deeplearning.ai

Image-to-Image Translation

Outline

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

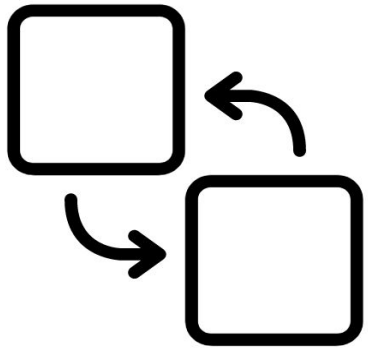
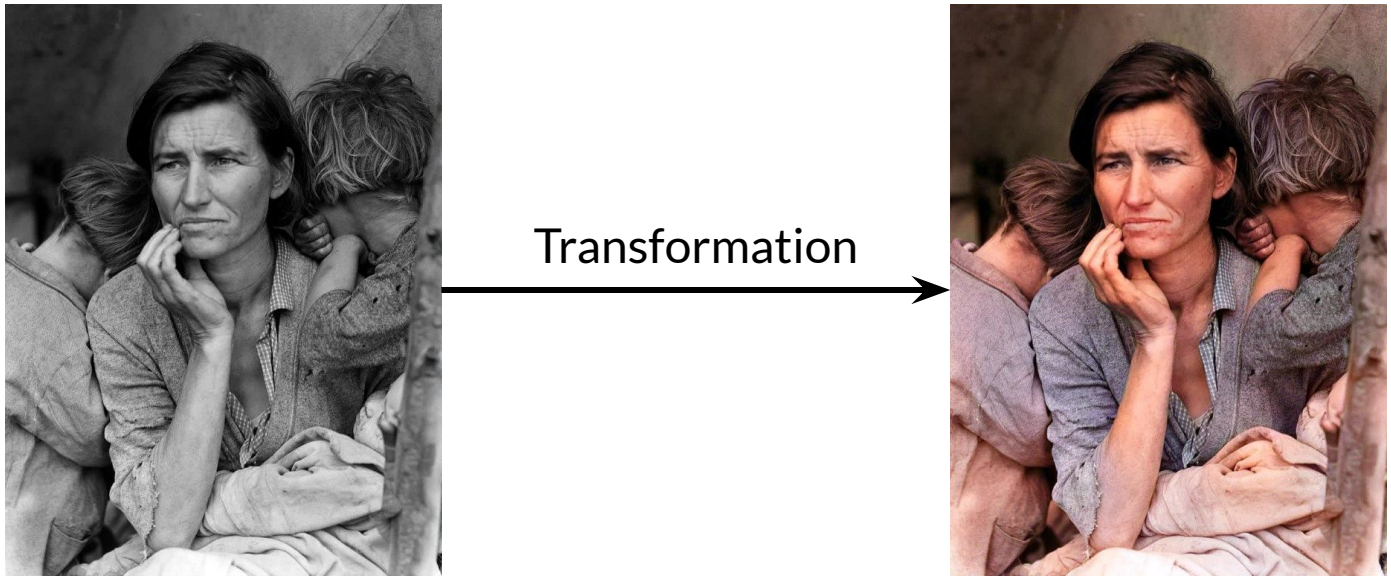
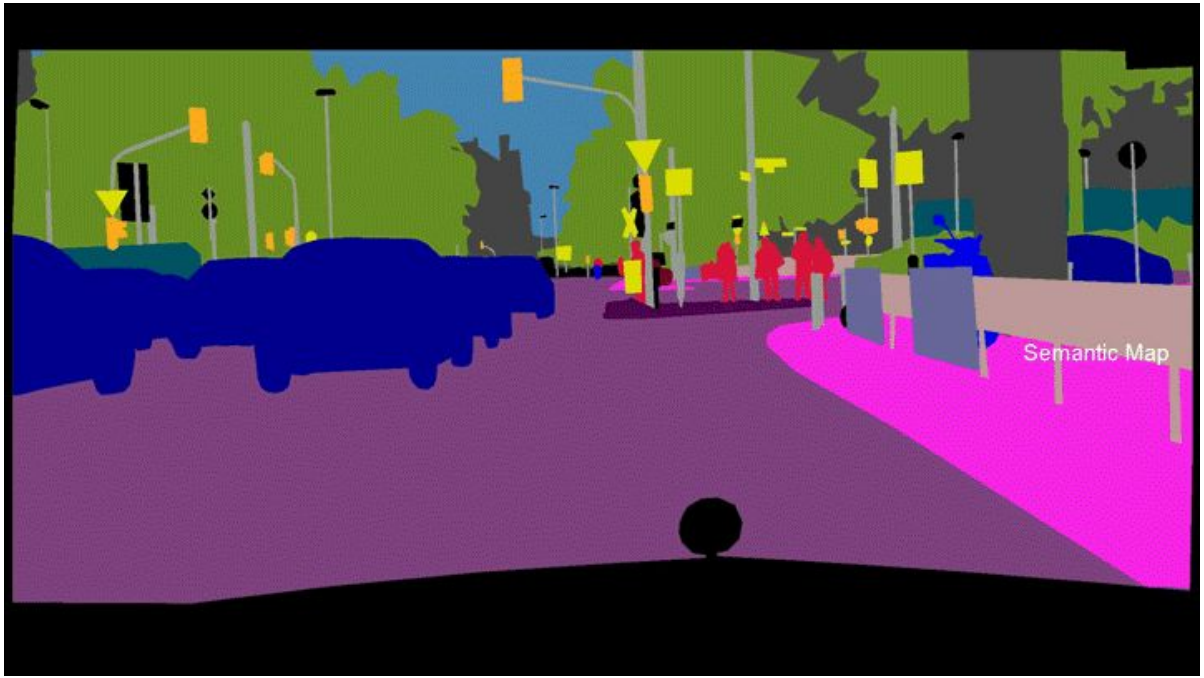


Image-to-Image Translation



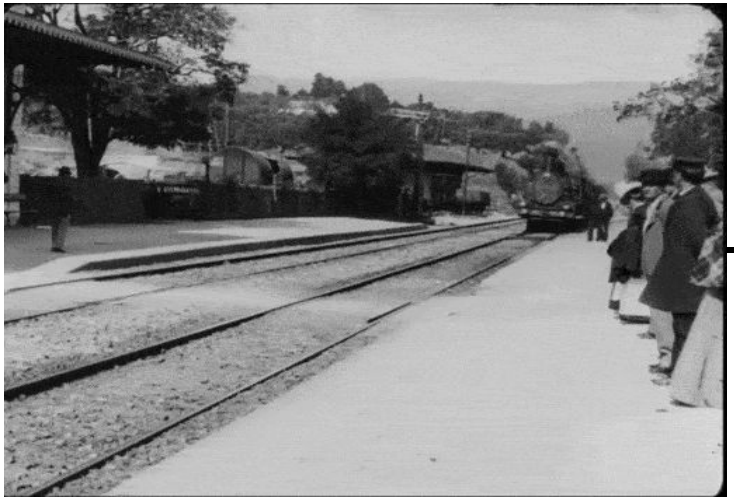
Available from: https://twitter.com/citnaj/status/1124904251128406016

Image-to-Image Translation



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

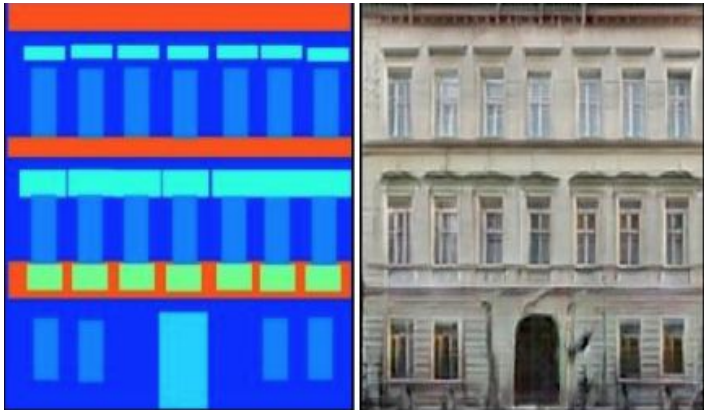
Image-to-Image Translation



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

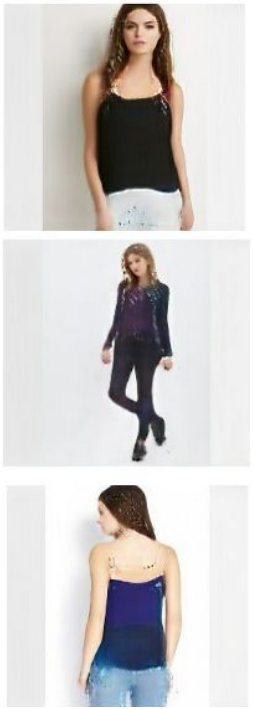
Day to night

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”

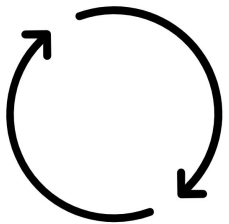
Available from: https://arxiv.org/abs/1711.10485

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

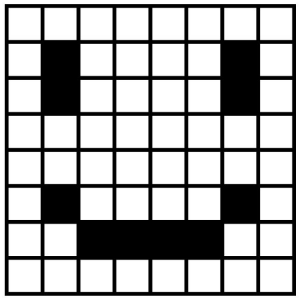


deeplearning.ai

Pix2Pix Overview

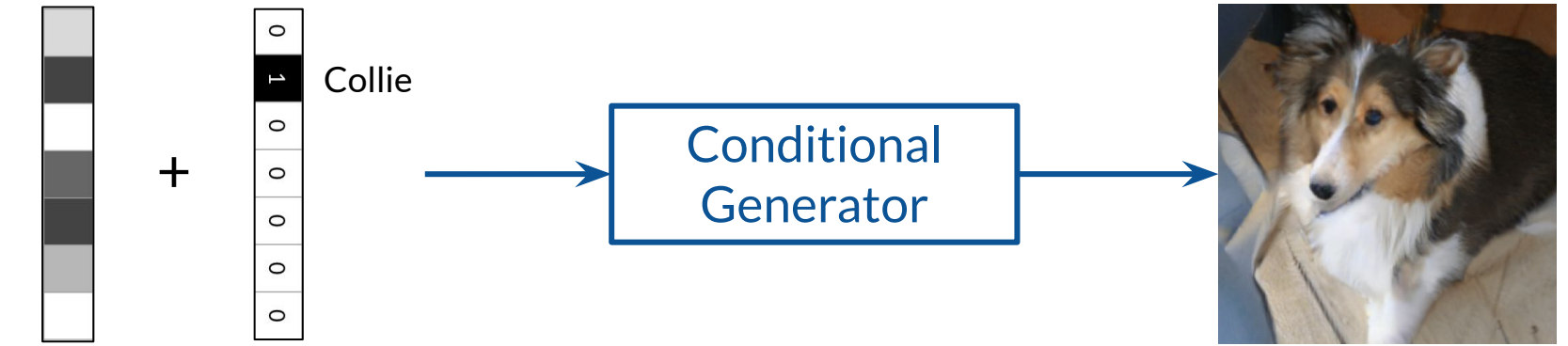
Outline

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

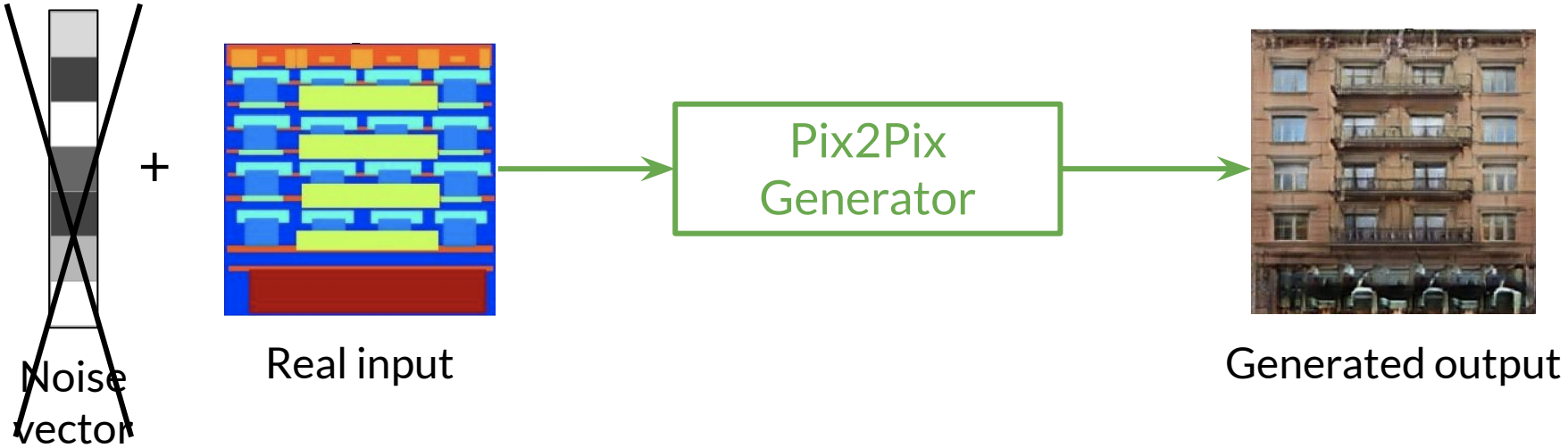


Pix2Pix for Paired Image-to-Image Translation

Pix2Pix Generator

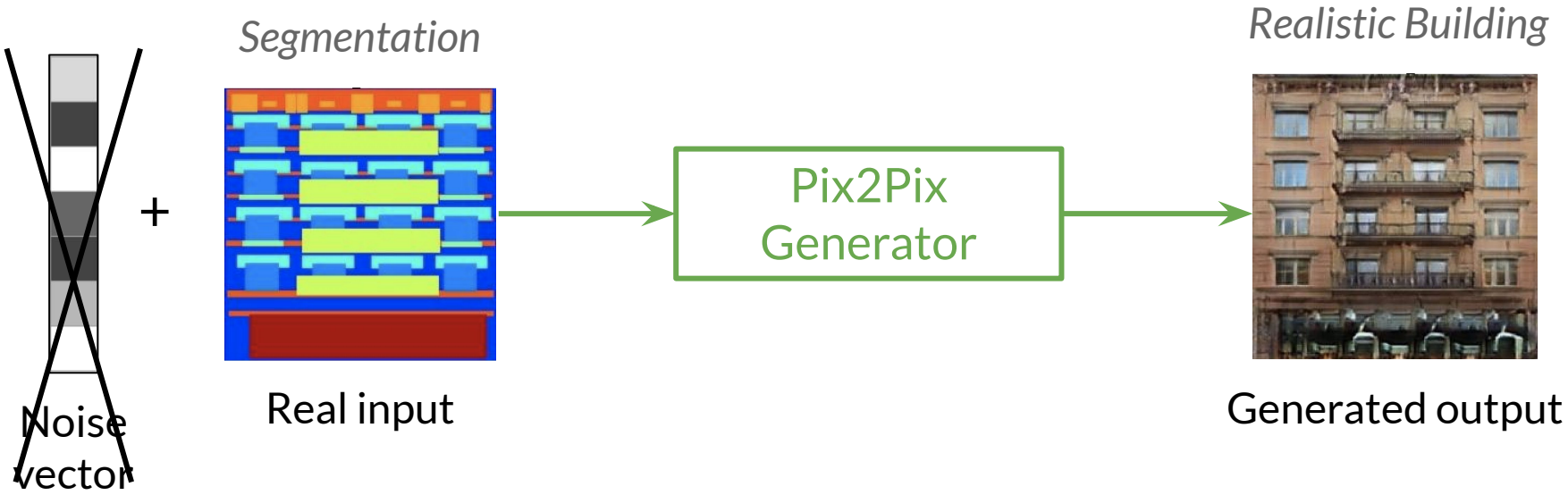


Pix2Pix Generator



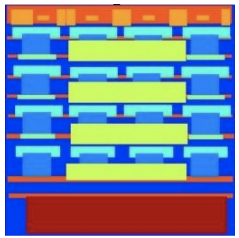
Available from: https://arxiv.org/abs/1611.07004

