Conditional Generation by GAN

Text-to-Image

a dog is running

a bird is flying

- **Traditional supervised approach**

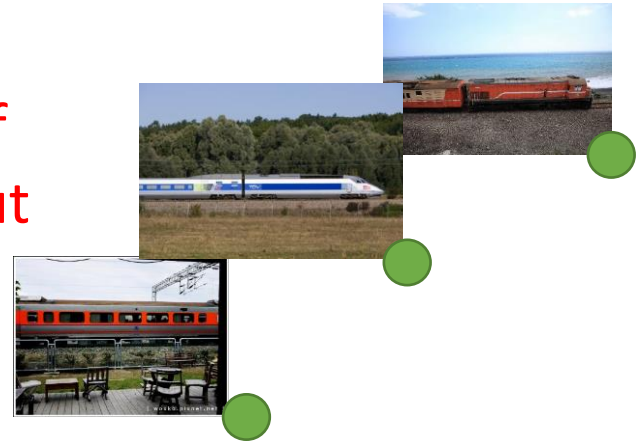
c^1: a dog is running → NN → Image ↔ 

as close as possible

Text: “train”

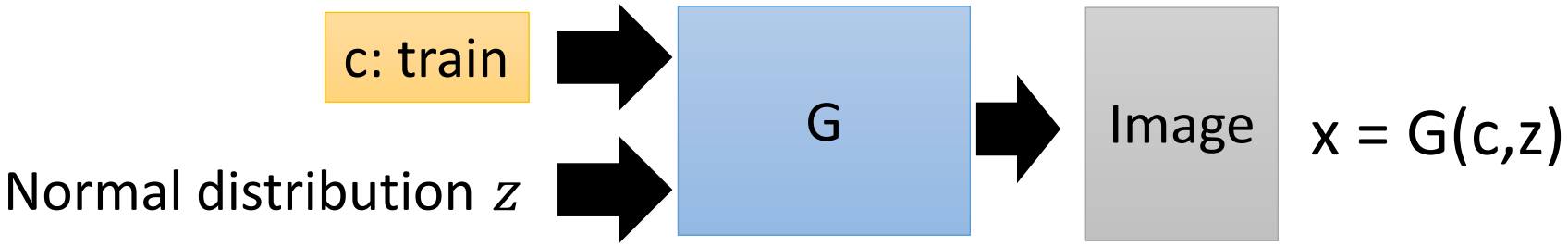
Target of NN output

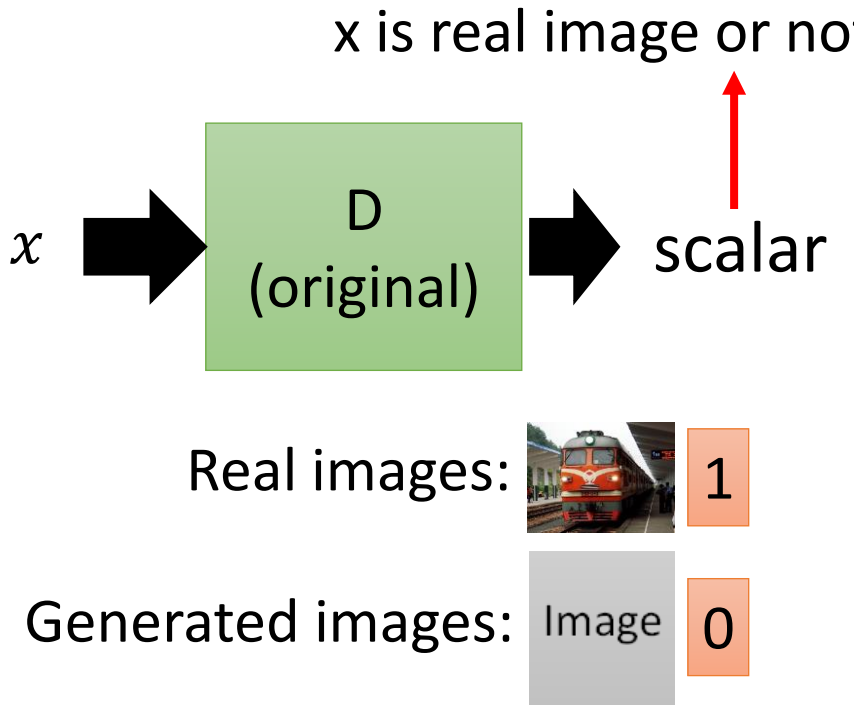
A blurry image!



