

# Copyright Notice

These slides are distributed under the Creative Commons License.

DeepLearning.AI makes these slides available for educational purposes. You may not use or distribute these slides for commercial purposes. You may make copies of these slides and use or distribute them for educational purposes as long as you cite DeepLearning.AI as the source of the slides.

For the rest of the details of the license, see

<https://creativecommons.org/licenses/by-sa/2.0/legalcode>

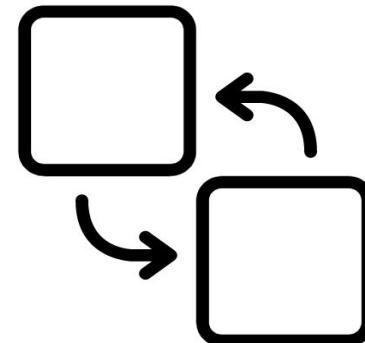


deeplearning.ai

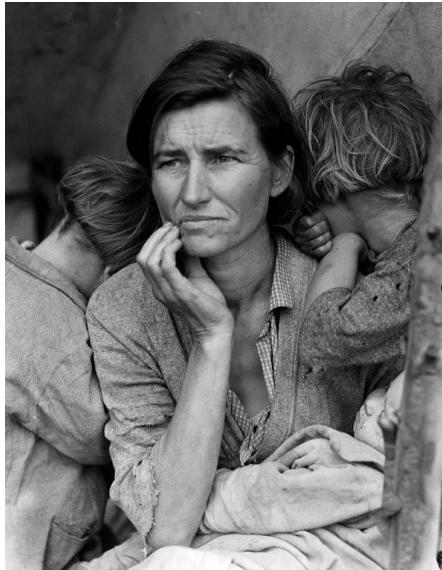
# Image-to-Image Translation

# Outline

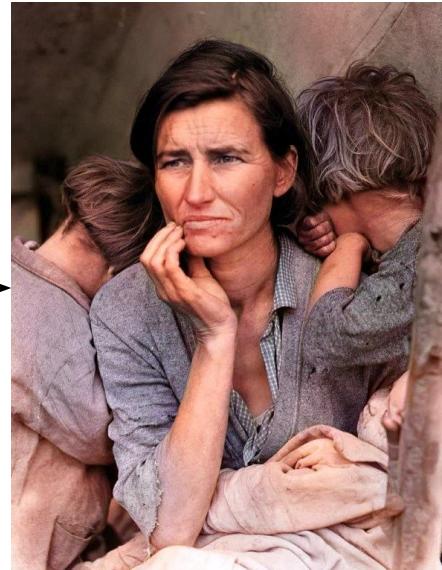
- Image-to-image translation
- Other types of translation



# Image-to-Image Translation



Transformation



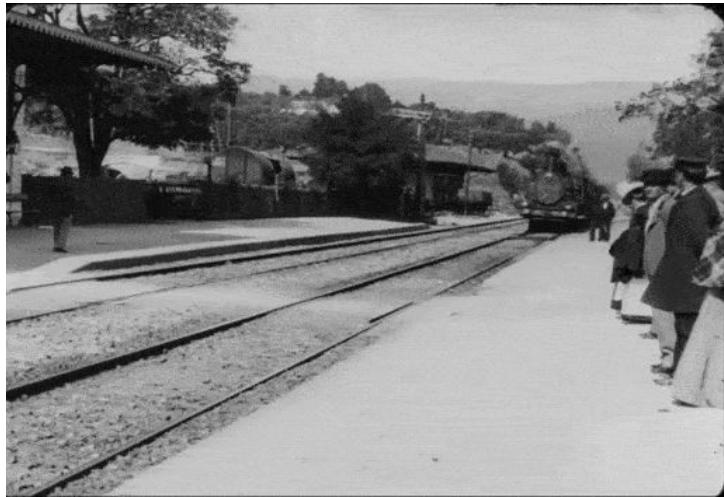
Available from: <https://twitter.com/citnaj/status/1124904251128406016>

# Image-to-Image Translation



Available from: <https://github.com/NVIDIA/pix2pixHD>

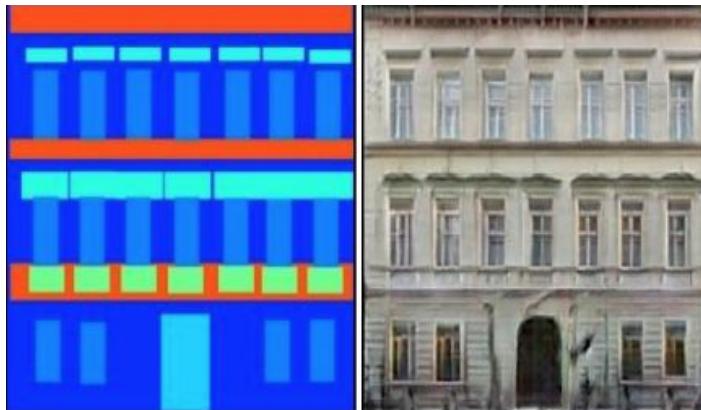
# Image-to-Image Translation



Available from: <https://youtu.be/3RYNThid23g>

# Paired Image-to-Image Translation

Labels to facade



Black-and-white to color



Available from: <https://arxiv.org/abs/1611.07004>

# Paired Image-to-Image Translation

Day to night



Edges to photo

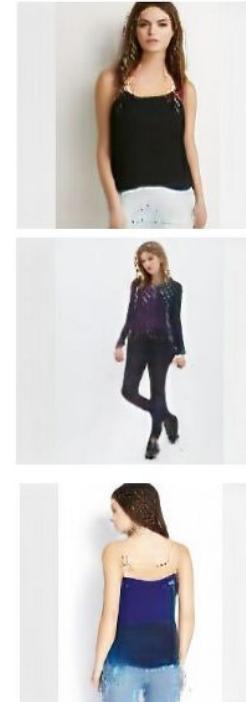


Available from: <https://arxiv.org/abs/1611.07004>

# Paired Image-to-Image Translation



Clothes and pose to  
*pose with clothes*



Available from: <https://arxiv.org/abs/1705.09368>

# Other Translations

“This bird is red with white and has a very short beak”



Available from: <https://arxiv.org/abs/1711.10485>

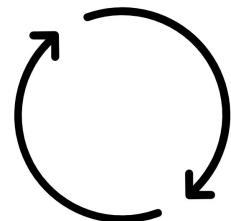
# Other Translations



Available from: <https://arxiv.org/abs/1905.08233>

# Summary

- Image-to-image translation transforms images into different styles
- GANs' realistic generation abilities are well-suited to image-to-image translation tasks
- Other types of translation include text-to-image or image-to-video



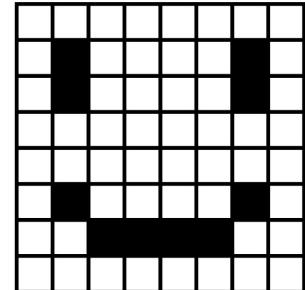


deeplearning.ai

# Pix2Pix Overview

# Outline

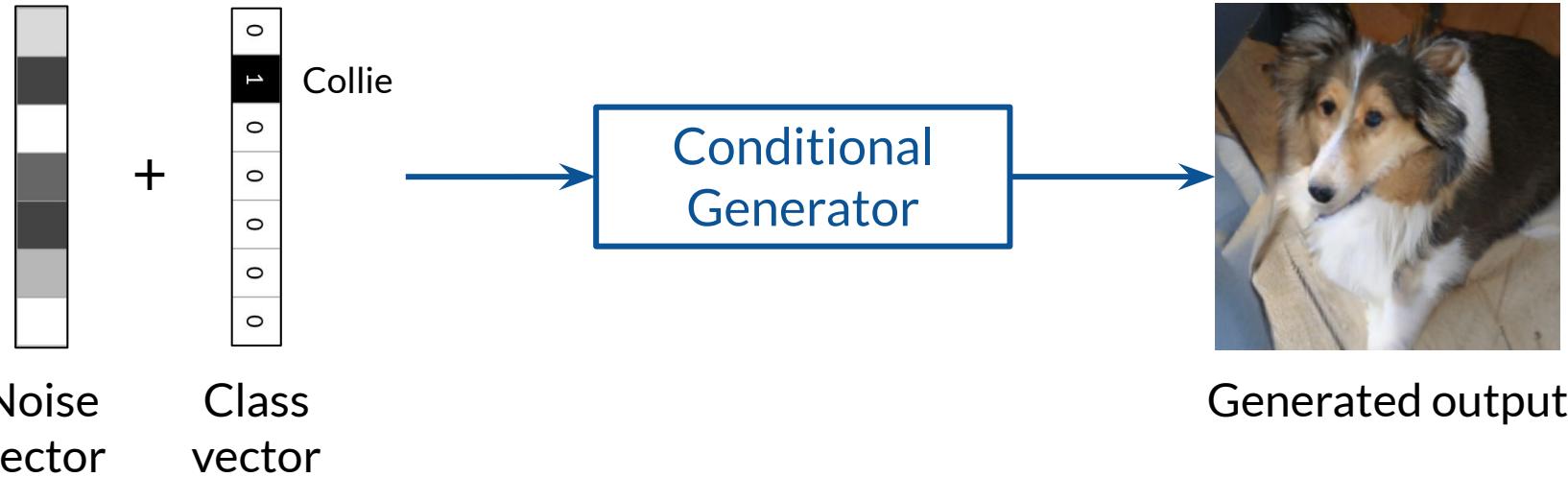
- Overview of Pix2Pix
- Comparison with conditional GAN
- Upgraded generator and discriminator architectures



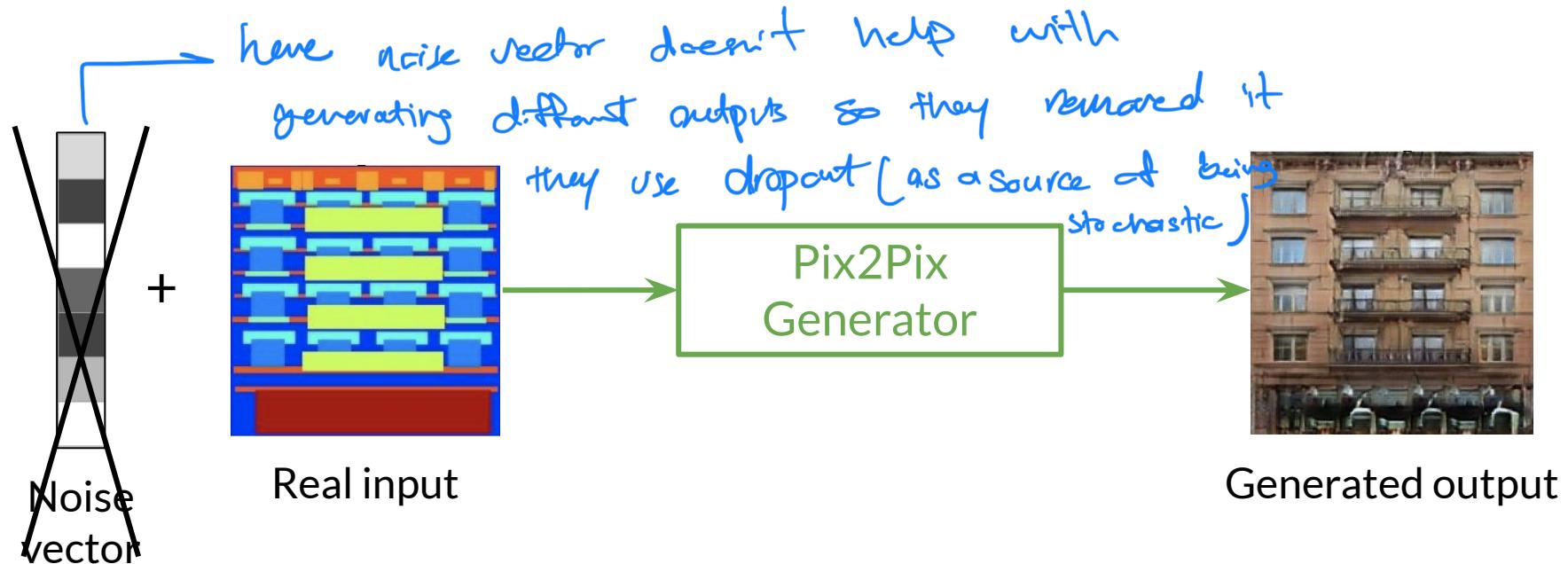
# Pix2Pix for Paired Image-to-Image Translation

Image-to-Image  $\longrightarrow$  *Pix-to-Pix*  $\longrightarrow$  *Pix2Pix*

# Pix2Pix Generator

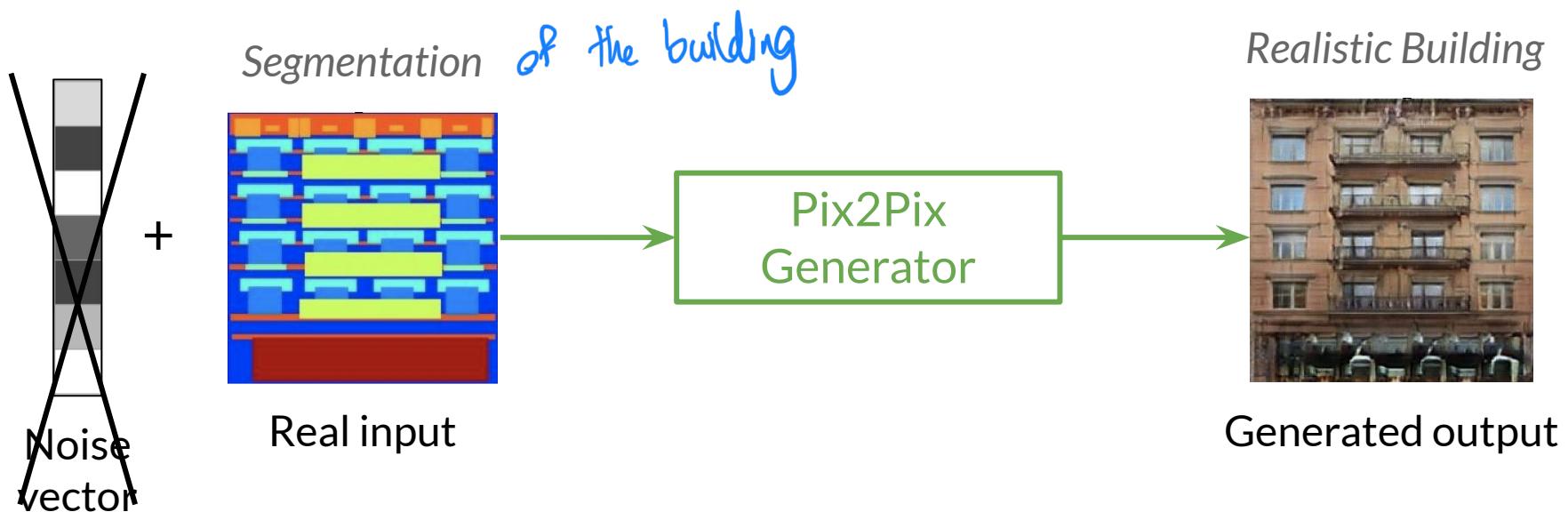


# Pix2Pix Generator



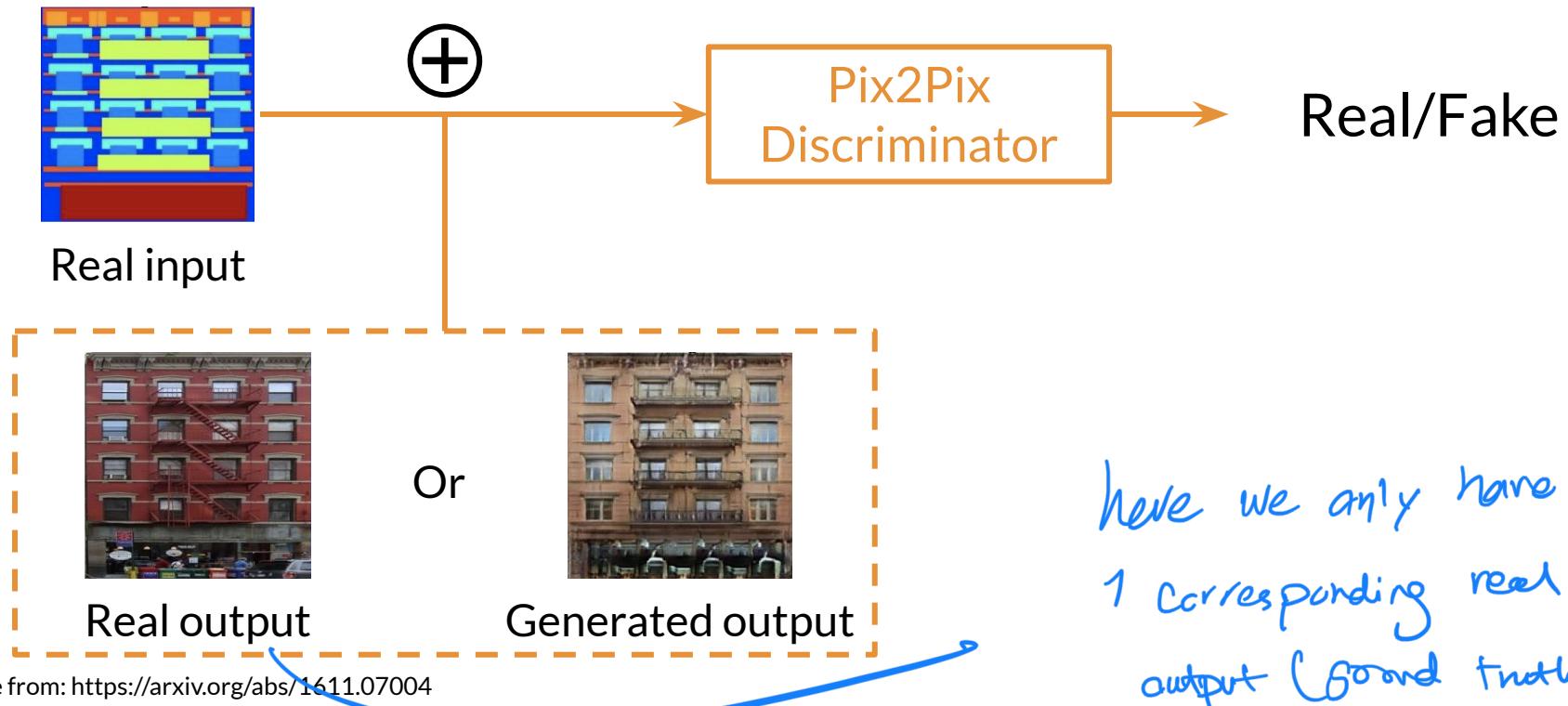
Available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Generator

