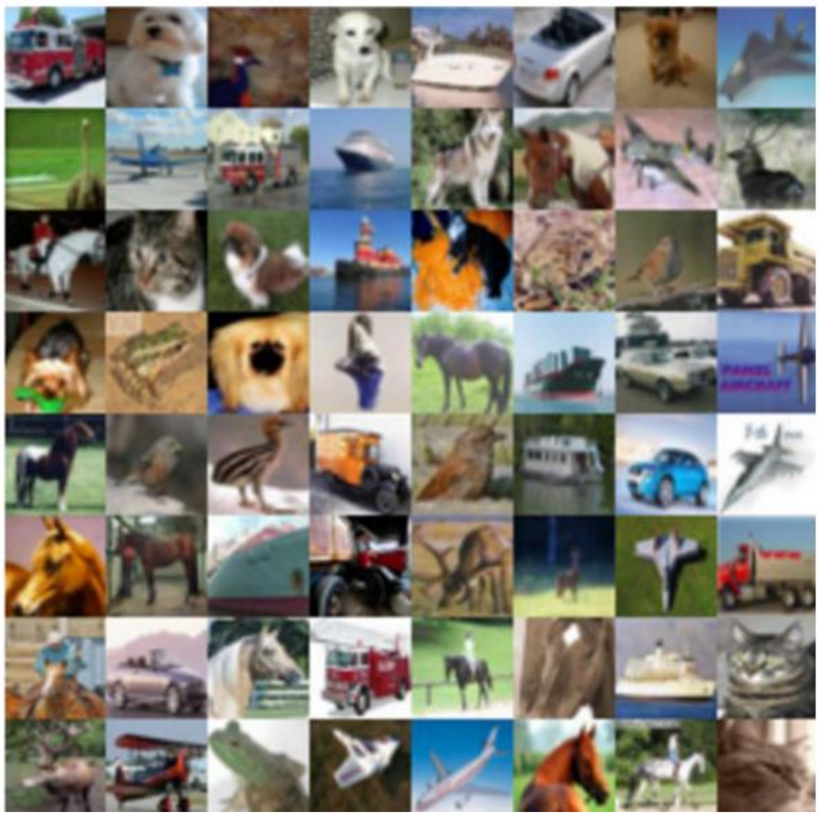
Autoencoder

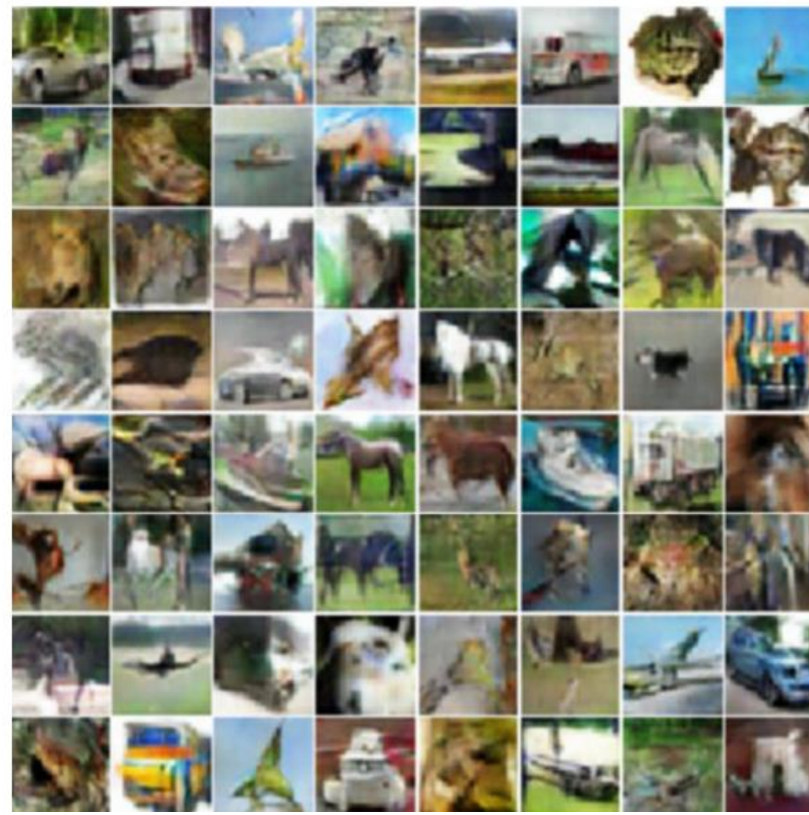
Generative Adversarial Network (GAN)

Ian J. Good fellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks, arXiv preprint 2014

Cifar-10

- Which one is machine-generated?





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

Yann LeCun’s comment

What are some recent and potentially upcoming breakthroughs in unsupervised 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 Huang Xiao

Adversarial training is the coolest thing since sliced bread.

I’ve listed a bunch of relevant papers in a previous answer.

Expect more impressive results with this technique in the coming years.

What’s missing at the moment is a good understanding of it so we can make it work reliably. It’s very finicky. Sort of like ConvNet were in the 1990s, when I had the reputation of being the only person who could make them work (which wasn’t true).

