Western
UNIVERSITY · CANADA

AAISS
Amirkabir Artificial Intelligence Summer Summit

Advanced GAN Workshop

Vahid Reza Khazaie

Western University

August, 2020

Overview

- GAN
- DCGAN
- Contex Encoder
- Pix2pix
- CycleGAN
- Progressive GAN
- BigGAN
- StyleGAN

Generative Adversarial Network

Want to generate sample from a complex, high-dimensional training distribution.
No direct way to do this!

Generative Adversarial Network

Start with a simple distribution like a random noise and then learn the transformations to make it like the distribution of real data set.

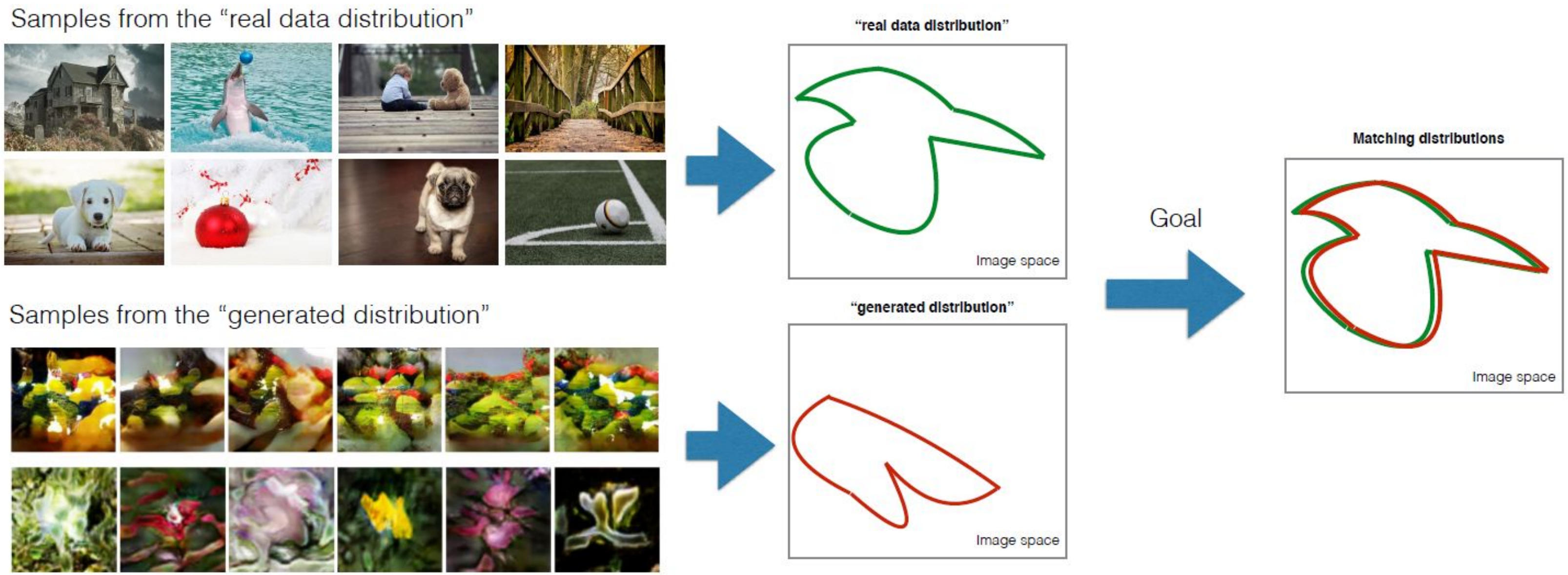
Generative Adversarial Network

Want to learn some complex transformation?

Use Neural Networks!

Generative Adversarial Network

Probability distributions:



Generative Adversarial Network

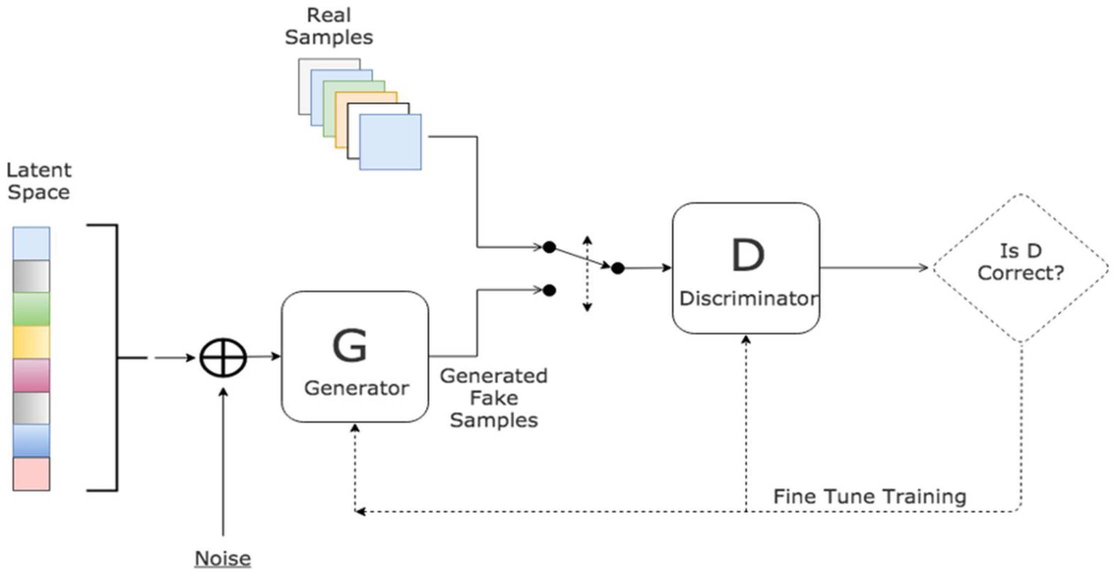
Discriminator

The discriminator, on the other hand, is a common binary classifier. It has two main jobs. First, it categorizes whether its received input comes from the true data distribution or from the Generator distribution. In addition, D also guides G to create more realistic samples by passing to G its gradients. In fact, taking the gradients from D is the only way G optimize its parameters.

Generator

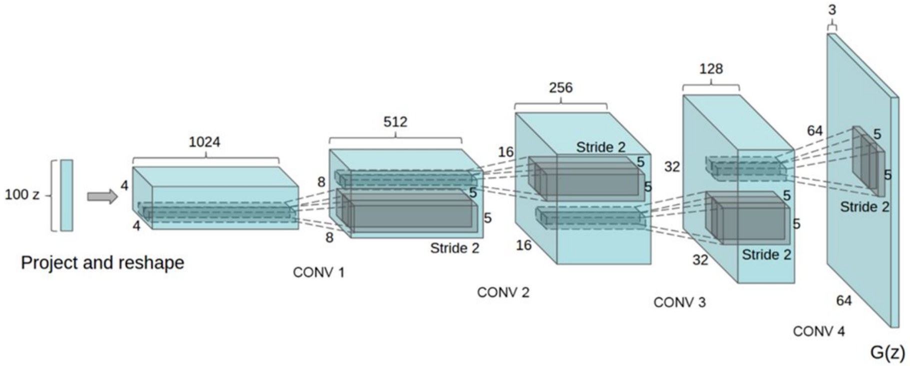
The generator creates samples as an attempt to mimic the ones from the same probability distribution.

