

Machine Learning, Spring 2020

Conditional Generative Adversarial Networks (GAN)

Instructor: Prof. Yi Fang

yfang@nyu.edu

Python tutorial: <http://learnpython.org/>

TensorFlow tutorial: <https://www.tensorflow.org/tutorials/>

PyTorch tutorial: <https://pytorch.org/tutorials/>

Acknowledge: The slides are partially referred to the online materials by Taegyun Joen, <https://www.slideshare.net/TaegyunJeon1/pr12-you-only-look-once-yolo-unified-realtime-object-detection> and online YOLO paper and other materials (from ECS289g by Prof. Lee)

Text-to-Image

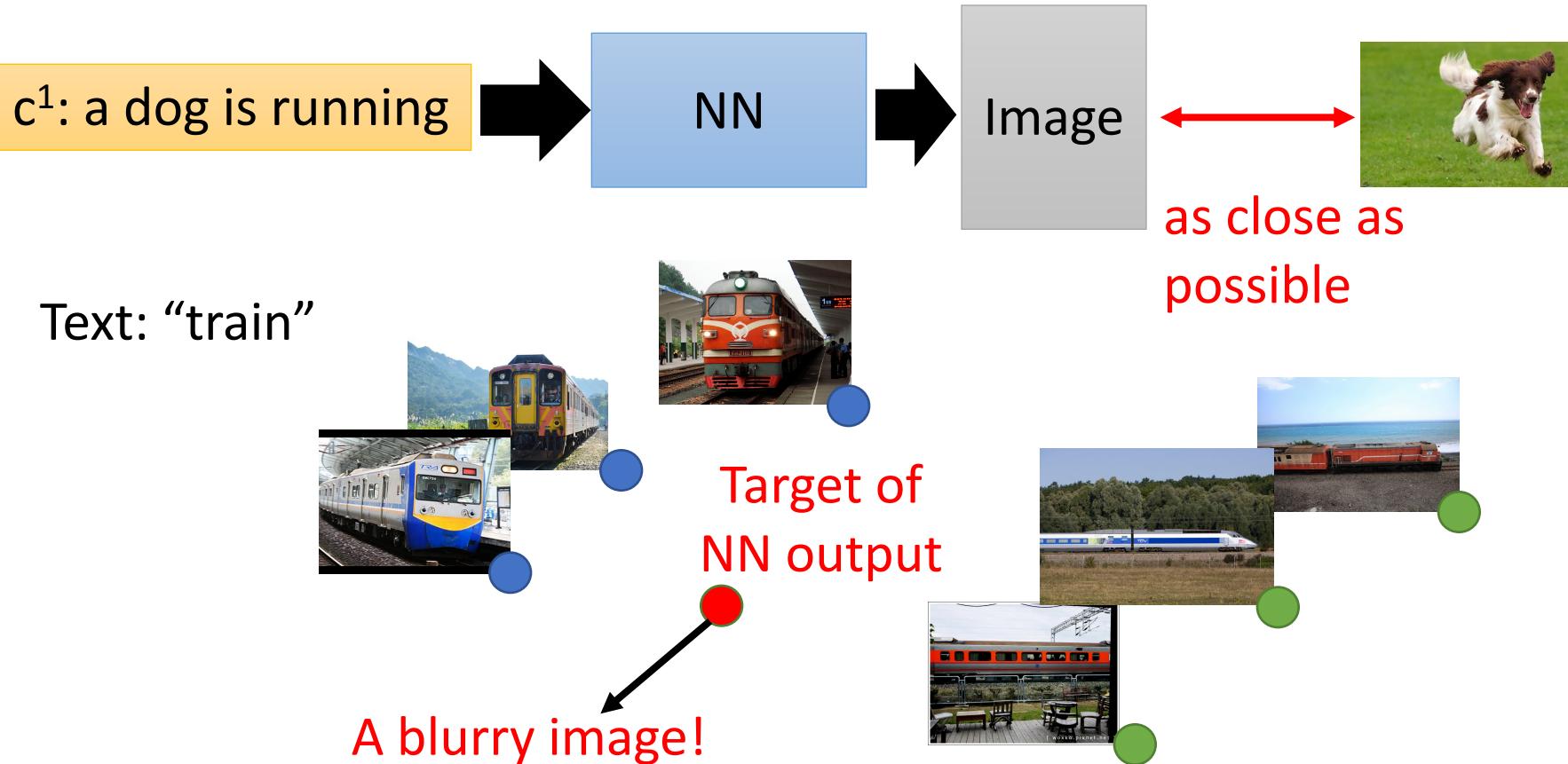
a dog is running



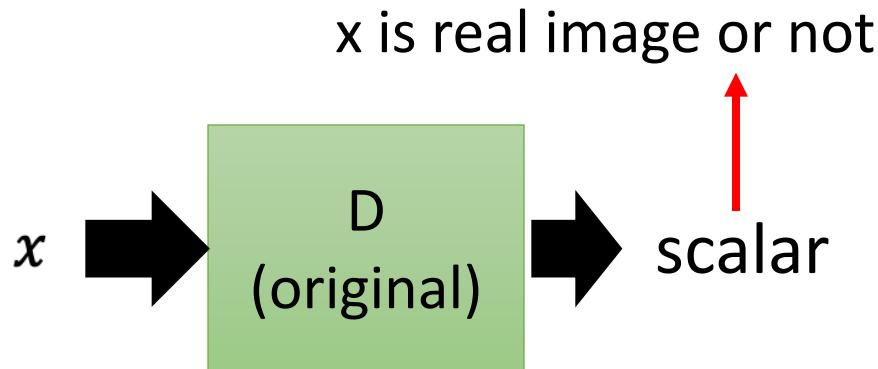
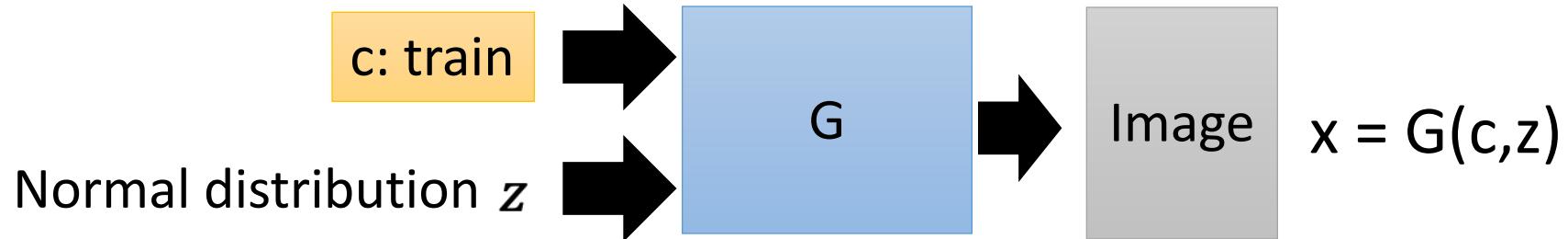
a bird is flying



- Traditional supervised approach

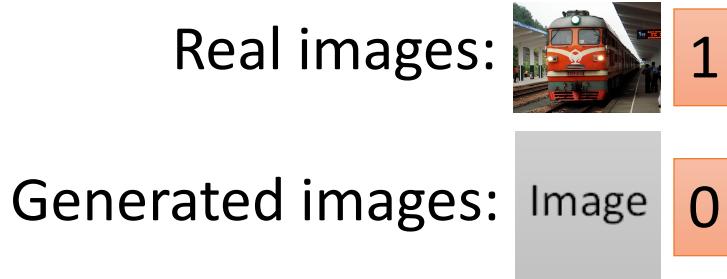


Conditional GAN

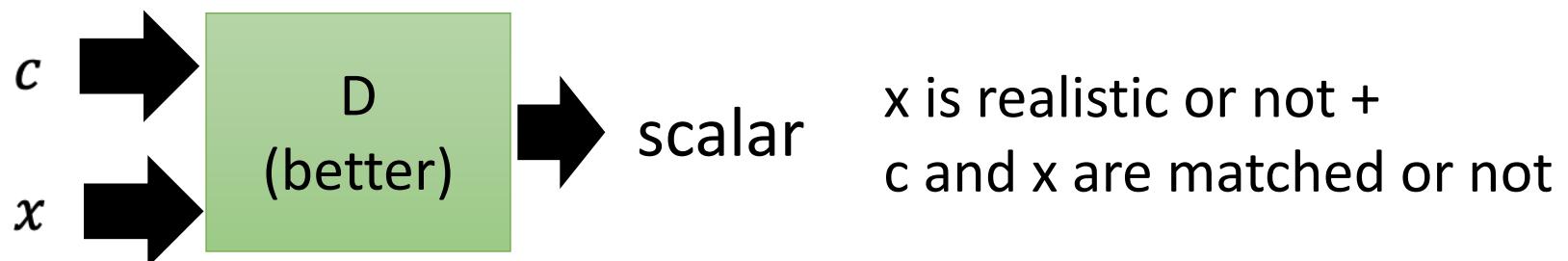
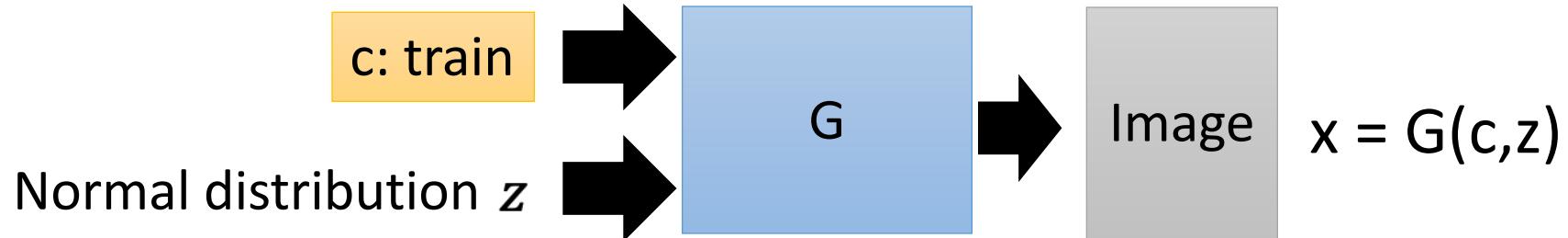


Generator will learn to generate realistic images

But completely ignore the input conditions.