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

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

擬態的演化

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

generator

Brown

Kallima inachus

veins

Butterflies are not brown

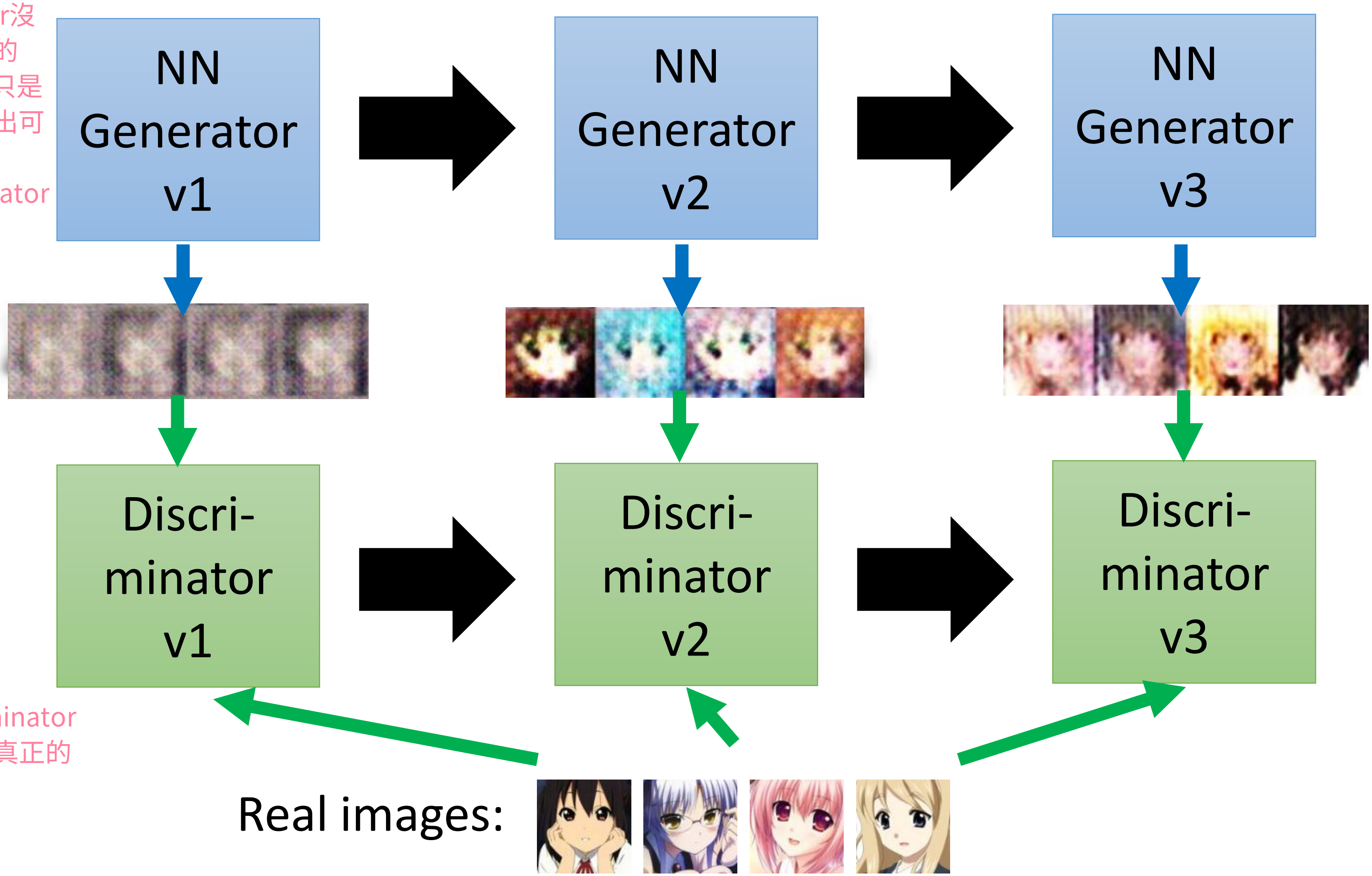
Butterflies do not have veins

....

discriminator

The evolution of generation

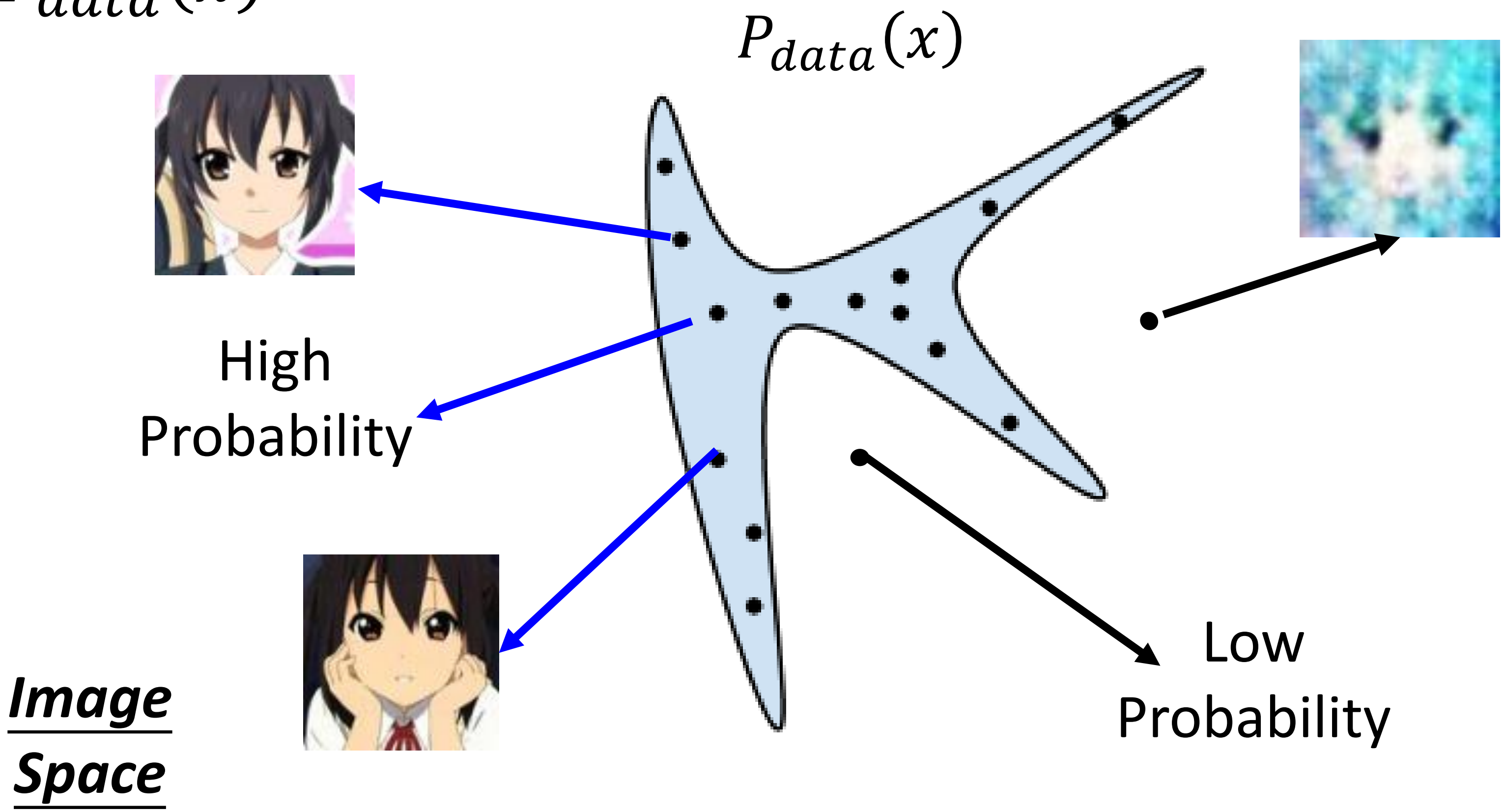
generator沒看過真正的image，只是想著要做出可以騙過discriminator的image



