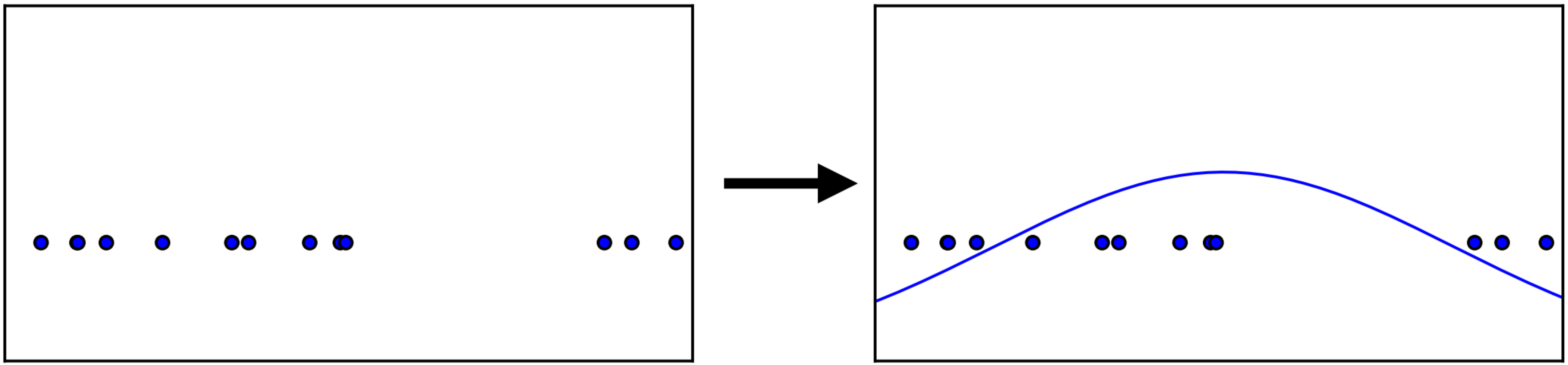
An Introduction to Generative Adversarial Networks

Sagie Benaim
Tel Aviv University

Generative Modeling: Density Estimation



Training Data

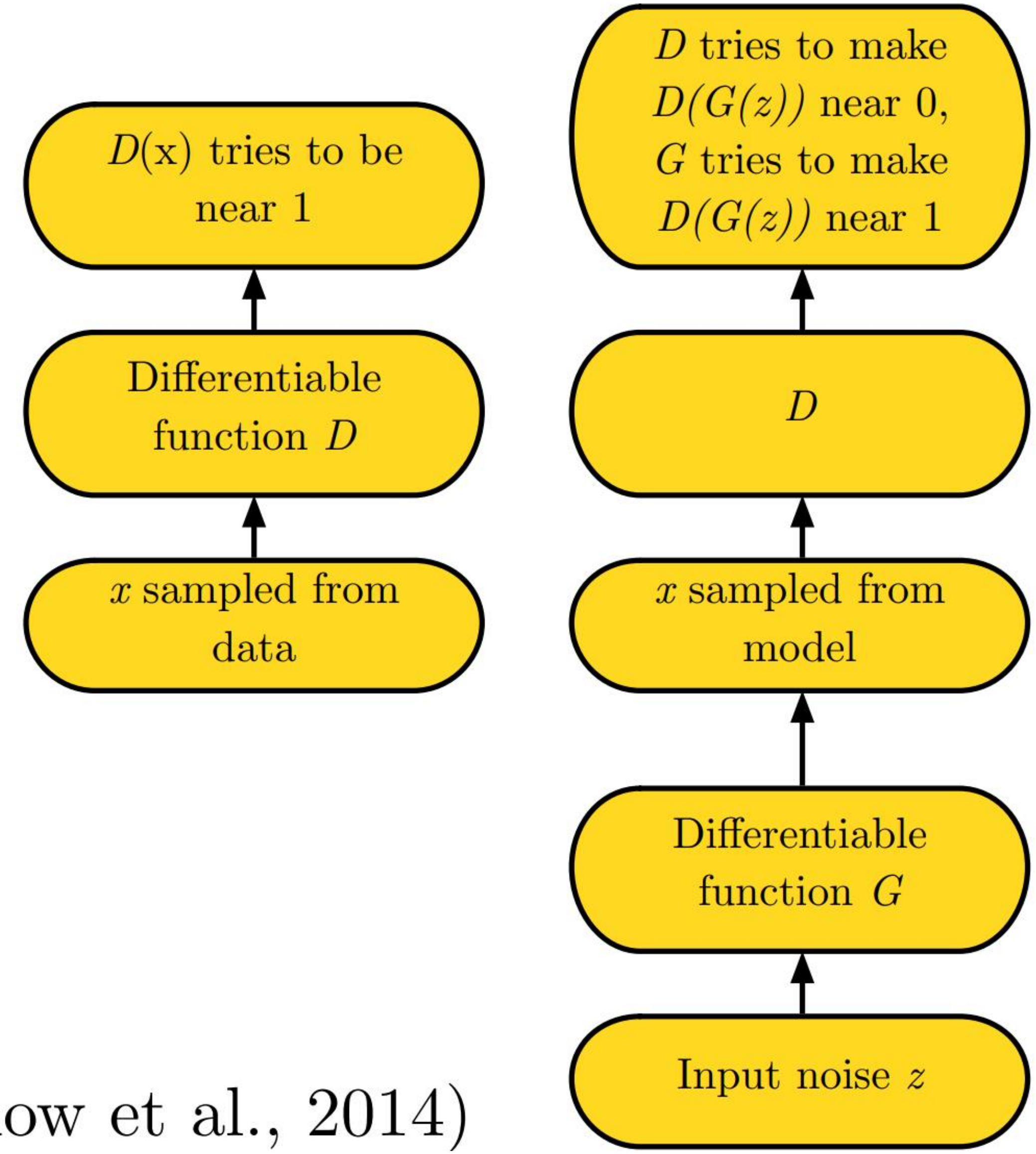
Density Function

Generative Modeling:
Sample Generation

Training Data
(CelebA)

Sample Generator
(Karras et al, 2017)

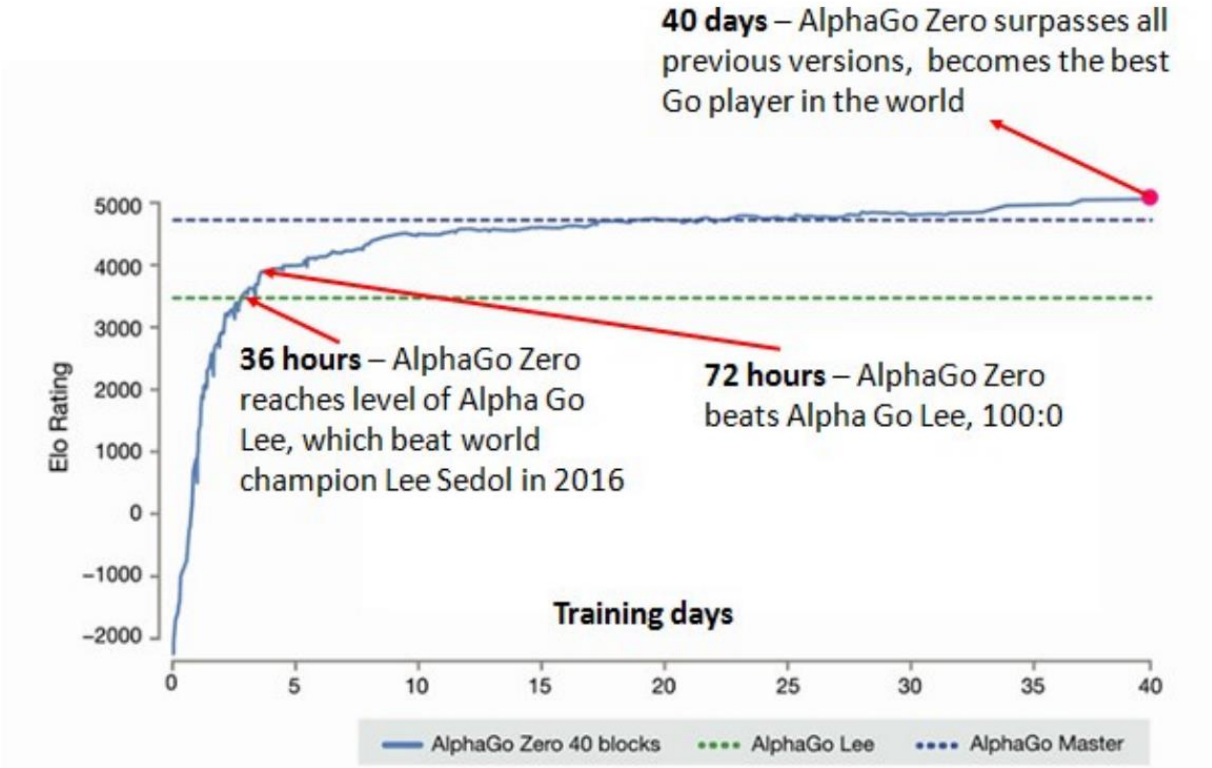
Adversarial Nets Framework



(Goodfellow et al., 2014)

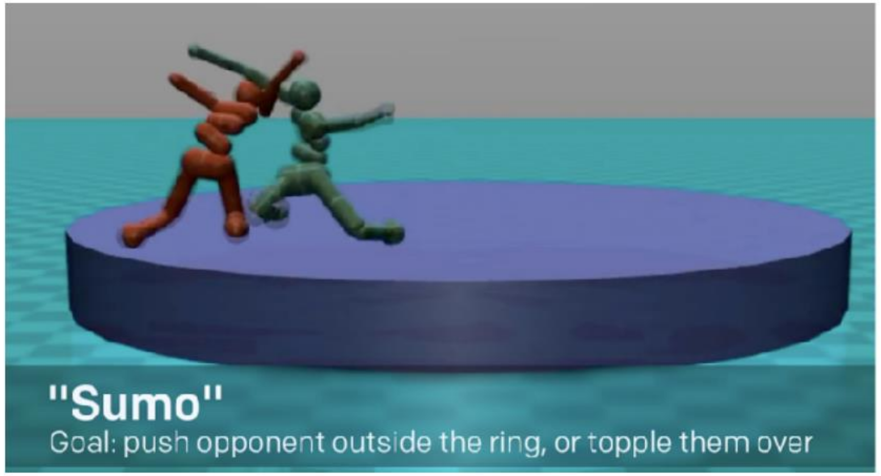
Self-Play

1959: Arthur Samuel's checkers agent



(Silver et al, 2017)

(OpenAI, 2017)