Conditional GAN



True text-image pairs: (train , ) 1

(cat , ) 0

(train , ) 0

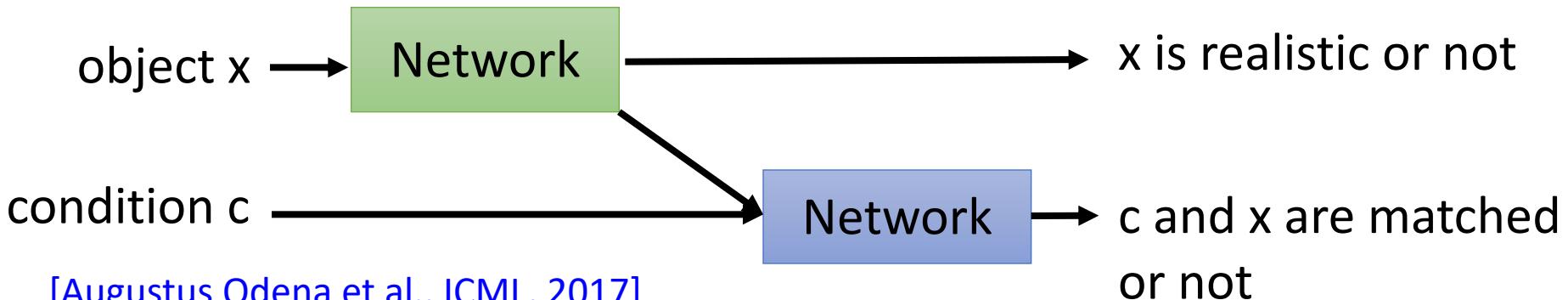
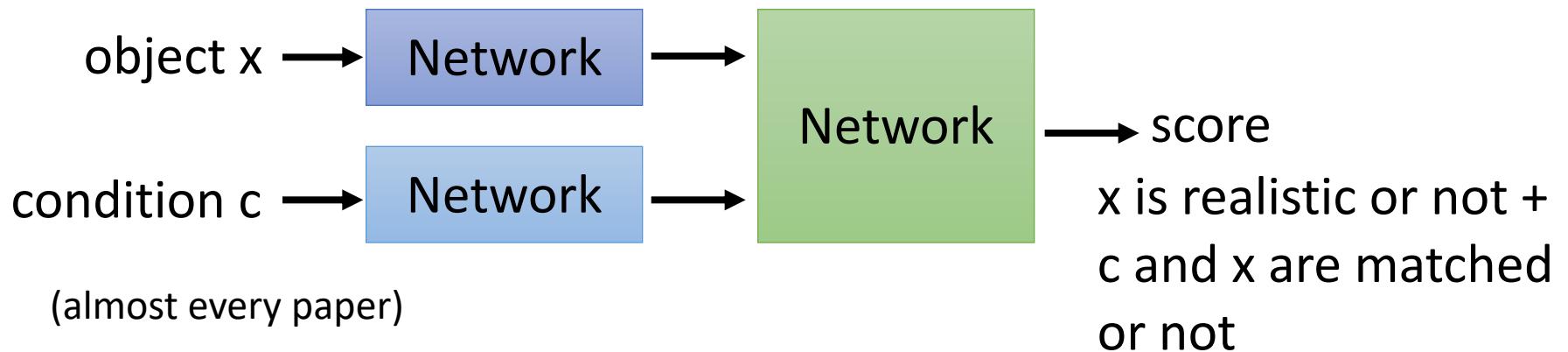
Learning
D

- In each training iteration:
 - 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
 - $\tilde{V} = \frac{1}{m} \sum_{i=1}^m \log D(c^i, x^i)$
 - $+ \frac{1}{m} \sum_{i=1}^m \log (1 - D(c^i, \tilde{x}^i)) + \frac{1}{m} \sum_{i=1}^m \log (1 - D(c^i, \hat{x}^i))$ •
 - $\theta_d \leftarrow \theta_d + \eta \nabla \tilde{V}(\theta_d)$

Learning
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
 - $\tilde{V} = \frac{1}{m} \sum_{i=1}^m \log (D(G(c^i, z^i)))$, $\theta_g \leftarrow \theta_g - \eta \nabla \tilde{V}(\theta_g)$

Conditional GAN - Discriminator



[Augustus Odena et al., ICML, 2017]

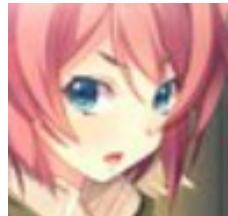
[Takeru Miyato, et al., ICLR, 2018]

[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



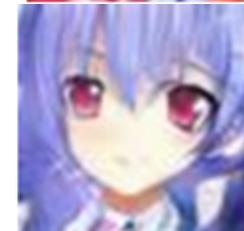
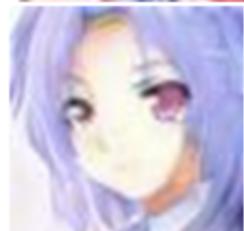
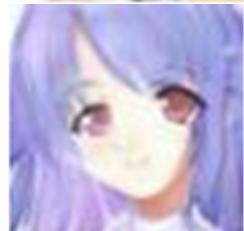
blue eyes
red hair
short hair