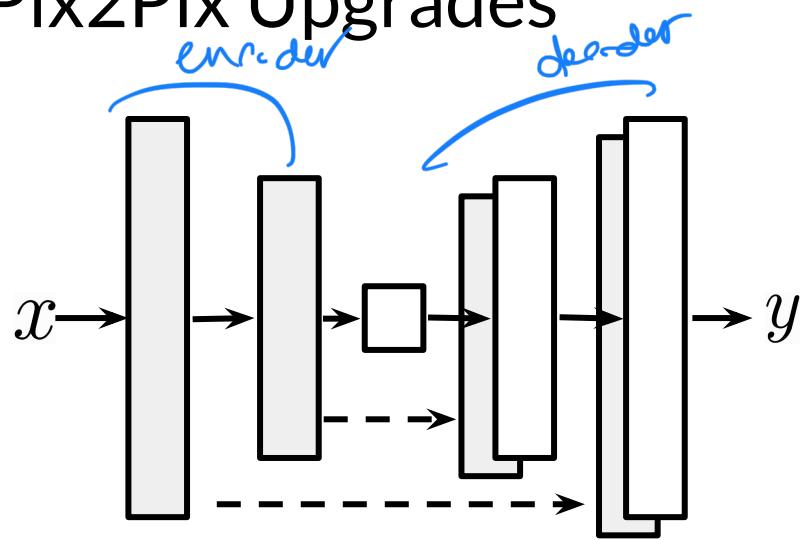


Available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Discriminator



# Pix2Pix Upgrades

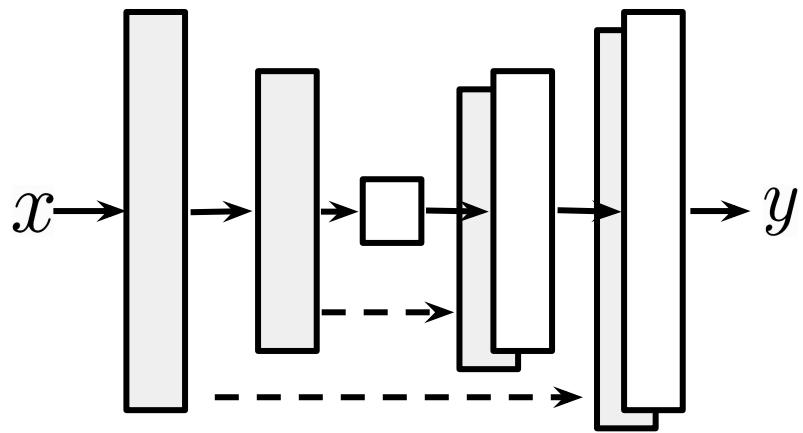


Pix2Pix  
Generator

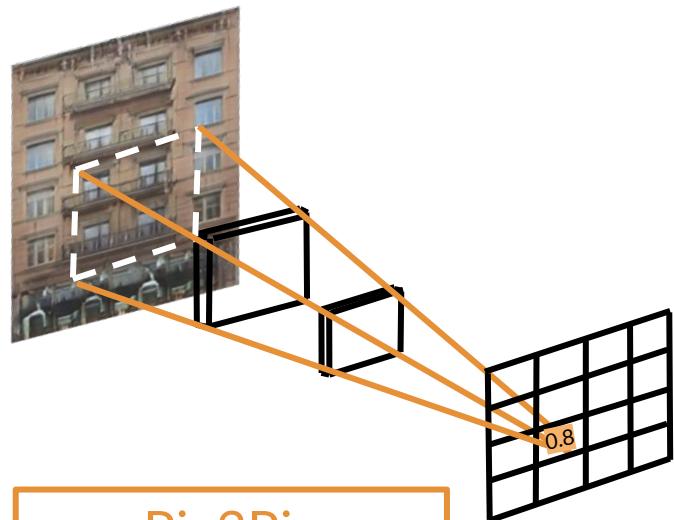
UNet  
for Segmentation

Based on: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Upgrades



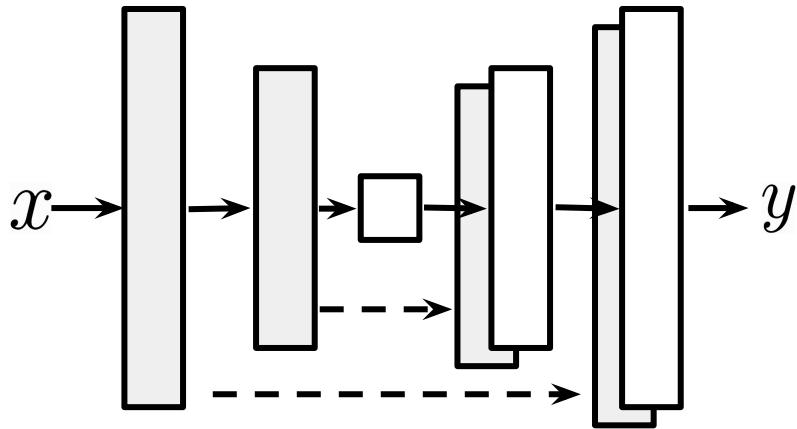
**Pix2Pix  
Generator**



**Pix2Pix  
Discriminator**

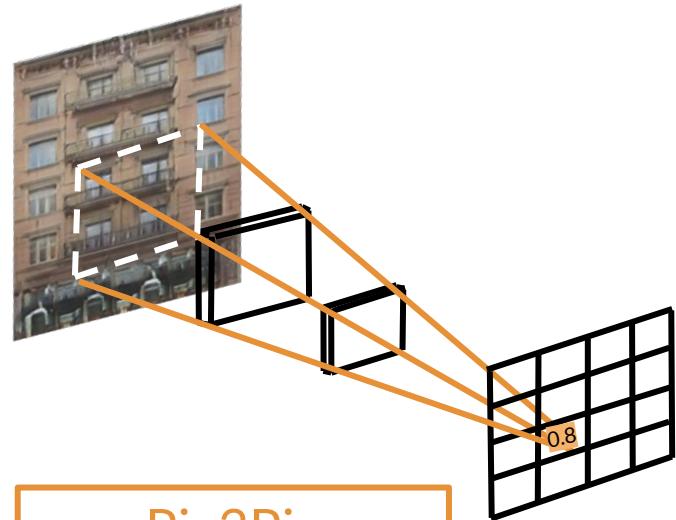
(Left) Based on: <https://arxiv.org/abs/1611.07004>  
(Right) Based on: <https://arxiv.org/abs/1803.07422>

# Pix2Pix Upgrades



Pix2Pix  
Generator

Goal is still to produce realistic outputs!

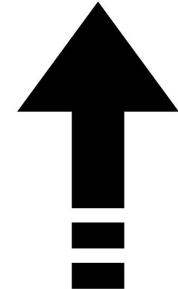


Pix2Pix  
Discriminator

(Left) Based on: <https://arxiv.org/abs/1611.07004>  
(Right) Based on: <https://arxiv.org/abs/1803.07422>

# Summary

- Inputs and outputs of Pix2Pix are similar to a conditional GAN
  - Take in the original image, instead of the class vector
  - No explicit noise as input
- Generator and discriminator models are upgraded





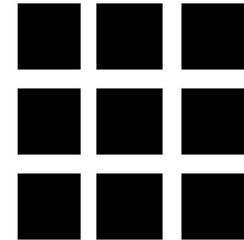
deeplearning.ai

# Pix2Pix: PatchGAN

discrim  
in Pix2Pix

# Outline

- PatchGAN discriminator architecture
- Matrix output vs. single output



# Pix2Pix Discriminator: PatchGAN

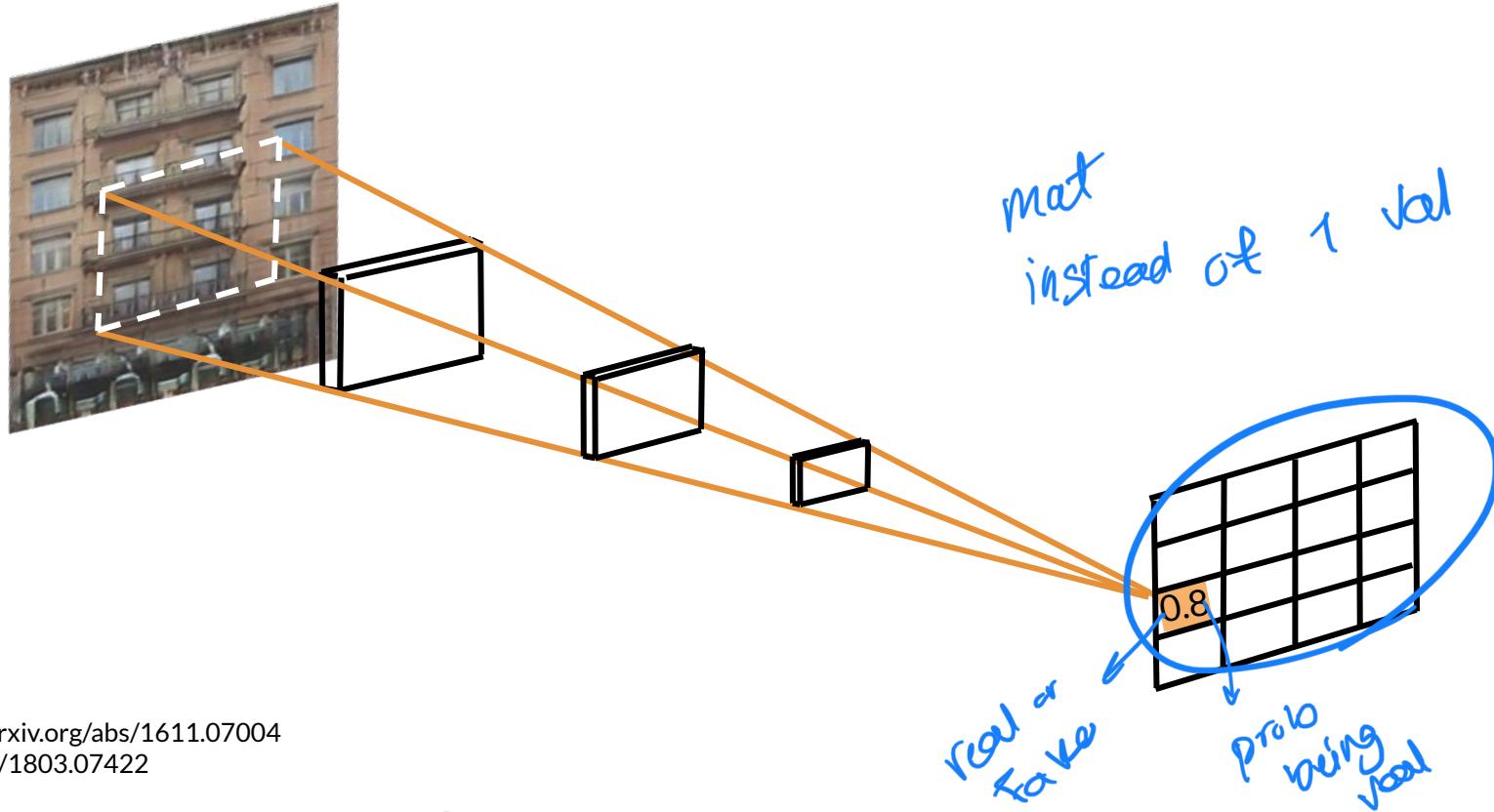


Image available from: <https://arxiv.org/abs/1611.07004>  
Based on: <https://arxiv.org/abs/1803.07422>