Pix2Pix Generator



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

Pix2Pix Discriminator

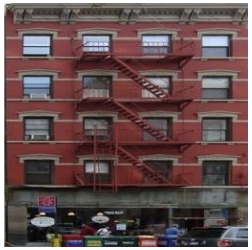


Real input

⊕

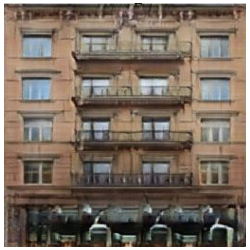
Pix2Pix Discriminator

Real/Fake



Real output

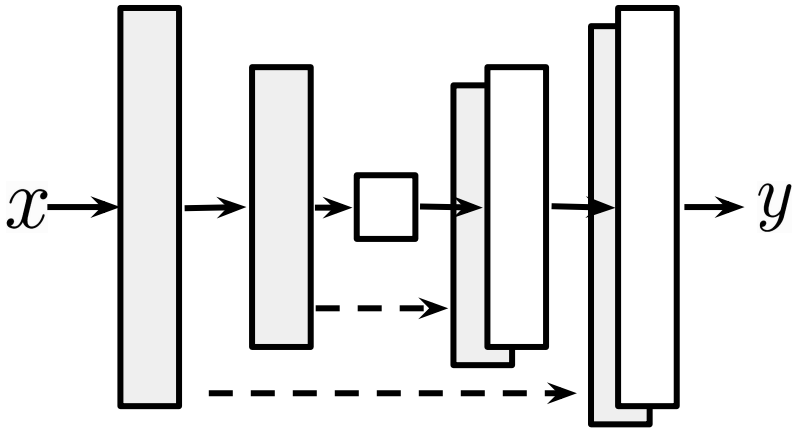
Or



Generated output

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

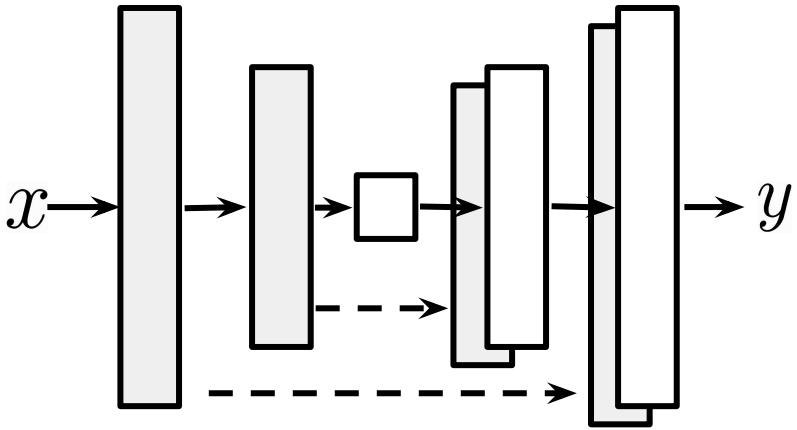
Pix2Pix Upgrades



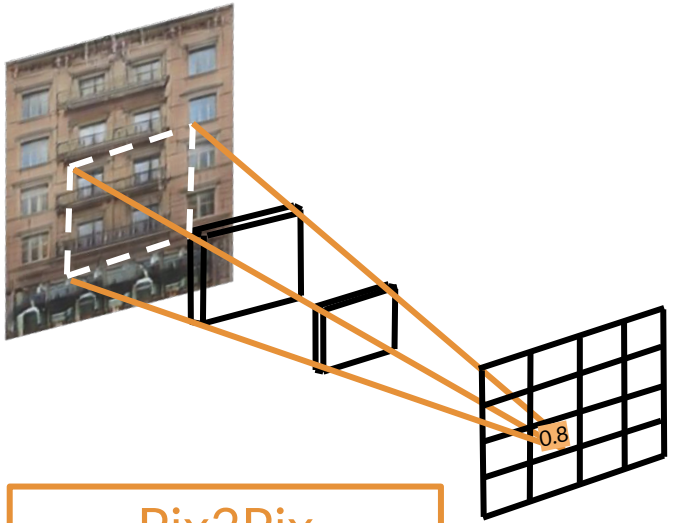
Pix2Pix Generator

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

Pix2Pix Upgrades



Pix2Pix
Generator

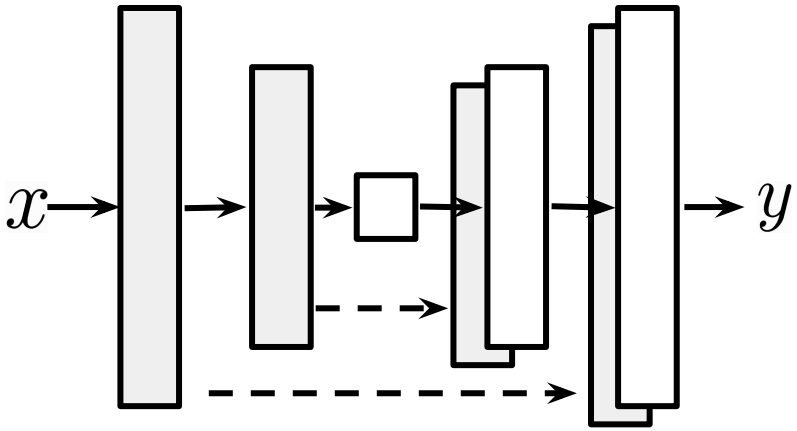


Pix2Pix
Discriminator

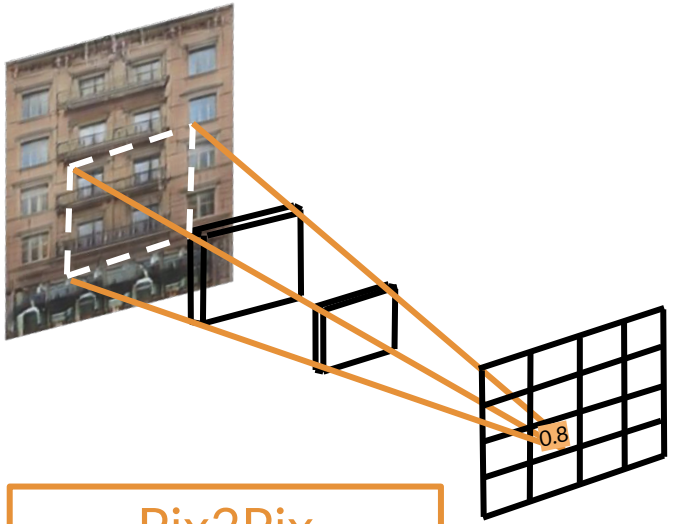
(Left) Based on: https://arxiv.org/abs/1611.07004
(Right) Based on: https://arxiv.org/abs/1803.07422

Pix2Pix Upgrades

Goal is still to produce realistic outputs!



Pix2Pix Generator

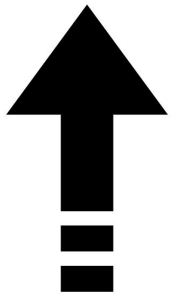


Pix2Pix Discriminator

(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

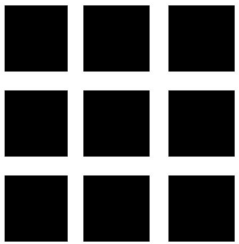


deeplearning.ai

Pix2Pix: PatchGAN

Outline

- PatchGAN discriminator architecture
- Matrix output vs. single output



Pix2Pix Discriminator: PatchGAN

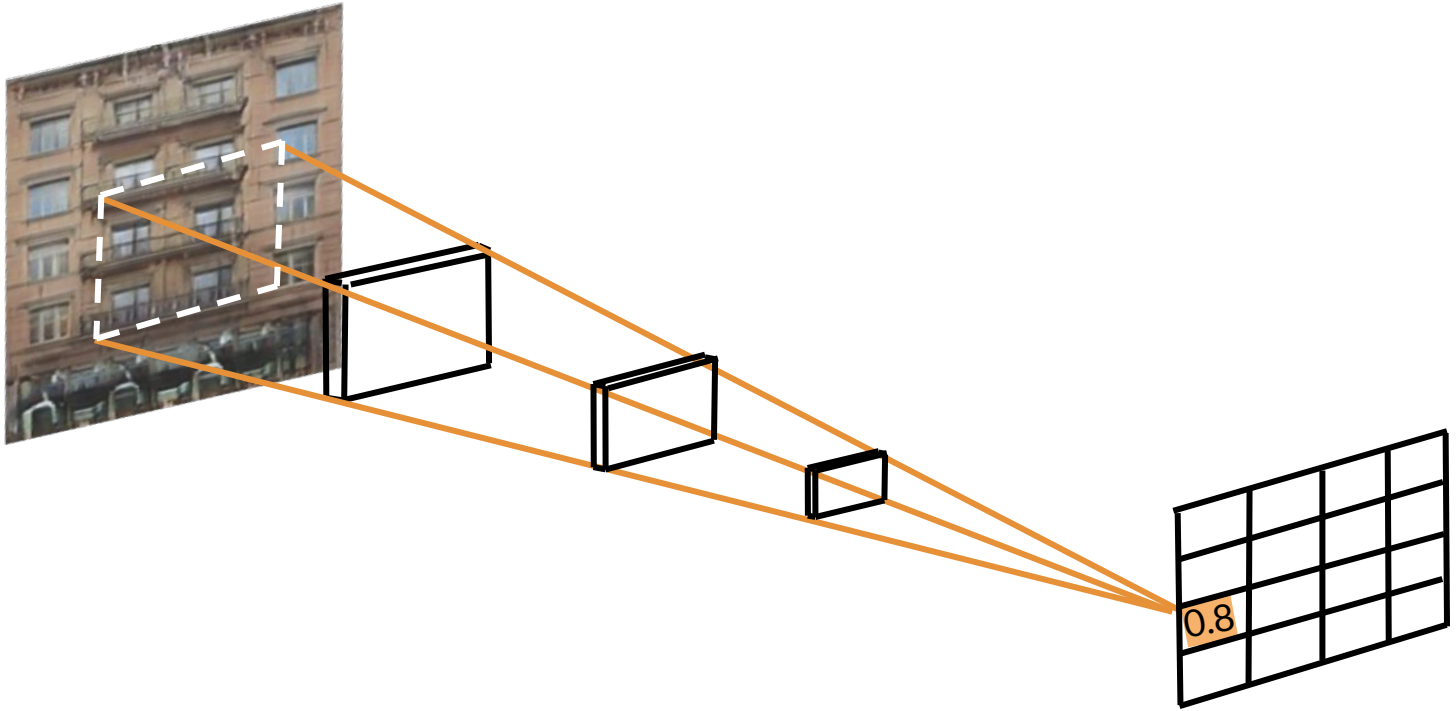


Image available from: https://arxiv.org/abs/1611.07004
Based on: https://arxiv.org/abs/1803.07422

Pix2Pix Discriminator: PatchGAN

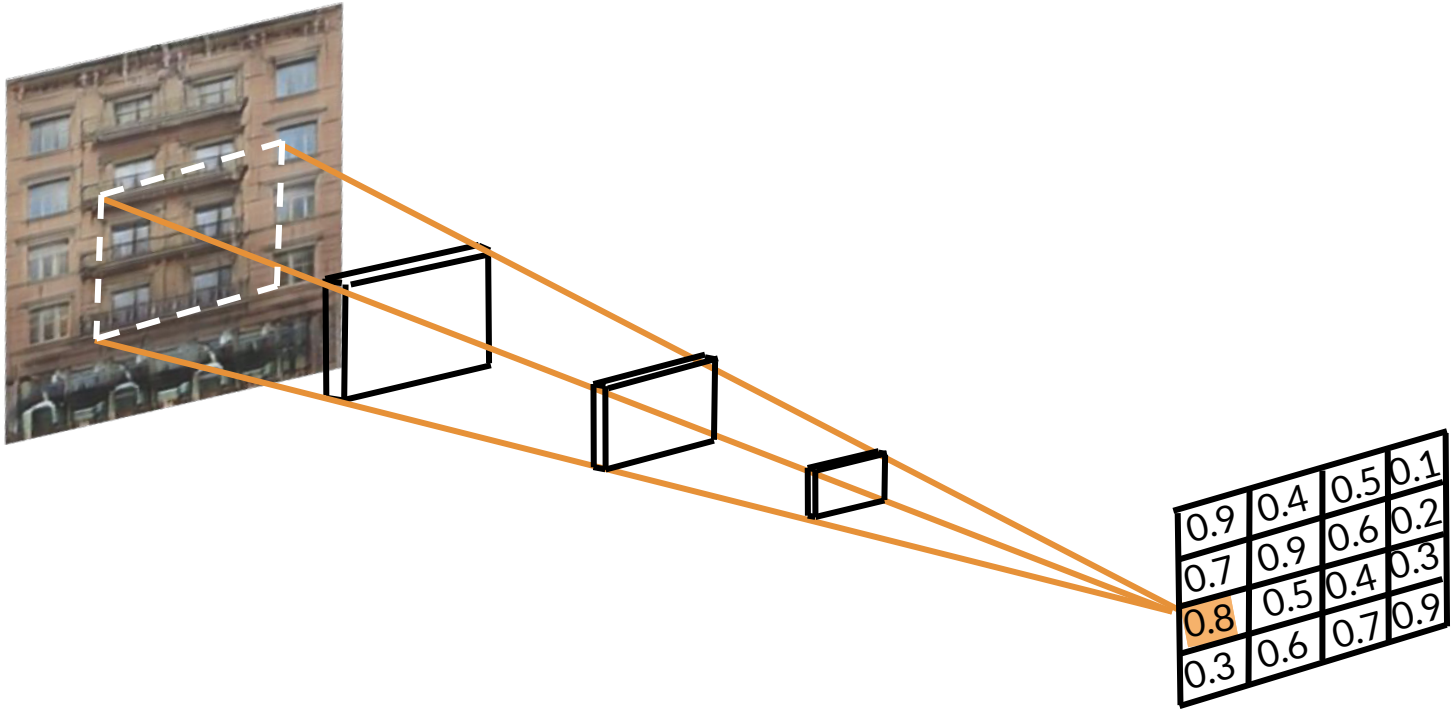


Image available from: https://arxiv.org/abs/1611.07004
Based on: https://arxiv.org/abs/1803.07422