discriminator有看過真正的image

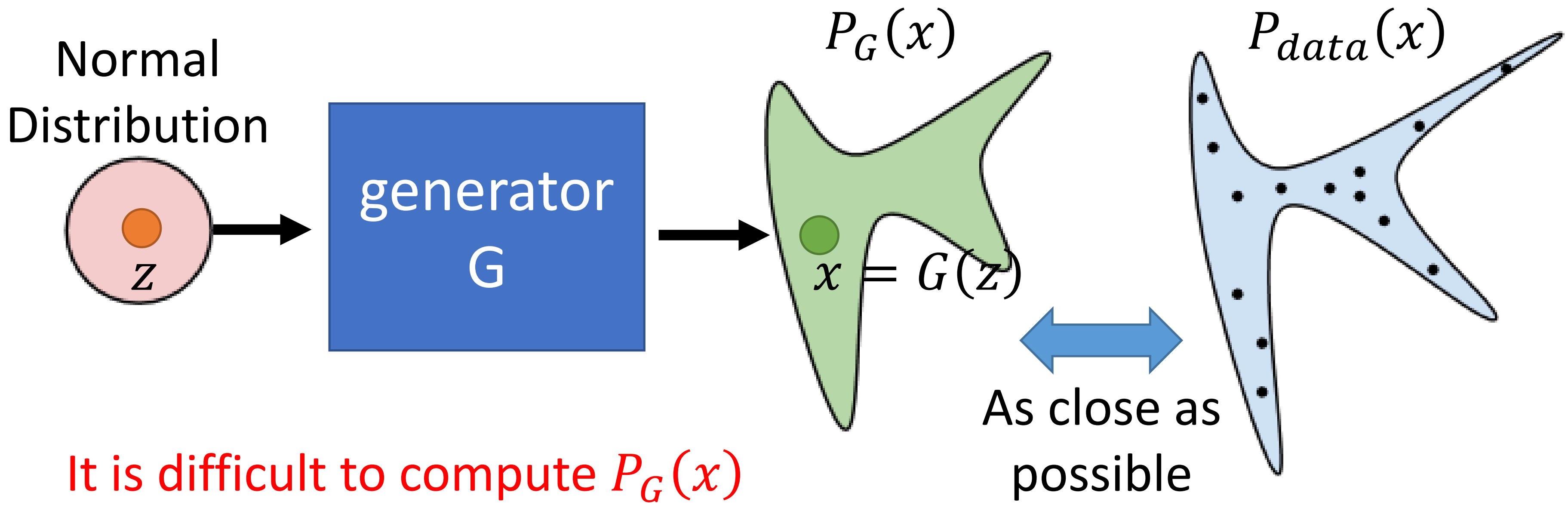
Basic Idea of GAN

- 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.

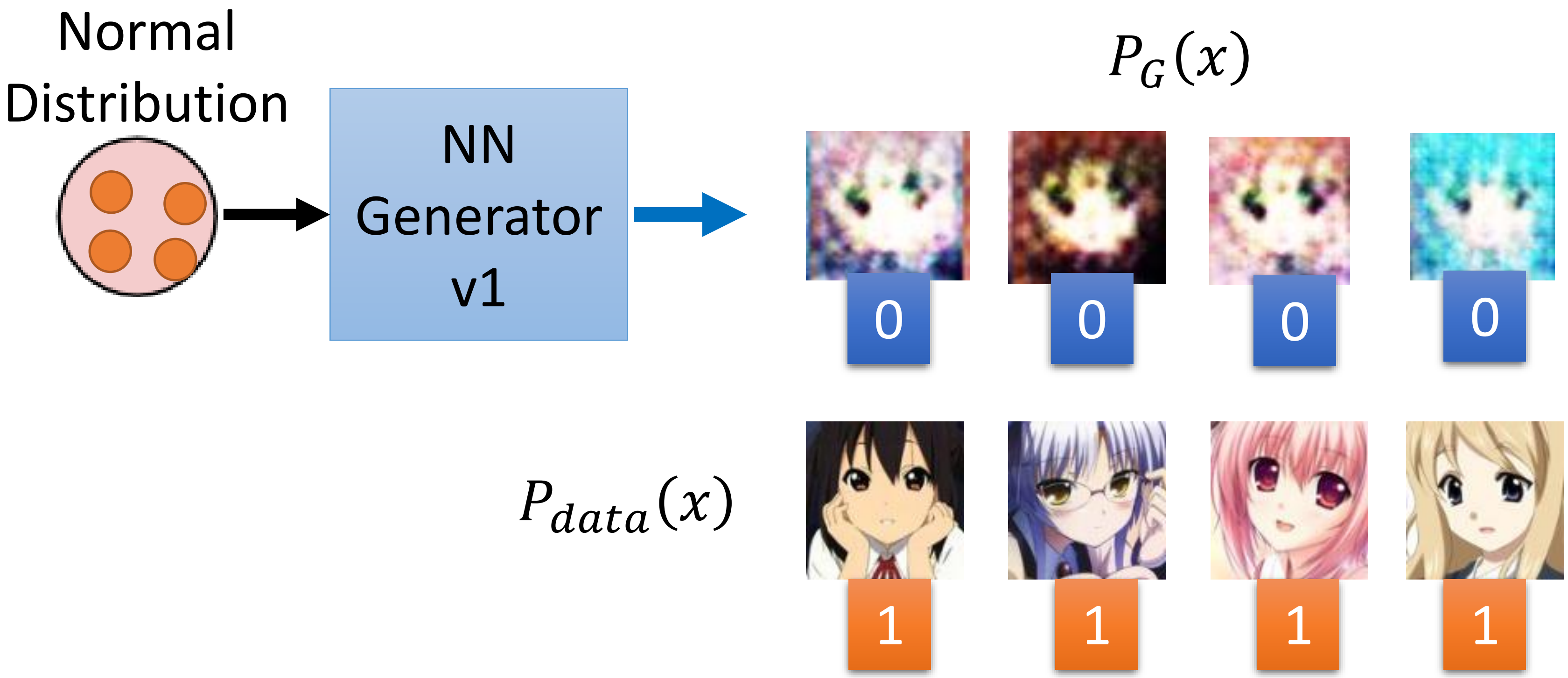


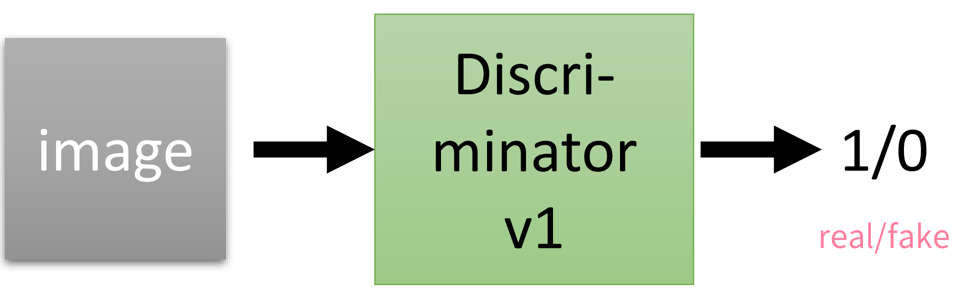
It is difficult to compute $P_G(x)$

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***.