# Pix2Pix Discriminator: PatchGAN

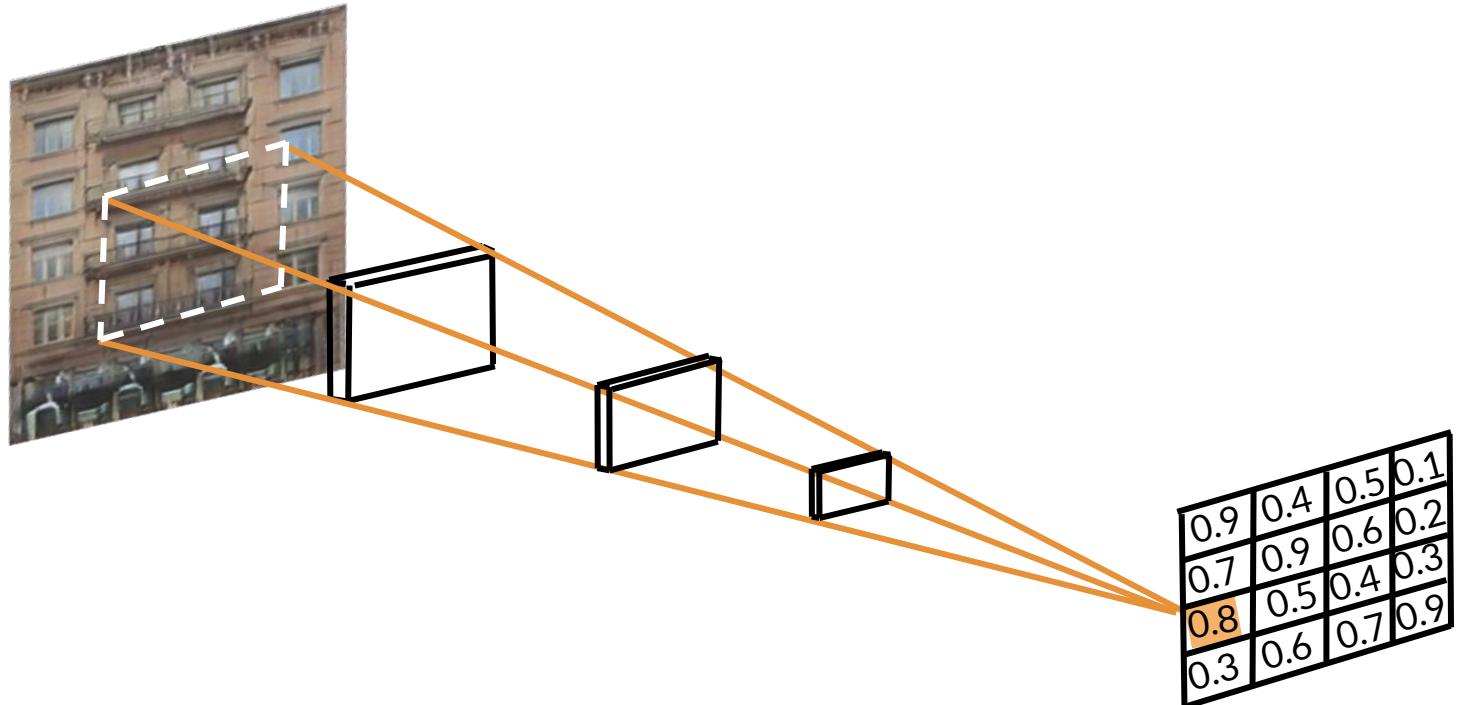


Image available from: <https://arxiv.org/abs/1611.07004>

Based on: <https://arxiv.org/abs/1803.07422>

# Pix2Pix Discriminator: PatchGAN

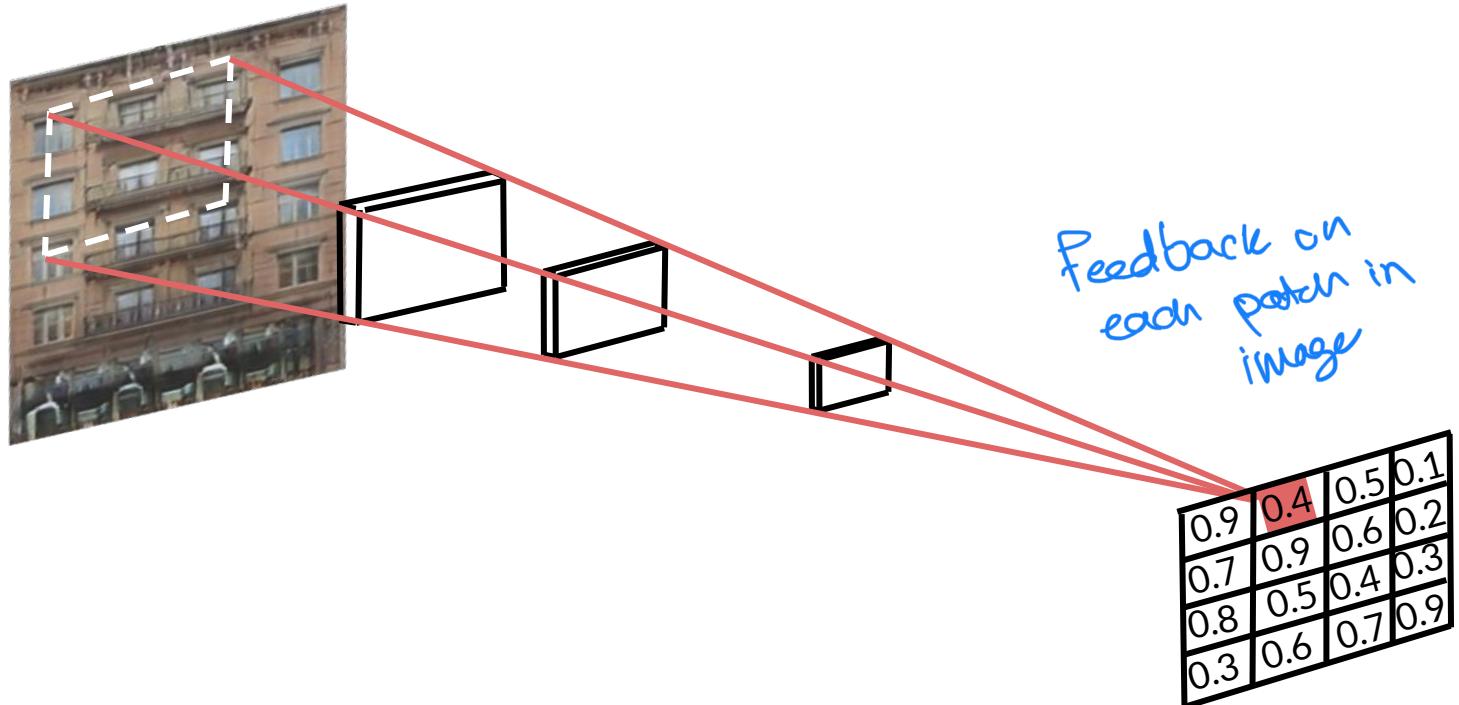


Image available from: <https://arxiv.org/abs/1611.07004>

Based on: <https://arxiv.org/abs/1803.07422>

# Pix2Pix Discriminator: PatchGAN

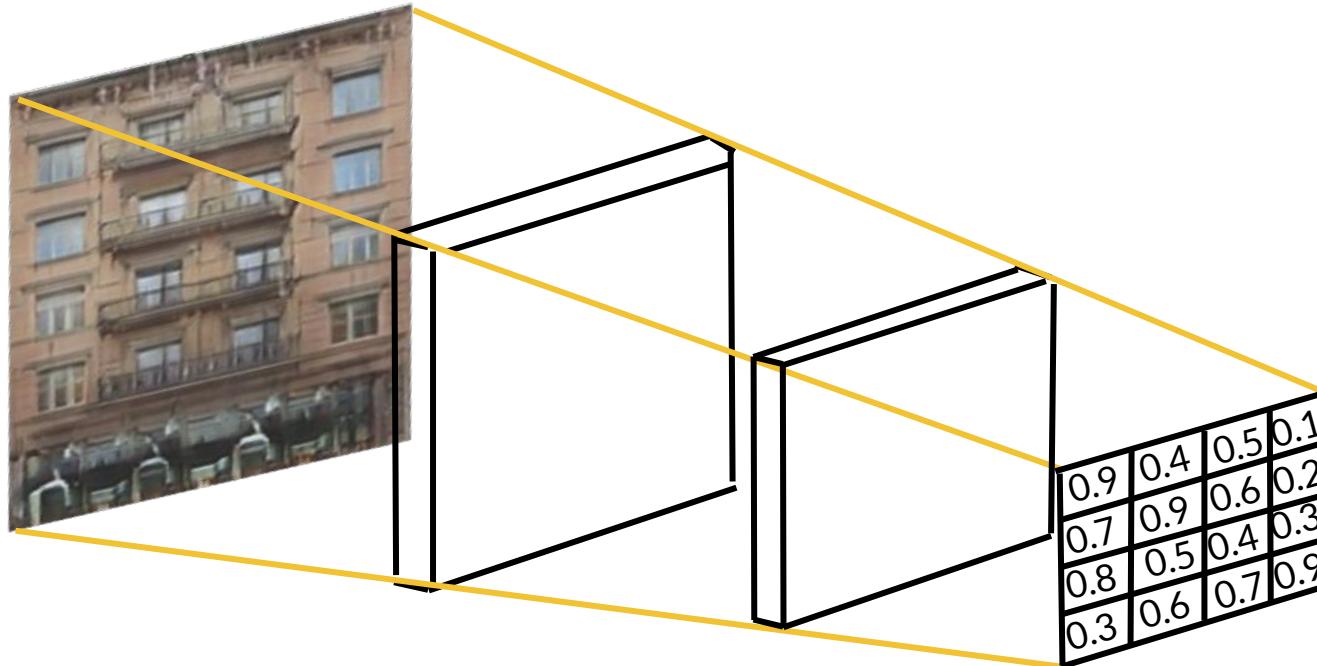
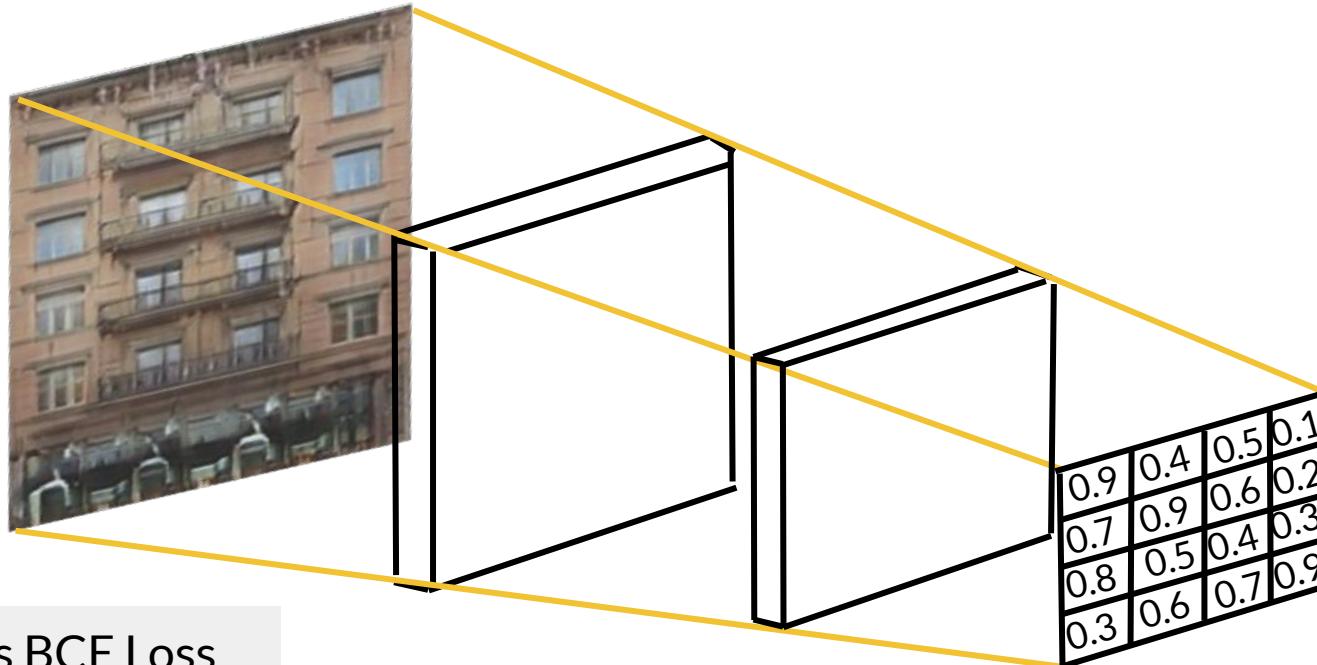


Image available from: <https://arxiv.org/abs/1611.07004>

Based on: <https://arxiv.org/abs/1803.07422>

# Pix2Pix Discriminator: PatchGAN



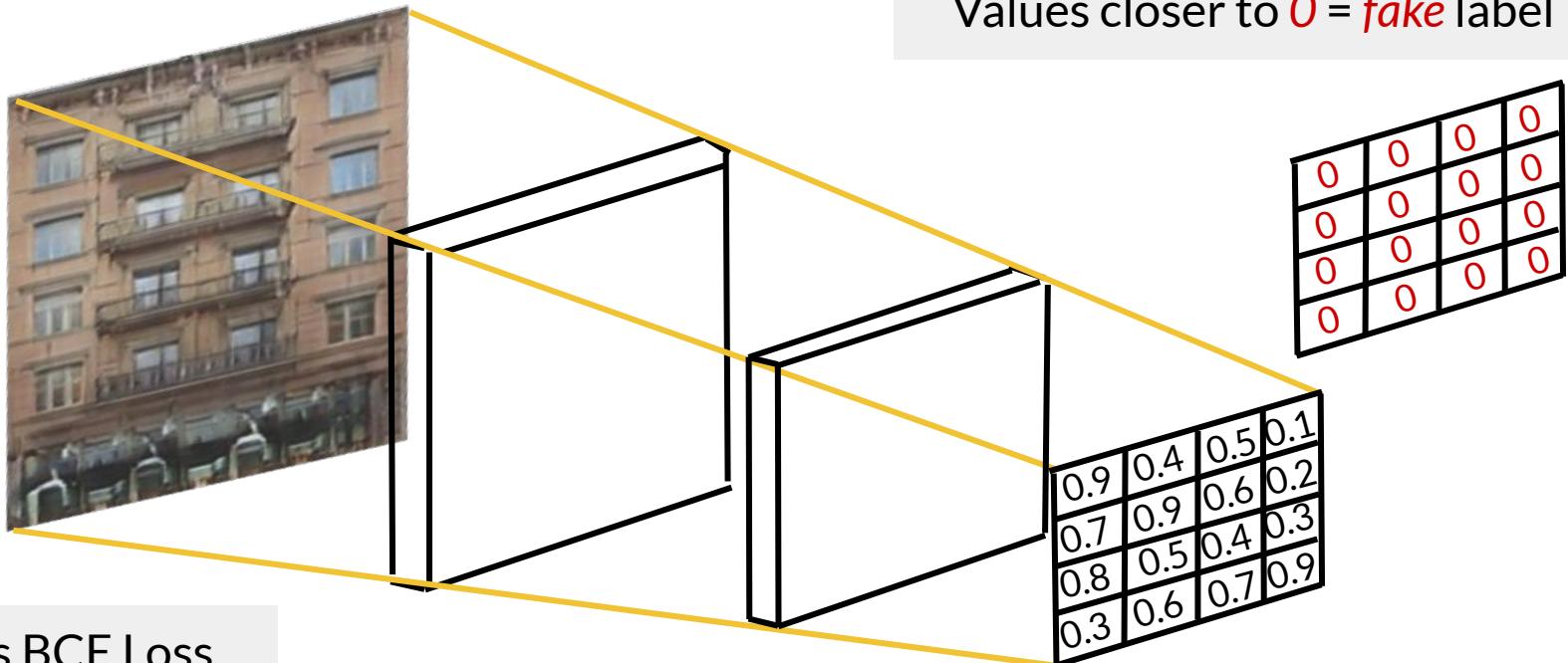
Still uses BCE Loss

Image available from: <https://arxiv.org/abs/1611.07004>

Based on: <https://arxiv.org/abs/1803.07422>

# Pix2Pix Discriminator: PatchGAN

Values closer to  $0 = \text{fake}$  label



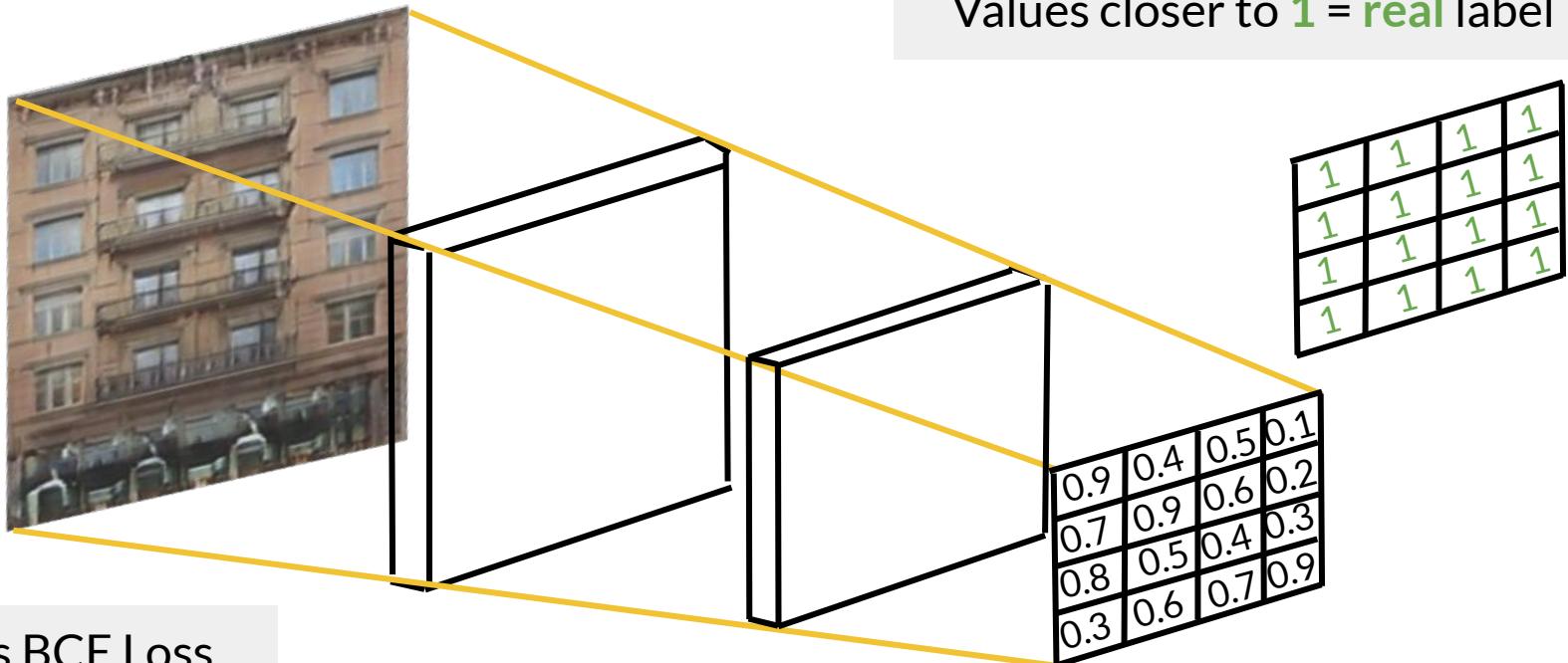
Still uses BCE Loss

Image available from: <https://arxiv.org/abs/1611.07004>

Based on: <https://arxiv.org/abs/1803.07422>

# Pix2Pix Discriminator: PatchGAN

Values closer to **1** = **real** label



Still uses BCE Loss

Image available from: <https://arxiv.org/abs/1611.07004>

Based on: <https://arxiv.org/abs/1803.07422>

# Summary

- PatchGAN discriminator outputs a matrix of values, each between 0 and 1
- Label matrices:
  - 0's = fake
  - 1's = real



