CSCE 636 Neural Networks (Deep Learning)

Lecture 17: VAE and GAN

Anxiao (Andrew) Jiang

Based on the interesting lecture of Prof. Hung-yi Lee "Unsupervised Learning: Generation"
https://www.youtube.com/watch?v=YNUek8ioAJk&list=PLJV_el3uVTsPy9oCRY30oBPNLCo89yu49&index=26

Unsupervised Learning: Generation

Creation

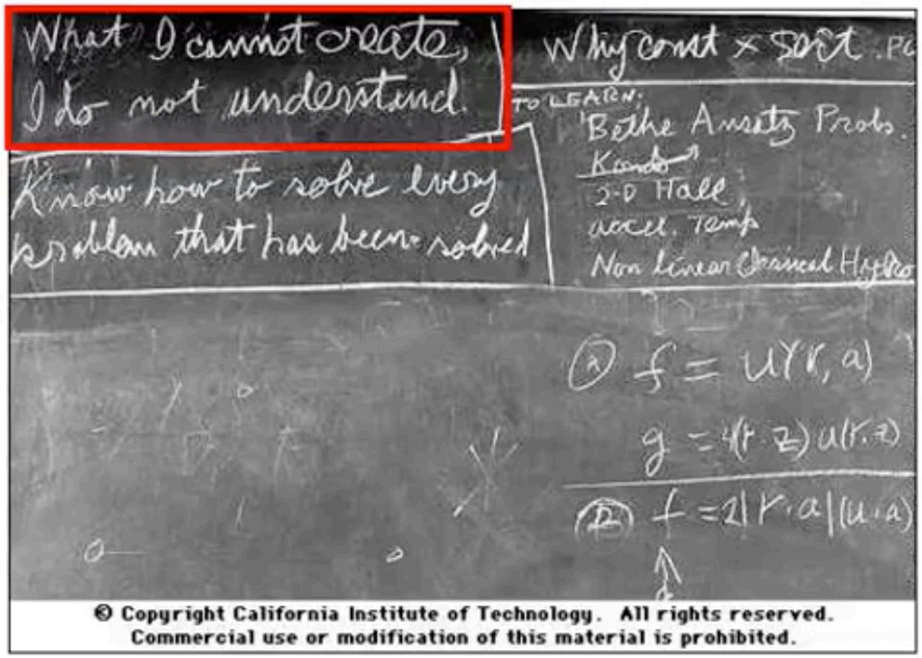
- Generative Models: A good review article: https://openai.com/blog/generative-models/

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

在這篇文章裡面呢

Creation

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



What I cannot create,
I do not understand.

Richard Feynman

Creation – Image Processing

Now

v.s.

In the future

Creation – Image Processing

Now

v.s.

In the future

Machine
draws a cat

Generative Models

PixelRNN

Variational Autoencoder (VAE)

Generative Adversarial Network (GAN)

Generative Models

PixelRNN

Variational Autoencoder (VAE)

Generative Adversarial Network (GAN)

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

E.g. 3 x 3 images

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

E.g. 3 x 3 images

接下來呢，你就 learn 一個 model

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

E.g. 3 x 3 images

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

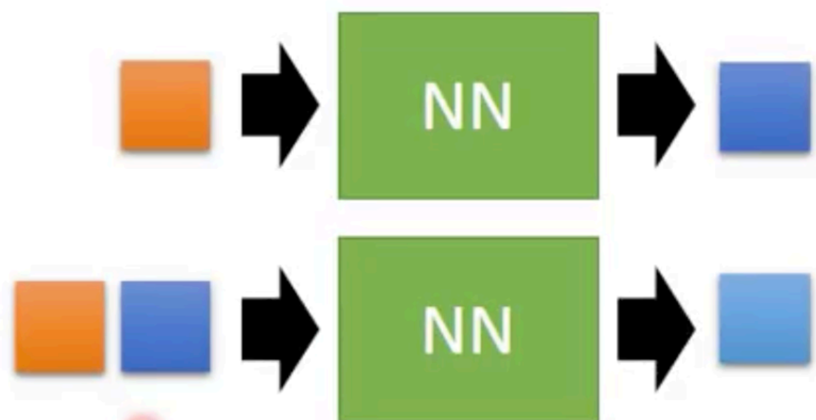
E.g. 3 x 3 images

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

E.g. 3 x 3 images

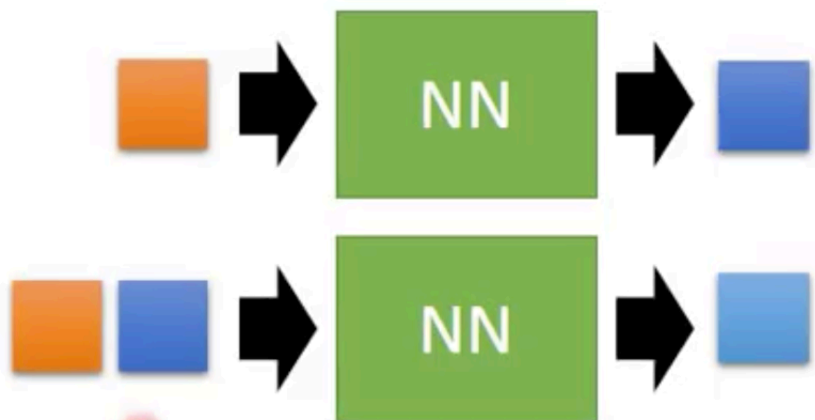


PixelRNN

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

E.g. 3 x 3 images

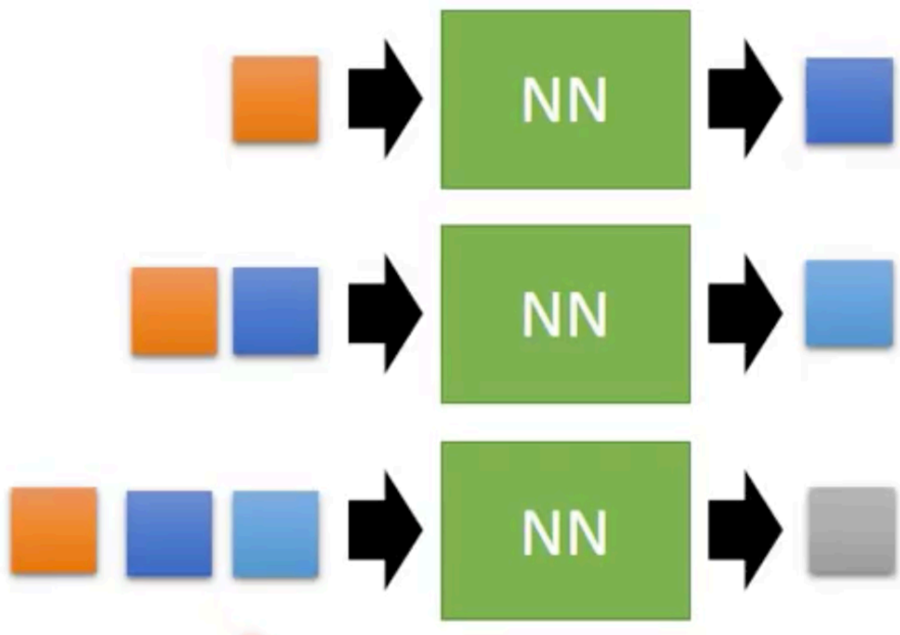


PixelRNN

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

E.g. 3 x 3 images

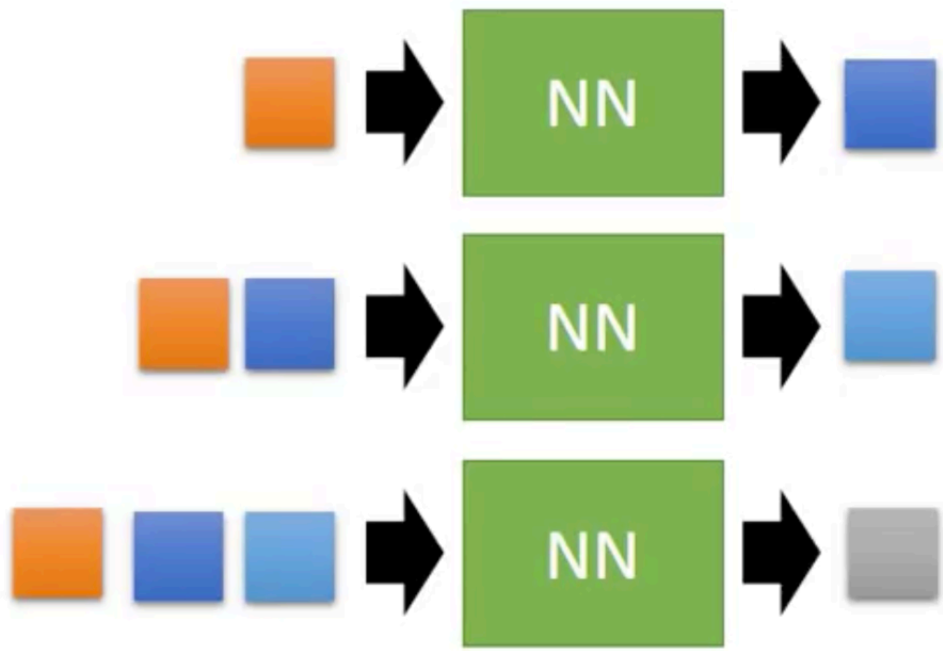


PixelRNN

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

E.g. 3 x 3 images

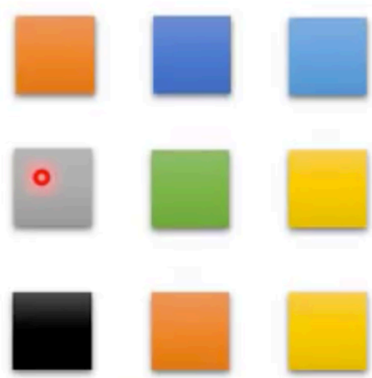


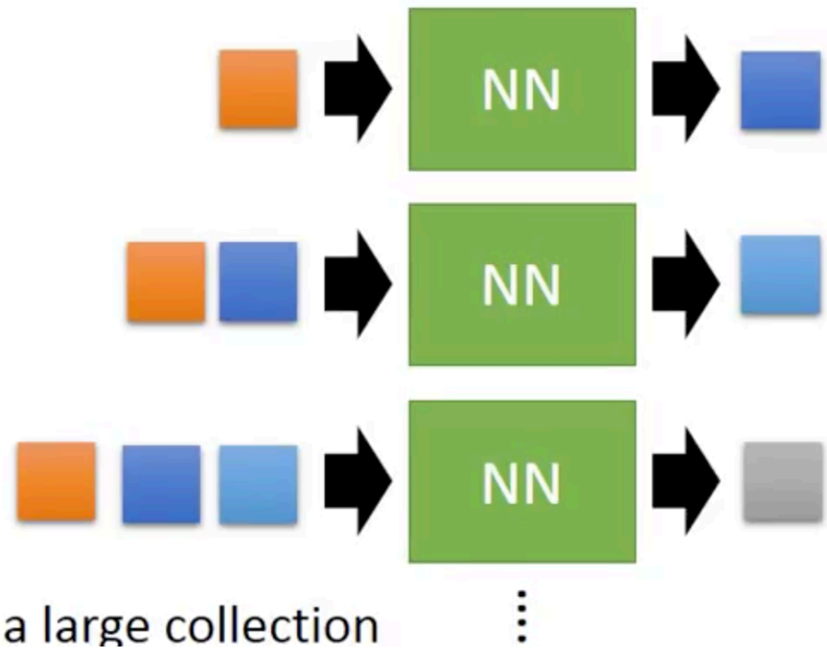
PixelRNN

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

E.g. 3 x 3 images





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

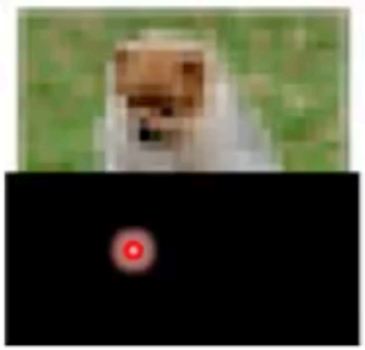
PixelRNN

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

Real World

PixelRNN

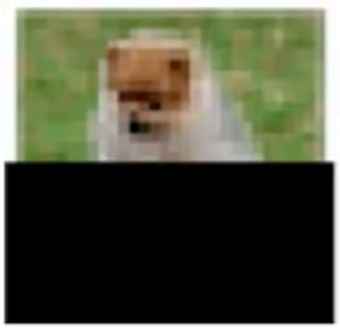
Ref: Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu, Pixel Recurrent Neural Networks, arXiv preprint, 2016



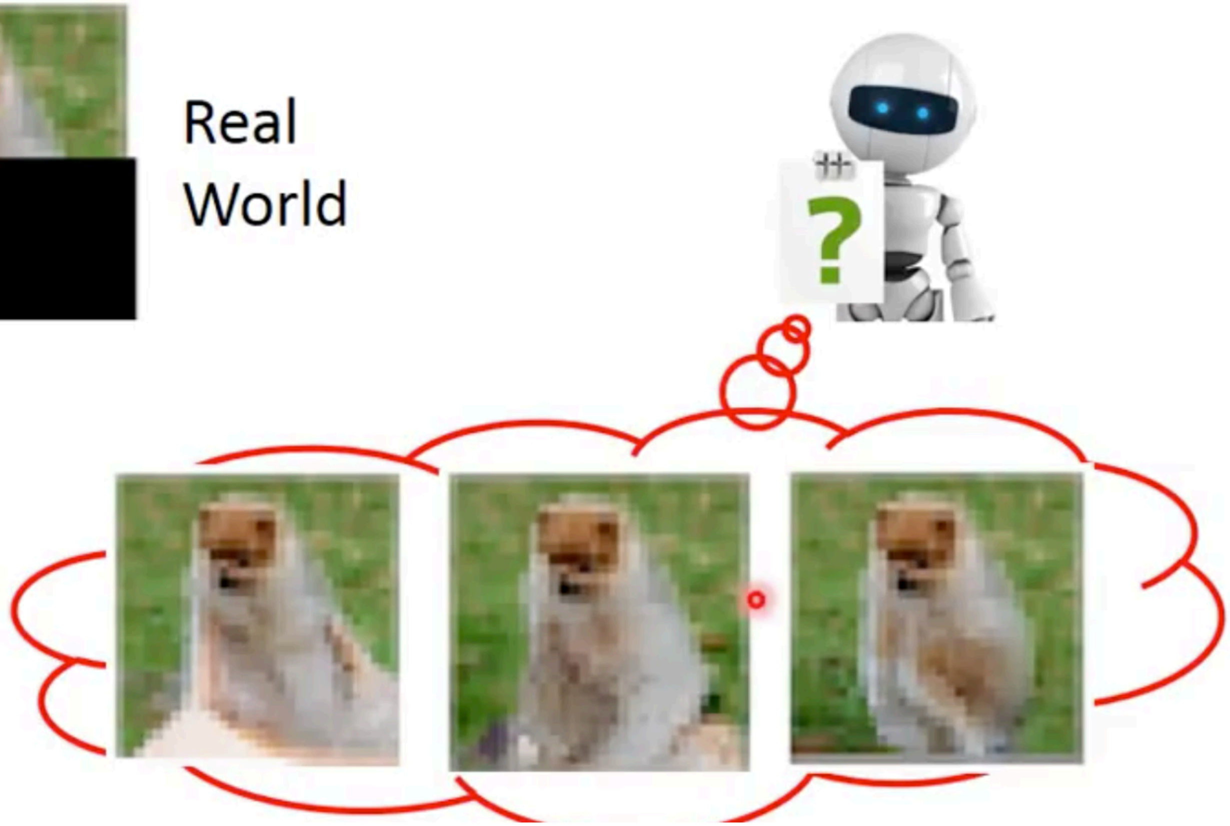
Real World

PixelRNN

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

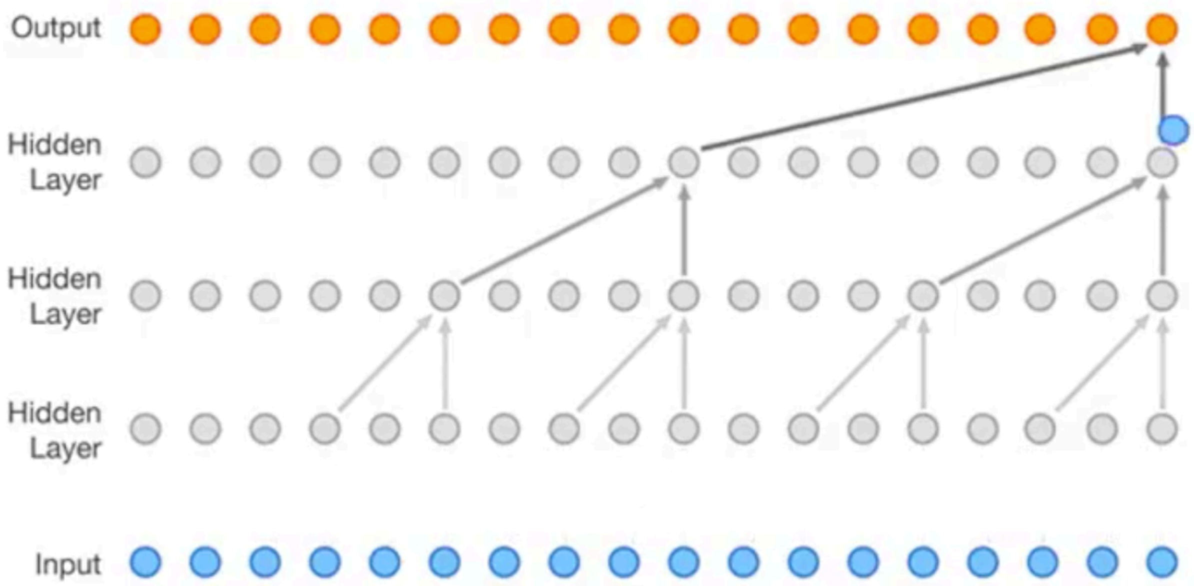


Real World



More than images ......

Speech generation: WaveNet



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 prepri

我們知道 WaveNet 可以做語音合成

Generative Models

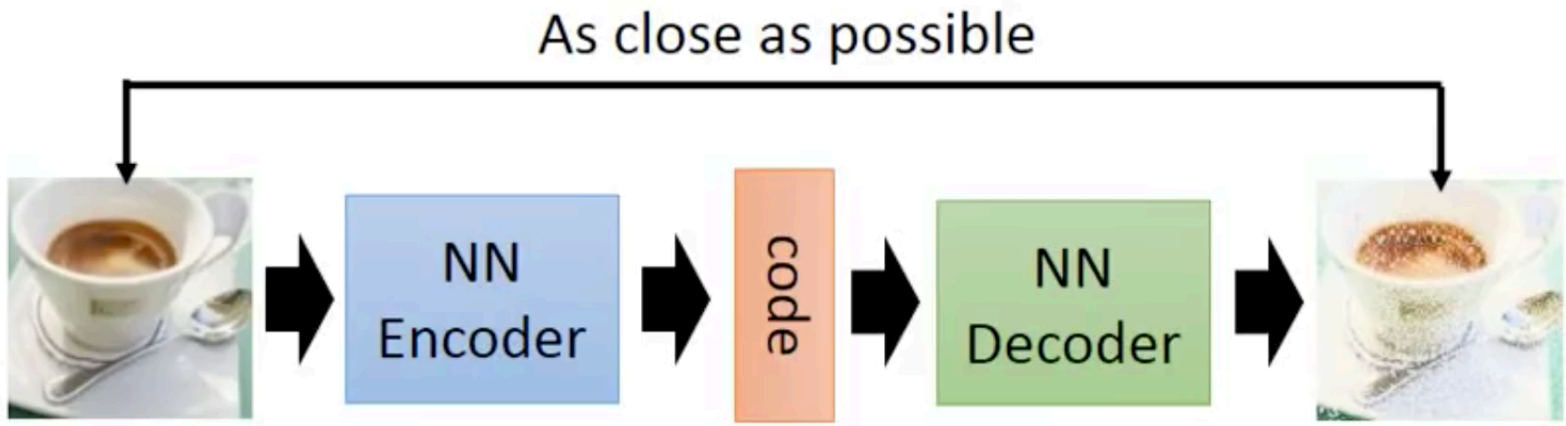
PixelRNN

Variational Autoencoder (VAE)

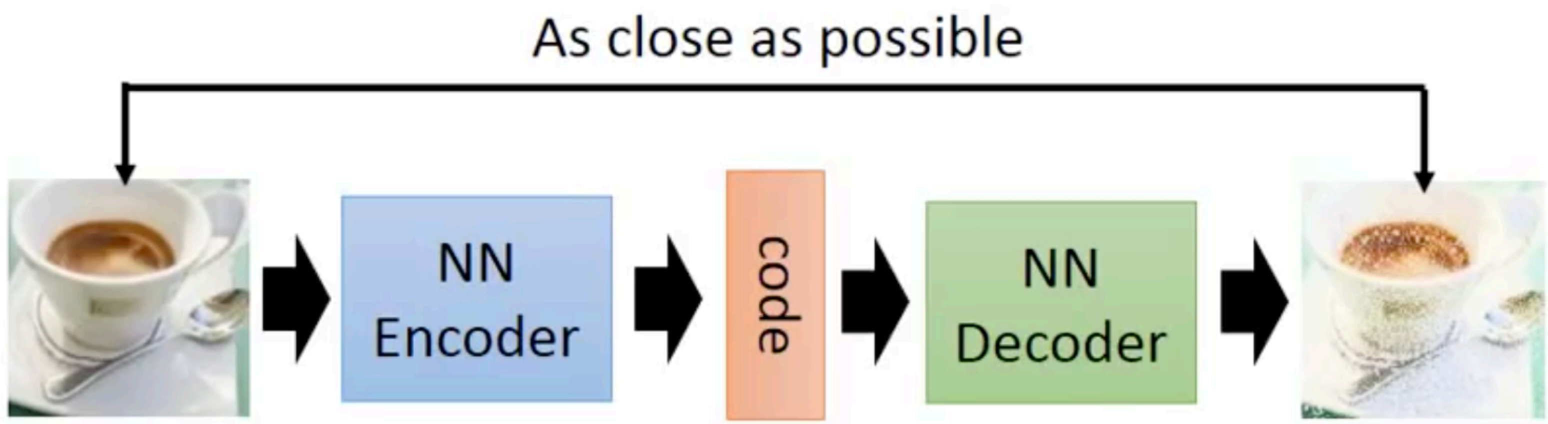
Diederik P Kingma, Max Welling, Auto-Encoding Variational Bayes, arXiv preprint, 2013

