Conditional Generation by GAN

李宏毅

Hung-yi Lee

Text-to-Image

a dog is running

a bird is flying

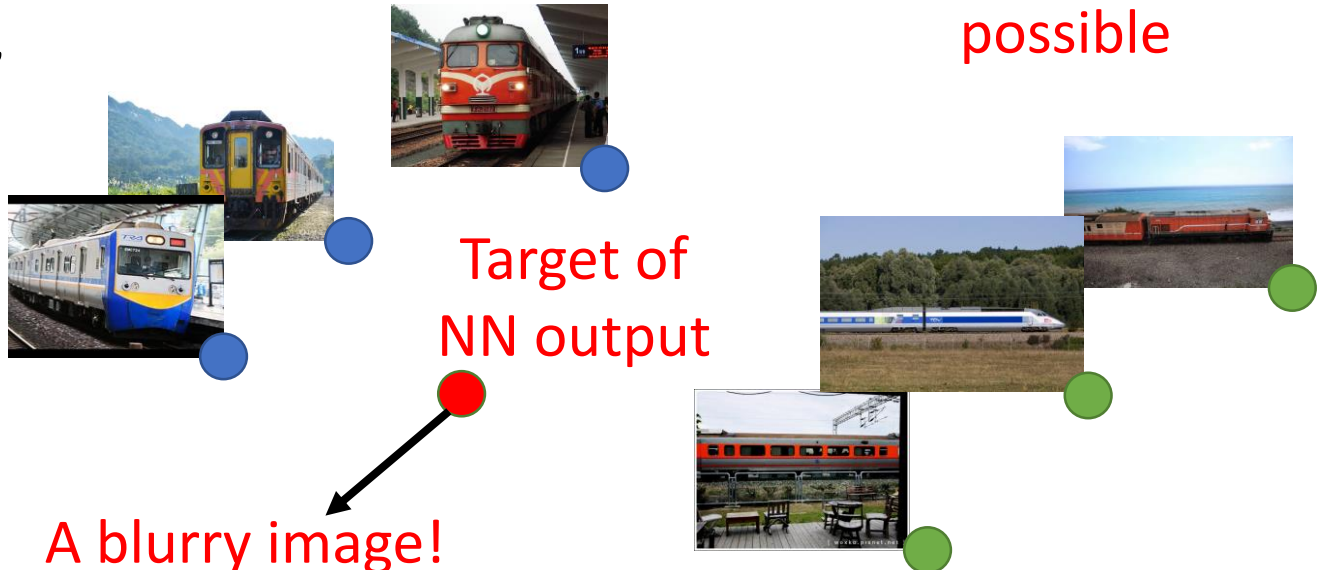- **Traditional supervised approach**

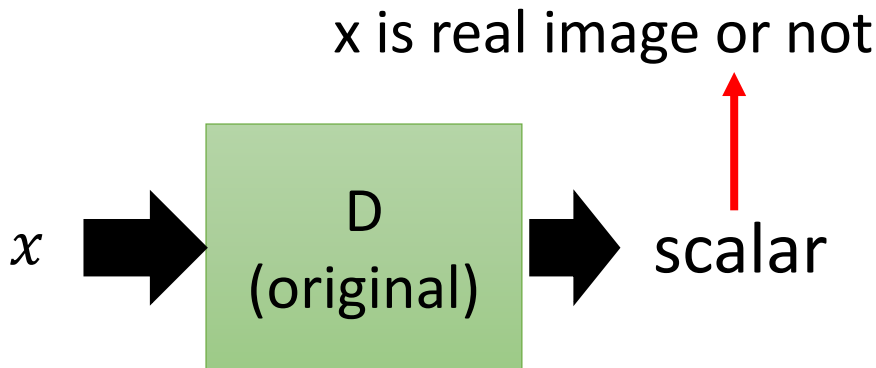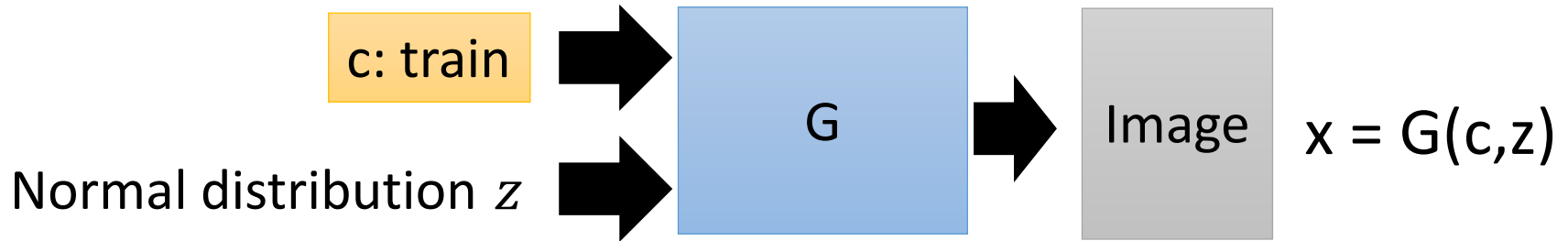c^1: a dog is running → NN → Image ↔ as close as possible
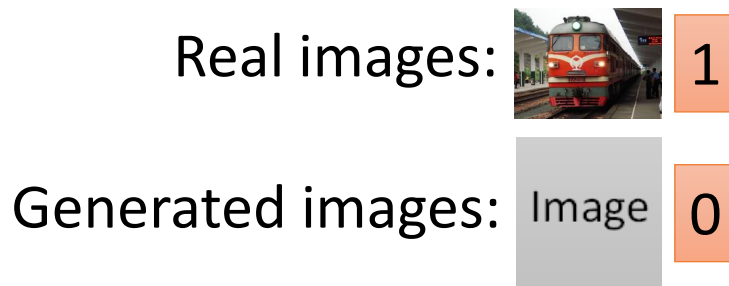
Text: "train"