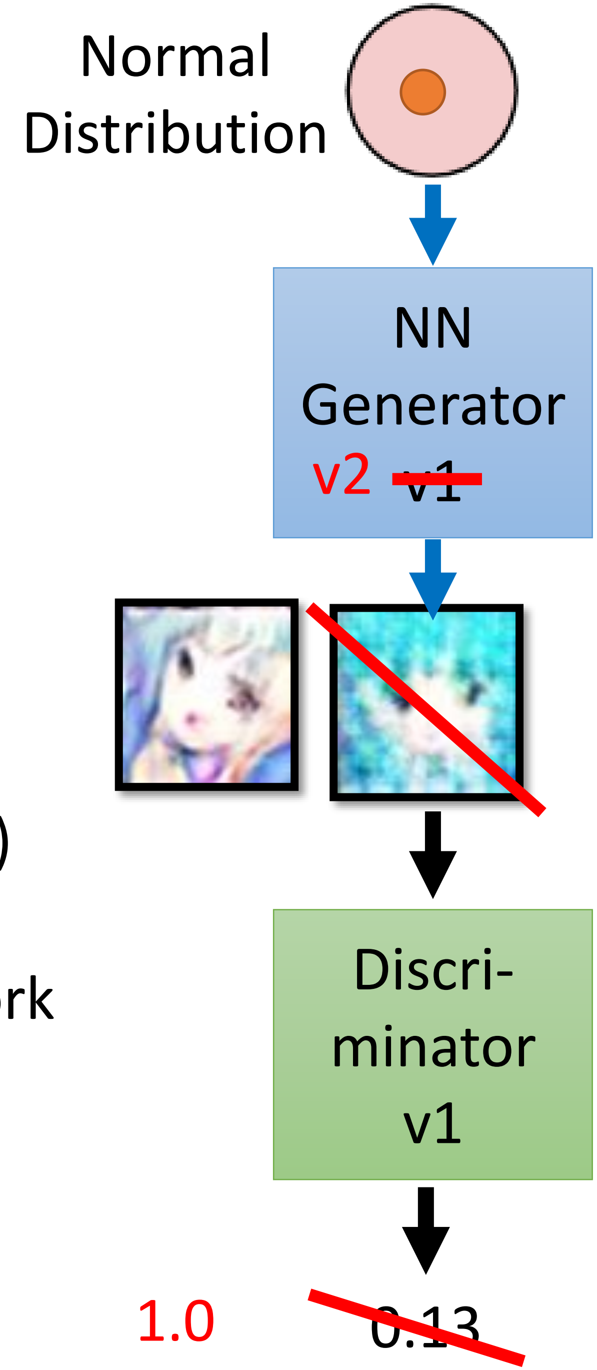
Basic Idea of GAN

- ***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

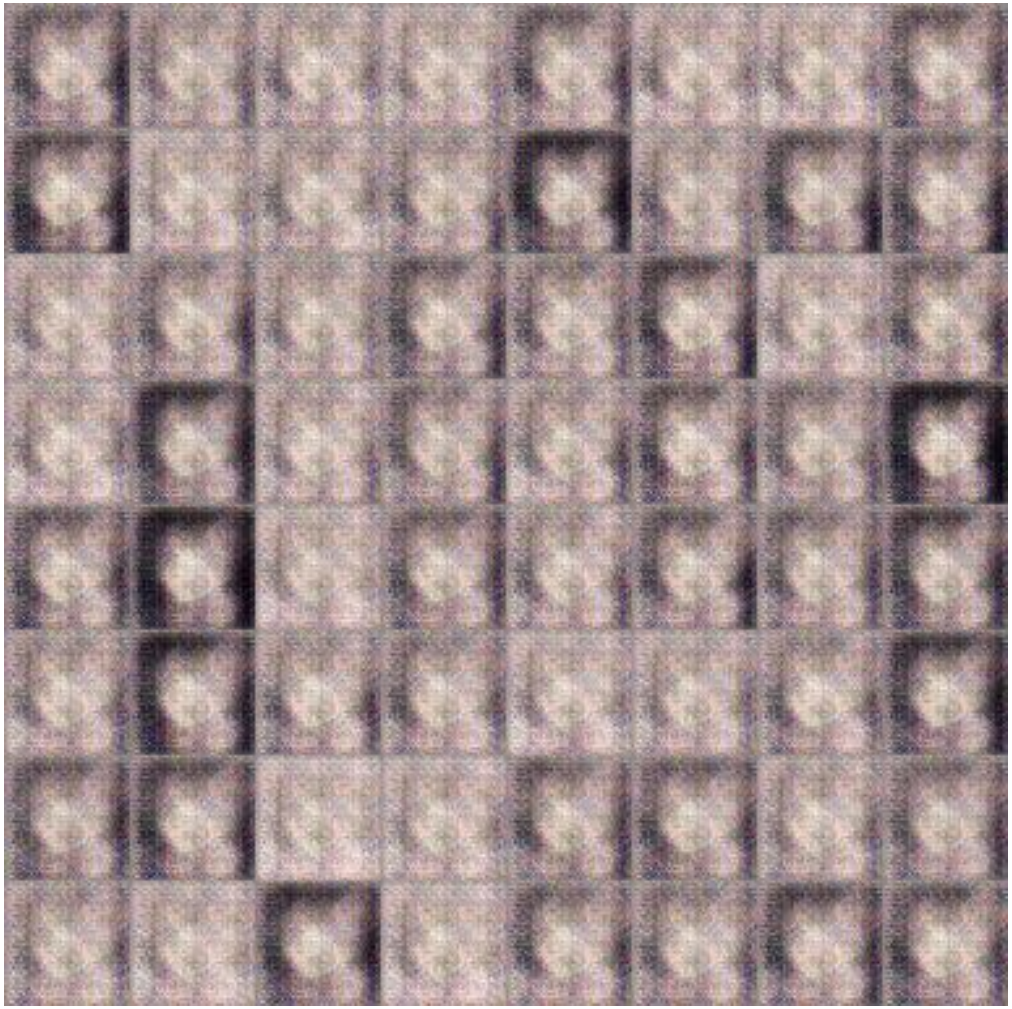


GAN – 二次元人物頭像鍊成

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

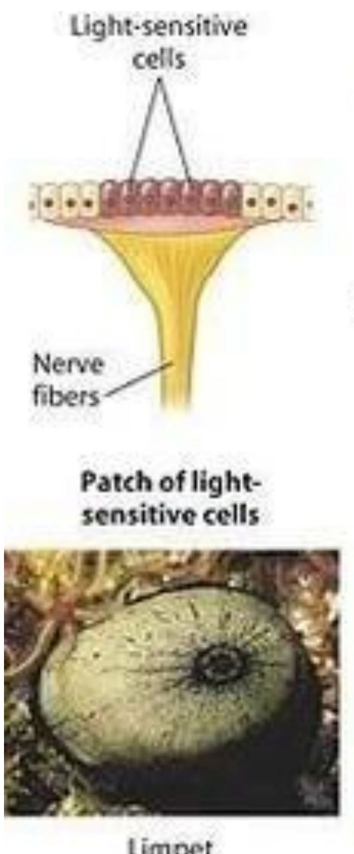
20,000 rounds

GAN – 二次元人物頭像鍊成

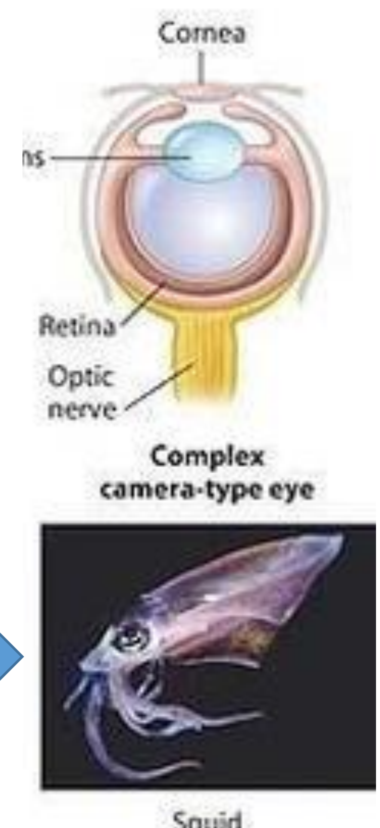
50,000 rounds

Why GAN is hard to train?