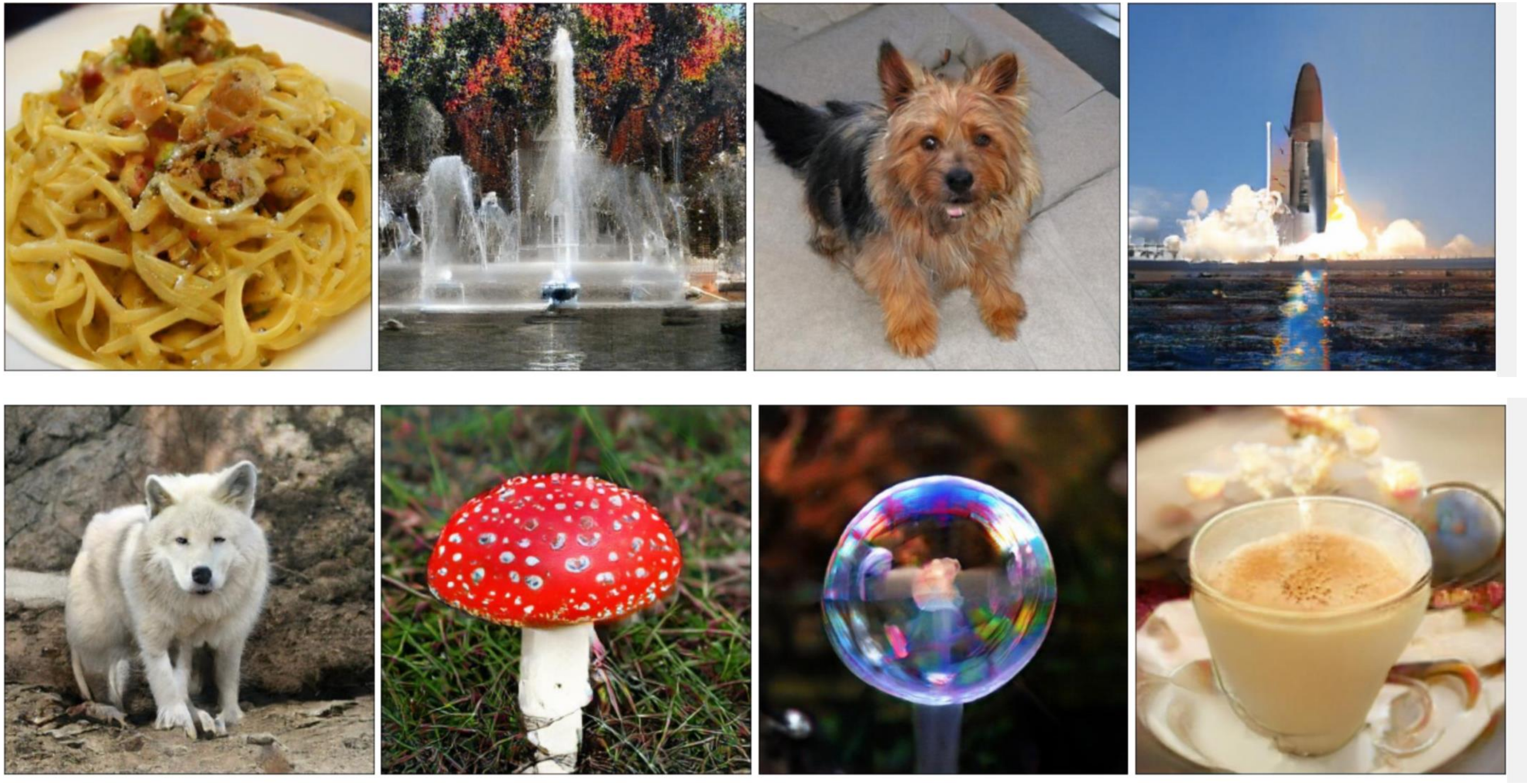
(Bansal et al, 2017)

Progress on Face Generation

2014 2015 2016 2017

(Brundage et al, 2018)

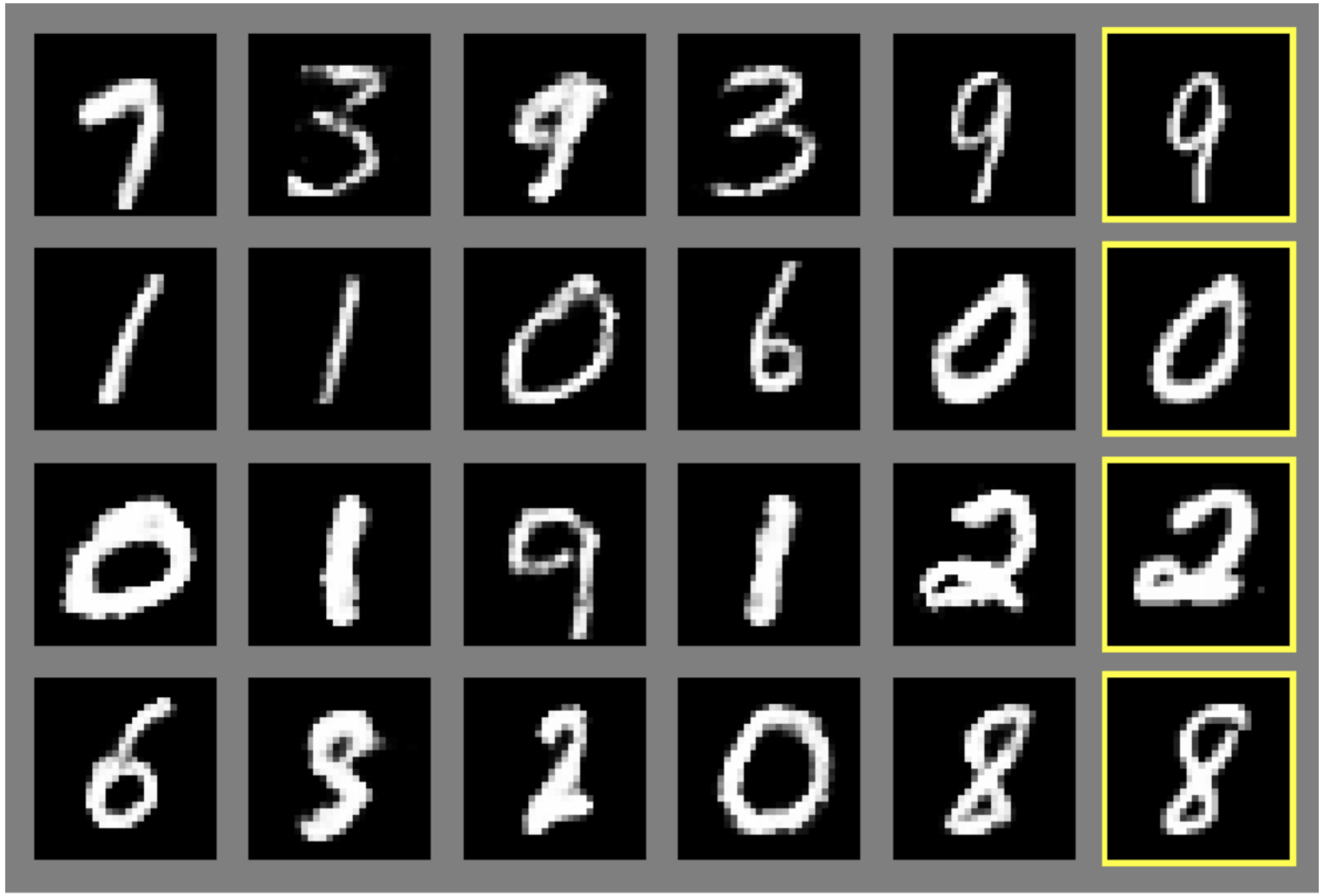
BigGAN – Late 2018



From GAN to BigGAN

- Depth and Convolution
- Class-conditional Generation
- Wasserstein GAN
- Self Attention
- BigGAN

No Convolution Needed to Solve Simple Tasks

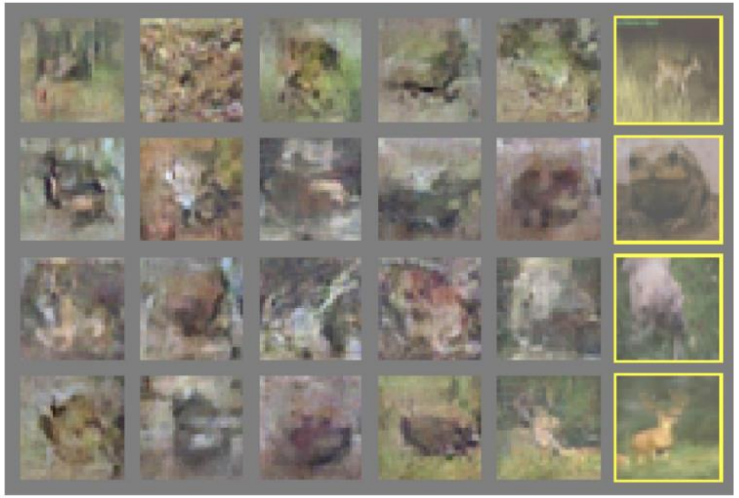


Original GAN, 2014

Depth and Convolution for Harder Tasks

Original GAN (CIFAR-10)

No convolution



One convolutional layer

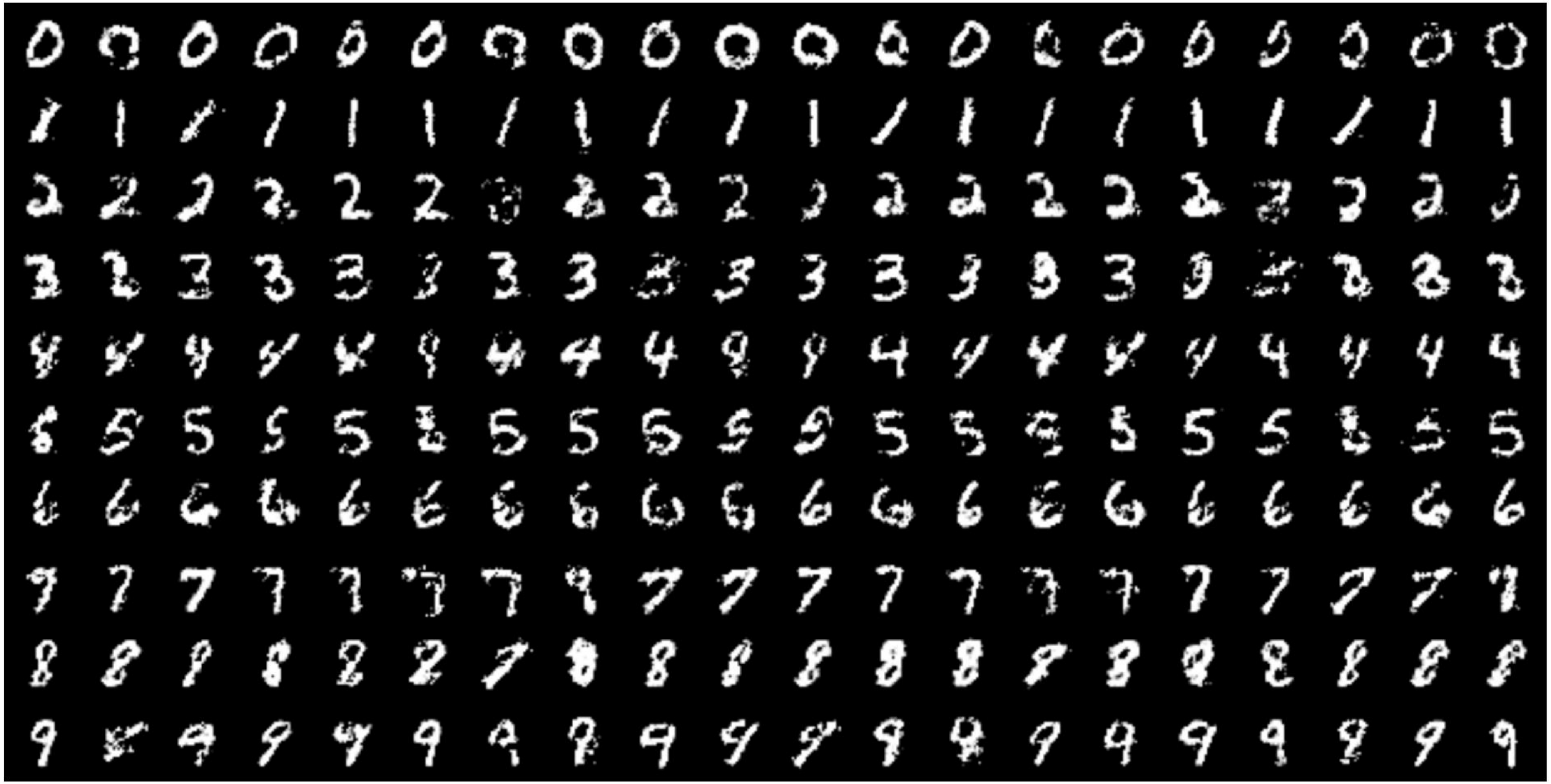
DCGAN (ImageNet)

