Conditional Generation by GAN

李宏毅

Hung-yi Lee

Text-to-Image

a dog is running

a bird is flying

- **Traditional supervised approach**

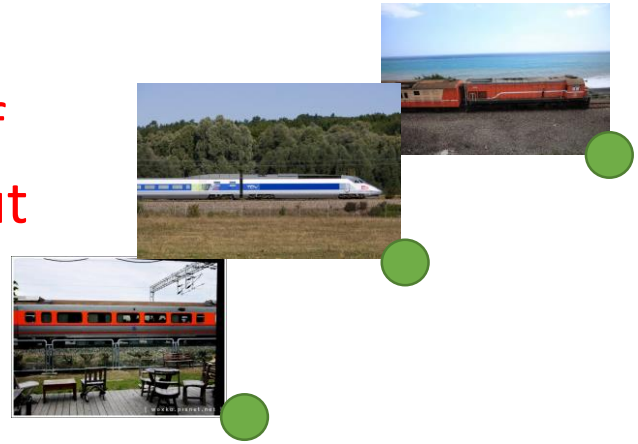
c^1: a dog is running → NN → Image ↔ 

as close as possible

Text: "train"

Target of NN output

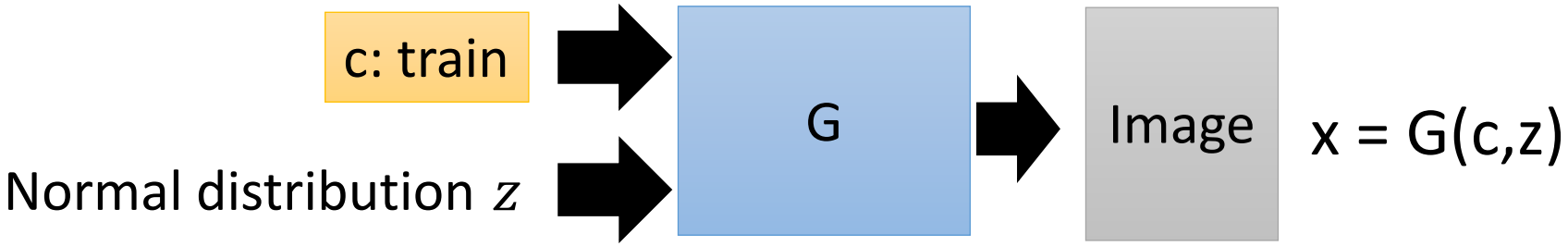
A blurry image!



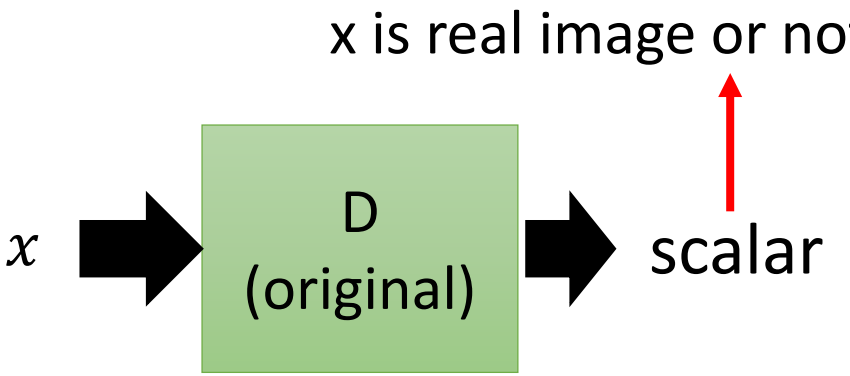
假設dataset有這六張火車照片，當最後train出來時model可能會學習產生這多張image的平均（模糊的結果）

[Scott Reed, et al, ICML, 2016]

Conditional GAN



對G而言，他只要輸出一個realistic image就可以讓D output高分，因此會忽略condition

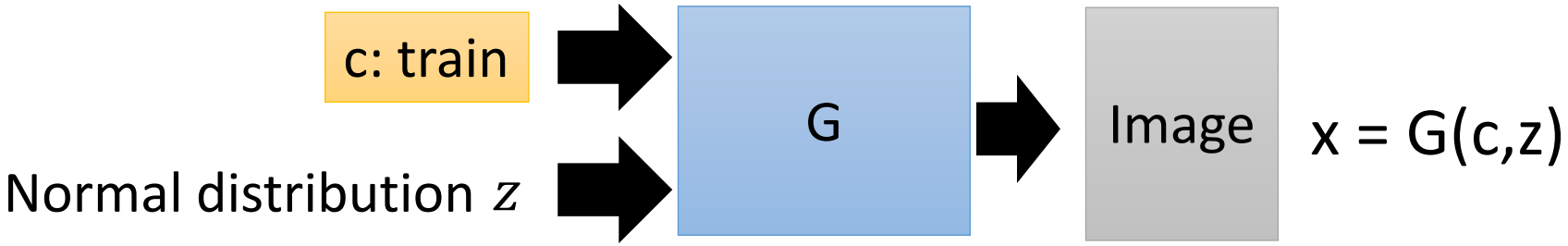


Generator will learn to generate realistic images ….

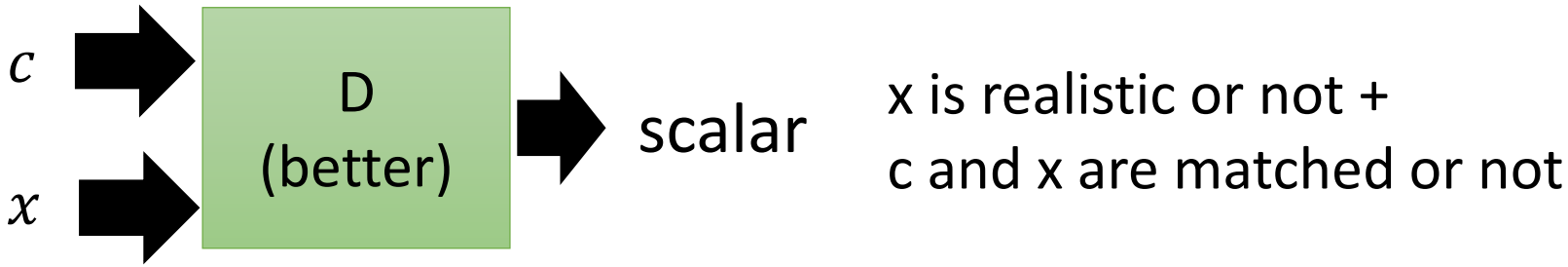
But completely ignore the input conditions.