Target of NN output
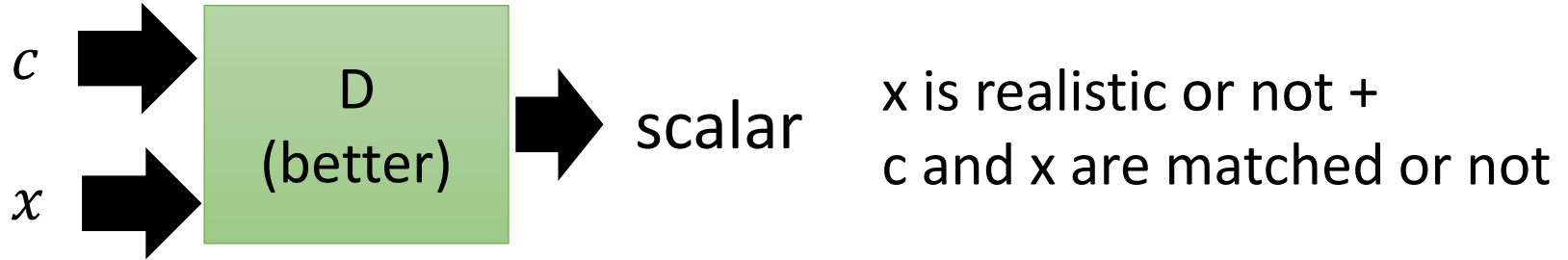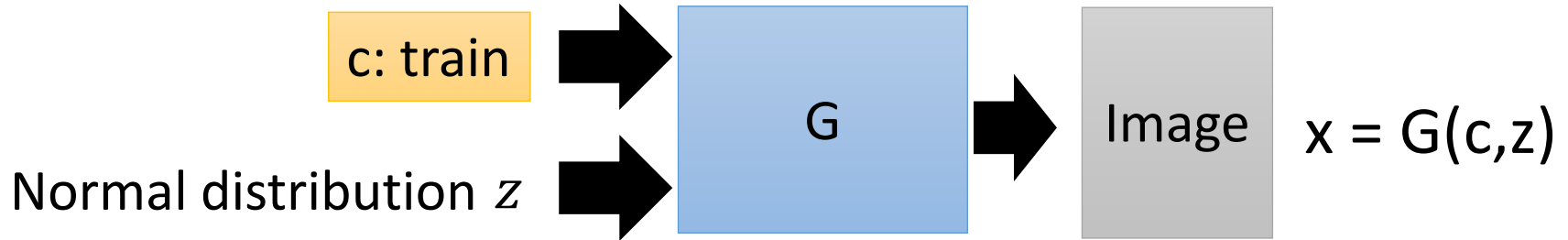
A blurry image!

Conditional GAN

c: train → G → Image   x = G(c,z)

Normal distribution z →

x is real image or not

x → D (original) → scalar

Generator will learn to generate realistic images ....

But completely ignore the input conditions.

Real images: 1

Generated images: Image 0

Conditional GAN

c: train $\rightarrow$ G $\rightarrow$ Image

Normal distribution z $\rightarrow$

x = G(c,z)

c $\rightarrow$ D (better) $\rightarrow$ scalar

x $\rightarrow$

x is realistic or not +
c and x are matched or not

True text-image pairs: (train , <image>) 1

(cat , <image>) 0   (train , Image) 0

Conditional GAN - Discriminator

object x → Network → Network → score

condition c → Network →

(almost every paper)

x is realistic or not +
c and x are matched
or not

object x → Network → x is realistic or not

condition c → Network → c and x are matched
or not
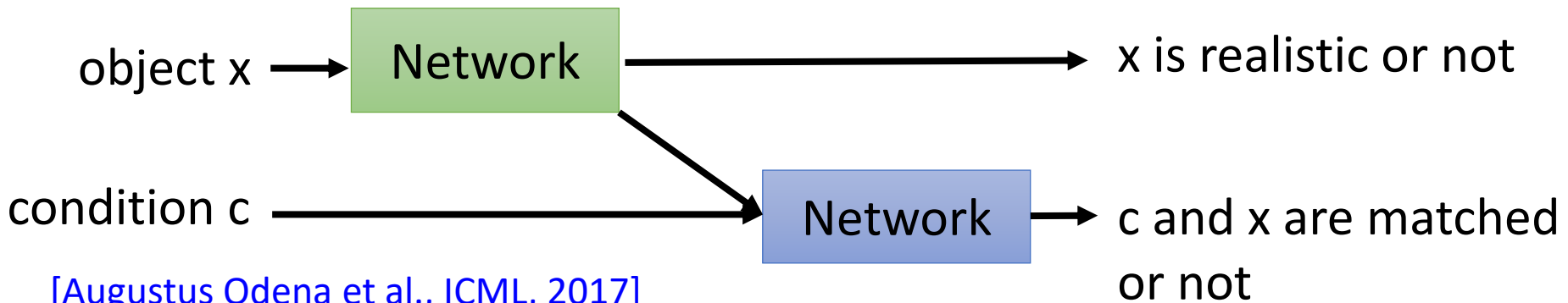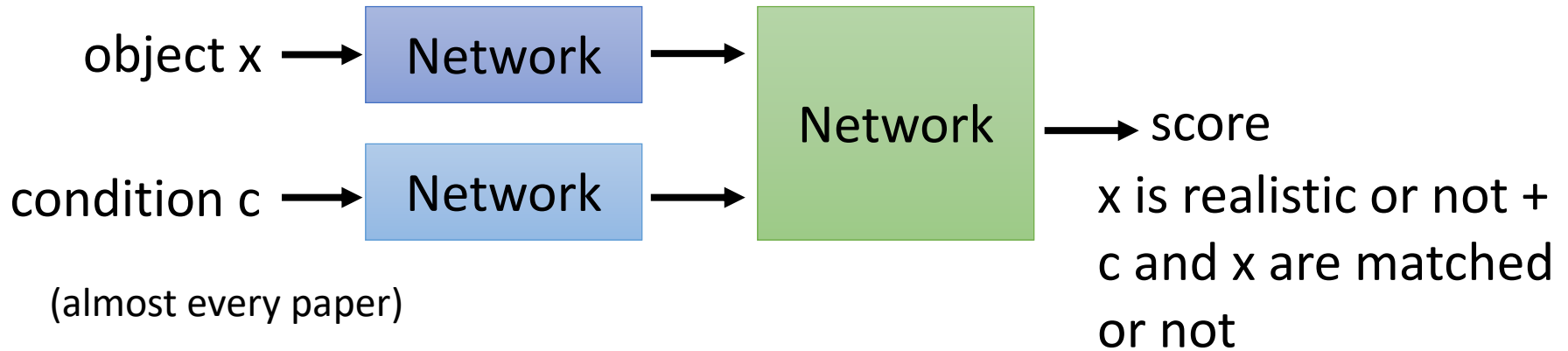
[Augustus Odena et al., ICML, 2017]
[Takeru Miyato, et al., ICLR, 2018]
[Han Zhang, et al., arXiv, 2017]

