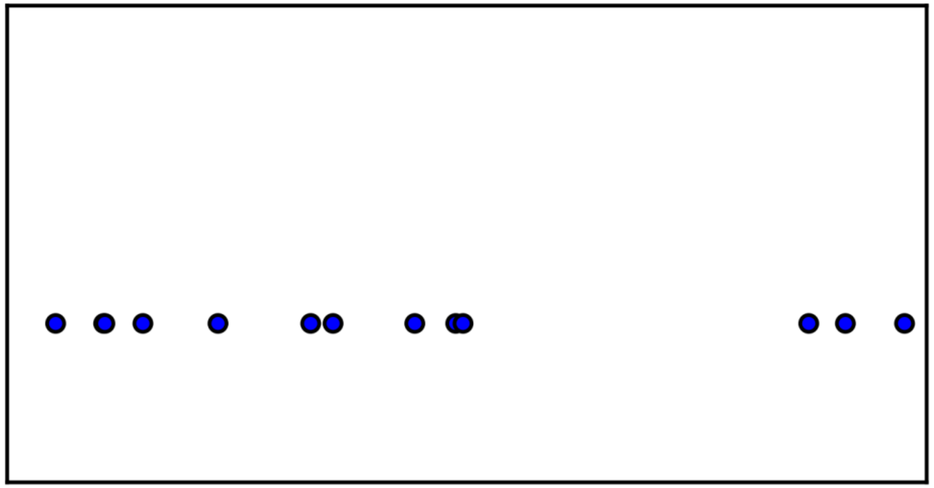
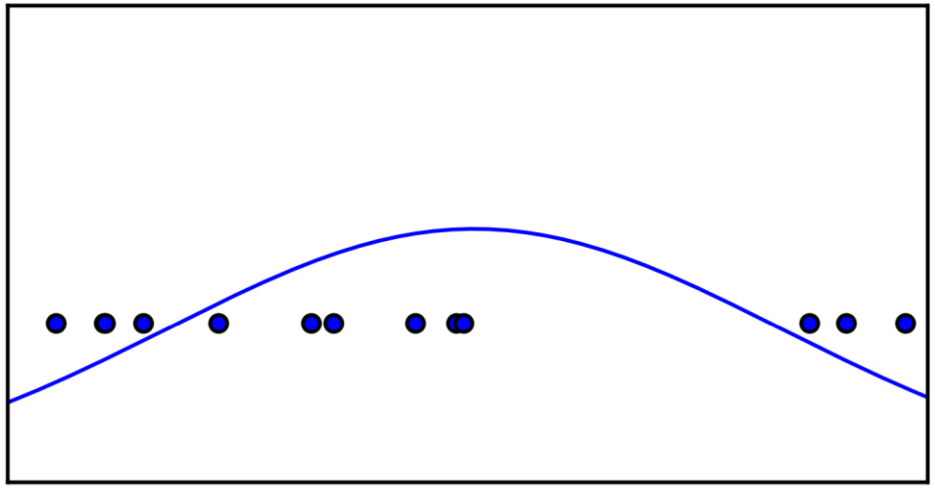
Generative Adversarial Networks For Image to Image Translation

Sagie Benaim
Tel Aviv University

Generative Modeling: Density Estimation



Training Data



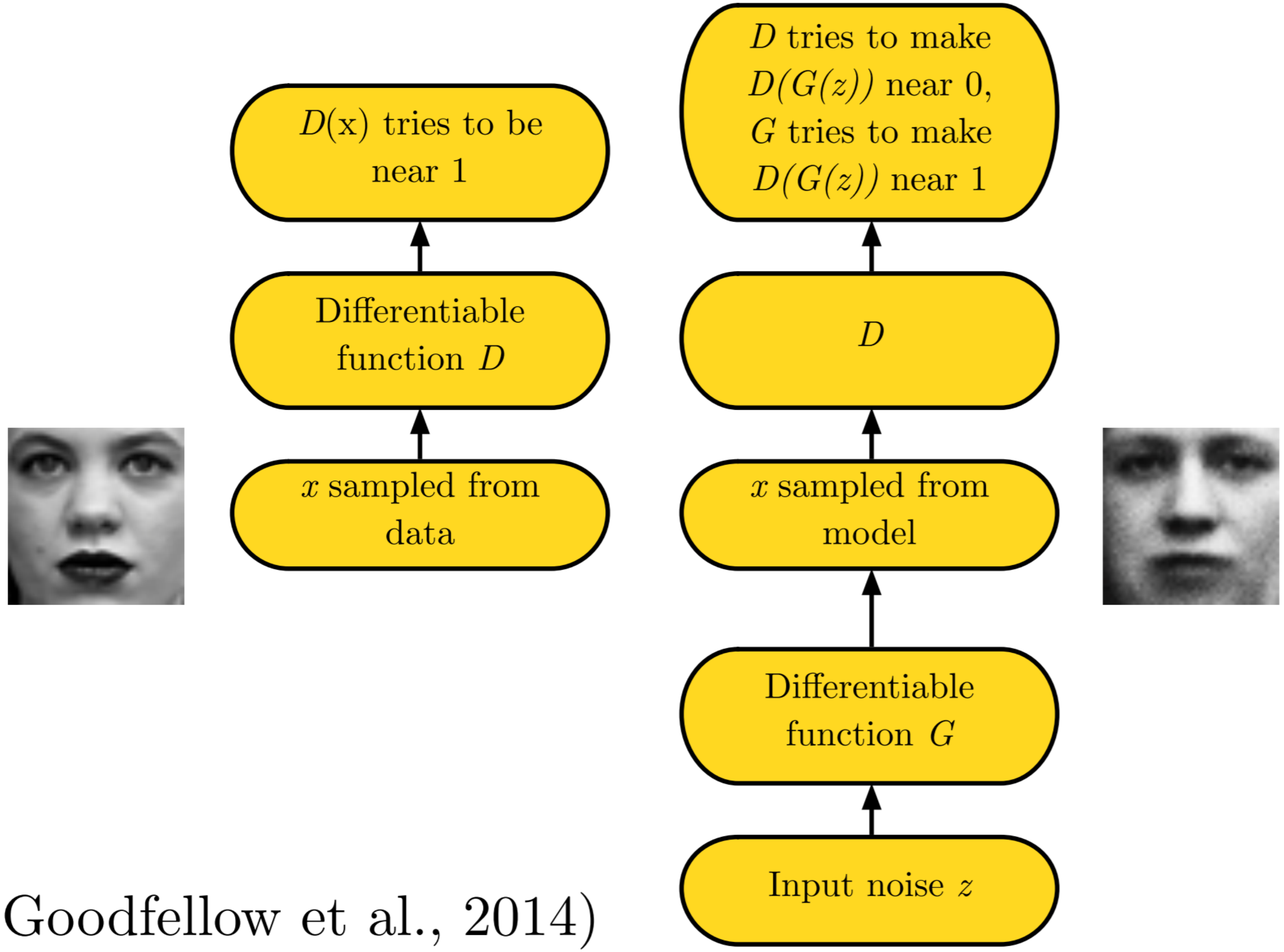
Density Function

Generative Modeling: Sample Generation

Training Data
(CelebA)

Sample Generator
(Karras et al, 2017)

Adversarial Nets Framework



(Goodfellow et al., 2014)