Many convolutional layers
(Radford et al, 2015)

From GAN to BigGAN

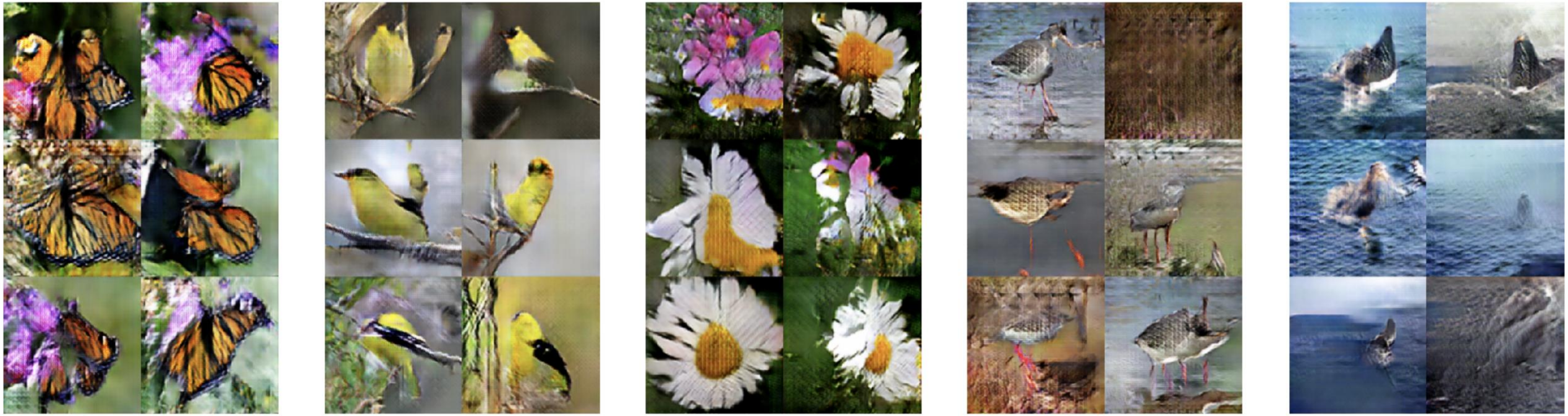
- Depth and Convolution
- Class-conditional Generation
- Wasserstein GAN
- Self Attention
- BigGAN

Class-Conditional GANs



(Mirza and Osindero, 2014)

AC-GAN: Specialist Generators

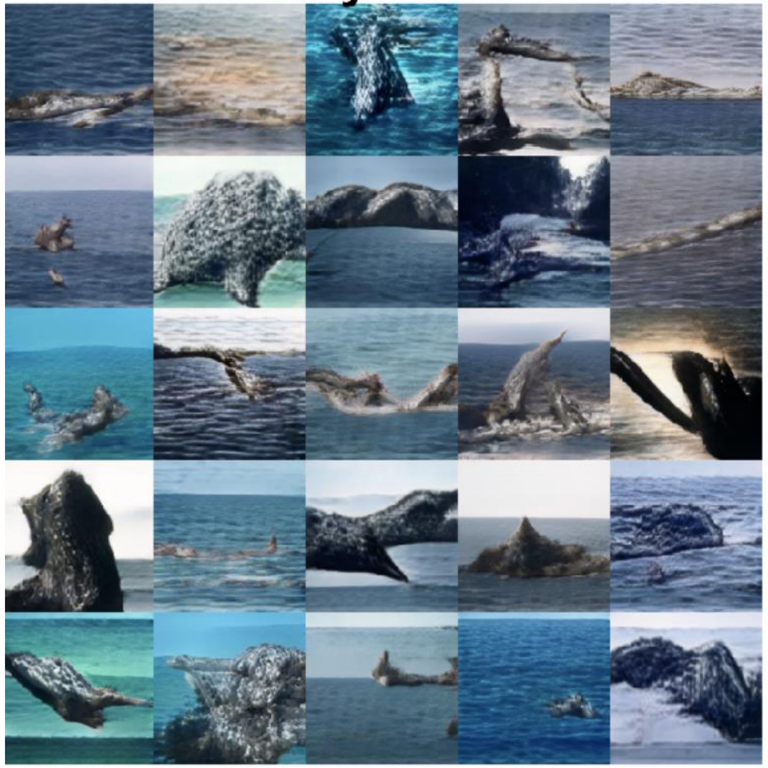


monarch butterfly | goldfinch | daisy | redshank | grey whale

(Odena et al, 2016)

SN-GAN: Shared Generator

Gray whale



Welsh springer spaniel

Persian cat

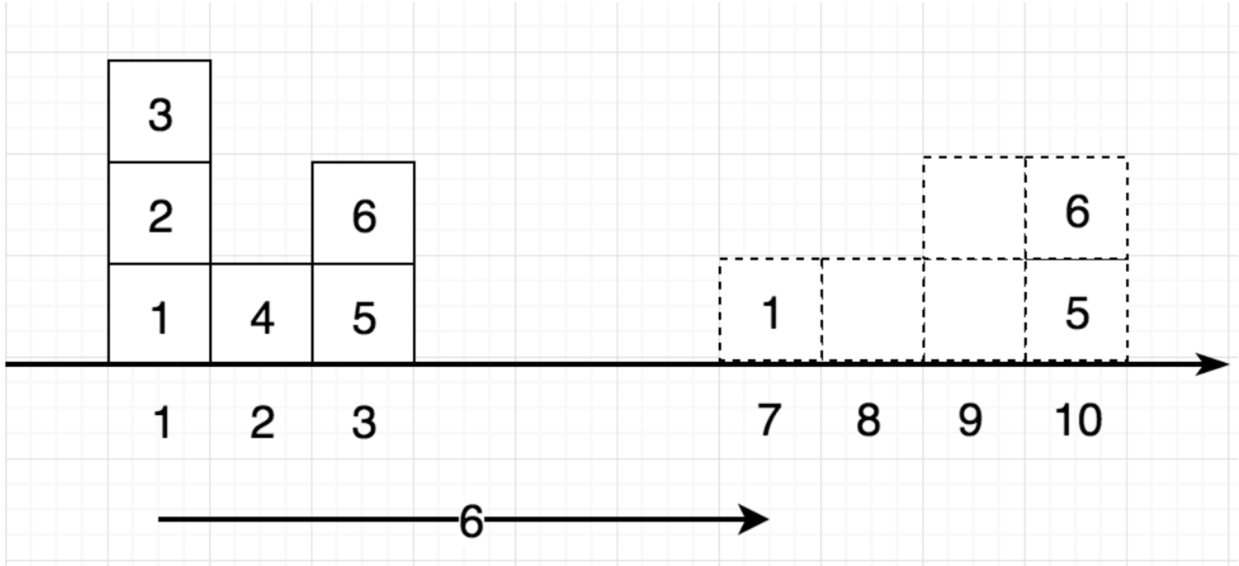
(Miyato et al, 2017)

From GAN to BigGAN

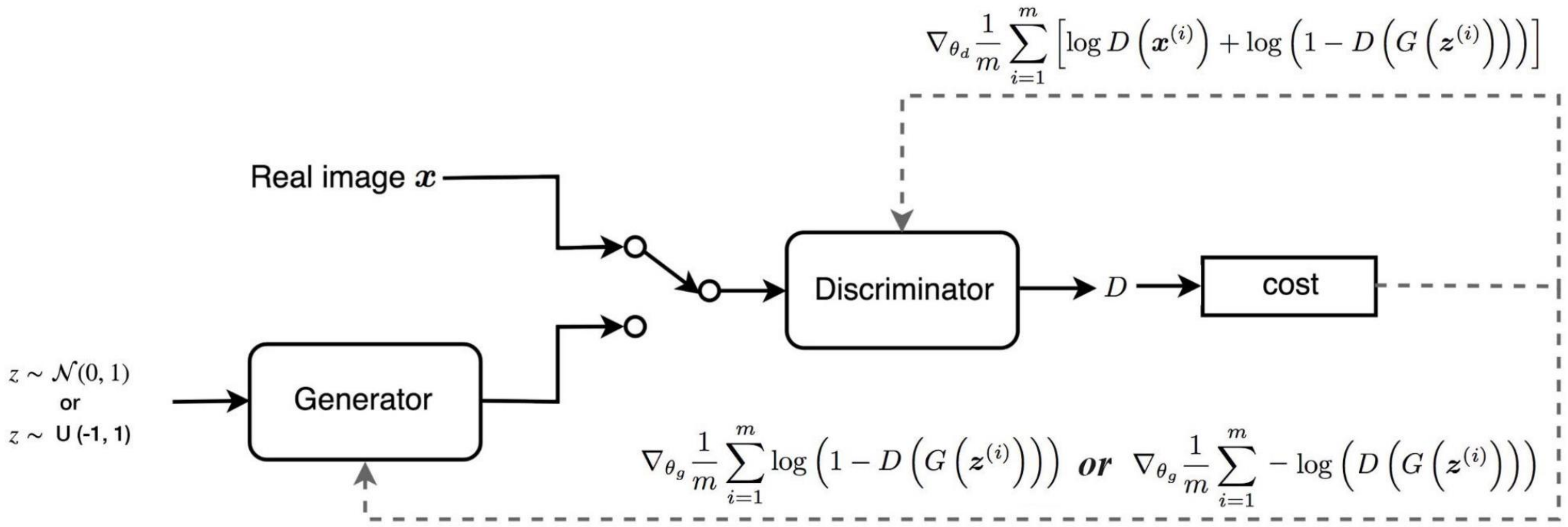
- Depth and Convolution
- Class-conditional Generation
- Wasserstein GAN
- Self Attention
- BigGAN

Wasserstein GAN

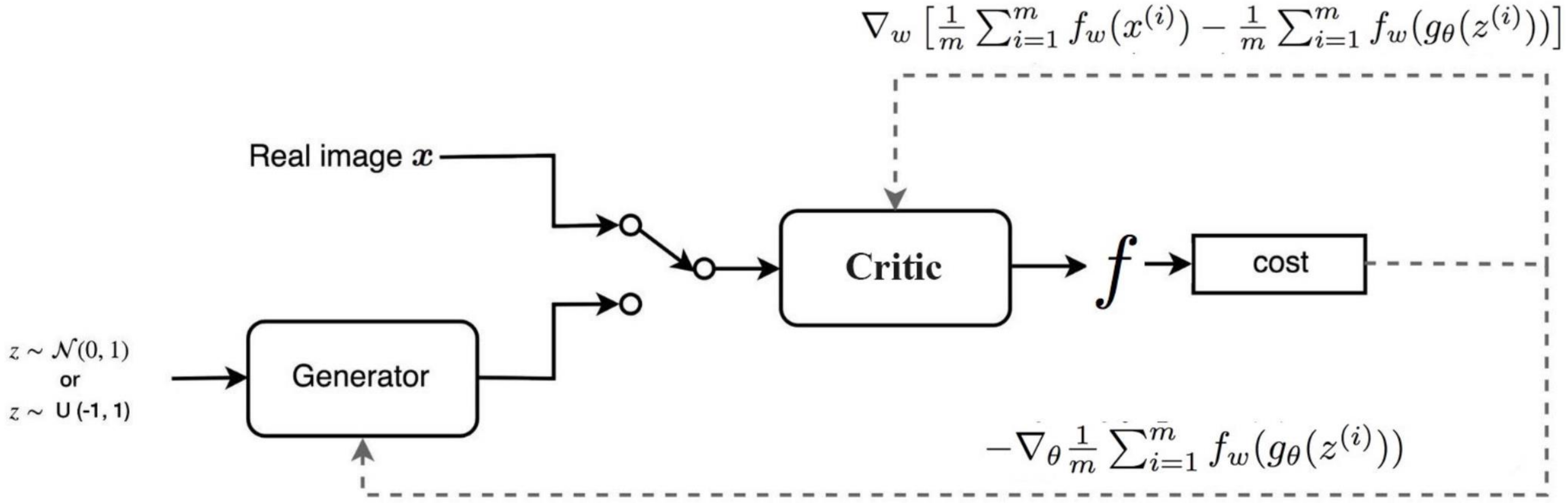
- Wasserstein Distance: Minimum cost of transporting mass in converting the data distribution q to the data distribution p.