Conditional GAN

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

Stack GAN

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
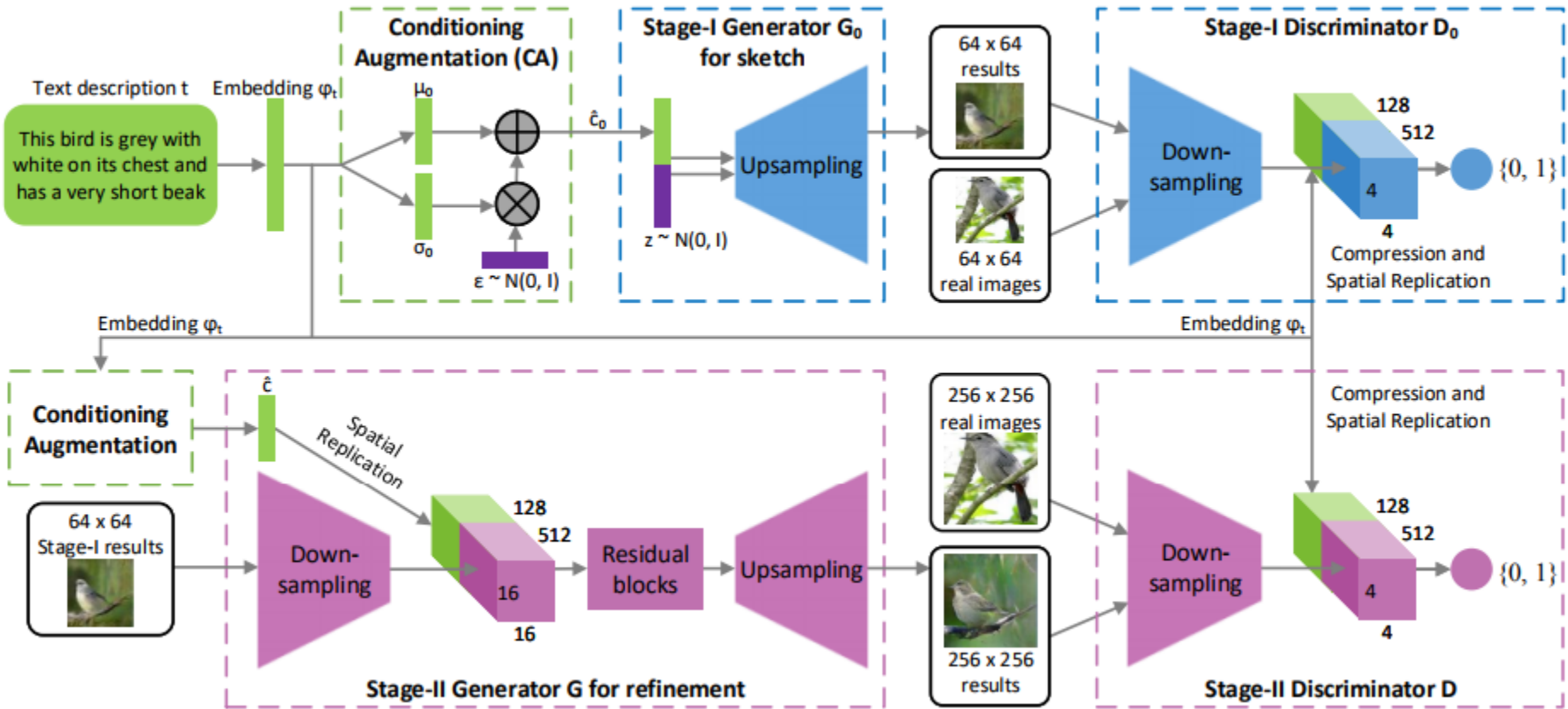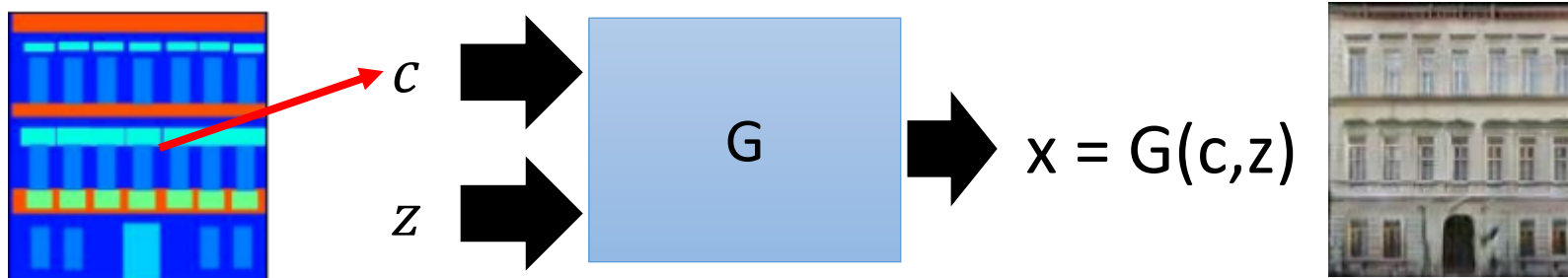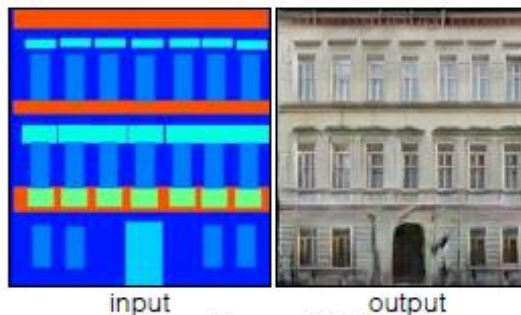
Image-to-image