[Scott Reed, et al, ICML, 2016]

Conditional GAN



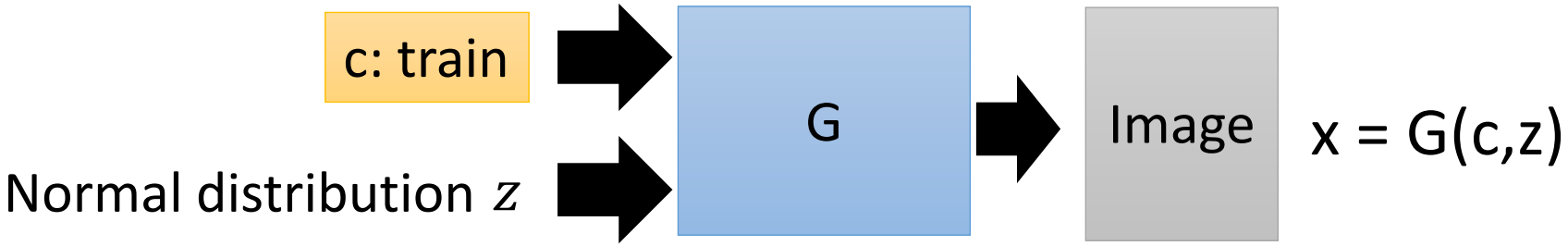


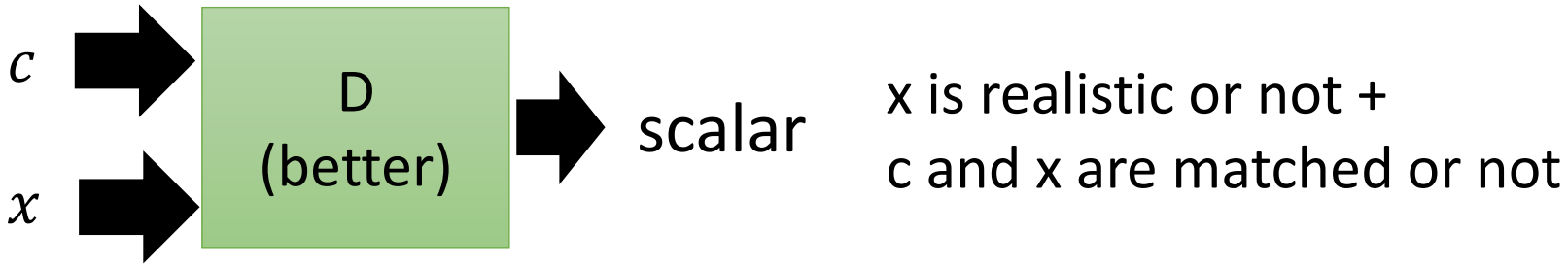
Generator will learn to generate realistic images ….

But completely ignore the input conditions.