Pix2Pix Discriminator: PatchGAN

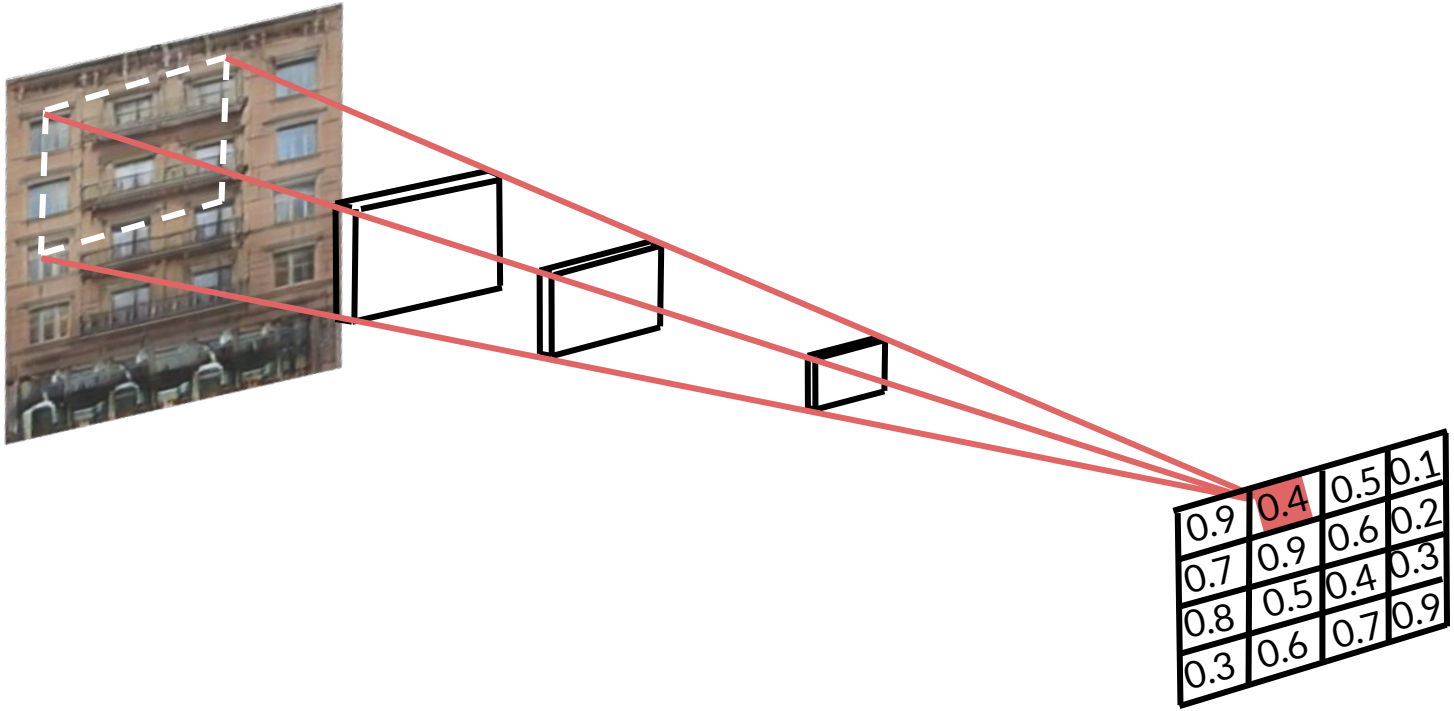


Image available from: https://arxiv.org/abs/1611.07004
Based on: https://arxiv.org/abs/1803.07422

Pix2Pix Discriminator: PatchGAN

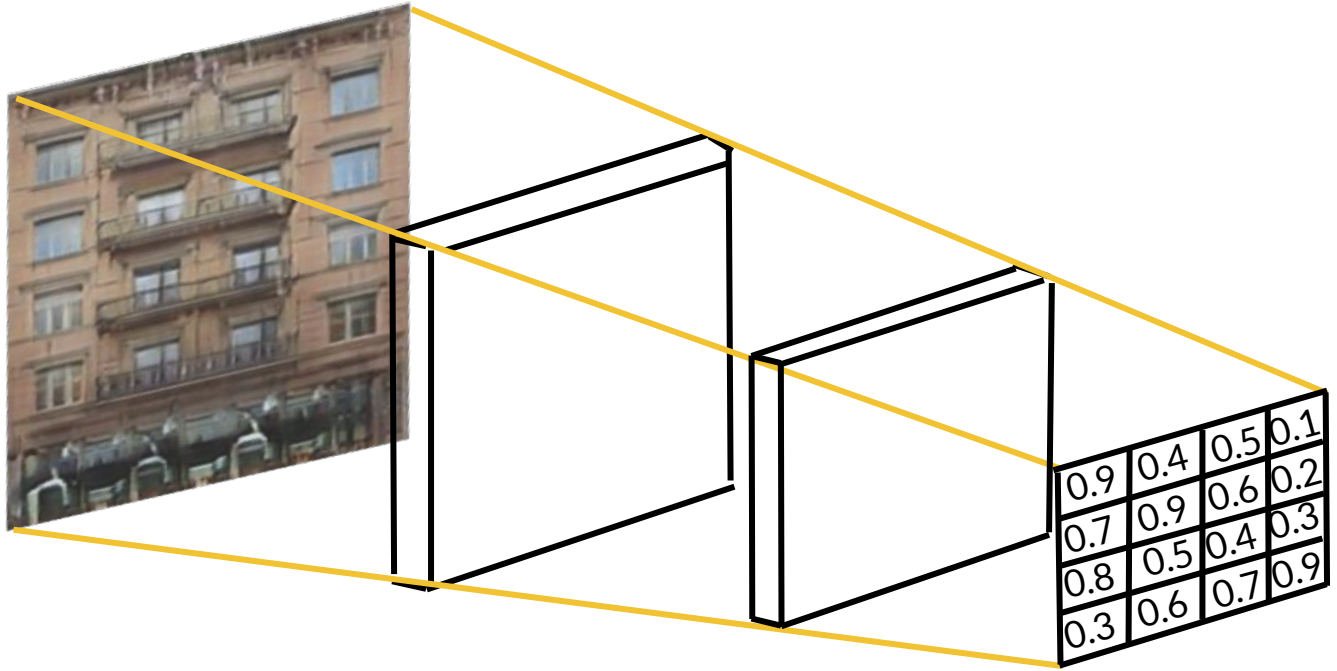
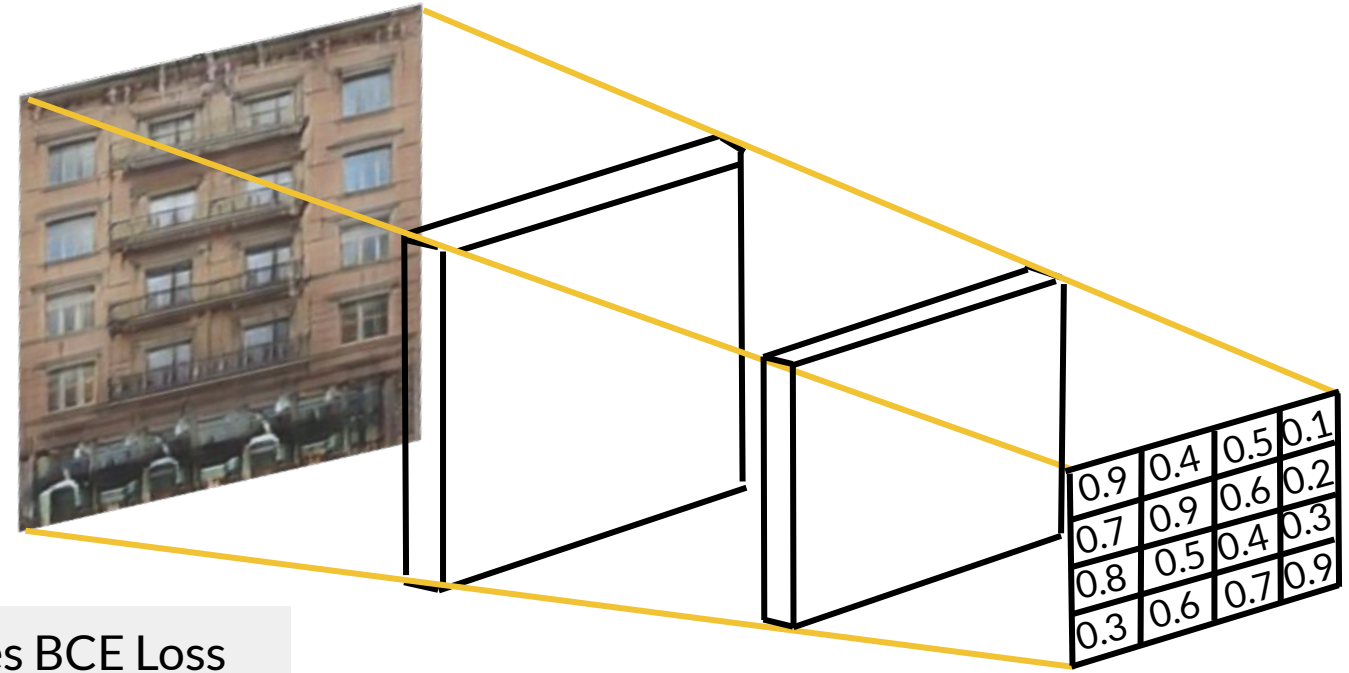


Image available from: https://arxiv.org/abs/1611.07004
Based on: https://arxiv.org/abs/1803.07422