Generative Adversarial Network (GAN)

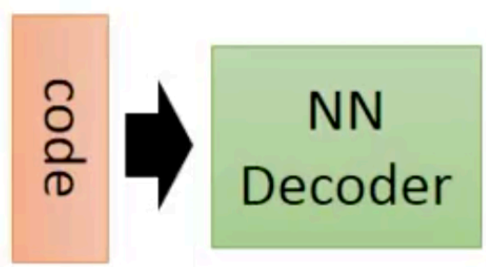
Auto-encoder



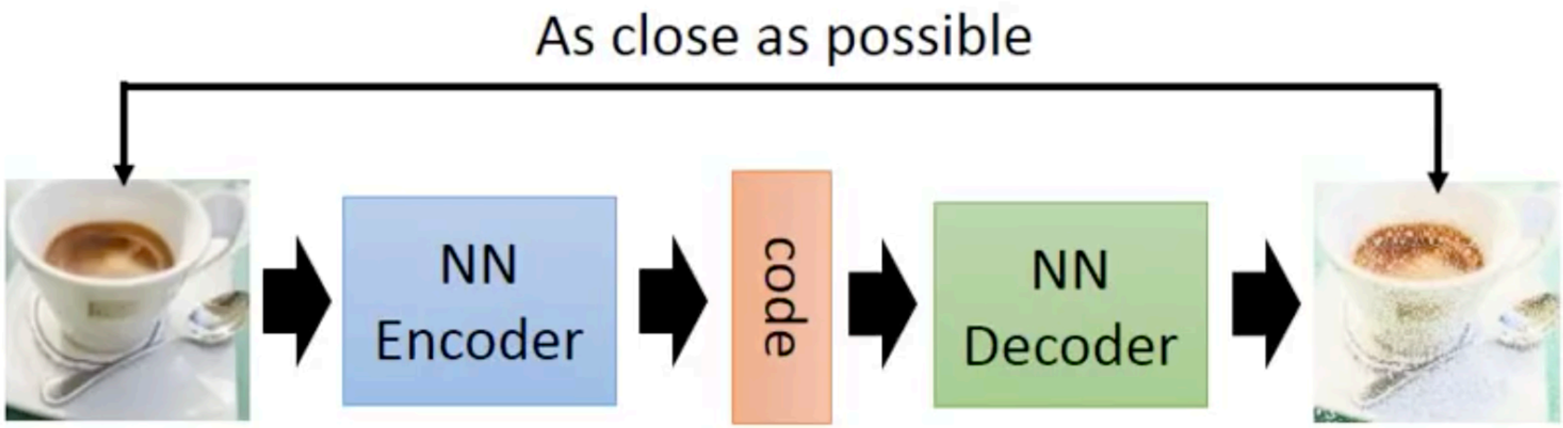
Auto-encoder



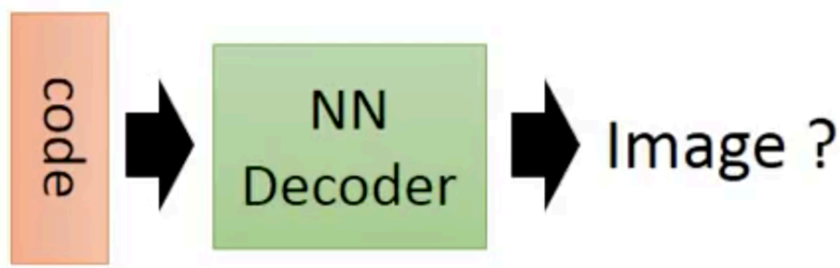
Randomly generate
a vector as code



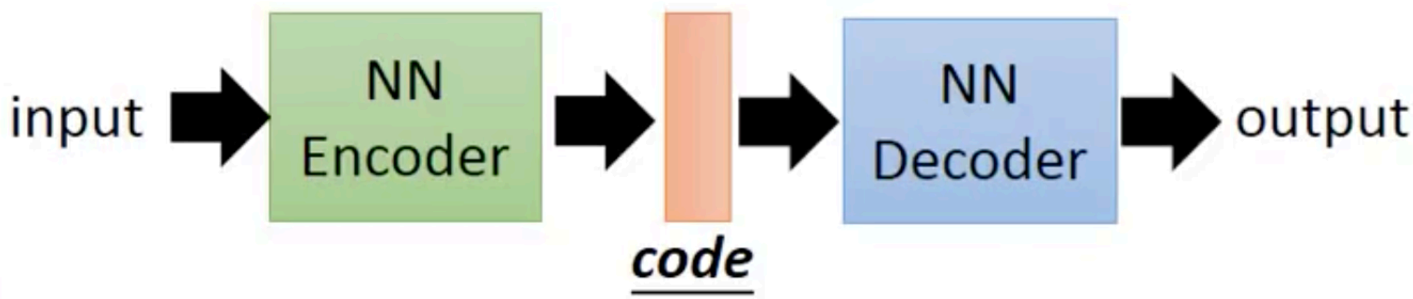
Auto-encoder



Randomly generate
a vector as code



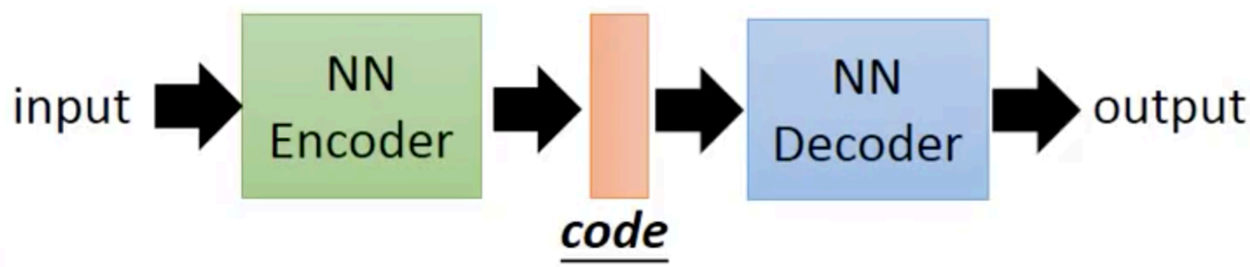
Auto-encoder



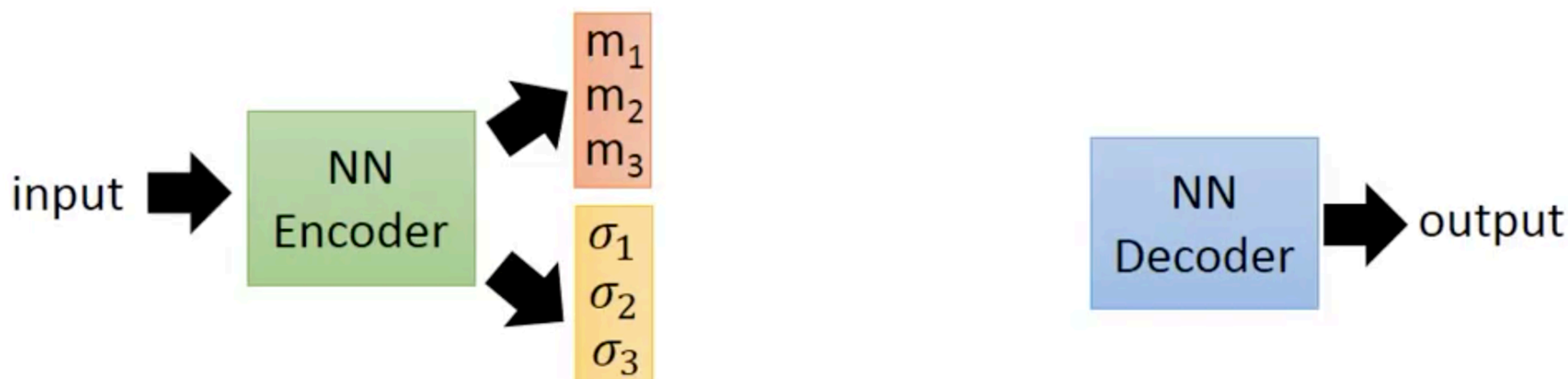
VAE

Better performance

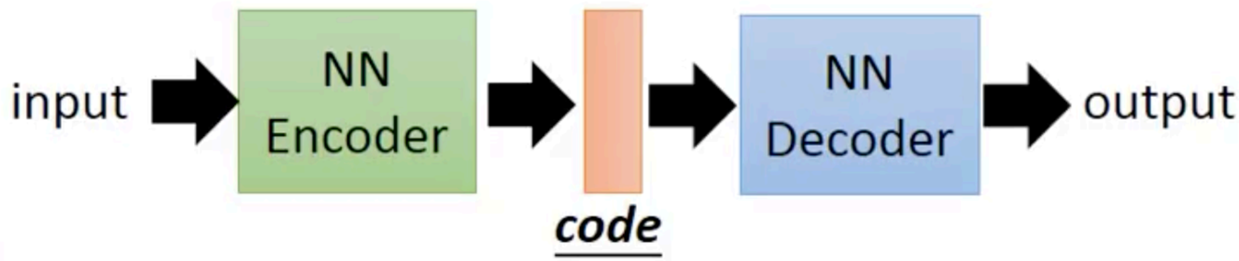
Auto-encoder



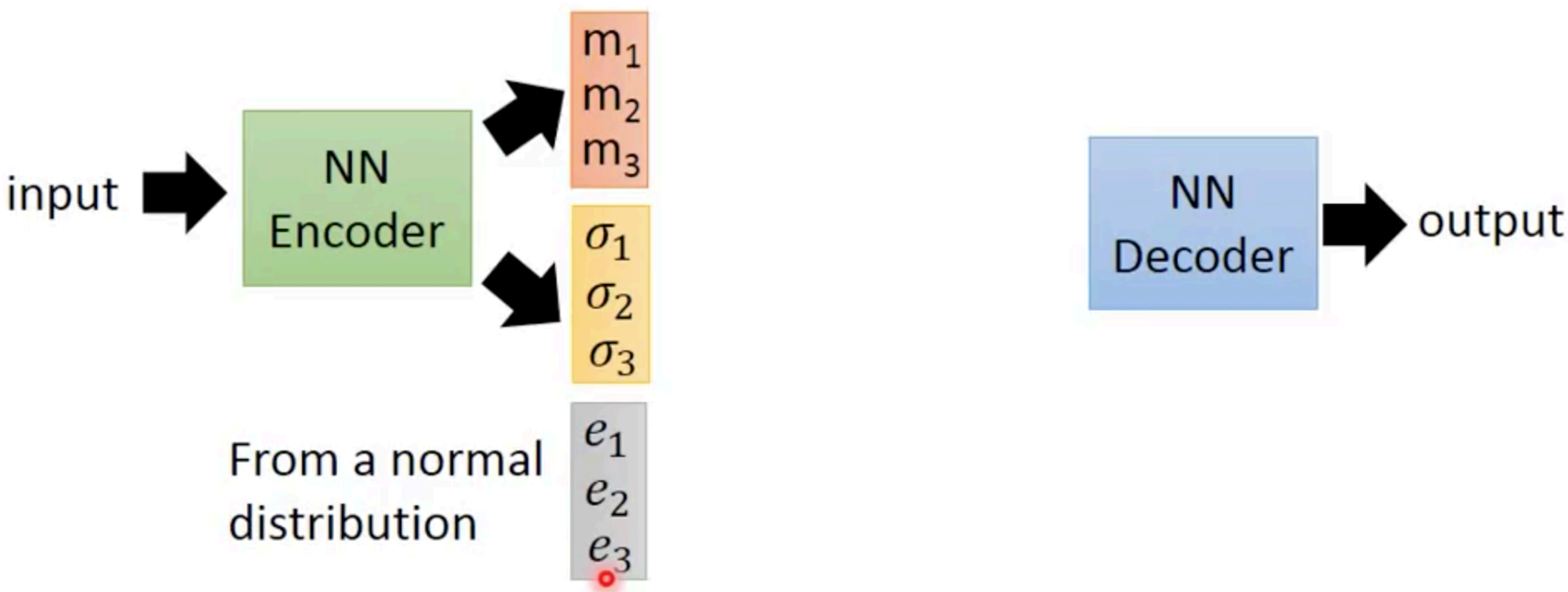
VAE



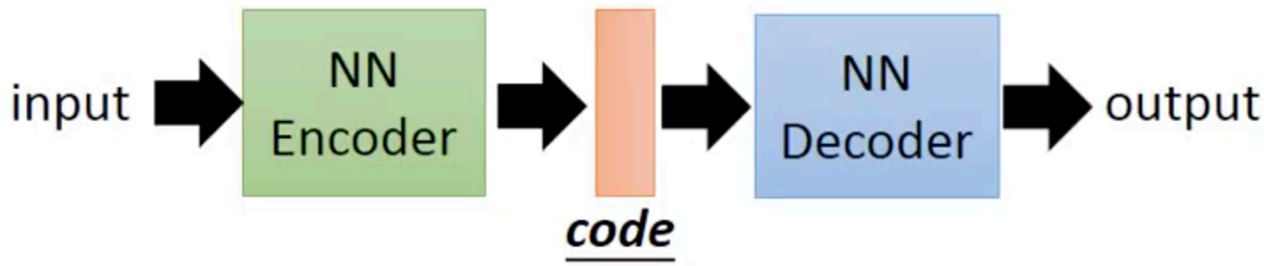
Auto-encoder



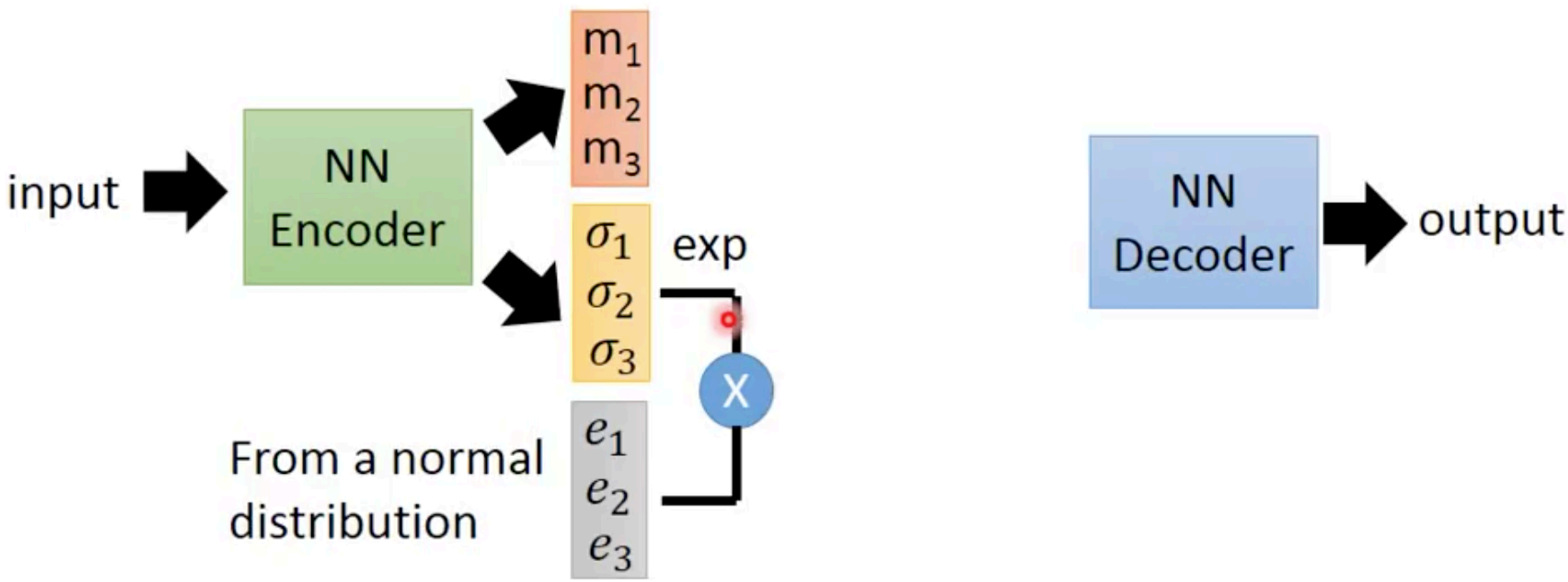
VAE



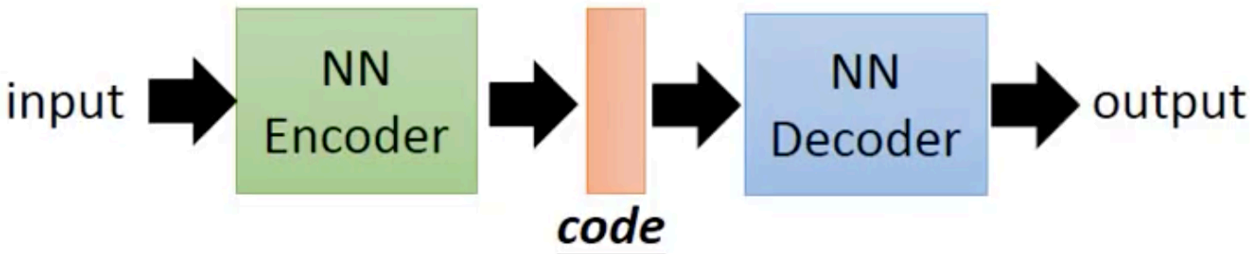
Auto-encoder



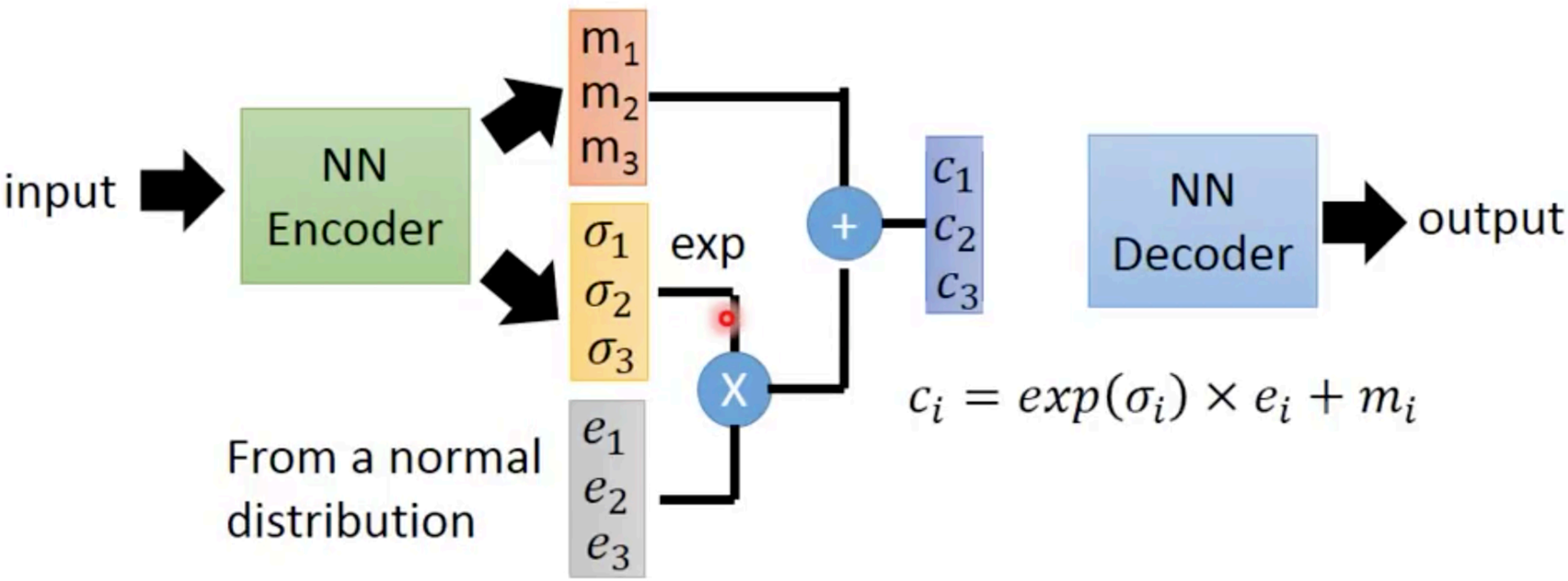
VAE

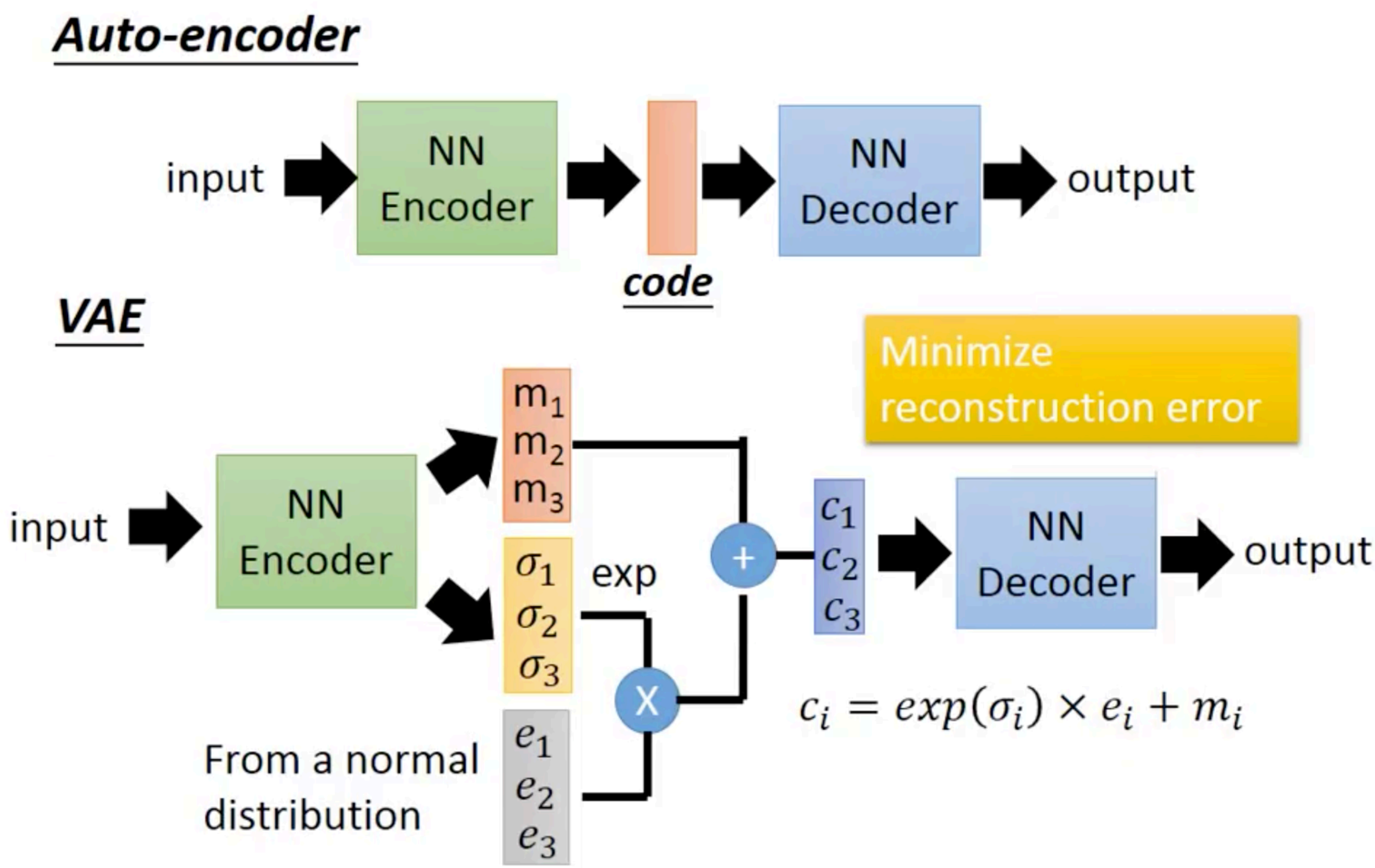


Auto-encoder

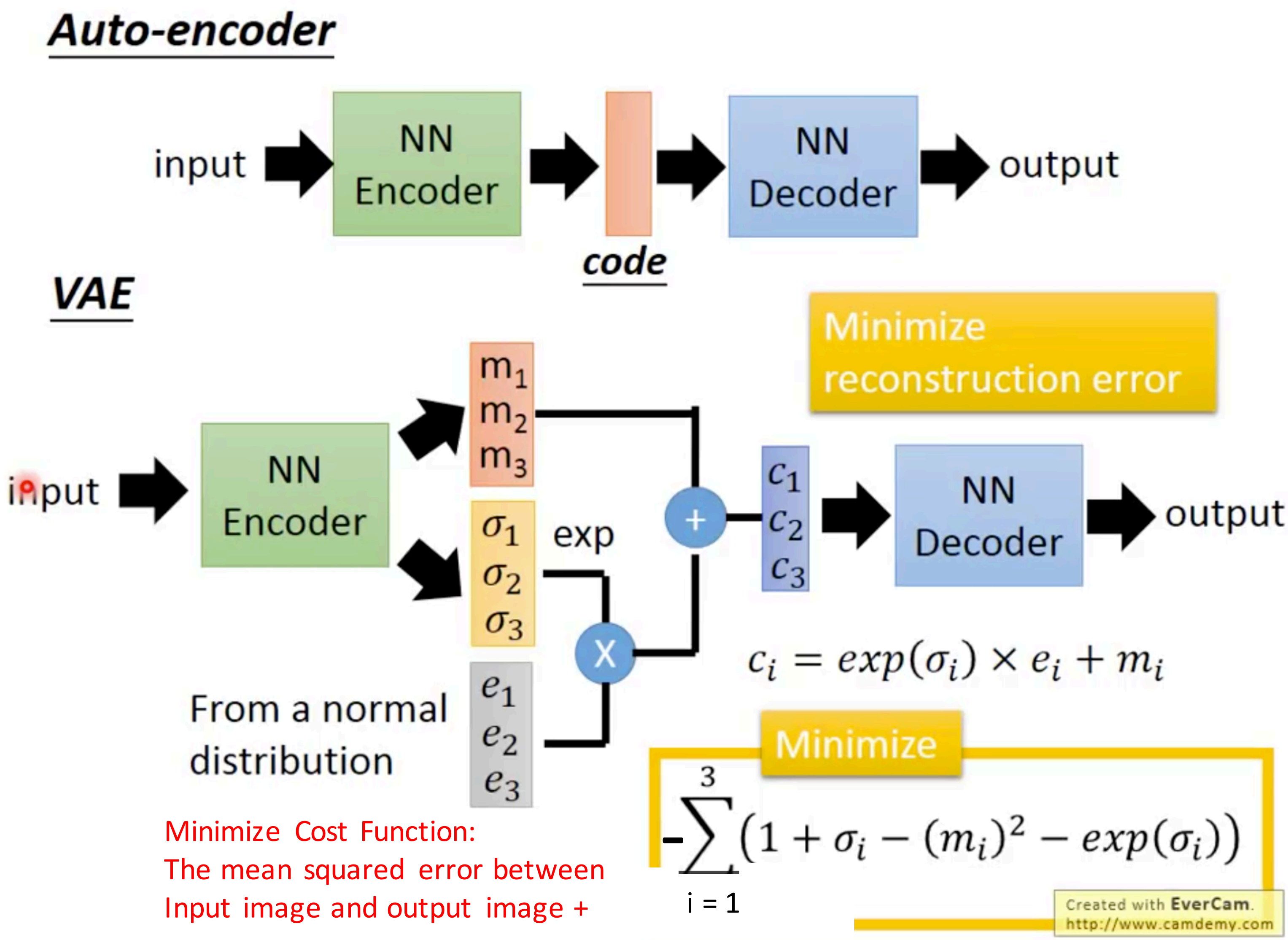


VAE





Auto-encoder
input
NN Encoder
code
NN Decoder
output
VAE
Minimize reconstruction error
input
NN Encoder
m_1 m_2 m_3
σ_1 σ_2 σ_3
exp
From a normal distribution
e_1 e_2 e_3
c_1 c_2 c_3
NN Decoder
output
$c_i = exp(\sigma_i) \times e_i + m_i$



Minimize Cost Function:
The mean squared error between
Input image and output image +

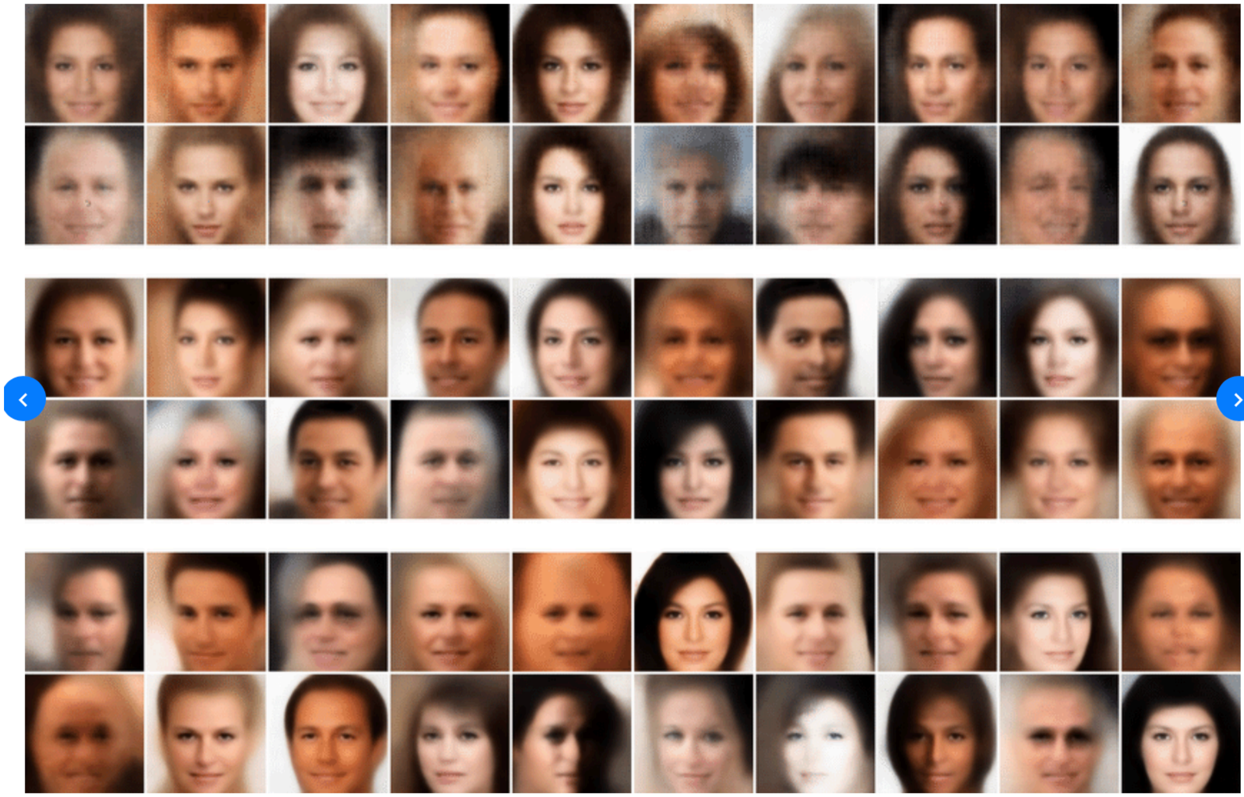
Cifar-10

https://github.com/openai/iaf

Figure 8 - uploaded by Hongyang Gao

Download

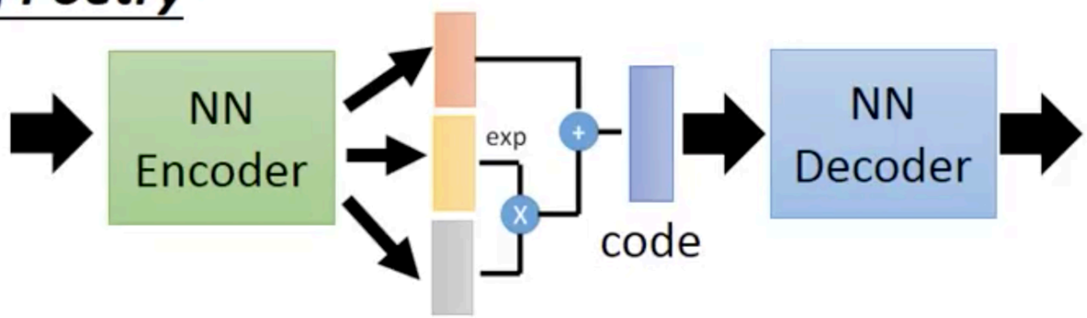
View publication



Sample images generated by different models when trained on the CelebA dataset. The first two rows are images generated by a standard VAE. The middle two rows are images generated by deep residual VAE. The last two rows are images generated by multi-stage VAE.

https://www.researchgate.net/figure/Sample-images-generated-by-different-models-when-trained-on-the-CelebA-dataset-The-first_fig5_317062169

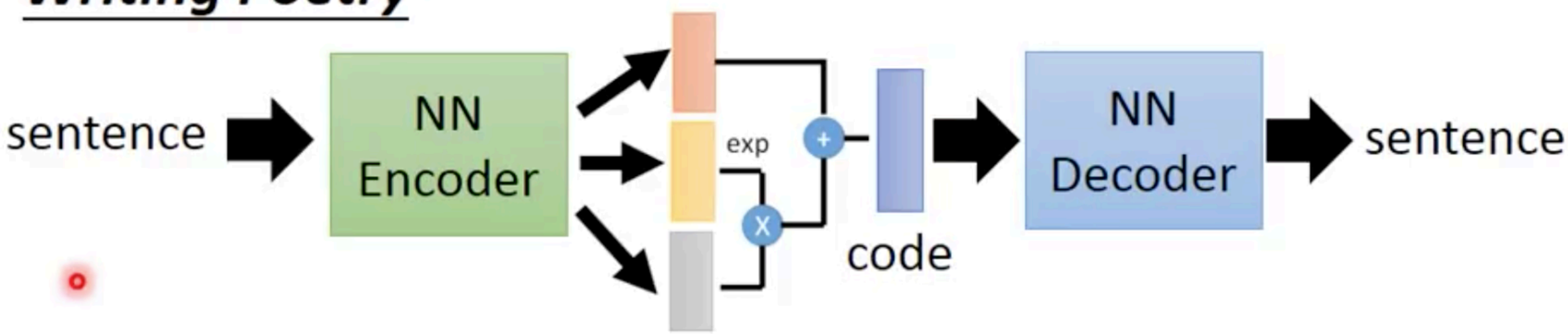
Writing Poetry



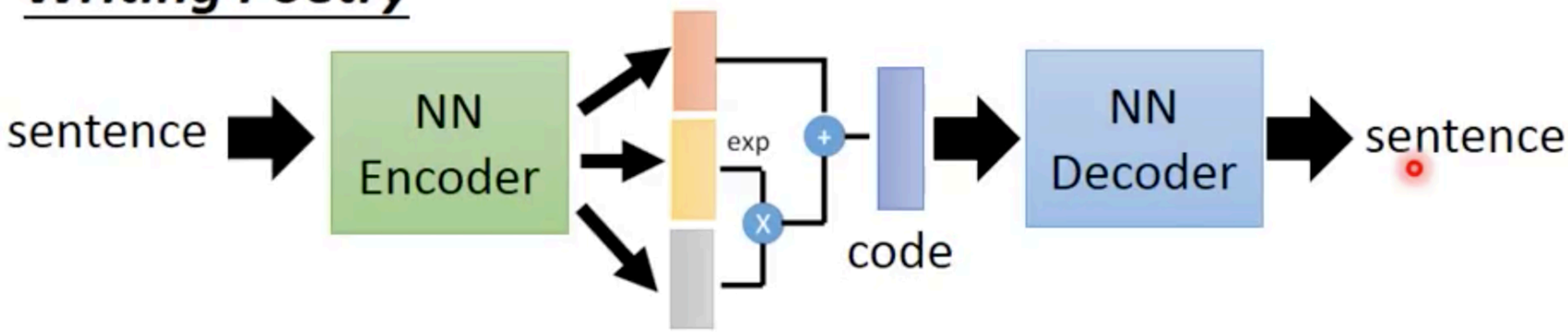
Ref: http://www.wired.co.uk/article/google-artificial-intelligence-poetry
Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, Samy Bengio, Generating Sente… prepring, 2015

怎麼用 VAE 來寫詩呢？

Writing Poetry



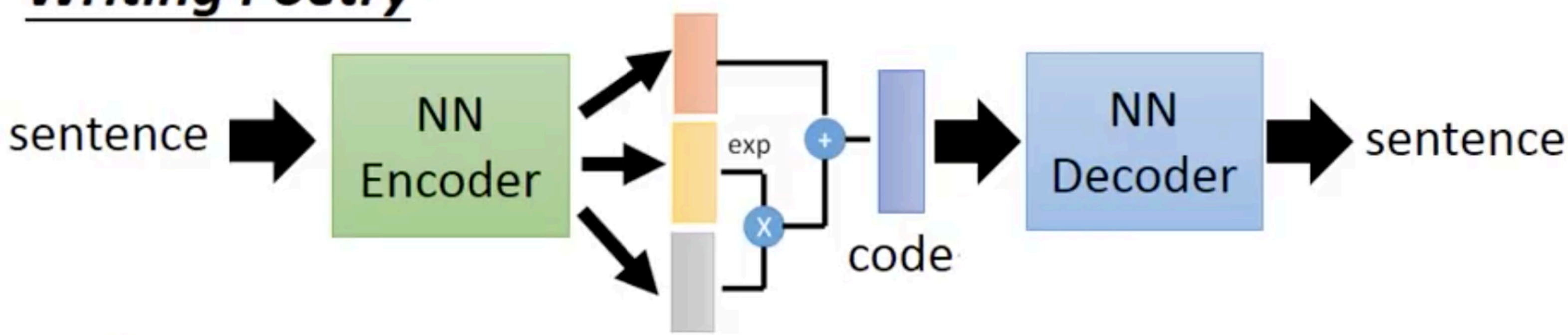
Writing Poetry



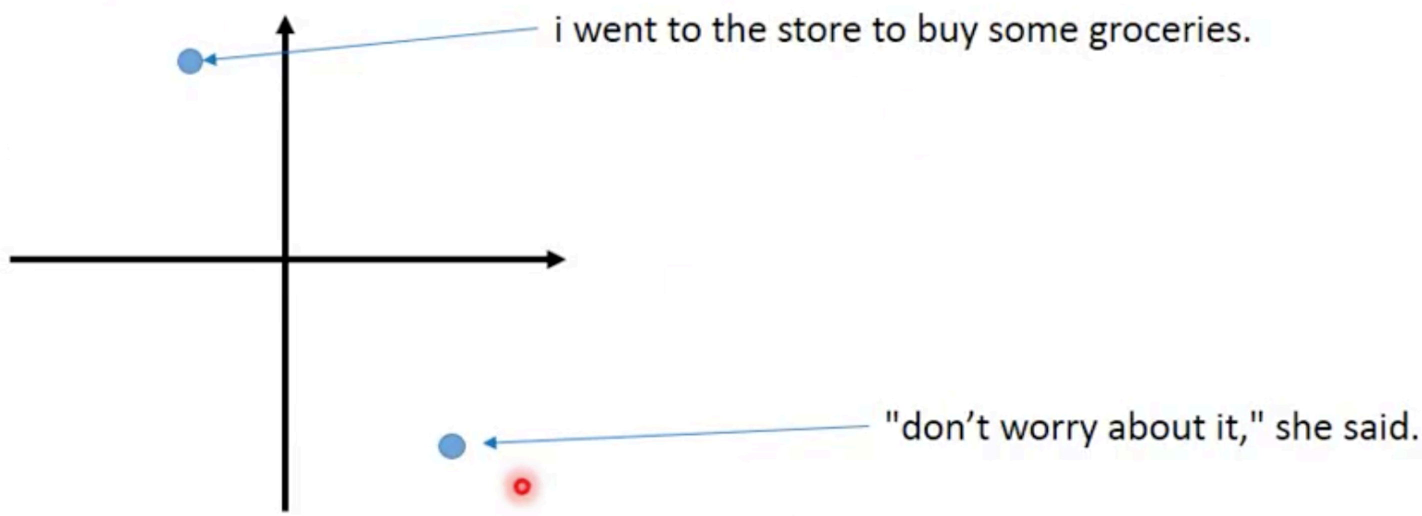
i went to the store to buy some groceries.

"don't worry about it," she said.

