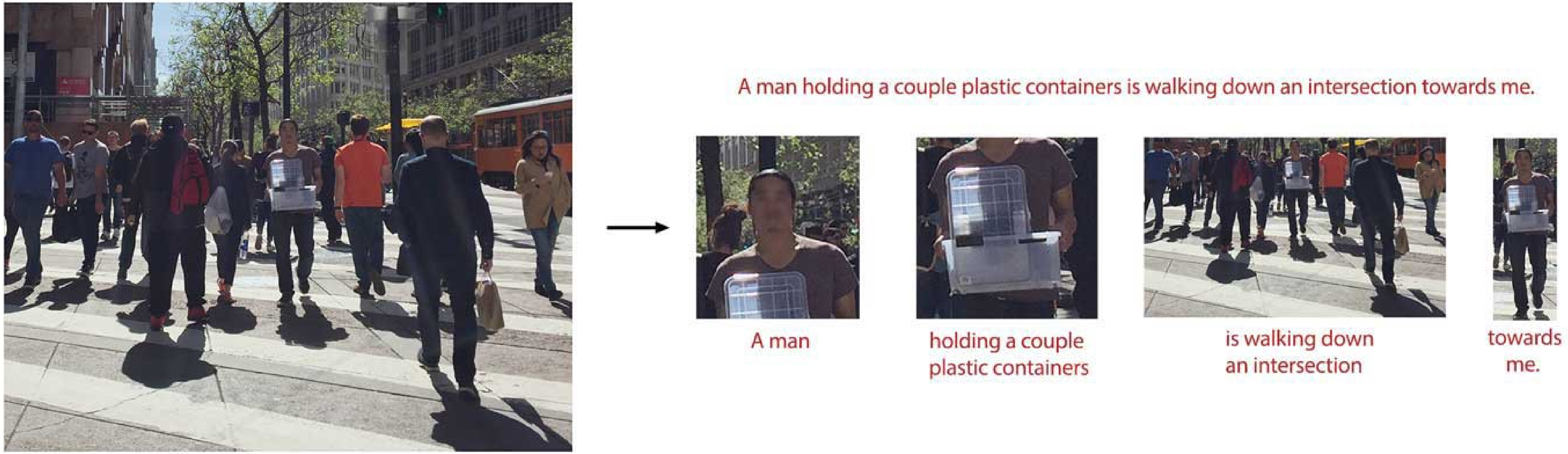
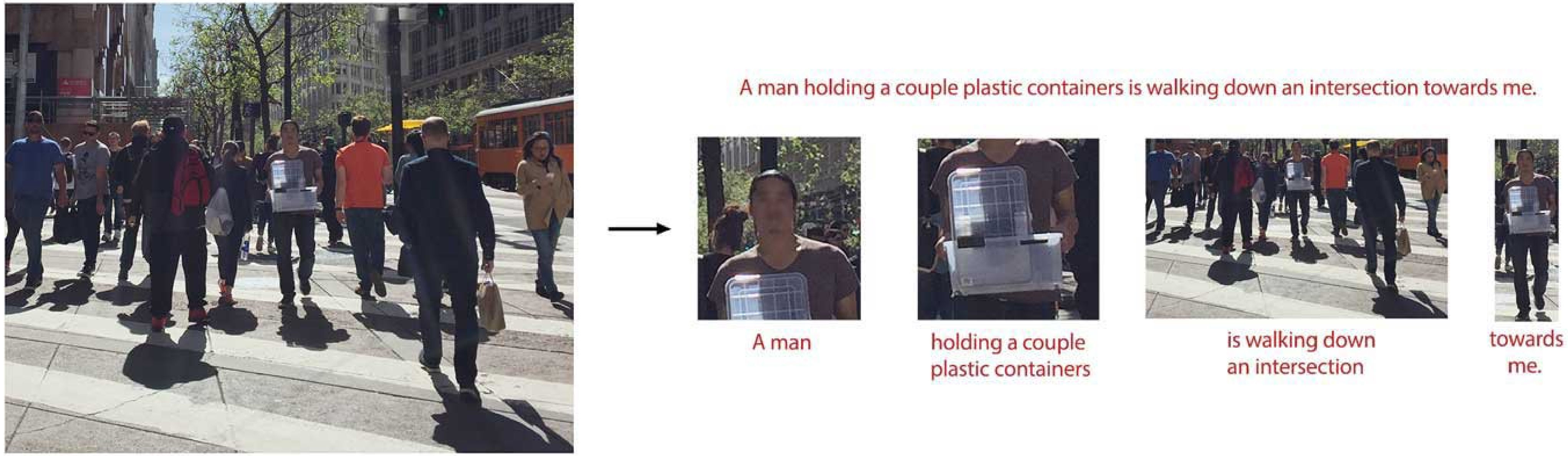
Self-Attention Generative Adversarial Networks (SAGAN)