c → G → $x = G(c,z)$

z →

Labels to Street Scene

input — output

Aerial to Map

input — output

Labels to Facade

input — output

Day to Night

input — output

BW to Color

input — output

Edges to Photo

input — output

https://arxiv.org/pdf/1611.07004
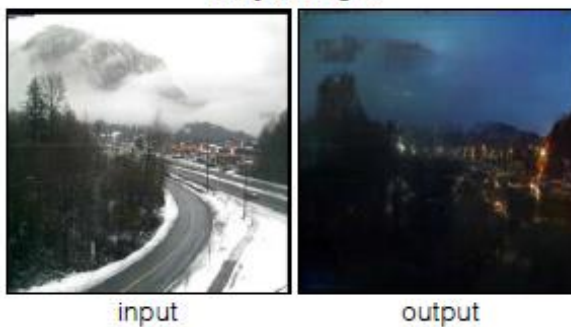
Image-to-image

- Traditional supervised approach

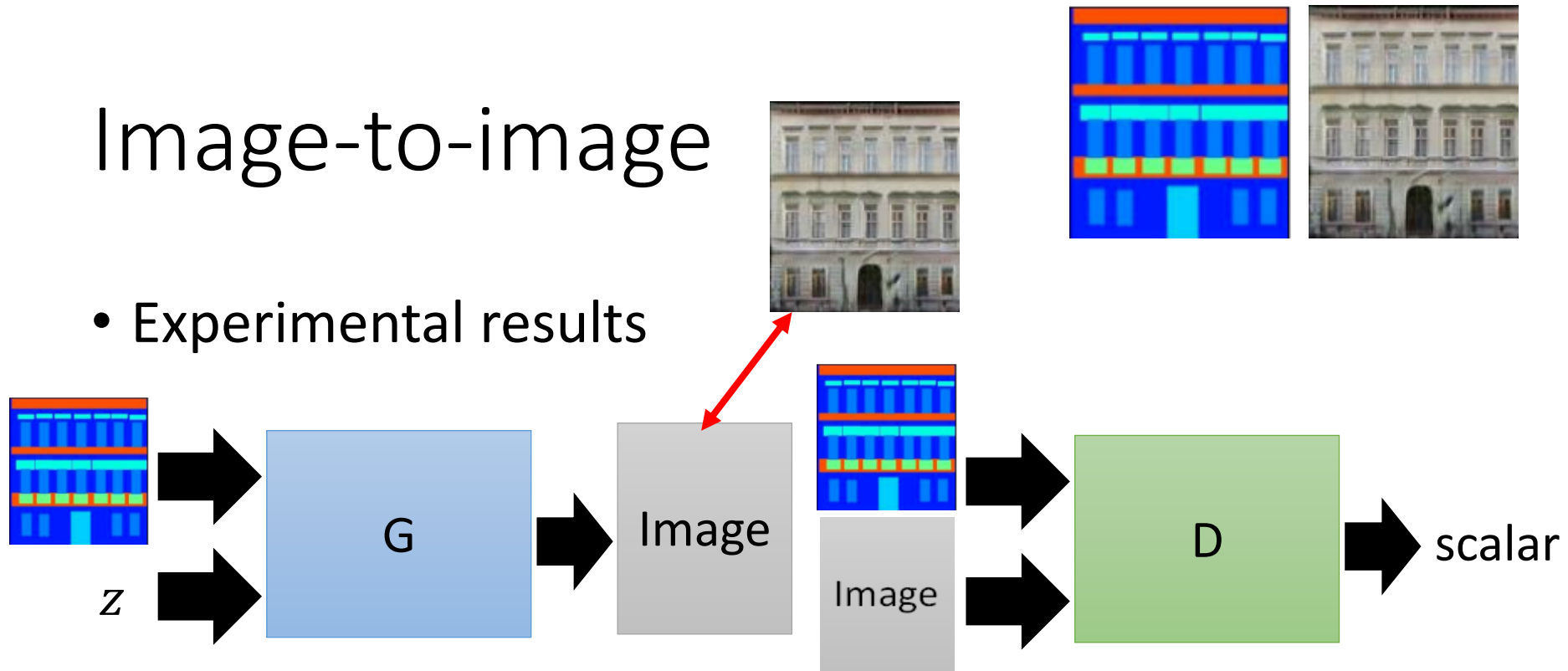NN → Image ←→ as close as possible
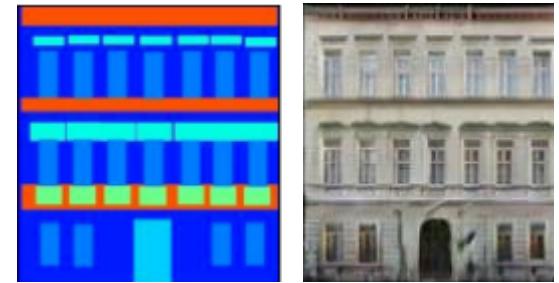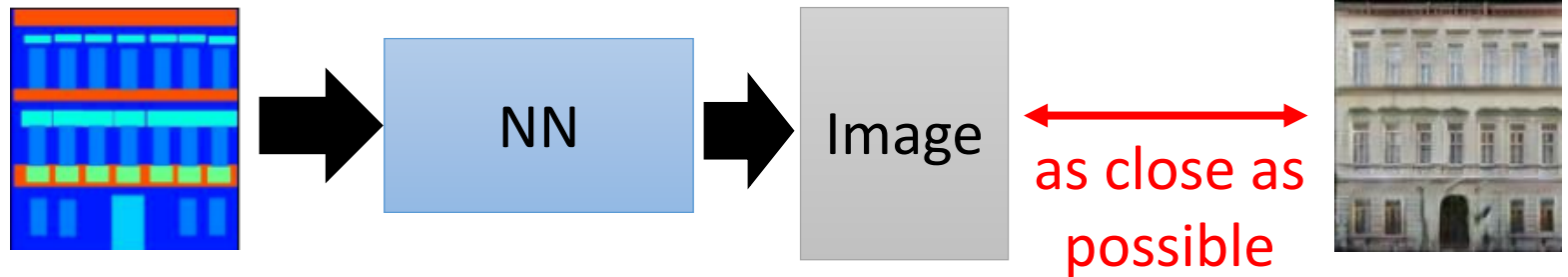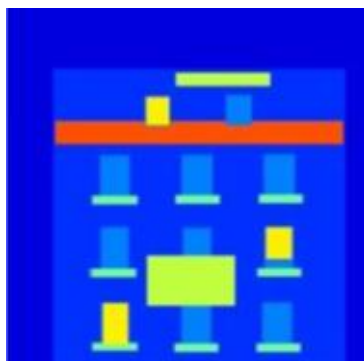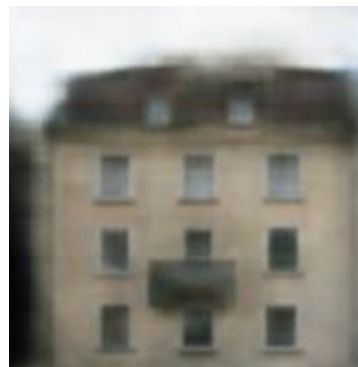
Testing:
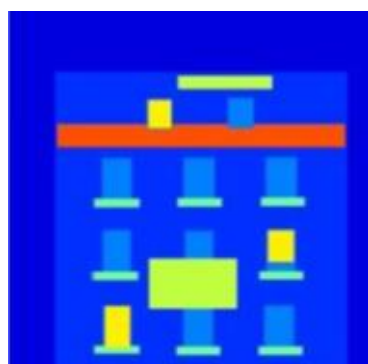