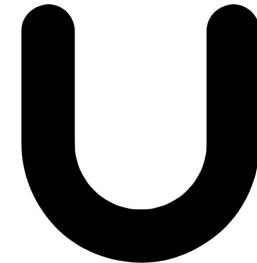


deeplearning.ai

# Pix2Pix: U-Net

# Outline

- Net framework
  - Encoder-Decoder
- U-Skip connections
- Pix2Pix generator



# Image Segmentation

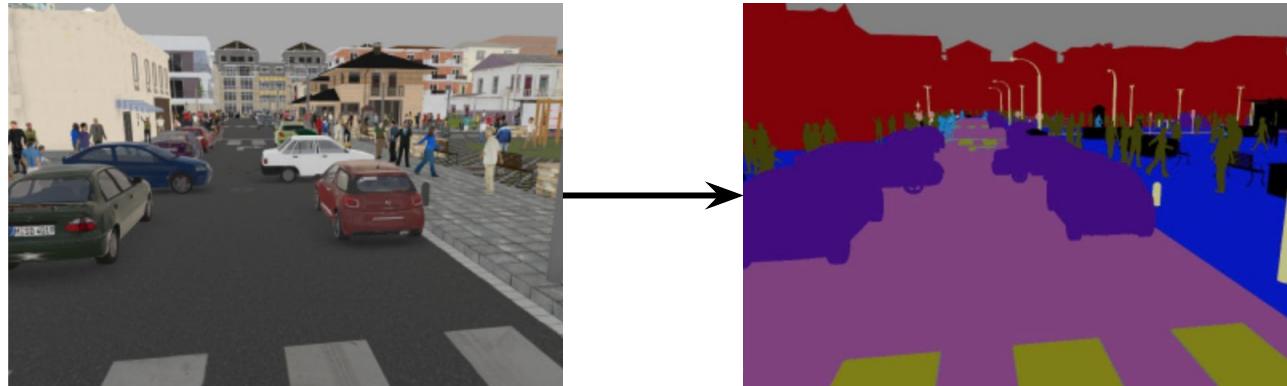


Image Segmentation

Available from: <https://developer.nvidia.com/blog/image-segmentation-using-digits-5/>

# Image Segmentation

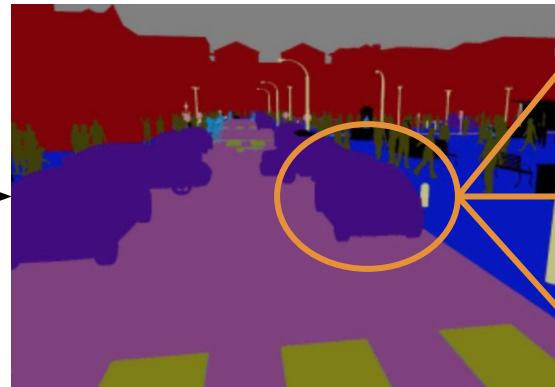
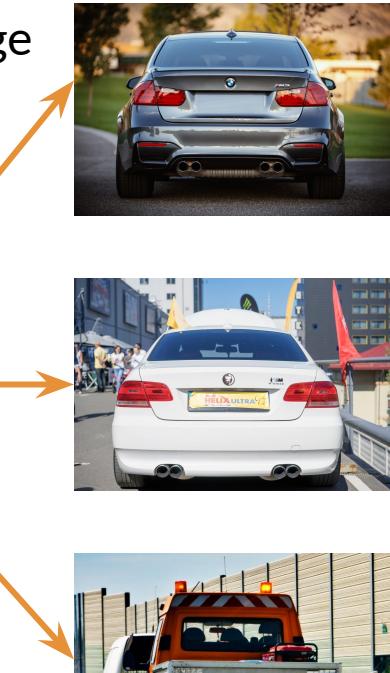


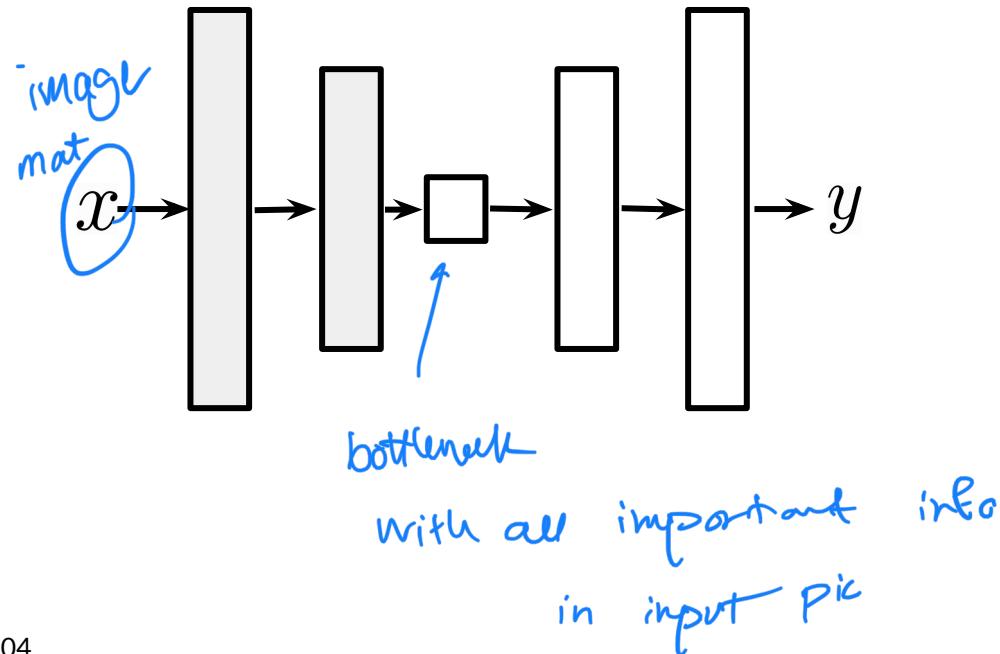
Image Segmentation

## Image-to-Image Translation



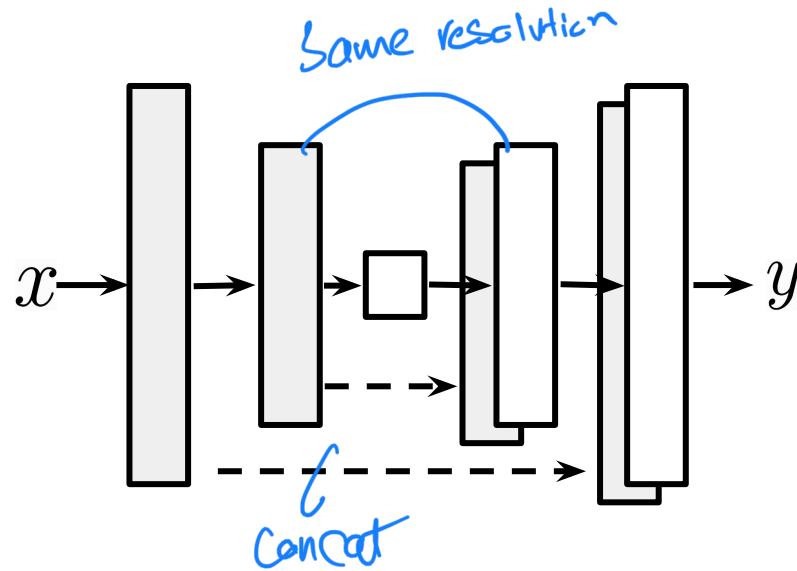
Available from: <https://developer.nvidia.com/blog/image-segmentation-using-digits-5/>

# U-Net Framework: Encoder-Decoder



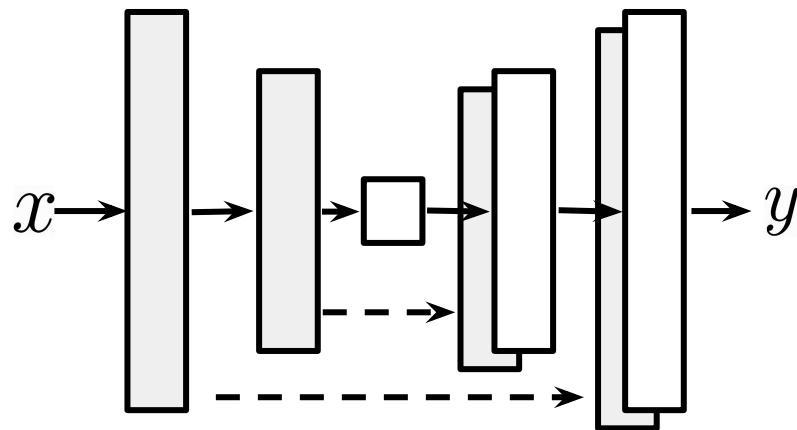
Based on: <https://arxiv.org/abs/1611.07004>

# U-Net Framework: Skip Connections



Based on: <https://arxiv.org/abs/1611.07004>

# U-Net Framework: Skip Connections



Forward pass

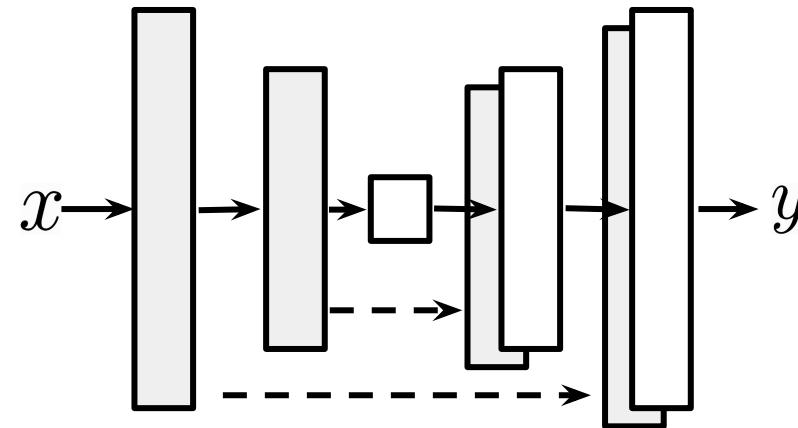
**Skip connections**  
allow information  
flow to the decoder

Based on: <https://arxiv.org/abs/1611.07004>

# U-Net Framework: Skip Connections

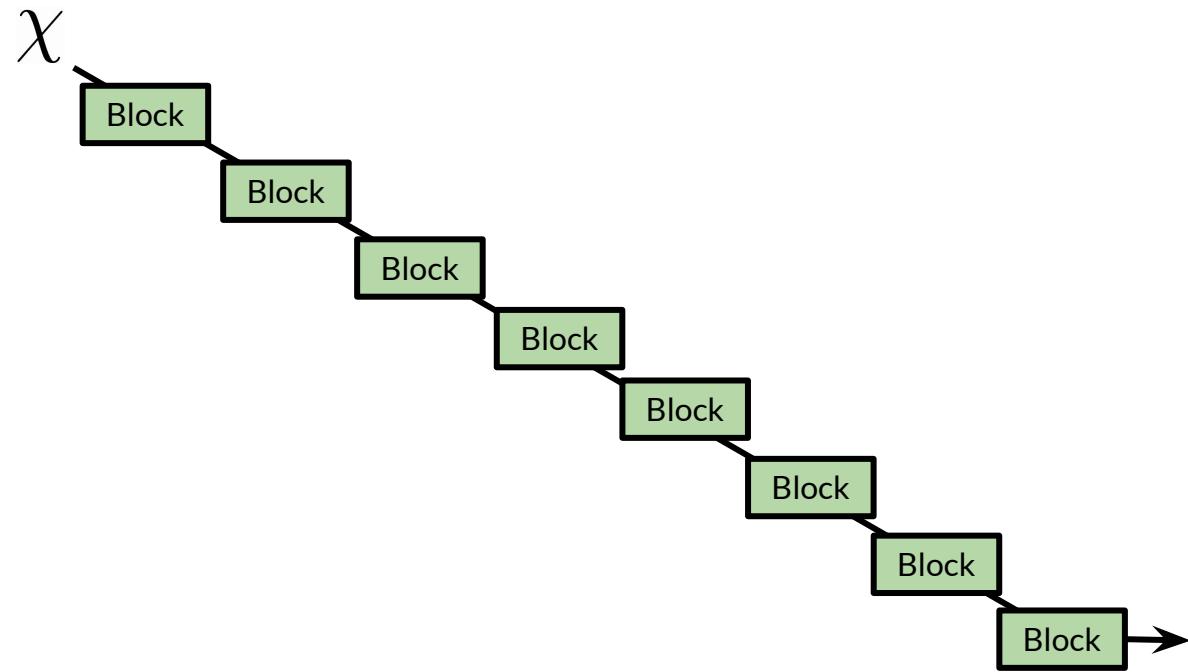
Backward pass

Skip connections  
improve gradient  
flow to encoder



Based on: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Encoder



# Pix2Pix Encoder

$\chi$

Input size: 256 x 256 x 3

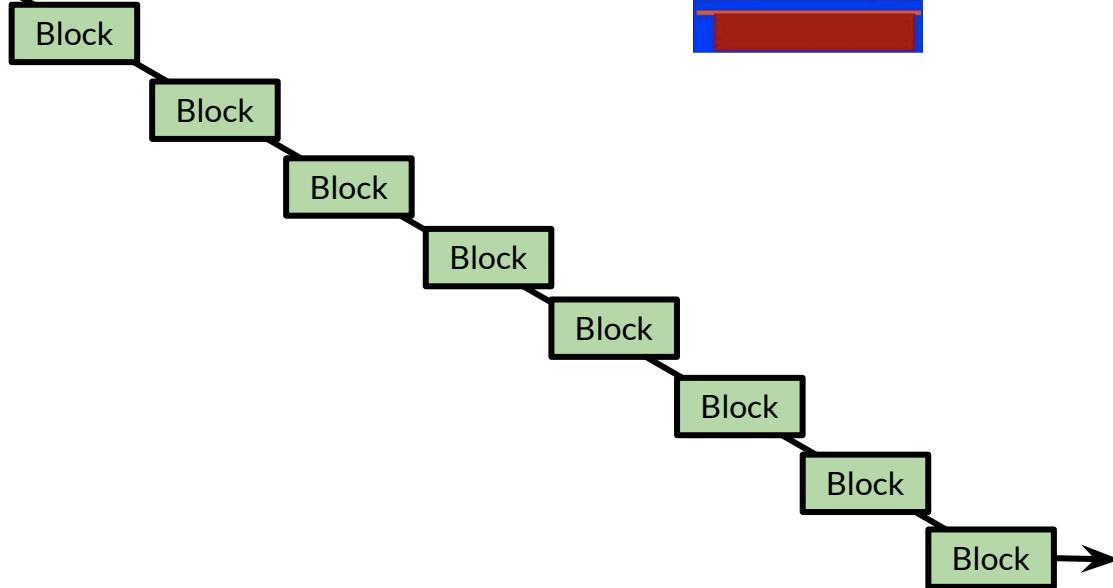
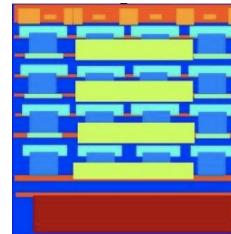
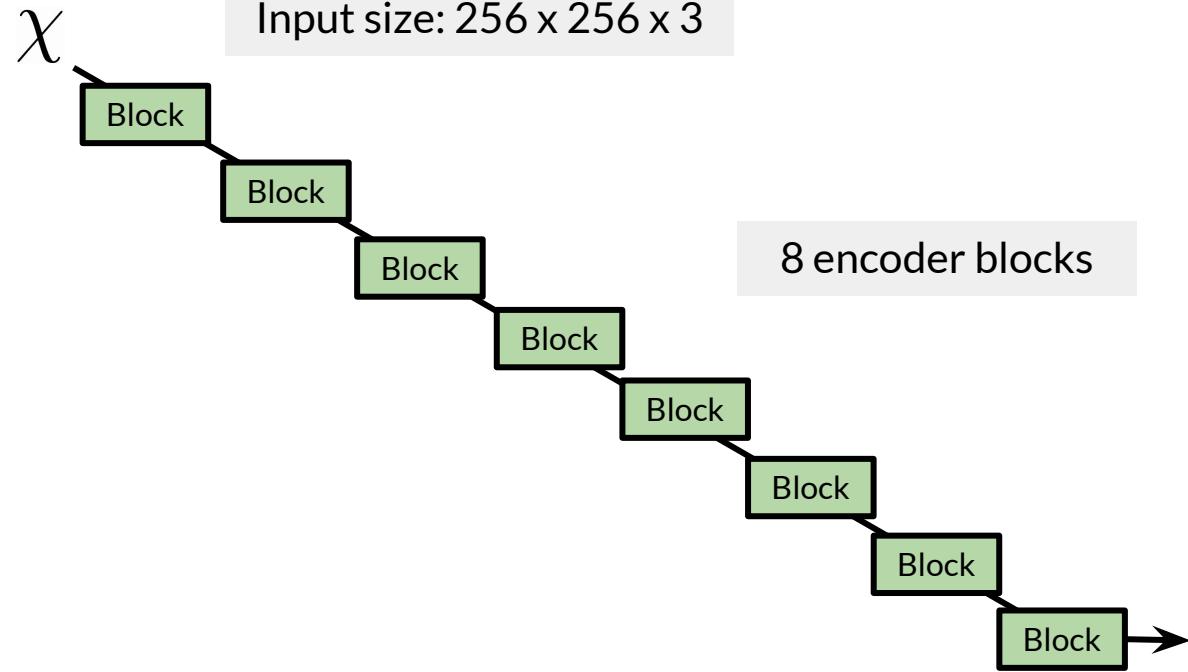