Collecting anime faces
and the description of its
characteristics

red hair,
green eyes



blue hair,
red eyes



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

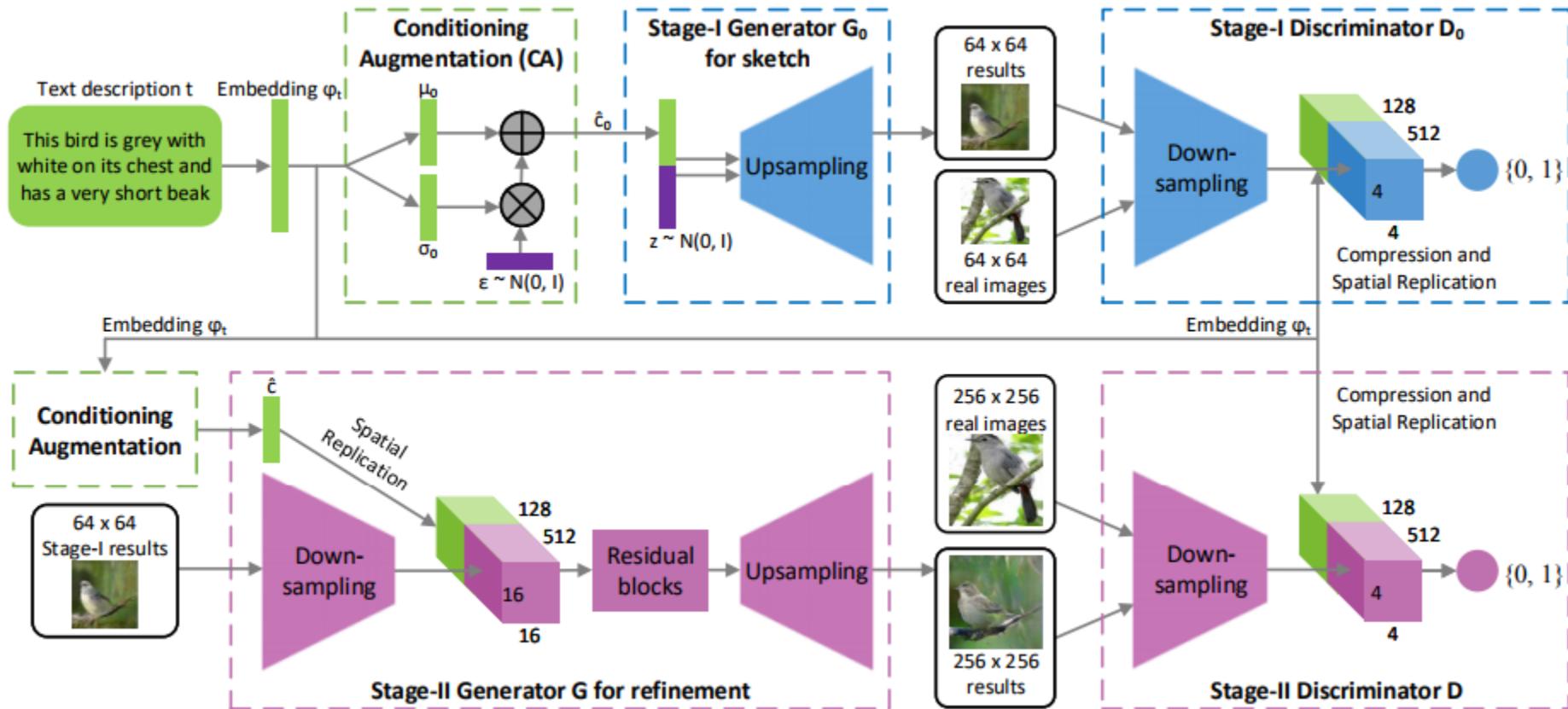
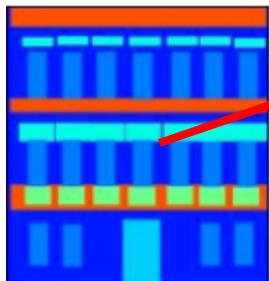