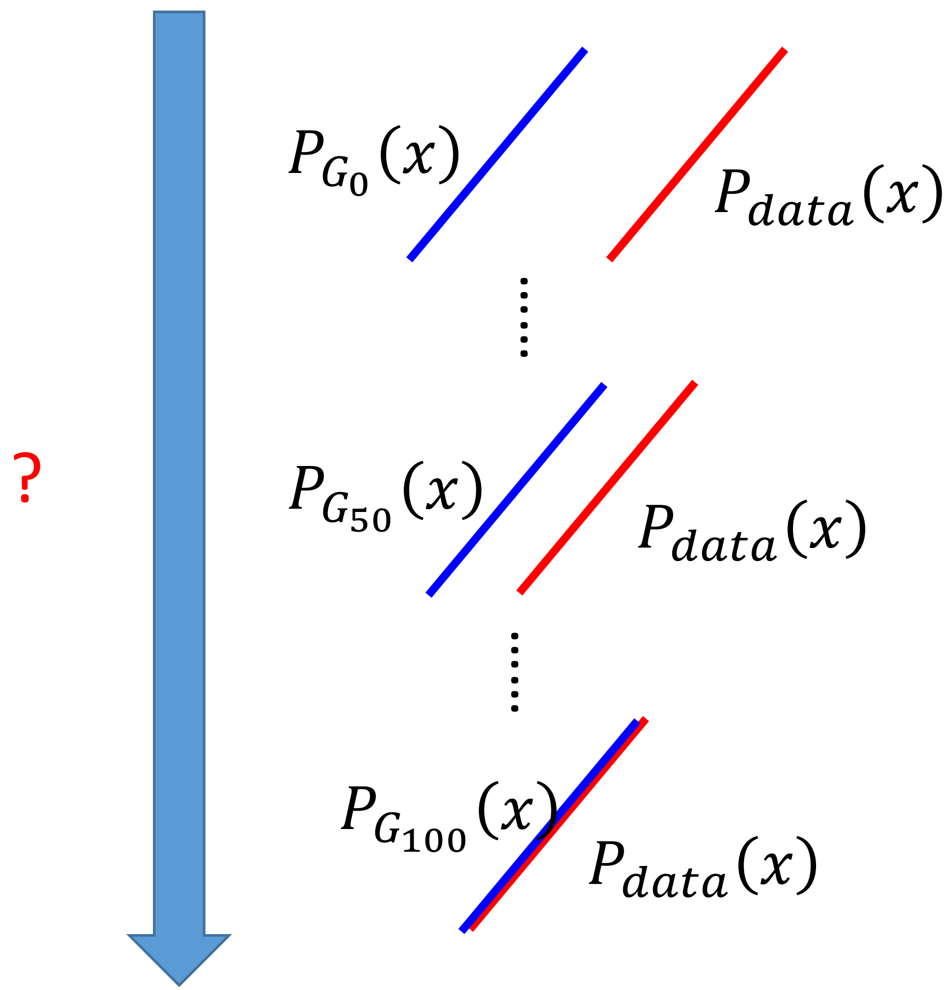
回到演化的比喻 ……



Better



Why GAN is hard to train?