Image available from: <https://arxiv.org/abs/1611.07004>

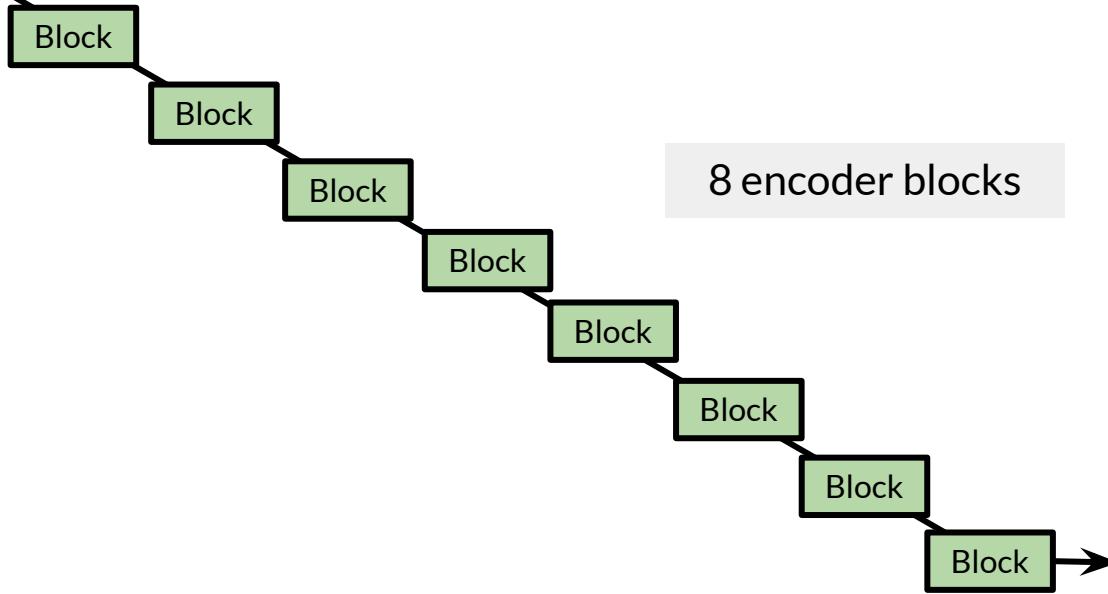
# Pix2Pix Encoder



# Pix2Pix Encoder

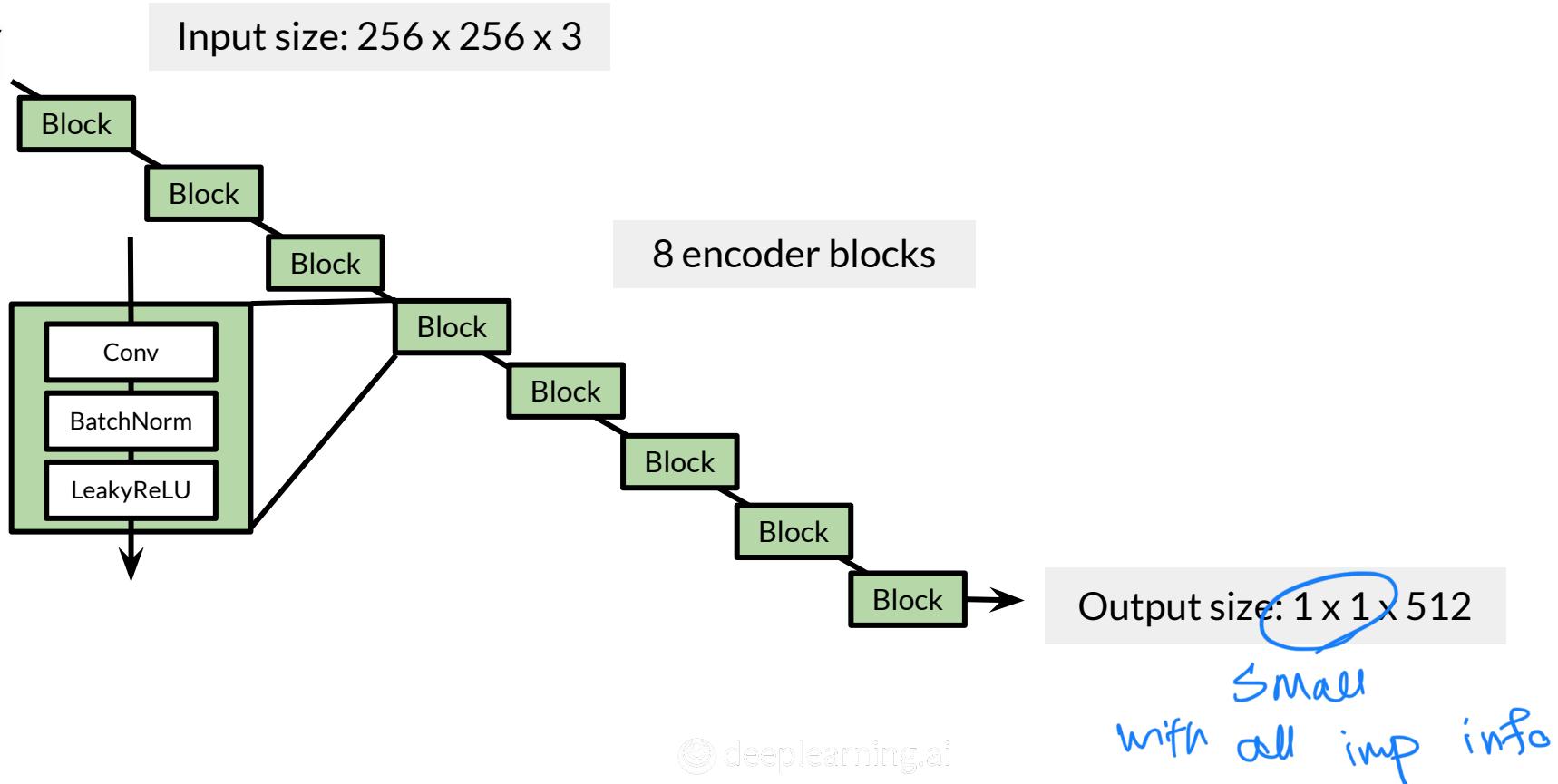
$\chi$

Input size: 256 x 256 x 3

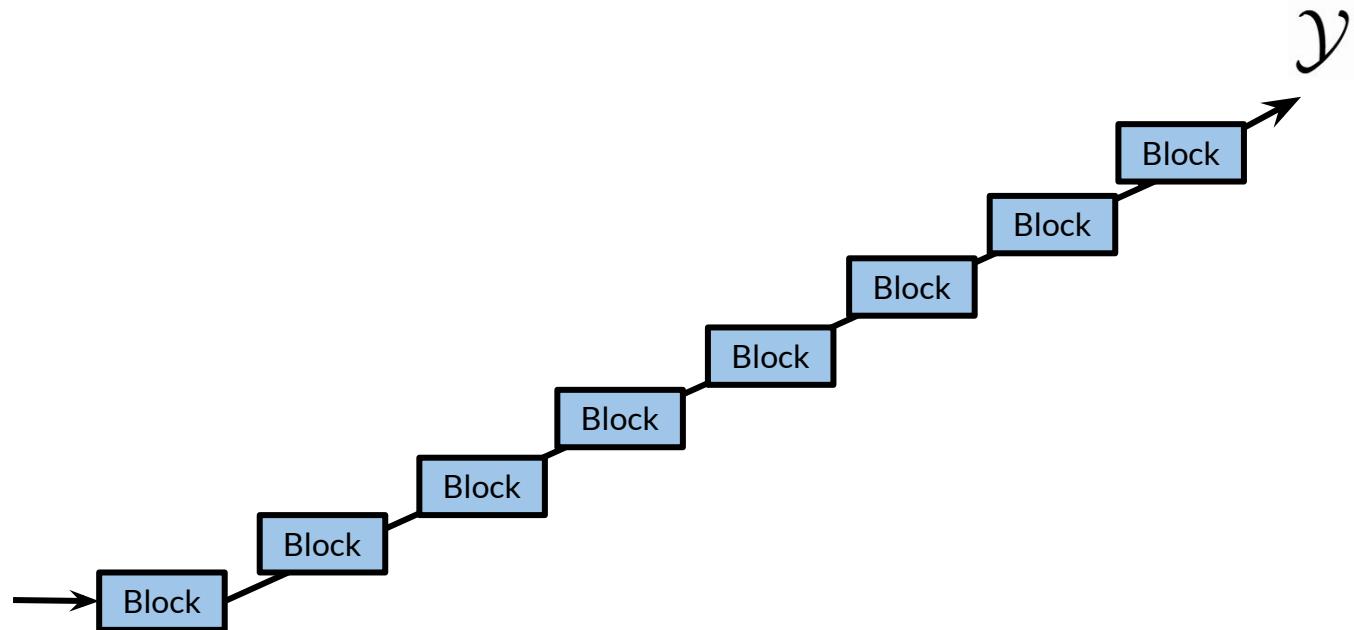


# Pix2Pix Encoder

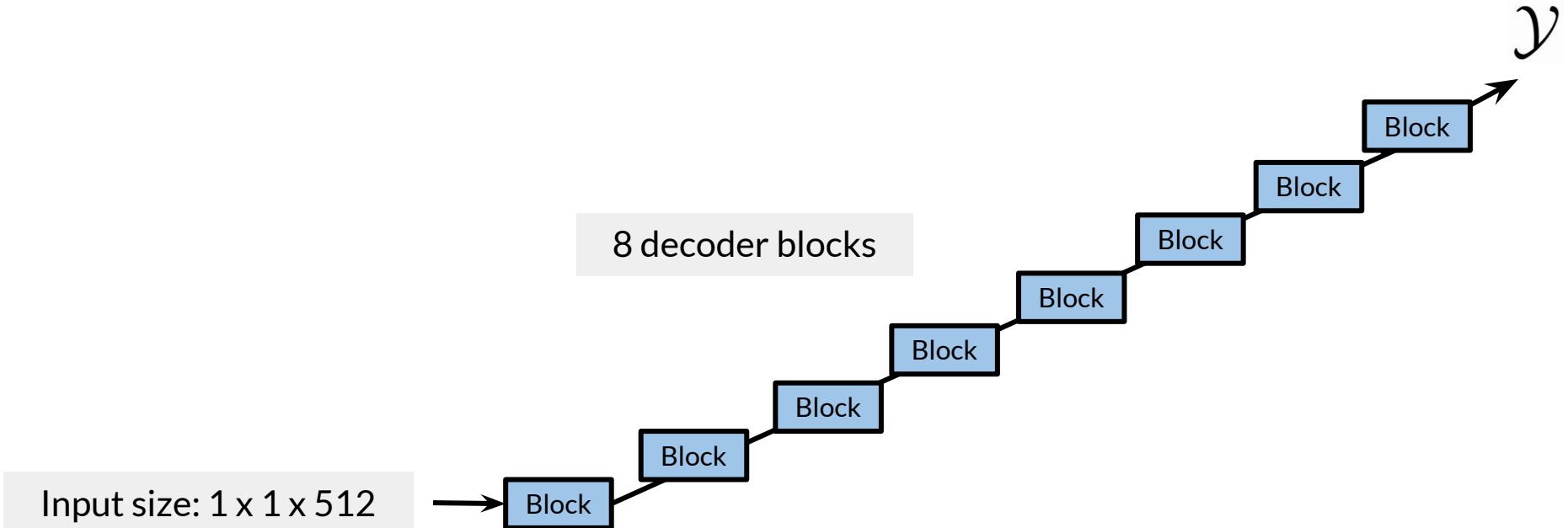
$\chi$



# Pix2Pix Decoder



# Pix2Pix Decoder

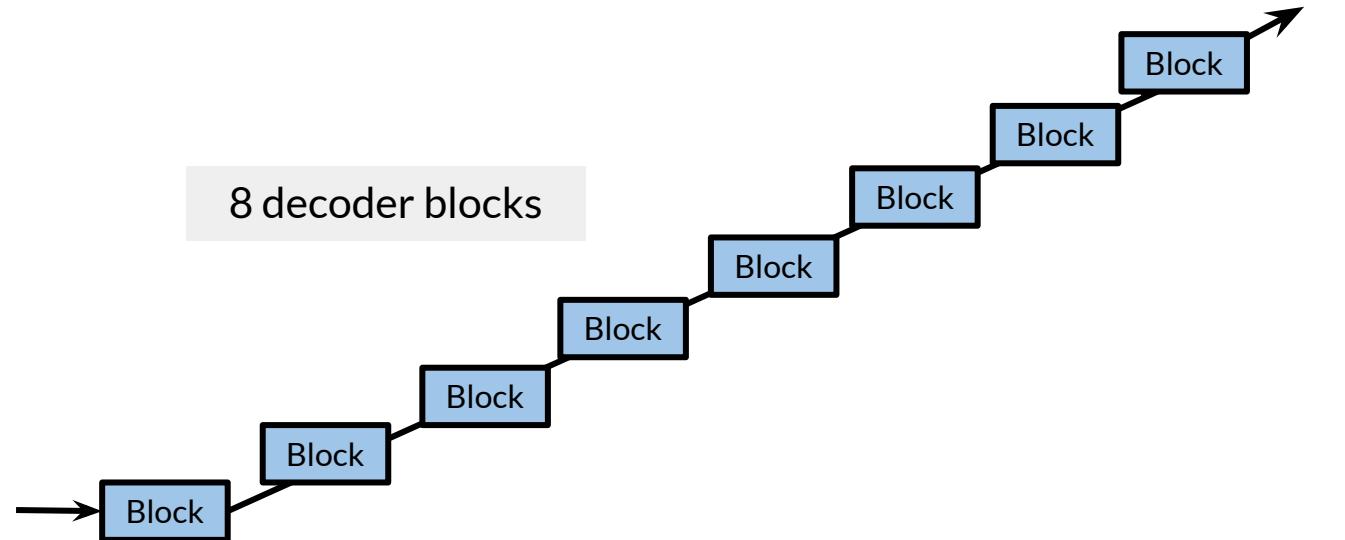


# Pix2Pix Decoder

Output size: 256 x 256 x 3



$y$



Input size:  $1 \times 1 \times 512$

Image available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Decoder