- attention improves accuracy in language translation and image captioning
  - an image captioning deep network focuses on different areas of the image to generate words in the caption.

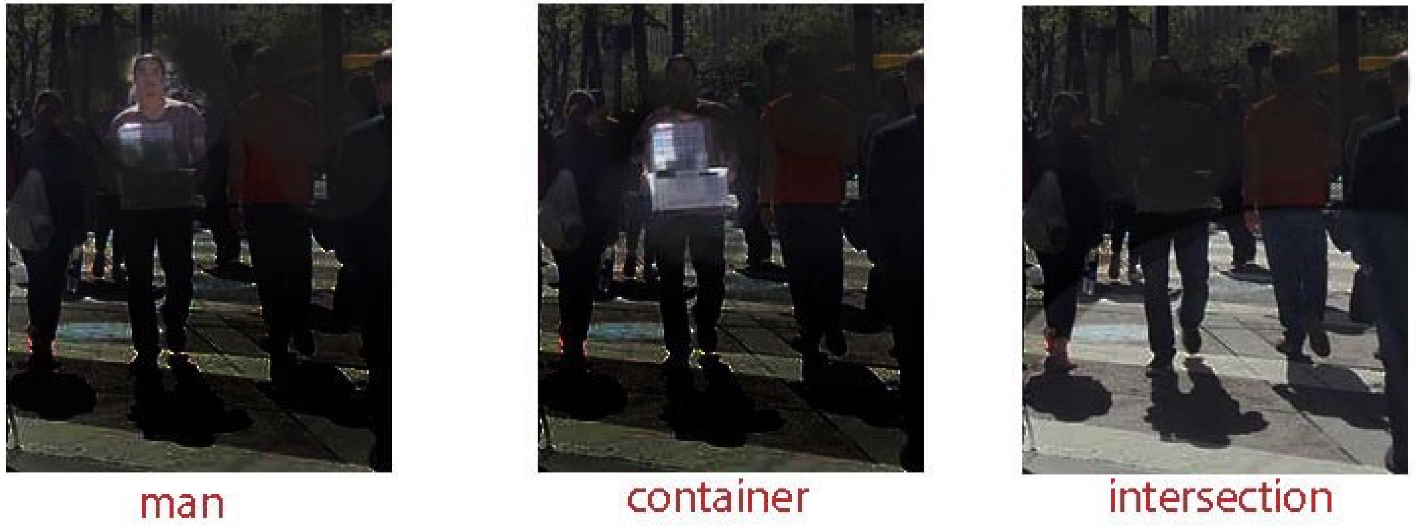


Motivation

- GAN models trained with ImageNet are good at classes with a lot of texture
- but perform much worse for structure.
- While convolutional filters are good at exploring spatial locality information, the receptive fields may not be large enough to cover larger structures.
- We can increase the filter size or the depth of the deep network but this will make GANs even harder to train.