GAN:


$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right]$
Real image $\boldsymbol{x}$
Discriminator
D
cost
$z \sim \mathcal{N}(0, 1)$
or
$z \sim$ U (-1, 1)
Generator
$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)$ *or* $\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} -\log\left(D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)$

WGAN


$\nabla_w \left[\frac{1}{m}\sum_{i=1}^m f_w(x^{(i)}) - \frac{1}{m}\sum_{i=1}^m f_w(g_\theta(z^{(i)}))\right]$
Real image x
Critic
f
cost
$z \sim \mathcal{N}(0,1)$
or
$z \sim$ U (-1, 1)
Generator
$-\nabla_\theta \frac{1}{m}\sum_{i=1}^m f_w(g_\theta(z^{(i)}))$

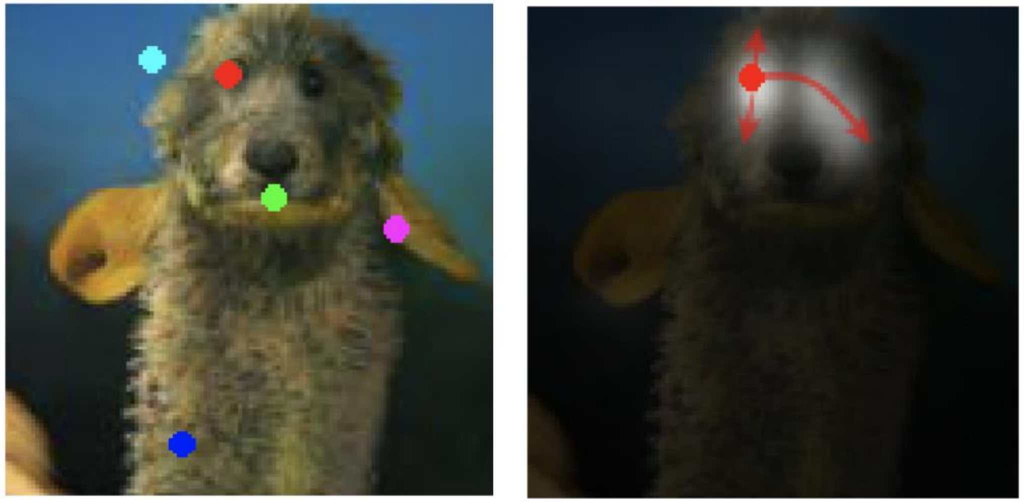
	Discriminator/Critic	Generator
GAN	$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right]$	$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} - \log\left(D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)$
WGAN	$\nabla_w \frac{1}{m} \sum_{i=1}^{m} \left[f(x^{(i)}) - f(G(z^{(i)})) \right]$	$\nabla_\theta \frac{1}{m} \sum_{i=1}^{m} -f(G(z^{(i)}))$

$$w \leftarrow w + \alpha \cdot \mathrm{RMSProp}(w, g_w)$$
$$w \leftarrow \mathrm{clip}(w, -c, c)$$

From GAN to BigGAN

- Depth and Convolution
- Class-conditional Generation
- Wasserstein GAN
- Self Attention
- BigGAN

Self-Attention



Use layers from
Wang et al 2018

From GAN to BigGAN

- Depth and Convolution
- Class-conditional Generation
- Wasserstein GAN
- Self Attention
- BigGAN

BigGAN

- Scalability: GANs benefit dramatically from scaling. Two architectural changes that improve scalability.
- Robustness: Fine control of the trade-offs between fidelity and variety is possible via the “truncation trick”
- Stability: Devises solutions that minimize the instabilities in Large Scale GANs

Figure 1: Class-conditional samples generated by our model.

Applying GANs

- Semi-supervised Learning
- Model-based optimization
- Extreme personalization
- Program synthesis

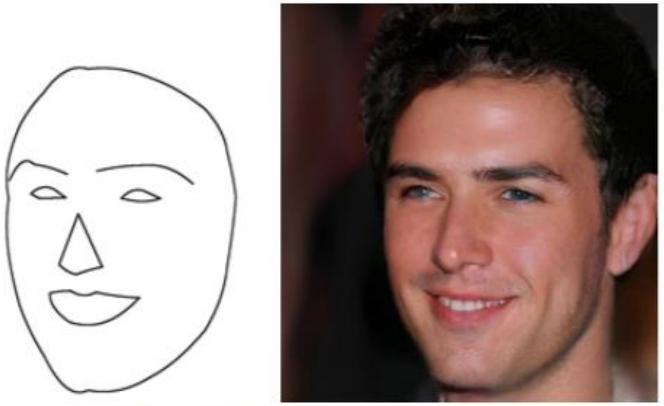
Image to Image Translation

Semantic label → Image

Day → Night

Winter → Summer

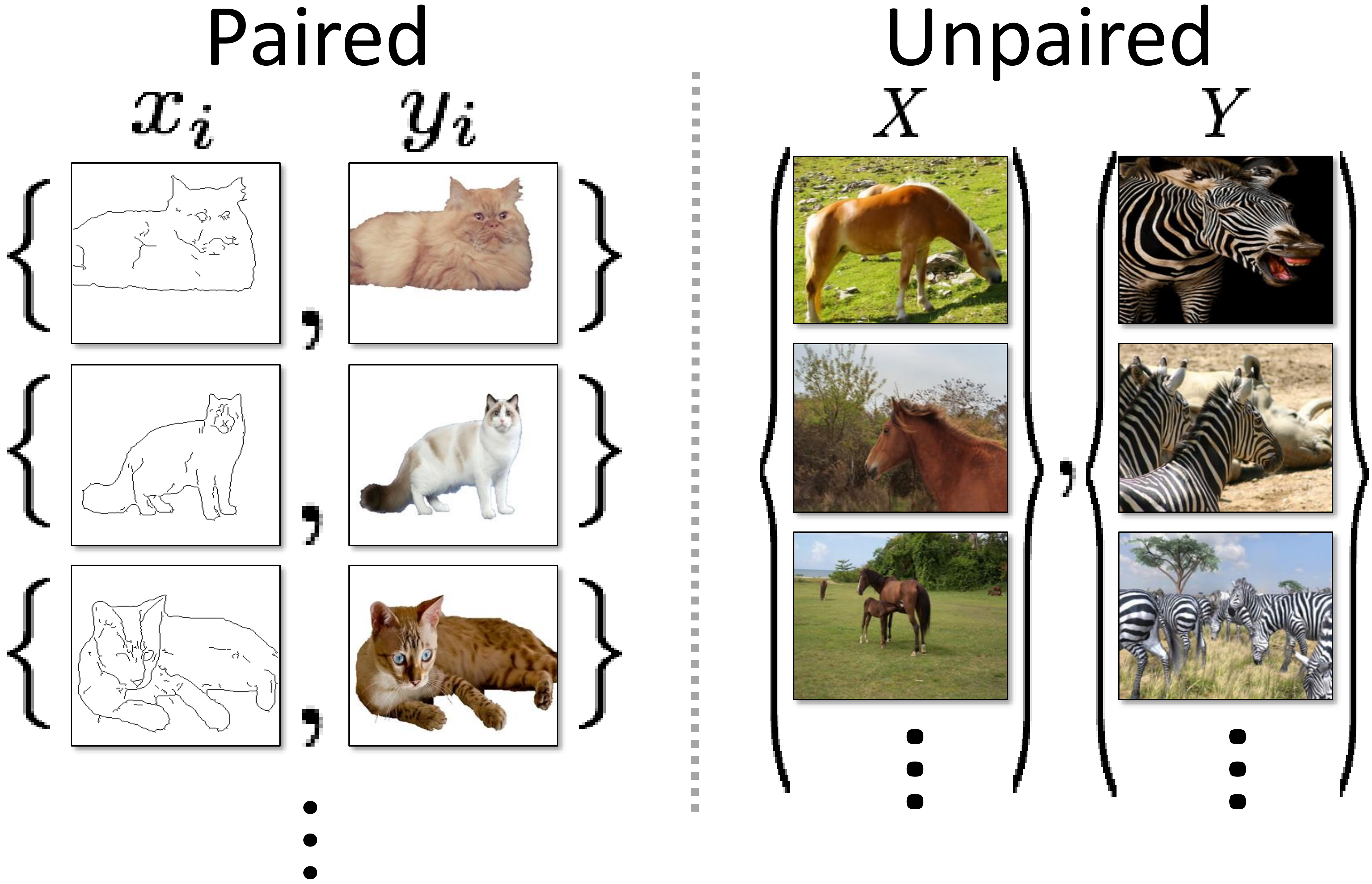
Artistic video gaming



Drawing → Image

Many other applications

	Supervised	Unsupervised
Unimodal	Pix2pix, CRN, SRGAN	DistanceGAN, CycleGAN, DiscoGAN, DualGAN, UNIT, DTN, StarGAN, OST
Multimodal	pix2pixHD, BicycleGAN	MUNIT, Augmented CycleGAN