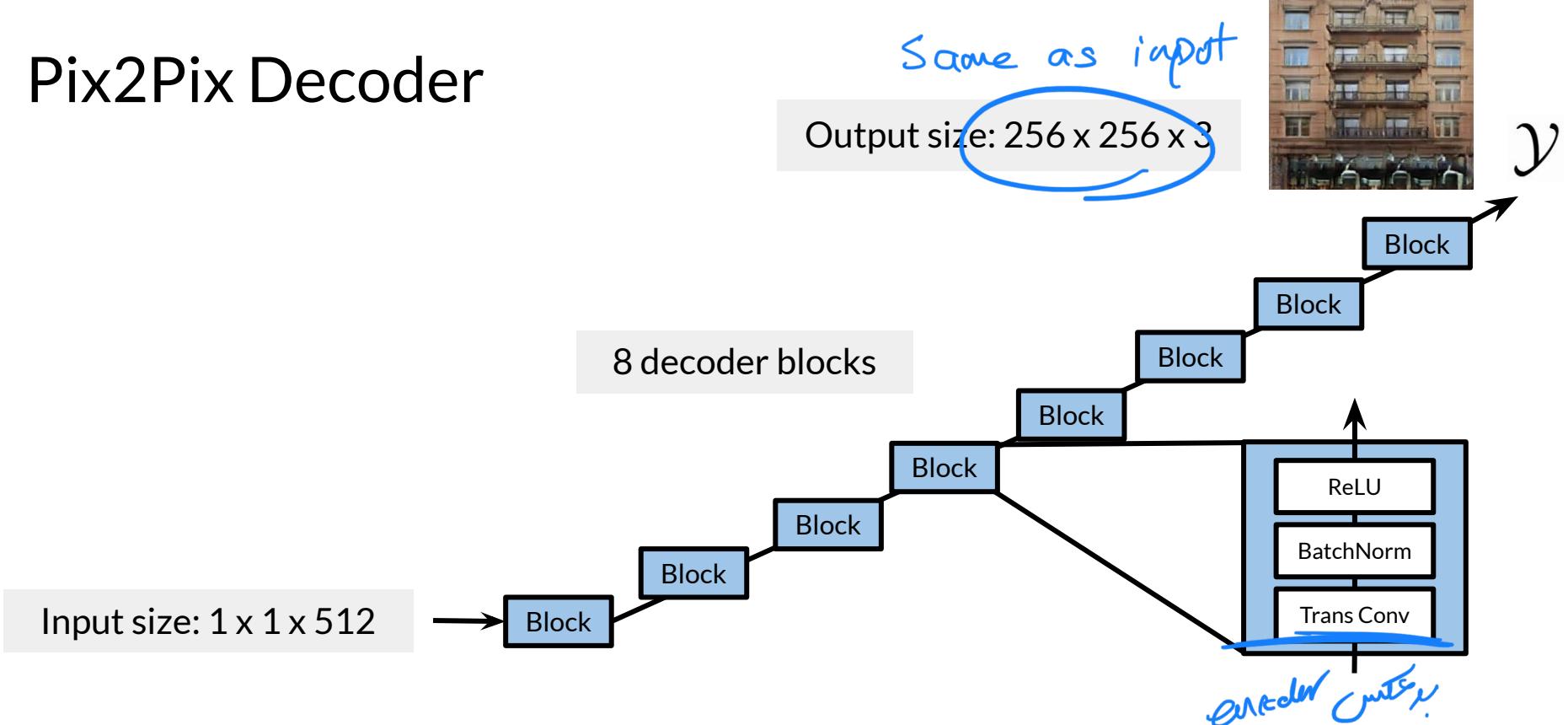


Image available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Decoder

Output size: 256 x 256 x 3



$y$

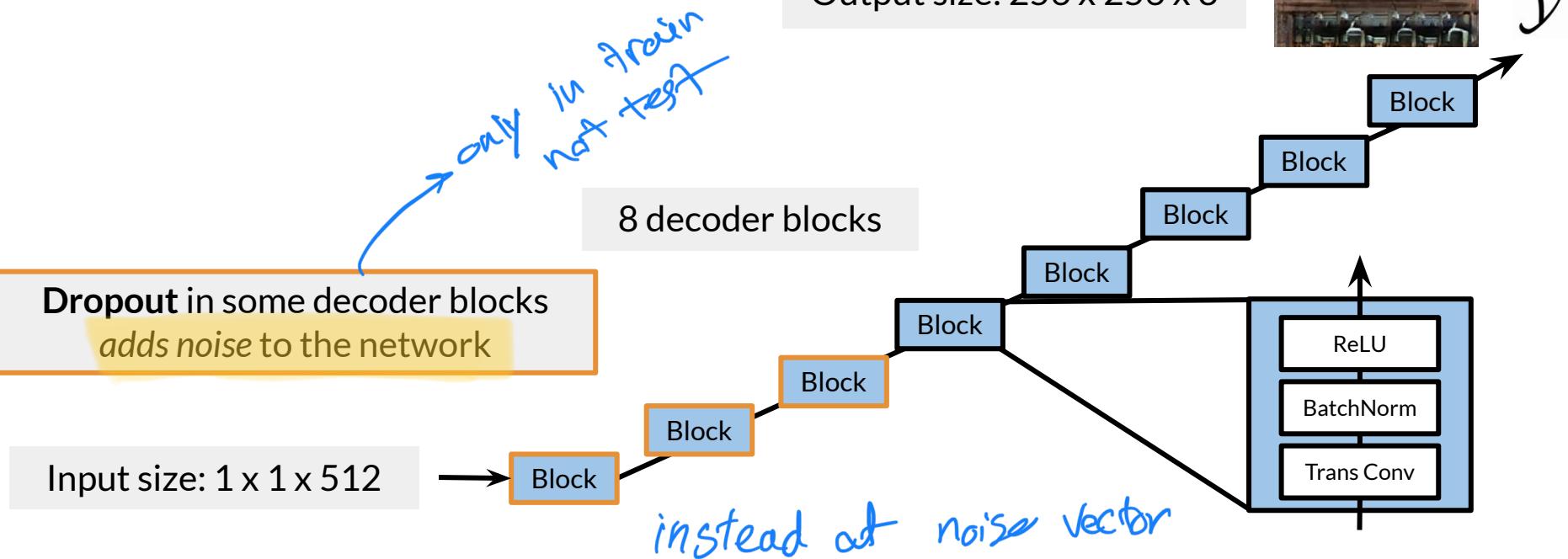
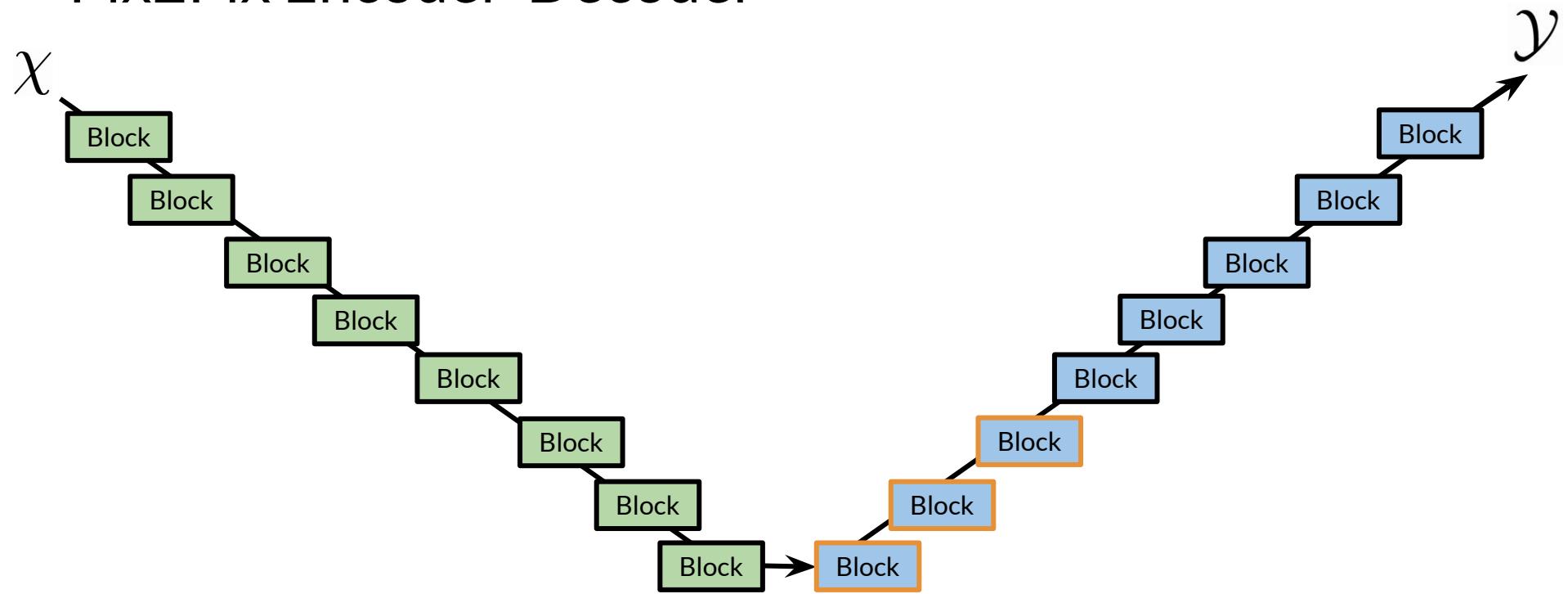
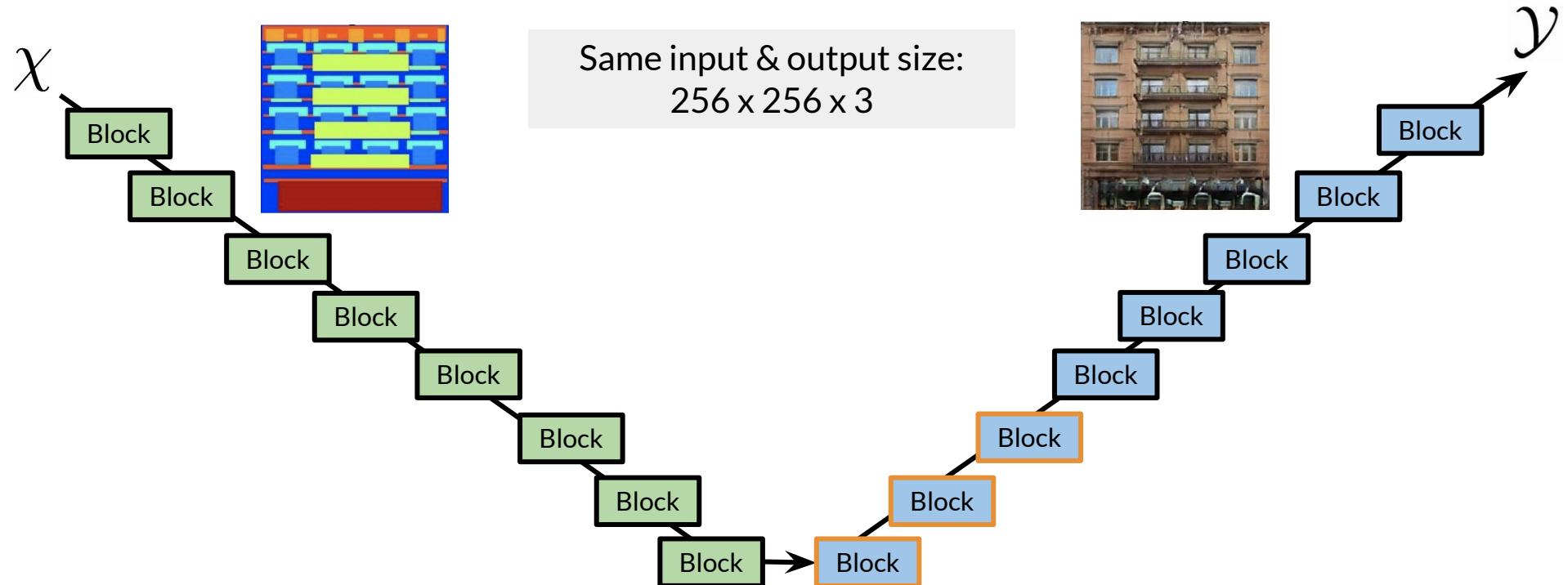


Image available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Encoder-Decoder

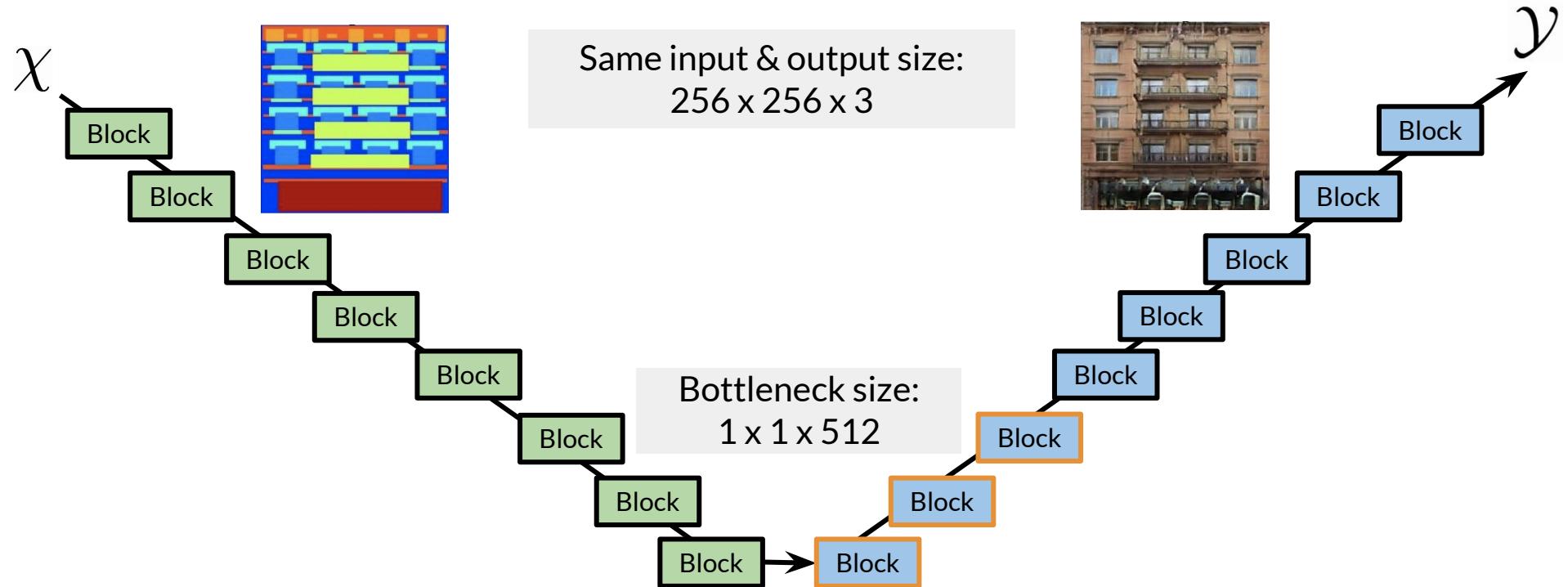


# Pix2Pix Encoder-Decoder



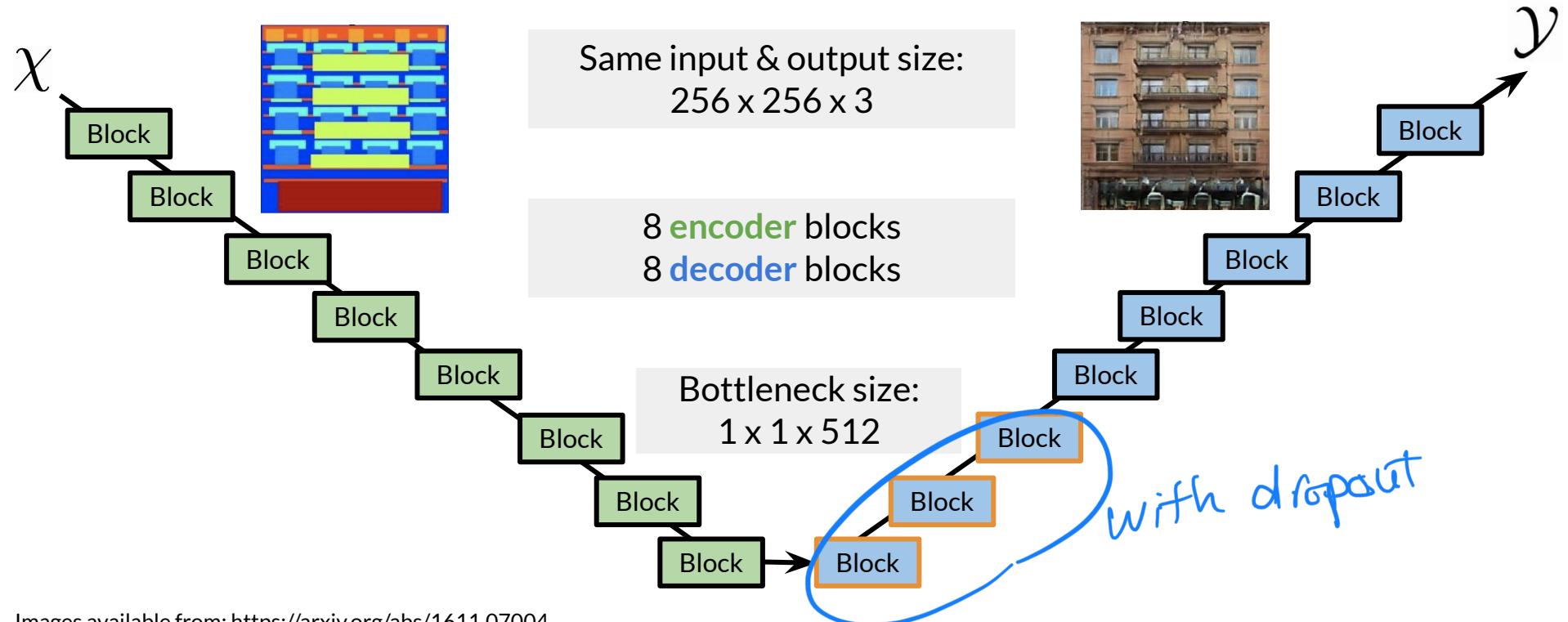
Images available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Encoder-Decoder