Conditional GAN

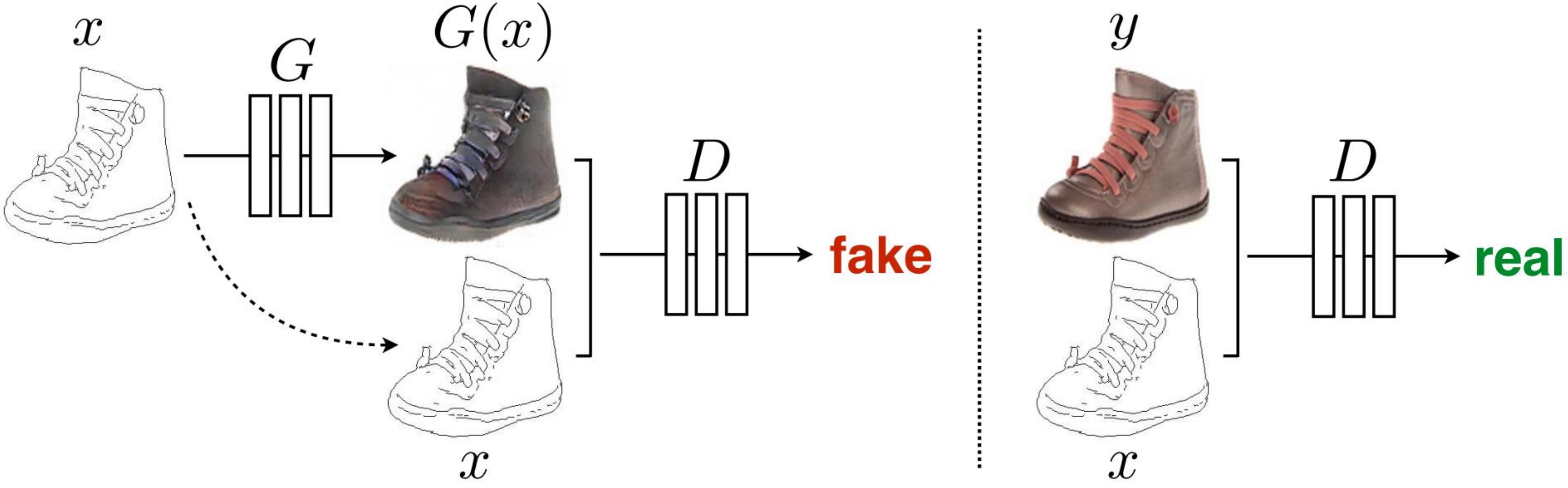
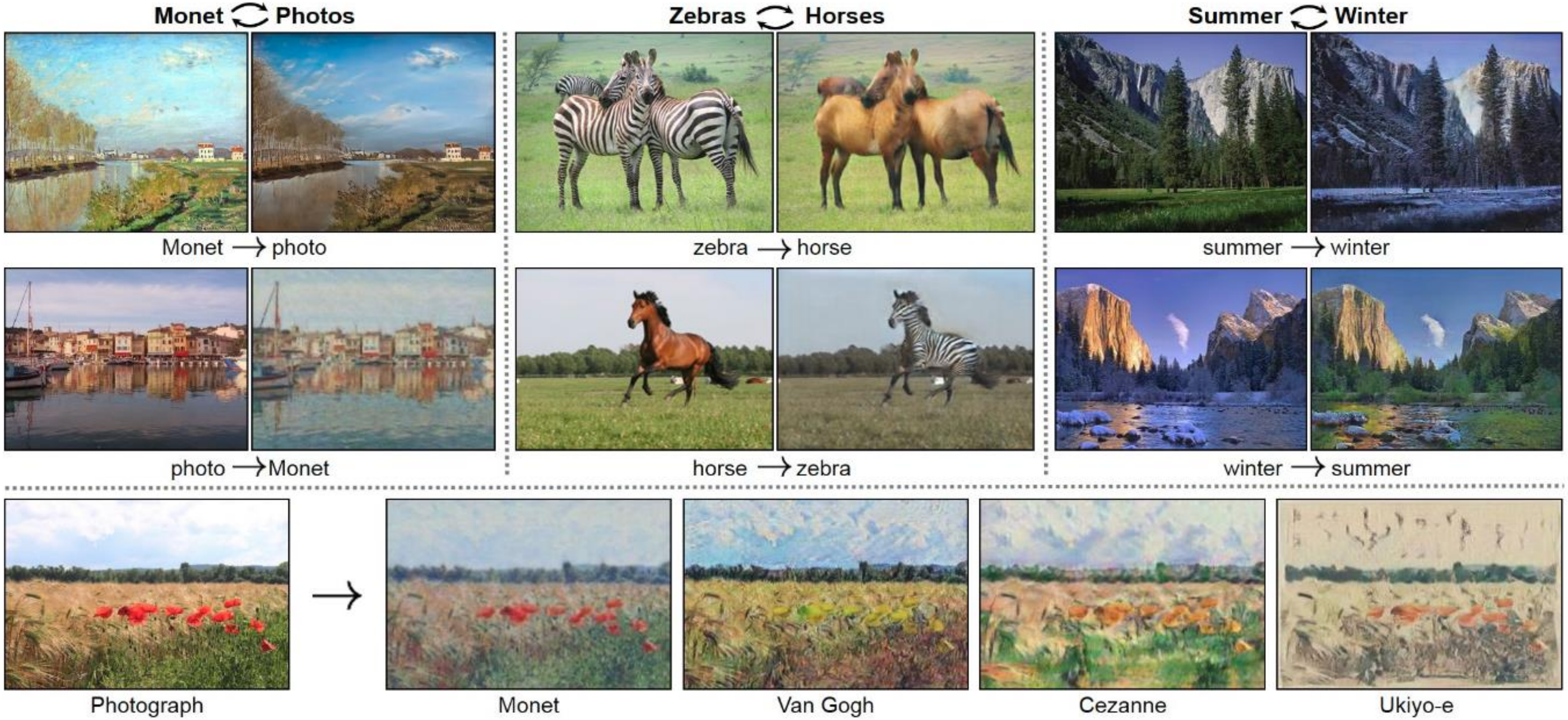
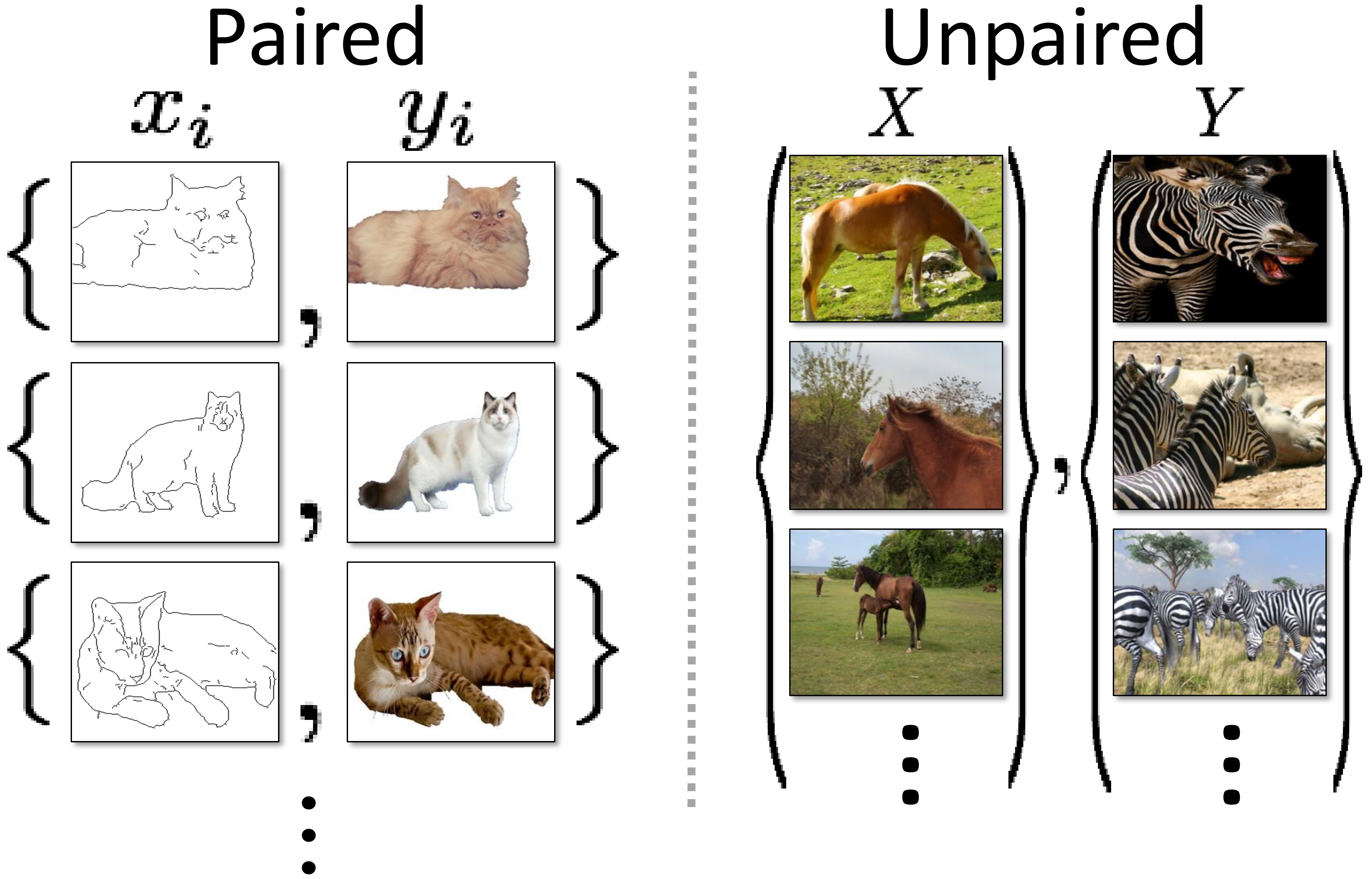