input   close

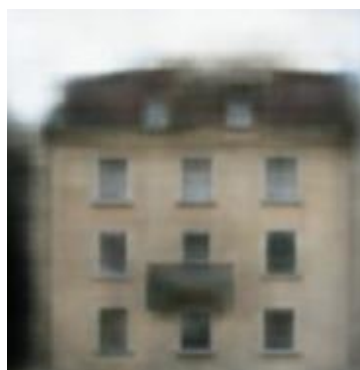It is blurry because it is the average of several images.

Image-to-image

- Experimental results

Testing:

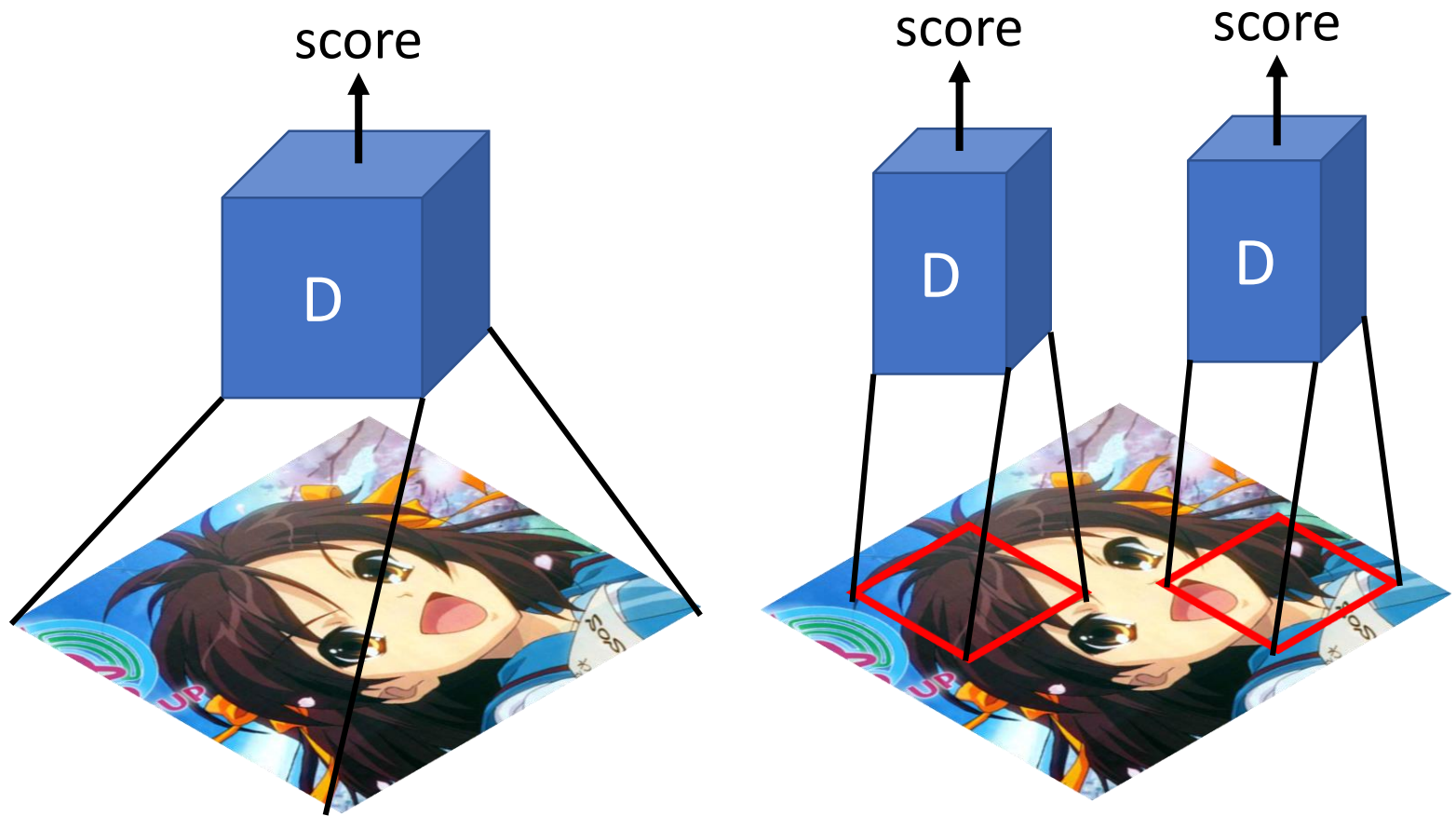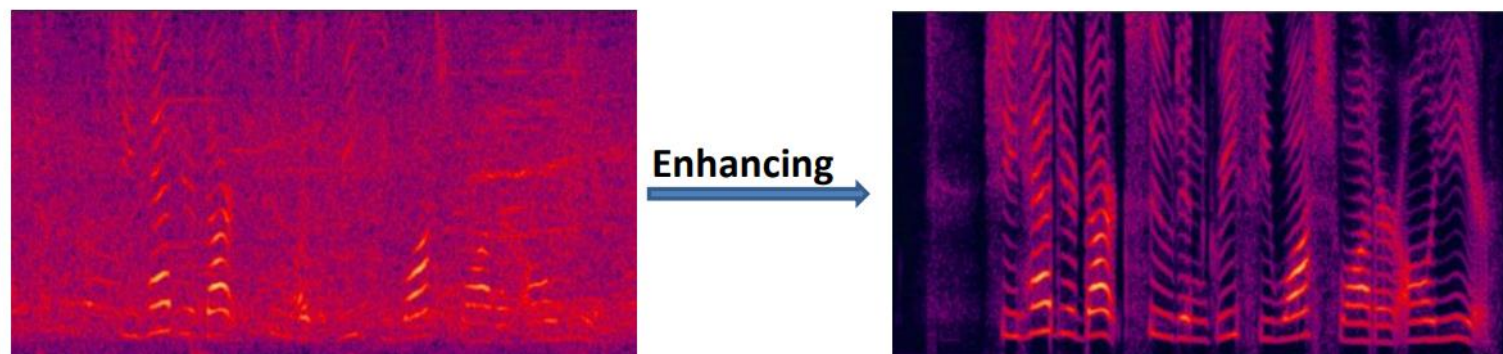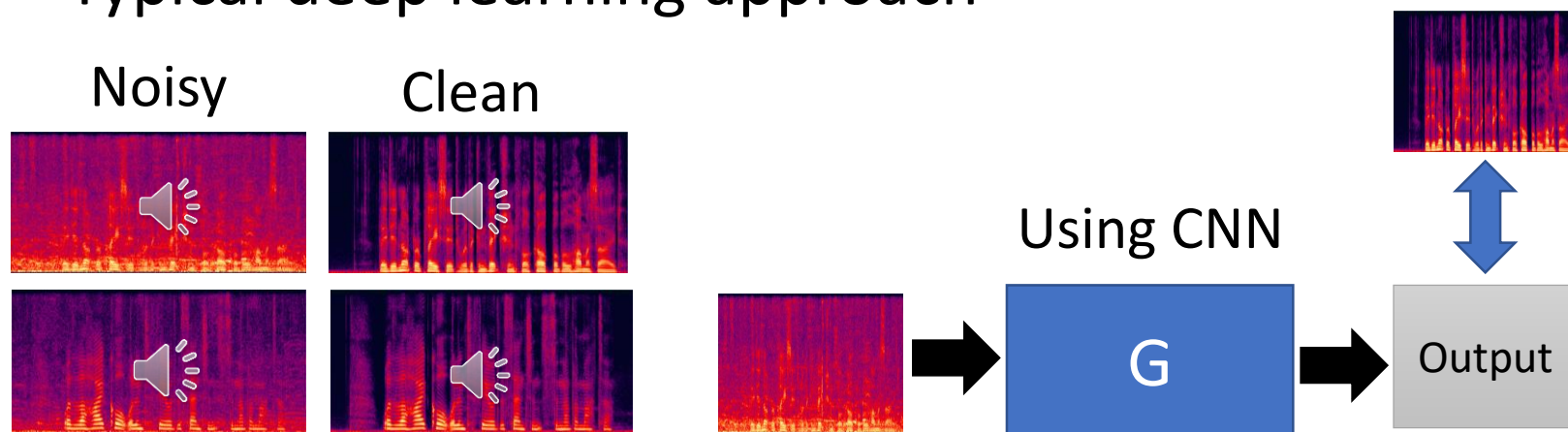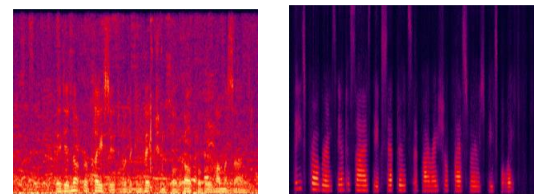| input | close | GAN | GAN + close |

Patch GAN

Speech Enhancement

Enhancing →

- Typical deep learning approach

Noisy   Clean

Using CNN

G

Output

Speech Enhancement

training data

noisy   clean

- Conditional GAN

noisy → G → output ↔ clean

output → D → scalar (fake pair or not)