Generative Adversarial Network

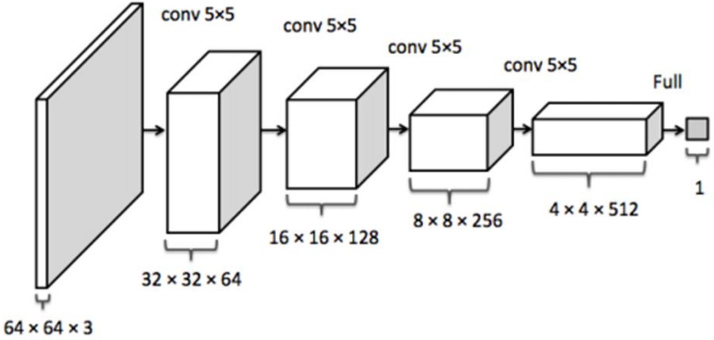


Different Types of GAN

Deep Convolutional Generative Adversarial Network (DCGAN)



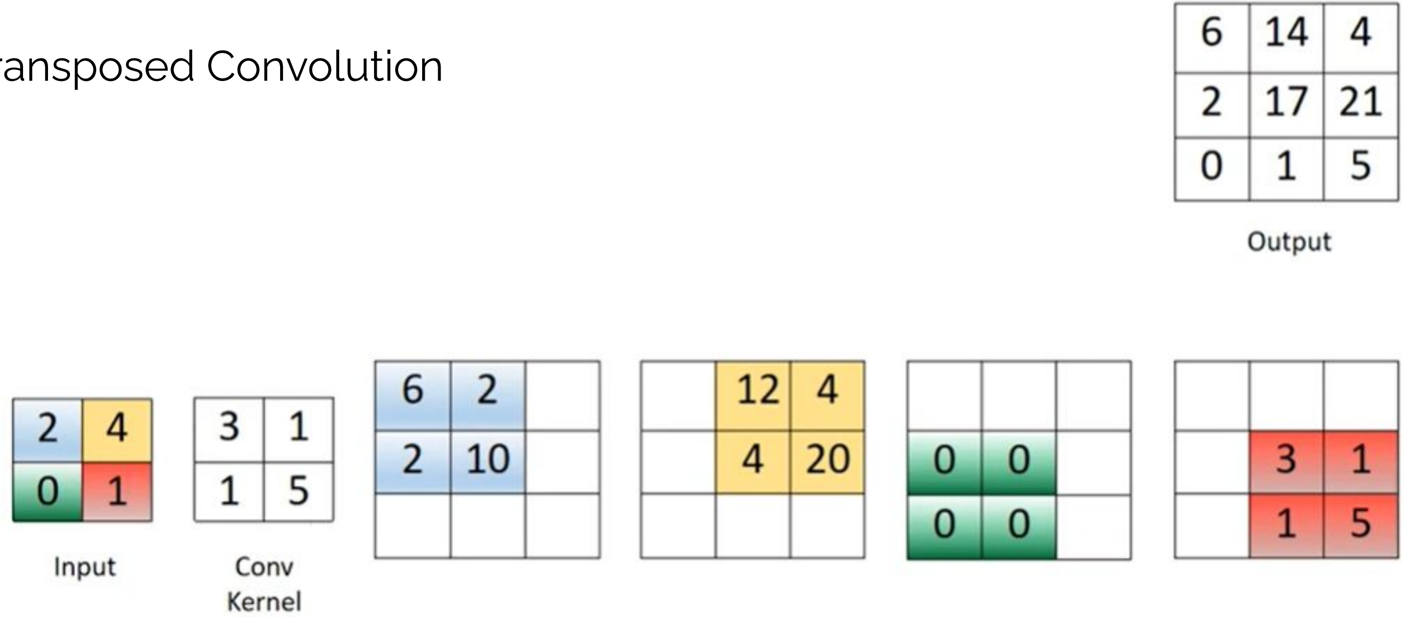
Generator



Discriminator

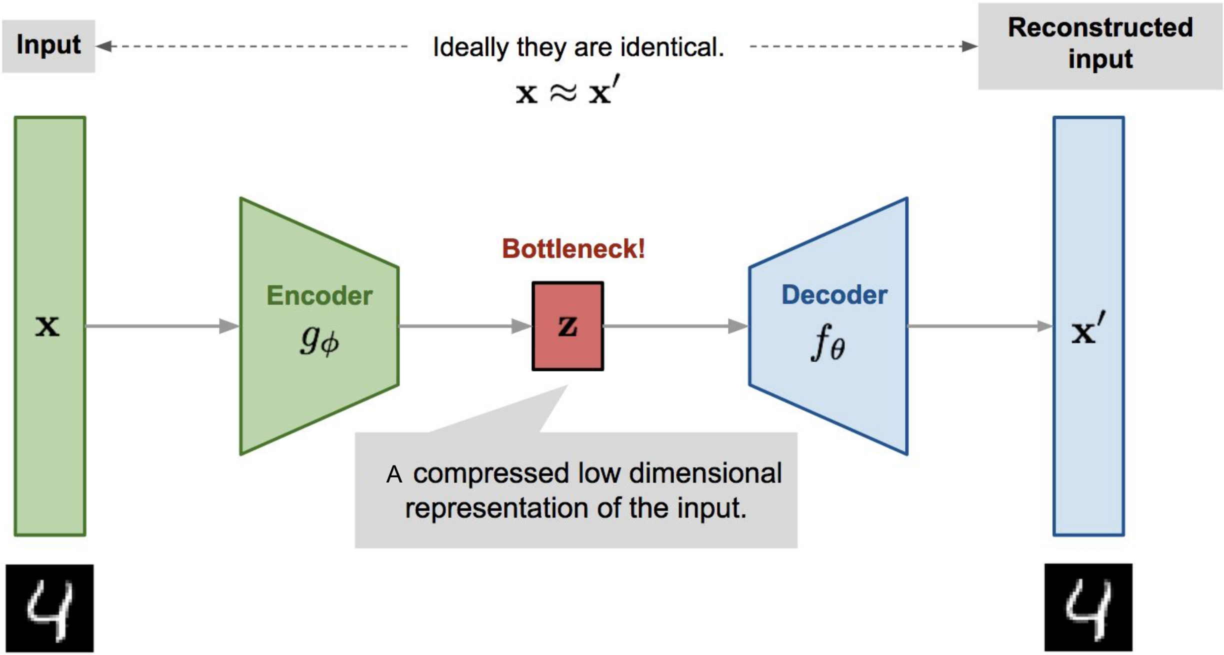
Deep Convolutional Generative Adversarial Network (DCGAN)