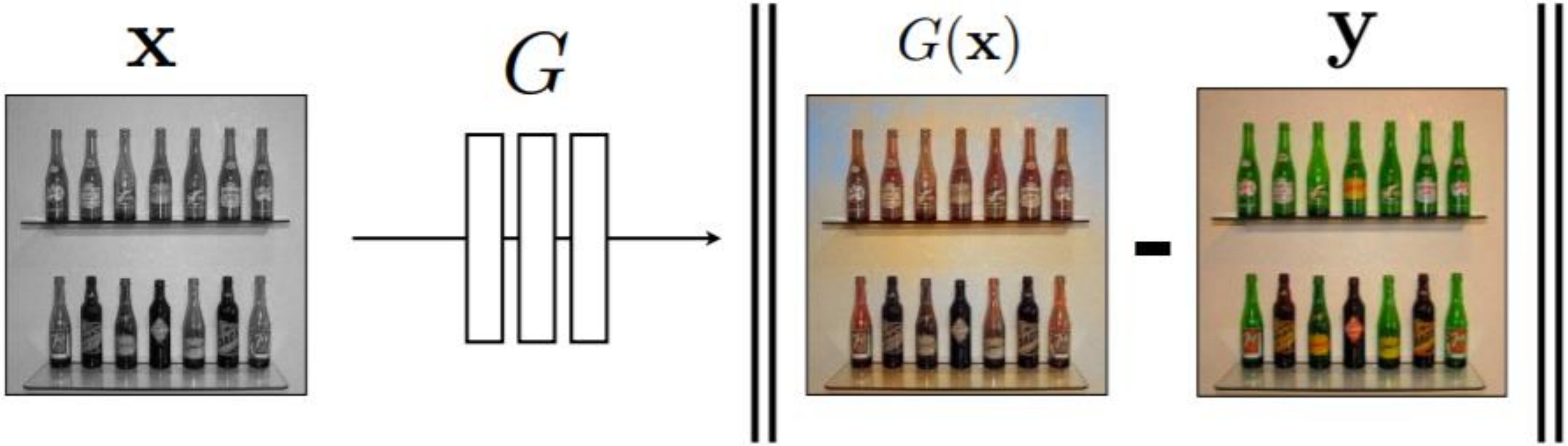


Paired
x_i
y_i
Unpaired
X
Y

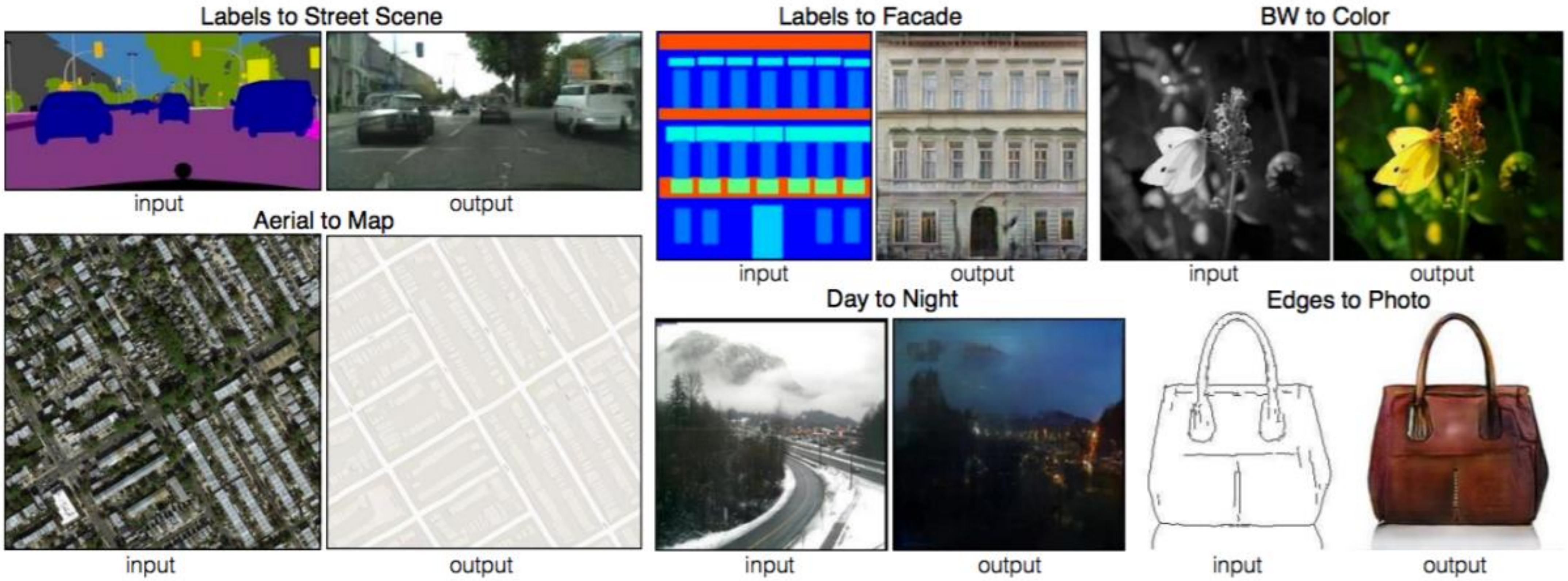
Fully Supervised: pix2pix

Conditional GAN

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$



[Isola et al., CVPR 2017]



[Isola et al., CVPR 2017]

Unsupervised: Circular GANs

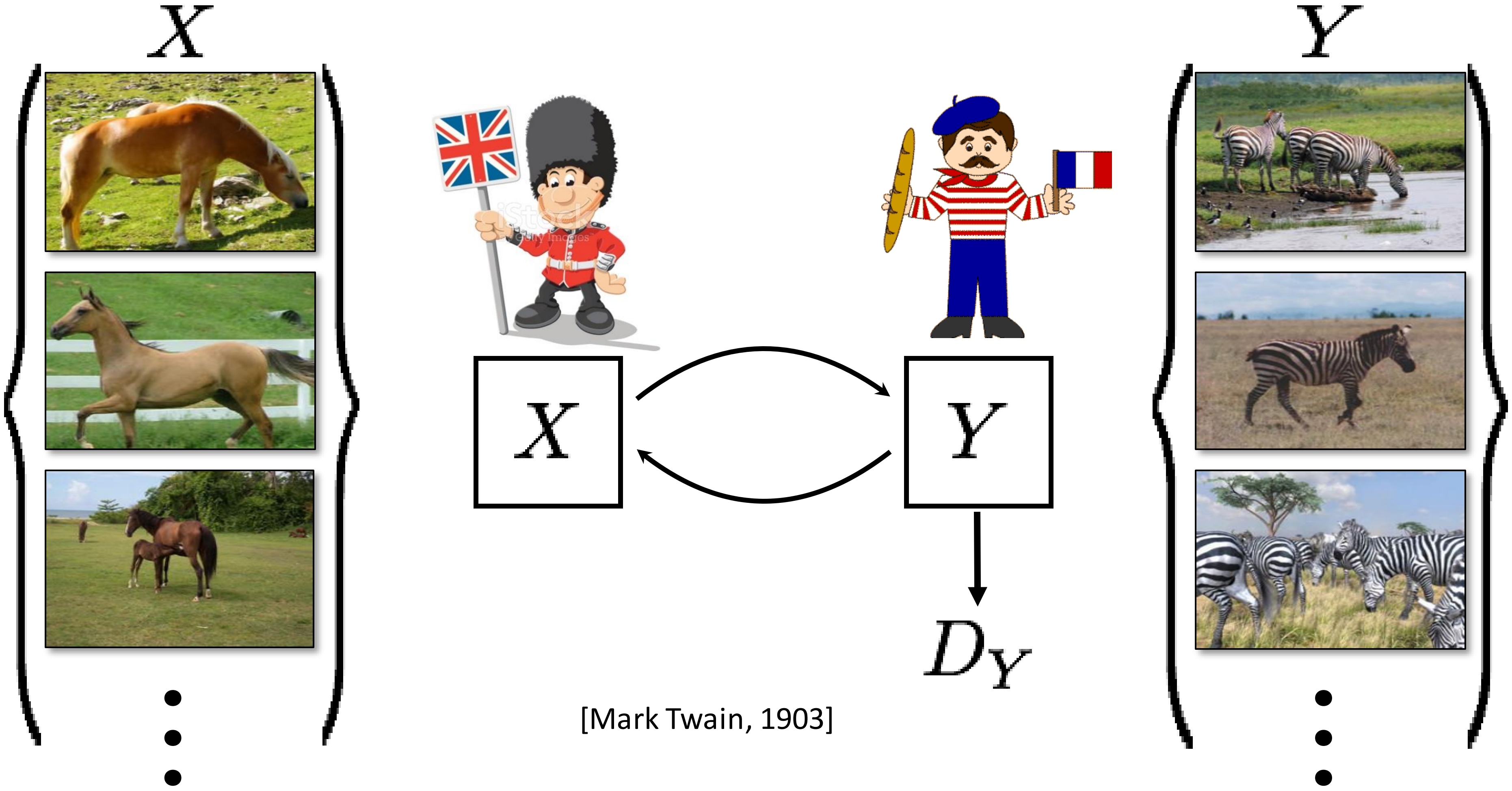
DiscoGAN: "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks". Kim et al. ICML'17.

CycleGAN: "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". Zhu et al. arXiv:1703.10593, 2017.

DualGAN: " Unsupervised Dual Learning for Image-to-Image Translation". Zili et al. arXiv:1704.02510, 2017.

.

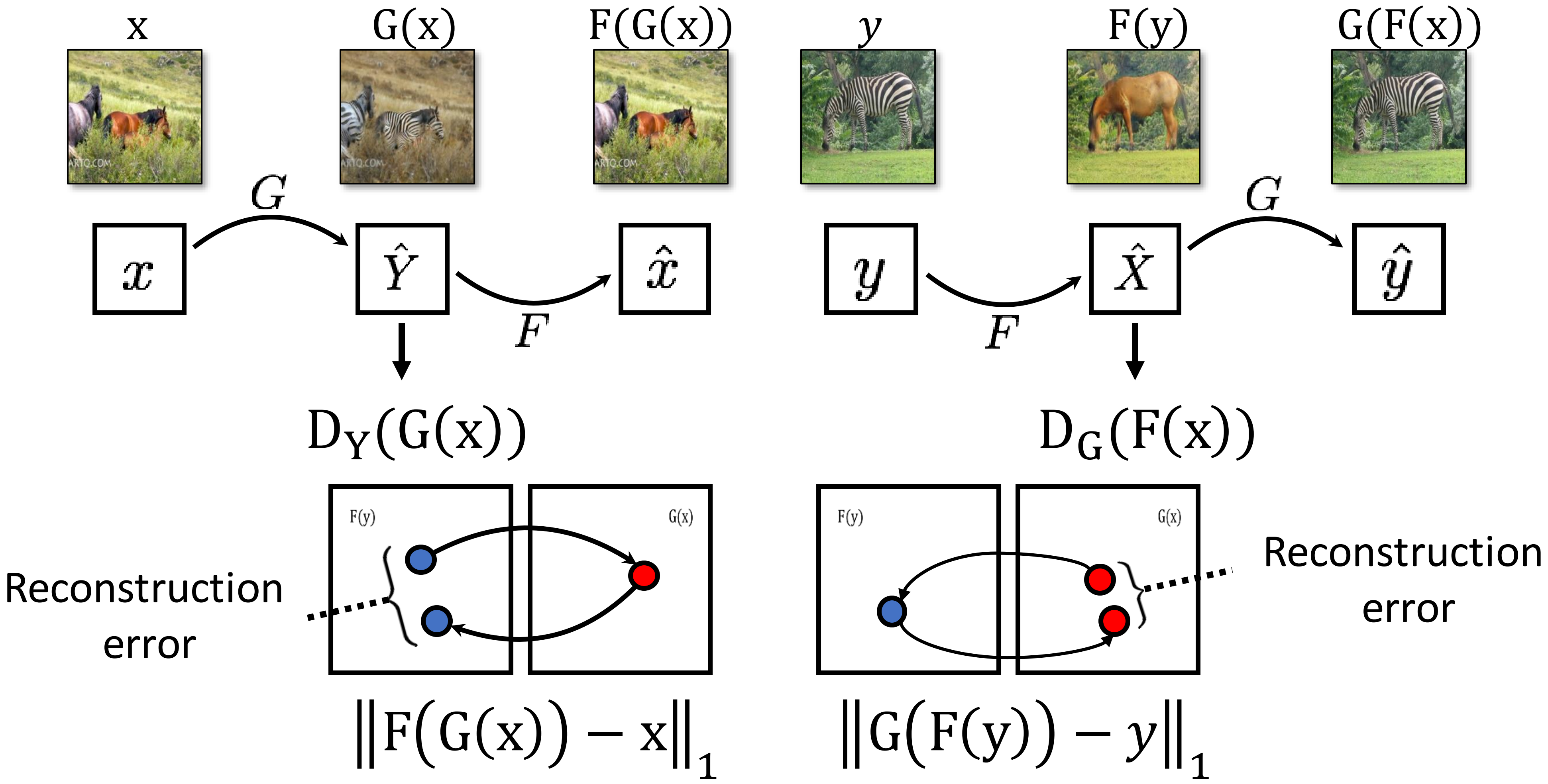
Cycle-Consistent Adversarial Networks



[Mark Twain, 1903]

[Zhu et al., ICCV 2017]

Cycle Consistency Loss



See similar formulations [Yi et al. 2017], [Kim et al. 2017]

[Zhu et al., ICCV 2017]