Image to Image Translation



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

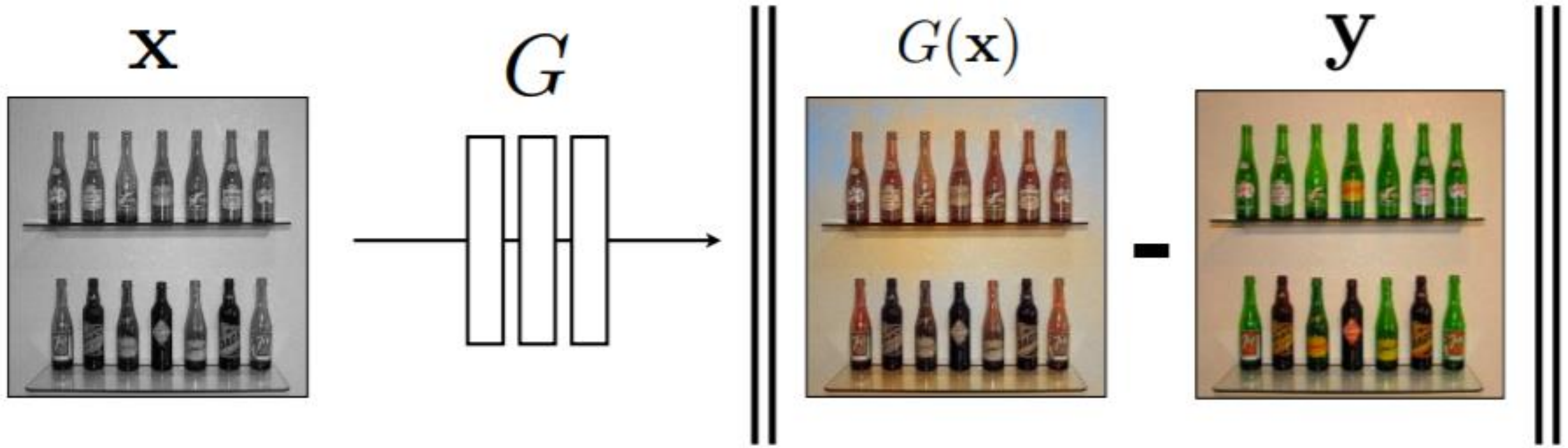


Paired
x_i
y_i
Unpaired
X
Y

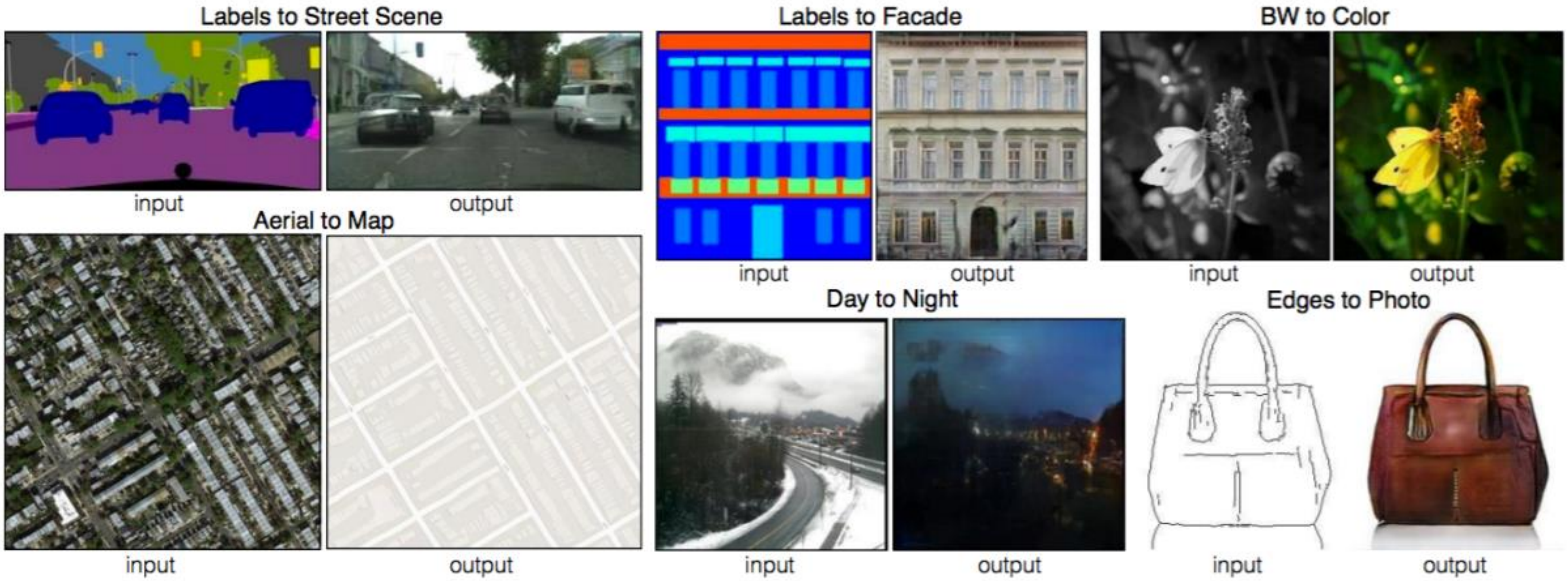
Fully Supervised: pix2pix

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]

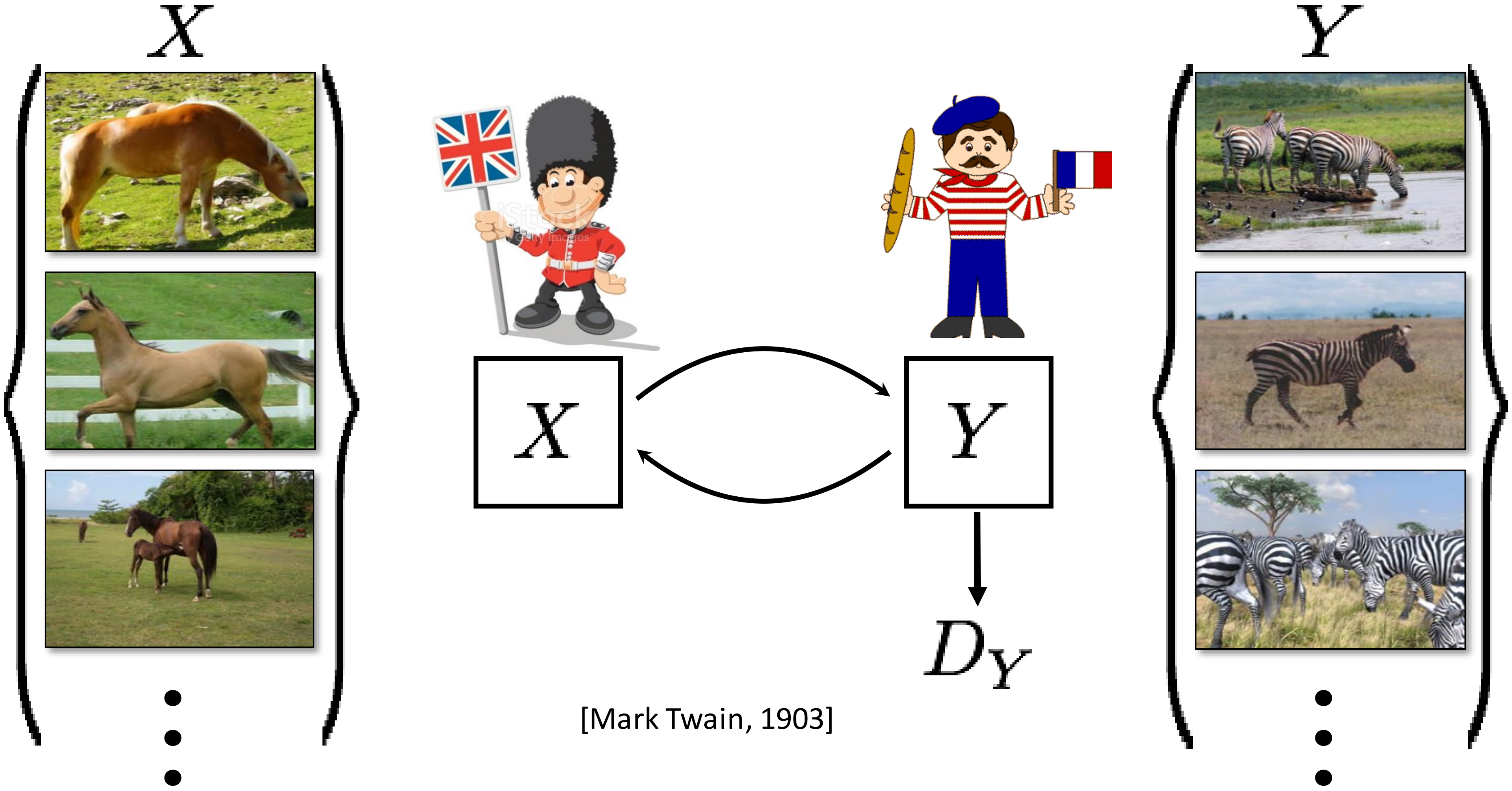
Unsupervised: Circular GANs

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.