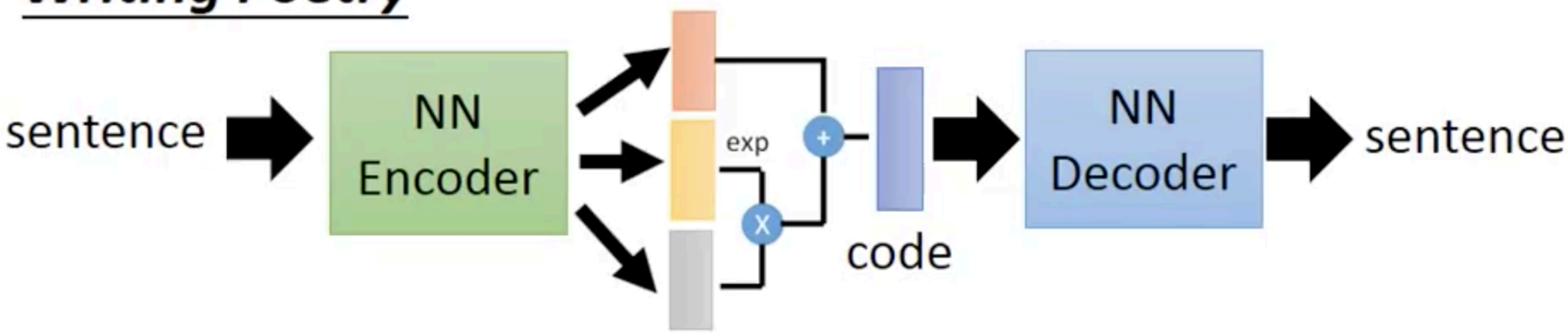
Writing Poetry

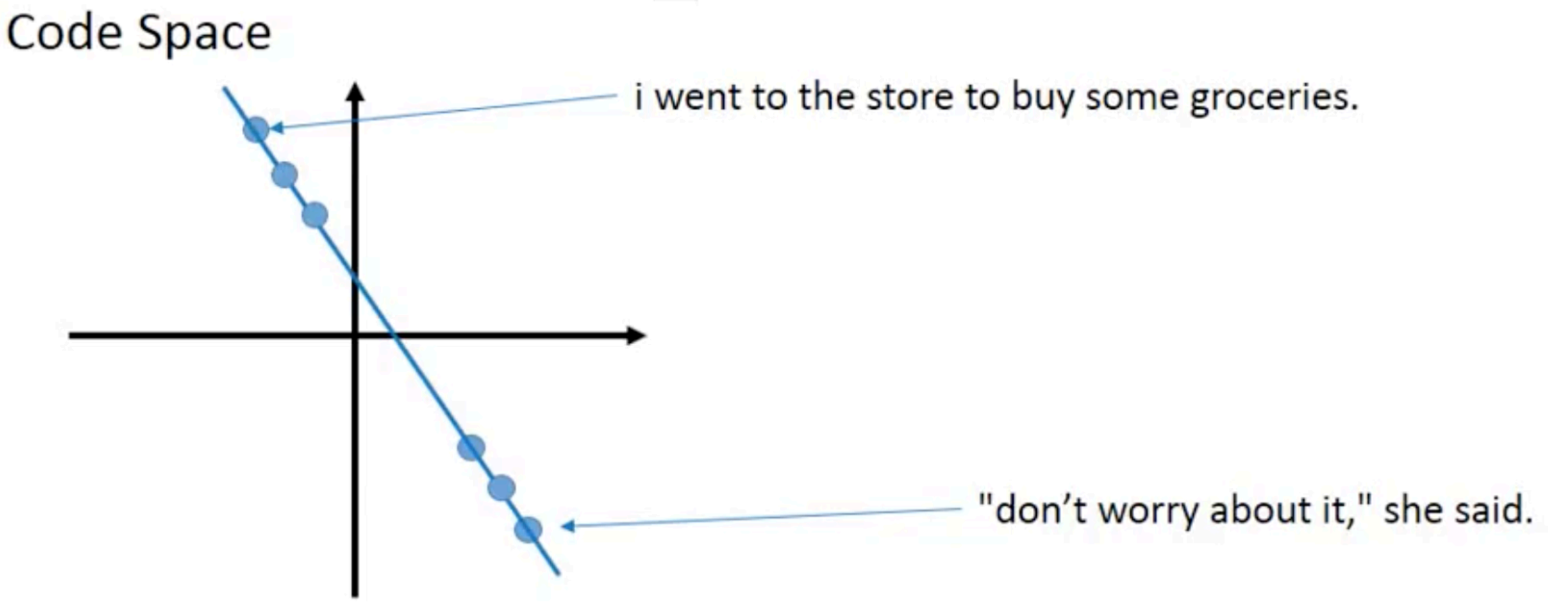


Code Space

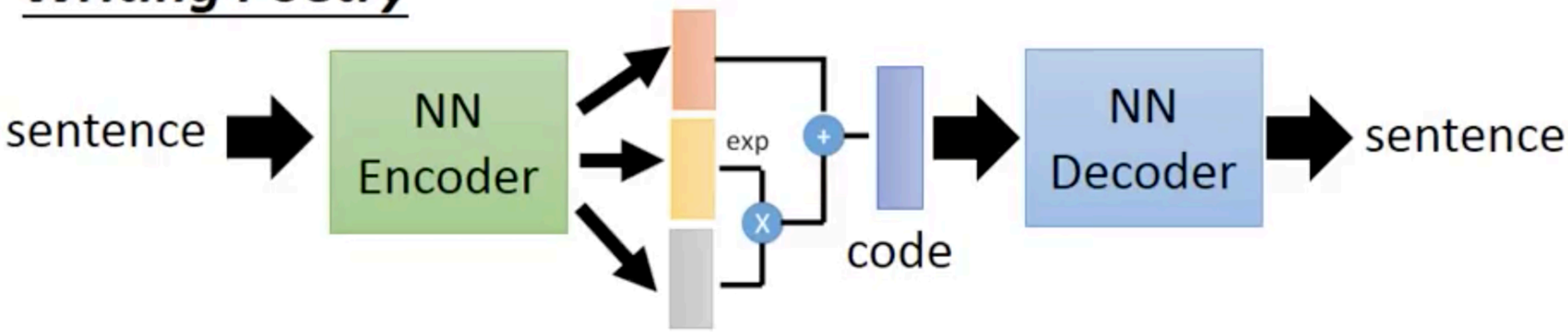


Writing Poetry

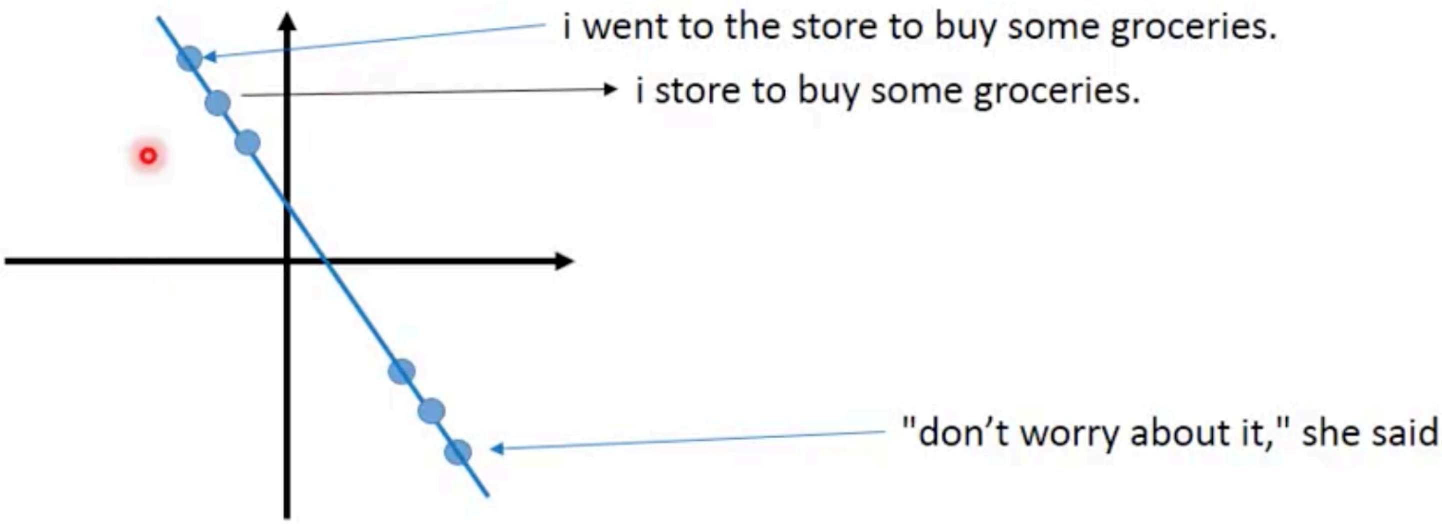




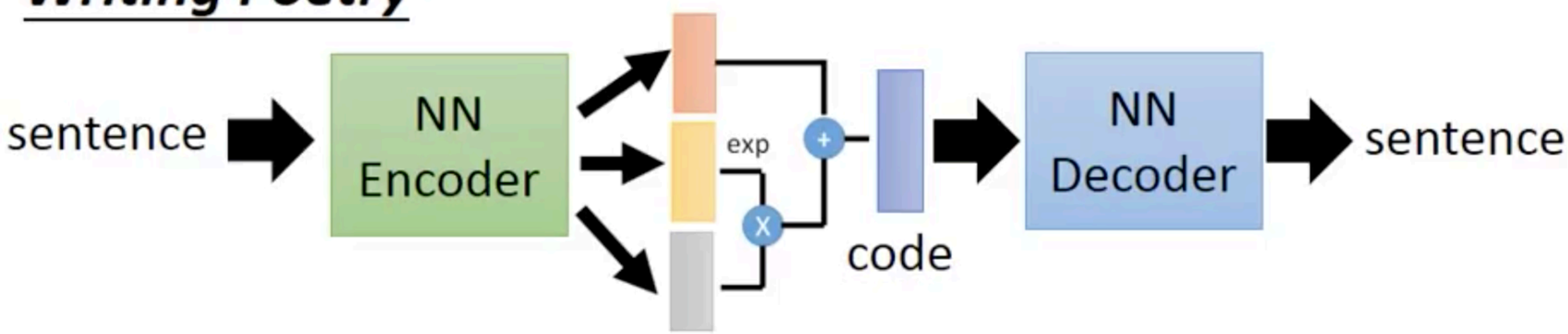
Writing Poetry

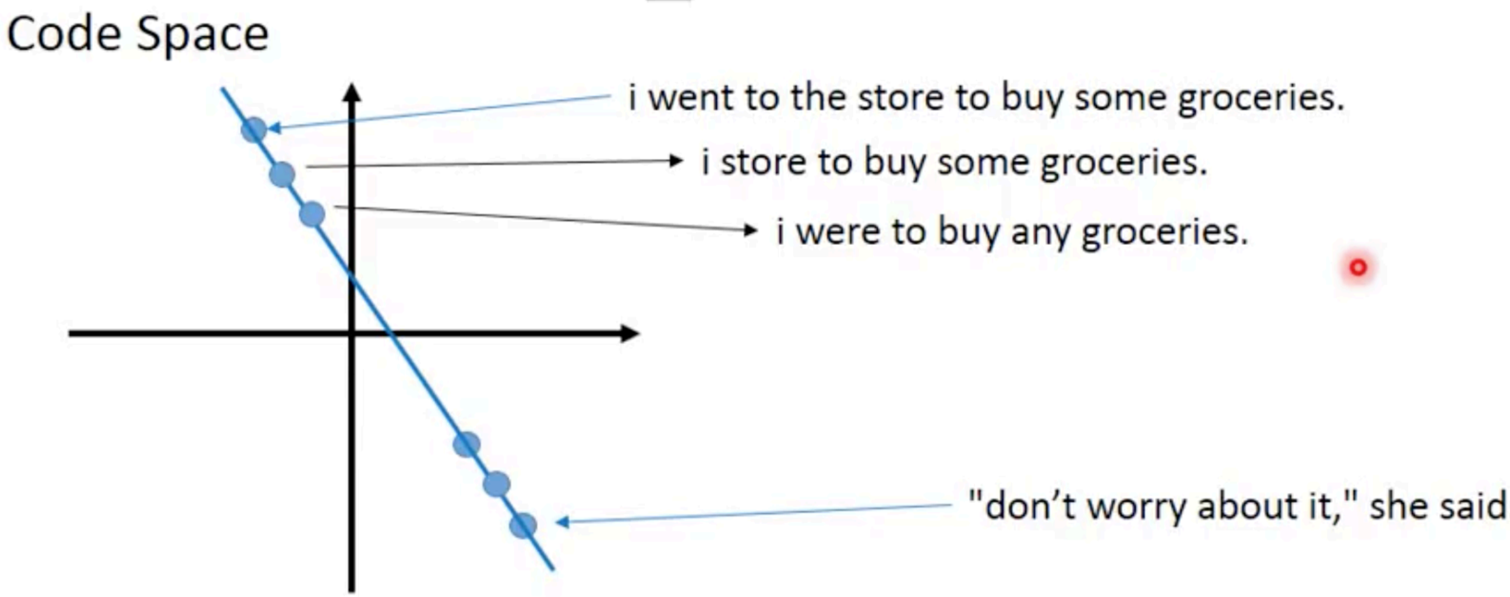


Code Space



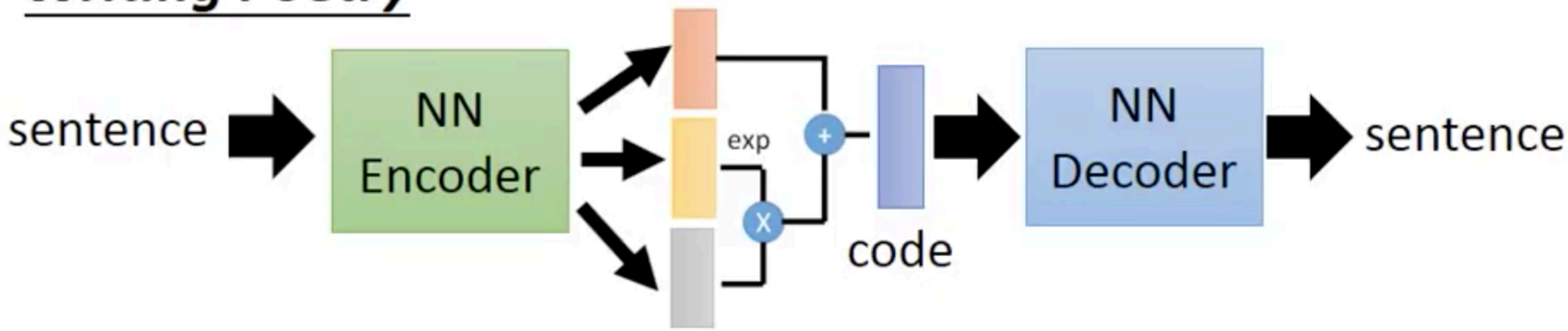
Writing Poetry



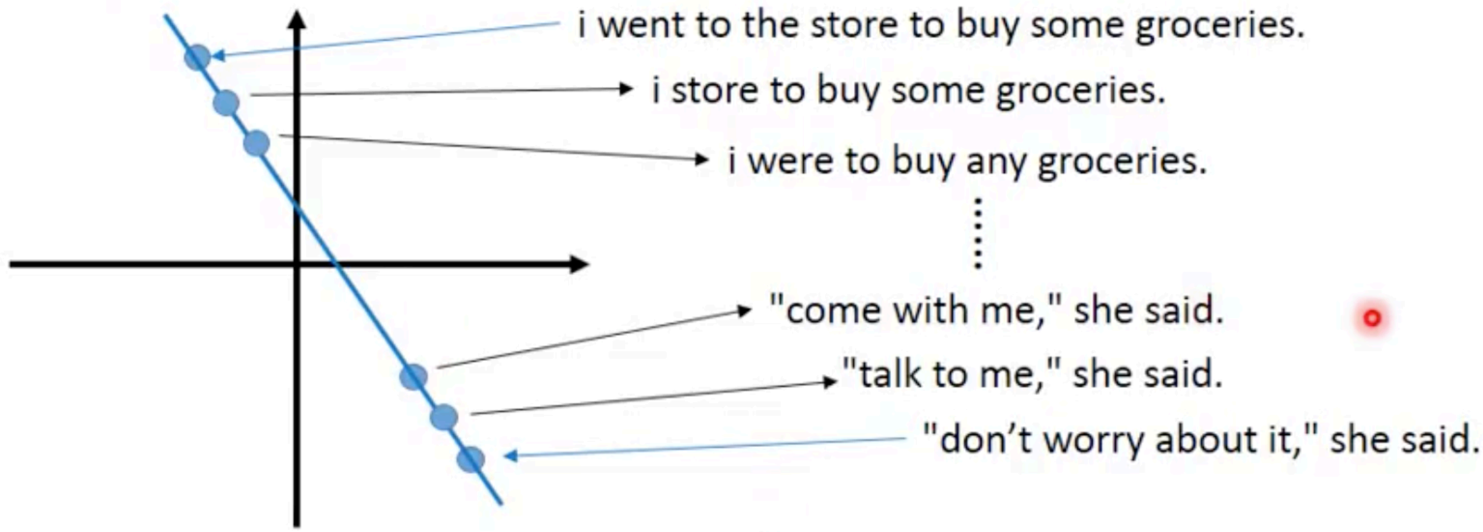


Ref: http://www.wired.co.uk/article/google-artificial-intelligence-poetry

Writing Poetry

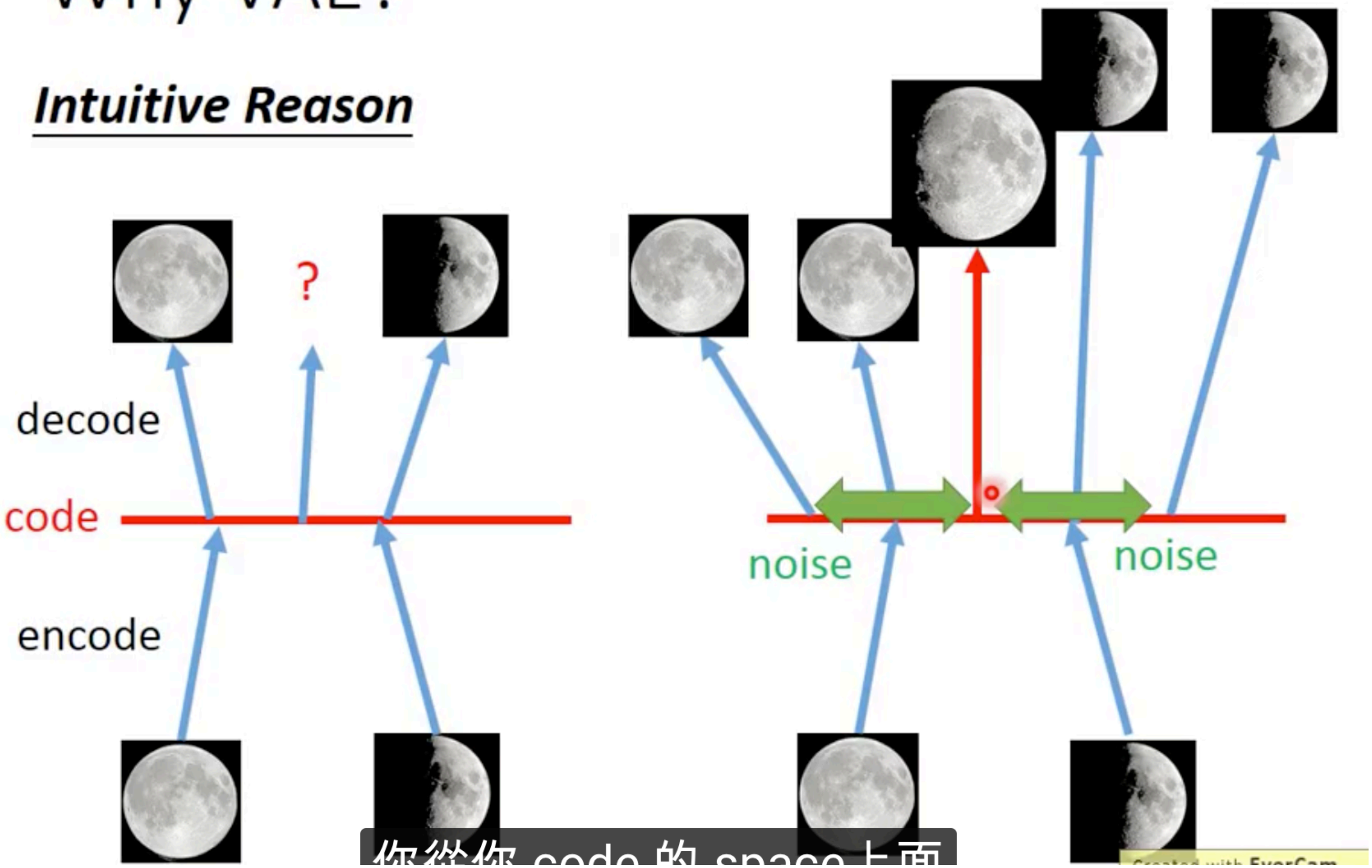


Code Space



Why VAE?

Intuitive Reason

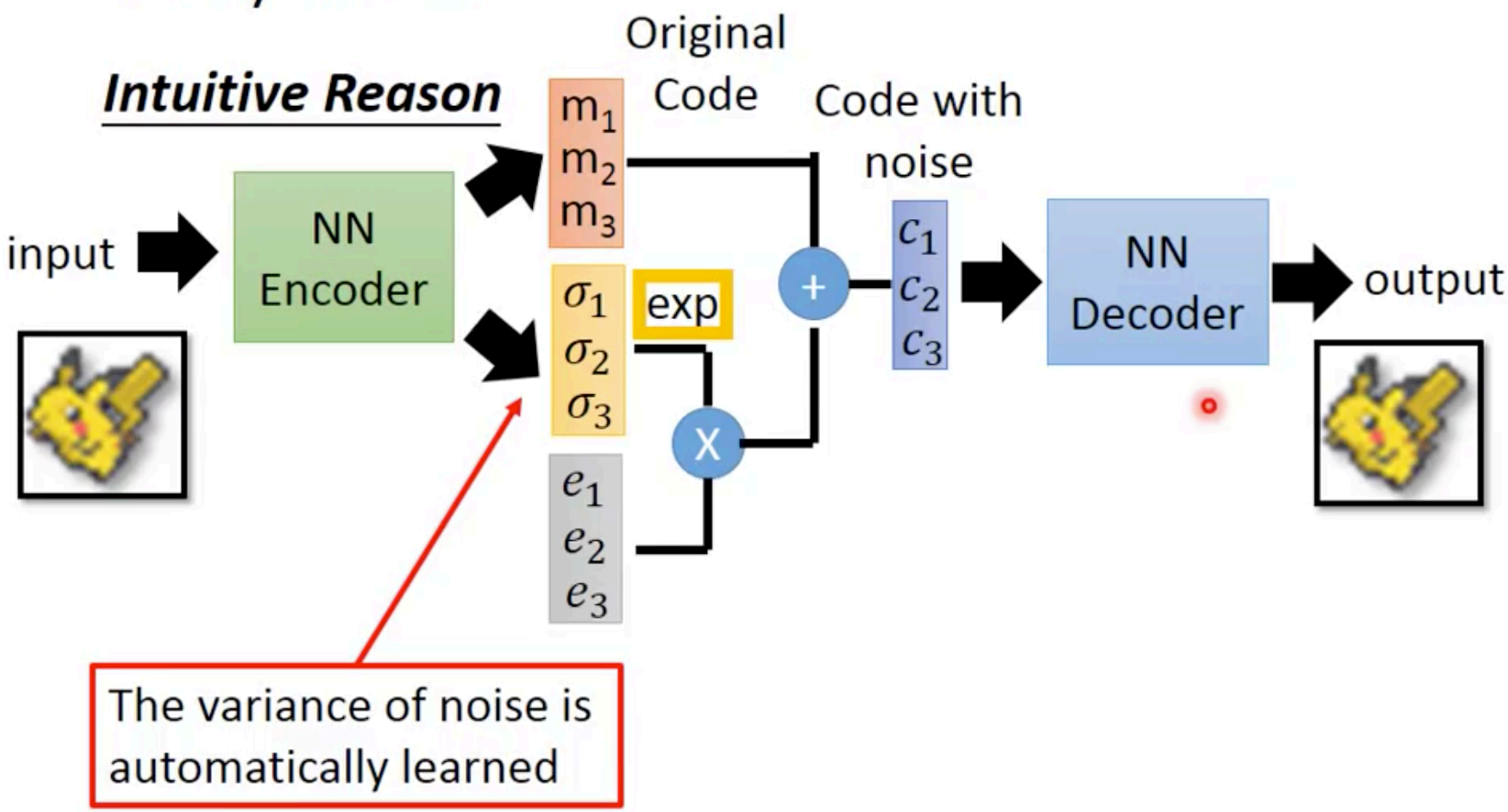


你從你 code 的 space 上面

Why VAE?

What will happen if we only minimize reconstruction error?