[Scott Reed, et al, ICML, 2016]

Conditional GAN



因此D要同時考慮到input image x是不是真實的，以及c跟x是不是match



True text-image pairs: (train , ) 1

低分的case

(cat , ) 0

圖片真實但text錯誤

(train , Image) 0

圖片不真實但text正確

- In each training iteration:

Learn D

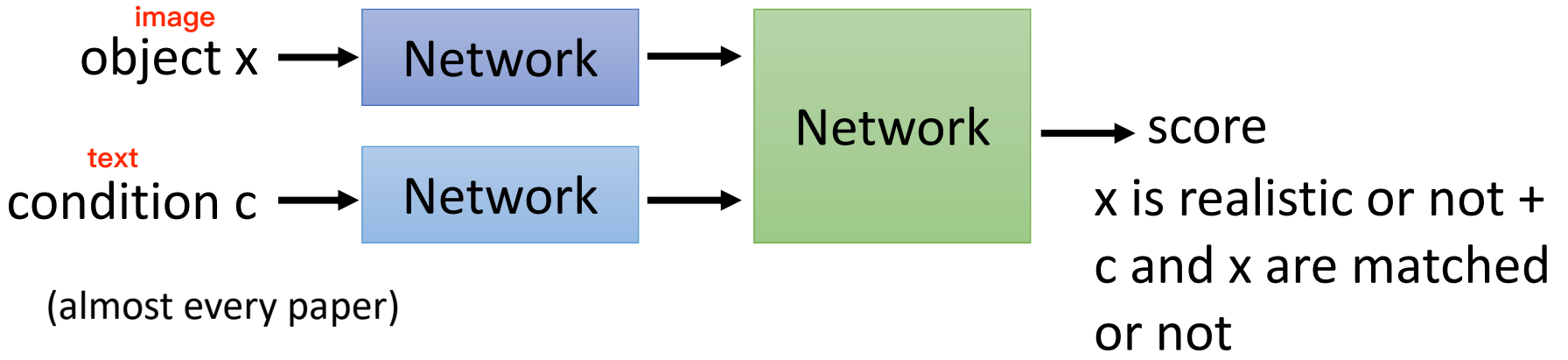
多考慮一組 negative example

- Sample m positive examples $\{(c^1, x^1), (c^2, x^2), \dots, (c^m, x^m)\}$ from database
- Sample m noise samples $\{z^1, z^2, \dots, z^m\}$ from a distribution
- Obtaining generated data $\{\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^m\}$, $\tilde{x}^i = G(c^i, z^i)$
- Sample m objects $\{\hat{x}^1, \hat{x}^2, \dots, \hat{x}^m\}$ from database
- Update discriminator parameters θ_d to maximize
  - $\tilde{V} = \frac{1}{m}\sum_{i=1}^{m} log D(c^i, x^i)$ 配上正確文字但模糊結果
    $+\frac{1}{m}\sum_{i=1}^{m} log\left(1 - D(c^i, \tilde{x}^i)\right) + \frac{1}{m}\sum_{i=1}^{m} log\left(1 - D(c^i, \hat{x}^i)\right)$ 正確文字但跟圖片無法match
  - $\theta_d \leftarrow \theta_d + \eta \nabla \tilde{V}(\theta_d)$

Learn G

- Sample m noise samples $\{z^1, z^2, \dots, z^m\}$ from a distribution
- Sample m conditions $\{c^1, c^2, \dots, c^m\}$ from a database
- Update generator parameters θ_g to maximize
  - $\tilde{V} = \frac{1}{m}\sum_{i=1}^{m} log\left(D\left(G(c^i, z^i)\right)\right)$, $\theta_g \leftarrow \theta_g - \eta \nabla \tilde{V}(\theta_g)$

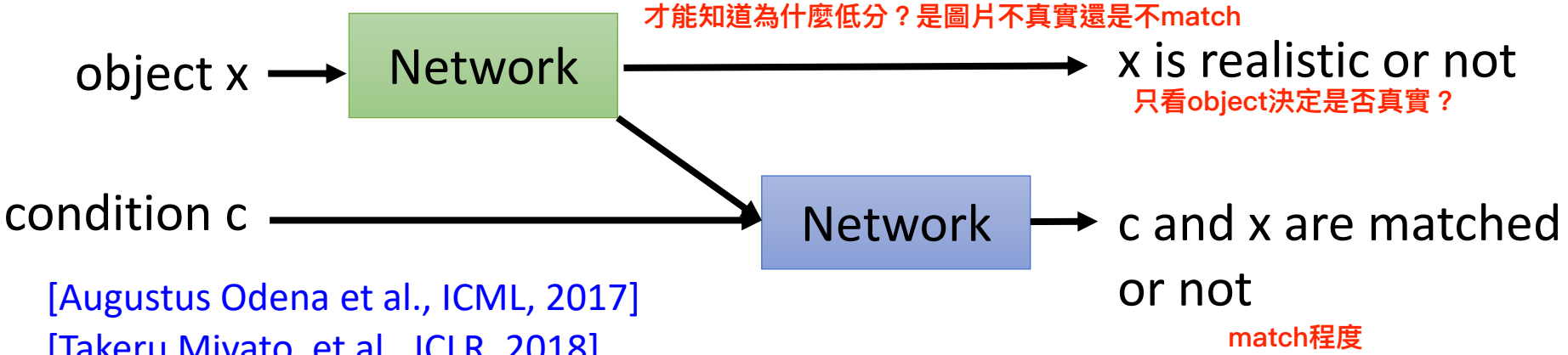
Conditional GAN - Discriminator



(almost every paper)