Image-to-image



c
 z



$$x = G(c, z)$$



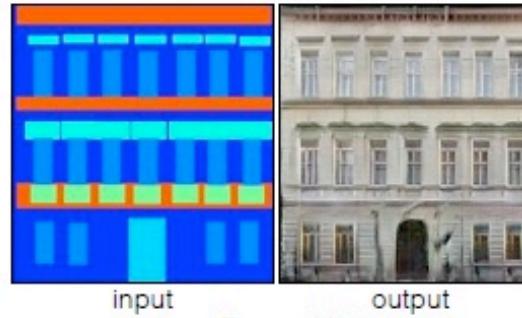
Labels to Street Scene



Aerial to Map



Labels to Facade



BW to Color



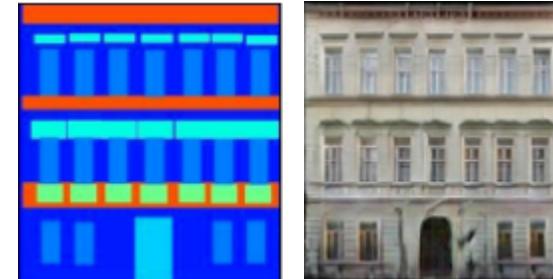
Day to Night



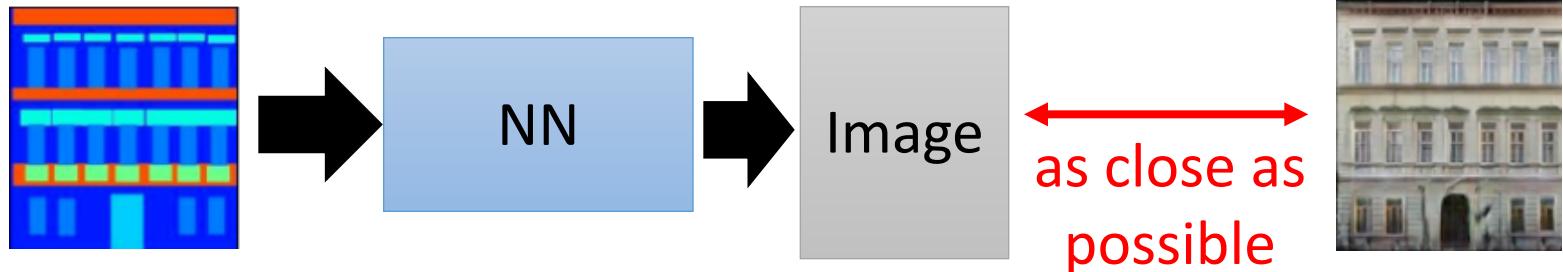
Edges to Photo



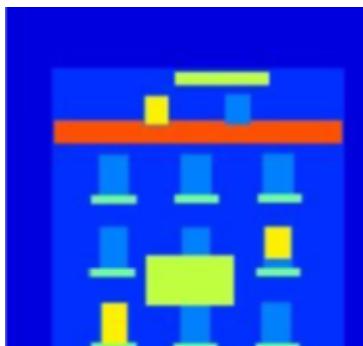
Image-to-image



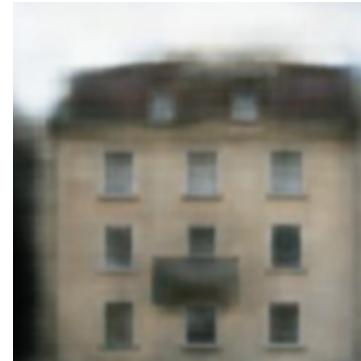
- Traditional supervised approach



Testing:



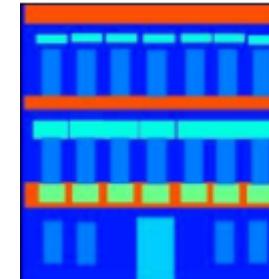
input



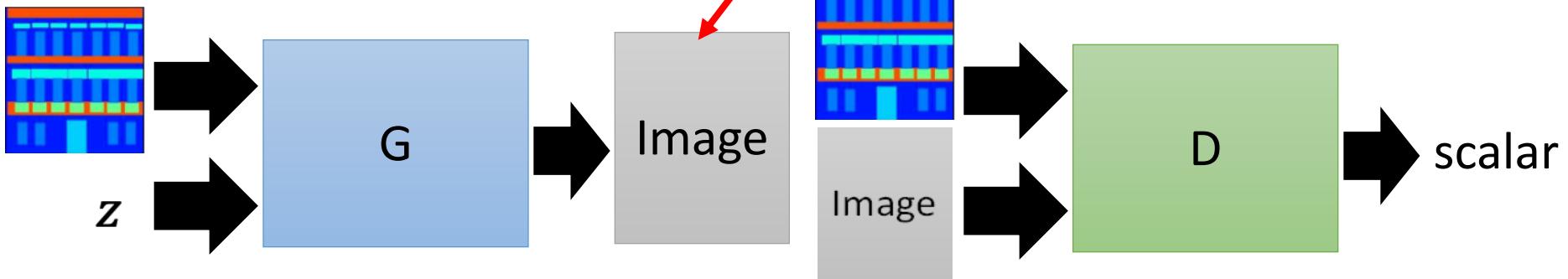
close

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

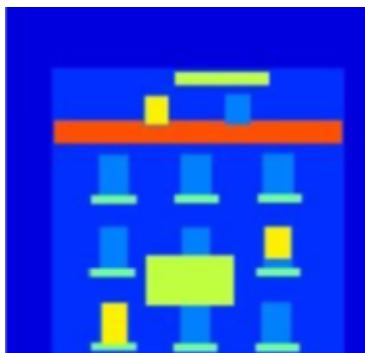
Image-to-image



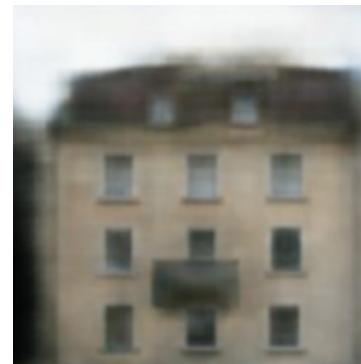
- Experimental results



Testing:



input



close



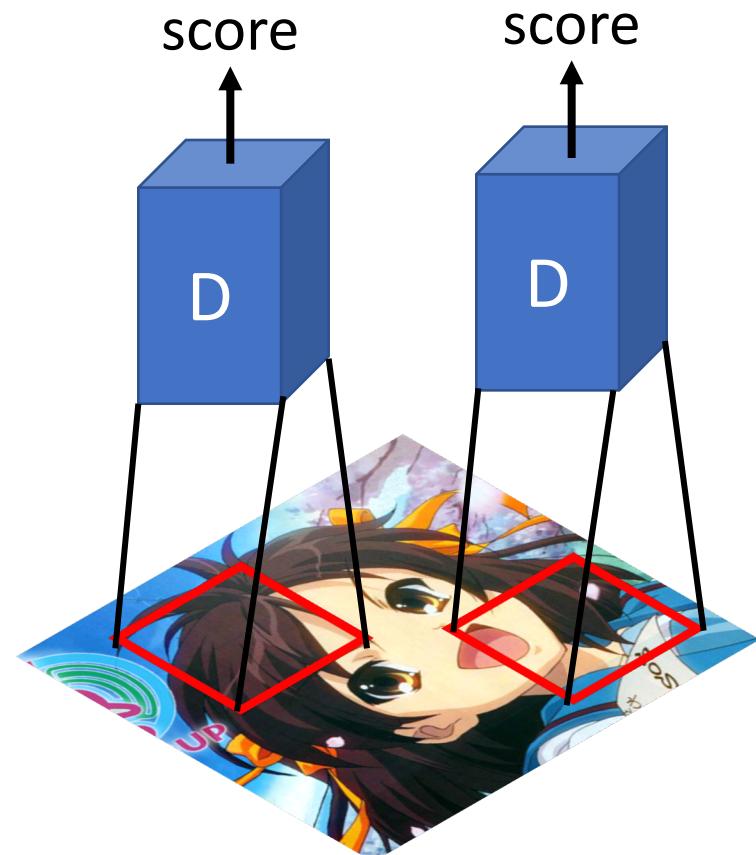
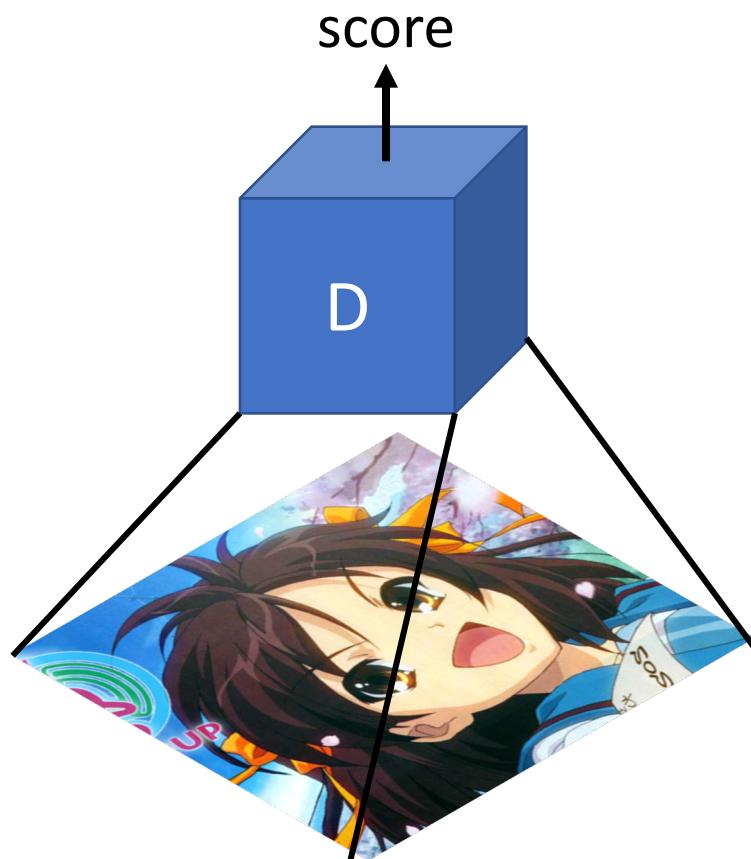
GAN