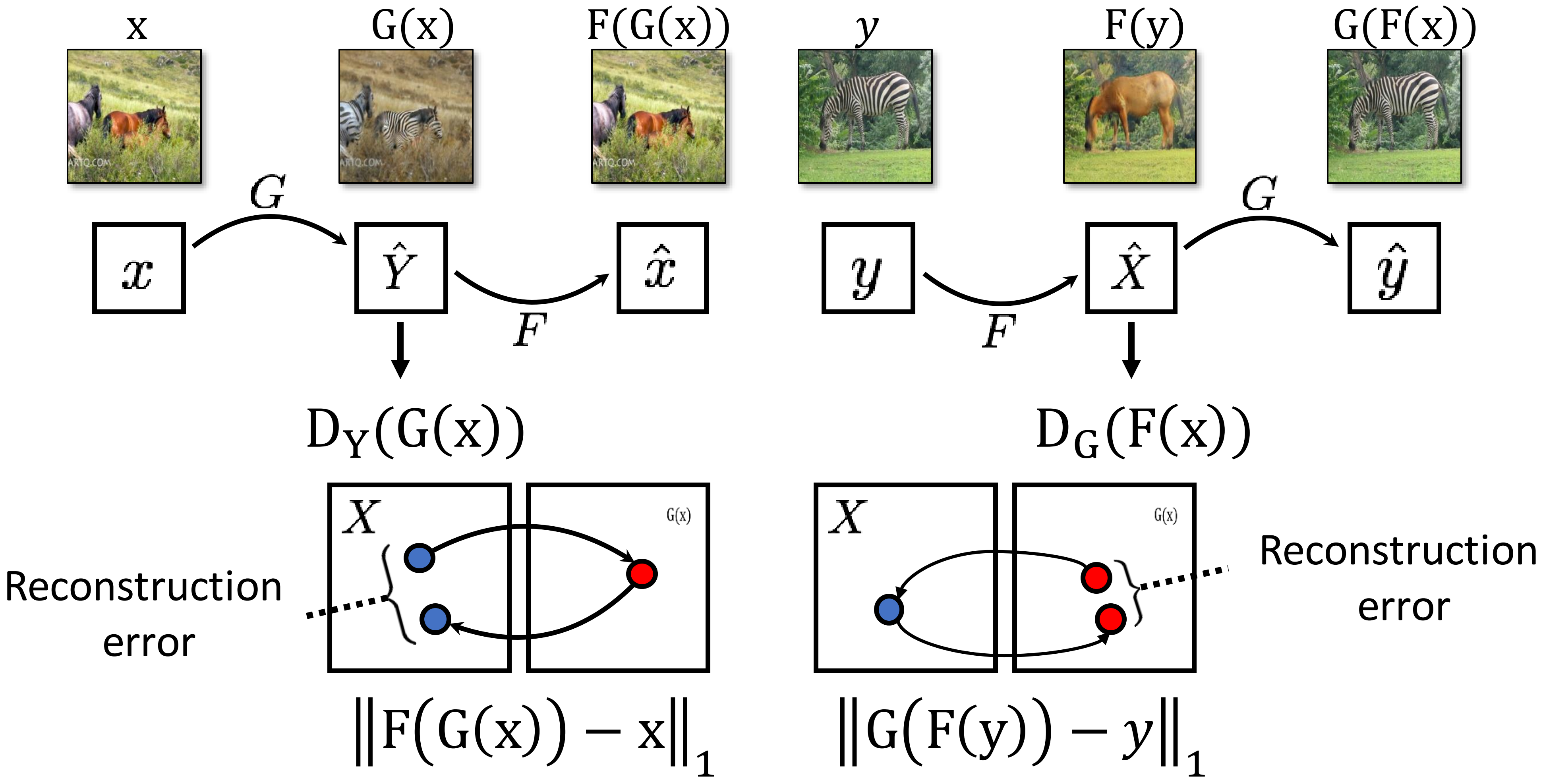
DualGAN: “ Unsupervised Dual Learning for Image-to-Image Translation”. Zili et al. arXiv:1704.02510, 2017.

Cycle-Consistent Adversarial Networks



[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

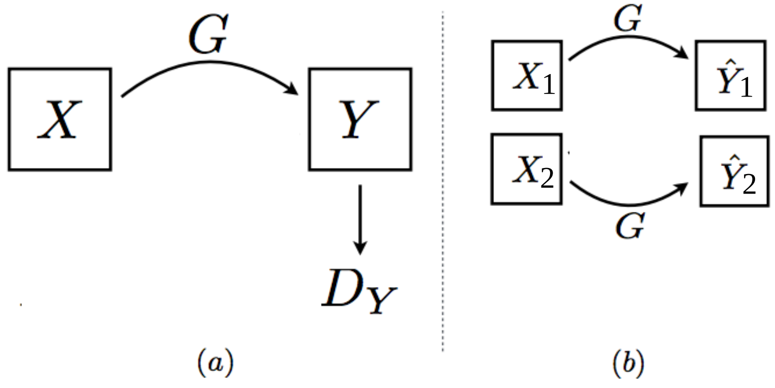
Van Gogh

Cezanne

Ukiyo-e

DistanceGAN

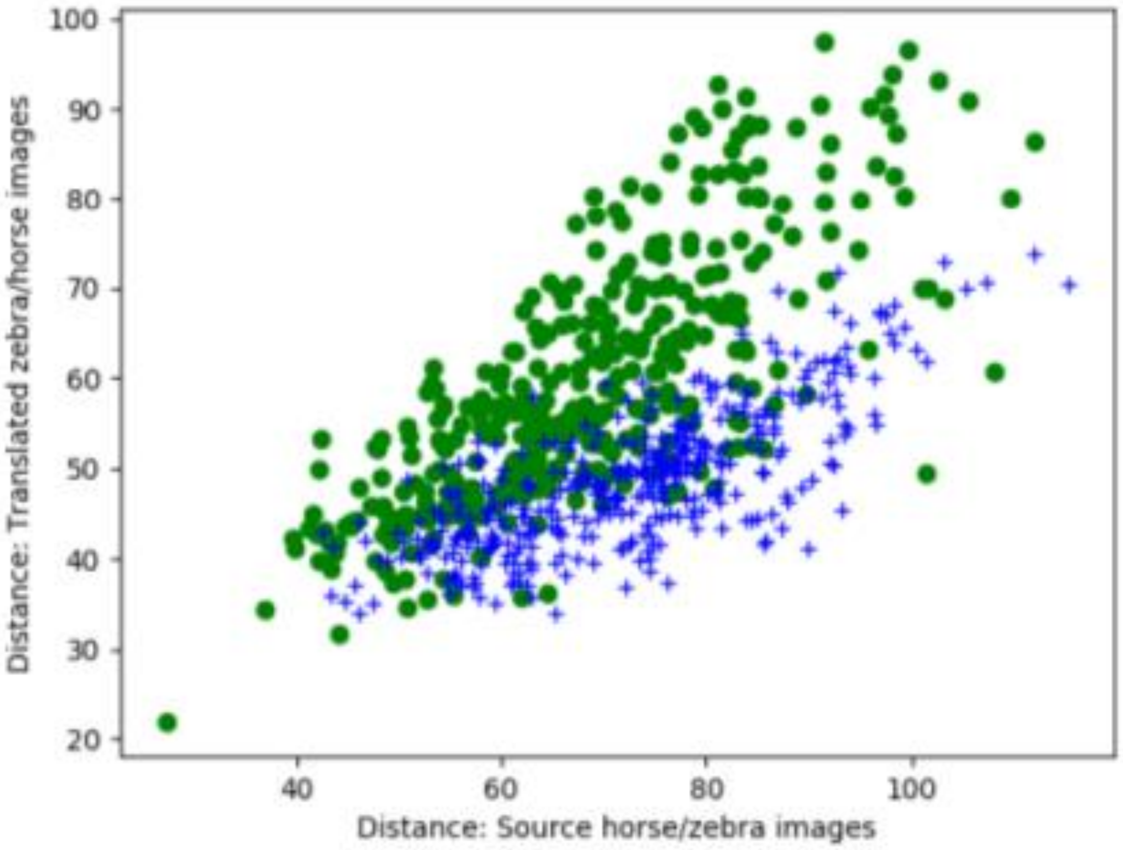
- A pair of images of a given distance are mapped to a pair of outputs with a similar distance
- $\left|x_i - x_j\right|_1$ and $\left|G(x_i) - G\left(x_j\right)\right|_1$ are highly correlated.