Transposed Convolution

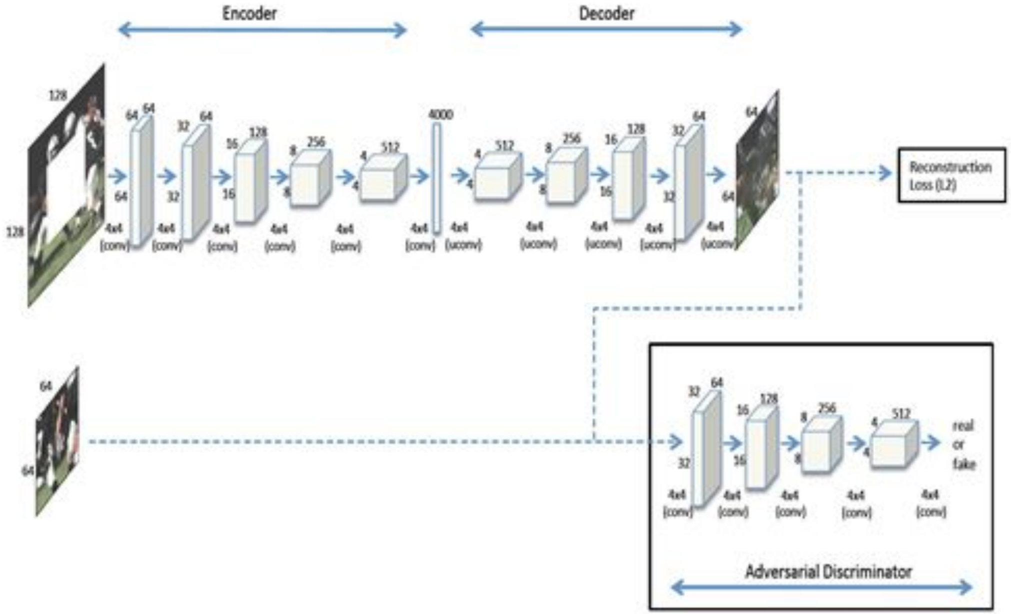


Unsupervised Feature Learning by Inpainting

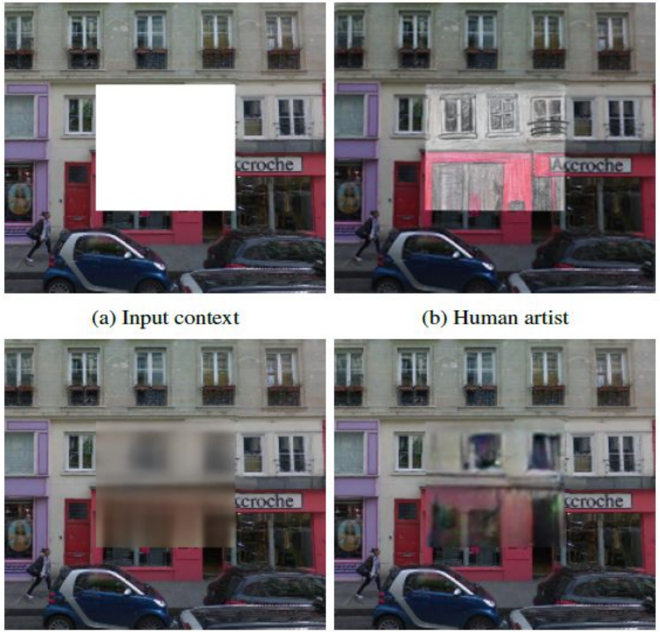
Traditional Autoencoders



Context Encoders



Context Encoders



(a) Input context

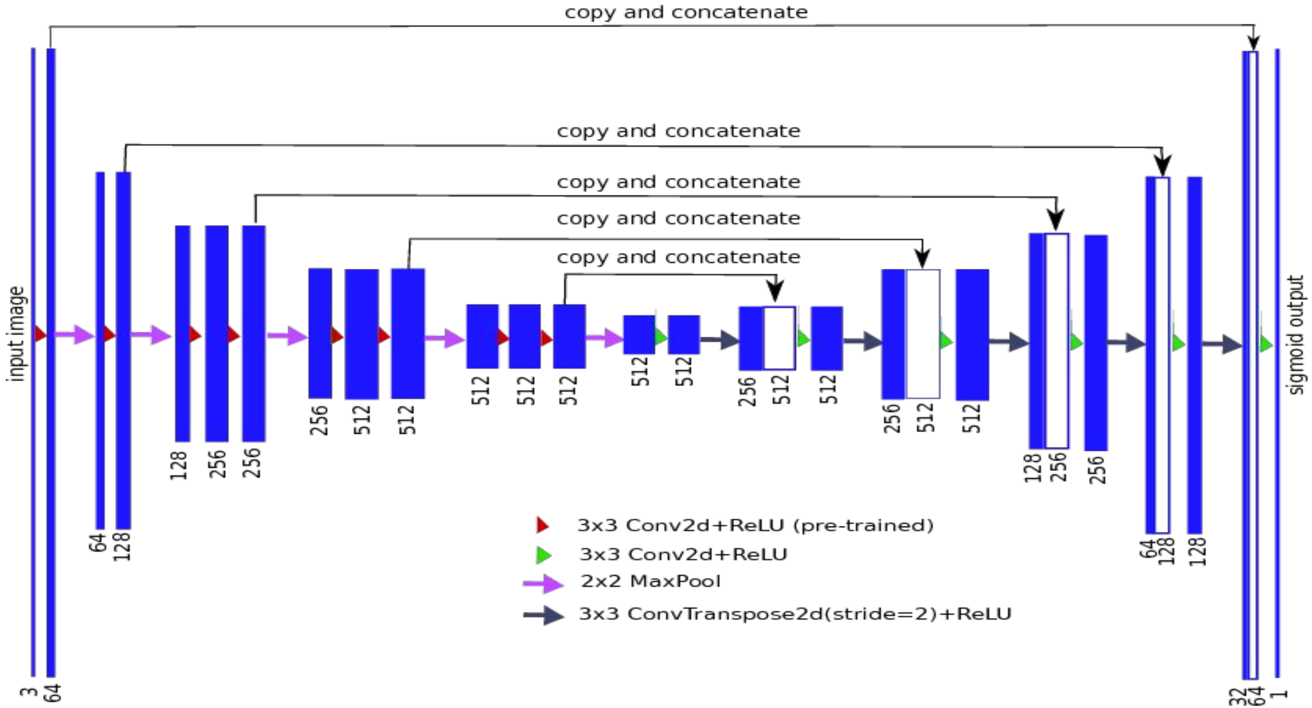
(b) Human artist

(c) Context Encoder ($L2$ loss)

(d) Context Encoder ($L2$ + Adversarial loss)

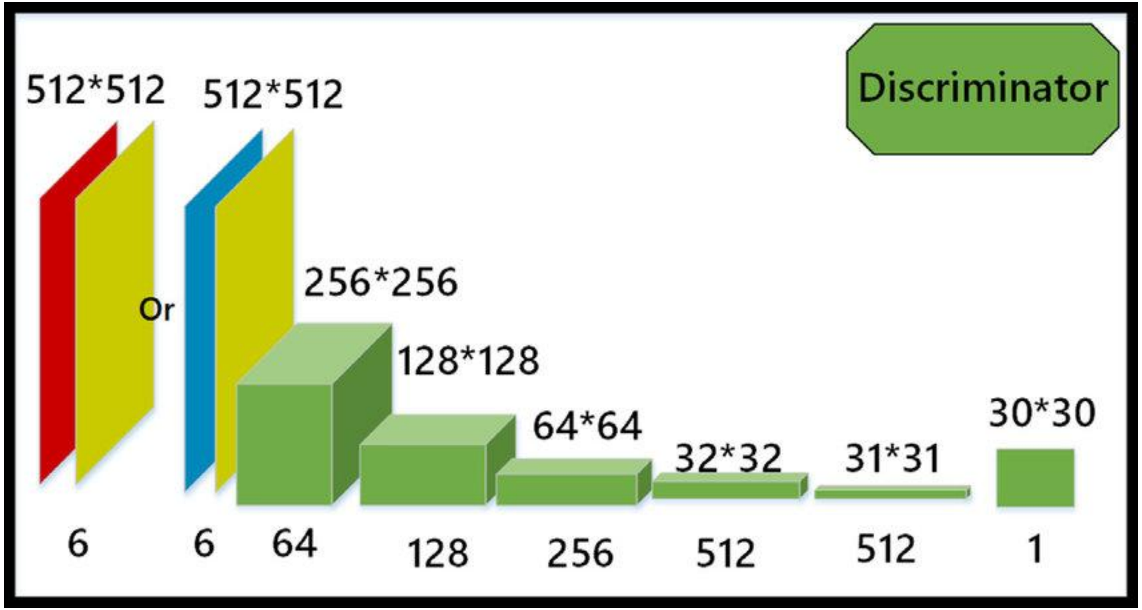
Image-to-Image Translation

U-Net



copy and concatenate
copy and concatenate
copy and concatenate
copy and concatenate
copy and concatenate
input image
sigmoid output
3
64
64
128
128
256
256
256
512
512
512
512
512
512
512
256
512
512
256
512
512
128
256
256
64
128
128
32
64
1
3x3 Conv2d+ReLU (pre-trained)
3x3 Conv2d+ReLU
2x2 MaxPool
3x3 ConvTranspose2d(stride=2)+ReLU

PatchGAN



Discriminator
512*512
512*512
Or
256*256
128*128
64*64
32*32
31*31
30*30
6
6
64
128
256
512
512
1

Image-to-Image Translation with Conditional Adversarial Networks (pix2pix)

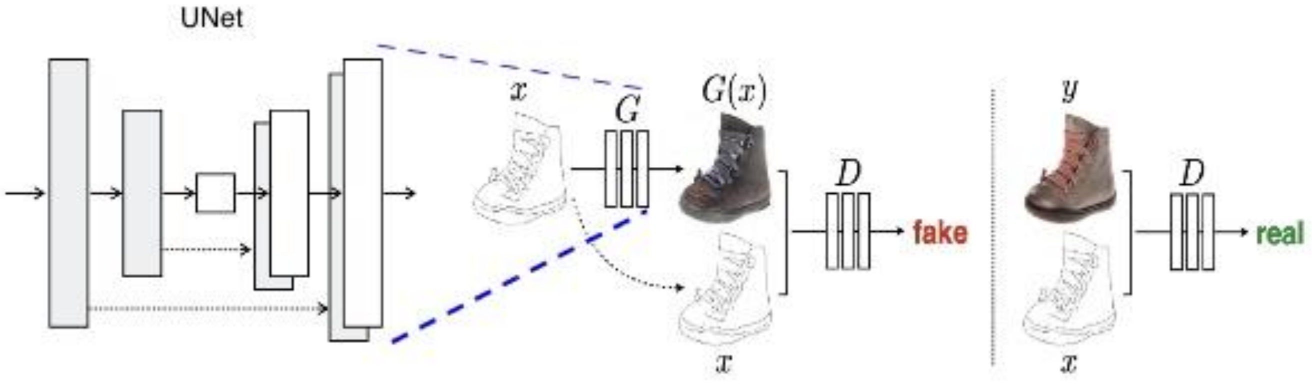


Image-to-Image Translation with Conditional Adversarial Networks (pix2pix)