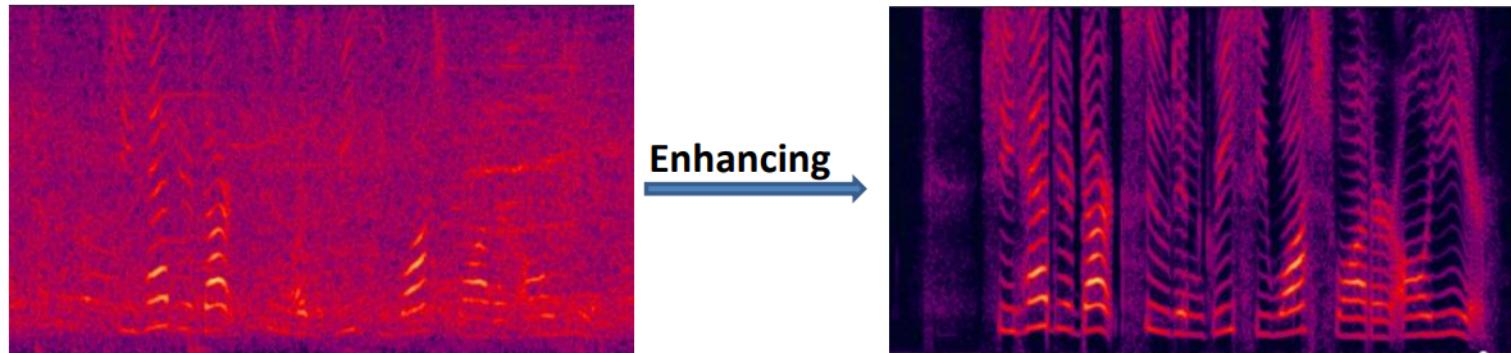
GAN + close

Patch GAN

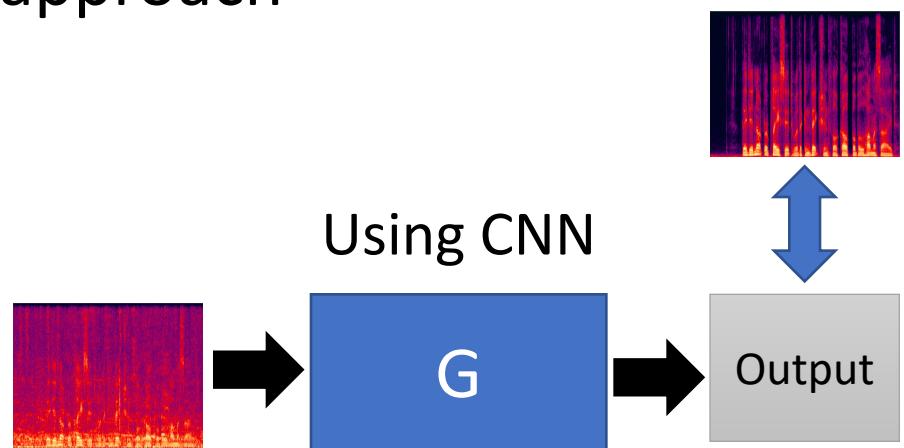
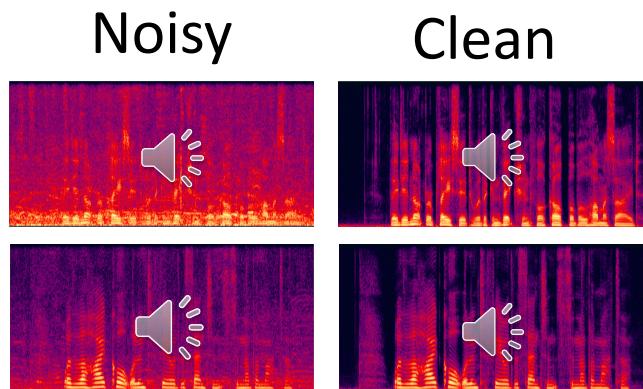
<https://arxiv.org/pdf/1611.07004.pdf>