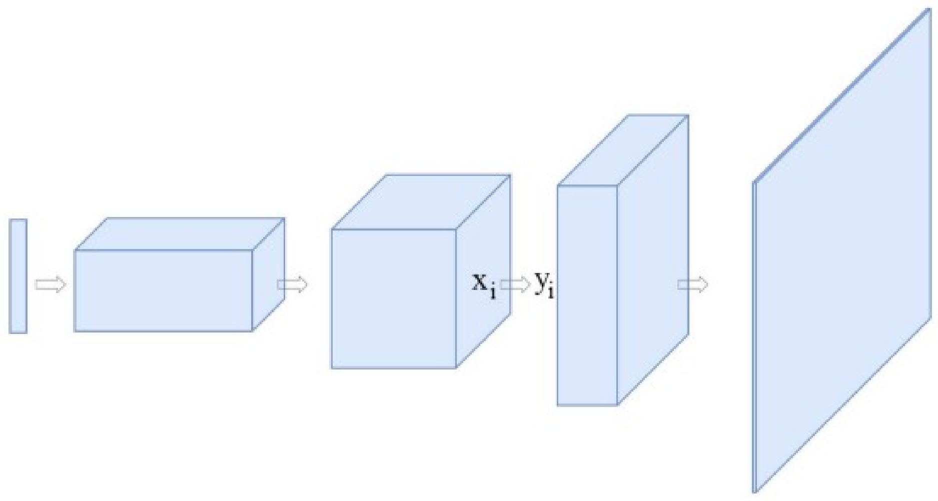
The highlighted area below is the attention area where the network focuses on in generating the specific word.

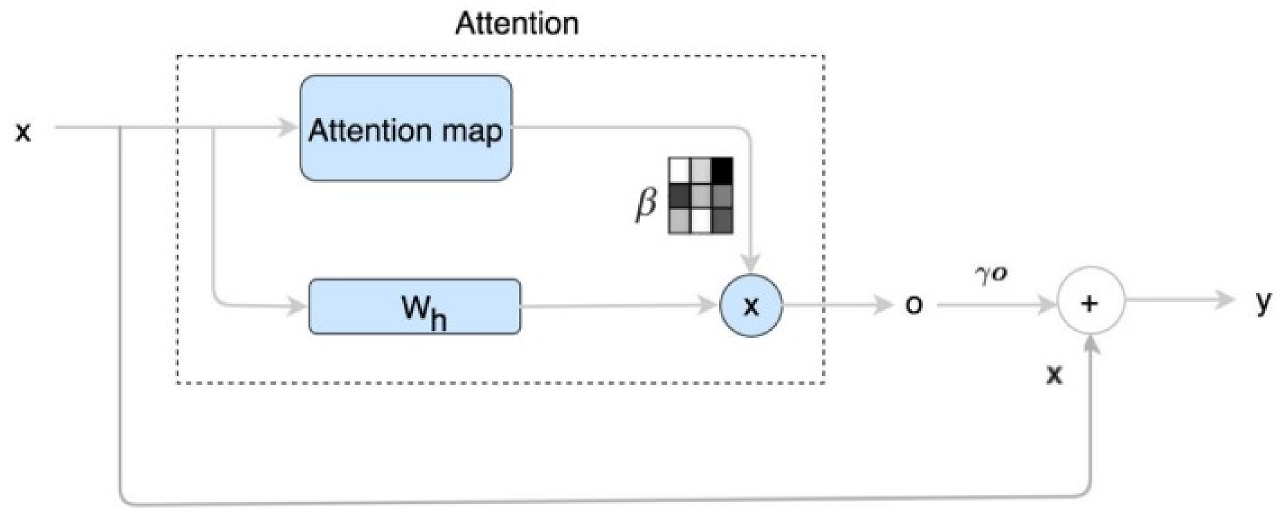


- To refine the image quality of the eye region (the red dot on the left figure), SAGAN only uses the feature map region on the highlighted area in the middle figure.

Design

- For each convolutional layer, they refine each spatial location output with an extra term computed by the self-attention mechanism. $y_i = \gamma o_i + x_i$
- SAGAN uses hinge loss to train the network:





$$L_D = -\mathbb{E}_{(x,y)\sim p_{data}}[\min(0, -1 + D(x, y))] - \mathbb{E}_{z\sim p_z, y\sim p_{data}}[\min(0, -1 - D(G(z), y))],$$
$$L_G = -\mathbb{E}_{z\sim p_z, y\sim p_{data}} D(G(z), y),$$