Intuitive Reason

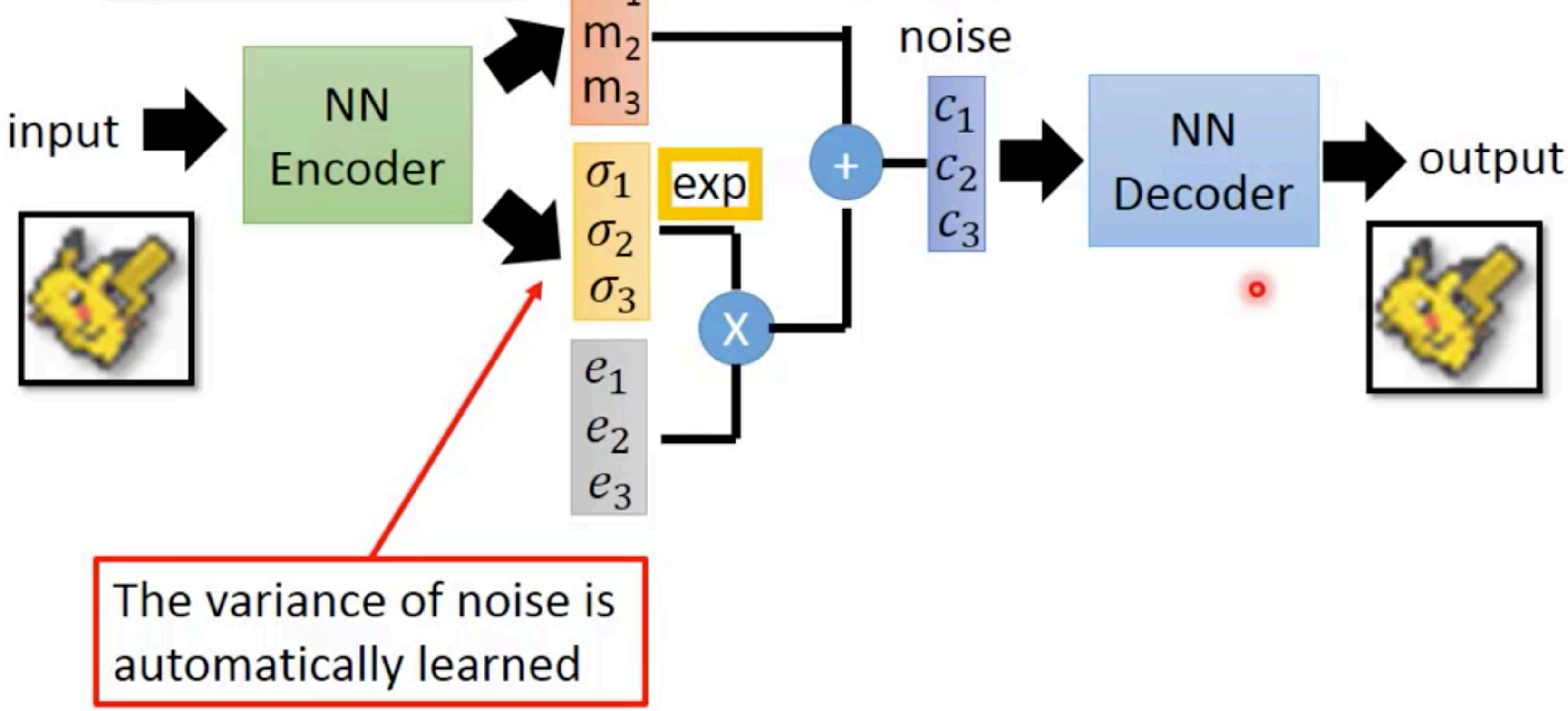


Why VAE?

What will happen if we only minimize reconstruction error?

NN will make variance be 0 during learning.

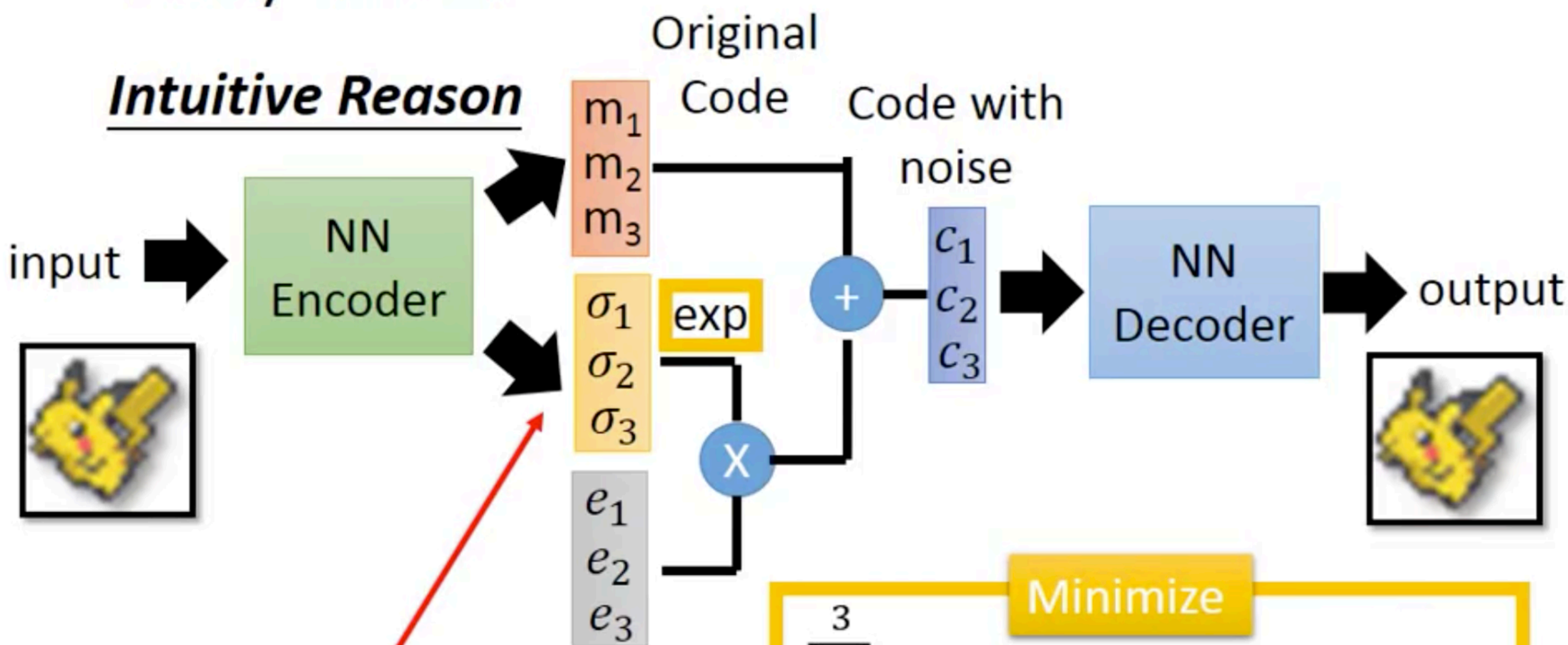
Intuitive Reason



Why VAE?

What will happen if we only minimize reconstruction error?

Intuitive Reason



The variance of noise is automatically learned

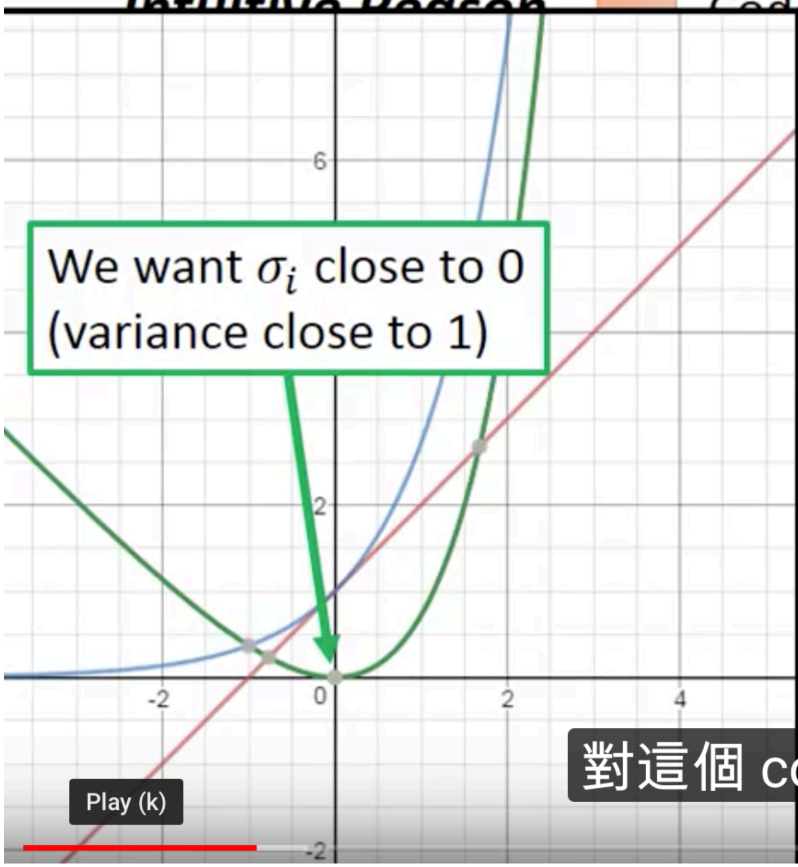
Minimize

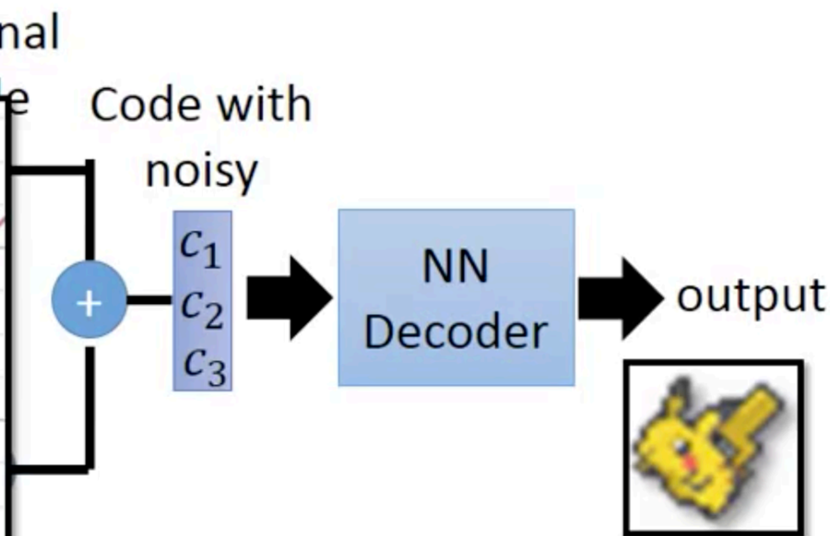
$$\sum_{i=1}^{3}\left(exp(\sigma_i) - (1+\sigma_i) + (m_i)^2\right)$$

Why VAE?

What will happen if we only minimize reconstruction error?

Intuitive Reason





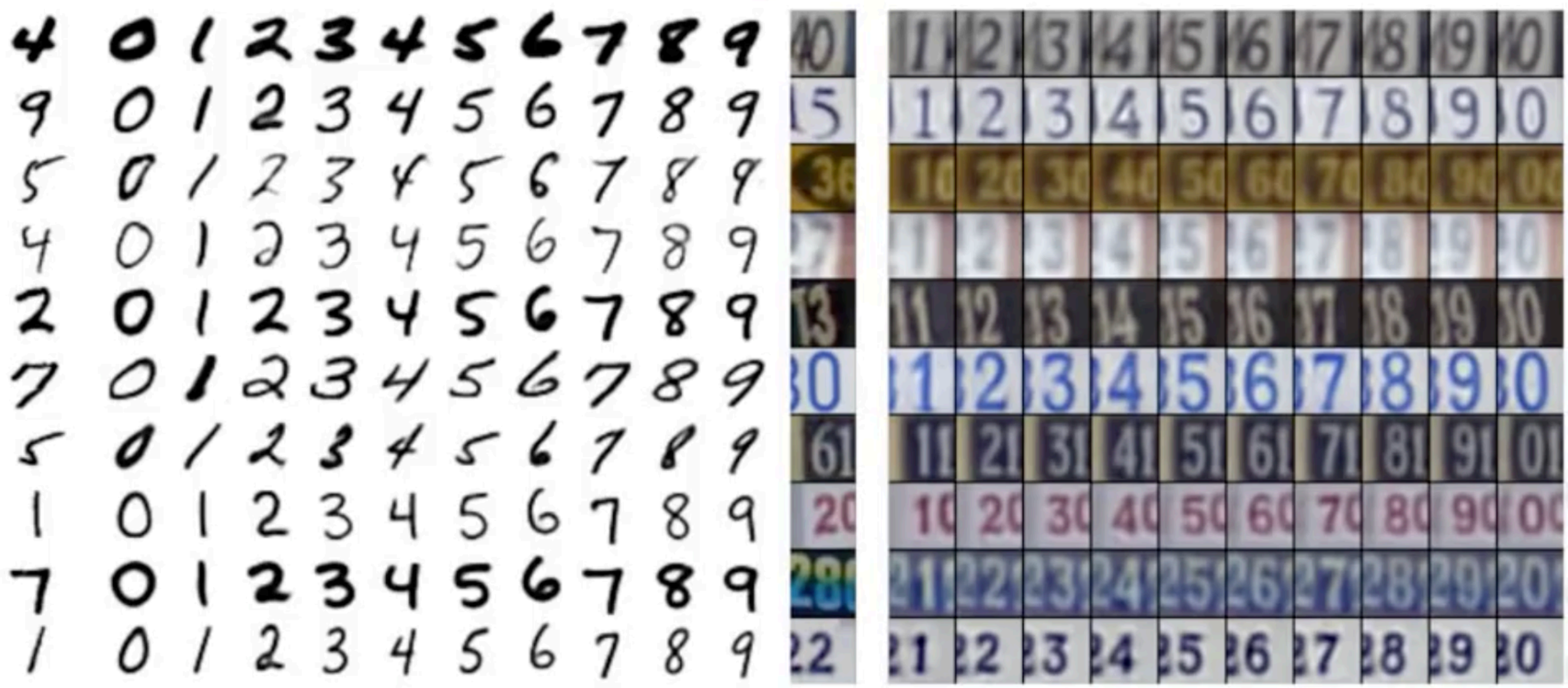
Minimize

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

L2 regularization

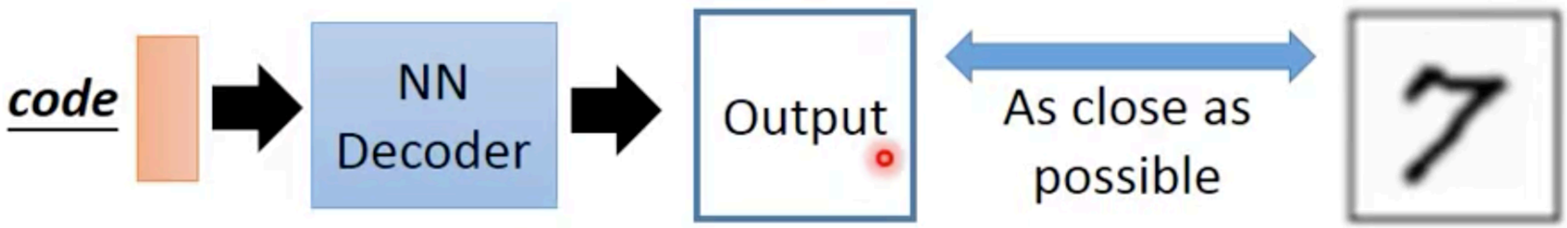
對這個 code 做這個

Conditional VAE



Problems of VAE

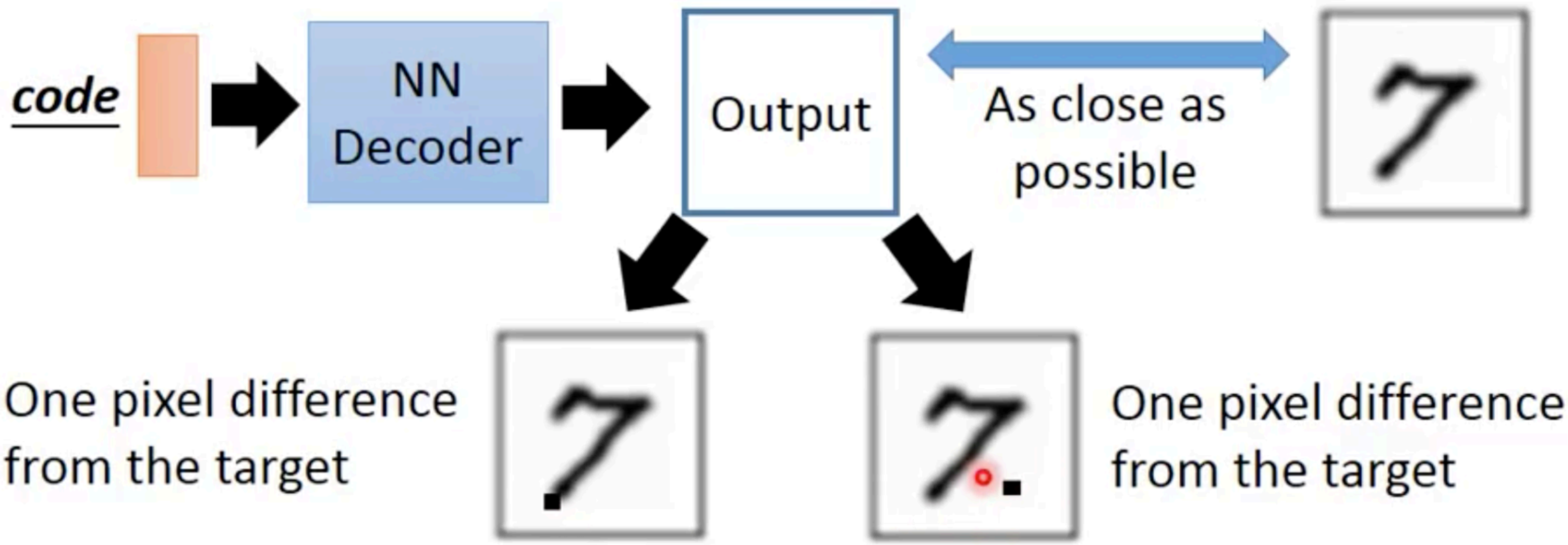
- It does not really try to simulate real images



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

Problems of VAE

- It does not really try to simulate real images



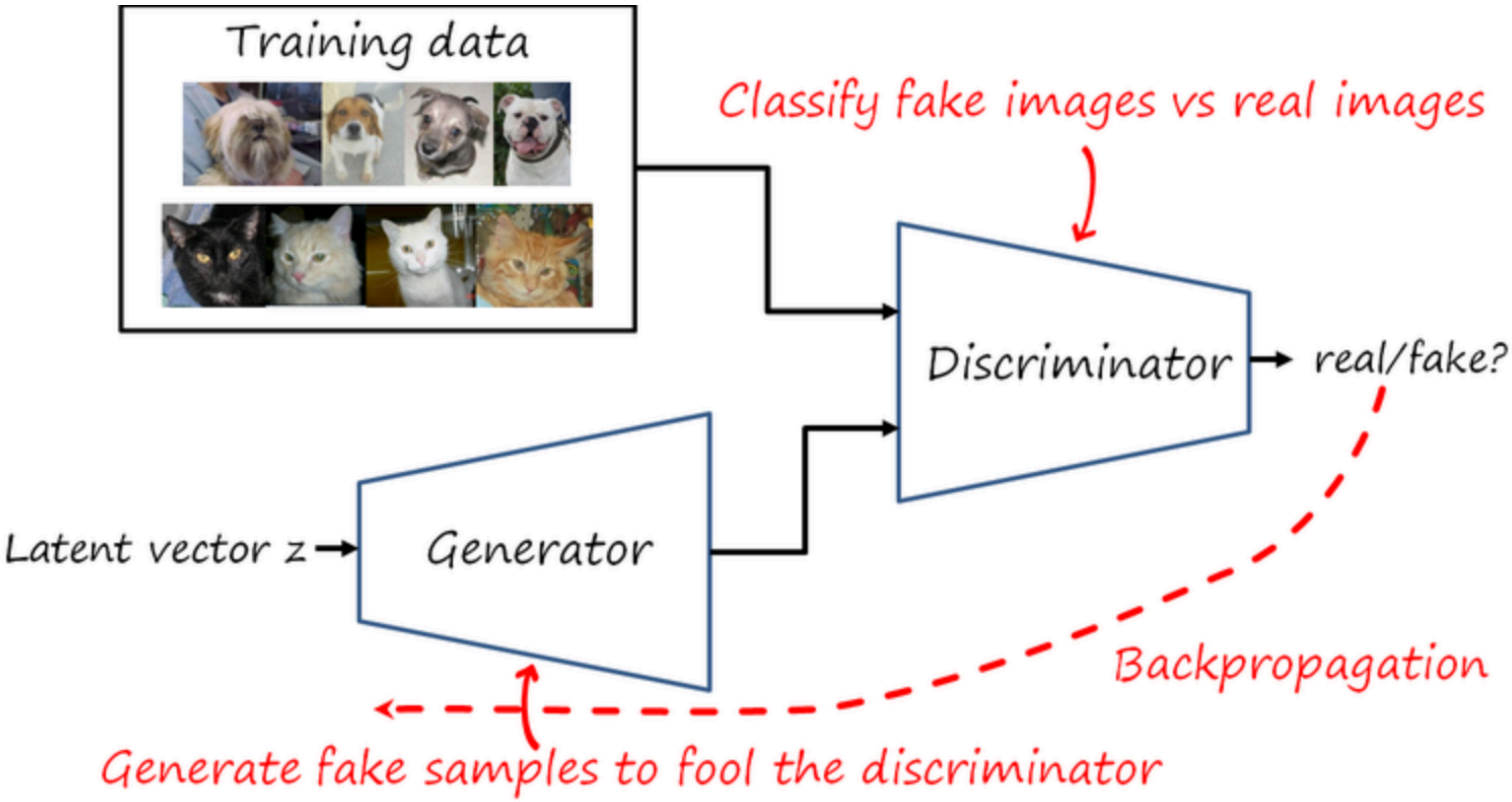
Generative Models

PixelRNN

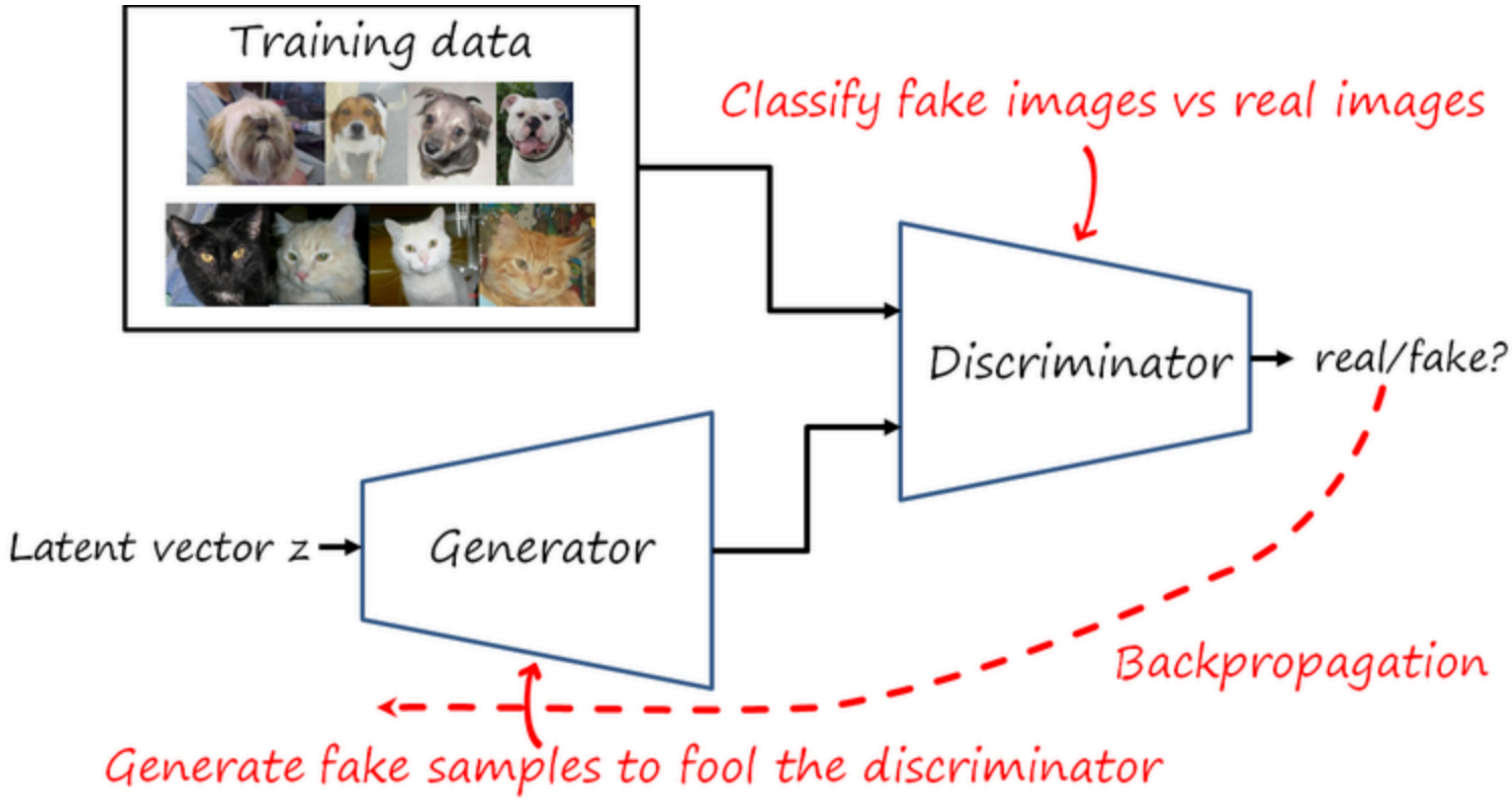
Variational Autoencoder (VAE)