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

Benaim et al., NIPS 2017

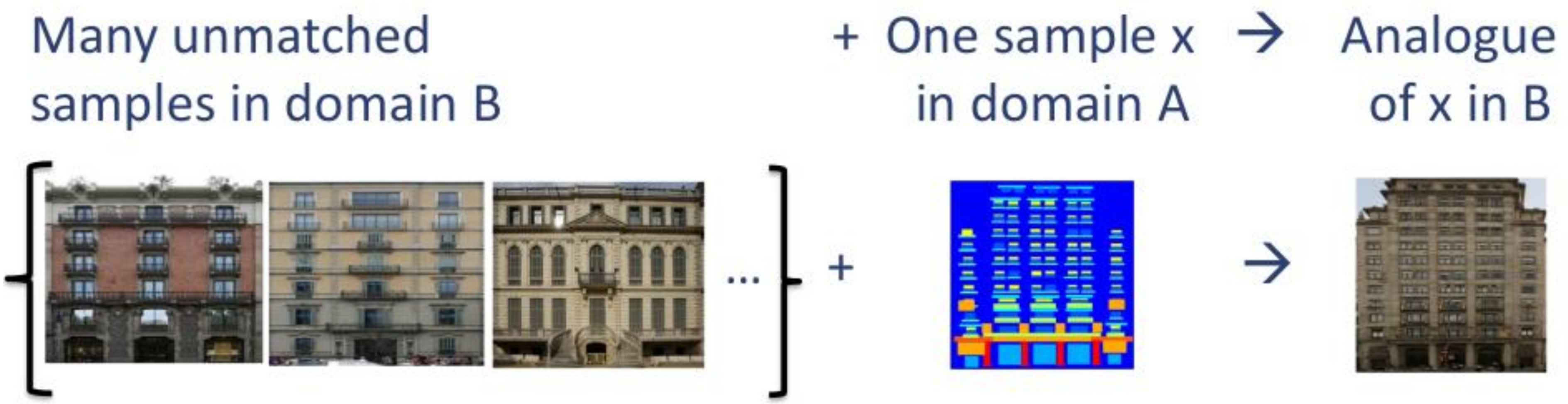
Motivating distance correlations



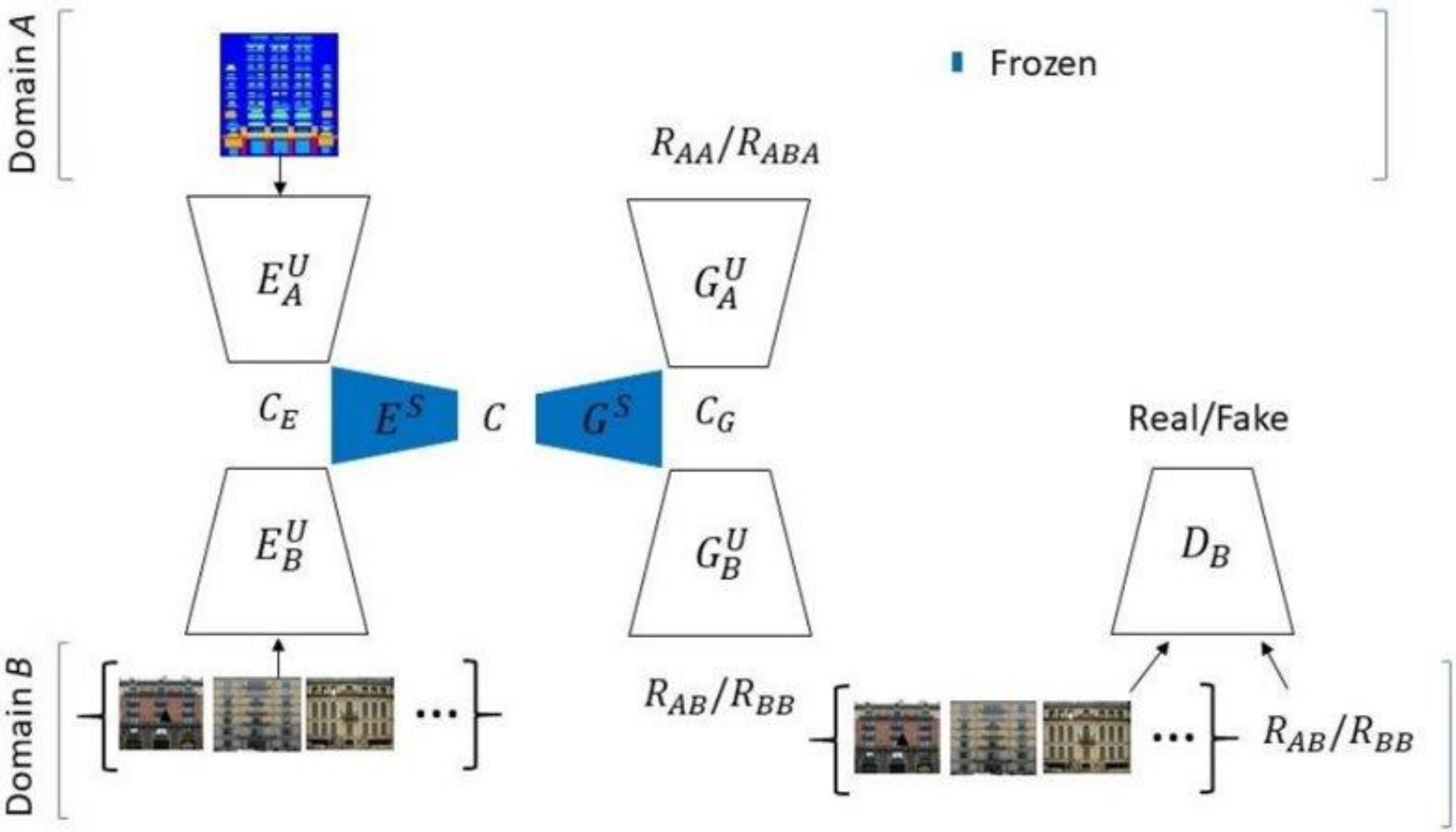
Analysis of CycleGAN's horse to zebra results

Benaim et al., NIPS 2017

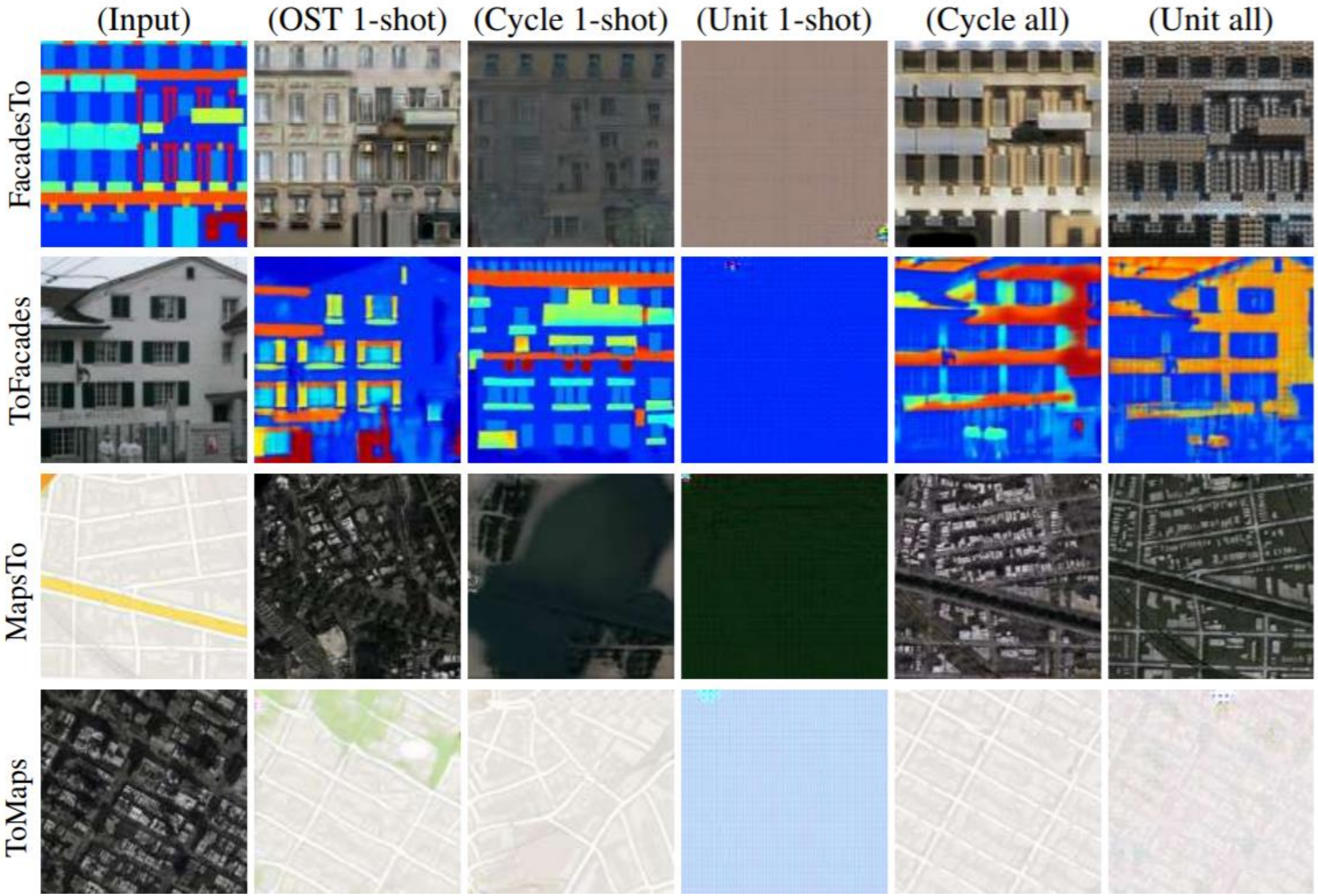
Less Supervision: Only a single image in domain A