下面這個架構比較好！！

把兩種evaluation的結果拆開來看比較合理，這樣
才能知道為什麼低分？是圖片不真實還是不match



[Augustus Odena et al., ICML, 2017]
[Takeru Miyato, et al., ICLR, 2018]
[Han Zhang, et al., arXiv, 2017]

Conditional GAN

The images are generated by Yen-Hao Chen, Po-Chun Chien, Jun-Chen Xie, Tsung-Han Wu.

paired data

blue eyes
red hair
short hair

Collecting anime faces and the description of its characteristics

red hair, green eyes

blue hair, red eyes

抓到了！一隻眼睛沒畫紅色！

登峰造極專用CGAN

先產生小圖，再產生大圖

Stack GAN

jointly train的

Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, Dimitris Metaxas, “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks”, ICCV, 2017

先train第一個小的，在疊第二個一起train，在疊第三個在一起train

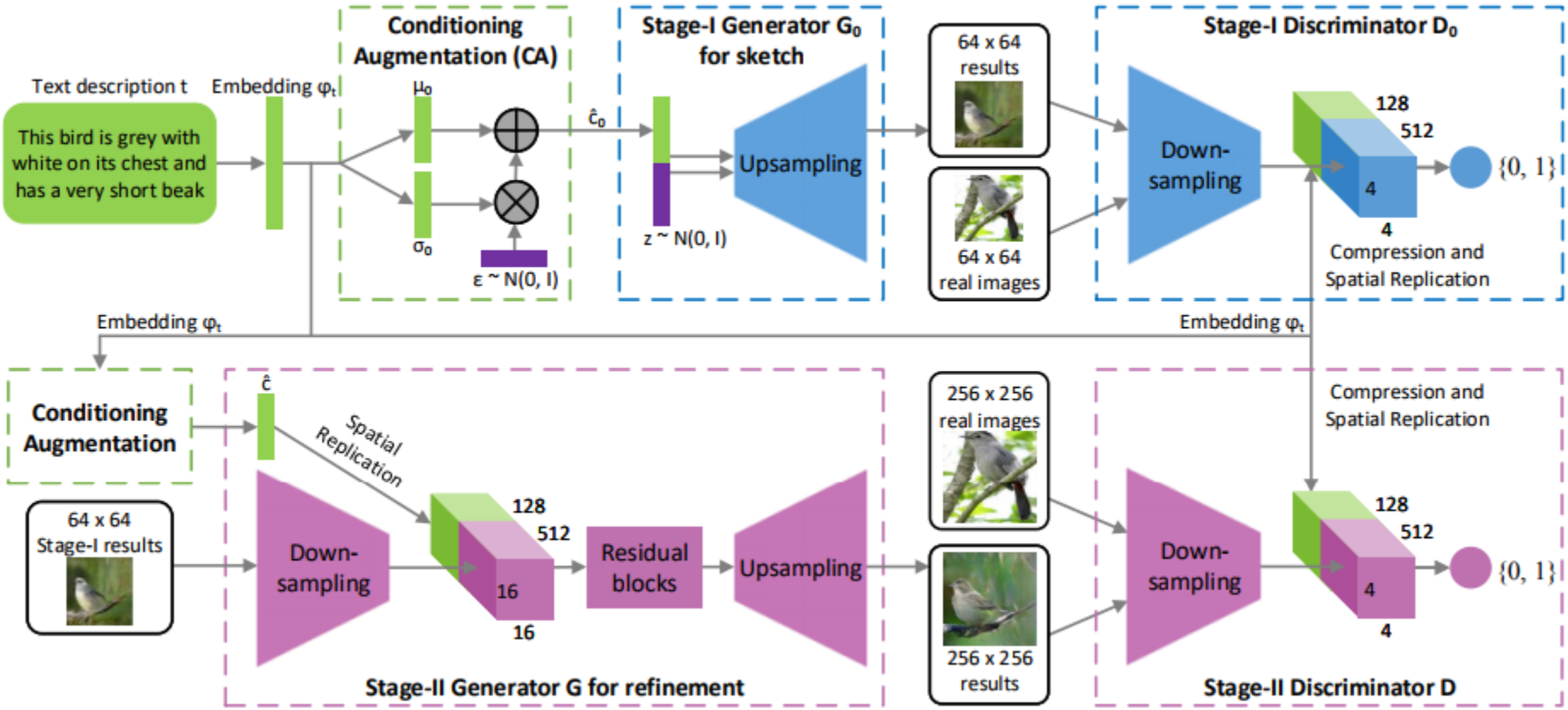
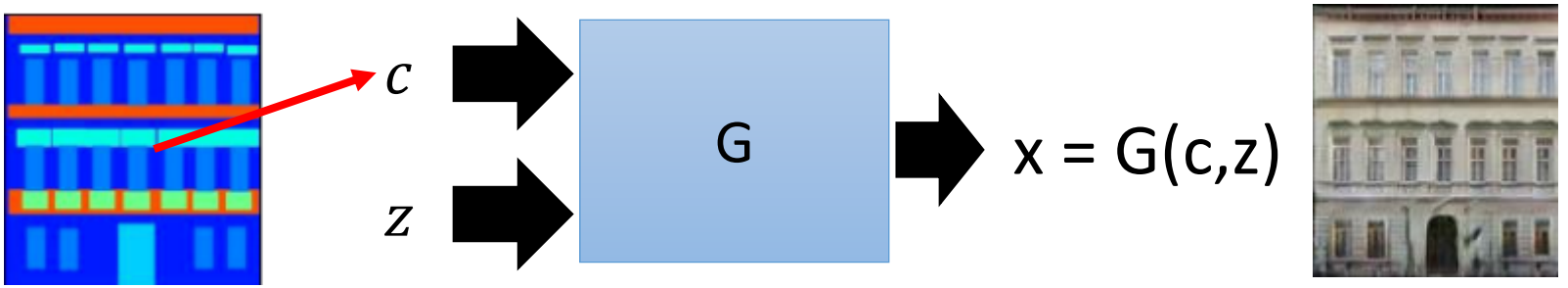
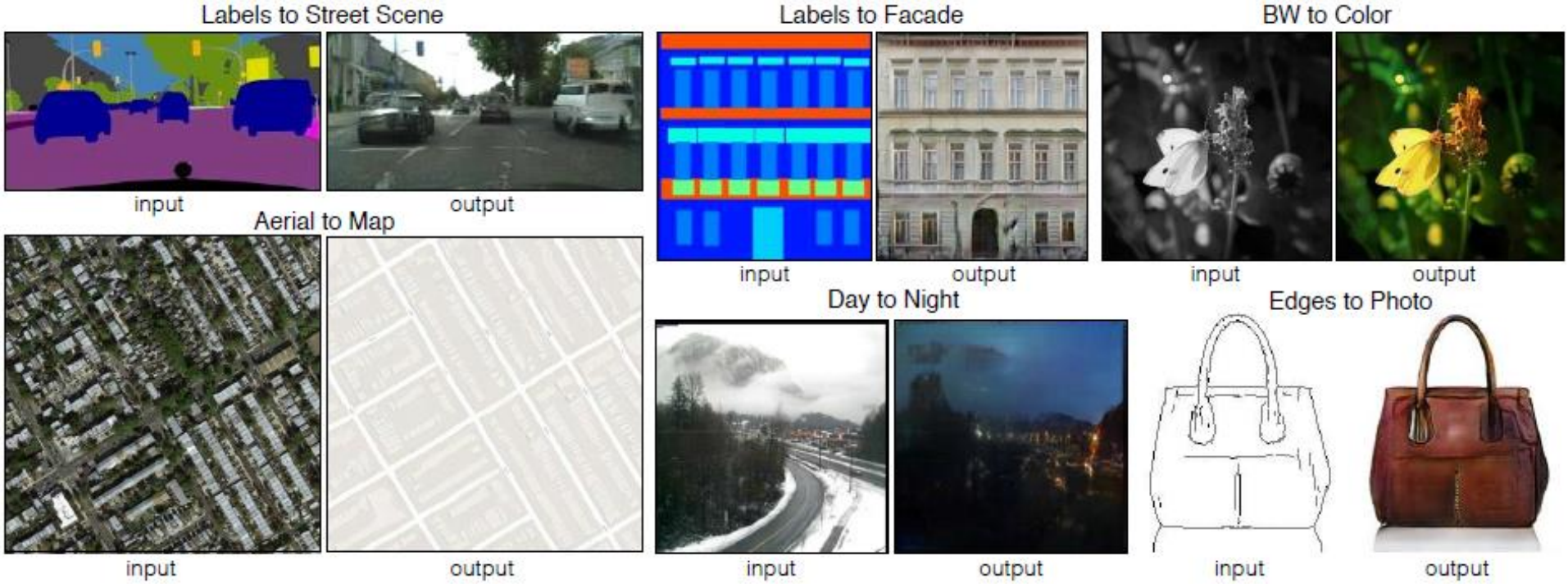