Pix2Pix Discriminator: PatchGAN

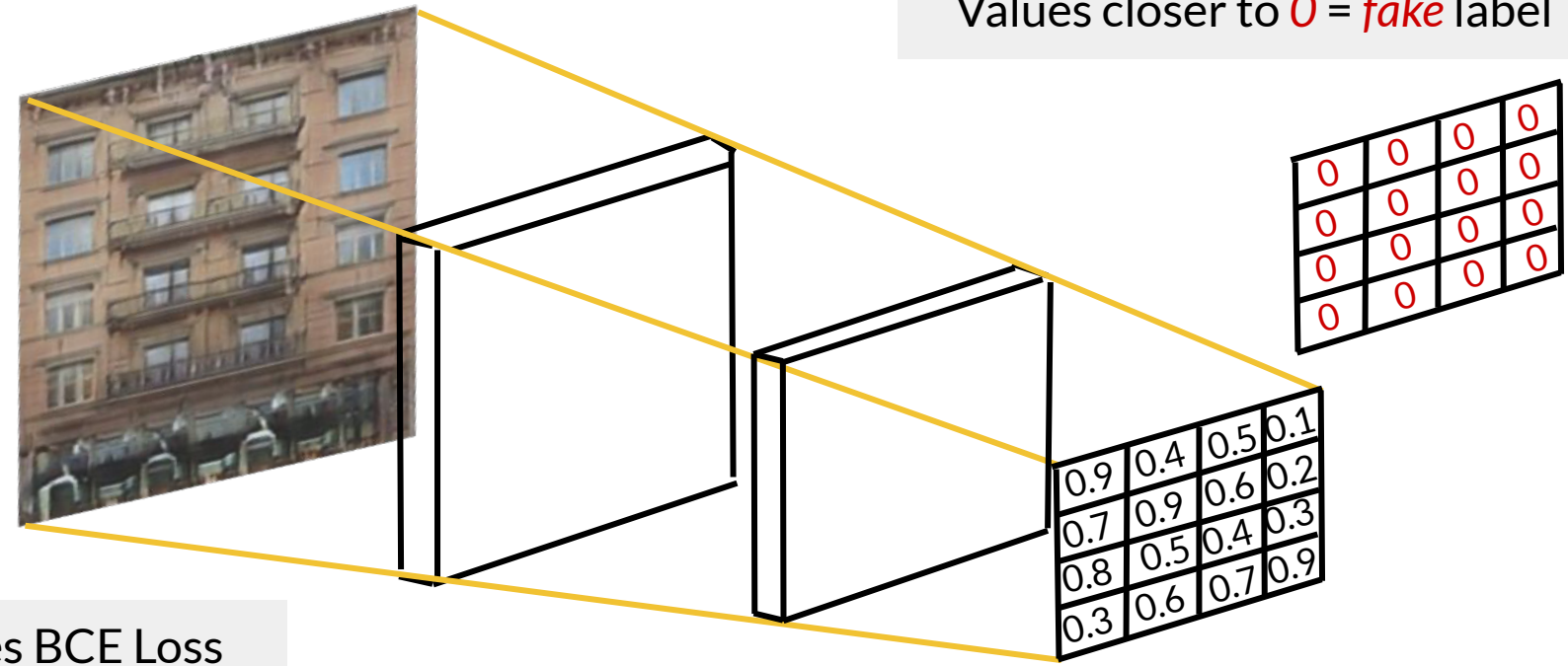


Still uses BCE Loss

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

Pix2Pix Discriminator: PatchGAN

Values closer to *0* = *fake* label

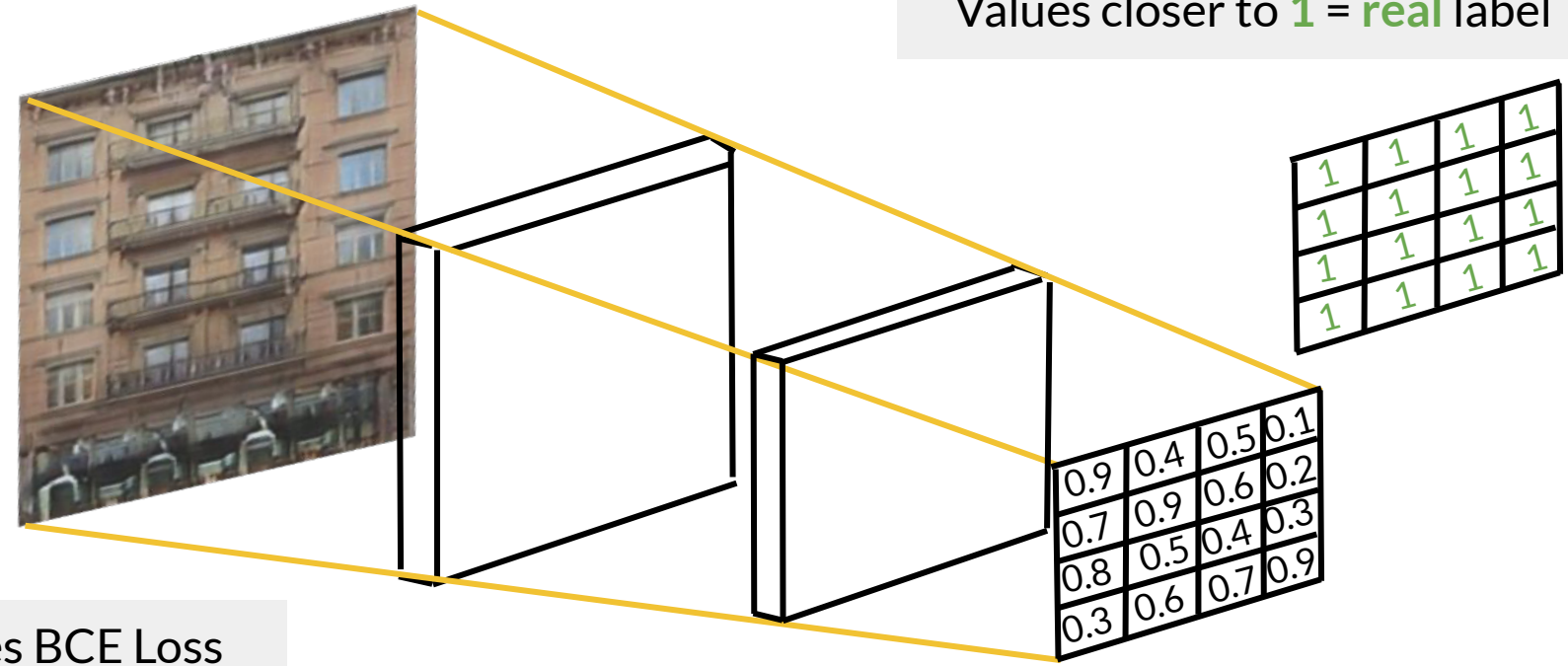


Still uses BCE Loss

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

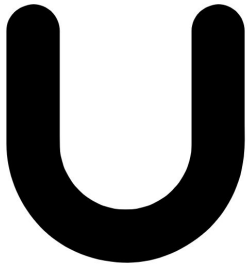


Image Segmentation

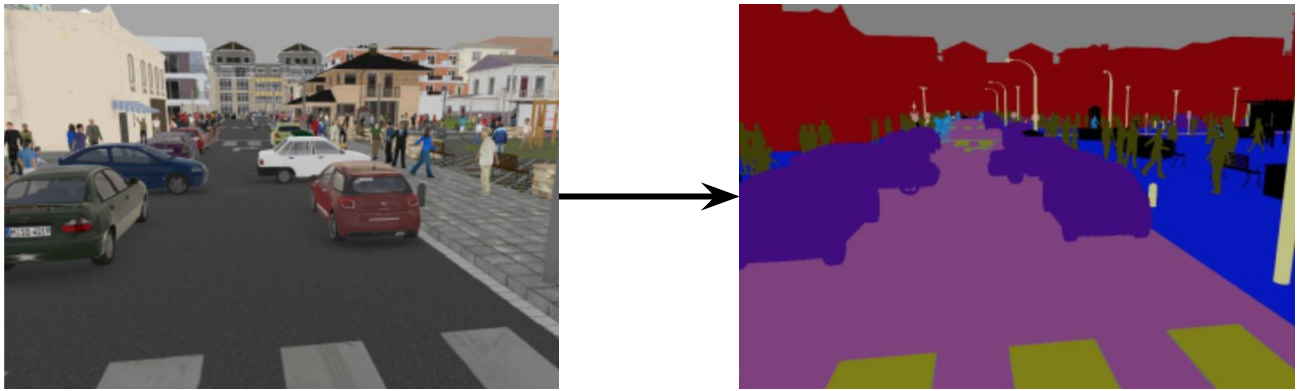


Image Segmentation

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

Image Segmentation

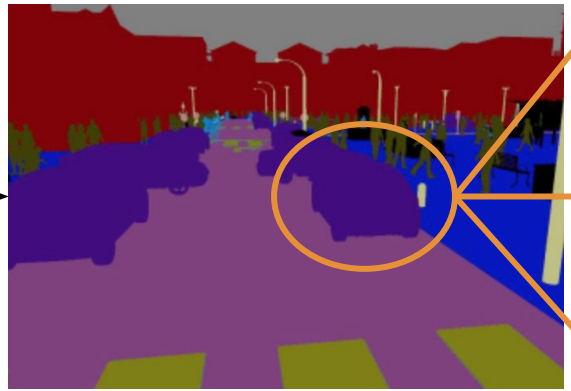
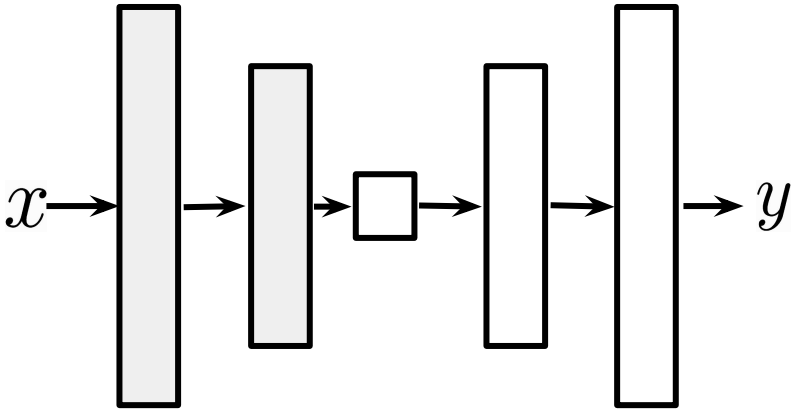


Image Segmentation

Image-to-Image Translation

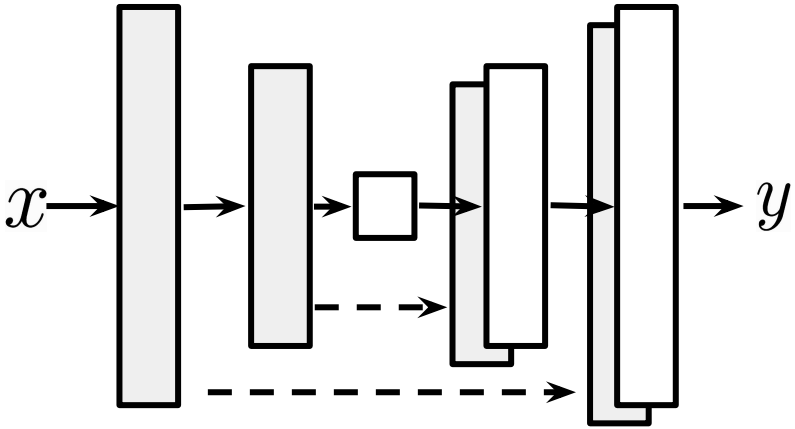
Available from: https://developer.nvidia.com/blog/image-segmentation-using-digits-5/

U-Net Framework: Encoder-Decoder



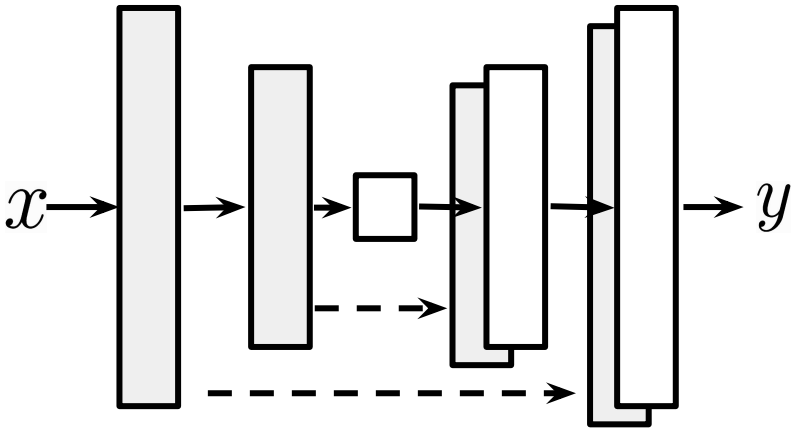
Based on: https://arxiv.org/abs/1611.07004

U-Net Framework: Skip Connections



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

U-Net Framework: Skip Connections



Forward pass

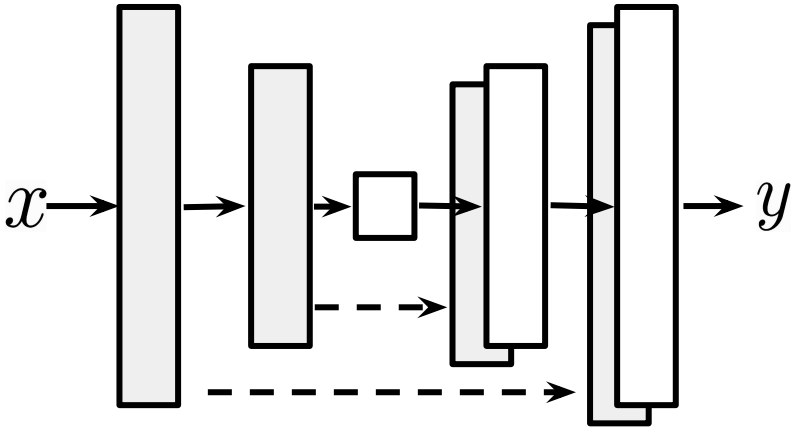
Skip connections allow information flow to the decoder