[Scott Reed, et al, ICML, 2016]

Conditional GAN



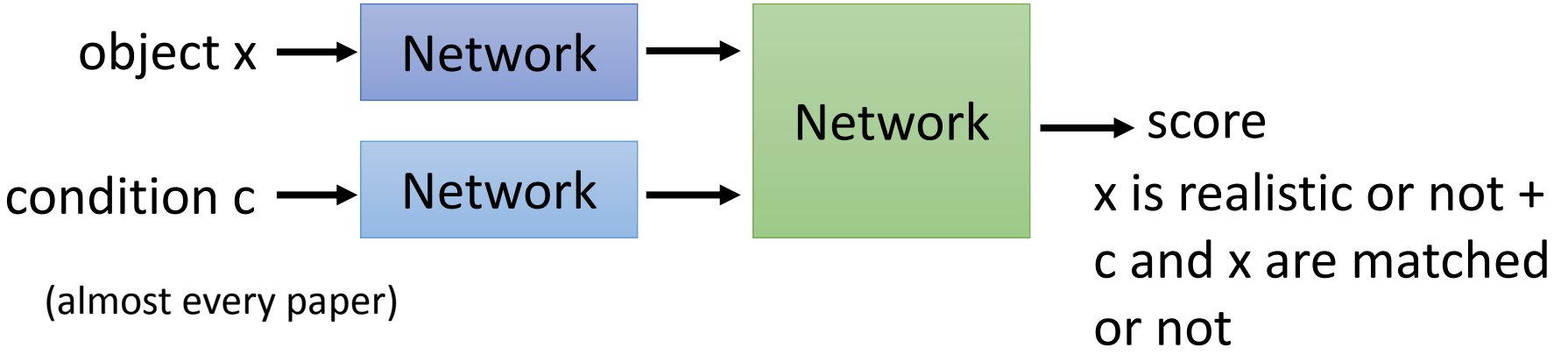


True text-image pairs: (train , ) 1

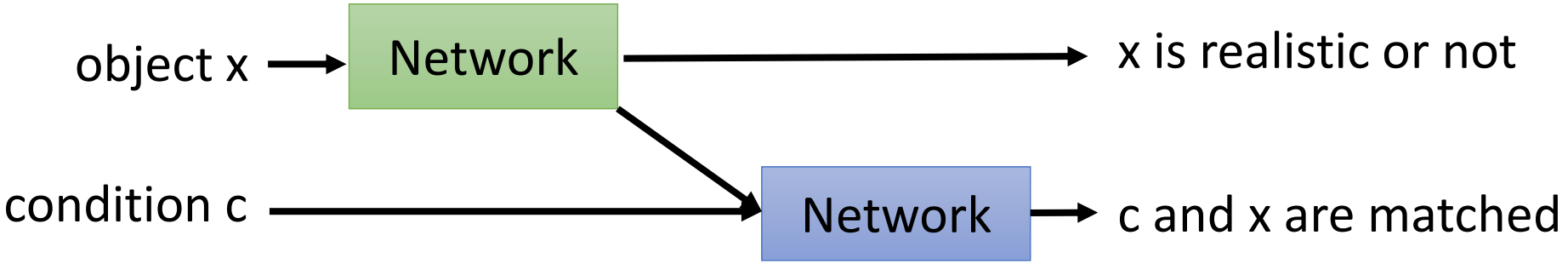