Image-to-image





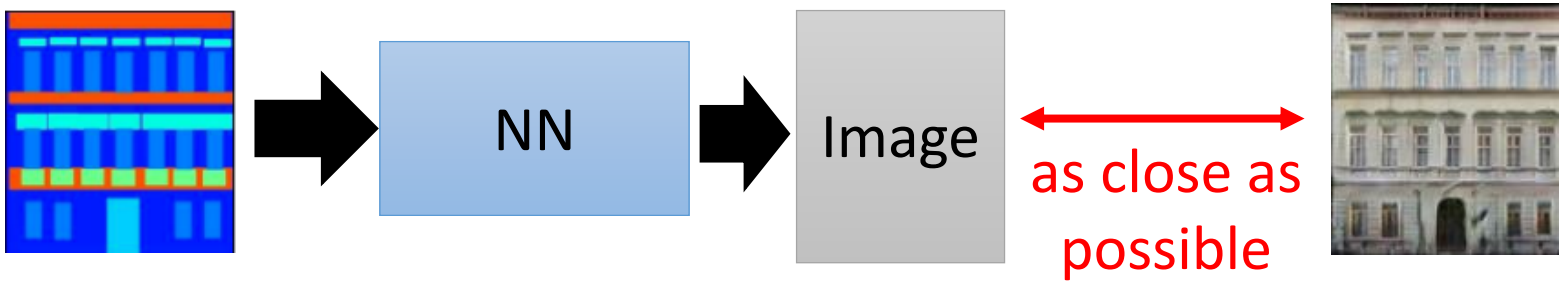
https://arxiv.org/pdf/1611.07004

Image-to-image

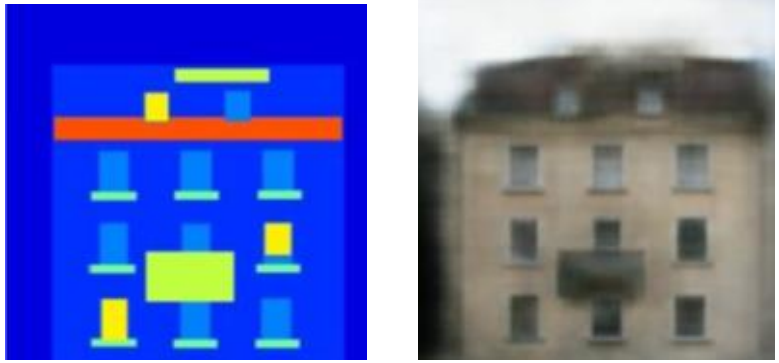
paired data

- Traditional supervised approach

supervised方法是模糊的



Testing:



input close

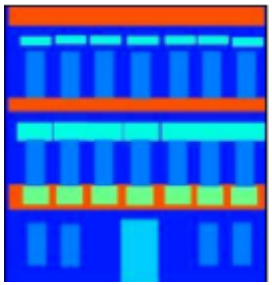
It is blurry because it is the average of several images.