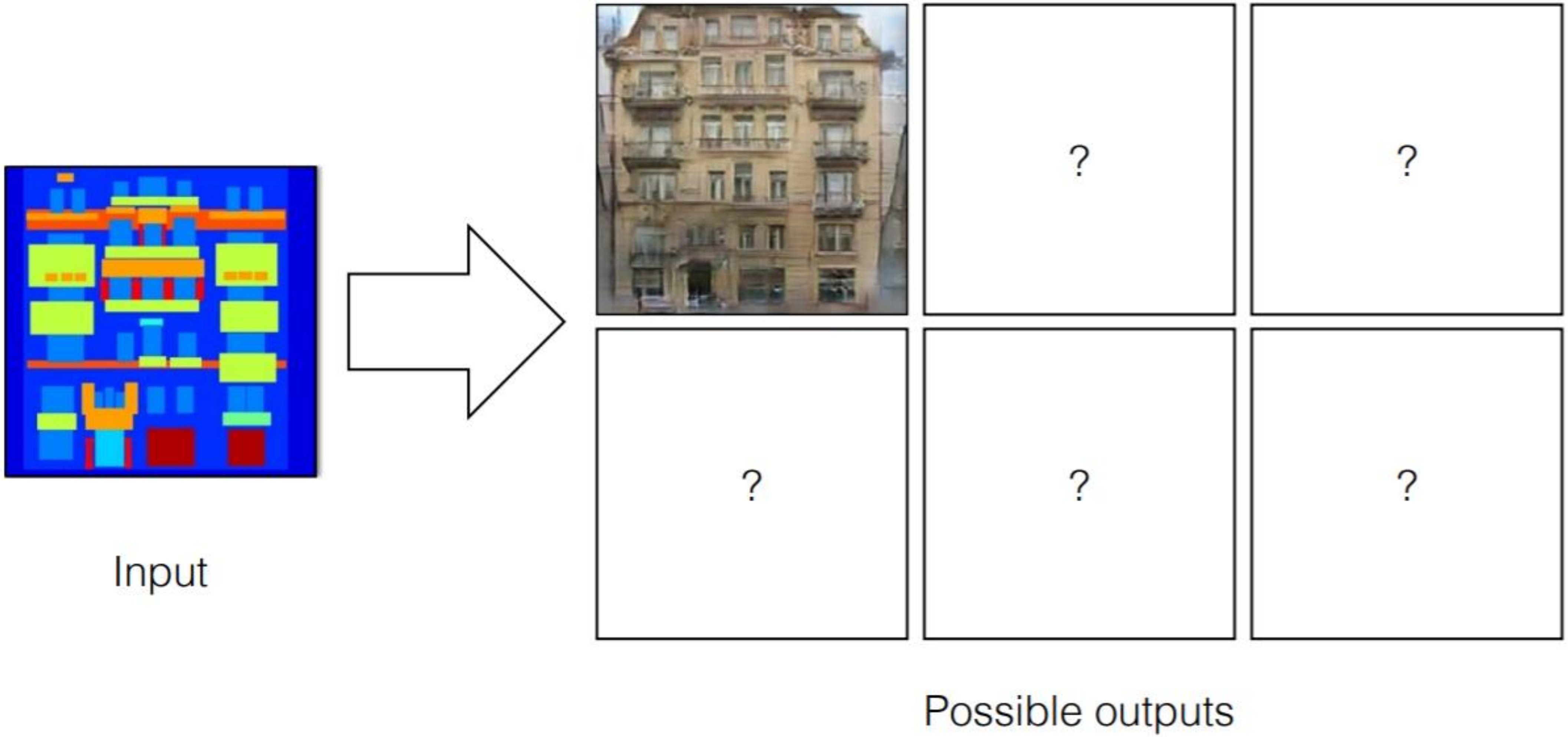
One Shot Unsupervised Cross Domain Translation (NeurIPS 2018)



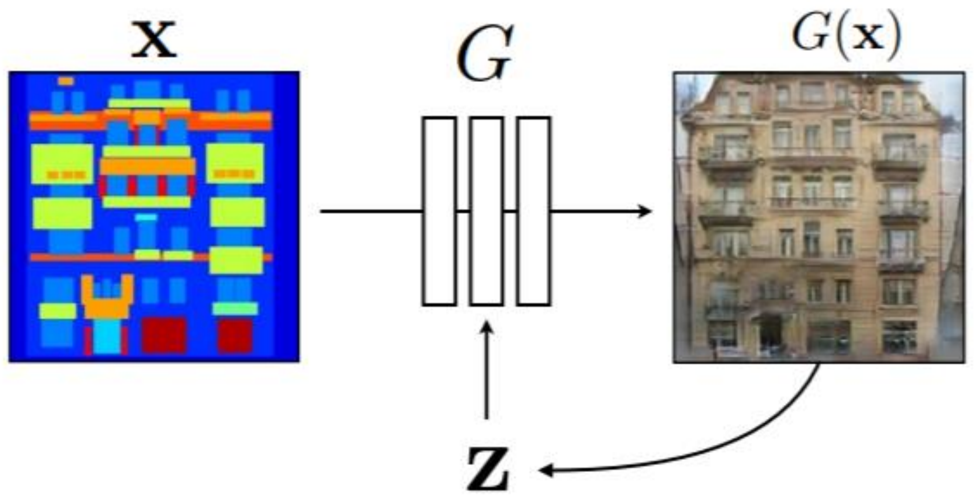
Domain A
Frozen
R_{AA}/R_{ABA}
E_A^U
G_A^U
C_E
E^S
C
G^S
C_G
Real/Fake
E_B^U
G_B^U
D_B
Domain B
R_{AB}/R_{BB}
R_{AB}/R_{BB}