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

U-Net Framework: Skip Connections

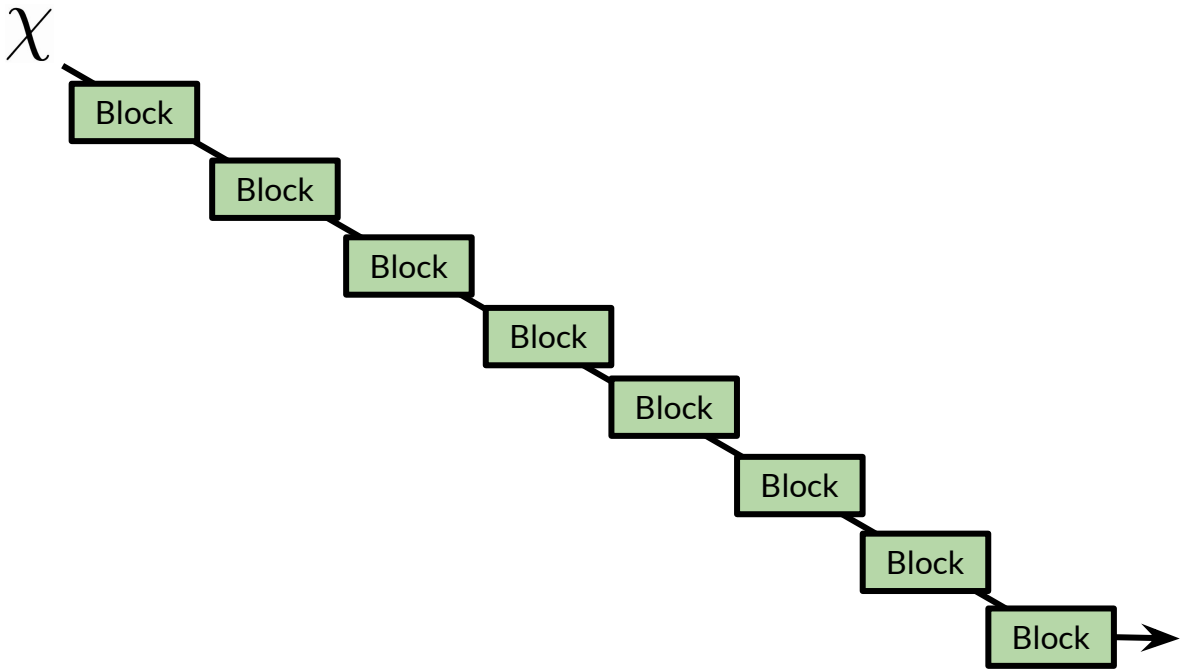
Backward pass

Skip connections improve gradient flow to encoder

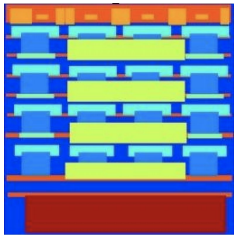


Based on: https://arxiv.org/abs/1611.07004

Pix2Pix Encoder



Pix2Pix Encoder

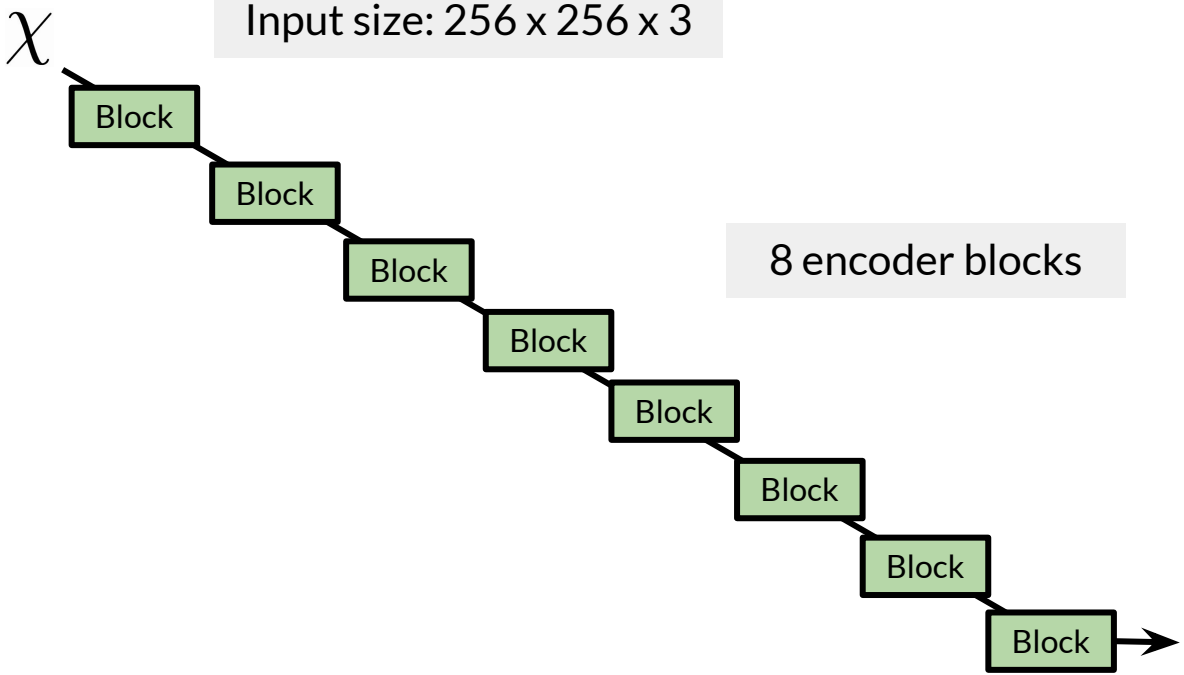


Input size: 256 x 256 x 3

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

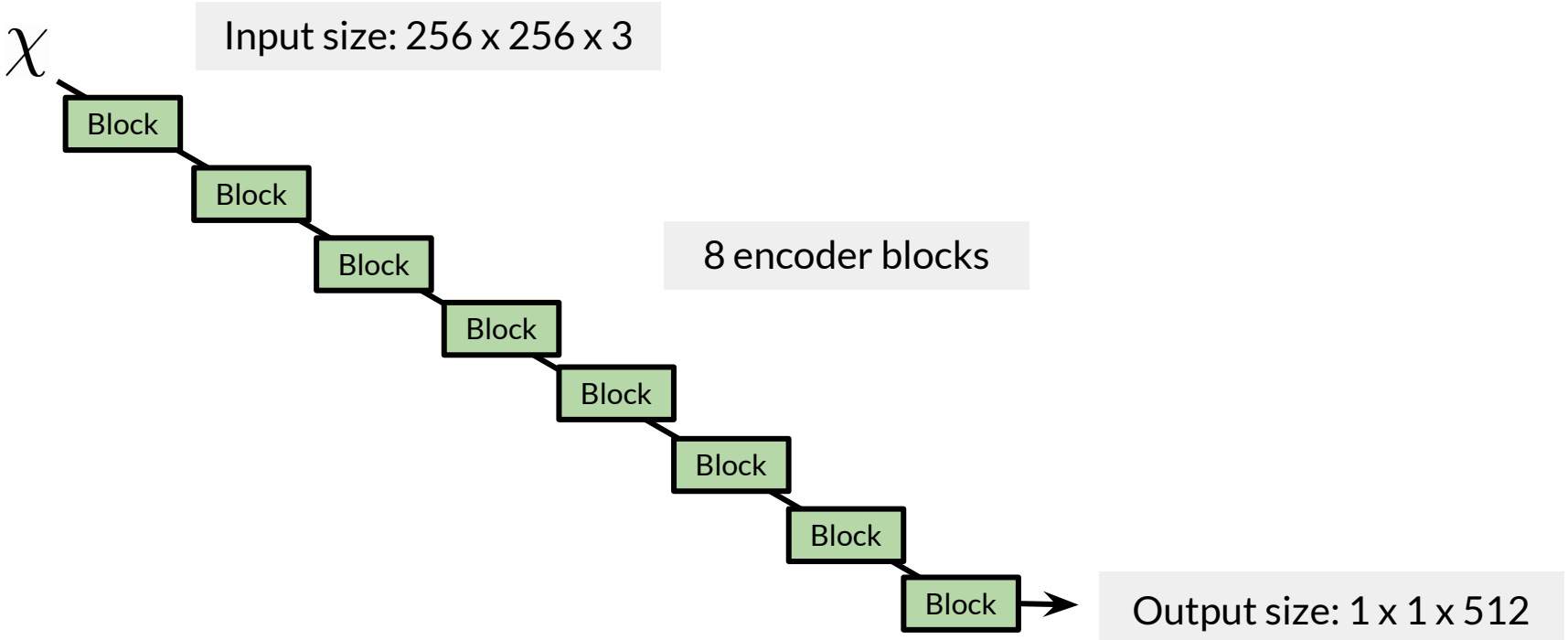
Pix2Pix Encoder

Input size: 256 x 256 x 3

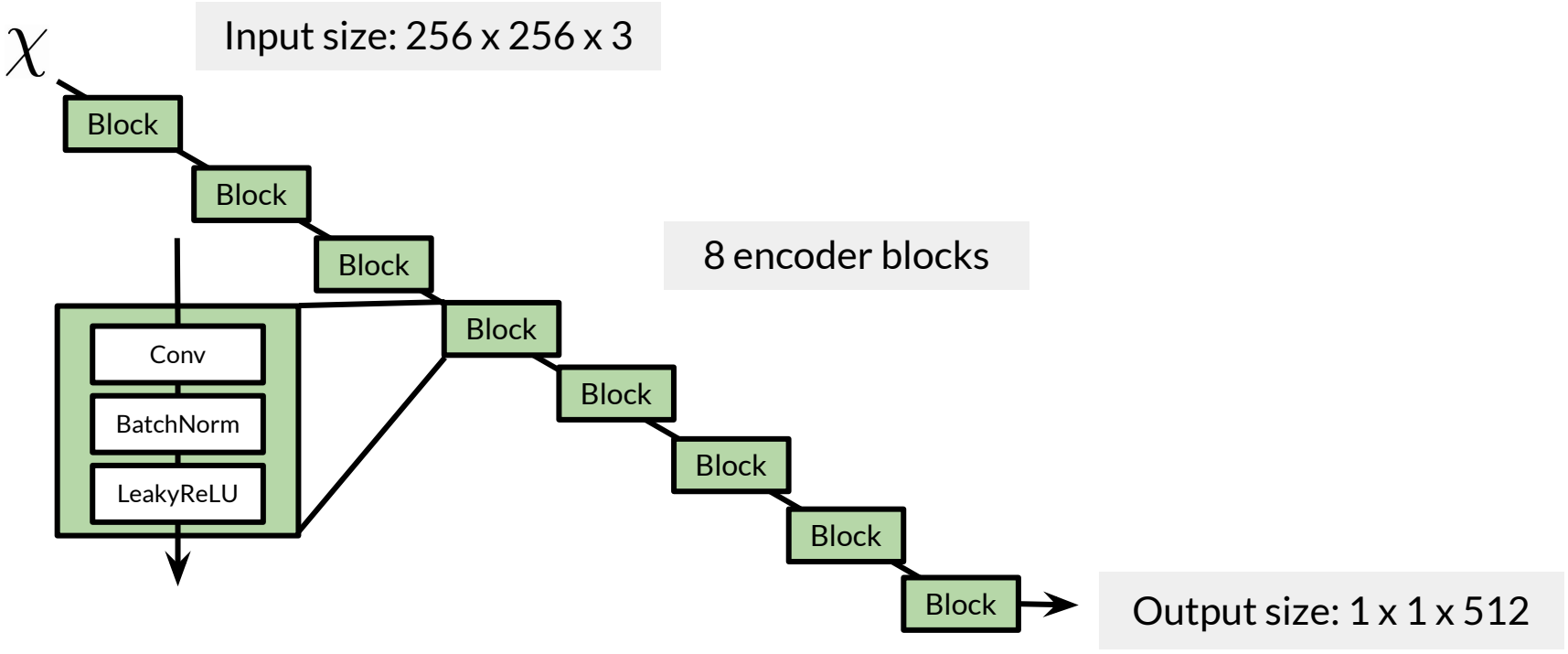


8 encoder blocks

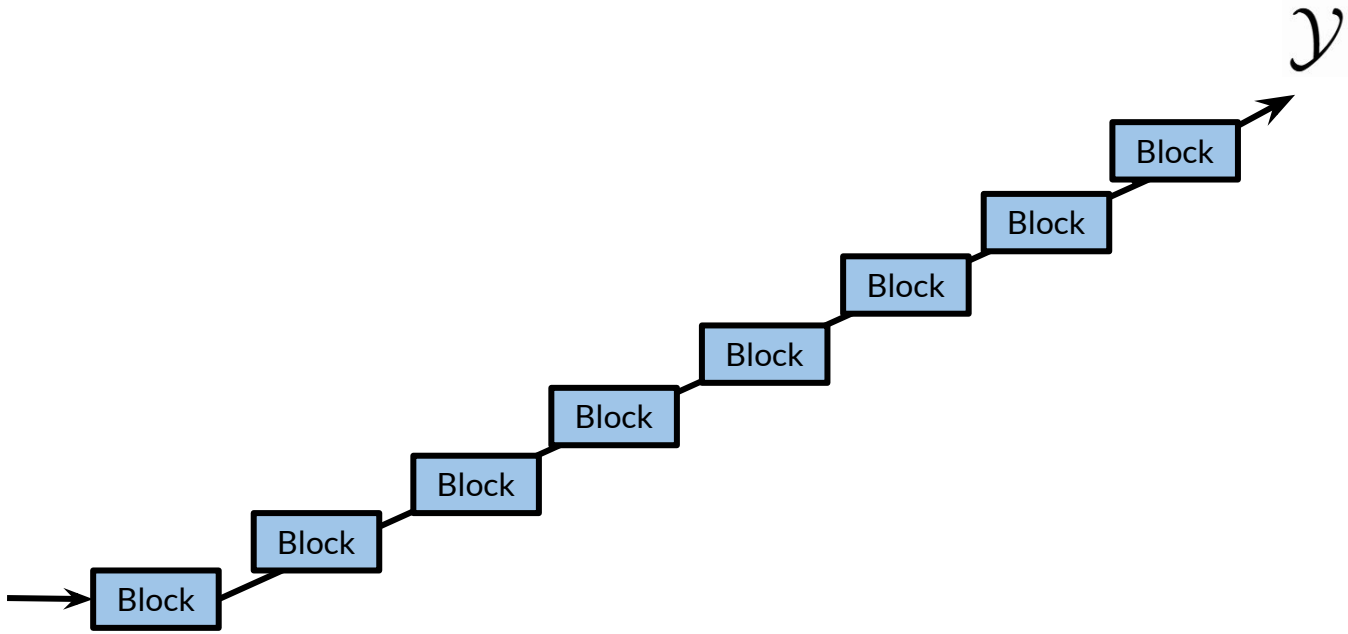
Pix2Pix Encoder



Pix2Pix Encoder

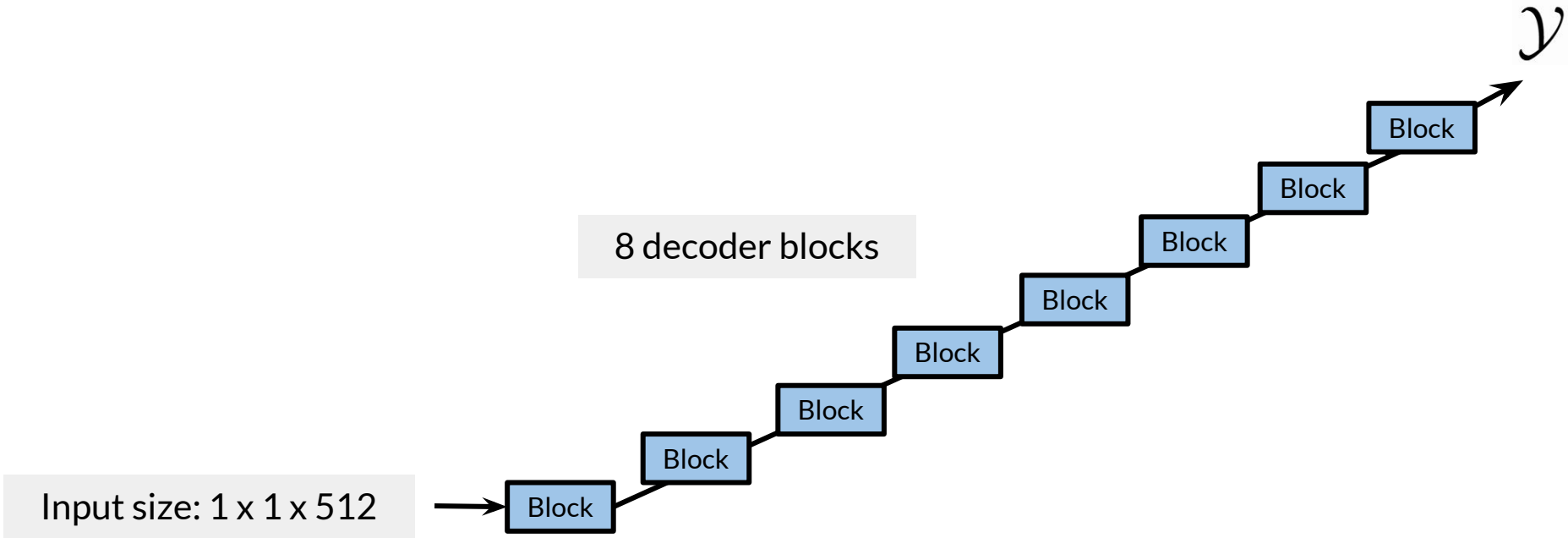


Pix2Pix Decoder

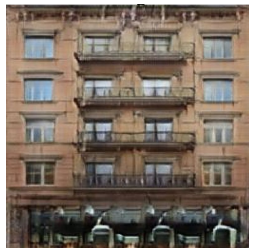


Input size: 1 x 1 x 512

Pix2Pix Decoder



Pix2Pix Decoder



y

Output size: 256 x 256 x 3

8 decoder blocks

Block

Block

Block

Block

Block

Block

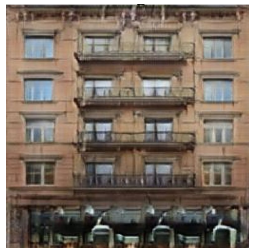
Block