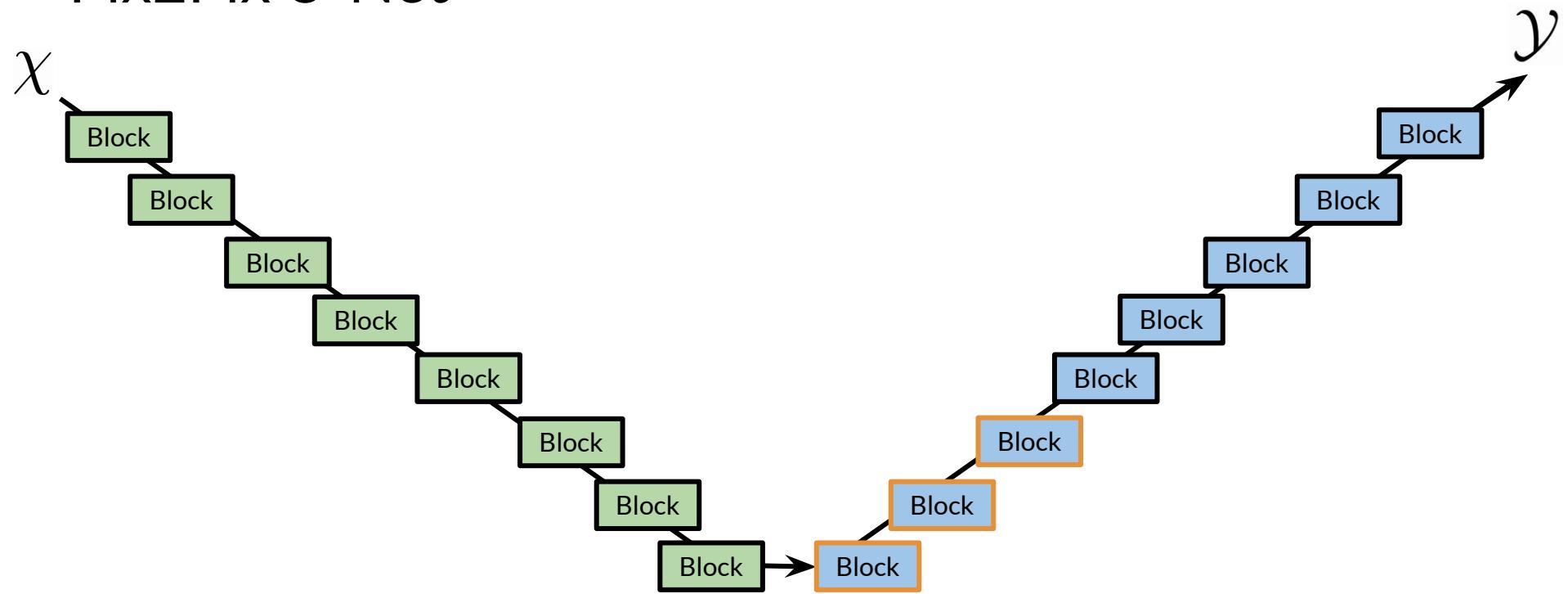
Images available from: <https://arxiv.org/abs/1611.07004>

## Pix2Pix Encoder-Decoder

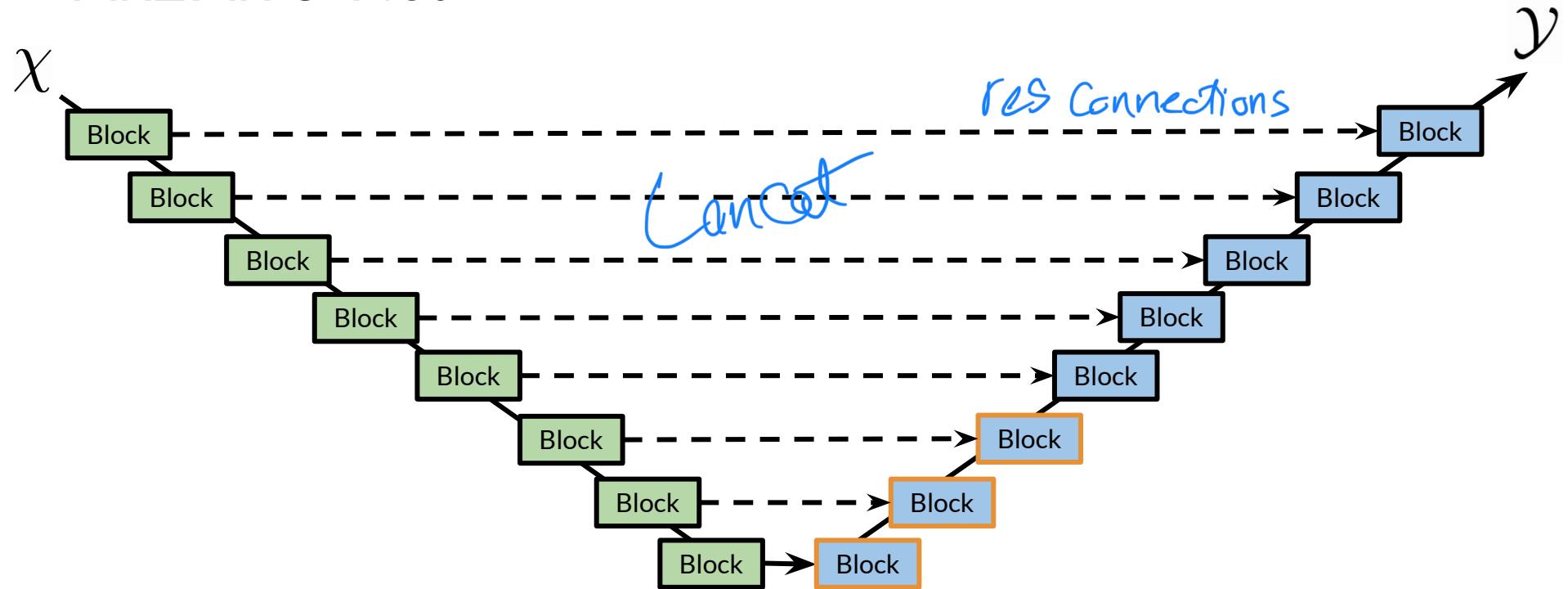


Images available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix U-Net

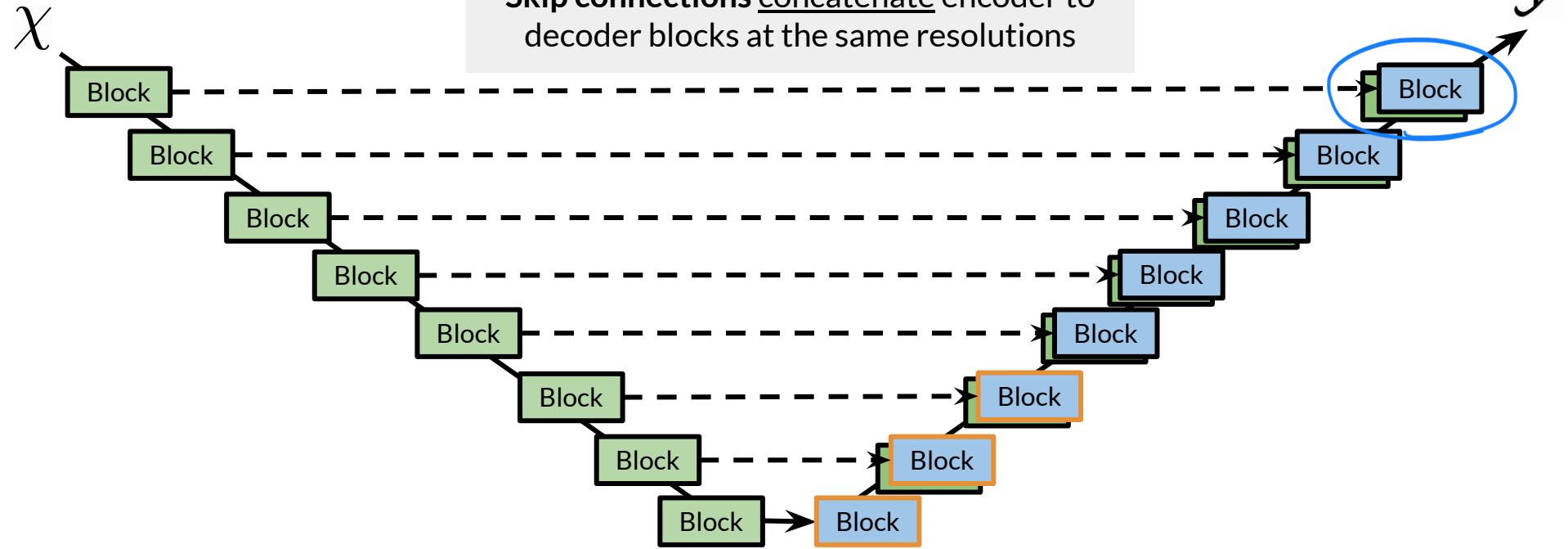


# Pix2Pix U-Net



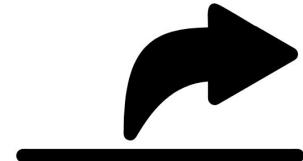
# Pix2Pix U-Net

Skip connections concatenate encoder to decoder blocks at the same resolutions



# Summary

- Pix2Pix's generator is a U-Net
- U-Net is an encoder-decoder, with same-size inputs and outputs
- U-Net uses skip connections
  - Skip connections help the decoder learn details from the encoder directly
  - Skip connections allow the encoder to learn from more gradients flowing from the decoder



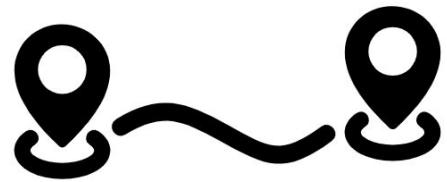


deeplearning.ai

# Pix2Pix: Pixel Distance Loss Term

# Outline

- Regularization and additional loss term
- Encourage pixel distance between generated and real outputs
- Additional loss term for Pix2Pix generator



# Additional Loss Term

$$\min_g \max_c \text{Adversarial Loss} + \underbrace{\lambda * \text{Other loss term}}_{\text{reg or penalty}}$$

main loss function

# Additional Loss Term

$$\min_g \max_c \text{Adversarial Loss} + \lambda * \text{Pixel loss term}$$

*to glue gener more info  
about real image to  
match a better output*

# Pixel Distance Loss Term

$$\sum_{i=1}^n$$



Generated output



Real output

Encourage generated output to be as close as possible to real output

Available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix Generator Loss

$$\text{BCE Loss} + \lambda \sum_{i=1}^n \left| \begin{array}{c} \text{Pixel loss term} \\ \text{Image A} - \text{Image B} \end{array} \right|$$

The equation illustrates the Pix2Pix Generator Loss function. It consists of two main components: a BCE Loss term and a weighted sum of pixel loss terms. The pixel loss term is represented by a vertical bar containing two images: the original image (Image A) and the generated image (Image B), separated by a minus sign. Above this bar, the text "Pixel loss term" is written in blue. The entire equation is enclosed in a purple rectangular border.

Available from: <https://arxiv.org/abs/1611.07004>

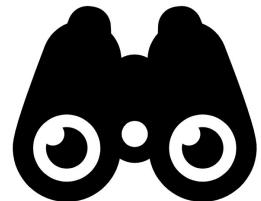
# Pix2Pix Generator Loss

$$\text{BCE Loss} + \lambda \left( \sum_{i=1}^n \left| \text{generated\_output} - \text{real\_output} \right| \right)$$

multiple samples

# Summary

- Pix2Pix adds a Pixel Distance Loss term to the generator loss function
- This loss term calculates the difference between the fake and the real target outputs
- Softly encourages the generator with this additional supervision
  - The target output labels are the supervision
  - Generator essentially “sees” these labels



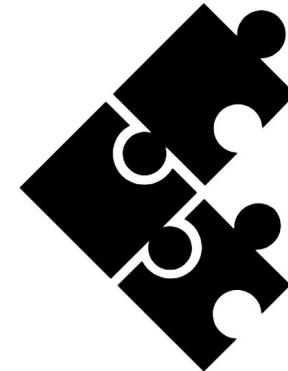


deeplearning.ai

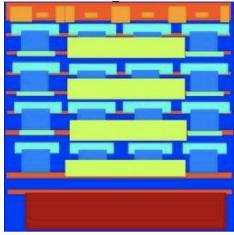
# Pix2Pix: Putting It All Together

# Outline

- Put the Pix2Pix architecture together!
  - U-Net generator
  - Pixel Distance Loss term
  - PatchGAN discriminator



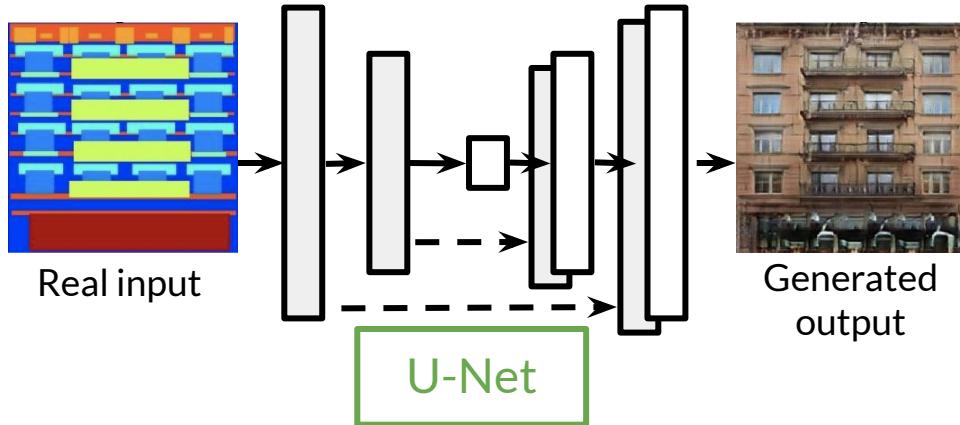
# Pix2Pix



Real input

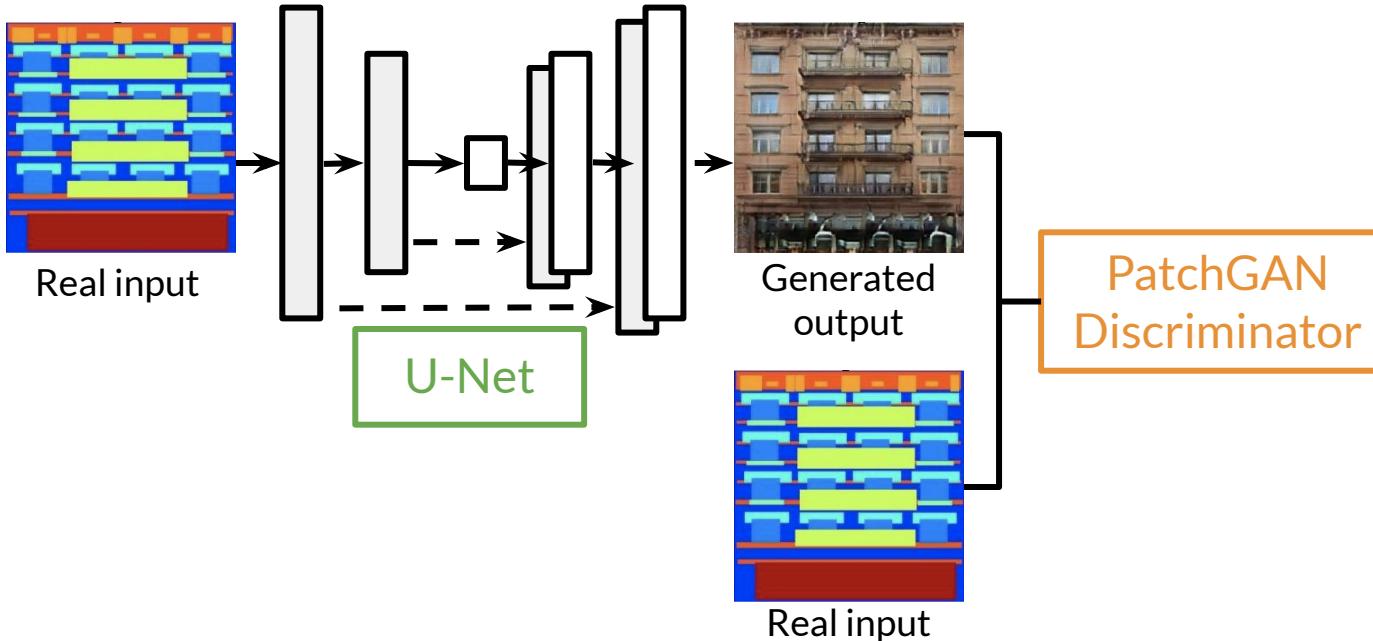
Image available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix



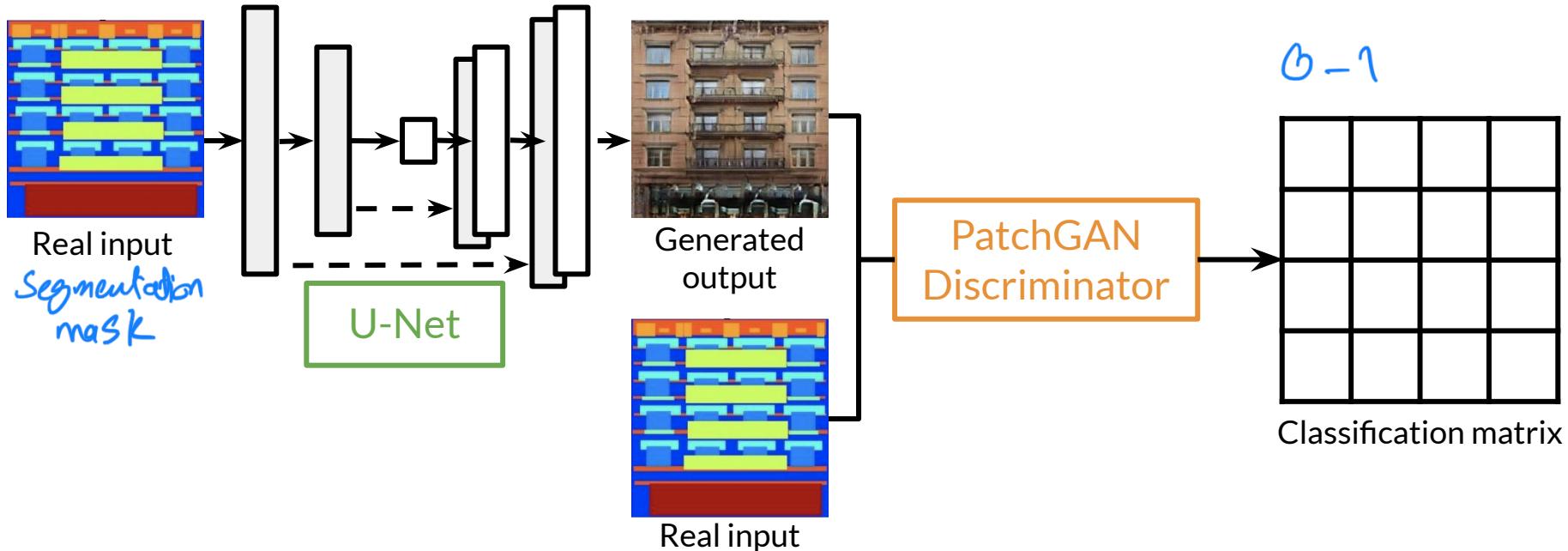
Images available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix



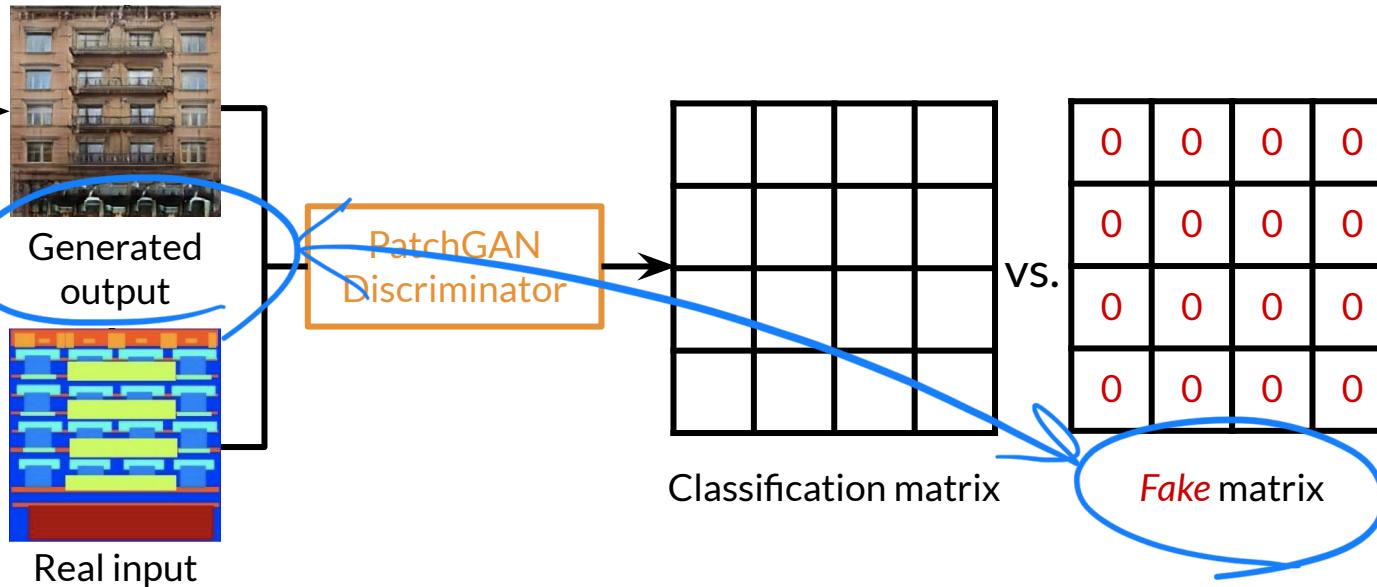
Images available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix



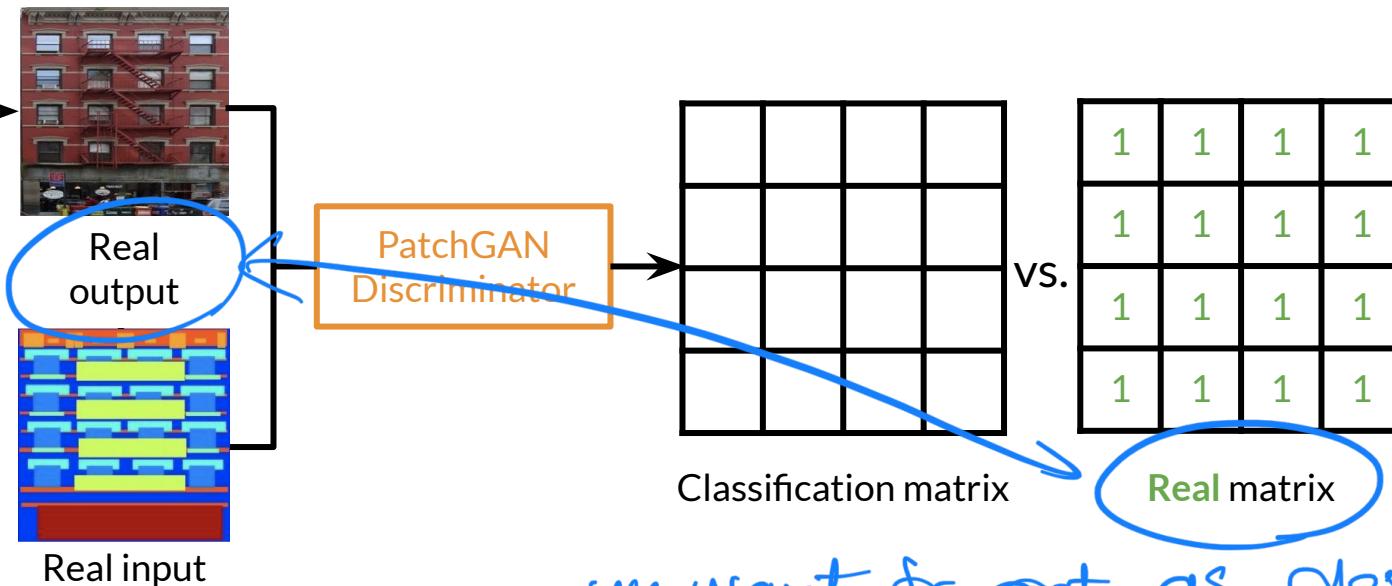
Images available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix: Discriminator Loss



Images available from: <https://arxiv.org/abs/1611.07004>

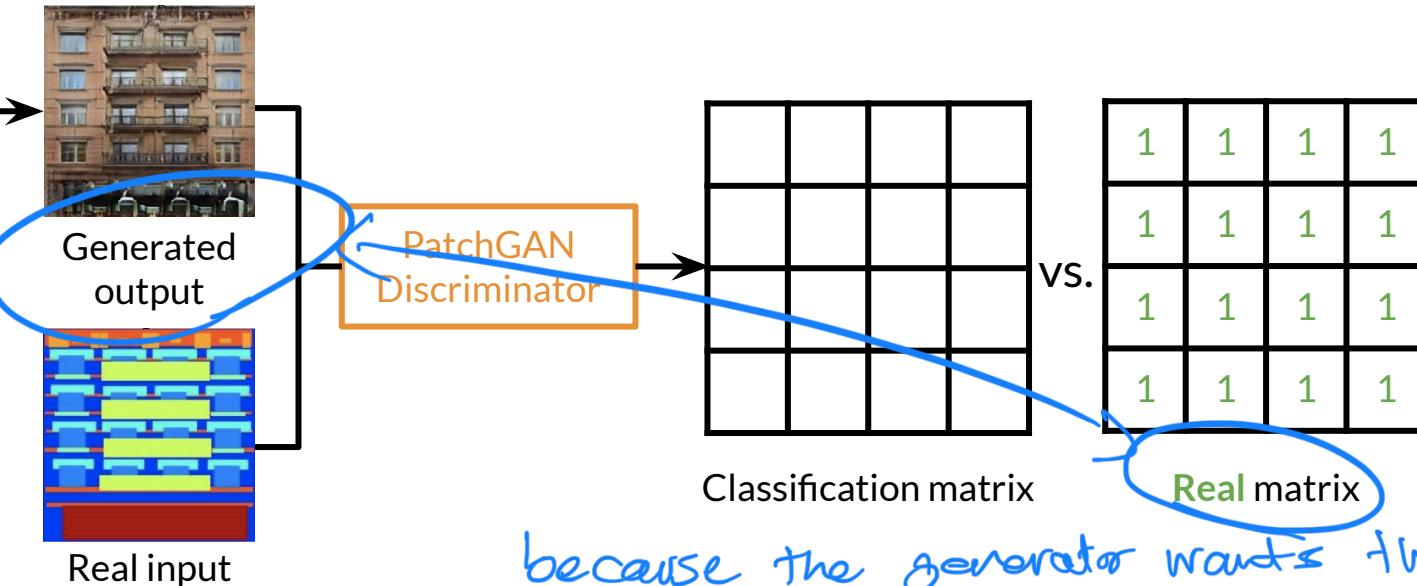
# Pix2Pix: Discriminator Loss



We want to get as close as possible to real image

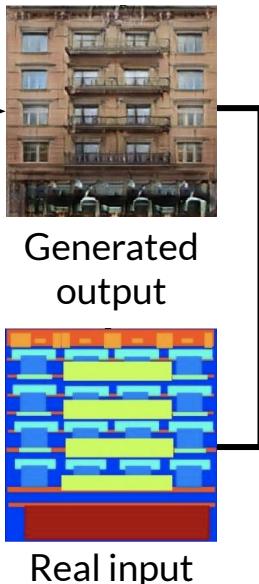
Images available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix: Generator Loss

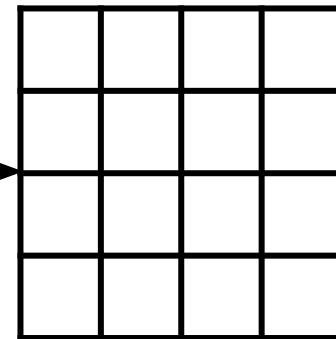


Images available from: <https://arxiv.org/abs/1611.07004>

# Pix2Pix: Generator Loss



PatchGAN  
Discriminator



vs.

1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

Real matrix

$$\lambda^*$$



Generated output



Real output

+  
Pixel Distance  
Loss Term

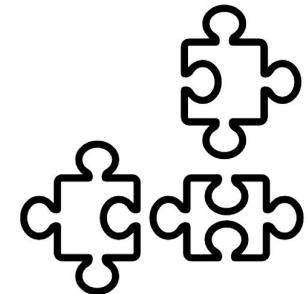
in generator  
loss only

not in discrim

Images available from: <https://arxiv.org/abs/1611.07004>

# Summary

- U-Net generator: image → image
- PatchGAN discriminator
  - Inputs input image and paired output (either real target or fake)
  - Outputs classification matrix
- Generator loss has a regularization term



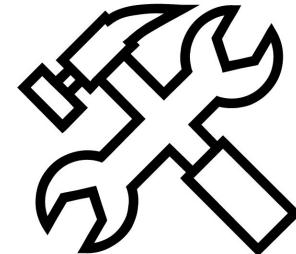


deeplearning.ai

# Pix2Pix Advancements

# Outline

- Improvements and extensions of Pix2Pix for paired image-to-image translation
  - Higher resolution images
  - Image editing

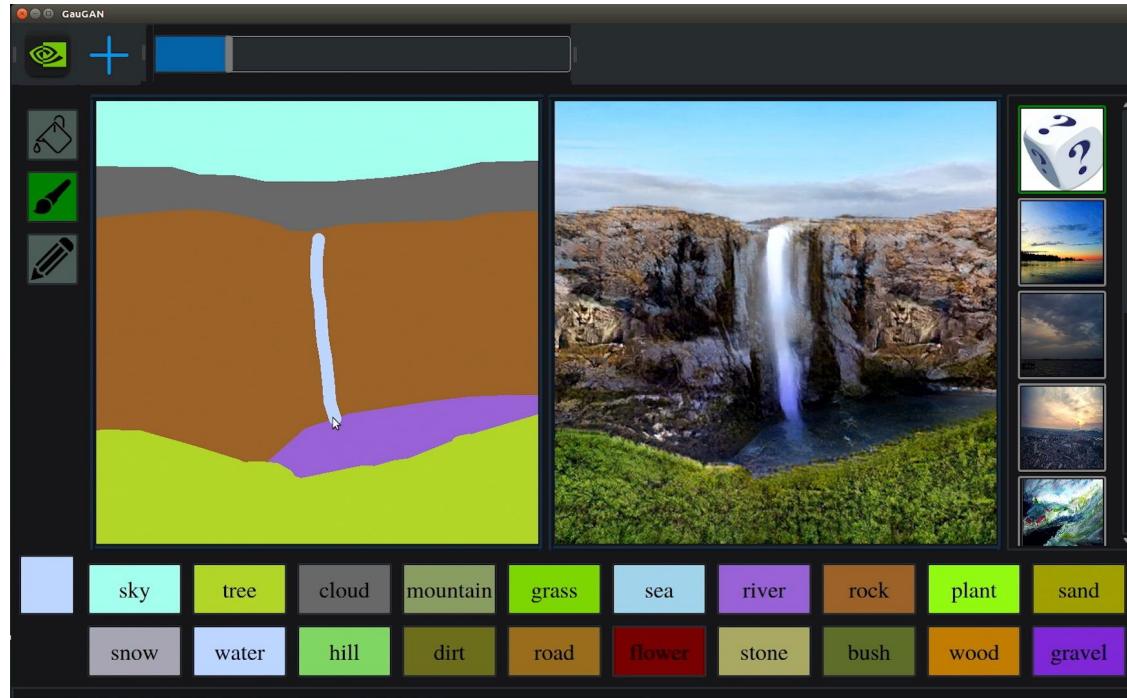


# Pix2PixHD



Available from: <https://github.com/NVIDIA/pix2pixHD>

# GauGAN



Available from: <https://blogs.nvidia.com/blog/2019/03/18/gaugan-photorealistic-landscapes-nvidia-research/>

# Summary

- Pix2PixHD and GauGAN are successors of Pix2Pix
- They are designed for higher resolution images
- They highlight opportunities for image editing using paired image-to-image translation
  - Pix2Pix can do this too, of course!