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

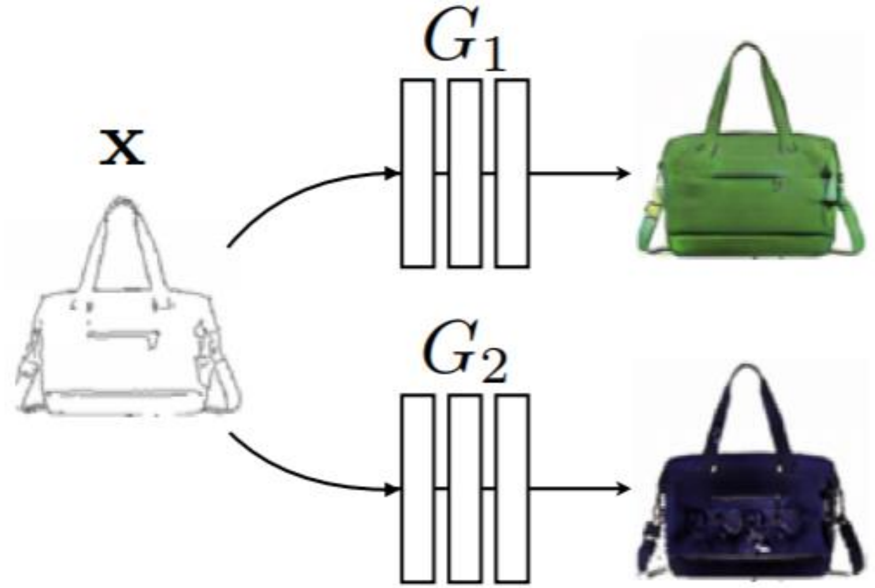
Modeling multiple possible outputs



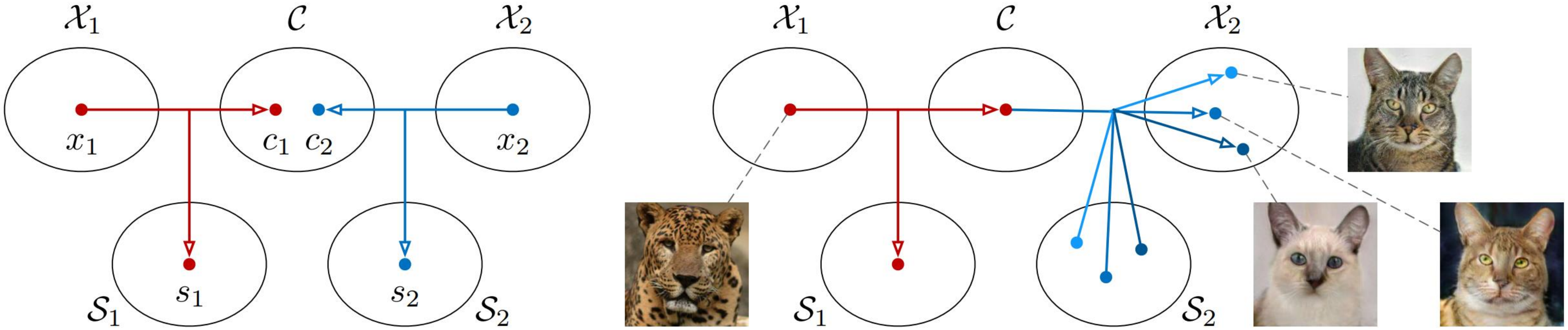
BiCycleGAN [Zhu et al., NIPS 2017]
(c.f. InfoGAN [Chen et al. 2016])