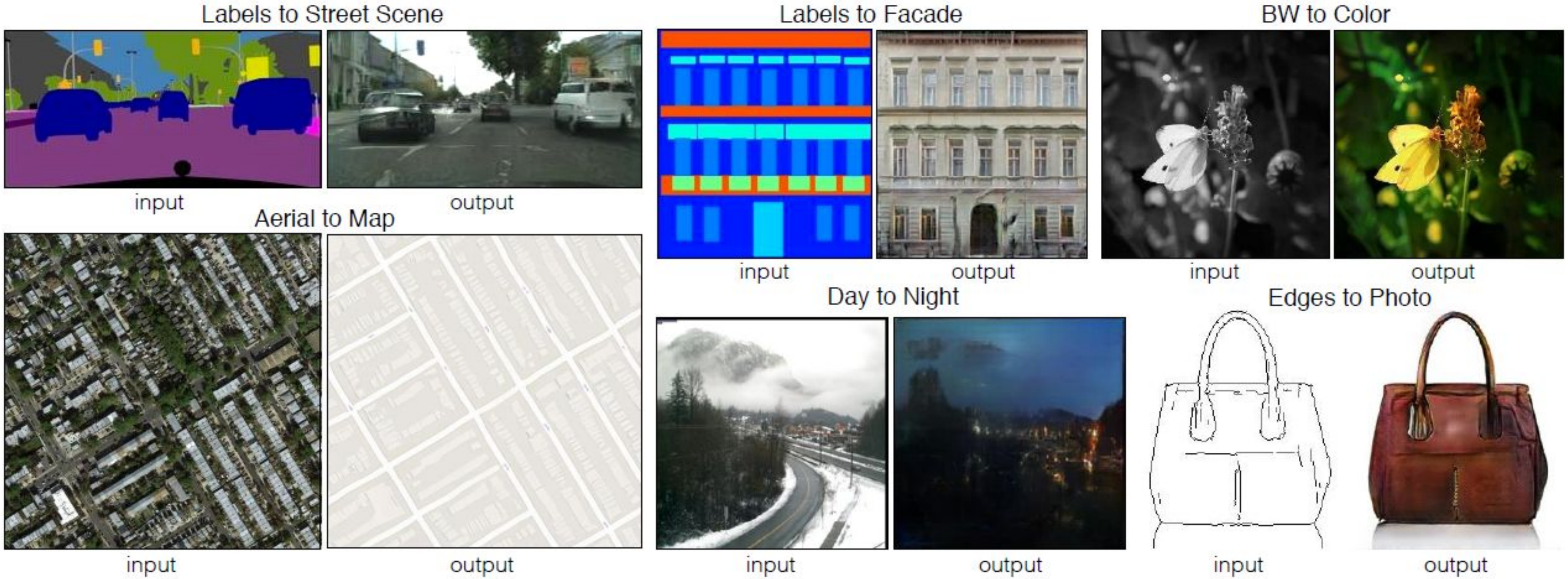
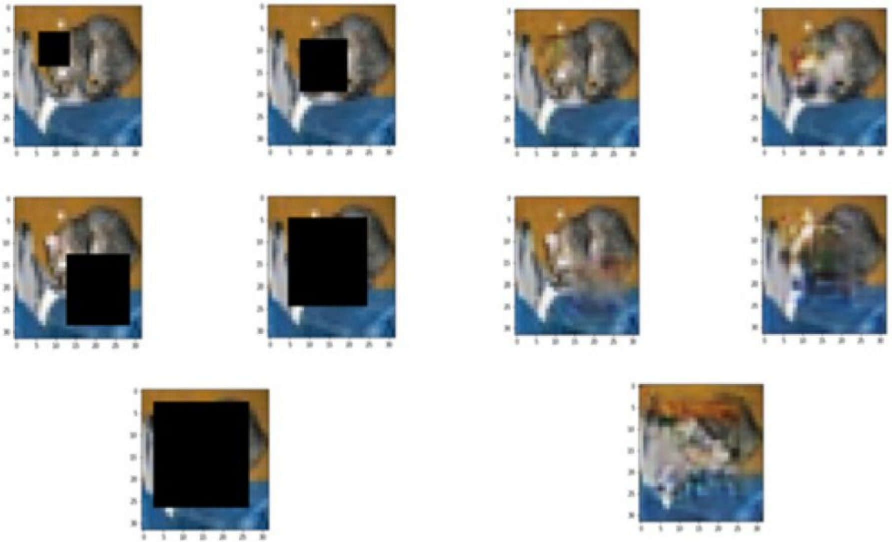


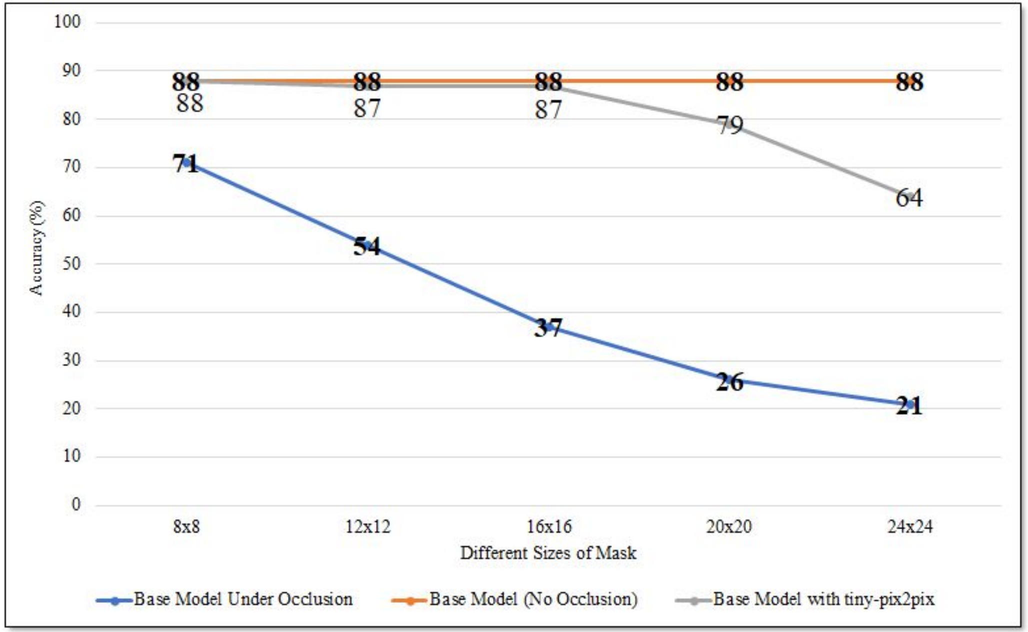
Image-to-Image Translation with Conditional Adversarial Networks (pix2pix)

Occluded Visual Object Recognition Using Deep Conditional Generative Adversarial Nets and Feedforward Convolutional Neural Networks



Khazaie and et al.