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

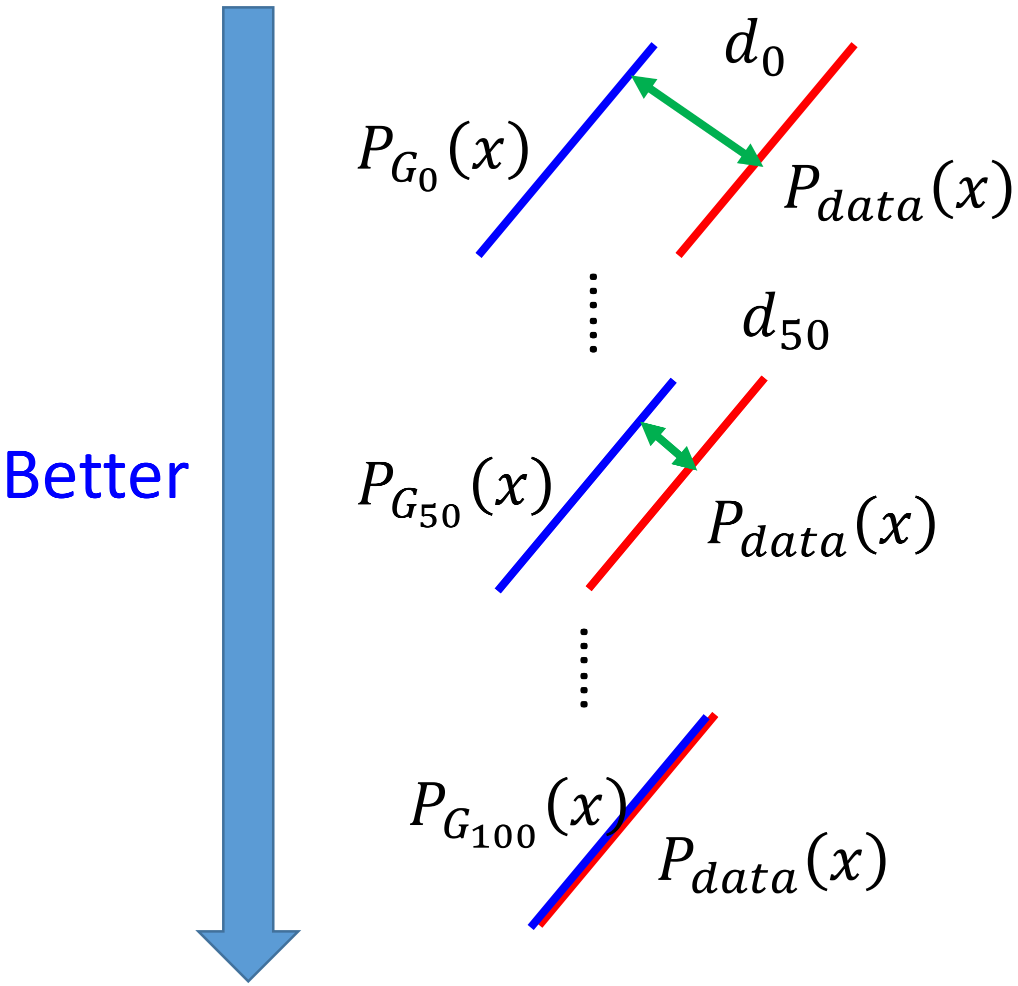
$JS(P_{G_2}||P_{data}) = log2$

Not really better ……

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

WGAN

Using Wasserstein distance instead of JS divergence

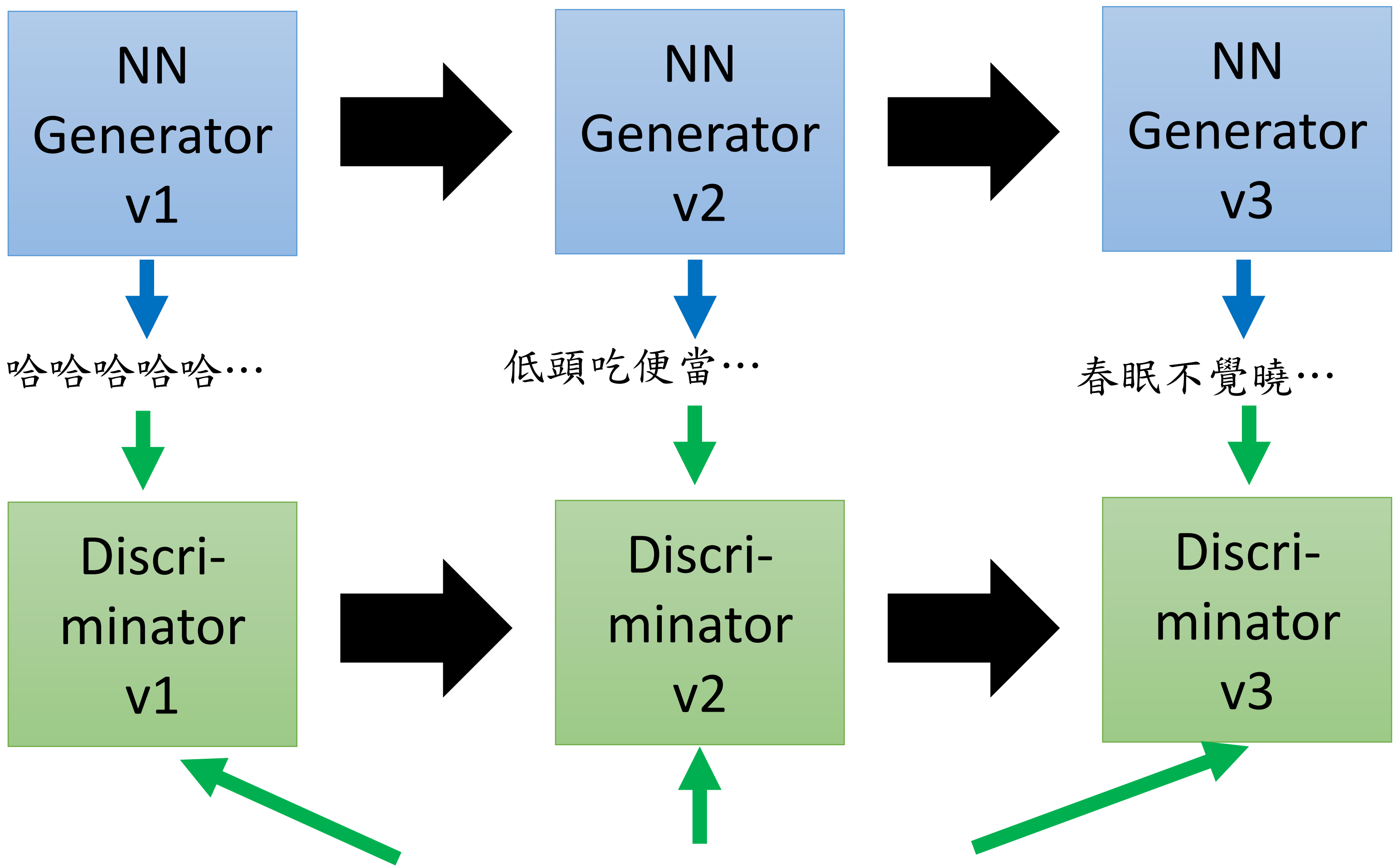


$$W\left(P_{G_1}||P_{data}\right) = d_0$$

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

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

WGAN – 唐詩鍊成



Real poems: 床前明月光，疑似地上霜，舉頭望明月，低頭思故鄉。

WGAN – 唐詩鍊成

由 李仲翊 同學提供實驗結果

Random generated

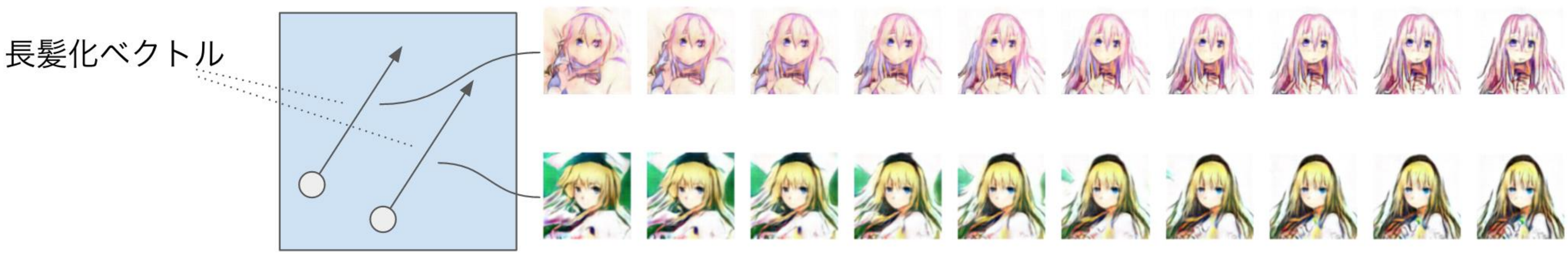
- 升雲白遲丹齋取，此酒新巷市入頭。黃道故海歸中後，不驚入得韻子門。
- 摢口容章蕃翎翎，邦貸無遊隔將毬。外蕭曾臺遶出畧，此計推上呂天夢。
- 新來寶伎泉，手雪泓臺蓑。曾子花路魏，不謀散薦船。
- 功持牧度機邈爭，不蹮官嬉牧涼散。不迎白旅今掩冬，盡蘸金祇可停。
- 玉十洪沄爭春風，溪子風佛挺橫鞋。盤盤稅焰先花齋，誰過飄鶴一丞憧。
- 海人依野庇，為阻例沉迴。座花不佐樹，弟闌十名儂。
- 入維當興日世瀕，不評皺。頭醉空其杯，駸園凋送頭。
- 鉢筌動春枝，寶絫潔長知。官爲窑爛去，絆粒薛一靜。
- 吾涼腕不楚，縱先待旅知。楚人縱酒待，一蔓飄聖猜。
- 折幕故癘應韻子，徑頭霜瓊老徑徑。尚錯春鏘熊悽梅，去吹依能九將香。
- 通可矯目鷚須浄，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。

Moving on the code space

Alec Radford, Luke Metz, Soumith Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, ICLR, 2016

Moving on the code space

- Ref: http://qiita.com/mattya/items/e5bfe5e04b9d2f0bbd47

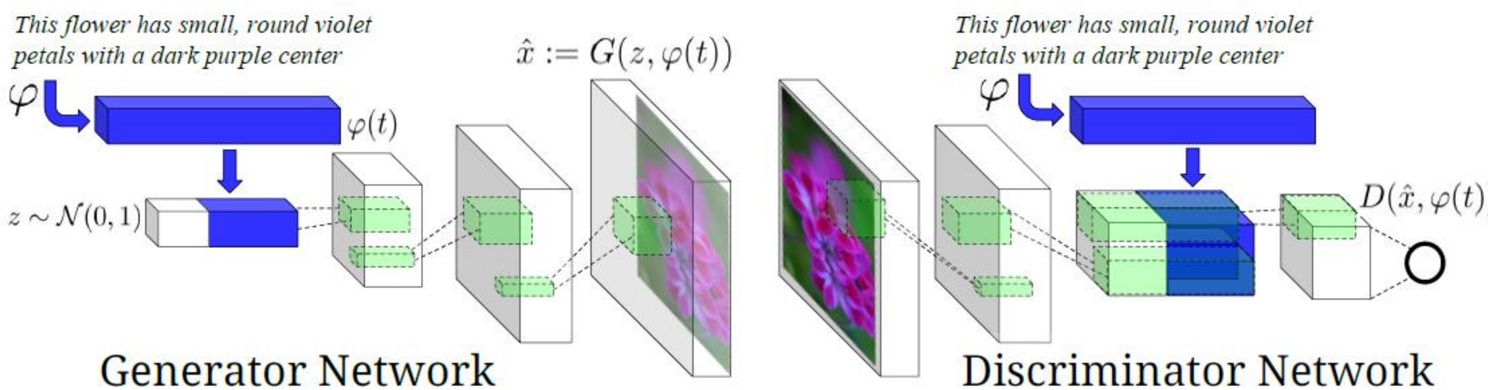


一番左のキャラクターが元画像で、
右に行くほど長髪化ベクトルを強く足している

元画像　－赤髪＋金髪　－赤目＋青目　＋制服＋セーラー　＋笑顔＋口開き　＋青背景

Text to Image

Text → Generator → Image



Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee, "Generative Adversarial Text-to-Image Synthesis", ICML 2016

Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, Dimitris Metaxas, "StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks", arXiv prepring, 2016

Scott Reed, Zeynep Akata, Santosh Mohan, Samuel Tenka, Bernt Schiele, Honglak Lee, "Learning What and Where to Draw", NIPS 2016

Text to Image

"red flower with black center"

Caption	Image
this flower has white petals and a yellow stamen	
the center is yellow surrounded by wavy dark purple petals	
this flower has lots of small round pink petals	

Text to Image

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

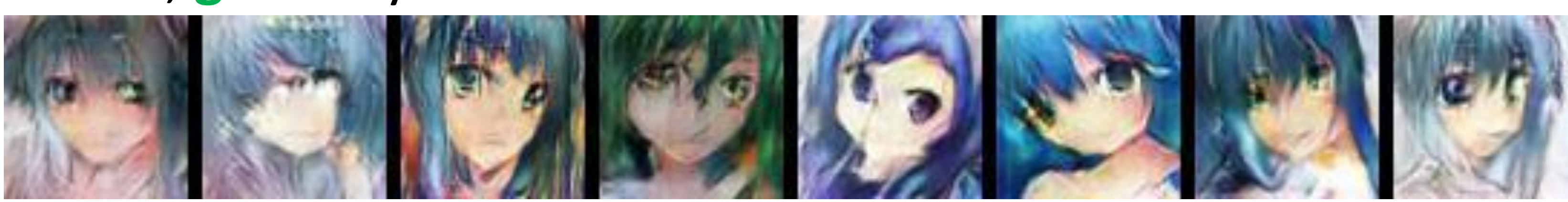
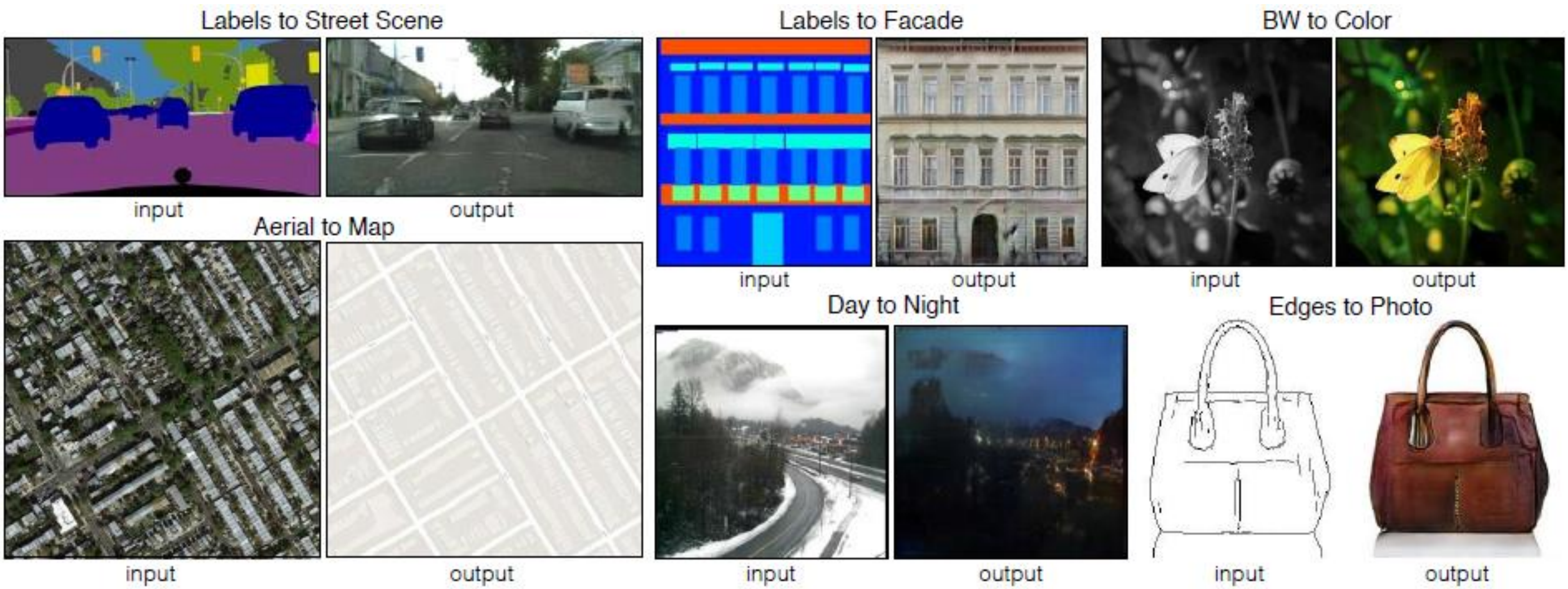
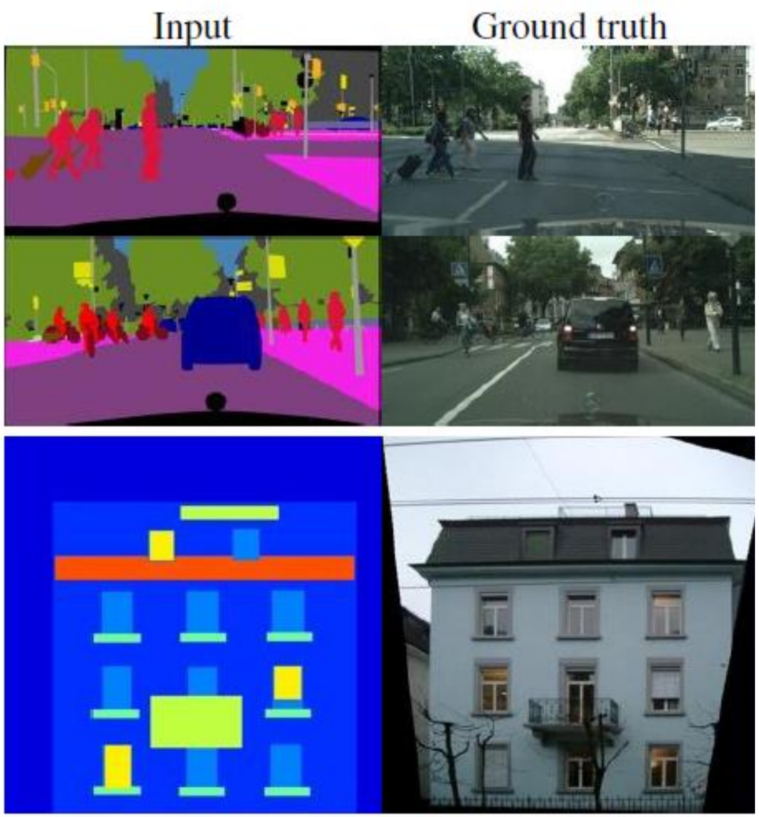