(cat , ) 0

(train , Image) 0

Conditional GAN - Discriminator

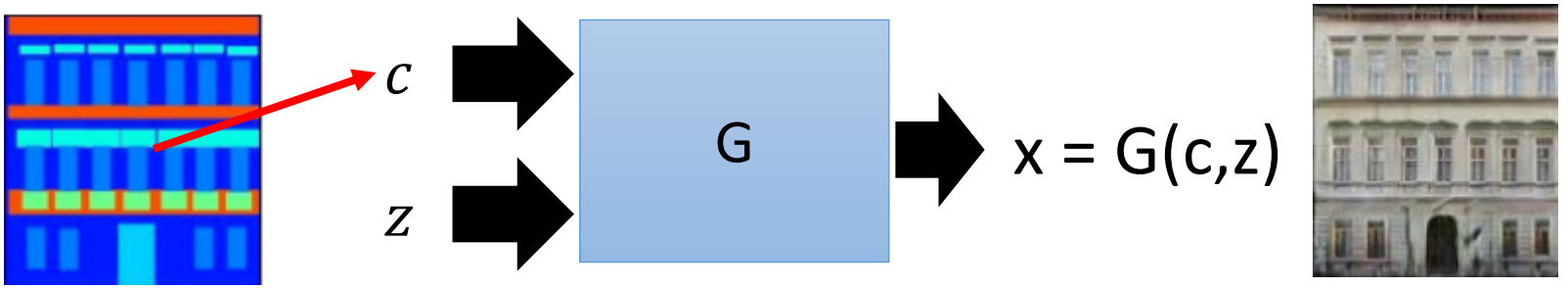


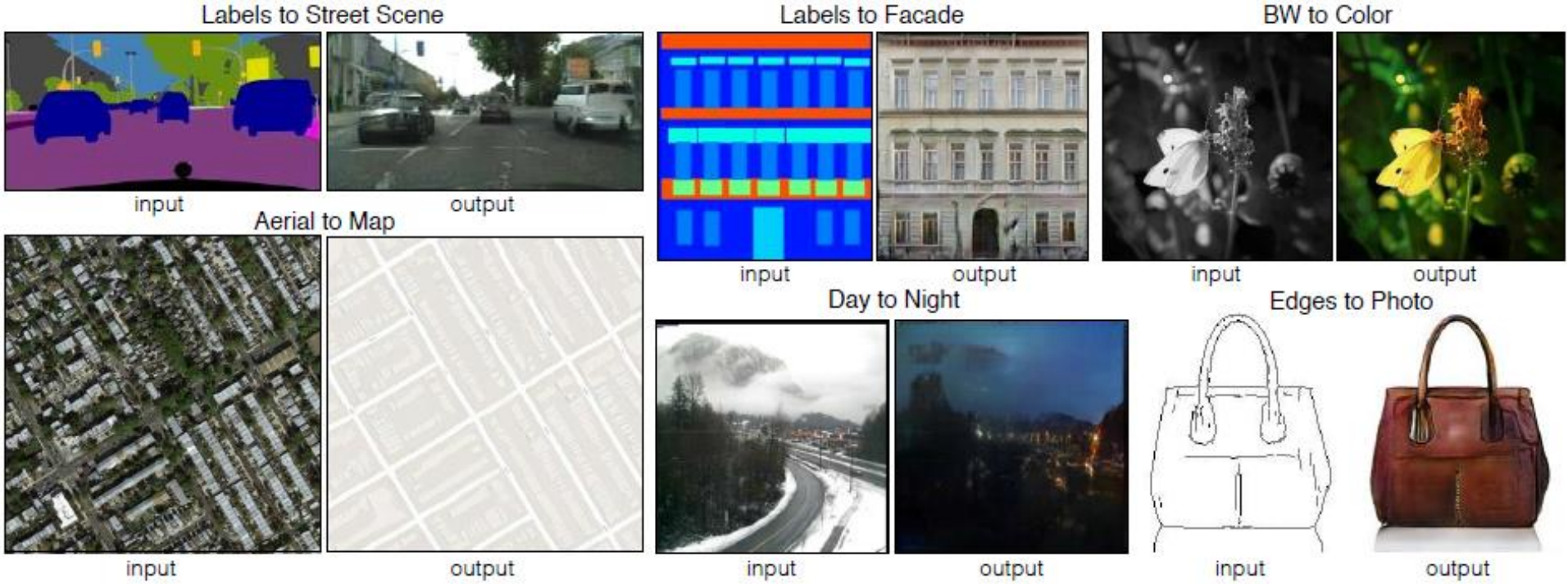
(almost every paper)