同一張image可以對應到許多不同的圖片，所以會產生平均的結果（模糊）

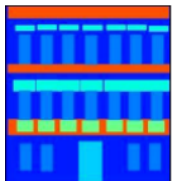
Image-to-image

額外的constrain

希望跟目標的圖片越接近越好



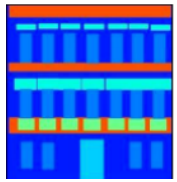
- Experimental results



z

G

Image



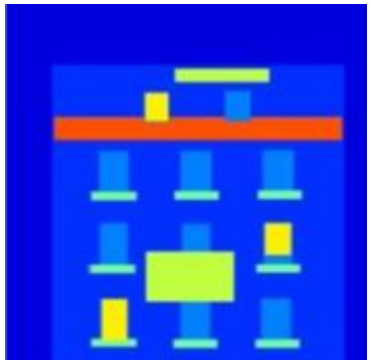
Image

要多考慮match程度！而非
單純只考慮realistic程度

D

scalar

Testing:



input

close

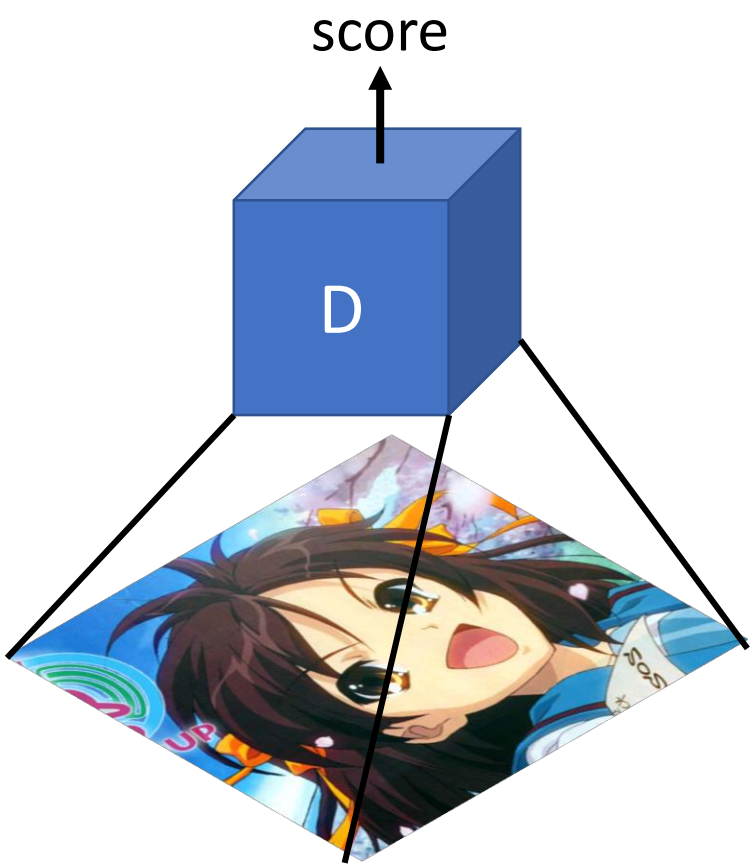
GAN

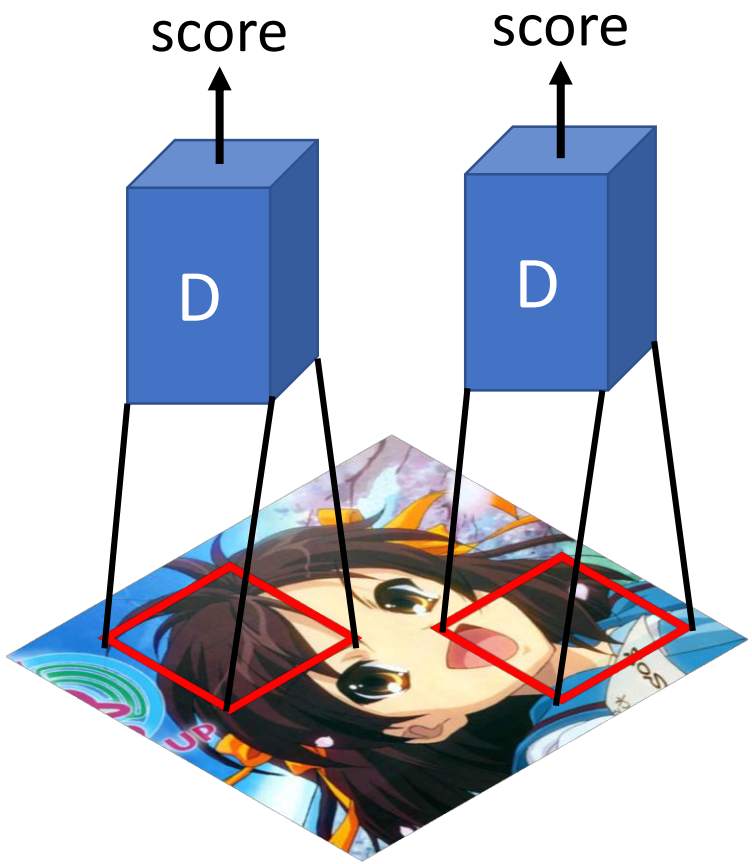
GAN + close

Patch GAN

當今天圖片很大張的時候，要是D一次檢查一大張
圖片很容易造成overfitting或是根本train不起來

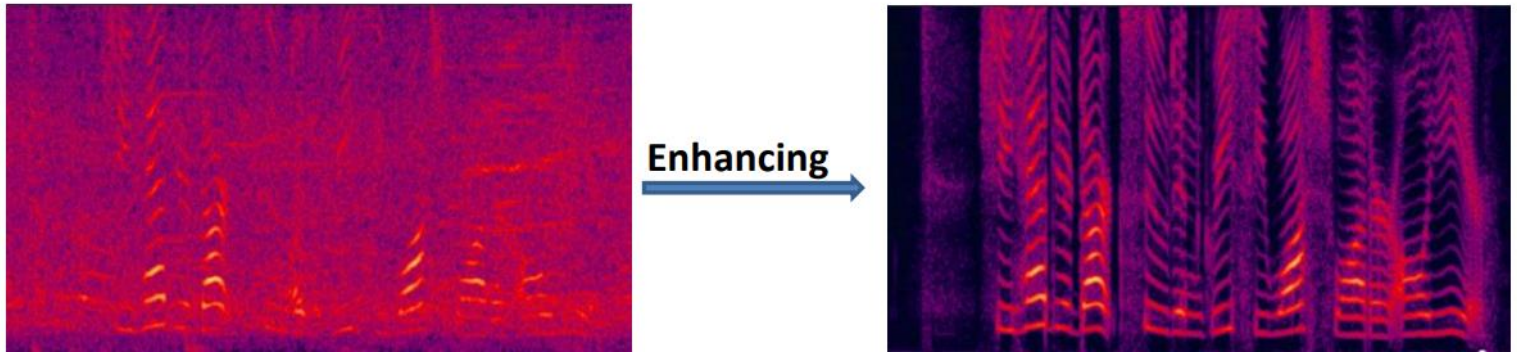