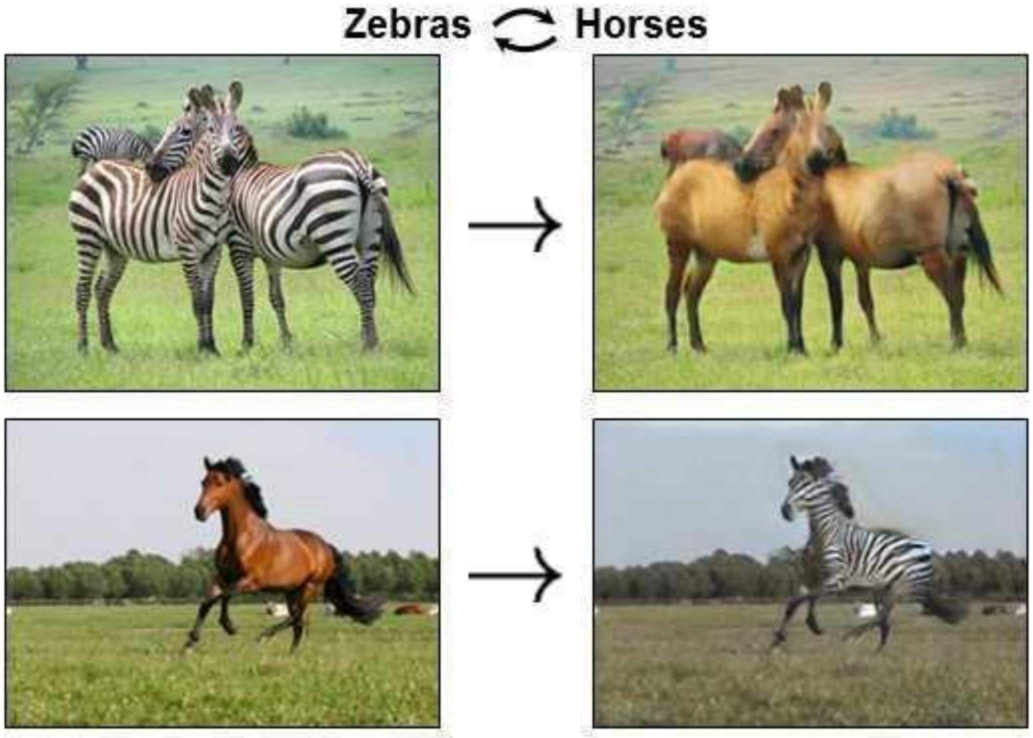
Occluded Visual Object Recognition Using Deep Conditional Generative Adversarial Nets and Feedforward Convolutional Neural Networks



Khazaie and et al.

What if we don't have pairs?

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks



Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

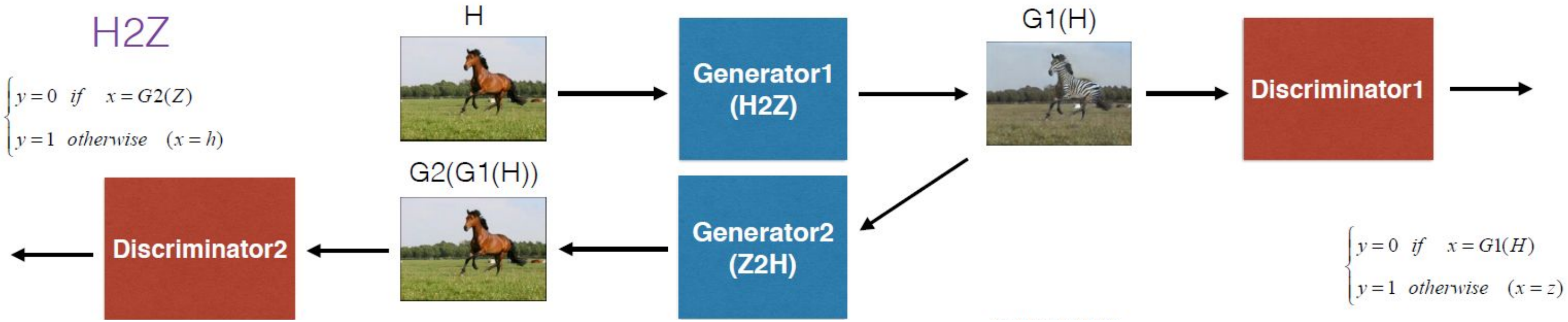
Architecture?

H2Z

$$\begin{cases} y = 0 & \text{if} \quad x = G2(Z) \\ y = 1 & \text{otherwise} \quad (x = h) \end{cases}$$

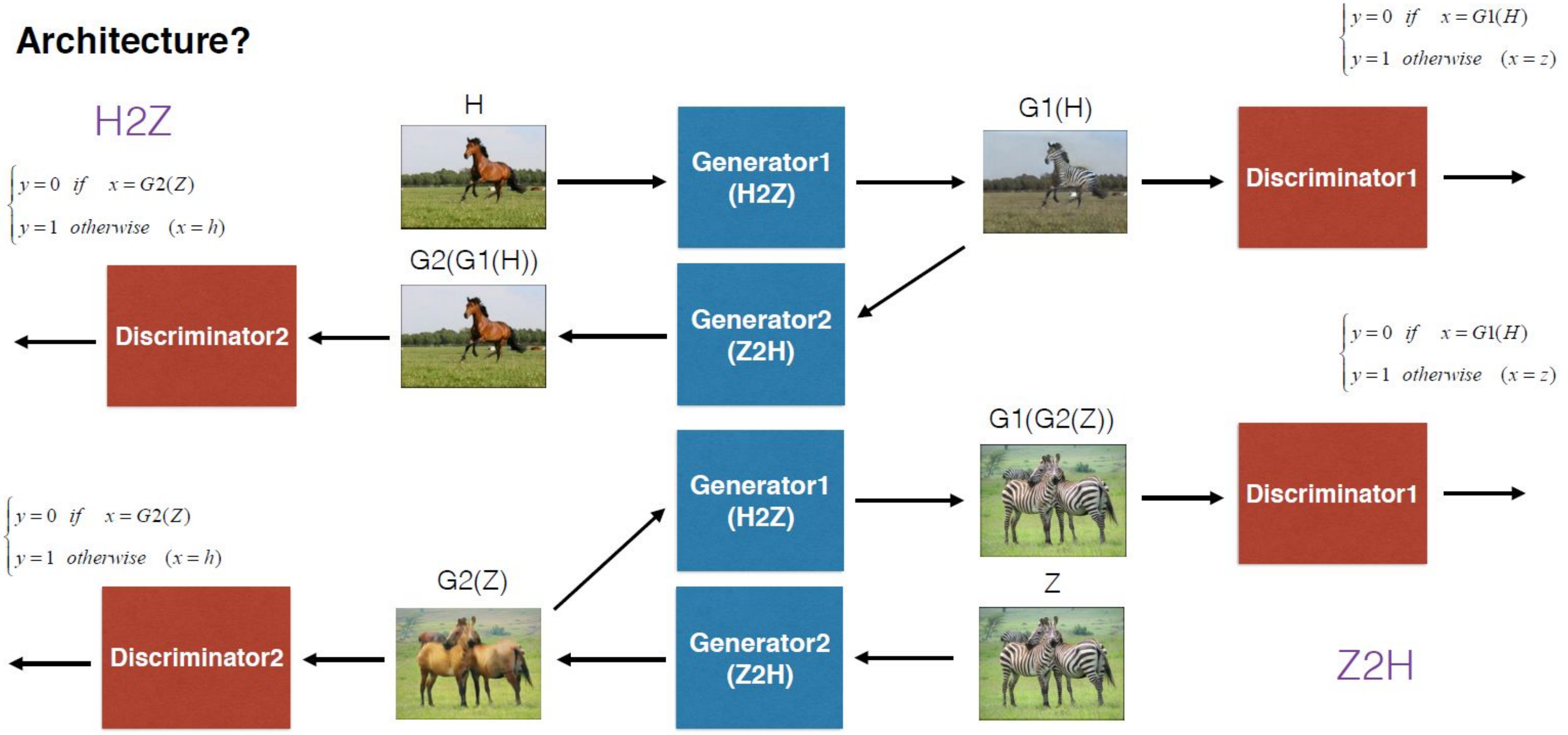
$$\begin{cases} y = 0 & \text{if} \quad x = G1(H) \\ y = 1 & \text{otherwise} \quad (x = z) \end{cases}$$

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks



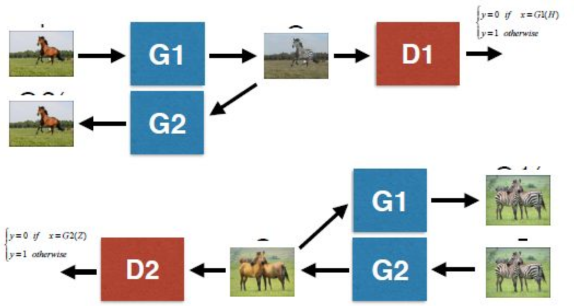
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Architecture?



Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Loss to minimize?



$$J^{(D1)} = -\frac{1}{m_{real}}\sum_{i=1}^{m_{real}} \log(D1(z^{(i)})) - \frac{1}{m_{gen}}\sum_{i=1}^{m_{gen}} \log(1 - D1(G1(H^{(i)})))$$

$$J^{(G1)} = -\frac{1}{m_{gen}}\sum_{i=1}^{m_{gen}} \log(D1(G1(H^{(i)})))$$

$$J^{(D2)} = -\frac{1}{m_{real}}\sum_{i=1}^{m_{real}} \log(D2(h^{(i)})) - \frac{1}{m_{gen}}\sum_{i=1}^{m_{gen}} \log(1 - D2(G2(Z^{(i)})))$$

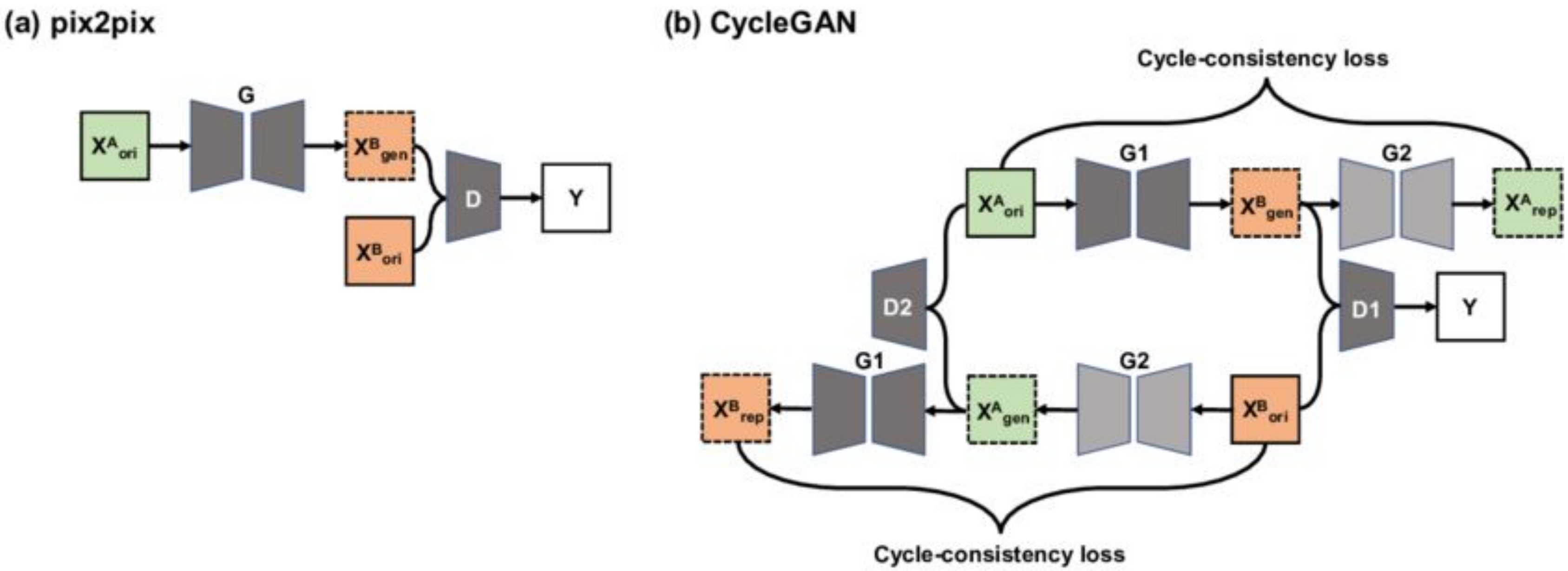
$$J^{(G2)} = -\frac{1}{m_{gen}}\sum_{i=1}^{m_{gen}} \log(D2(G2(Z^{(i)})))$$

$$J^{cycle} = \frac{1}{m_{gen}}\sum_{i=1}^{m_{gen}} \| G2(G1(H^{(i)}) - H^{(i)} \|_1 + \frac{1}{m_{gen}}\sum_{i=1}^{m_{gen}} \| G1(G2(Z^{(i)}) - Z^{(i)} \|_1$$

$$\boxed{J = J^{(D1)} + J^{(G1)} + J^{(D2)} + J^{(G2)} + \lambda J^{cycle}}$$

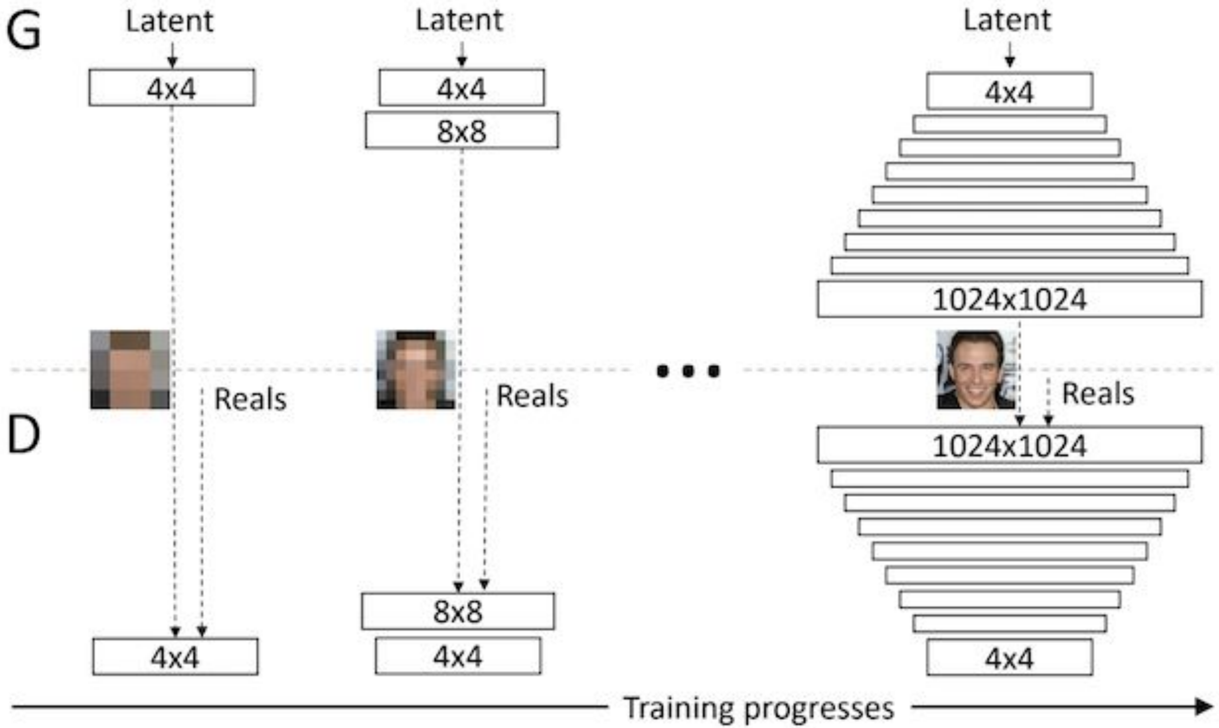
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Pix2pix vs CycleGAN