Block

Input size: 1 x 1 x 512

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

Pix2Pix Decoder



Output size: 256 x 256 x 3

y

8 decoder blocks

Input size: 1 x 1 x 512

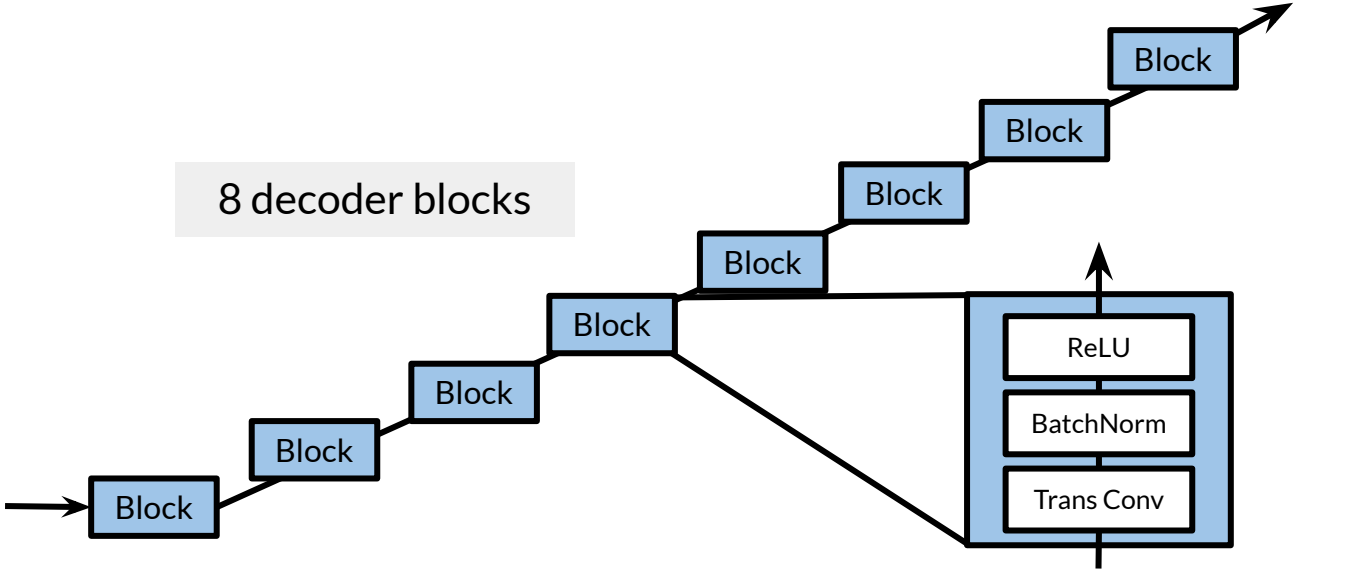


Image available from: https://arxiv.org/abs/1611.07004

Pix2Pix Decoder

Output size: 256 x 256 x 3

y

8 decoder blocks

Dropout in some decoder blocks *adds noise* to the network

Input size: 1 x 1 x 512

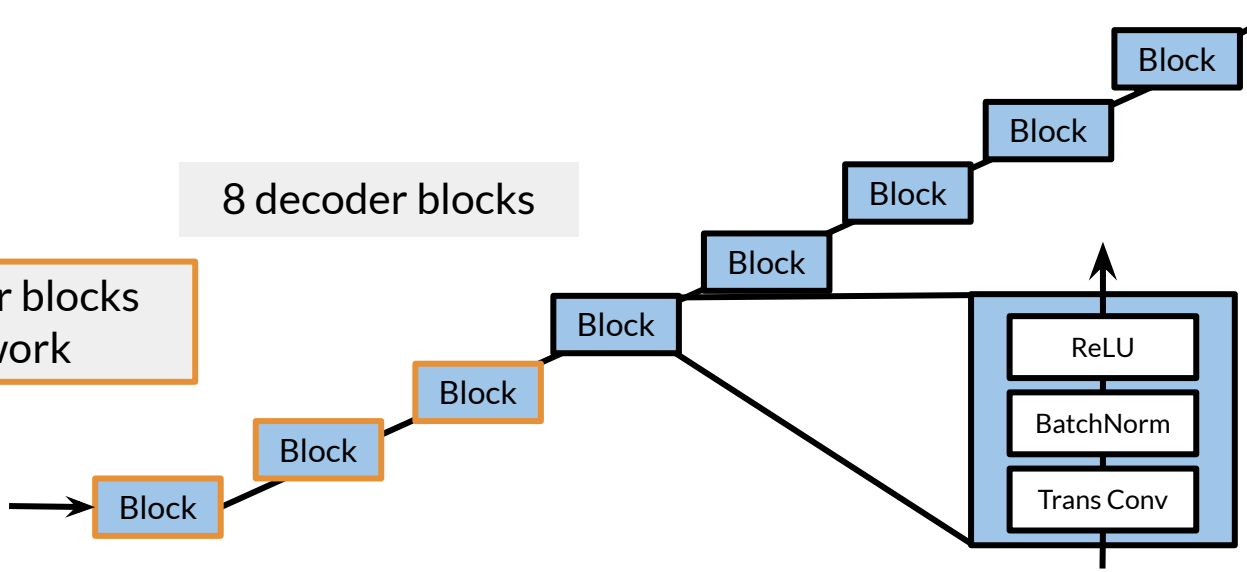
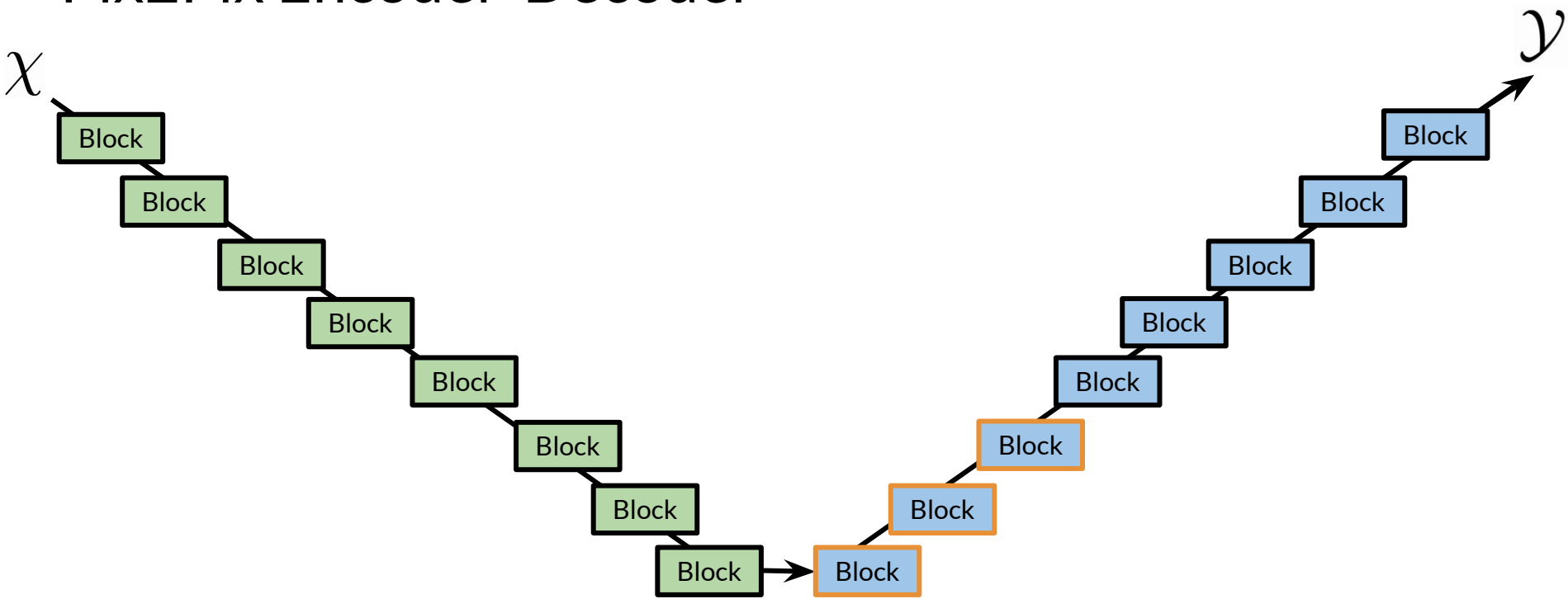
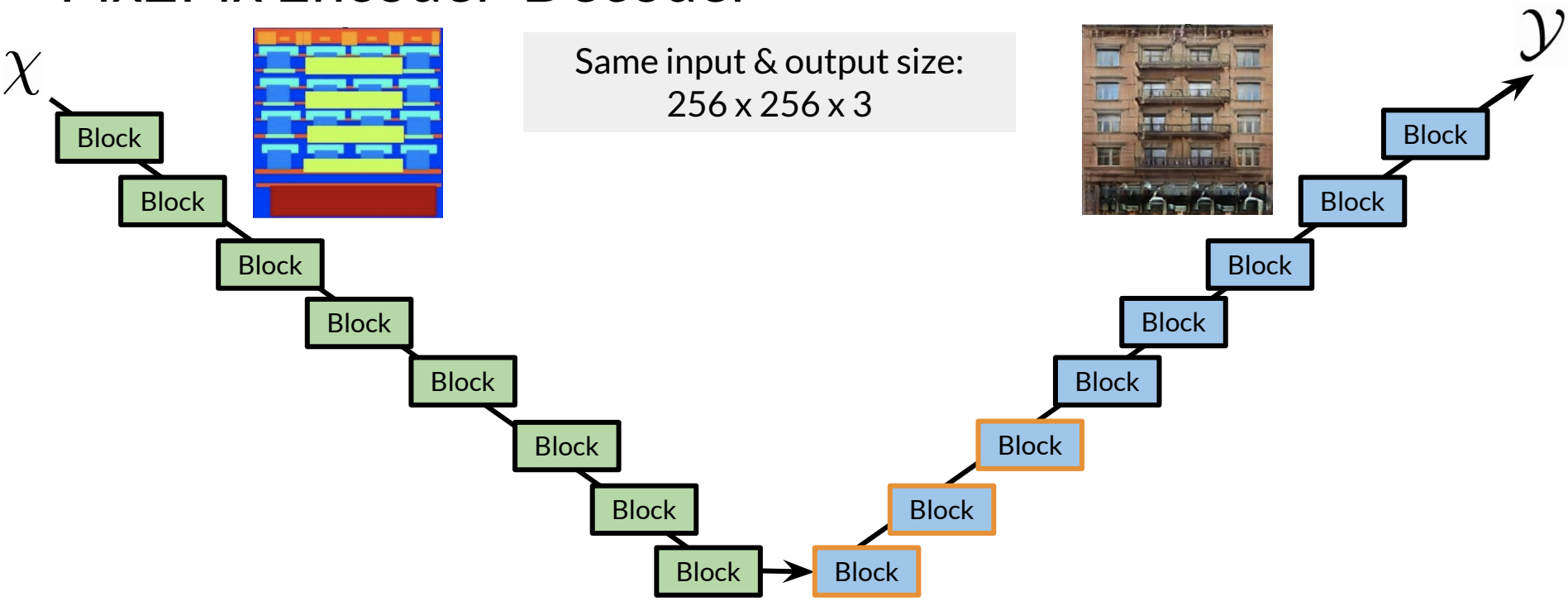


Image available from: https://arxiv.org/abs/1611.07004

Pix2Pix Encoder-Decoder



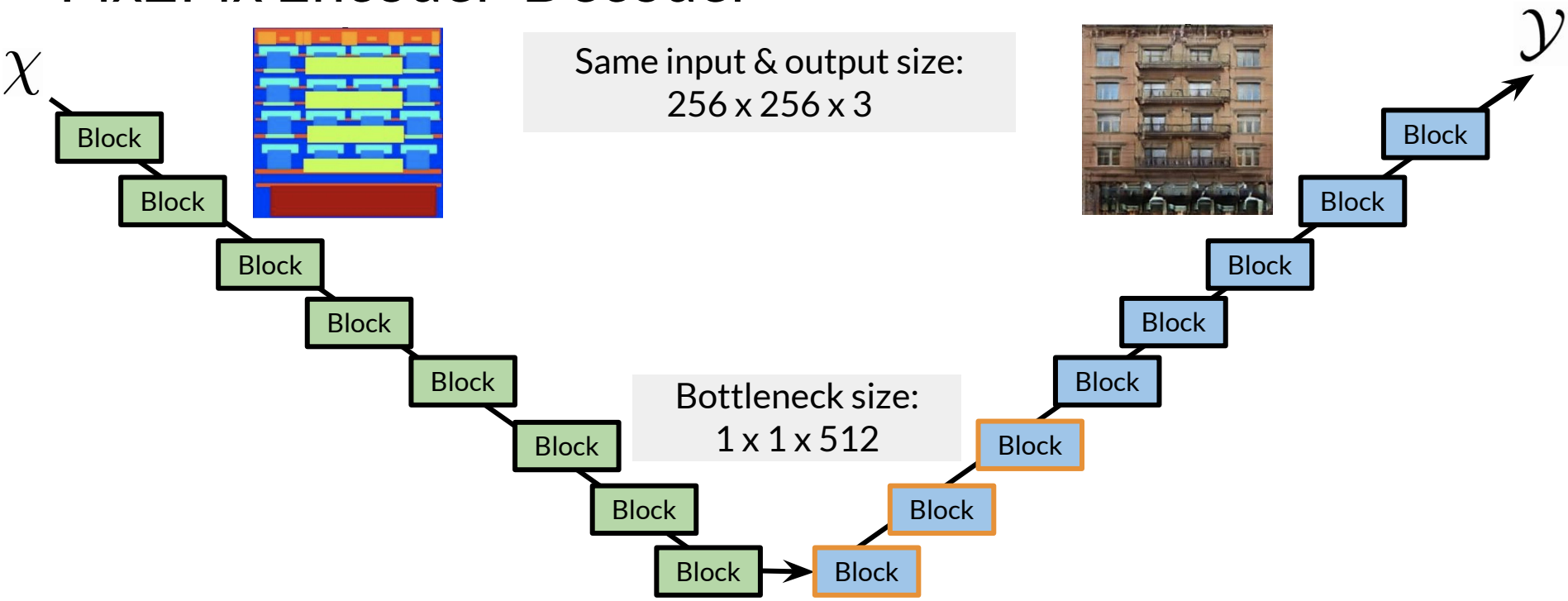
Pix2Pix Encoder-Decoder



Same input & output size:
256 x 256 x 3

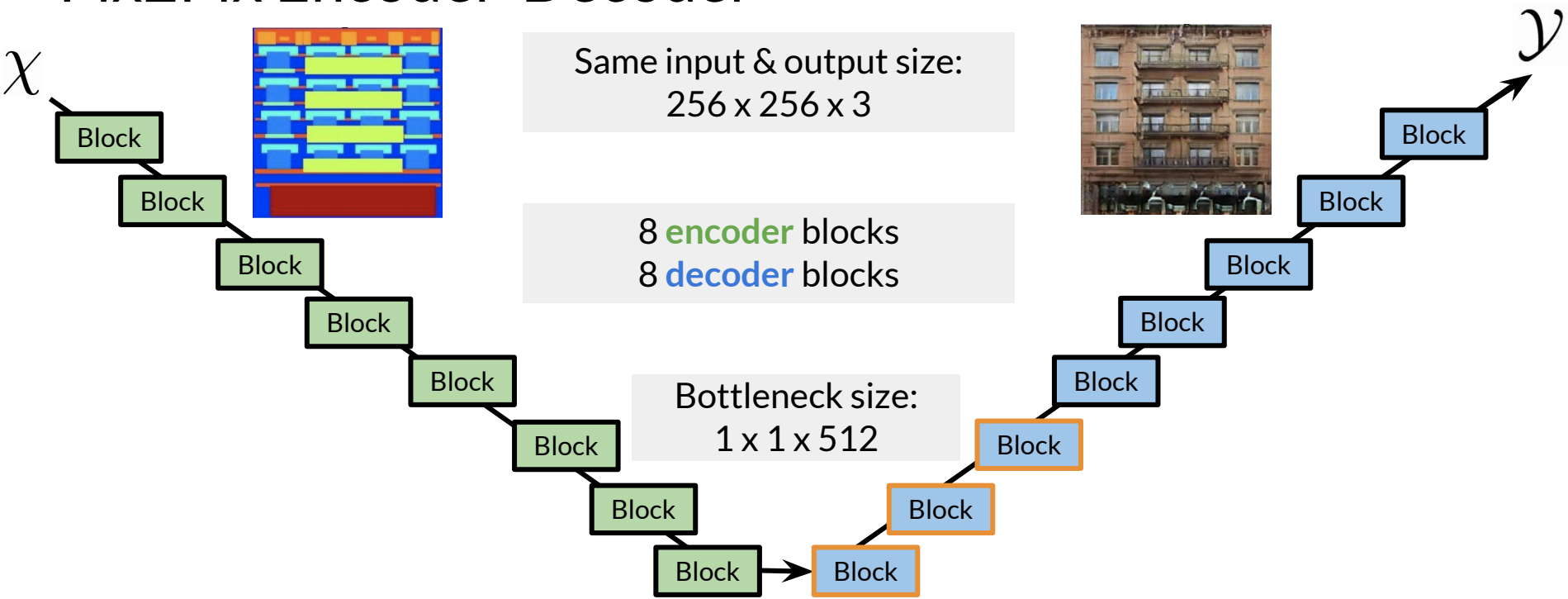
Images available from: https://arxiv.org/abs/1611.07004

Pix2Pix Encoder-Decoder



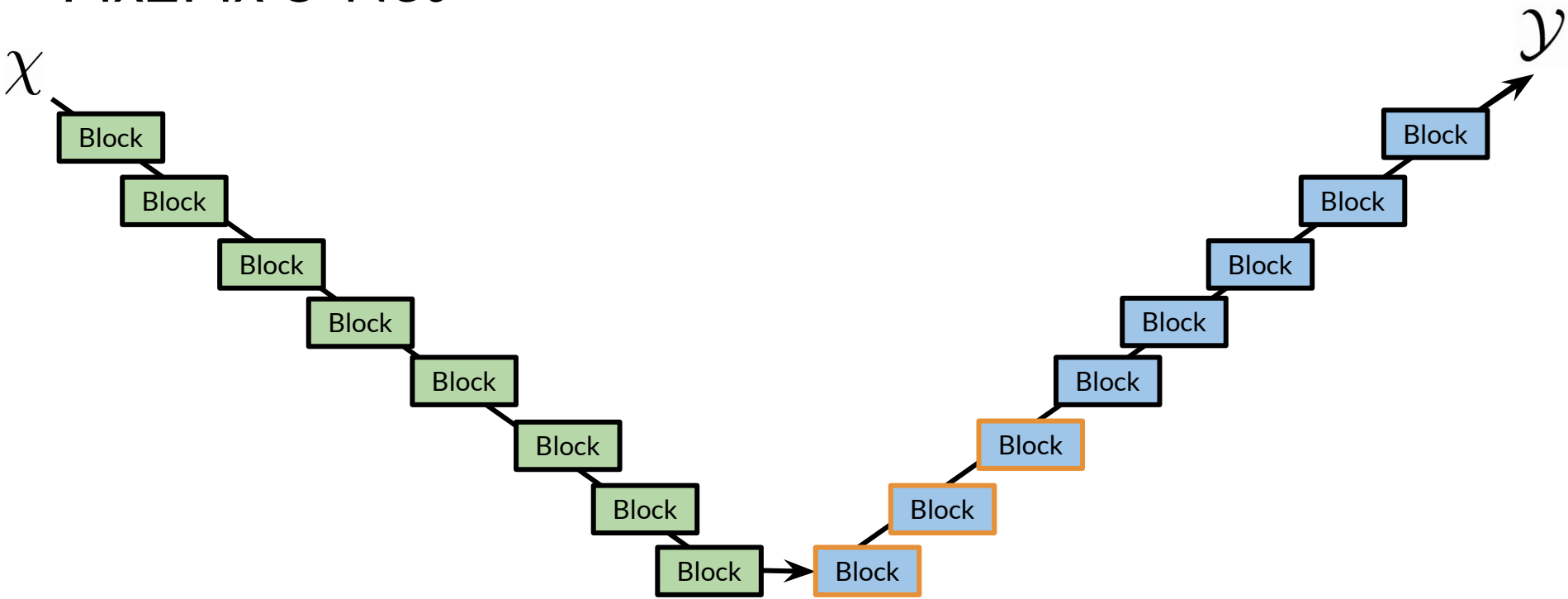
Images available from: https://arxiv.org/abs/1611.07004