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



MUNIT: Multimodal Translation



(a) Auto-encoding (b) Translation

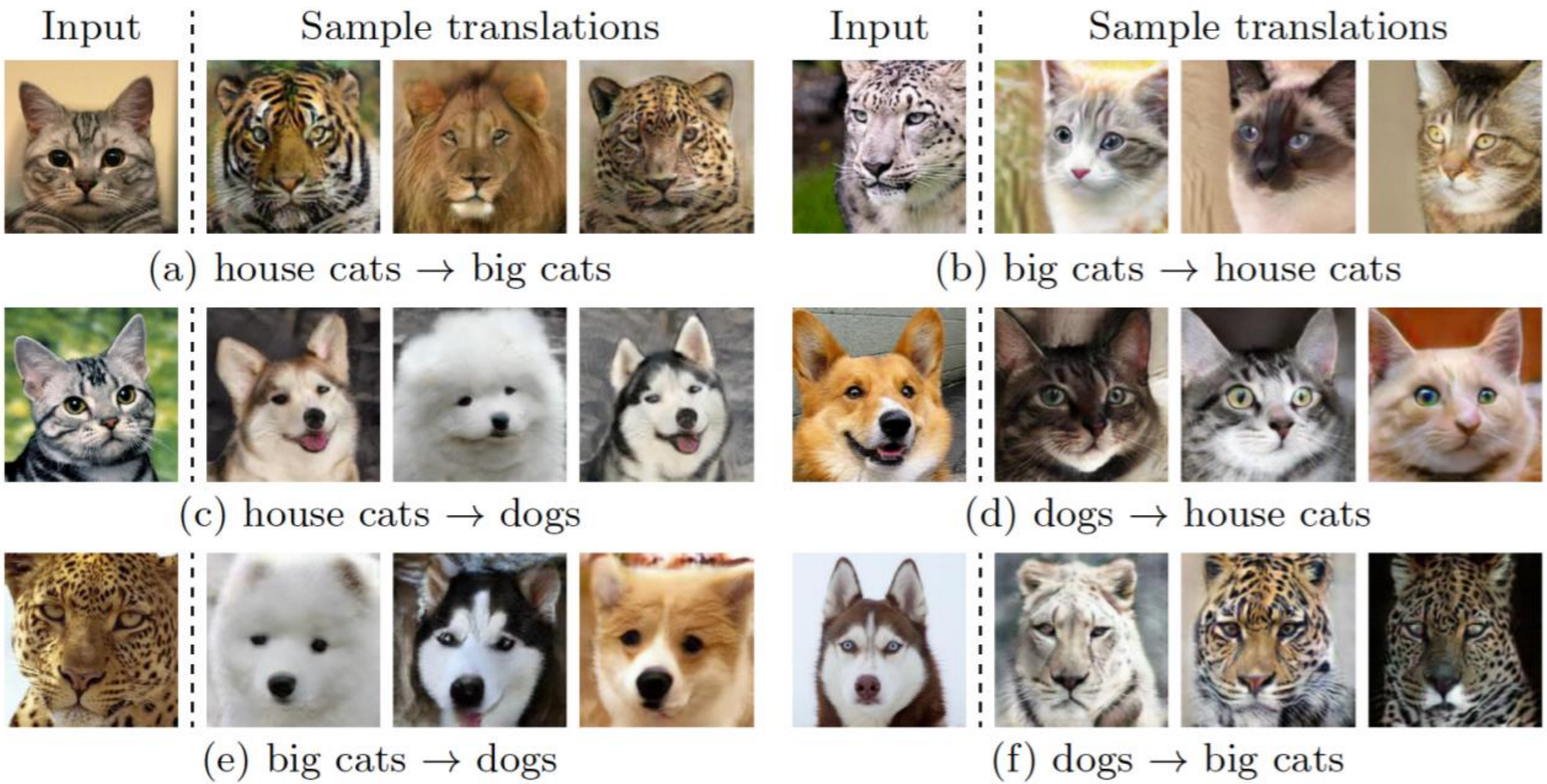
Huang et al., ECCV 2018

Sketch to Image Translation

(a) edges ↔ shoes (b) edges ↔ handbags

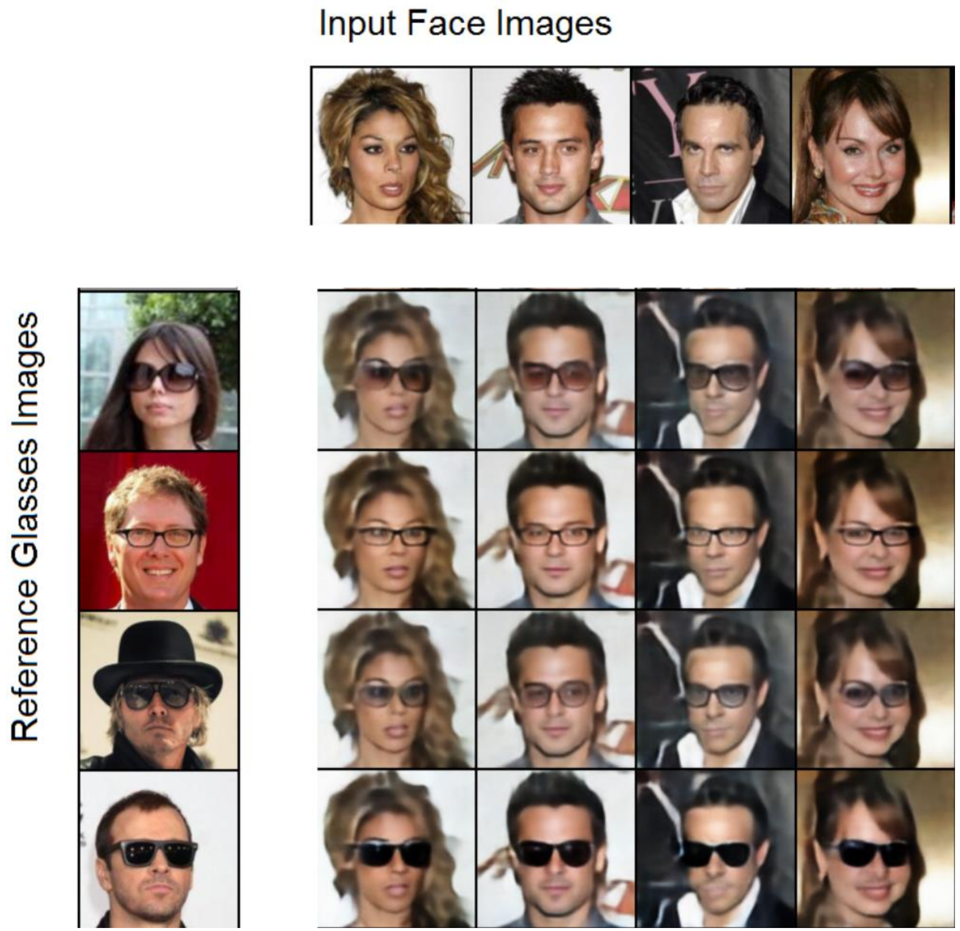
Huang et al., ECCV 2018

Animal Image Translation

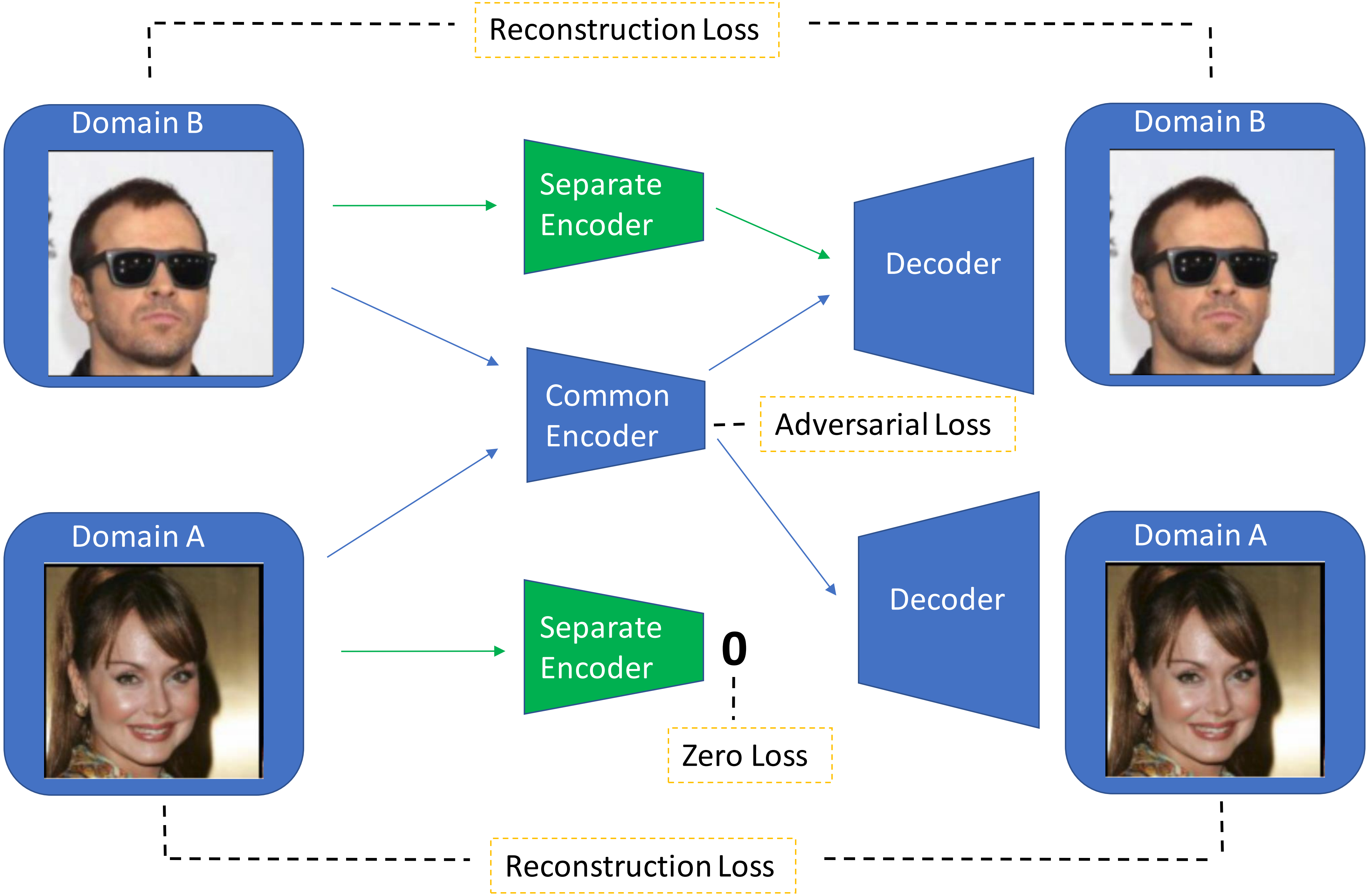


Huang et al., ECCV 2018

Full Content Disentanglement

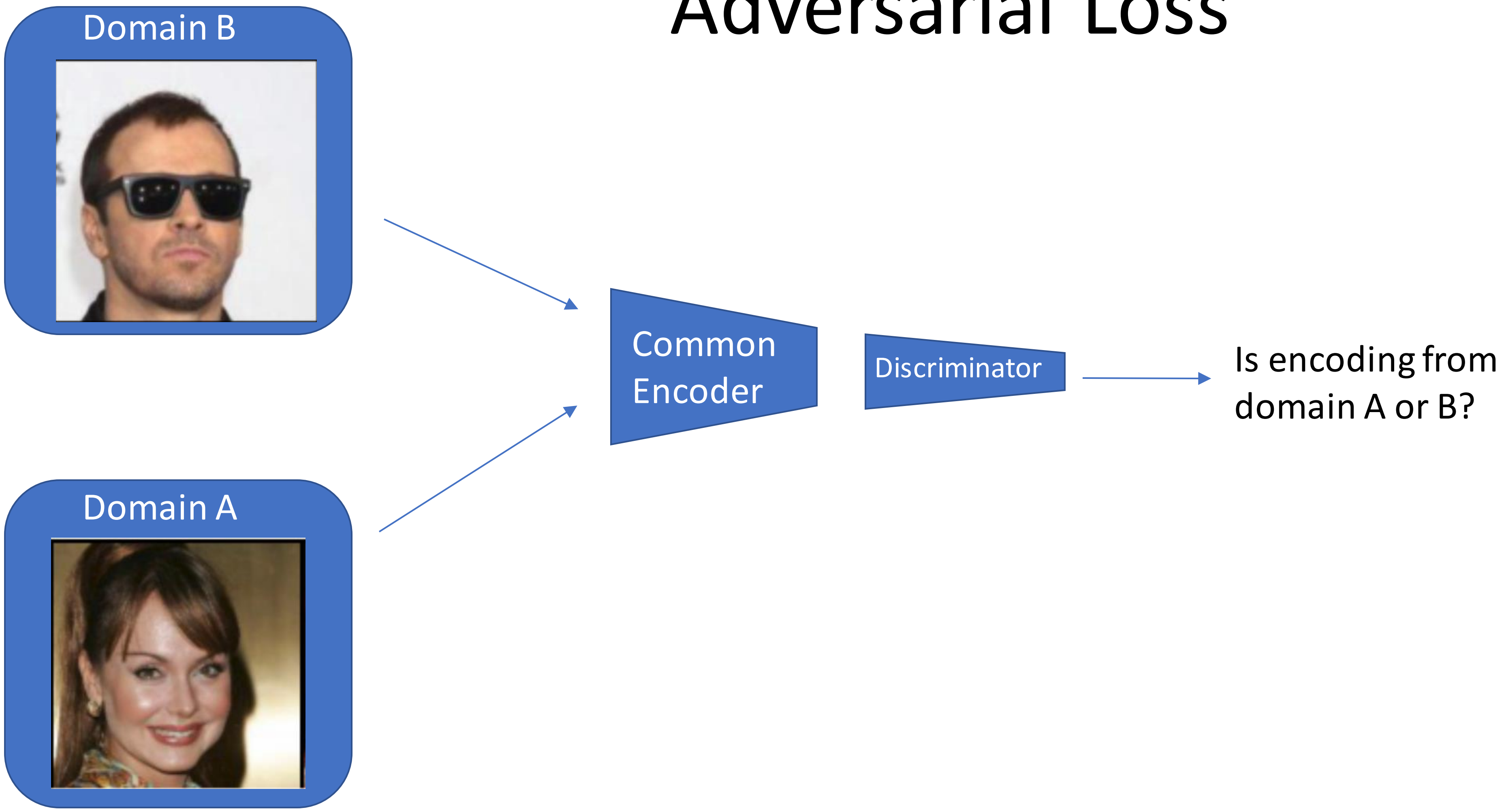


"Emerging Disentanglement in Auto-Encoder Based Unsupervised Image Content Transfer", ICLR 2019



Reconstruction Loss
Domain B
Separate Encoder
Decoder
Domain B
Common Encoder
Adversarial Loss
Domain A
Separate Encoder
0
Decoder
Domain A
Zero Loss
Reconstruction Loss

Adversarial Loss



Other Domains?

- Audio Separation: Training data consists of a set of samples of mixed music and an unmatched set of instrumental music.
- Given a mixed sample, wish the separate the voice from the background instrumental music.
- After mapping the audio sample to a Spectrogram, can subtract the “background” from the “mixed” sample in “pixel space”, to get the “voice” only sample.
- Samples at: https://sagiebenaim.github.io/Singing/

"Semi-Supervised Monaural Singing Voice Separation With a Masking Network Trained on Synthetic Mixtures." ICASSP 2019

Video to Video

- Use GAN to generate each from in a video
- Use optical flow to further constrain the generator
- Samples at: https://github.com/NVIDIA/vid2vid

"High Resolution photorealistic video to video translation." NeurIPS 2018

Many More Applications

- 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?