Generative Adversarial Network (GAN)

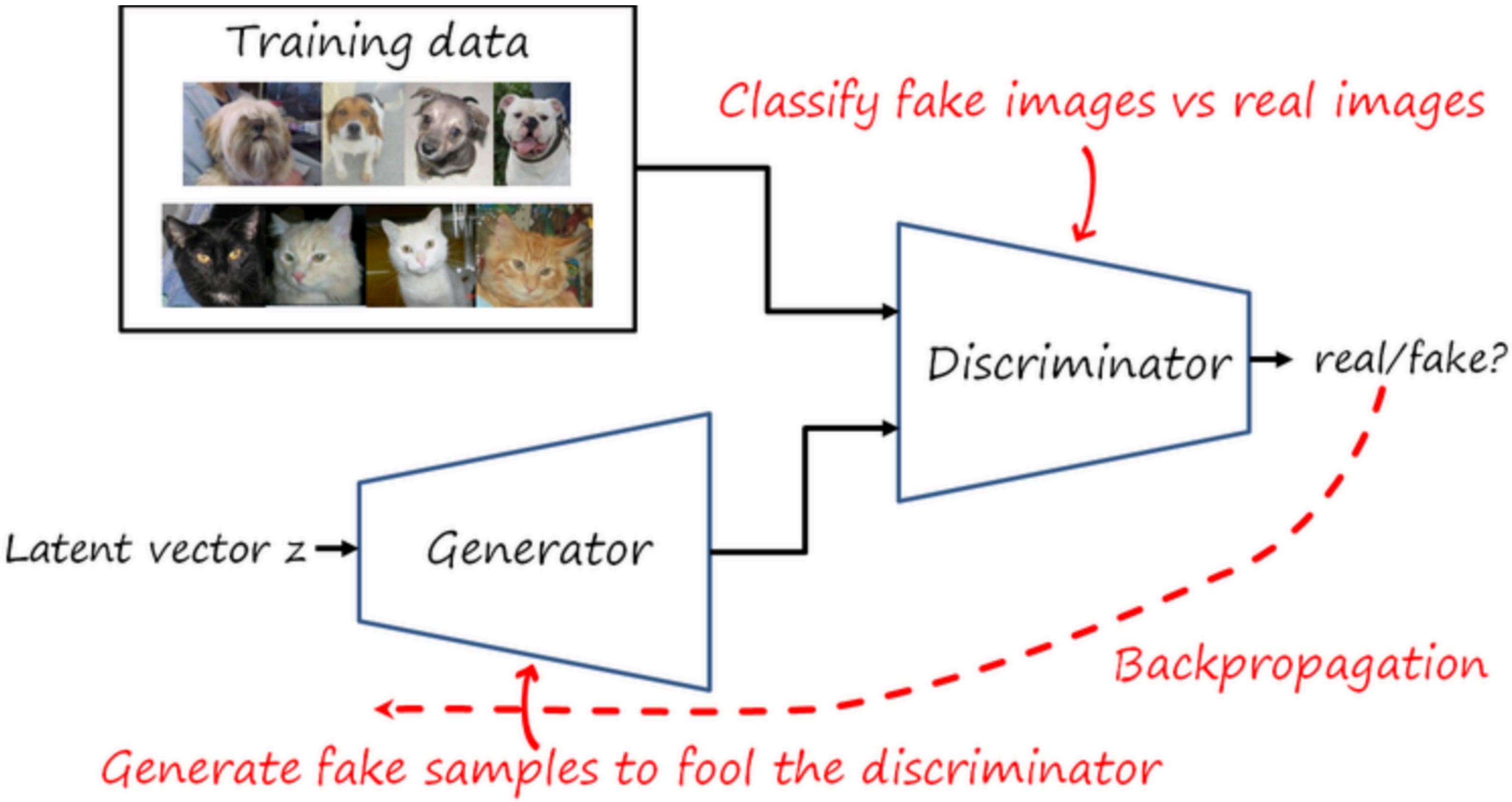
Ian J. Good fellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks, arXiv preprint 2014



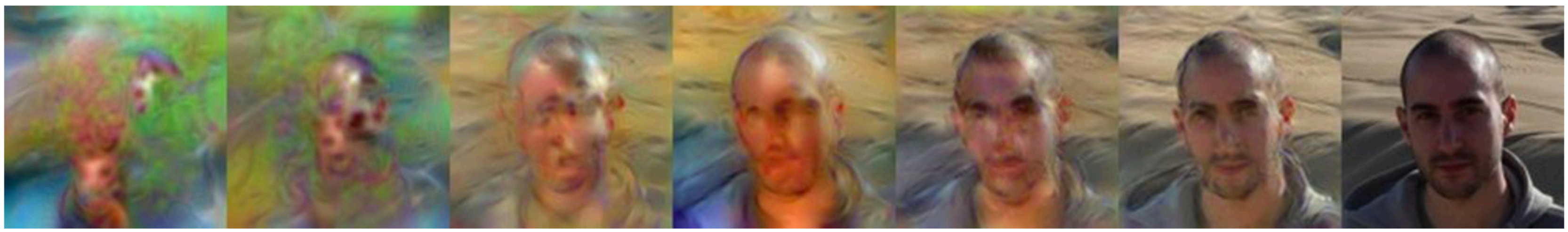
http://www.lherranz.org/2018/08/07/imagetranslation/



Iterate: (1) Fix generator, train discriminator. (Discriminator gets better.)



Iterate: (1) Fix generator, train discriminator. (Discriminator gets better.)
(2) Fox discriminator, train generator. (Generator gets better.)

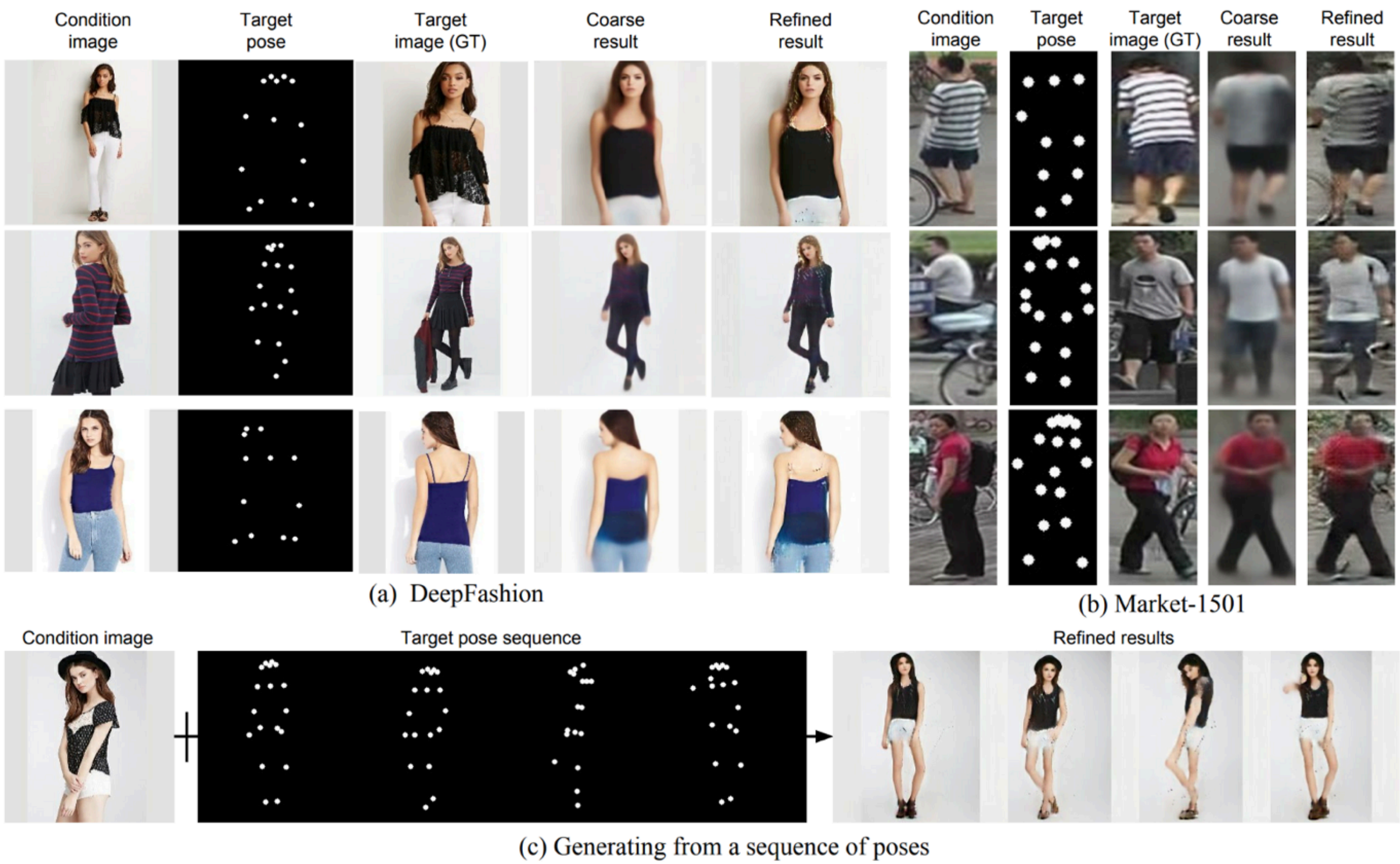


http://www.lherranz.org/2018/08/07/imagetranslation/

Images generated using Progressive GAN

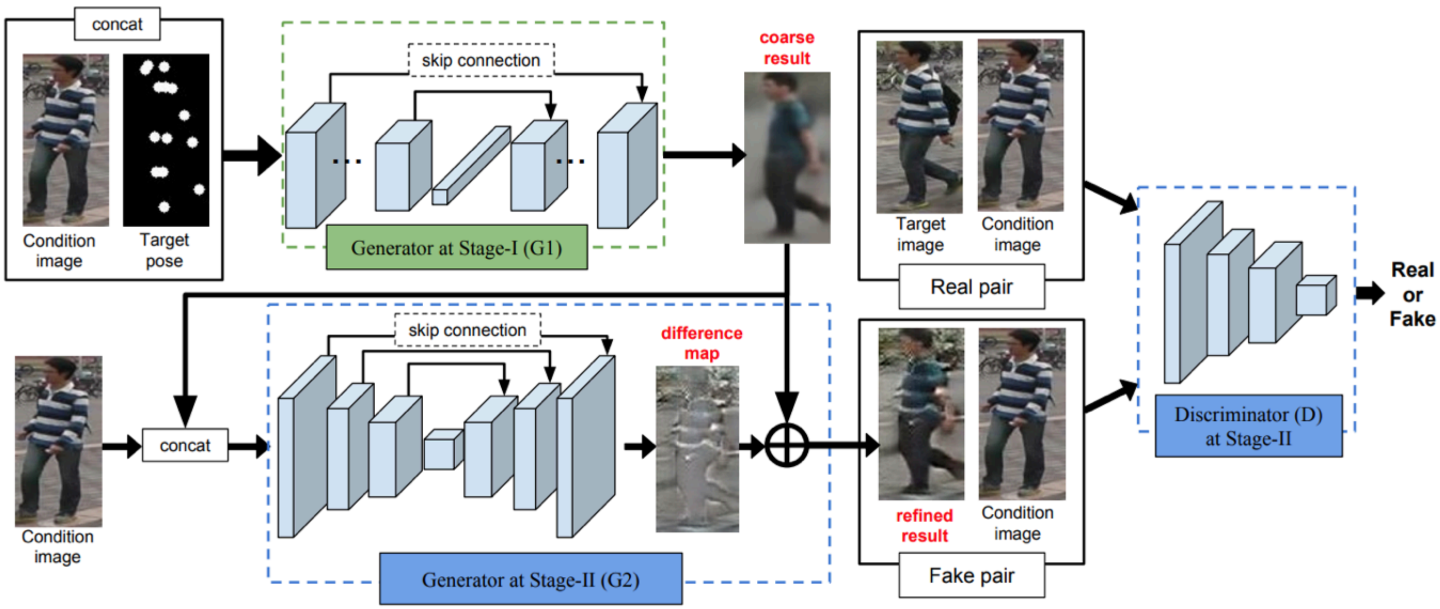
Figure 5: 1024×1024 images generated using the CELEBA-HQ dataset. See Appendix F for a larger set of results, and the accompanying video for latent space interpolations.

https://medium.com/@jonathan_hui/gan-some-cool-applications-of-gans-4c9ecca35900



Pose Guided Person Image Generation

https://medium.com/@jonathan_hui/gan-some-cool-applications-of-gans-4c9ecca35900

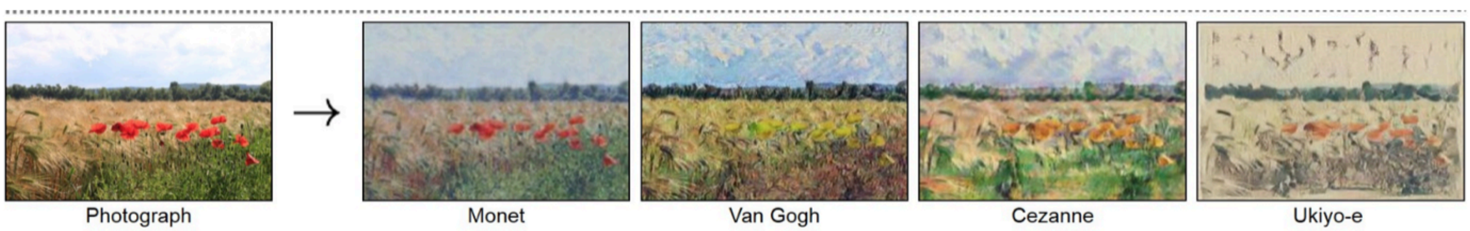


Pose Guided Person Image Generation

https://medium.com/@jonathan_hui/gan-some-cool-applications-of-gans-4c9ecca35900

CycleGAN

Cross-domain transfer GANs will be likely the first batch of commercial applications. These GANs transform images from one domain (say real scenery) to another domain (Monet paintings or Van Gogh).

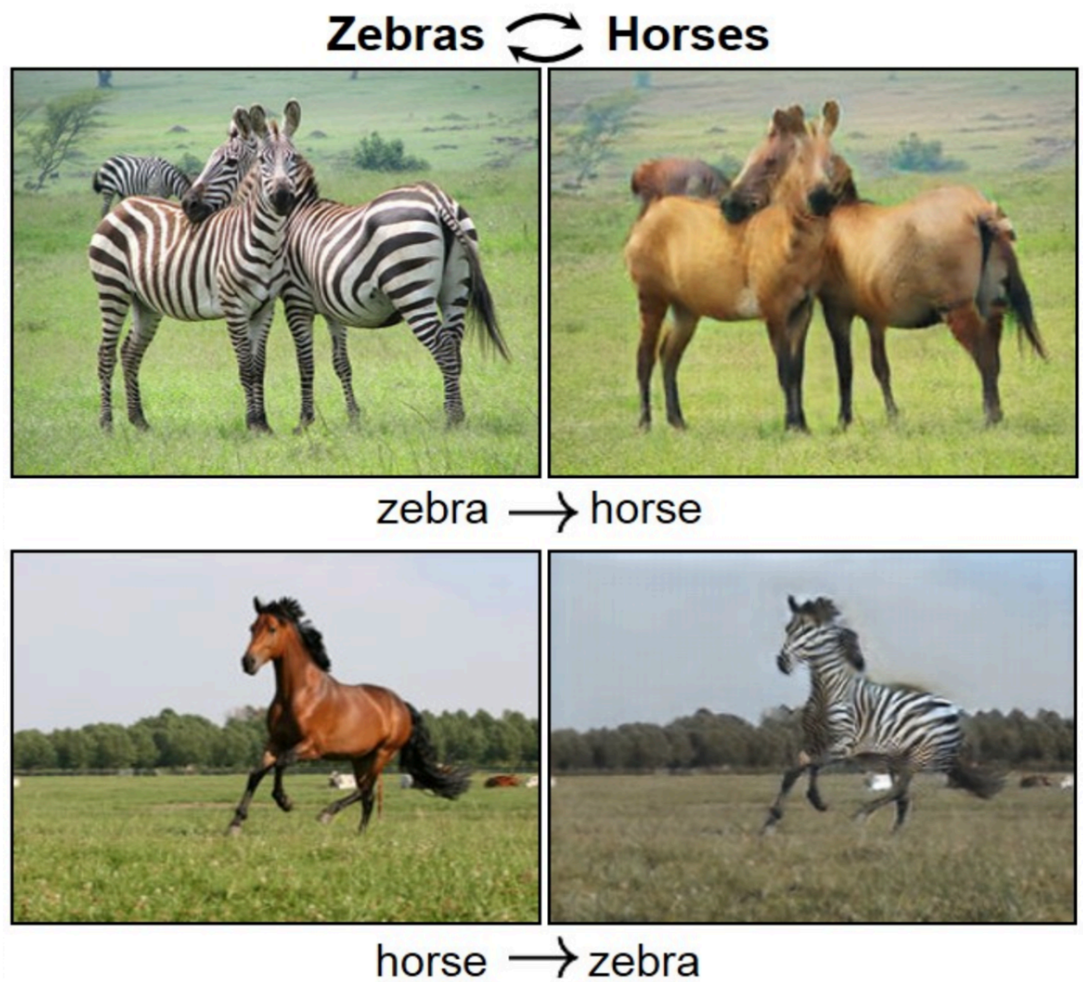


CycleGAN

https://medium.com/@jonathan_hui/gan-some-cool-applications-of-gans-4c9ecca35900


Photograph
Monet
Van Gogh
Cezanne
Ukiyo-e

For example, it can transform pictures between zebras and horses.



CycleGAN

PixelDTGAN

Suggesting merchandise based on celebrity pictures has been popular for fashion blogger and e-commerce. PixelDTGAN creates clothing images and styles from an image.

A source image. Possible target images.

Super resolution

Create super-resolution images from the lower resolution. This is one area where GAN shows very impressive result with immediate commercial possibility.