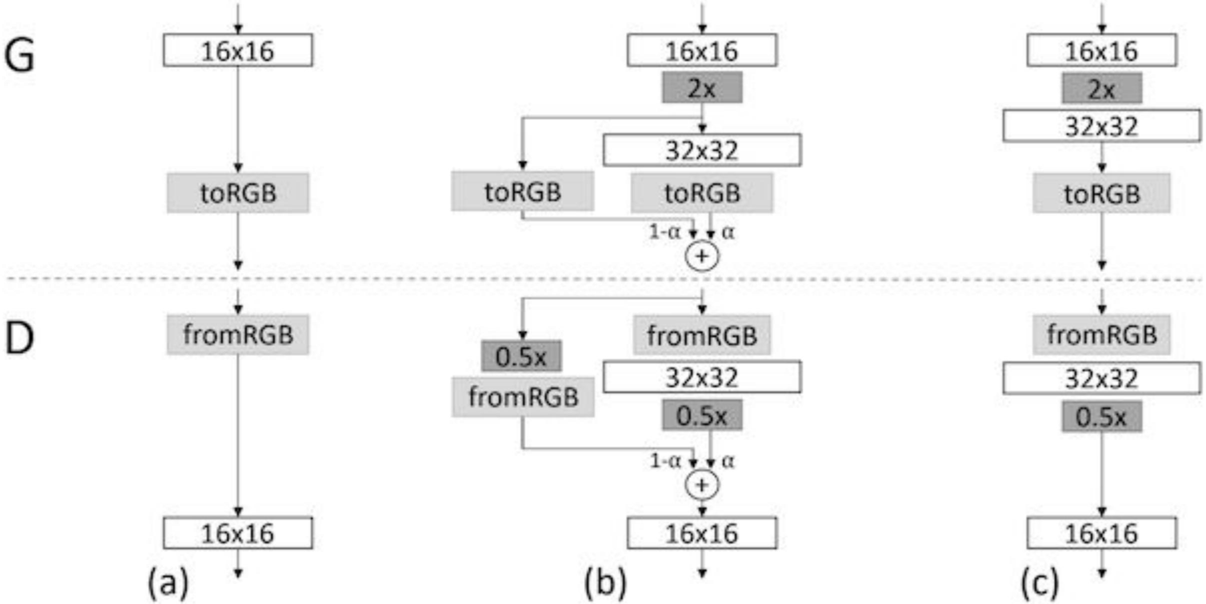


Progressive Growing GAN

Progressive Growing GAN



Progressive Growing GAN



Progressive Growing GAN

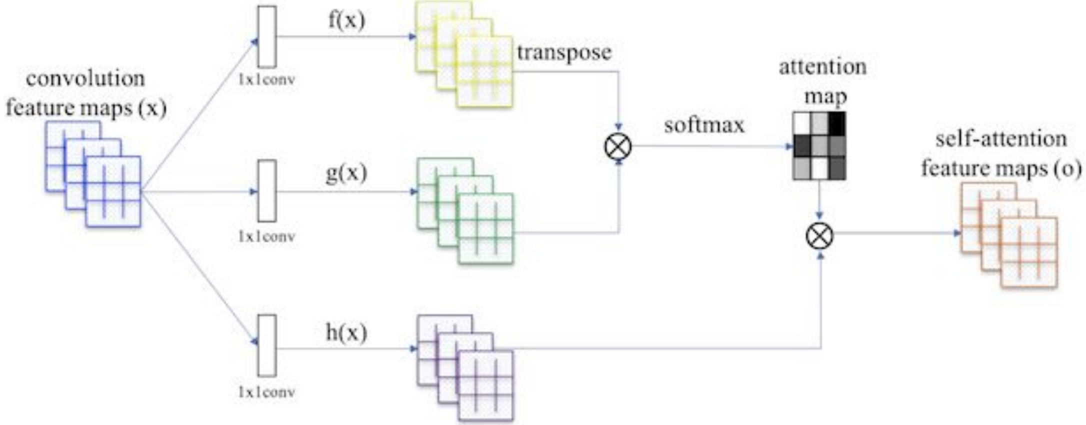
Generator	Act.	Output shape	Params
Latent vector	–	512 × 1 × 1	–
Conv 4 × 4	LReLU	512 × 4 × 4	4.2M
Conv 3 × 3	LReLU	512 × 4 × 4	2.4M
Upsample	–	512 × 8 × 8	–
Conv 3 × 3	LReLU	512 × 8 × 8	2.4M
Conv 3 × 3	LReLU	512 × 8 × 8	2.4M
Upsample	–	512 × 16 × 16	–
Conv 3 × 3	LReLU	512 × 16 × 16	2.4M
Conv 3 × 3	LReLU	512 × 16 × 16	2.4M
Upsample	–	512 × 32 × 32	–
Conv 3 × 3	LReLU	512 × 32 × 32	2.4M
Conv 3 × 3	LReLU	512 × 32 × 32	2.4M
Upsample	–	512 × 64 × 64	–
Conv 3 × 3	LReLU	256 × 64 × 64	1.2M
Conv 3 × 3	LReLU	256 × 64 × 64	590k
Upsample	–	256 × 128 × 128	–
Conv 3 × 3	LReLU	128 × 128 × 128	295k
Conv 3 × 3	LReLU	128 × 128 × 128	148k
Upsample	–	128 × 256 × 256	–
Conv 3 × 3	LReLU	64 × 256 × 256	74k
Conv 3 × 3	LReLU	64 × 256 × 256	37k
Upsample	–	64 × 512 × 512	–
Conv 3 × 3	LReLU	32 × 512 × 512	18k
Conv 3 × 3	LReLU	32 × 512 × 512	9.2k
Upsample	–	32 × 1024 × 1024	–
Conv 3 × 3	LReLU	16 × 1024 × 1024	4.6k
Conv 3 × 3	LReLU	16 × 1024 × 1024	2.3k
Conv 1 × 1	linear	3 × 1024 × 1024	51
Total trainable parameters			**23.1M**

Discriminator	Act.	Output shape	Params
Input image	–	3 × 1024 × 1024	–
Conv 1 × 1	LReLU	16 × 1024 × 1024	64
Conv 3 × 3	LReLU	16 × 1024 × 1024	2.3k
Conv 3 × 3	LReLU	32 × 1024 × 1024	4.6k
Downsample	–	32 × 512 × 512	–
Conv 3 × 3	LReLU	32 × 512 × 512	9.2k
Conv 3 × 3	LReLU	64 × 512 × 512	18k
Downsample	–	64 × 256 × 256	–
Conv 3 × 3	LReLU	64 × 256 × 256	37k
Conv 3 × 3	LReLU	128 × 256 × 256	74k
Downsample	–	128 × 128 × 128	–
Conv 3 × 3	LReLU	128 × 128 × 128	148k
Conv 3 × 3	LReLU	256 × 128 × 128	295k
Downsample	–	256 × 64 × 64	–
Conv 3 × 3	LReLU	256 × 64 × 64	590k
Conv 3 × 3	LReLU	512 × 64 × 64	1.2M
Downsample	–	512 × 32 × 32	–
Conv 3 × 3	LReLU	512 × 32 × 32	2.4M
Conv 3 × 3	LReLU	512 × 32 × 32	2.4M
Downsample	–	512 × 16 × 16	–
Conv 3 × 3	LReLU	512 × 16 × 16	2.4M
Conv 3 × 3	LReLU	512 × 16 × 16	2.4M
Downsample	–	512 × 8 × 8	–
Conv 3 × 3	LReLU	512 × 8 × 8	2.4M
Conv 3 × 3	LReLU	512 × 8 × 8	2.4M
Downsample	–	512 × 4 × 4	–
Minibatch stddev	–	513 × 4 × 4	–
Conv 3 × 3	LReLU	512 × 4 × 4	2.4M
Conv 4 × 4	LReLU	512 × 1 × 1	4.2M
Fully-connected	linear	1 × 1 × 1	513
Total trainable parameters			**23.1M**

Progressive Growing GAN

Big GAN

1. Self-Attention Module
2. Adversarial Hinge Loss
3. Class-conditional batch normalization
4. Update discriminator more
5. Larger Batch Size
6. Much more parameters
7. Skip-z connection
8. Truncation Trick
9. Weight Initialization



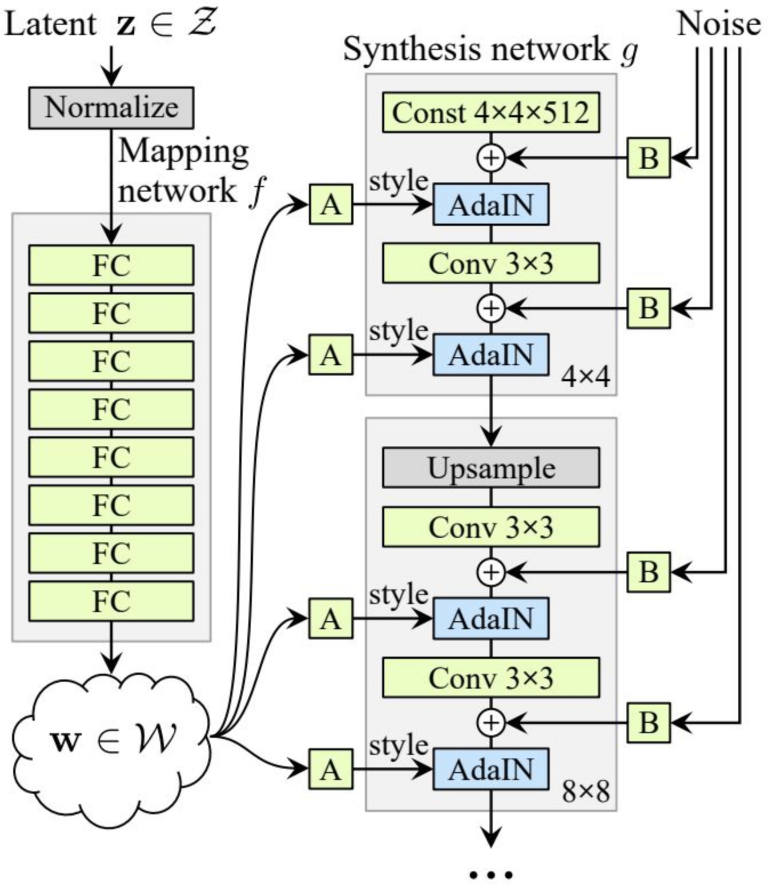
BigGAN

(a) 128×128 (b) 256×256 (c) 512×512

BigGAN

StyleGAN

1. **Progressive Growing Method**
2. **Mapping Network**
3. **AdaIn**
4. **Constant Latent Point**
5. **Noise Addition**
6. **Mixing regularization**