Collection Style Transfer

Photograph
@ Alexei Efros

Monet

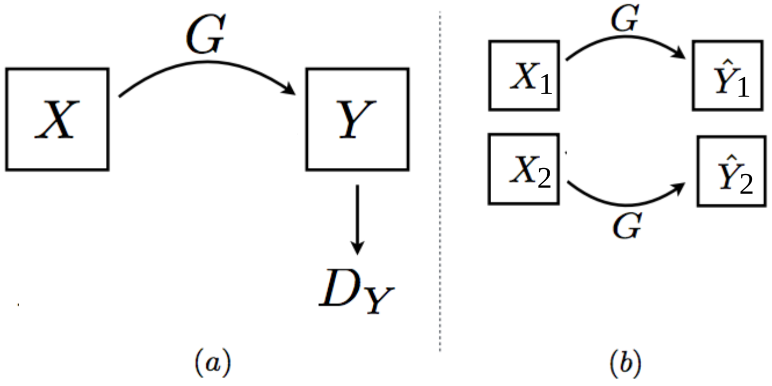
Van Gogh

Cezanne

Ukiyo-e

DistanceGAN

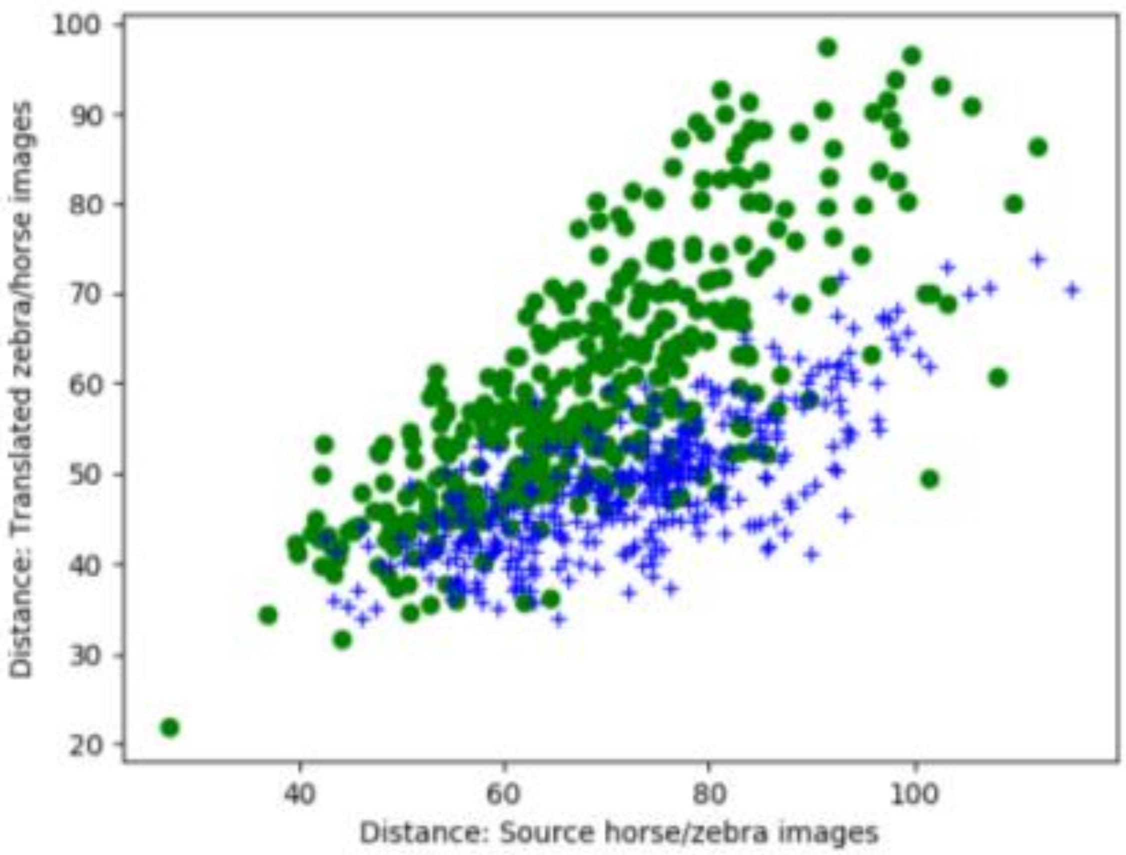
- A pair of images of a given distance are mapped to a pair of outputs with a similar distance
- $\left|x_i - x_j\right|_1$ and $\left|G(x_i) - G\left(x_j\right)\right|_1$ are highly correlated.



$$|x_1 - x_2|_1 \sim |G(x_1) - G(x_2)|_1$$

Benaim et al., NIPS 2017

Motivating distance correlations

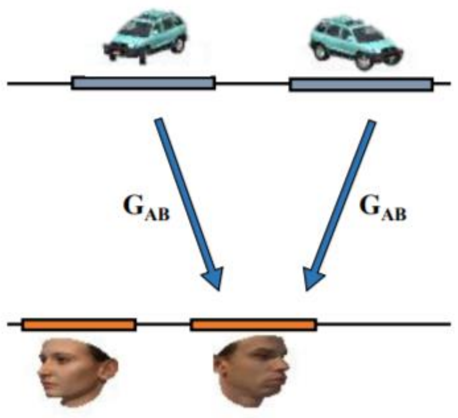


Analysis of CycleGAN's horse to zebra results

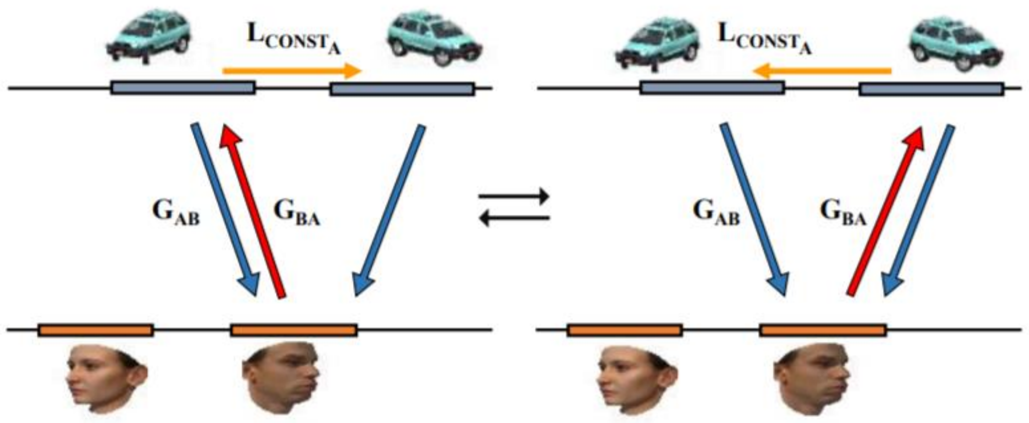
Benaim et al., NIPS 2017

Mode Collapse

- GAN:

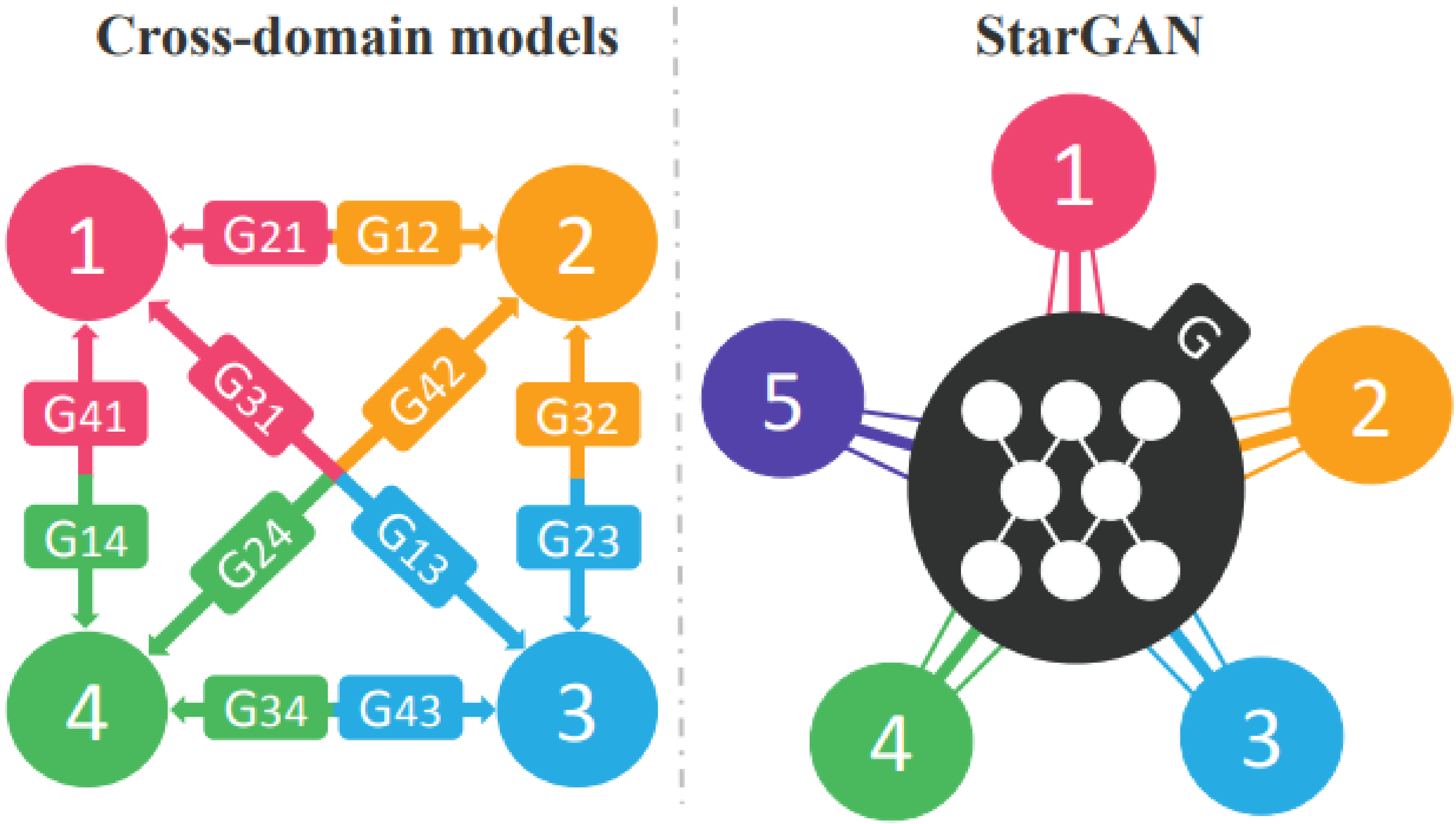


Cycle:



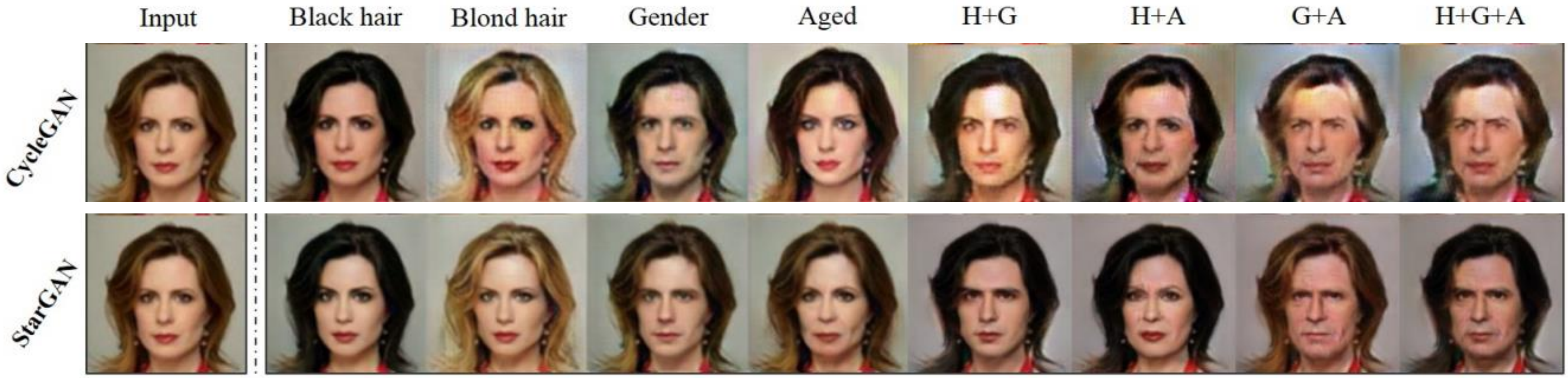
Benaim et al., NIPS 2017

More than 2 domains



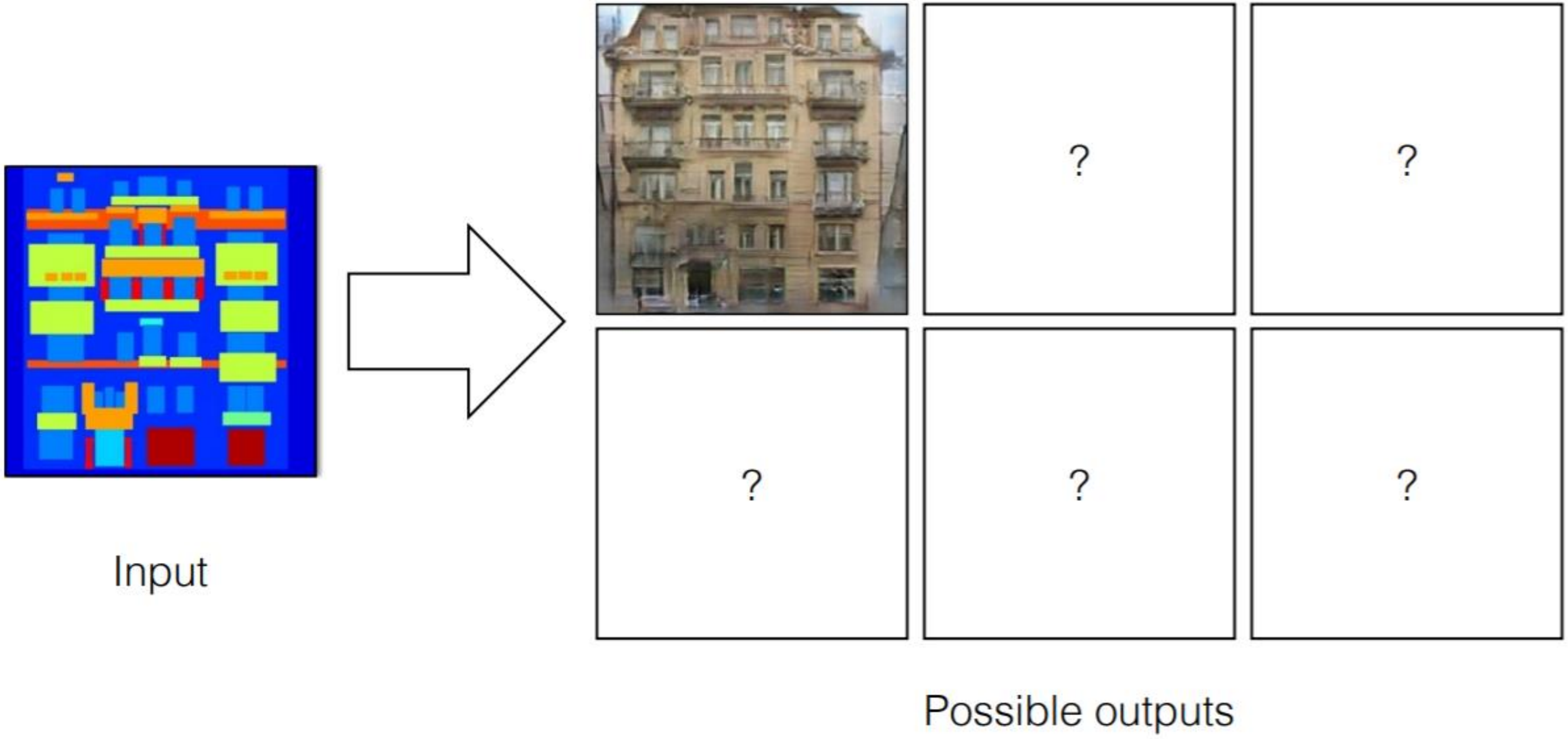
Choi et al., CVPR 2018

More than 2 domains

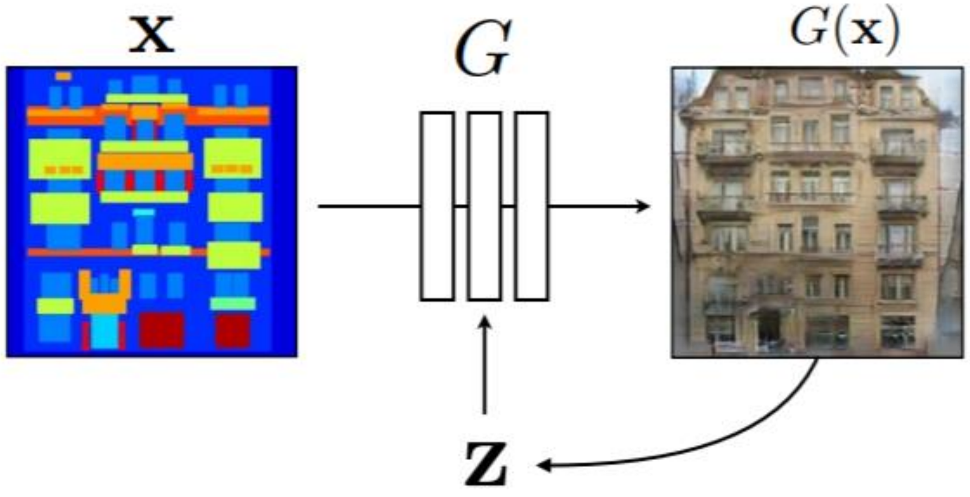


Choi et al., CVPR 2018

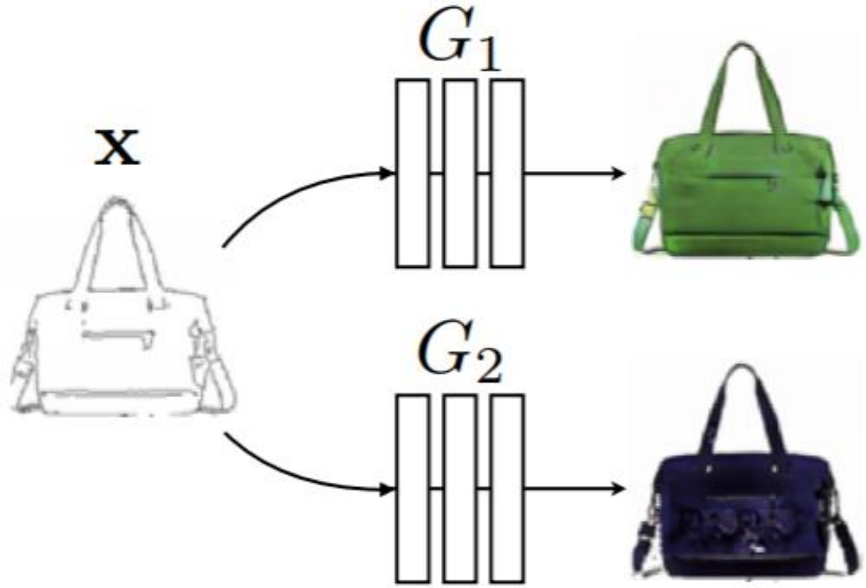
Modeling multiple possible outputs



BiCycleGAN [Zhu et al., NIPS 2017]
(c.f. InfoGAN [Chen et al. 2016])

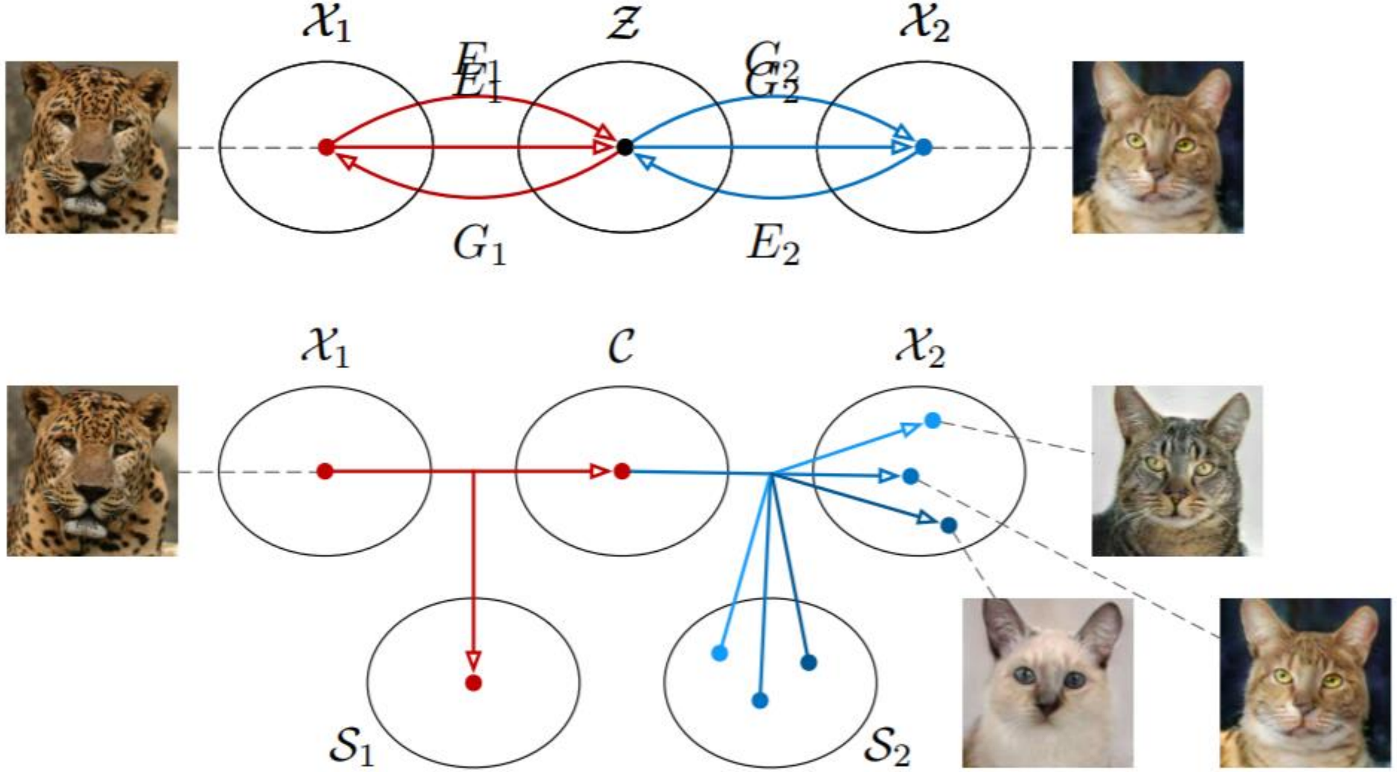


MAD-GAN [Ghosh et al., CVPR 2018]