noisy →

Video Generation

target

Minimize distance

Generator

Discriminator thinks it is real

Discriminator

Last frame is real or generated
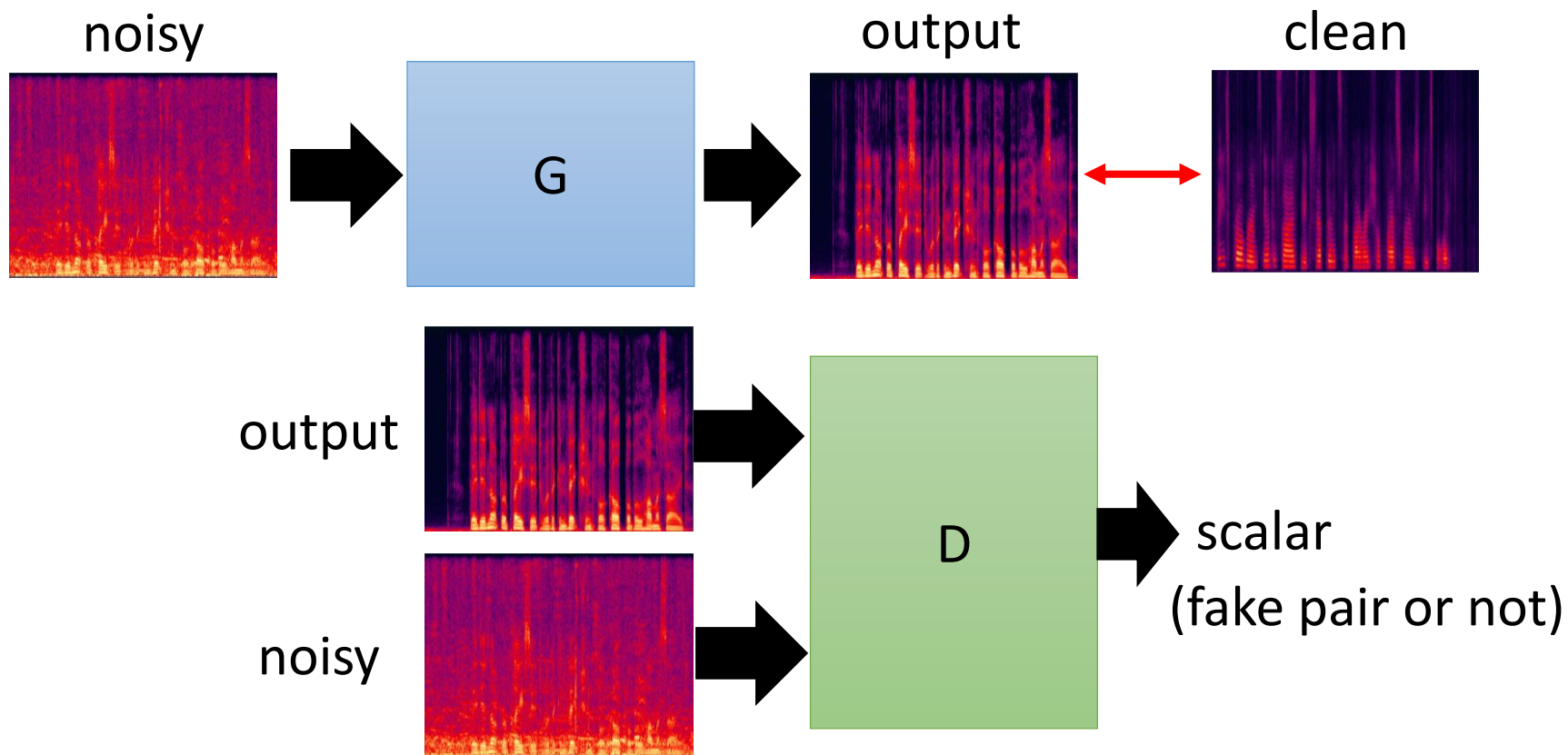