Speech Enhancement

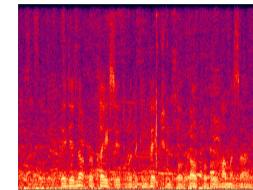


- Typical deep learning approach

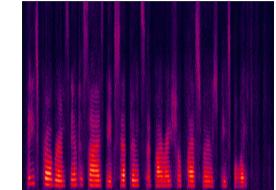


Speech Enhancement

training data

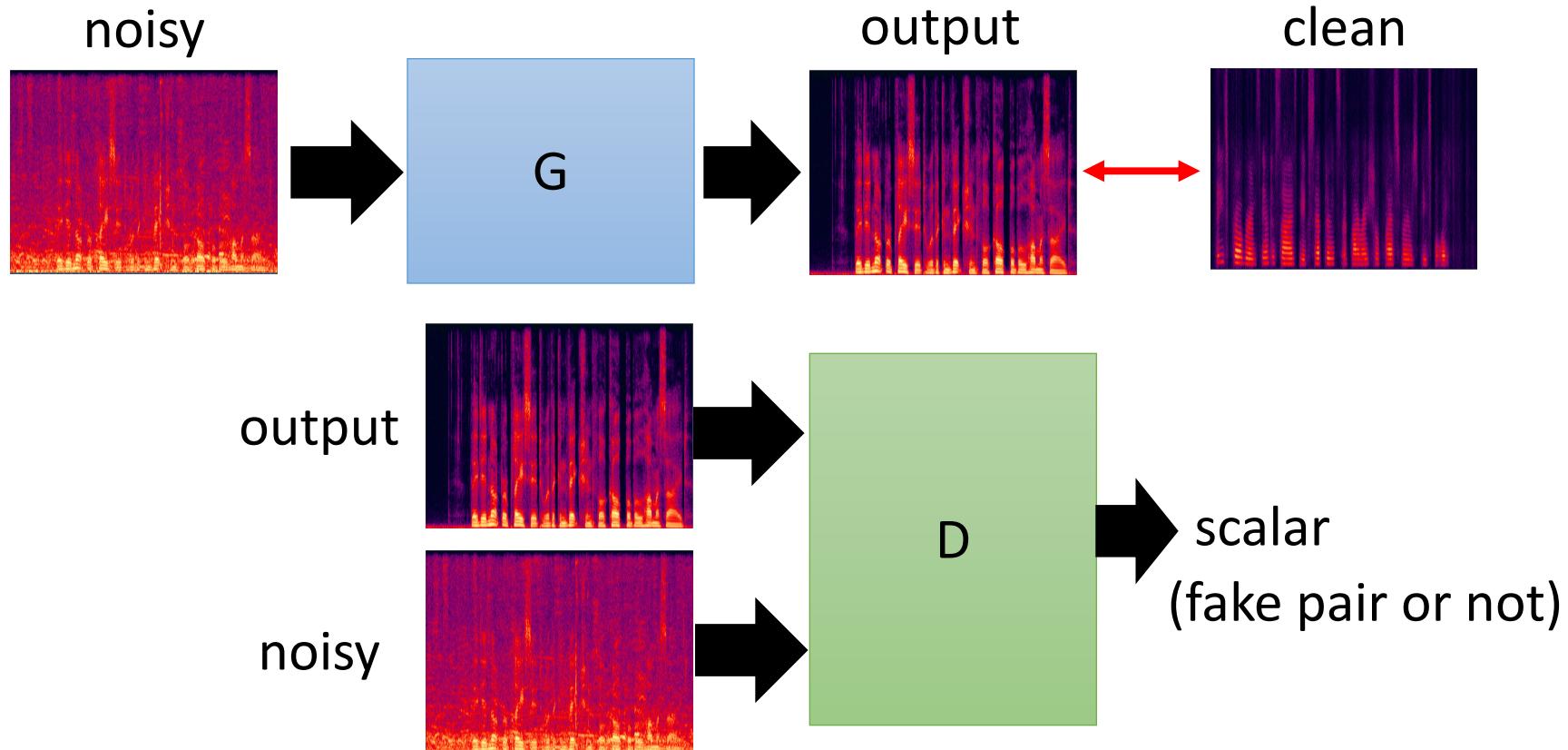


noisy

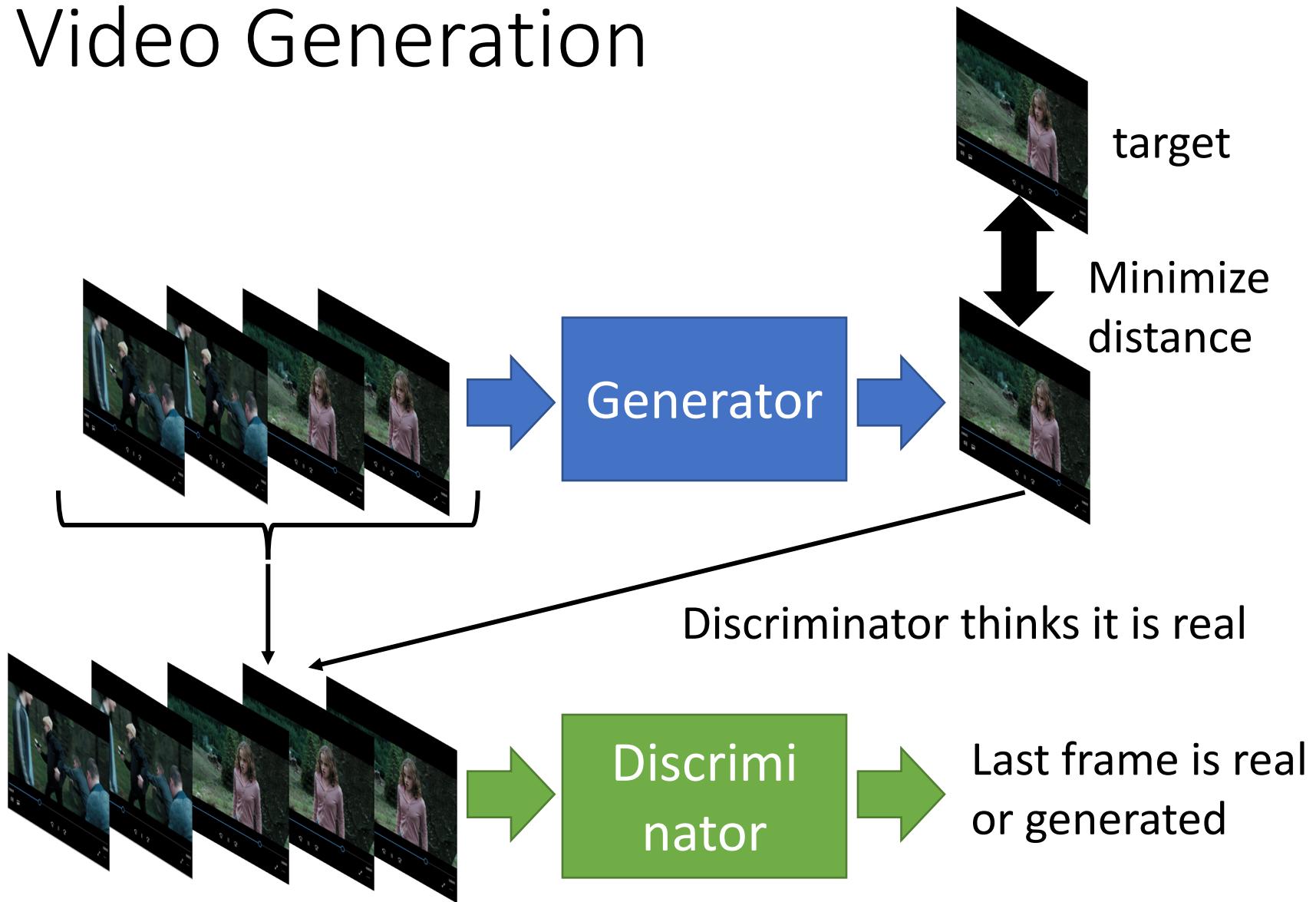


clean

- Conditional GAN



Video Generation



References

- <http://slazebni.cs.illinois.edu/spring17/>
- <https://cs.uwaterloo.ca/~mli/Deep-Learning-2017-Lecture7GAN.ppt>