UNIT: unimodal

MUNIT: multimodal



Huang et al., ECCV 2018

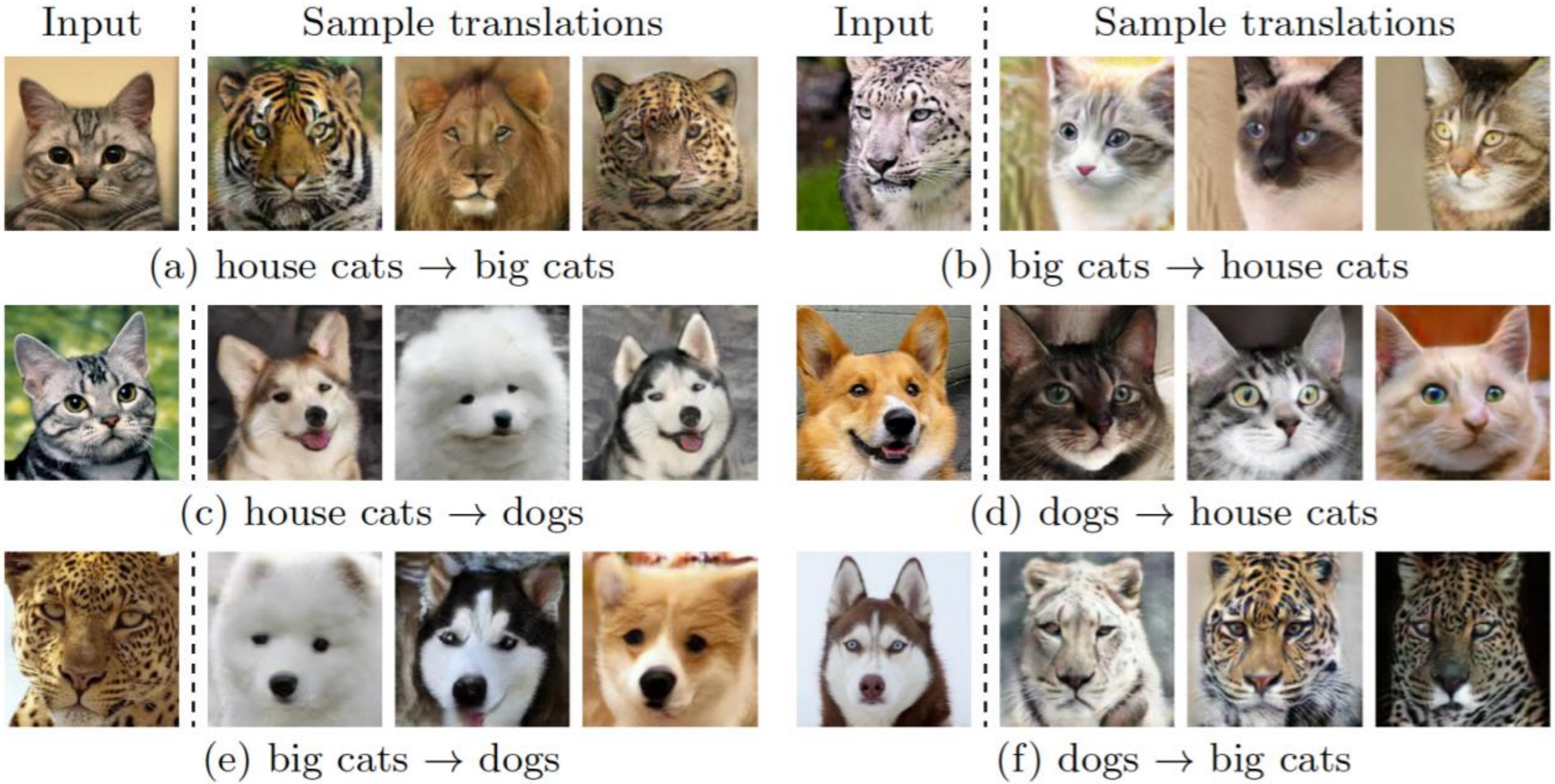
Sketch to Image Translation

(a) edges $\leftrightarrow$ shoes

(b) edges $\leftrightarrow$ handbags

Huang et al., ECCV 2018

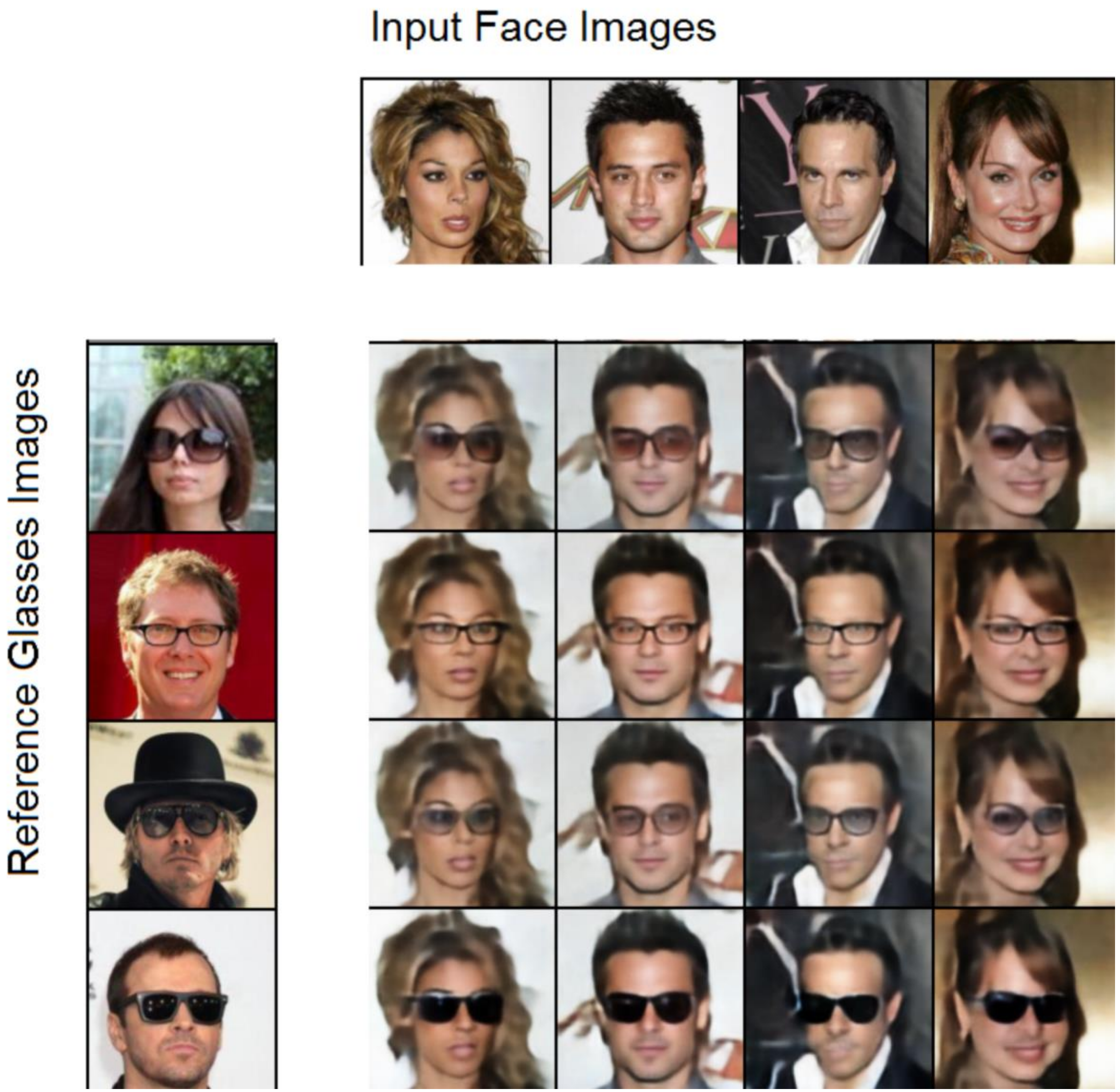
Animal Image Translation



(a) house cats → big cats
(b) big cats → house cats
(c) house cats → dogs
(d) dogs → house cats
(e) big cats → dogs
(f) dogs → big cats

Huang et al., ECCV 2018

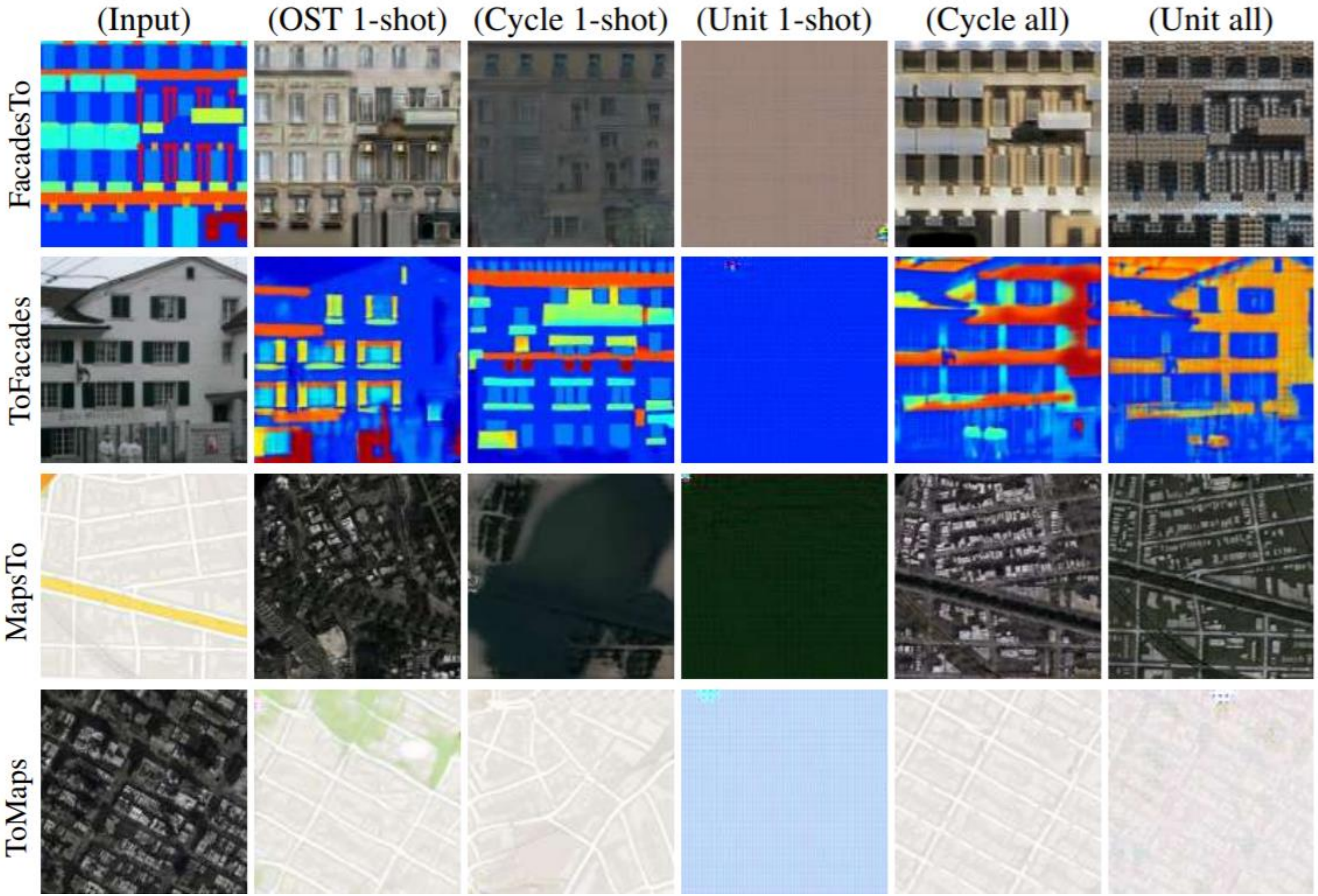
Content Transfer?



One Shot?

- Not only are we unsupervised, but we have only a single sample in the input domain!

"One Shot Unsupervised Cross Domain Mapping", Benaim, et al. In Submission



(Input)
(OST 1-shot)
(Cycle 1-shot)
(Unit 1-shot)
(Cycle all)
(Unit all)
FacadesTo
ToFacades
MapsTo
ToMaps

Applications Beyond Computer Vision

- Many other Vision Applications: Photo Enhancement, Image Dehazing
- Medical Imaging and Biology [Wolterink et al., 2017]
- Voice conversion [Fang et al., 2018, Kaneko et al., 2017]
- Cryptography [CipherGAN: Gomez et al., ICLR 2018]
- Robotics
- NLP: Unsupervised machine translation.
- NLP: Text style transfer.
- …

Thank You! Questions?