Image-to-image Translation



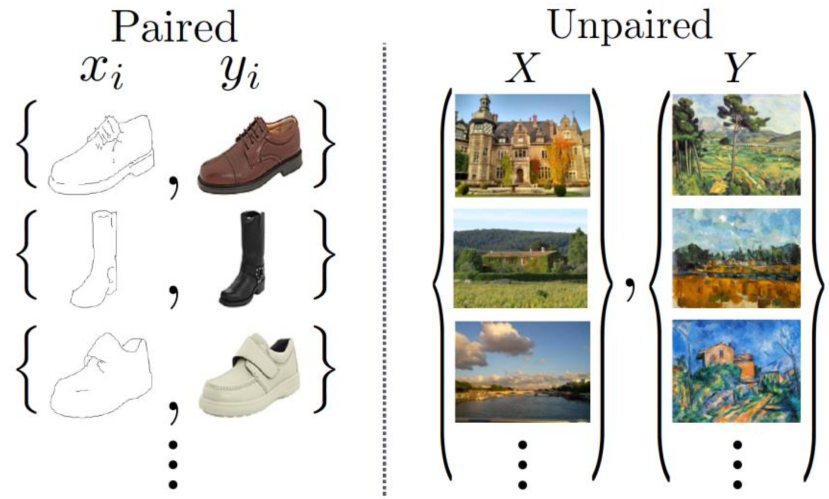
Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks”, arXiv preprint, 2016

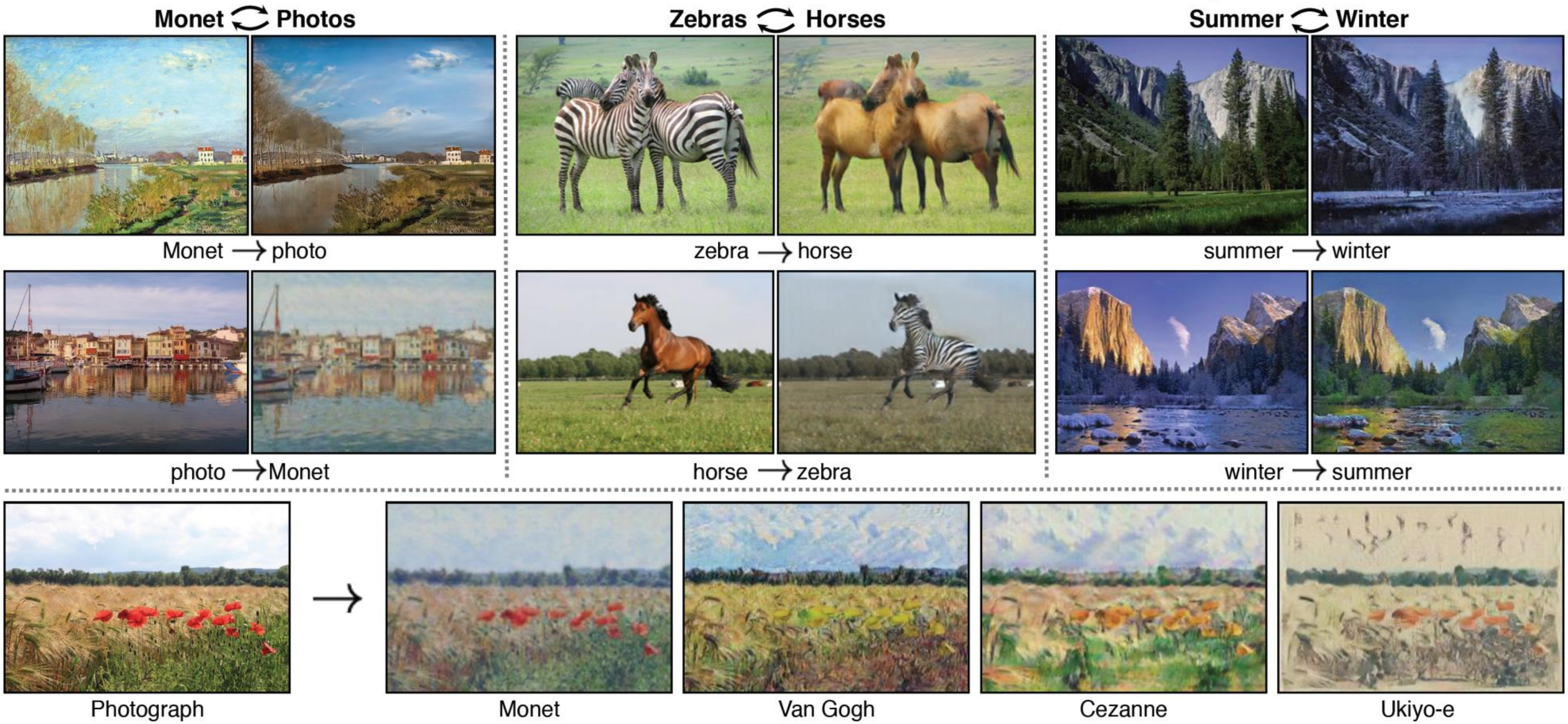
Image-to-image Translation
- Results



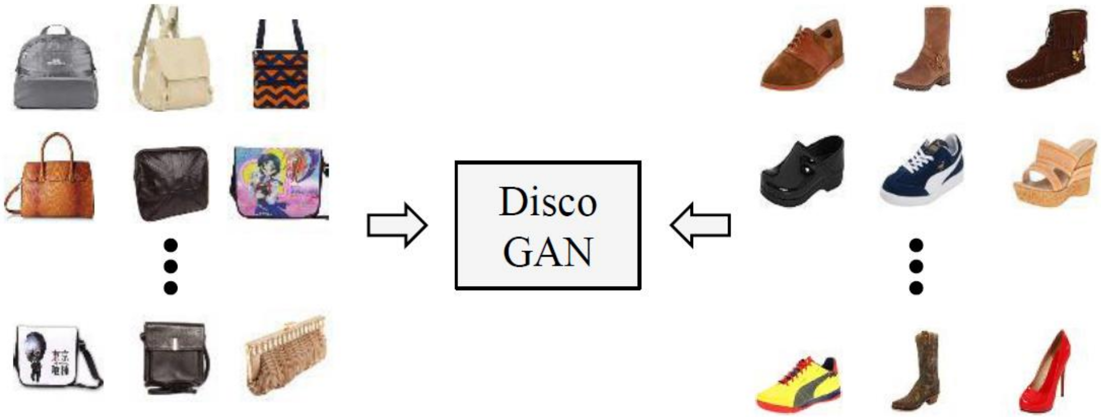
Cycle GAN

https://arxiv.org/abs/1703.10593





Disco GAN



(a) Learning cross-domain relations **without any extra label**

(b) Handbag images (input) & **Generated** shoe images (output)

(c) Shoe images (input) & **Generated** handbag images (output)

https://arxiv.org/abs/1703.05192

機械学習で美少女化~あるいはNEW GAME! の世界

- http://qiita.com/Hi-king/items/8d36d9029ad1203aac55

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
……	……

In practical ......

- GANs are difficult to optimize.
- No explicit signal about how good the generator is
  - In standard NNs, we monitor loss
  - In GANs, we have to keep "well-matched in a contest"
- When discriminator fails, it does not guarantee that generator generates realistic images
  - Just because discriminator is stupid
  - Sometimes generator find a specific example that can fail the discriminator
  - Making discriminator more robust may be helpful.