Pix2Pix Encoder-Decoder

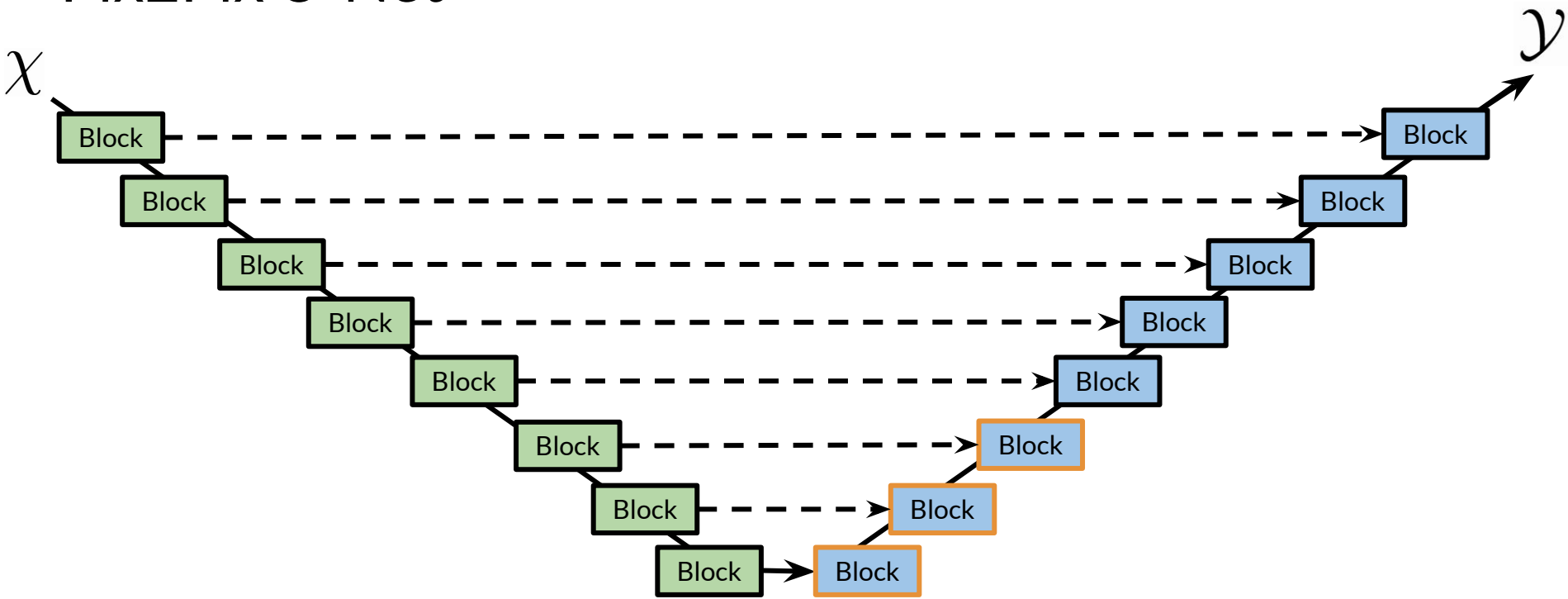


Images available from: https://arxiv.org/abs/1611.07004

Pix2Pix U-Net

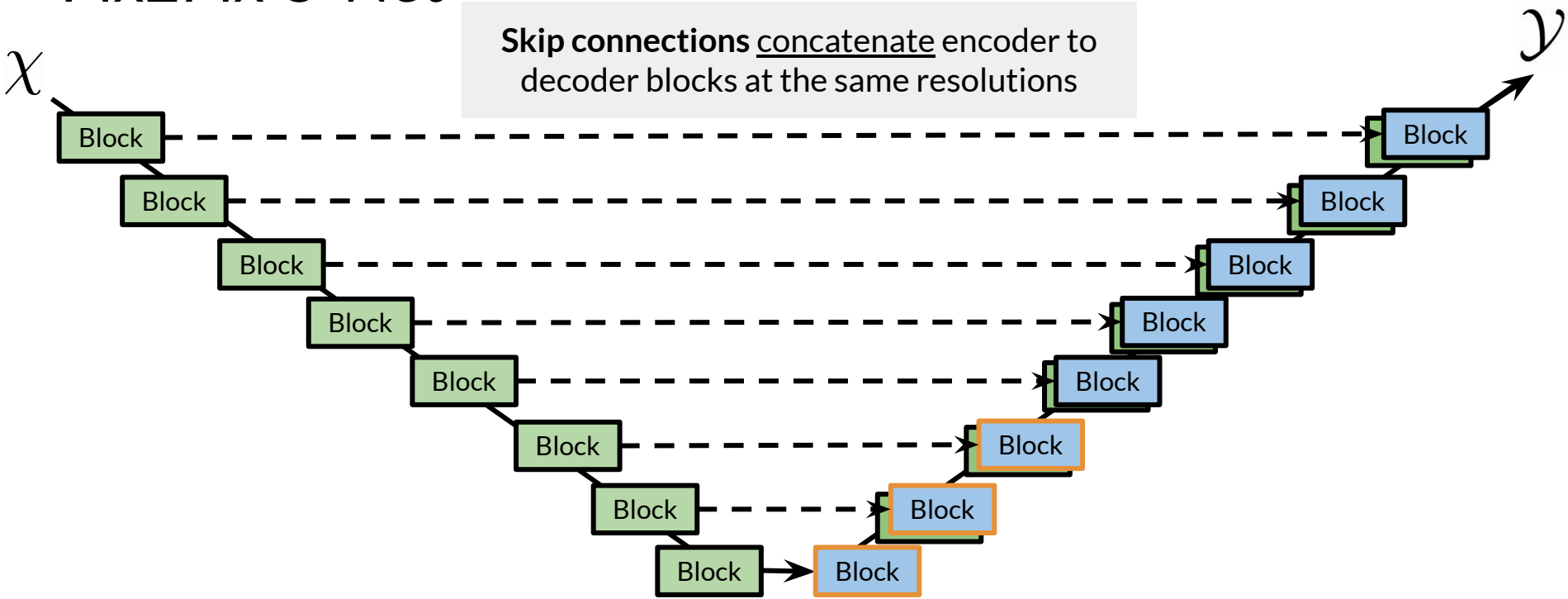


Pix2Pix U-Net



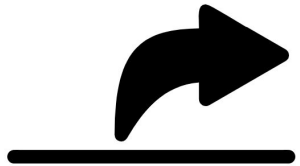
Pix2Pix U-Net

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 the encoder learn from more gradients flowing from decoder

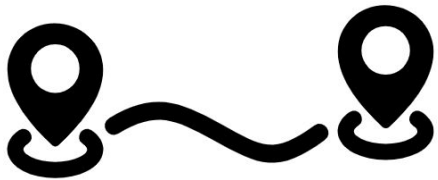


deeplearning.ai

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

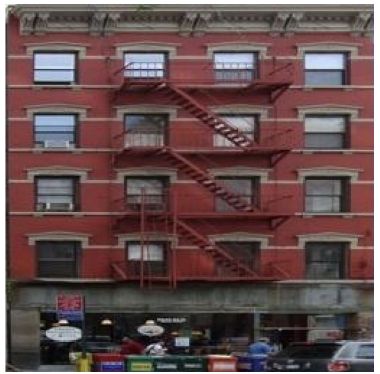
$$\min_{g} \max_{c} \text{ Adversarial Loss} + \lambda * \text{Other loss term}$$

Additional Loss Term

$$\min_{g} \max_{c} \text{ Adversarial Loss} + \lambda * \text{Pixel loss term}$$

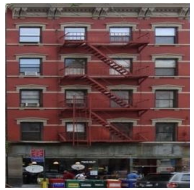
Pixel Distance Loss Term

$$\sum_{i=1}^{n} \left| \text{Generated output} - \text{Real output} \right|$$



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

Pix2Pix Generator Loss

BCE Loss + $\lambda \sum_{i=1}^{n} \left| \text{  } - \text{  } \right|$

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

Pix2Pix Generator Loss

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

Summary

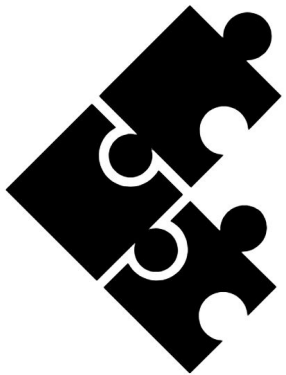
- Pix2Pix adds a Pixel Distance Loss term to the generator loss function
- This loss term calculates the difference between the fake and the real target outputs
- Softly encourages the generator with this additional supervision
    - The target output labels are the supervision
    - Generator essentially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

deeplearning.ai

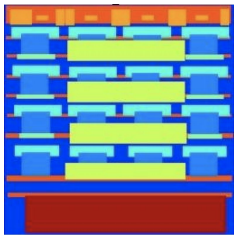
Pix2Pix: Putting It All Together

Outline

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



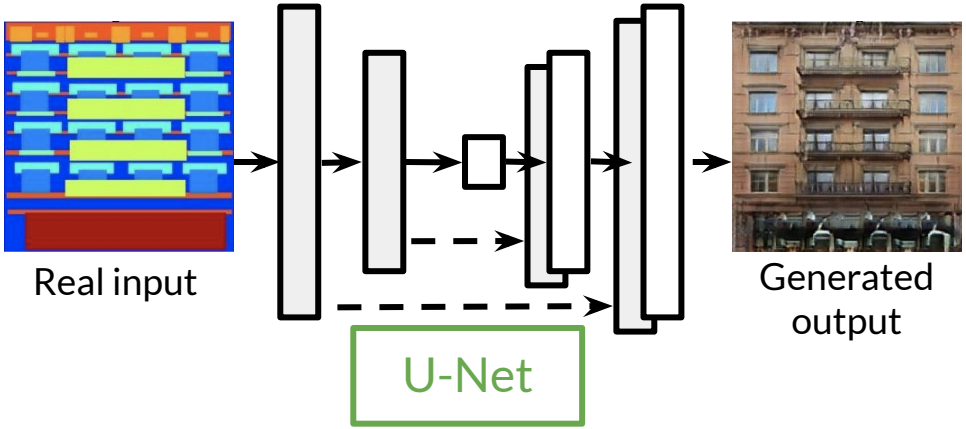
Pix2Pix



Real input

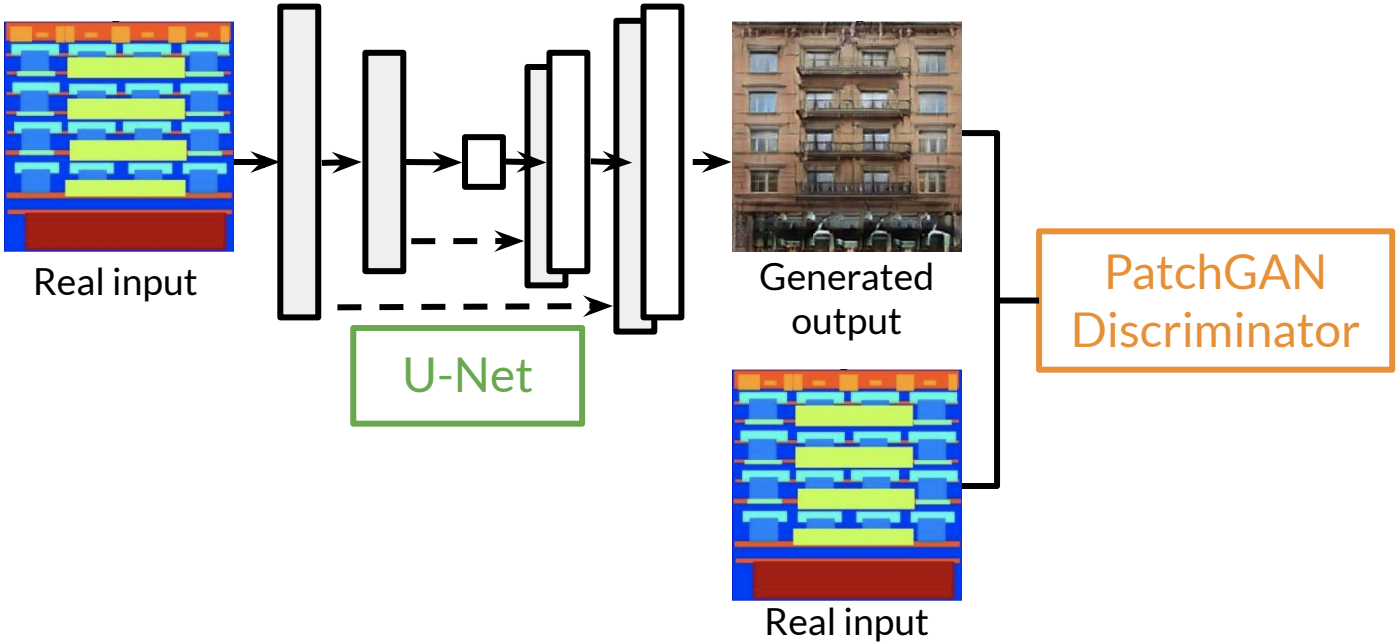
Image available from: https://arxiv.org/abs/1611.07004

Pix2Pix



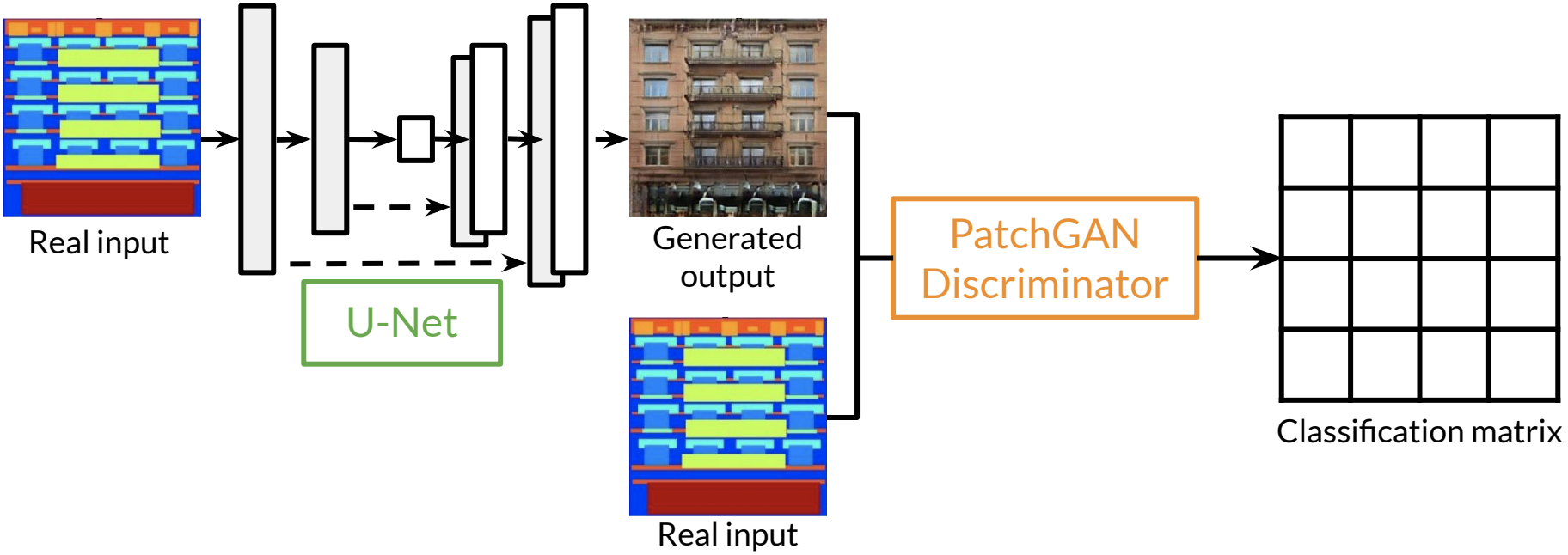
Images available from: https://arxiv.org/abs/1611.07004