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

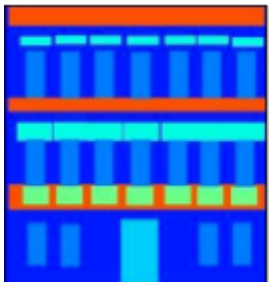
Image-to-image



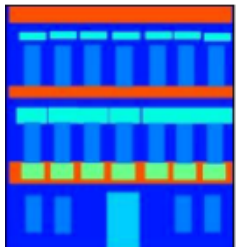


https://arxiv.org/pdf/1611.07004

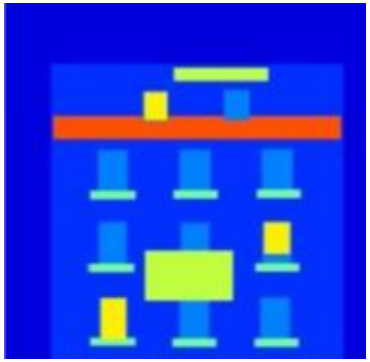
Image-to-image




- Traditional supervised approach



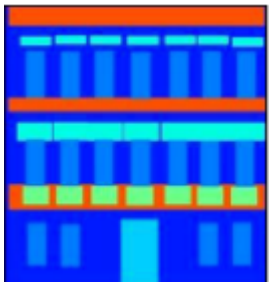
Testing:


input

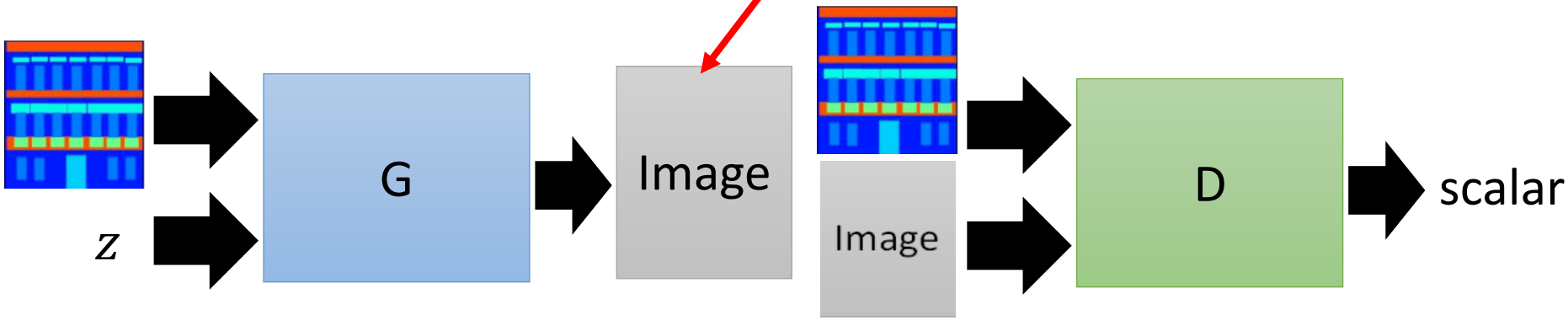

close

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

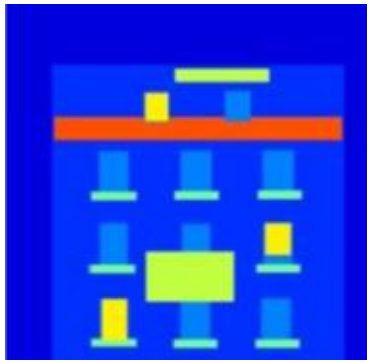
Image-to-image



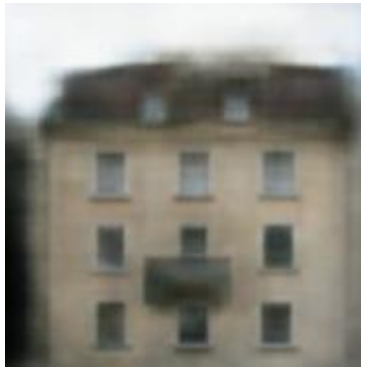
- Experimental results



Testing:



input



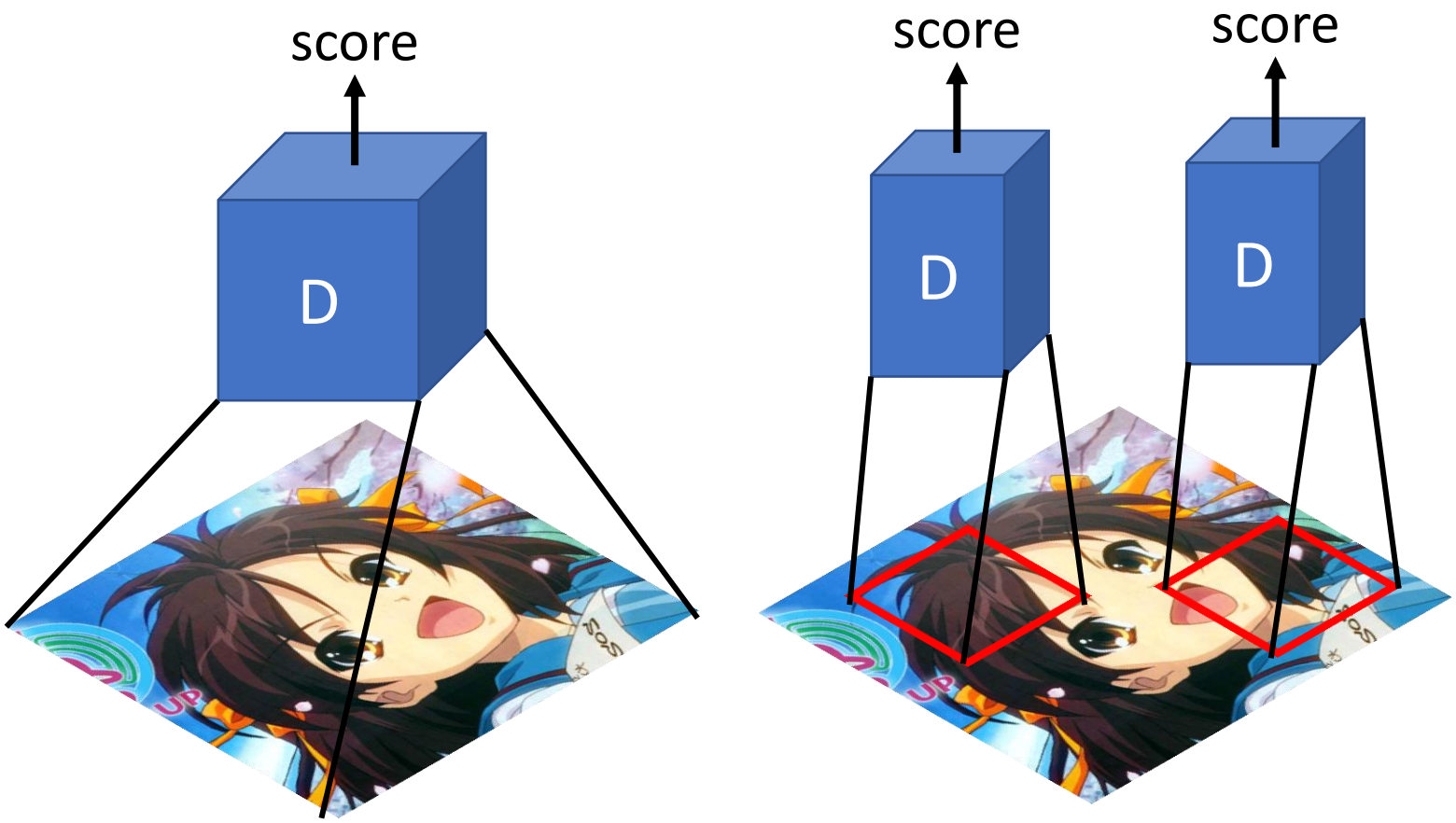
close

GAN

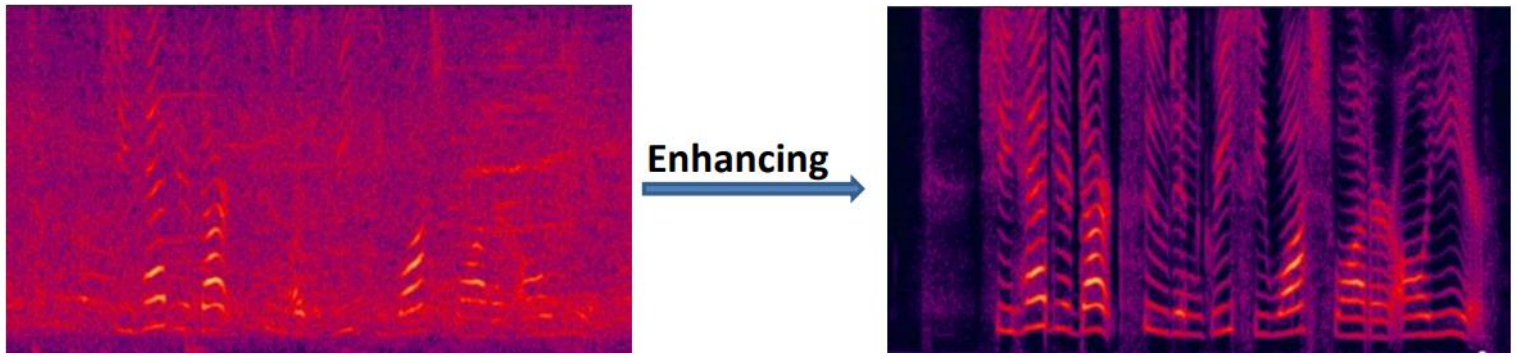
GAN + close