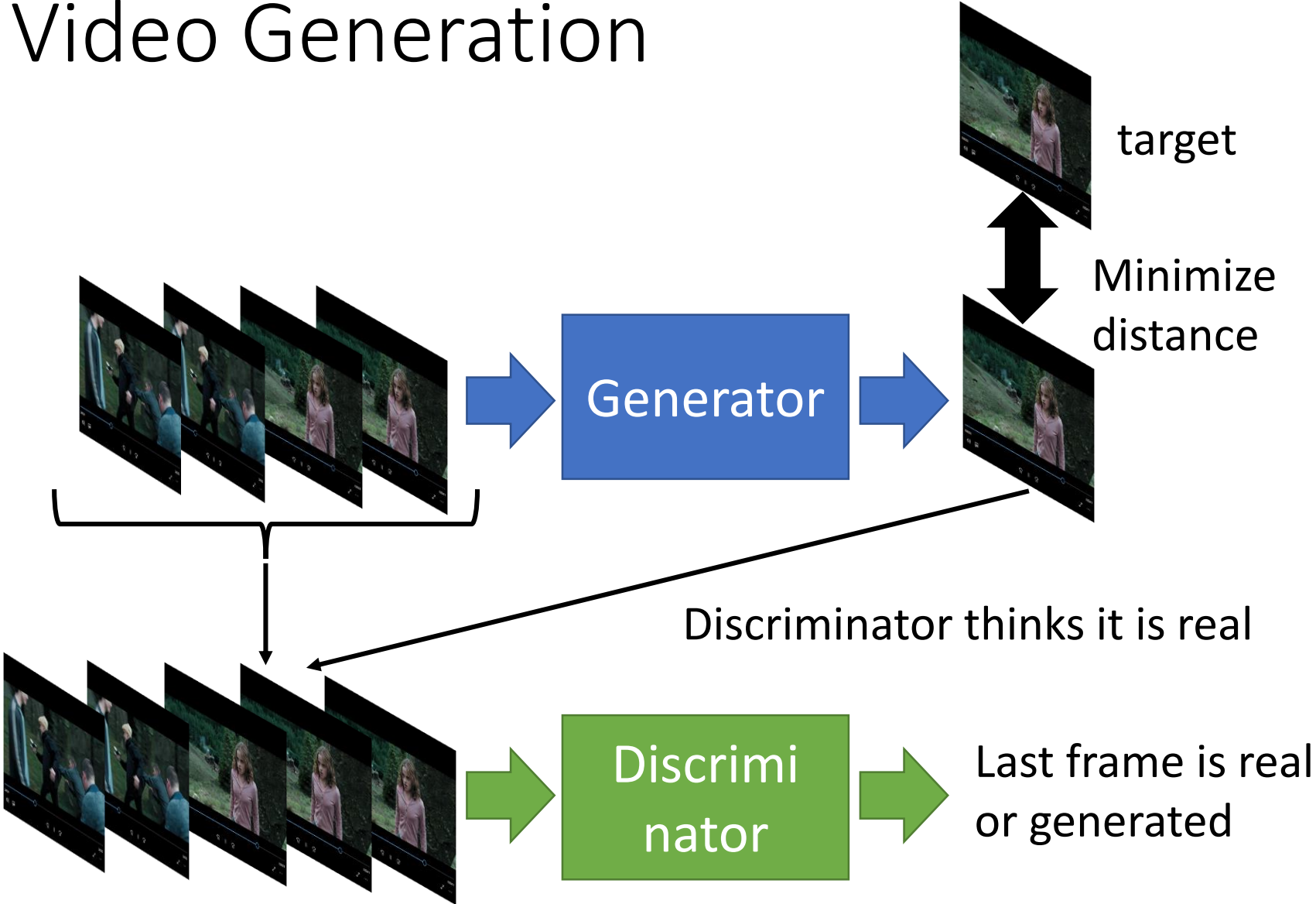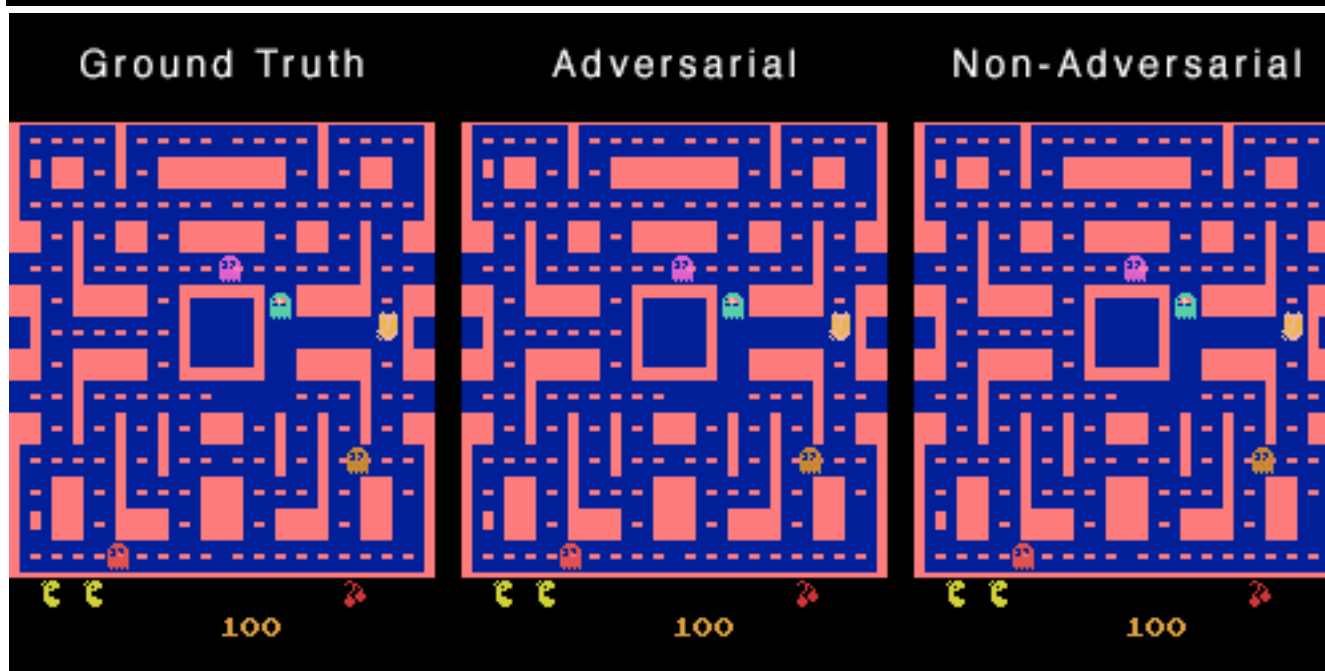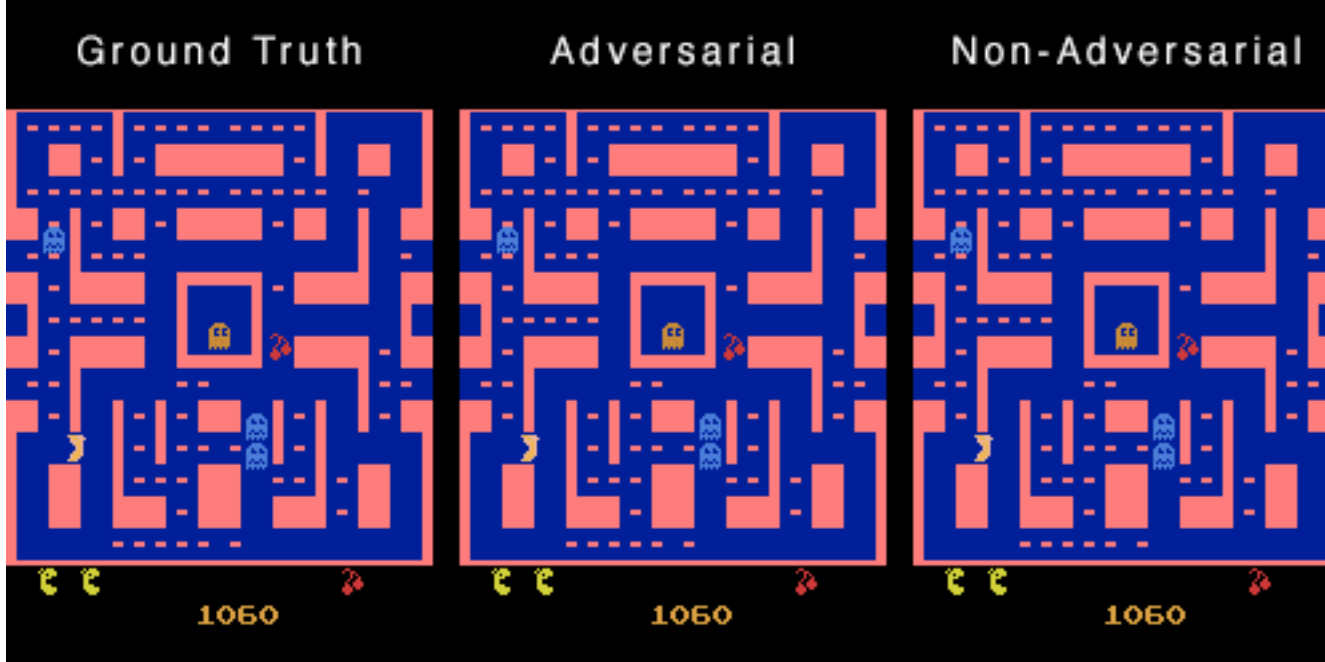
https://github.com/dyelax/Adversarial_Video_Generation