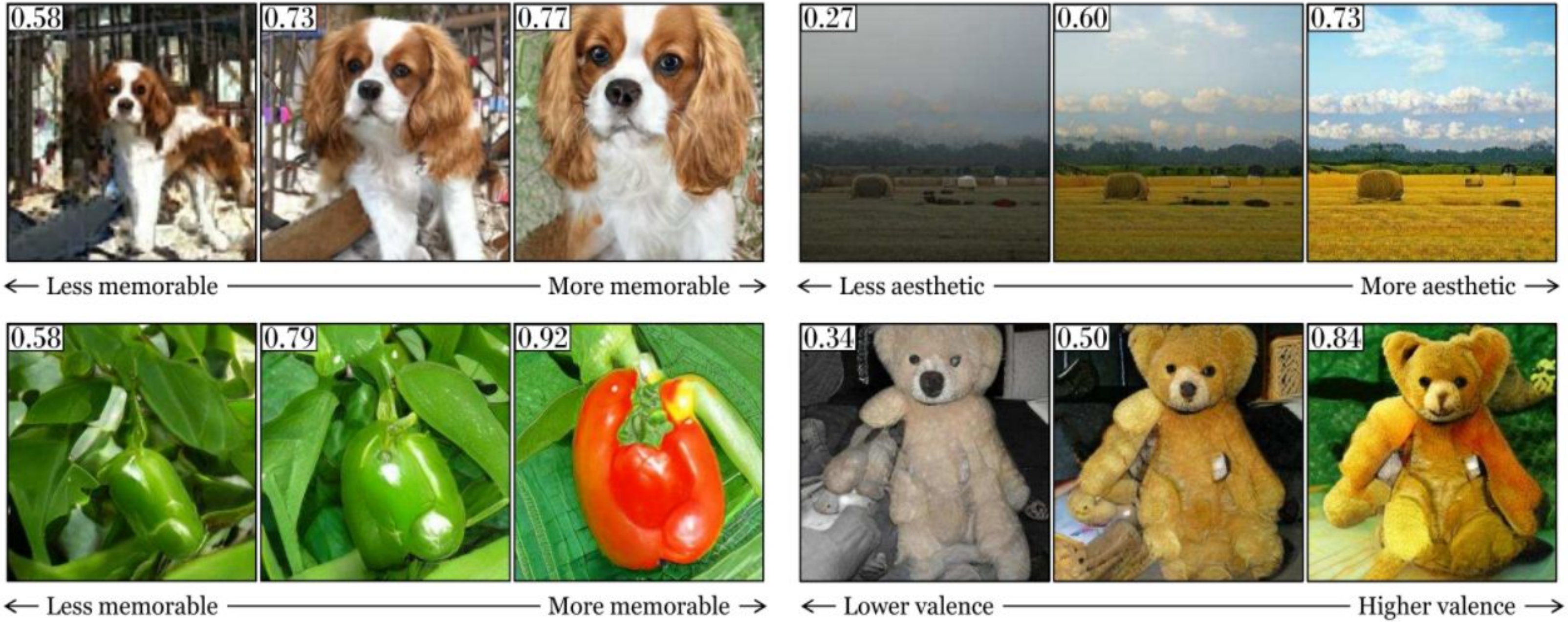


StyleGAN

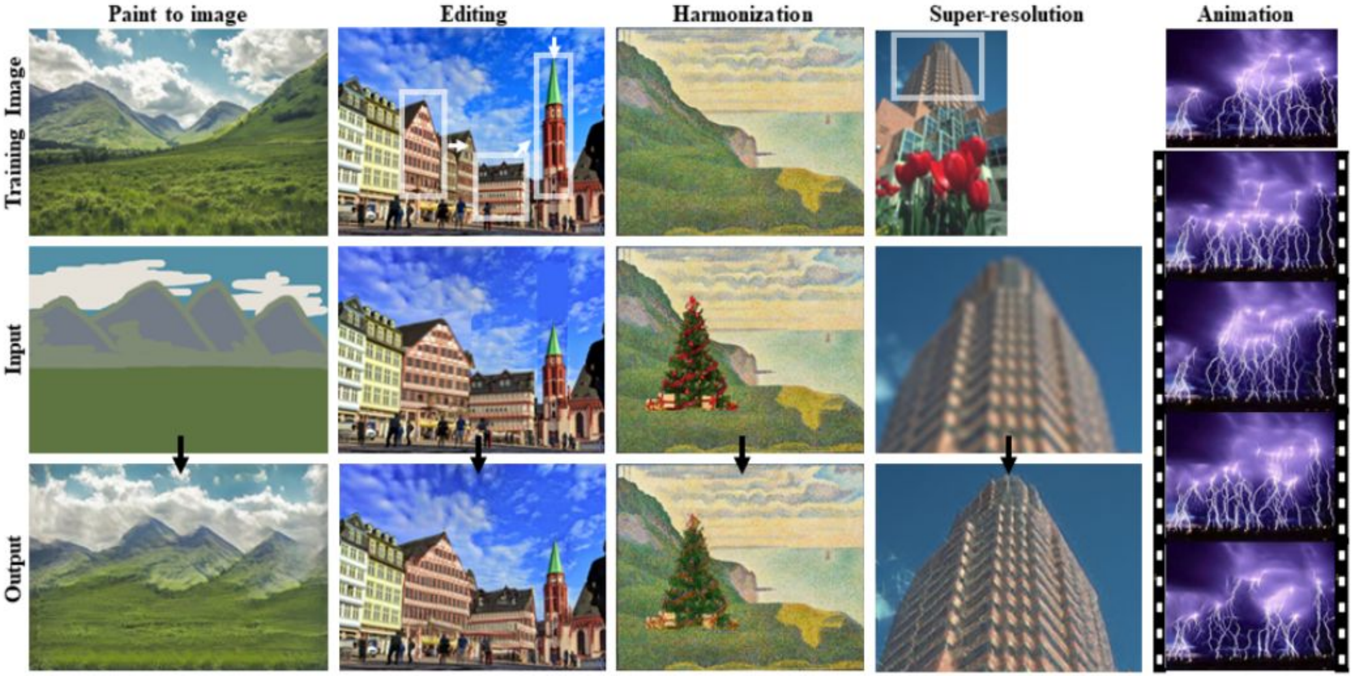
StyleGAN

Still More?

GANalyze: Toward Visual Definitions of Cognitive Image Properties



SinGAN: Learning a Generative Model from a Single Natural Image



Conclusion

- What is GAN?
- Different types of GAN
- GAN Applications

Thank you everyone!

Vahid Reza Khazaie

Contact: vkhazaie@uwo.ca

References

- CS230 Stanford University
- MIT Introduction to Deep Learning Course
- Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks Paper
- Context Encoders: Feature Learning by Inpainting
- Image-to-Image Translation with Conditional Adversarial Networks Paper
- Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks
- Progressive Growing of GANs for Improved Quality, Stability, and Variation
- Large Scale GAN Training for High Fidelity Natural Image Synthesis
- A Style-Based Generator Architecture for Generative Adversarial Networks