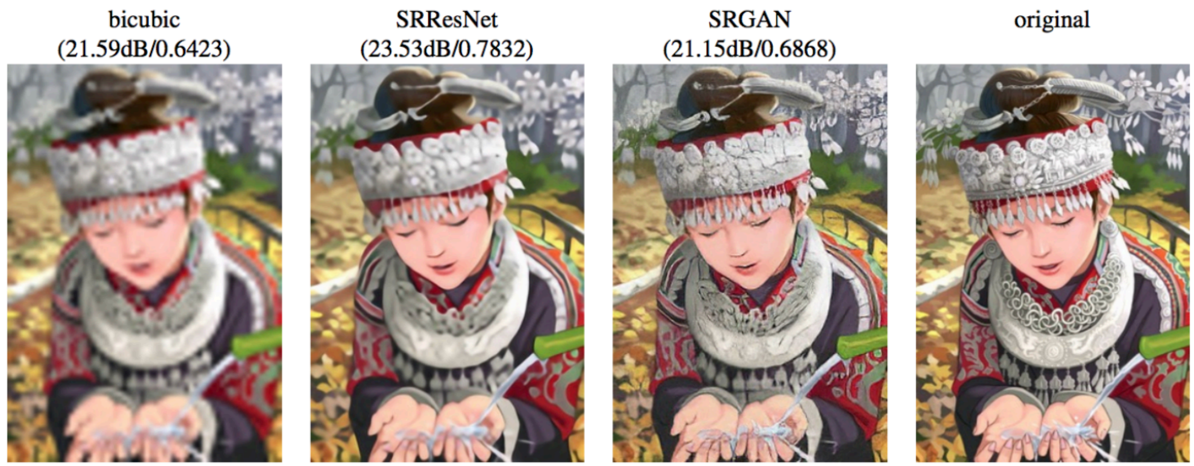


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

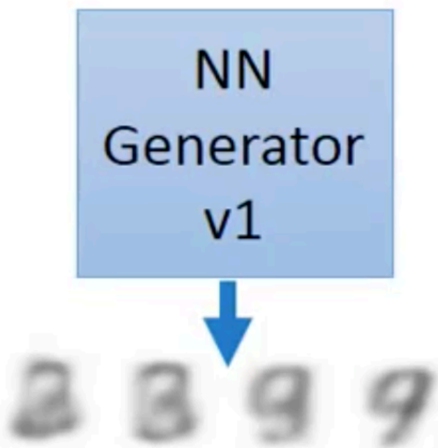
SRGAN

https://medium.com/@jonathan_hui/gan-some-cool-applications-of-gans-4c9ecca35900

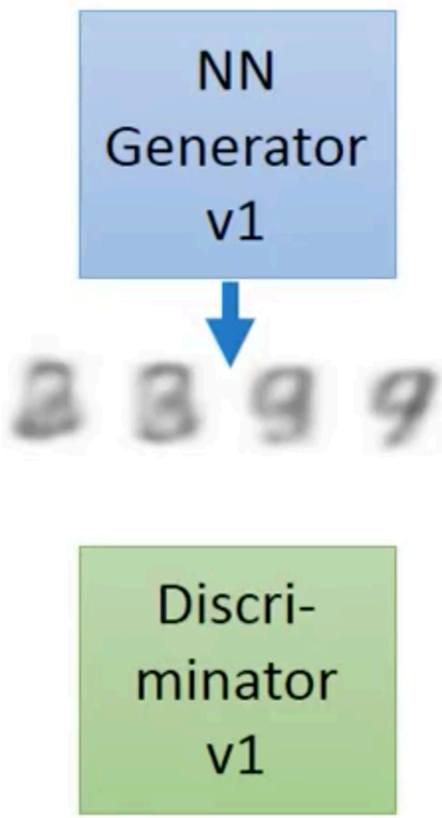
The evolution of generation

NN
Generator
v1

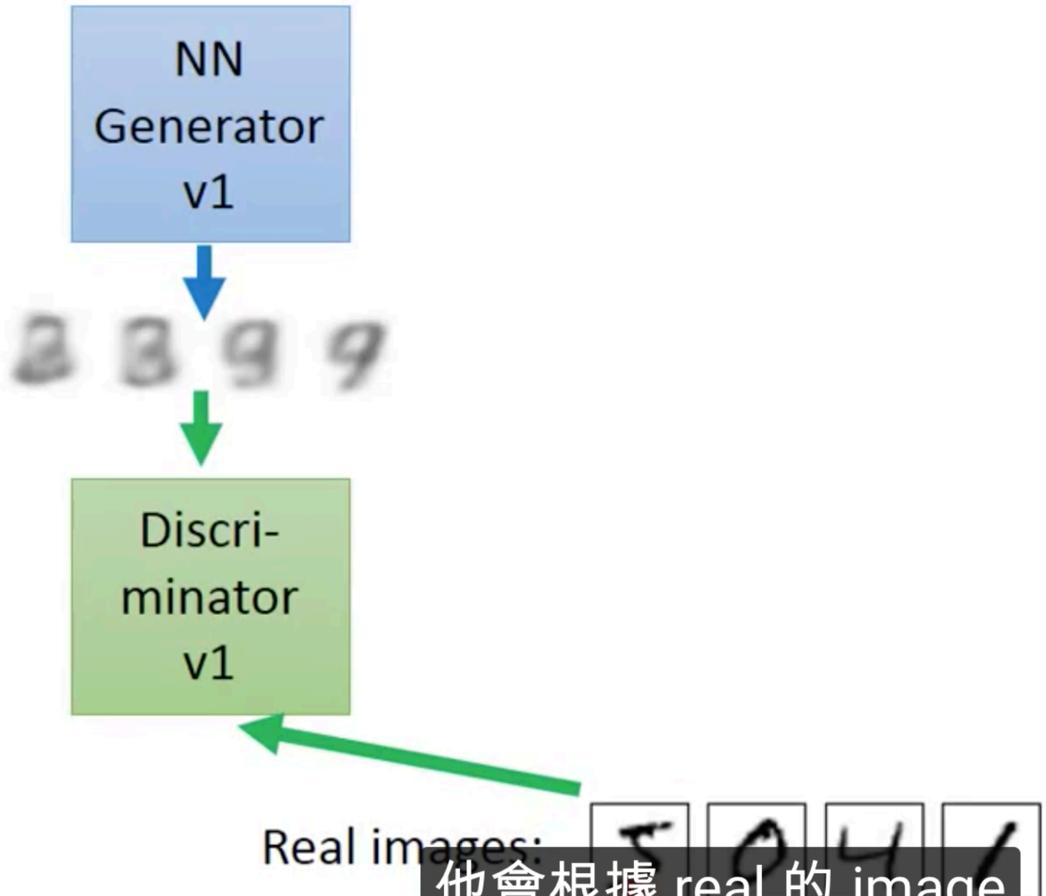
The evolution of generation



The evolution of generation

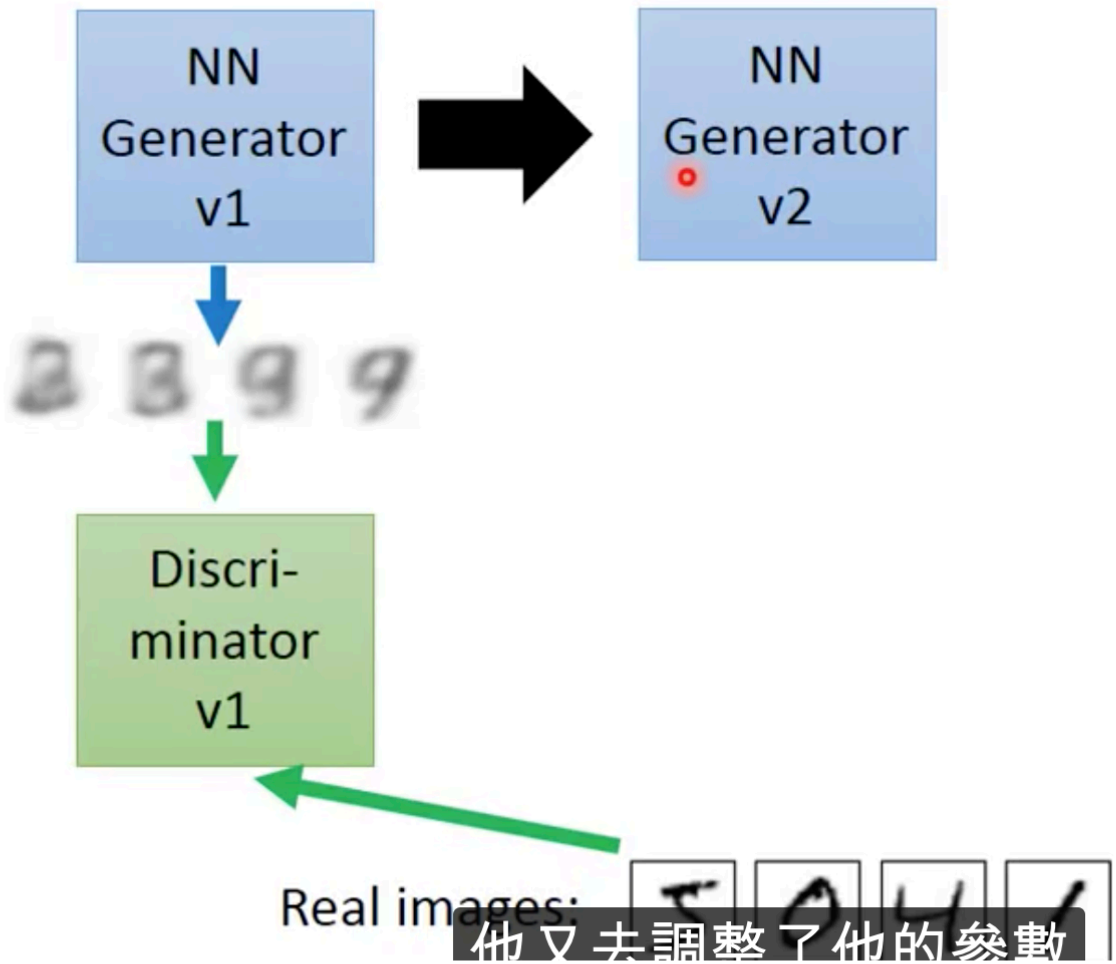


The evolution of generation



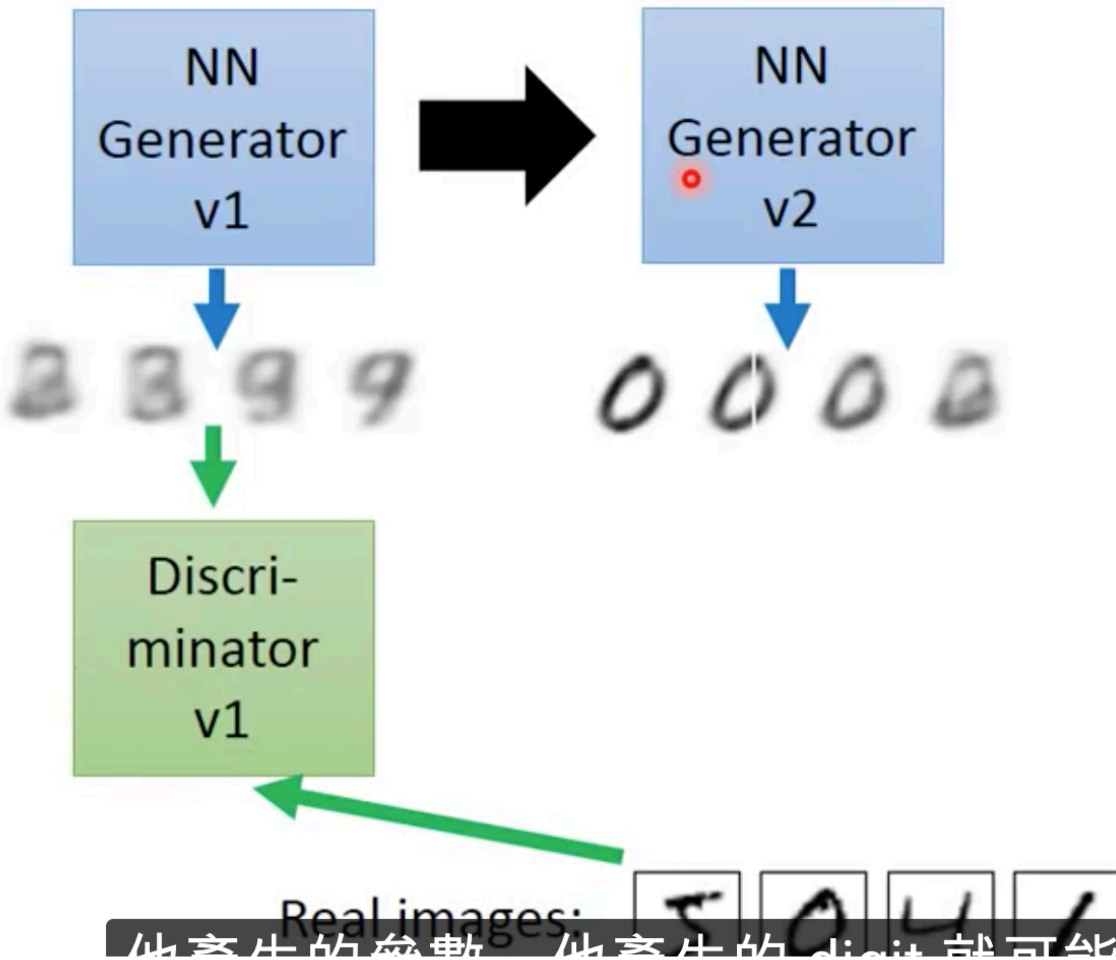
他會根據 real 的 image

The evolution of generation

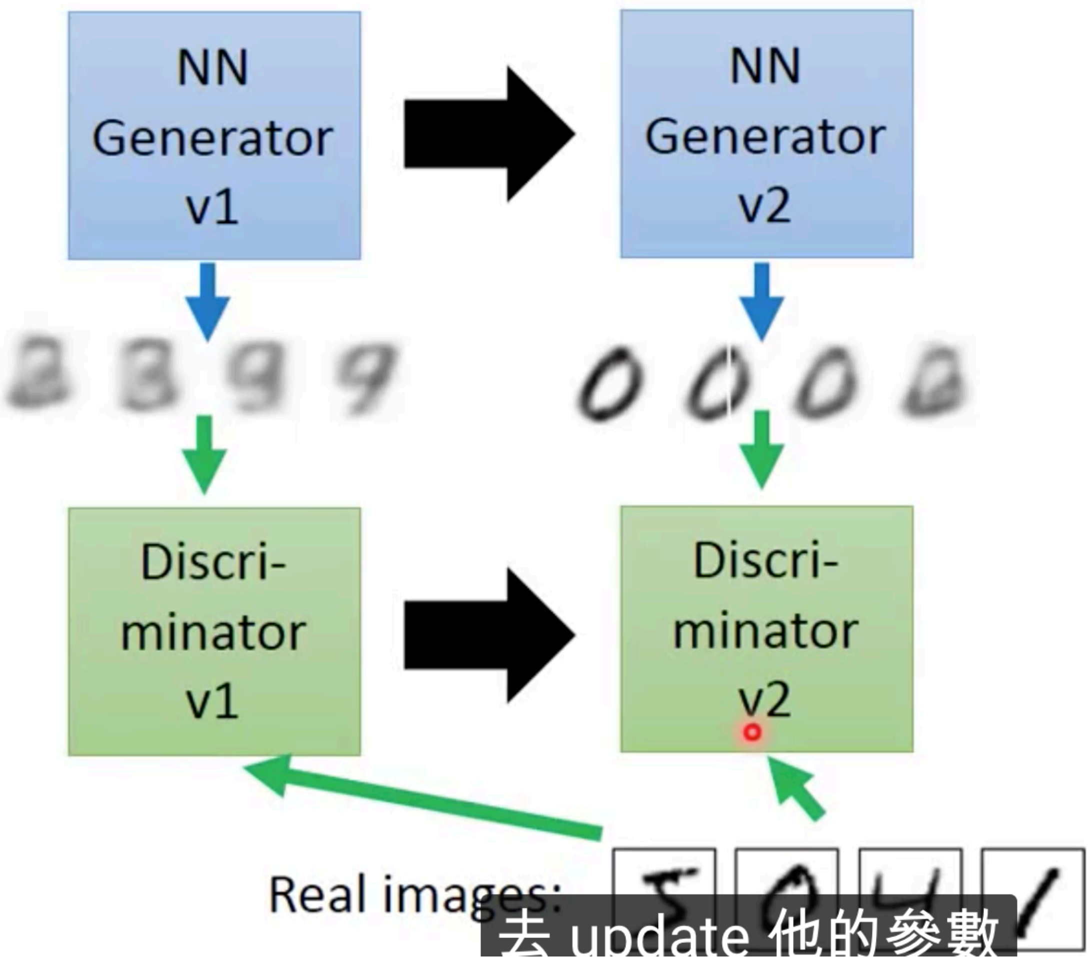


他又去調整了他的參數

The evolution of generation

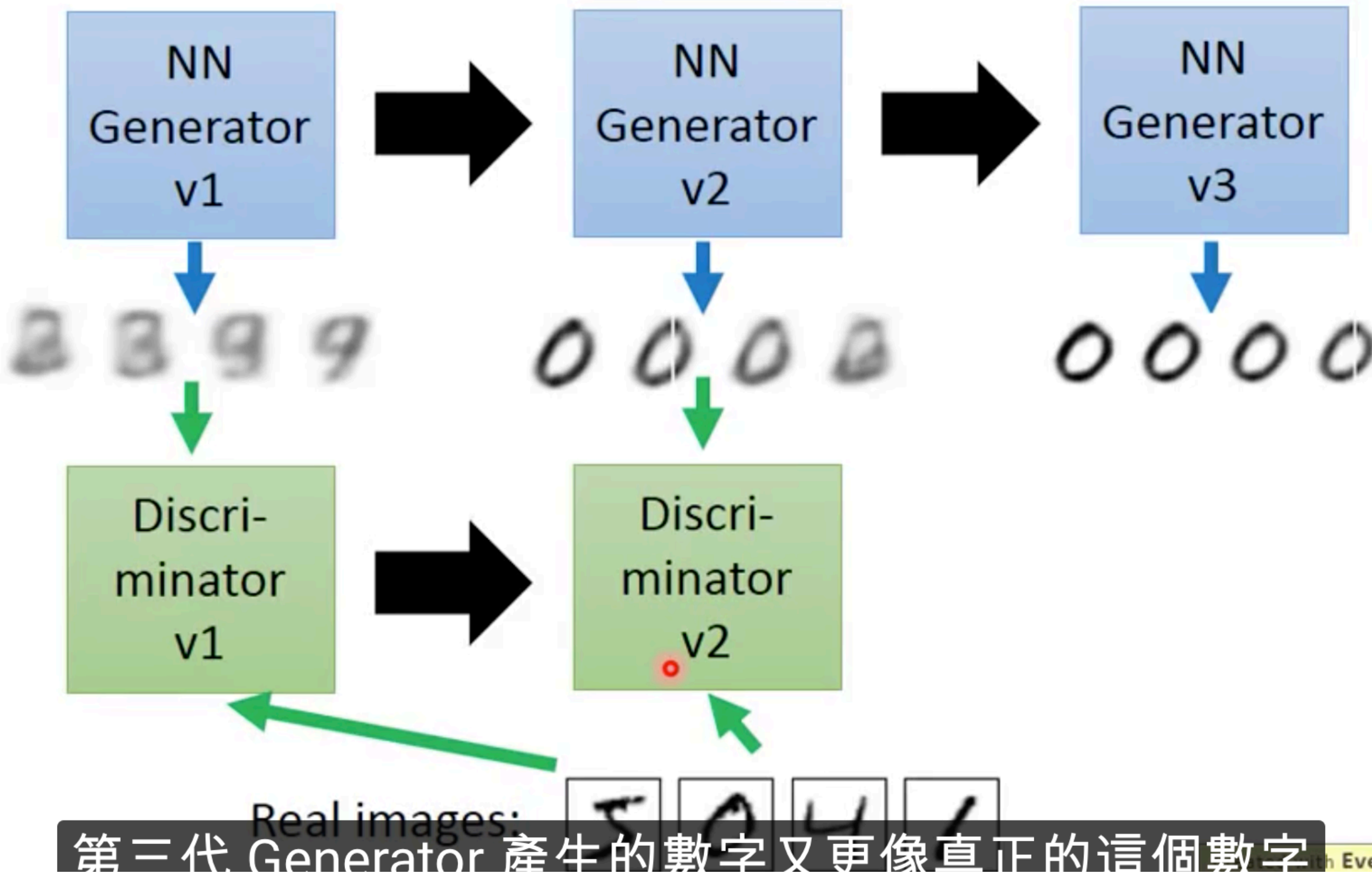


The evolution of generation



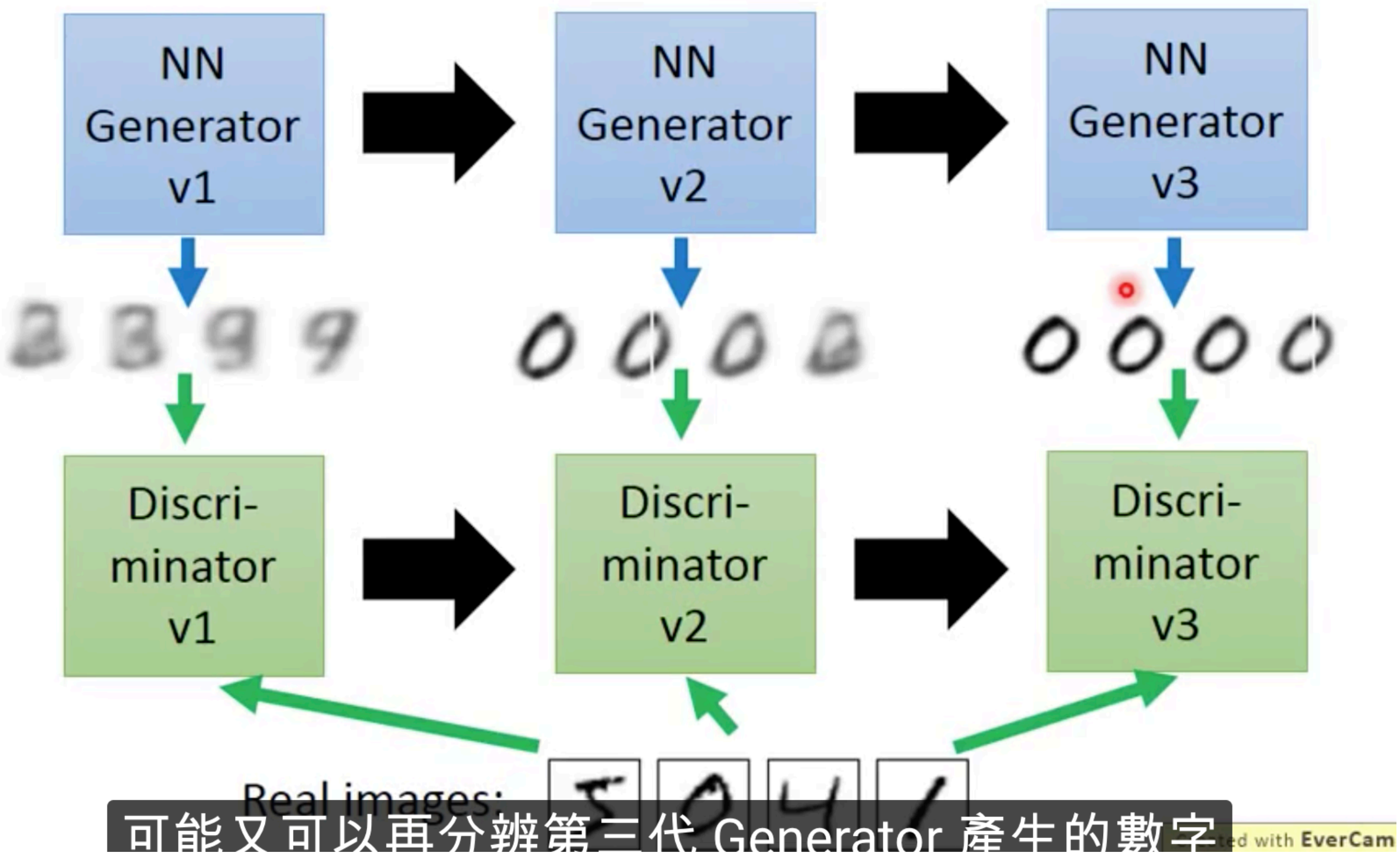
去 update 他的參數

The evolution of generation



第三代 Generator 產生的數字又更像真正的這個數字

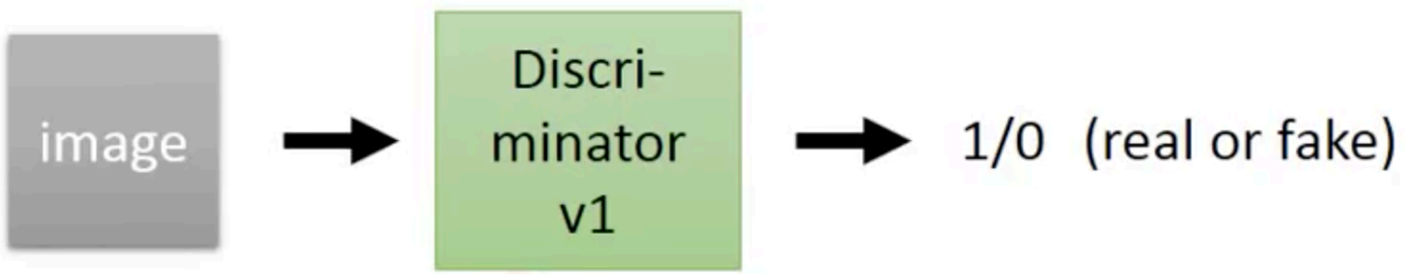
The evolution of generation



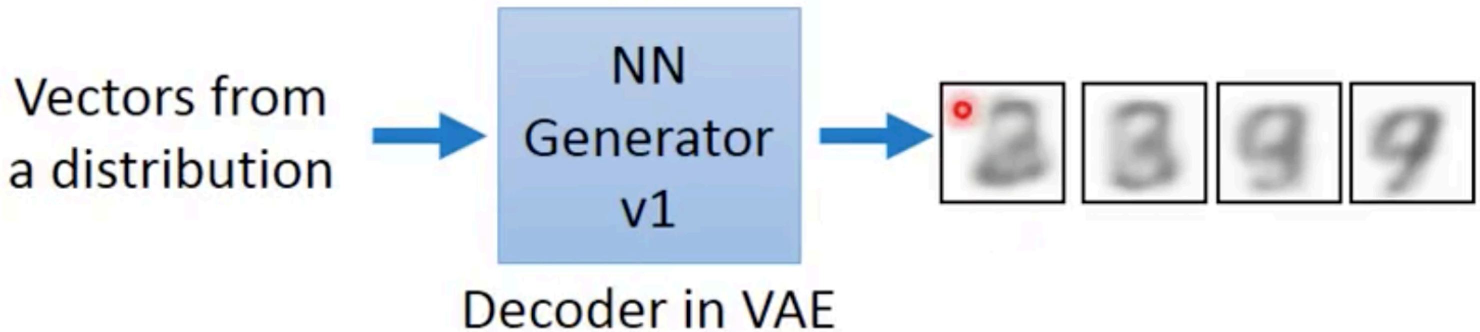
可能又可以再分辨第三代 Generator 產生的數字

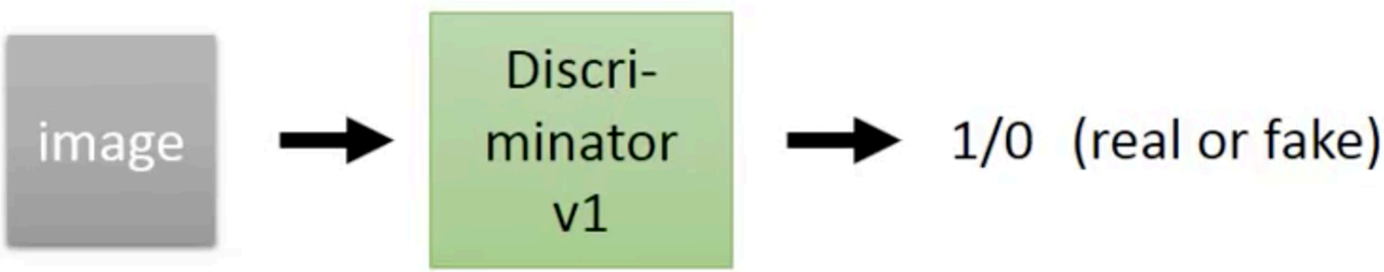
GAN - Discriminator

How to train the discriminator

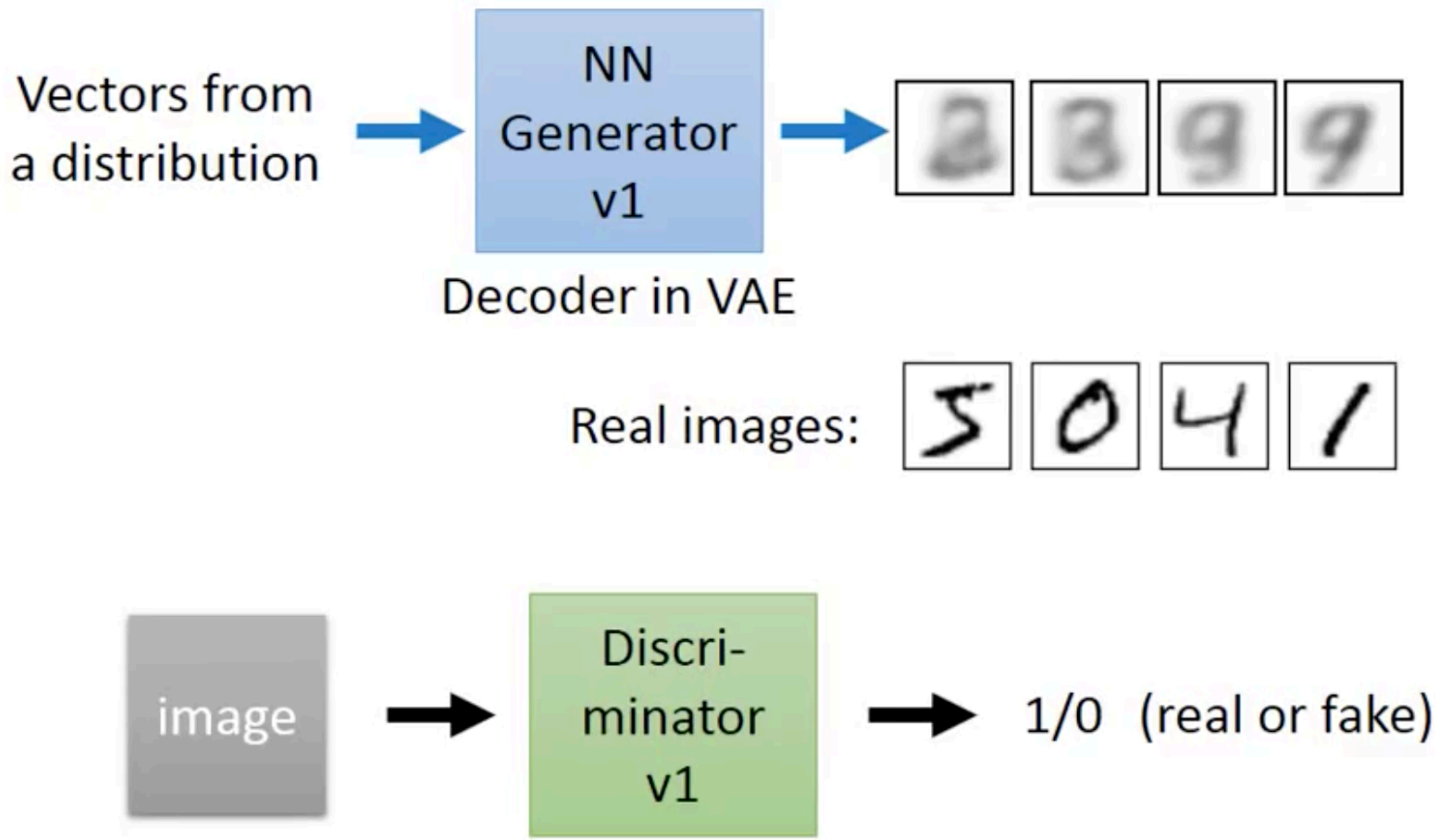


GAN - Discriminator

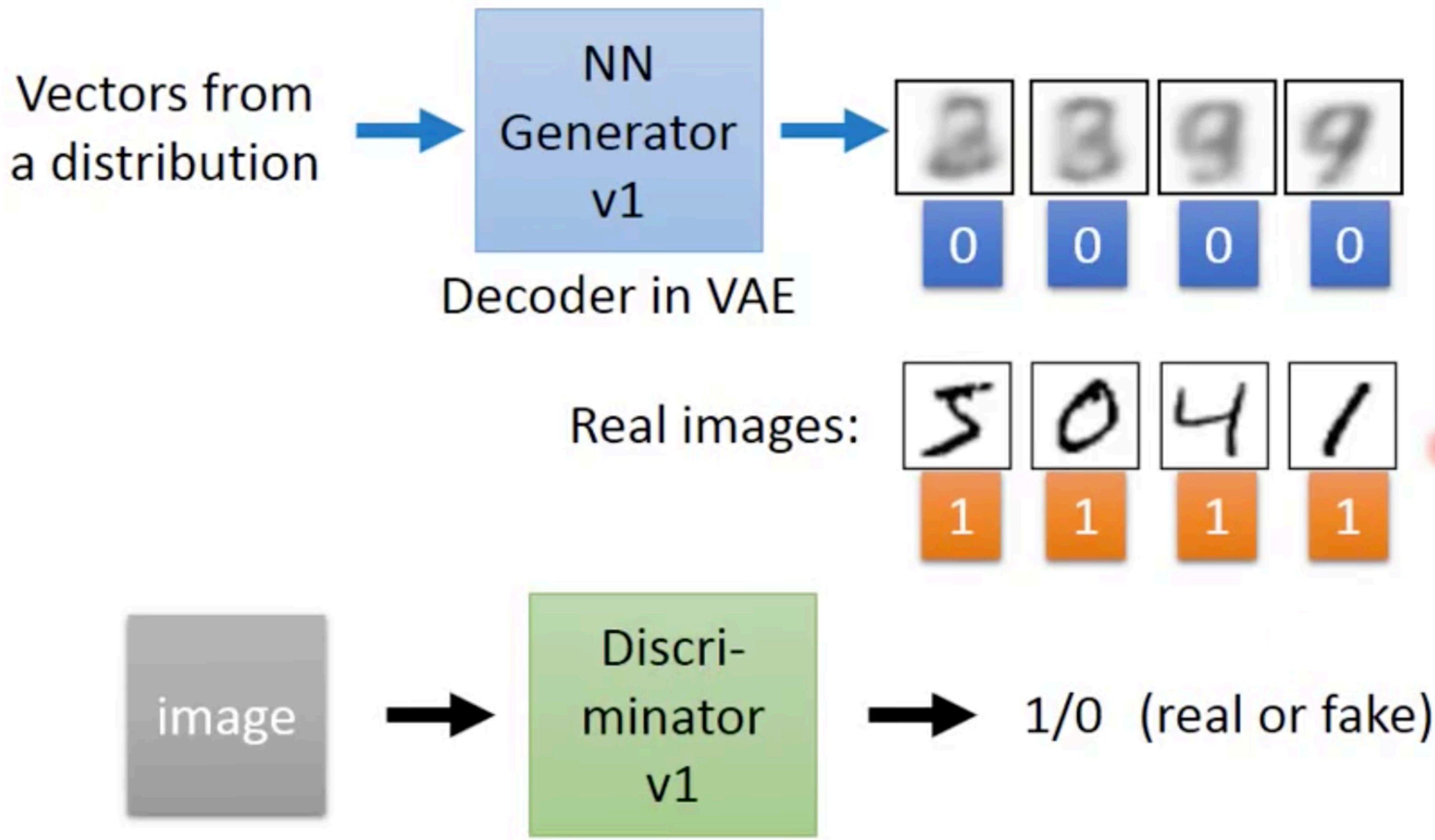




GAN - Discriminator

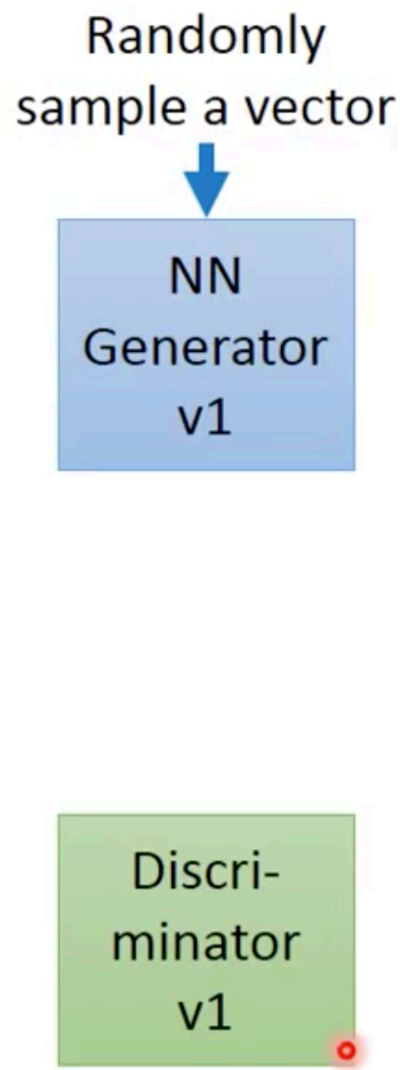