https://arxiv.org/pdf/1611.07004.pdf



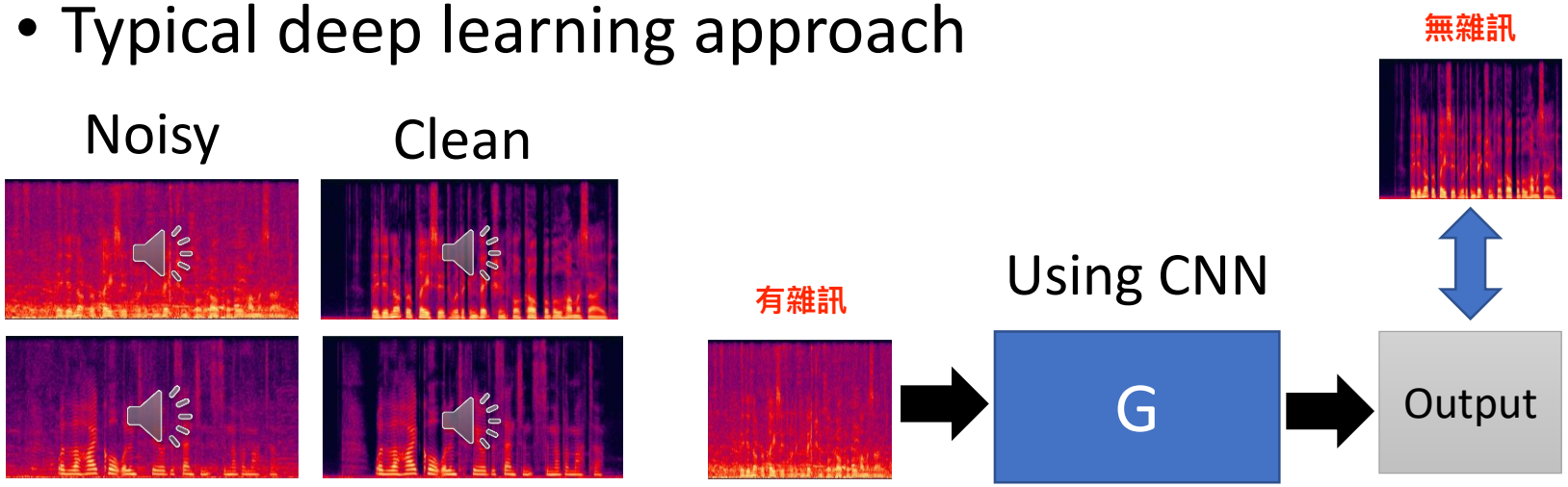


因此改成D一次檢查一個部分的區域即可

Speech Enhancement



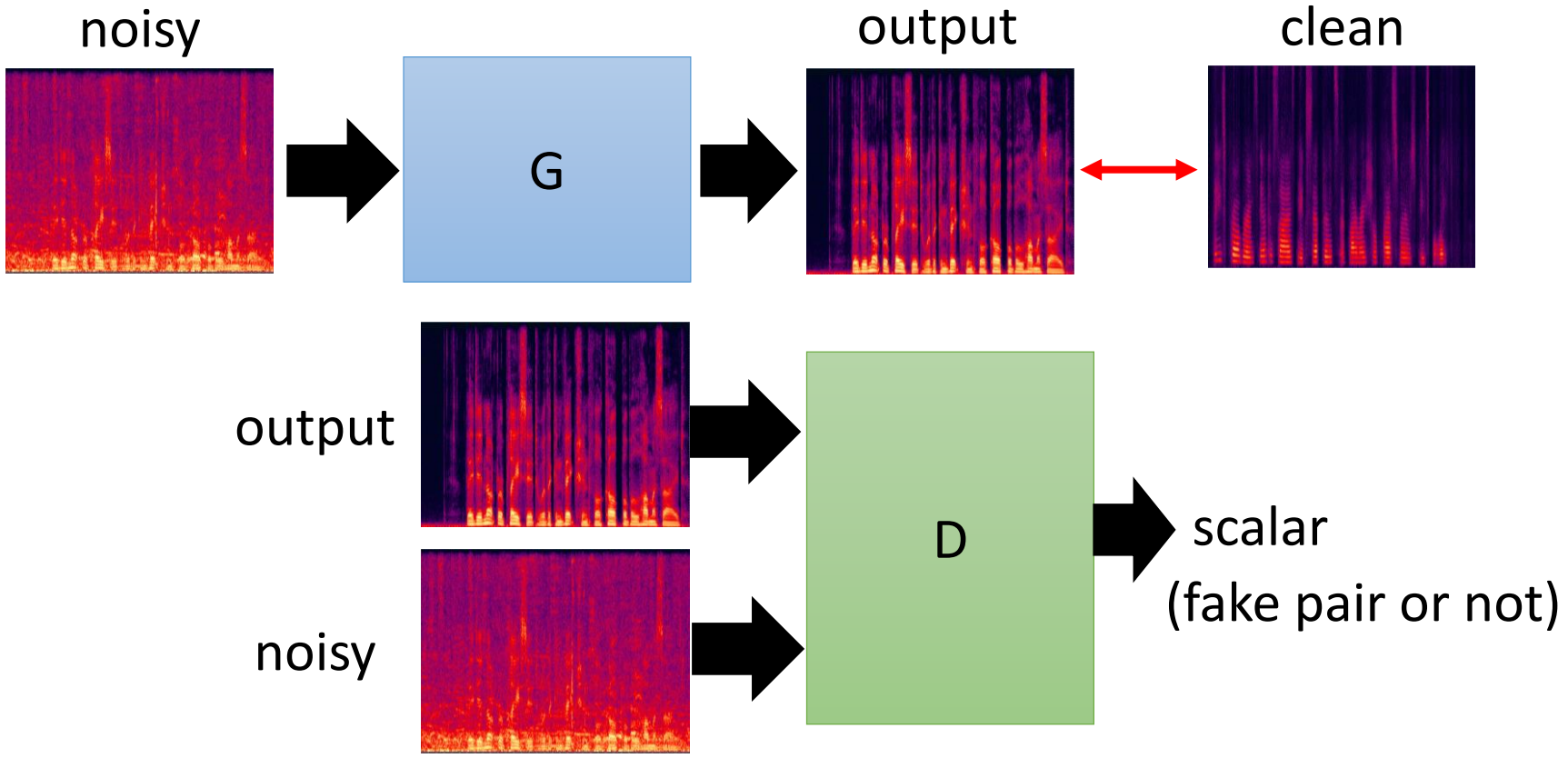
- Typical deep learning approach



強調很多次了conditionalGAN的D要多考慮match關係！

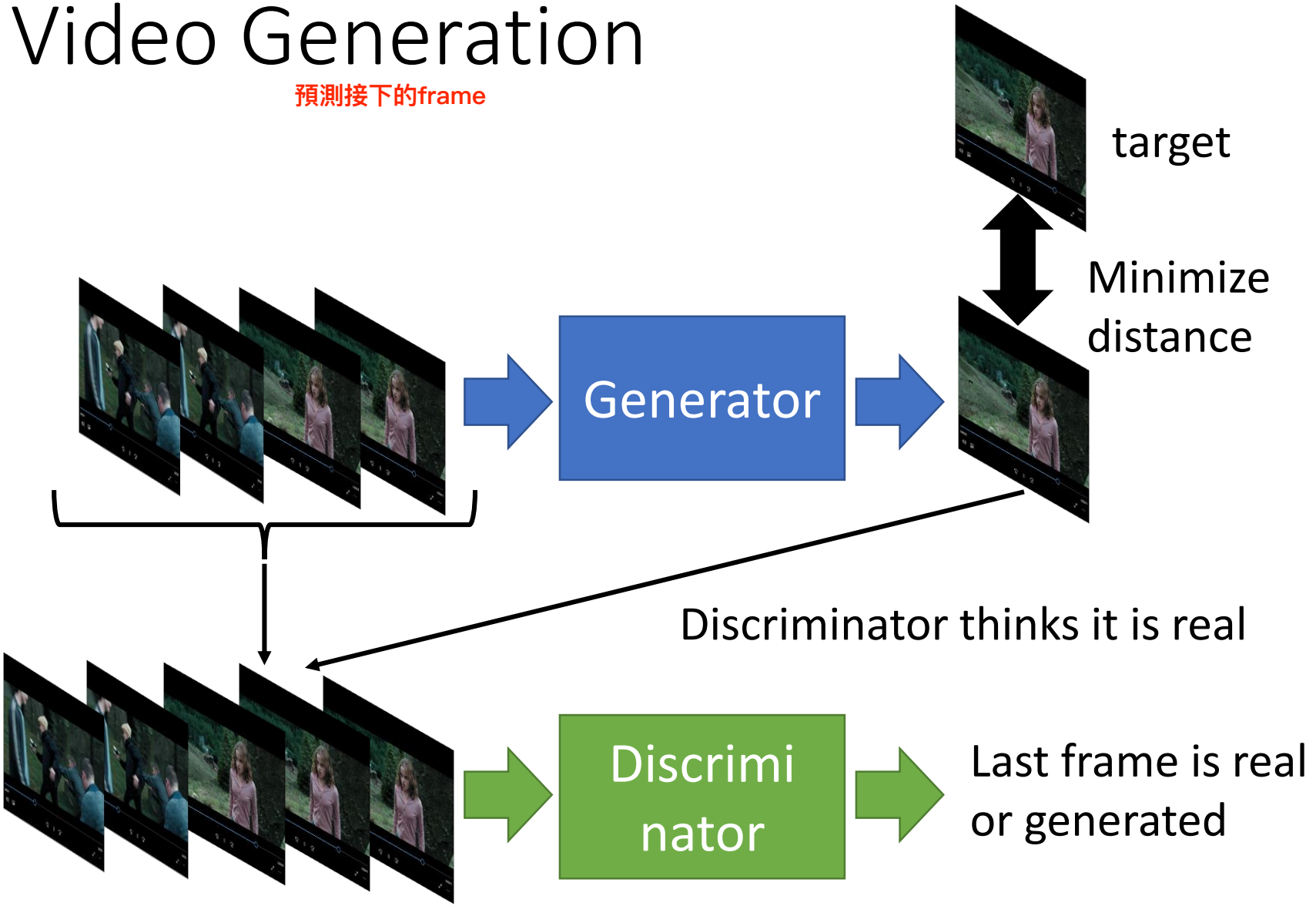