Unpaired Image-to-Image Translation with CycleGAN

Jun-Yan Zhu and Taesung Park
Joint work with Phillip Isola and Alexei A. Efros

BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH
CSAIL

Image-to-Image Translation with **pix2pix**

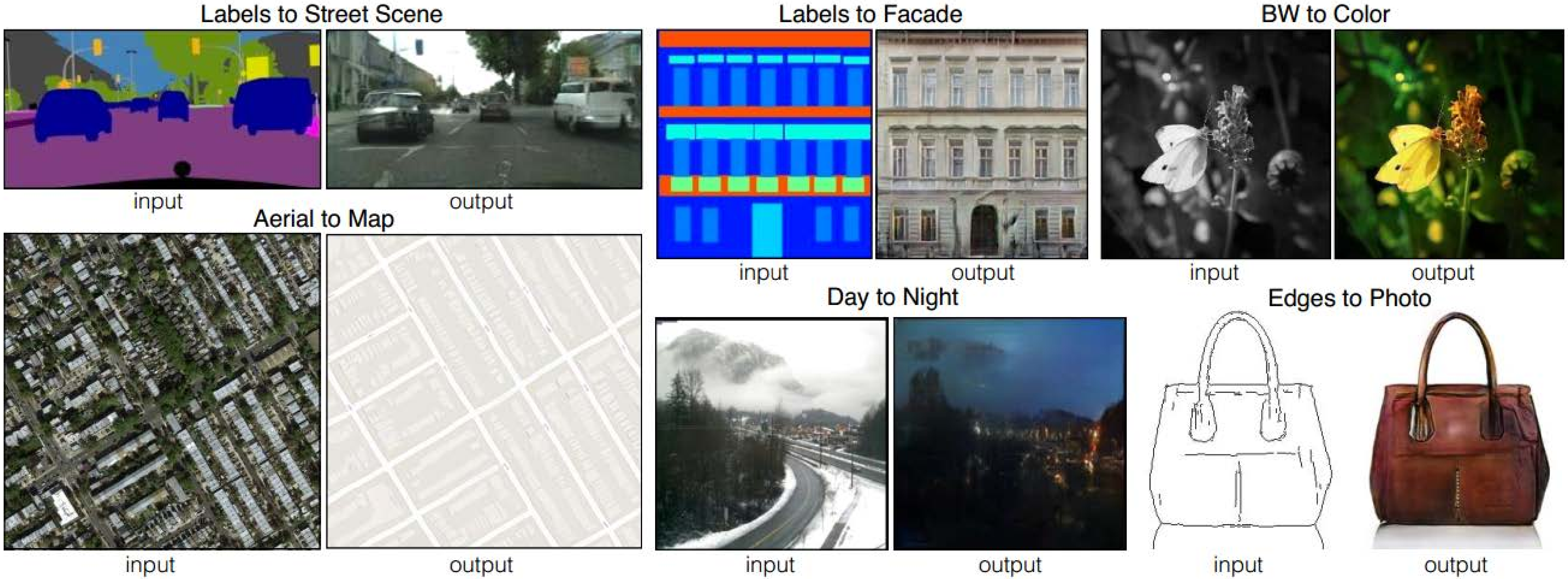
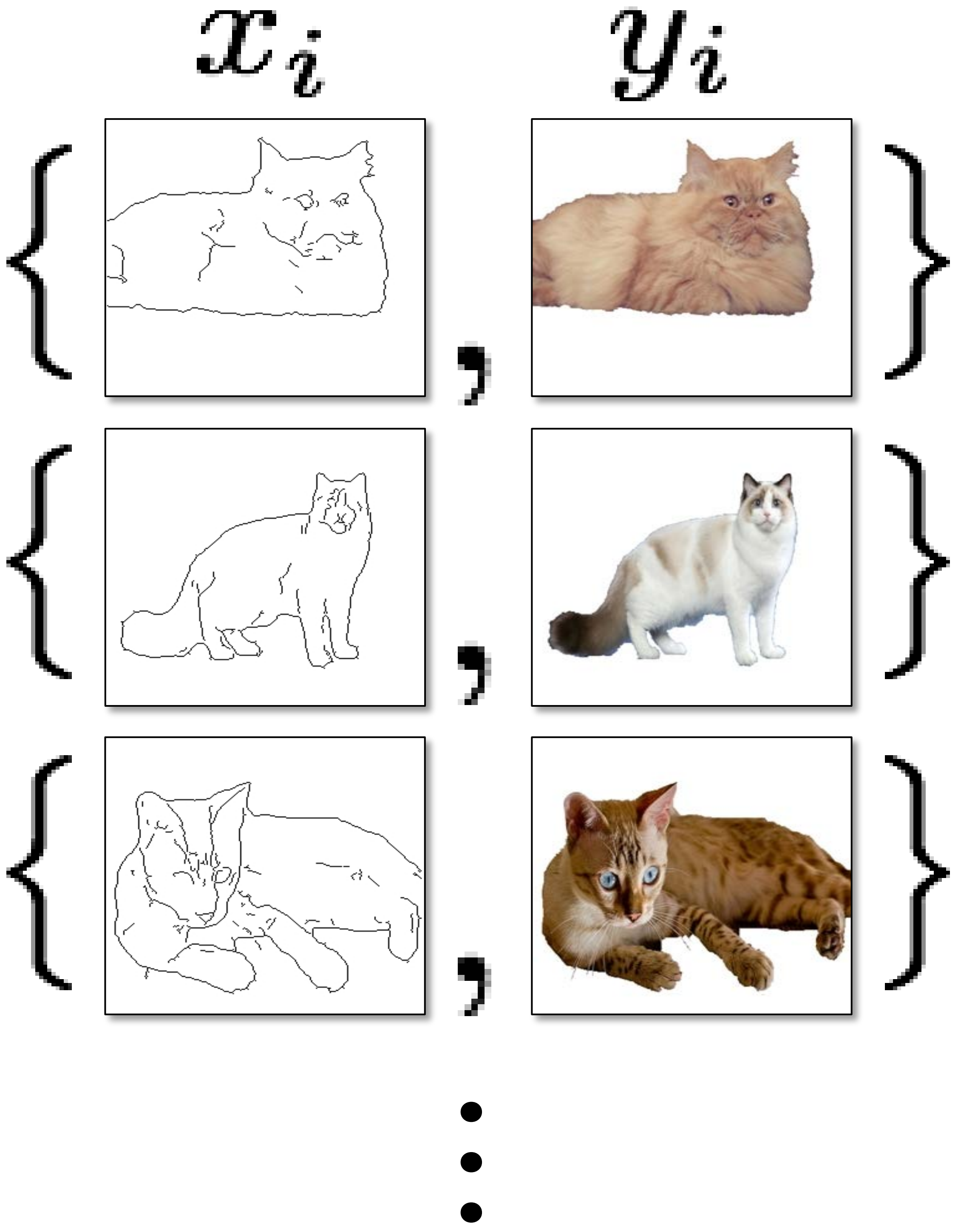
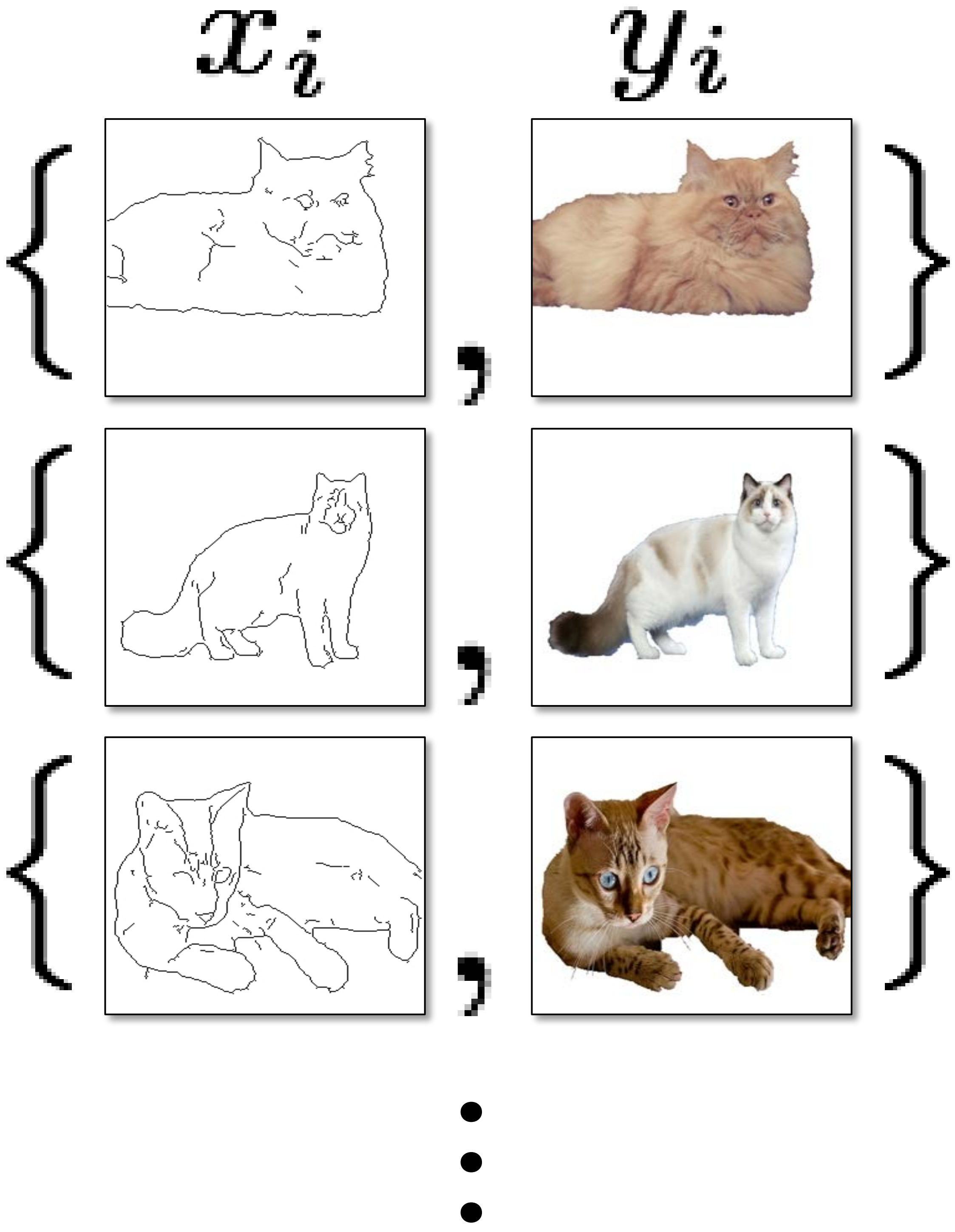