Pix2Pix



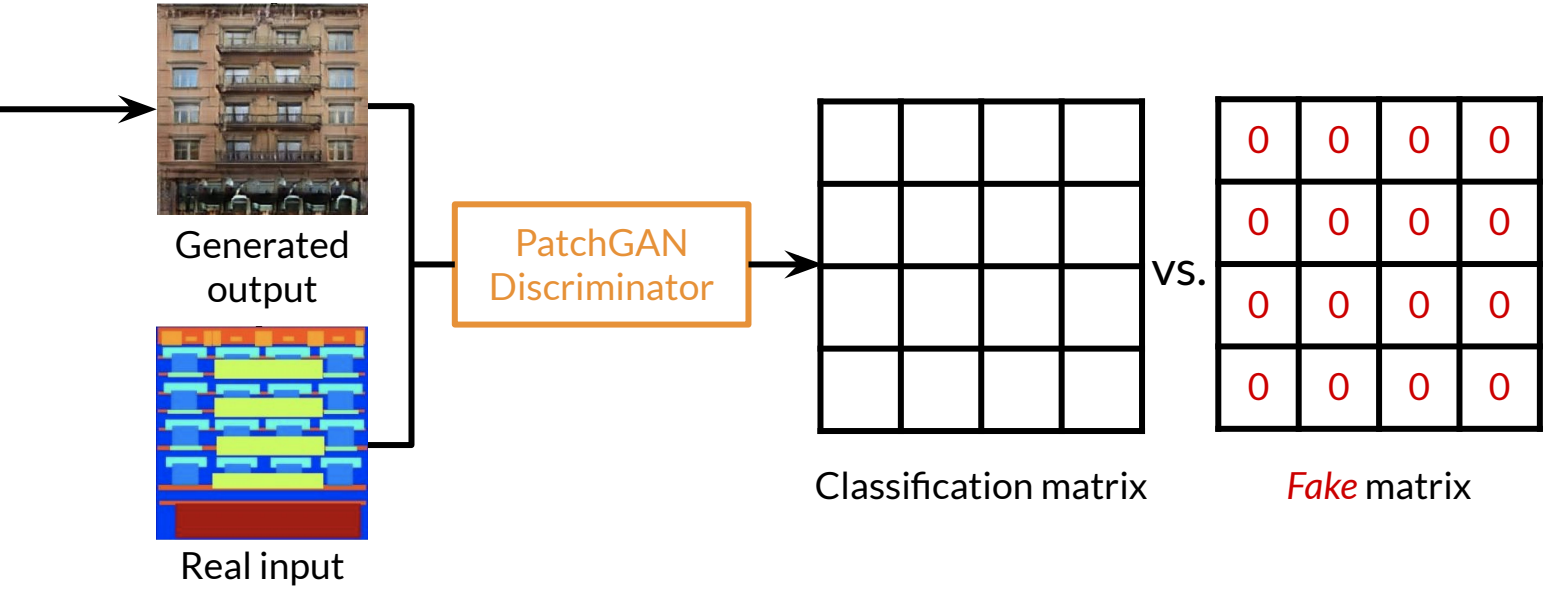
Images available from: https://arxiv.org/abs/1611.07004

Pix2Pix



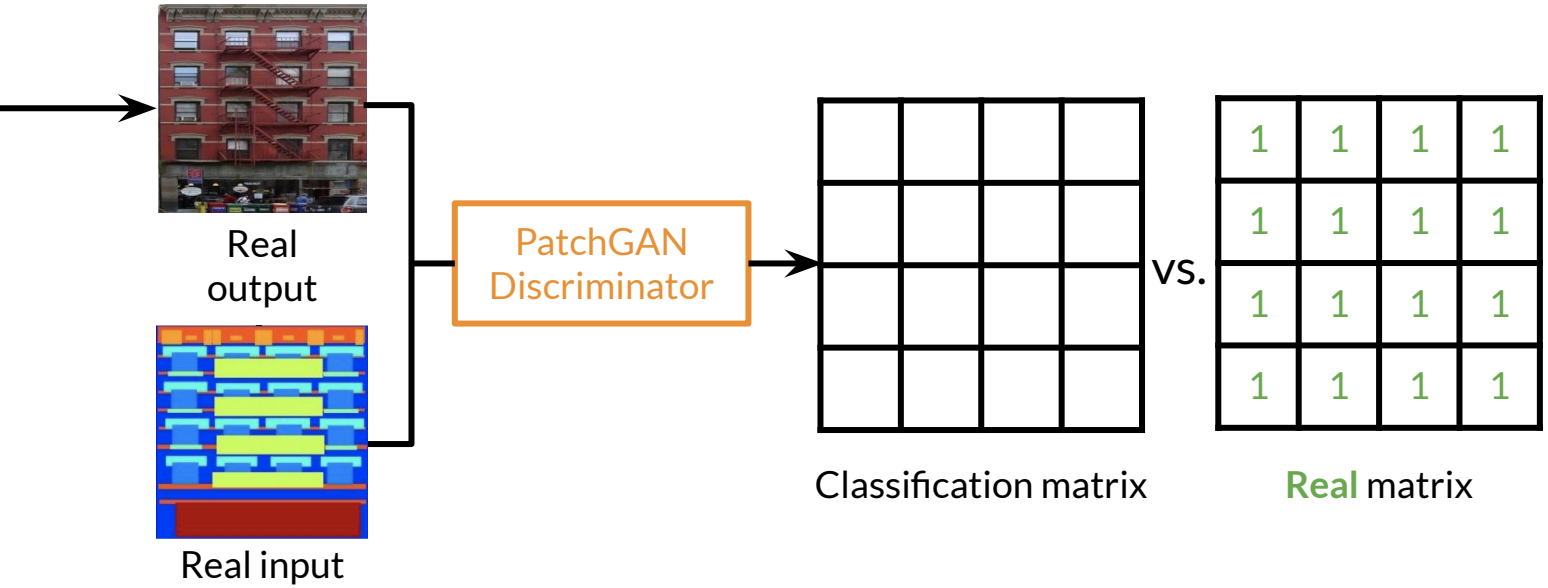
Images available from: https://arxiv.org/abs/1611.07004

Pix2Pix: Discriminator Loss



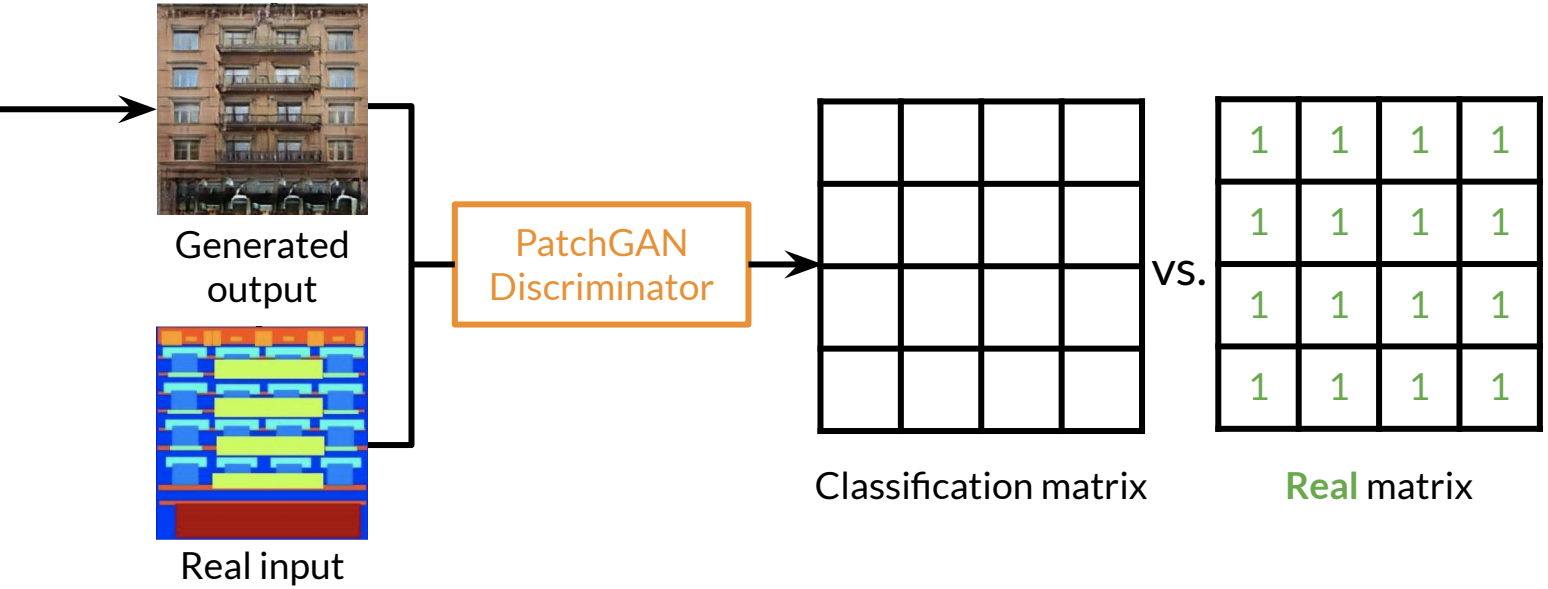
Images available from: https://arxiv.org/abs/1611.07004

Pix2Pix: Discriminator Loss



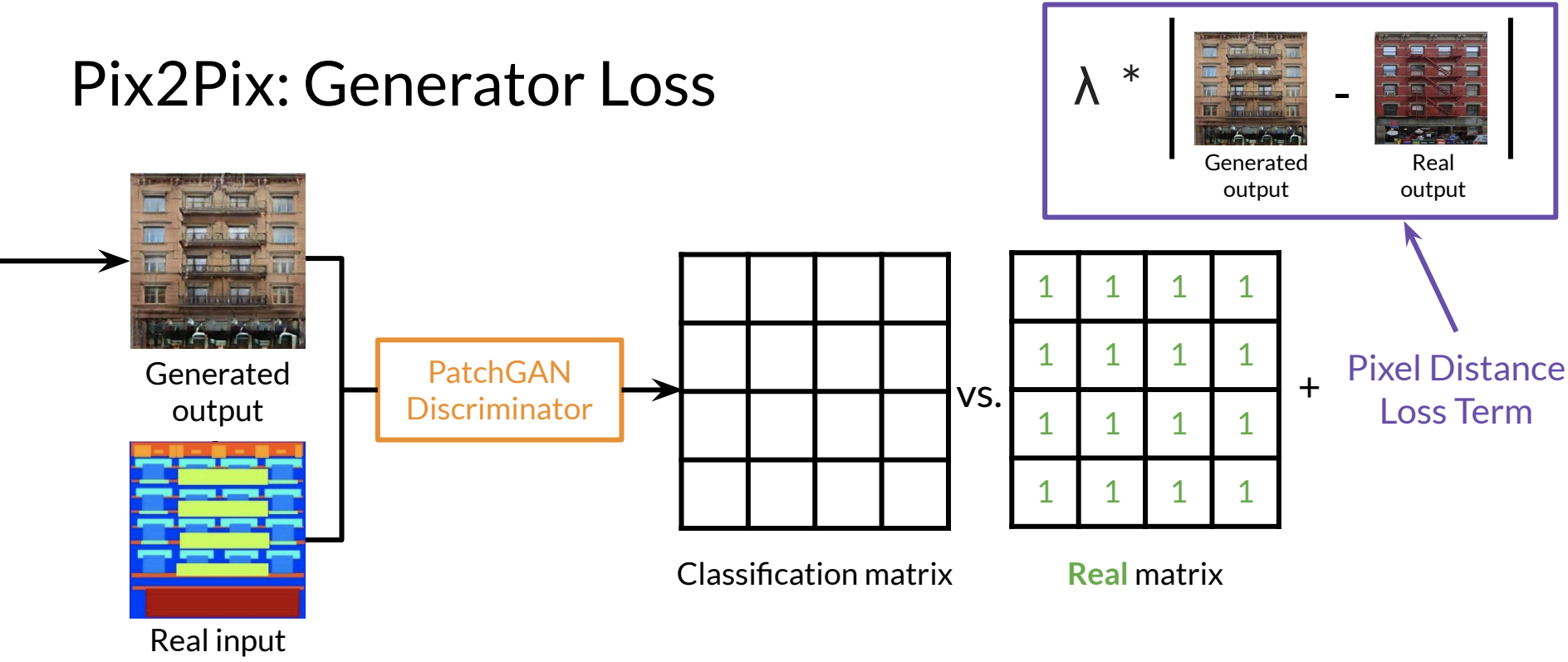
Images available from: https://arxiv.org/abs/1611.07004

Pix2Pix: Generator Loss



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

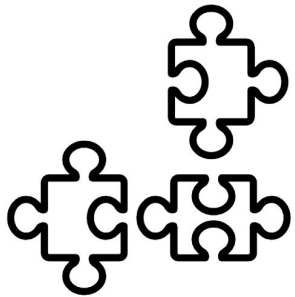
Pix2Pix: Generator Loss



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

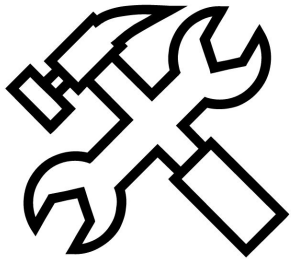


deeplearning.ai

Pix2Pix Advancements

Outline

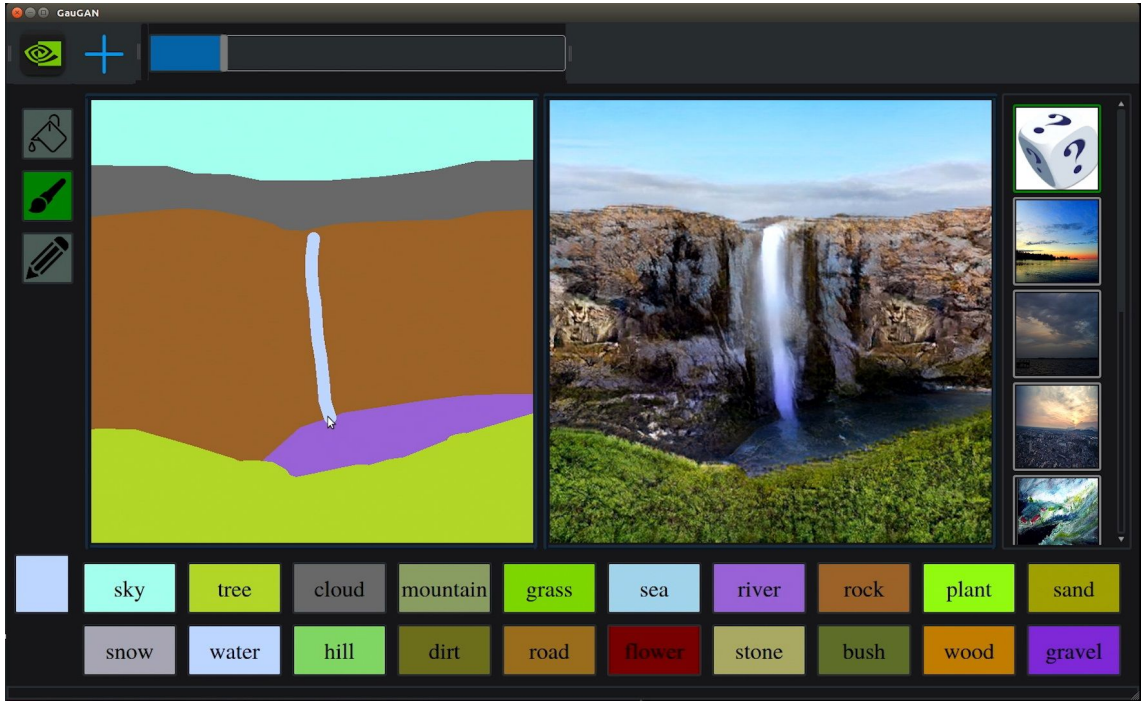
- Improvements and extensions of Pix2Pix for paired image-to-image translation
  - Higher resolution images
  - Image editing



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!