Speech Enhancement

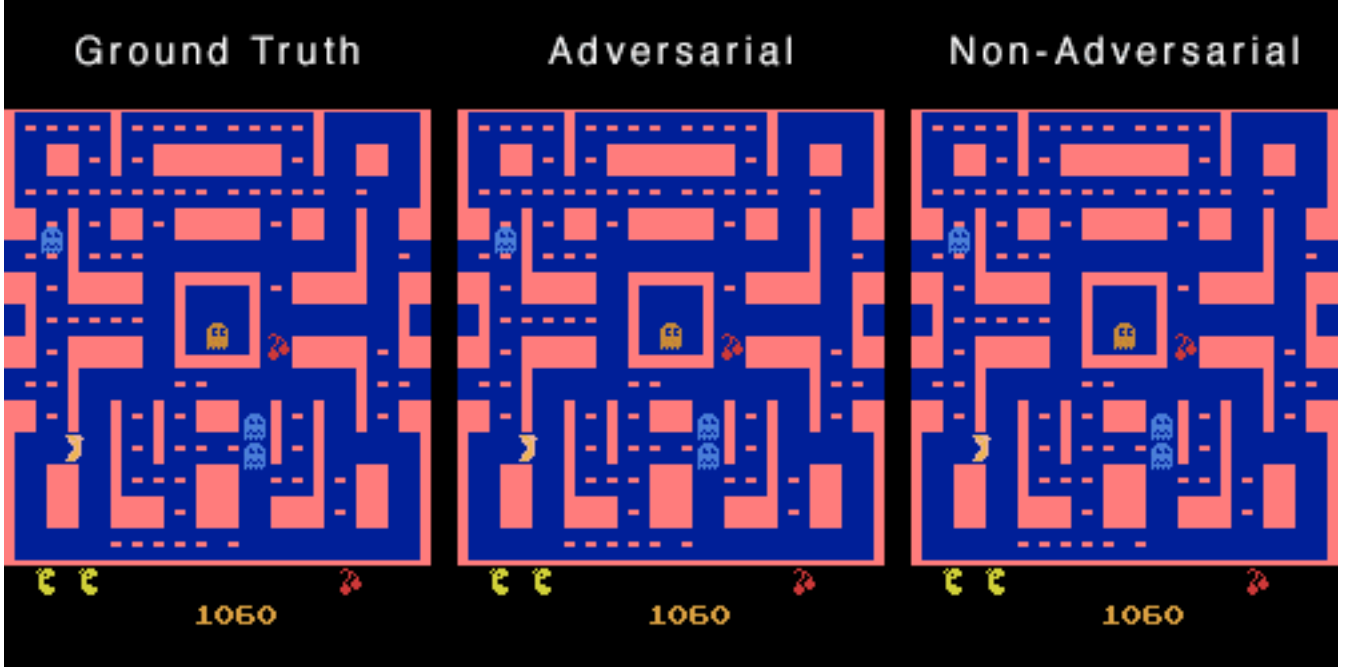
- Conditional GAN

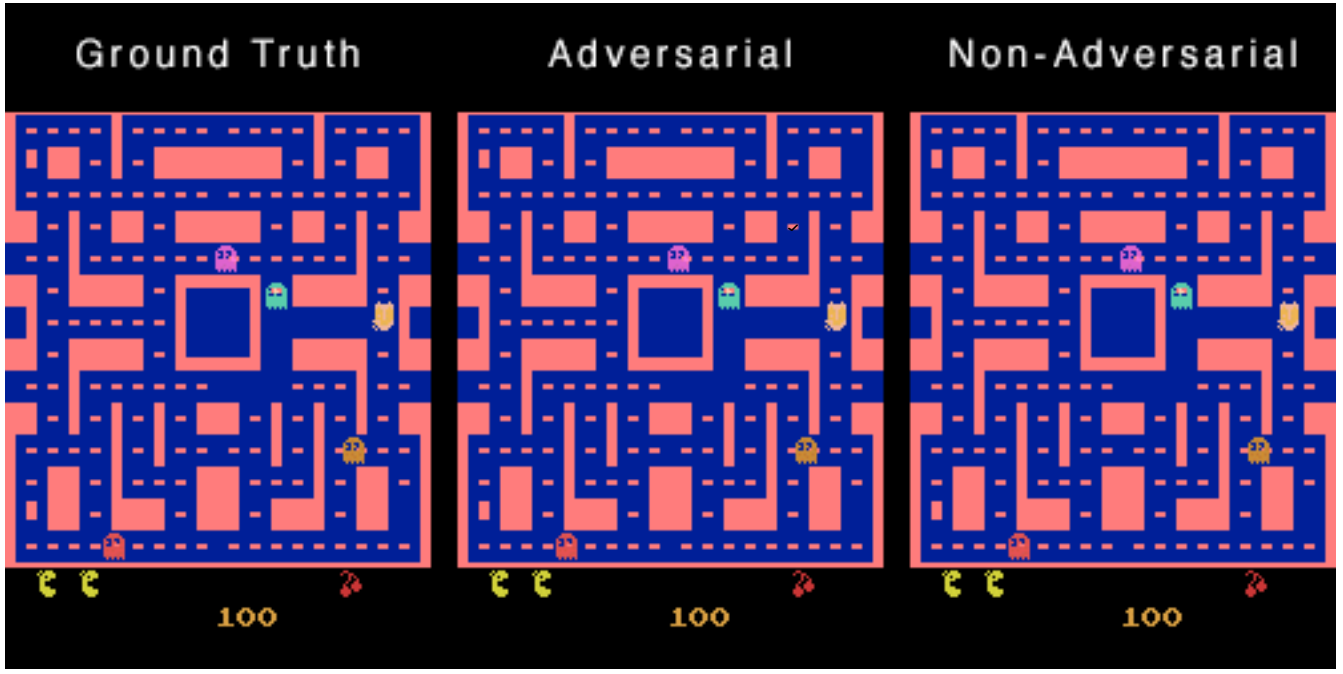


Video Generation

預測接下的frame







右下角沒有用GAN，造成如果碰到轉角小精靈會分裂，因為有時候是往左走有時候是往右走，他取平均就分開了哈哈哈

https://github.com/dyelax/Adversarial_Video_Generation