Patch GAN

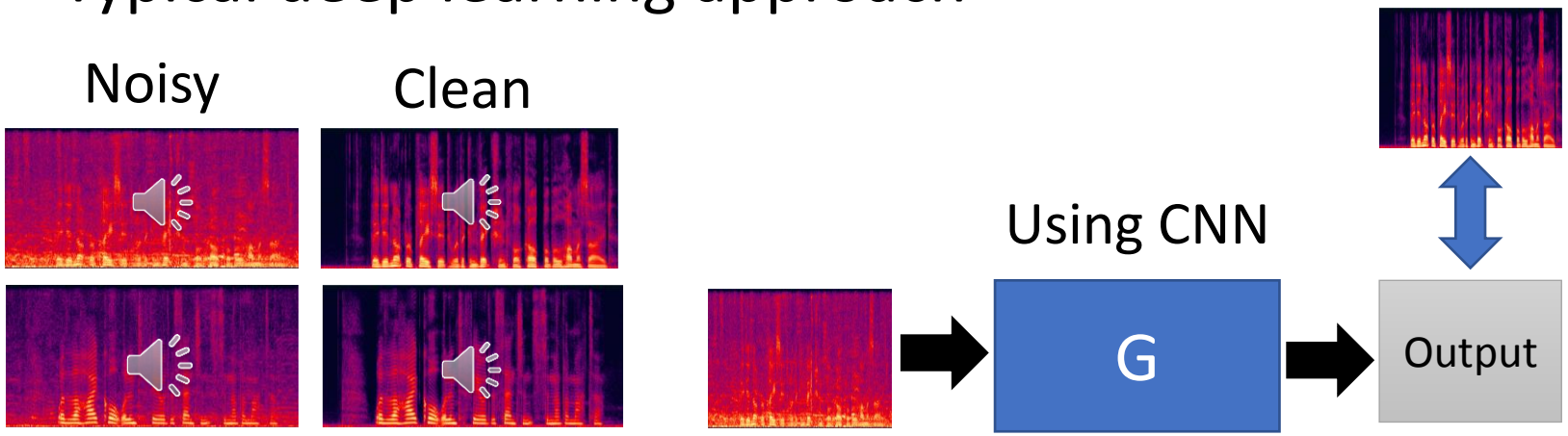
https://arxiv.org/pdf/1611.07004.pdf



Speech Enhancement