GAN - Discriminator

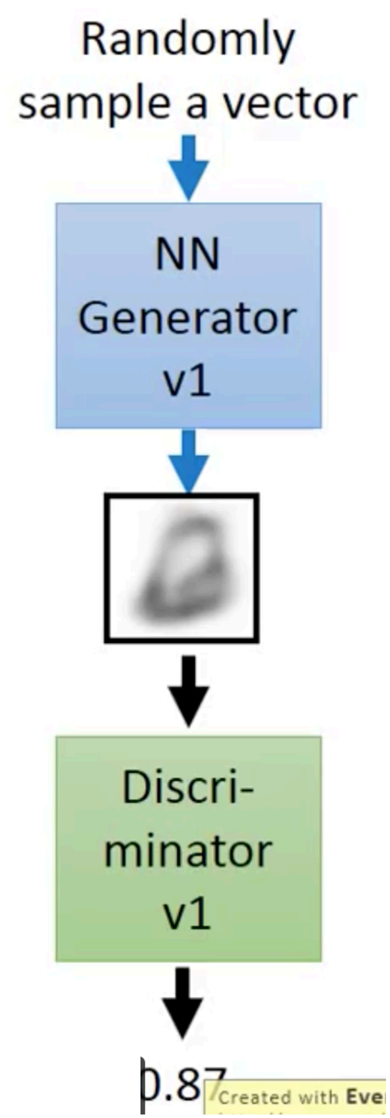


How to train the generator

GAN - Generator



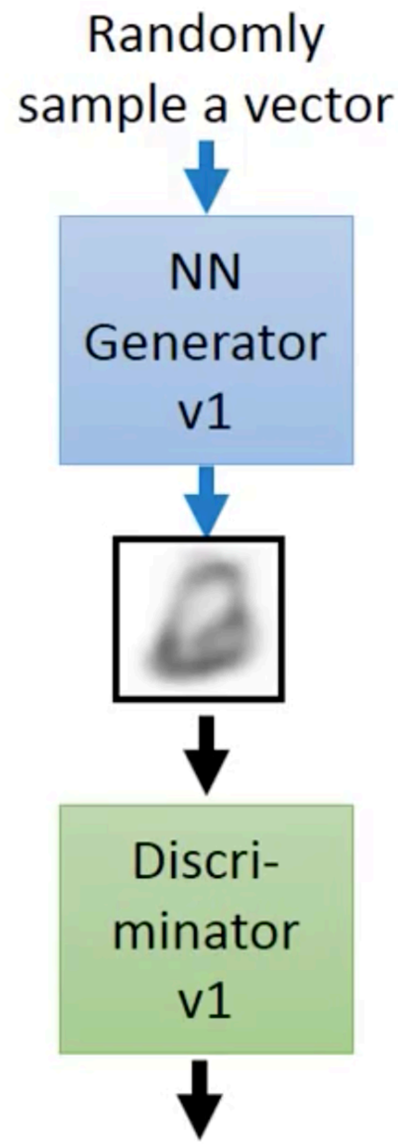
GAN - Generator



GAN - Generator

"Tuning" the parameters of generator

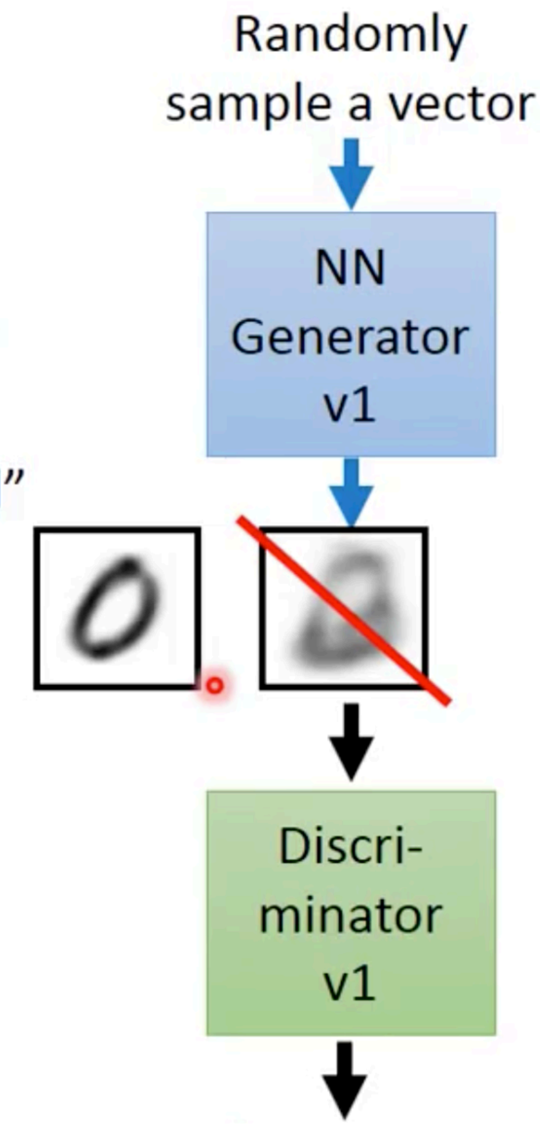
➡ The output be classified as "real" (as close to 1 as possible)



GAN - Generator

"Tuning" the parameters of generator

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

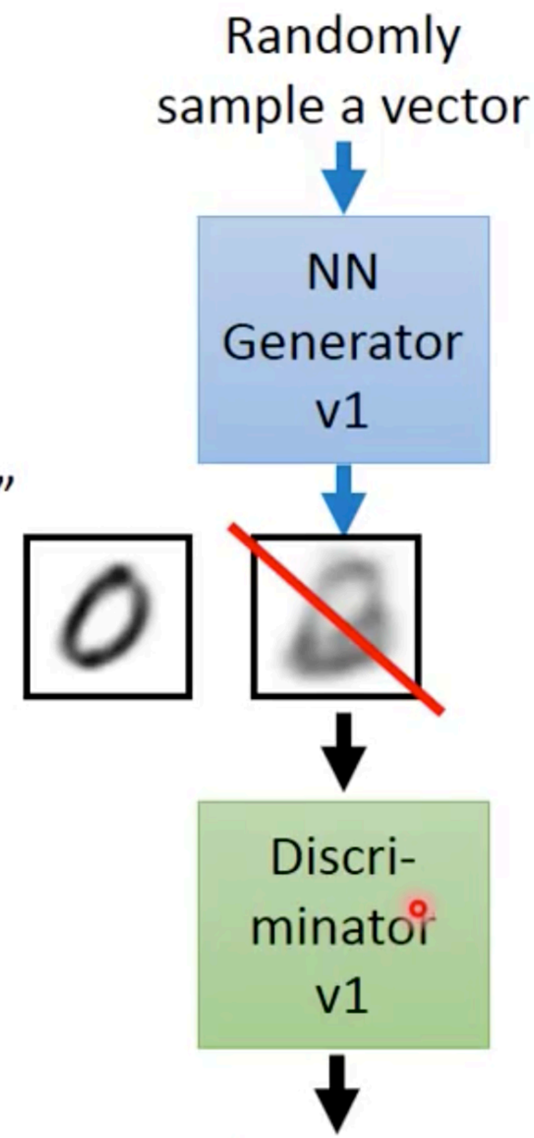


GAN - Generator

"Tuning" the parameters of generator

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

Generator + Discriminator
= a network



GAN - Generator

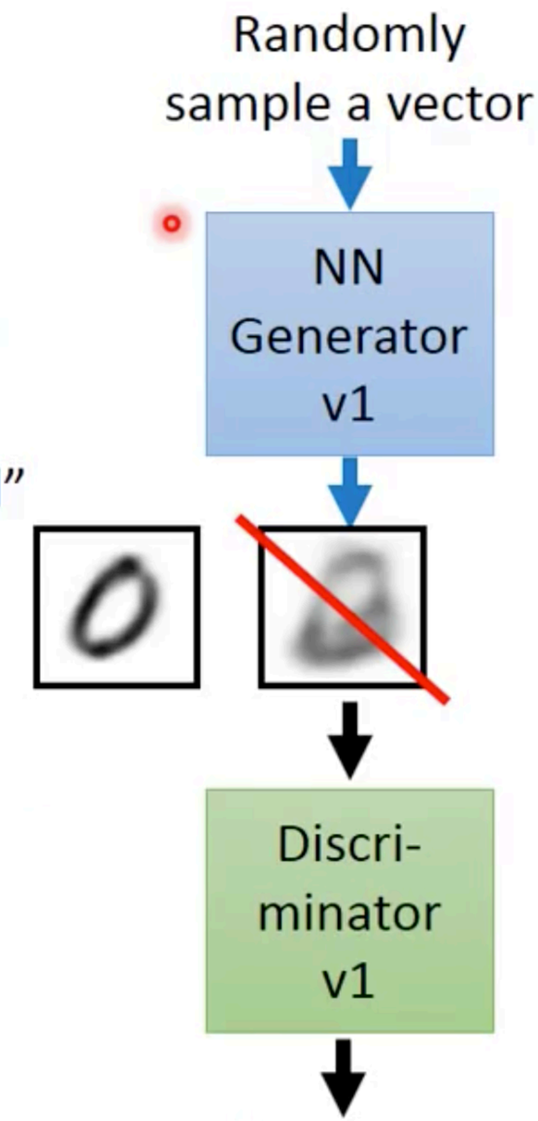
“Tuning” the parameters of generator

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

Generator + Discriminator
= a network

Using gradient descent to find the parameters of generator

Fix the discriminator



In practical ……

- GANs are difficult to optimize.

In practical ……

- GANs are difficult to optimize.
- No explicit signal about how good the generator is

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"

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

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