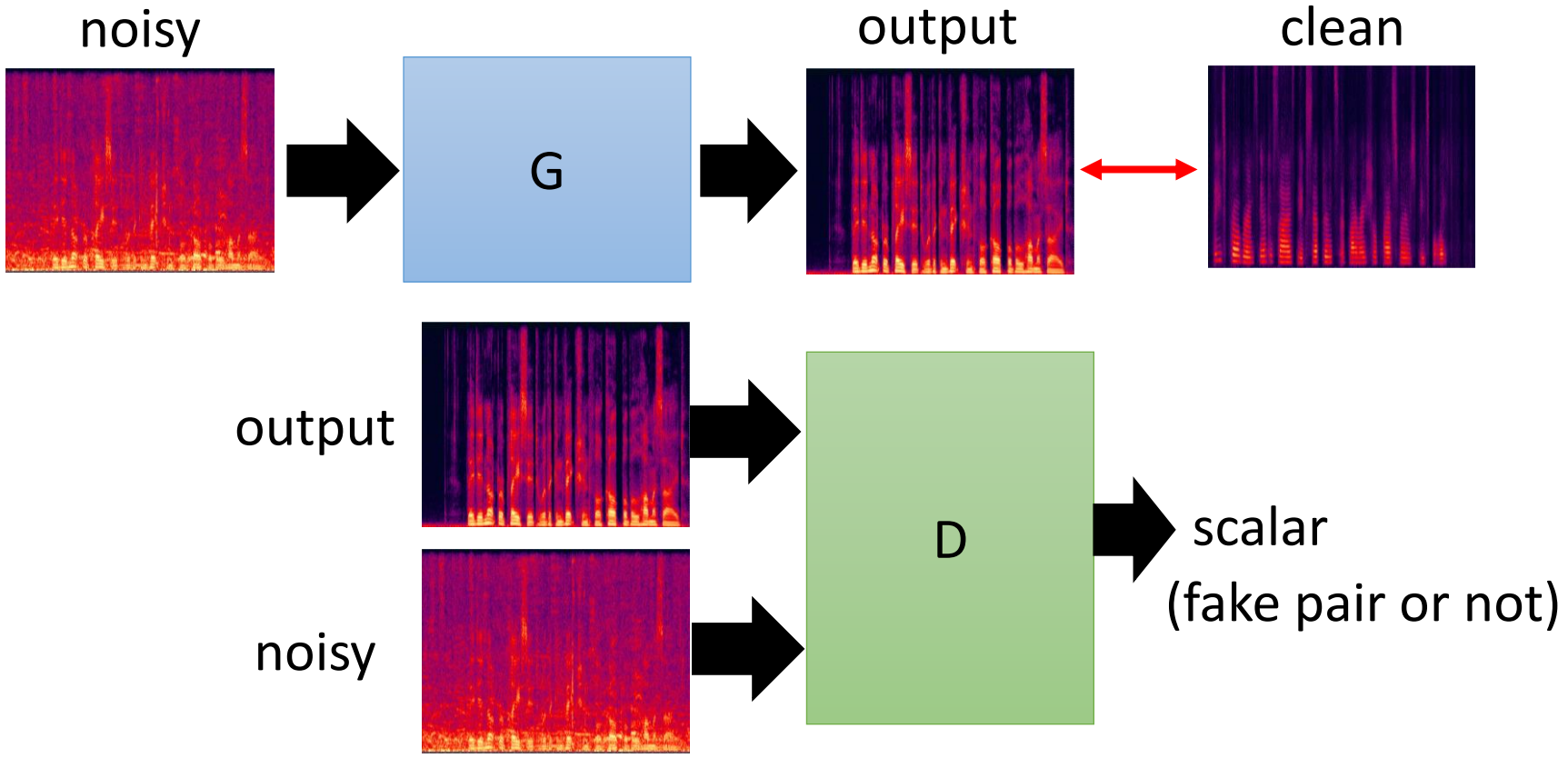


- Typical deep learning approach



Speech Enhancement

- Conditional GAN



Video Generation