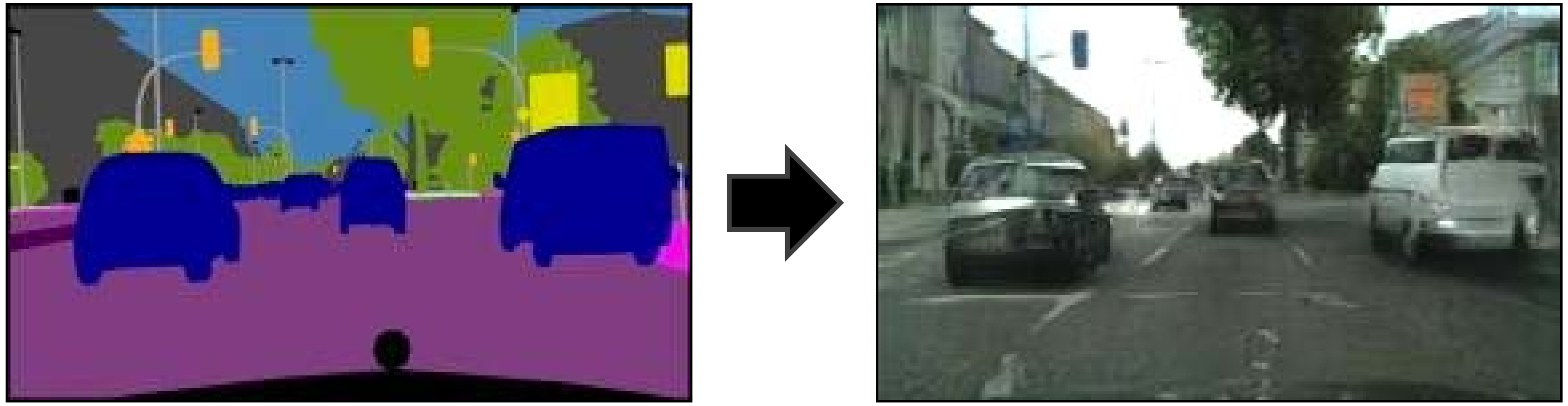
Image-to-image Translation with Conditional Adversarial Nets
Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. CVPR 2017

Paired

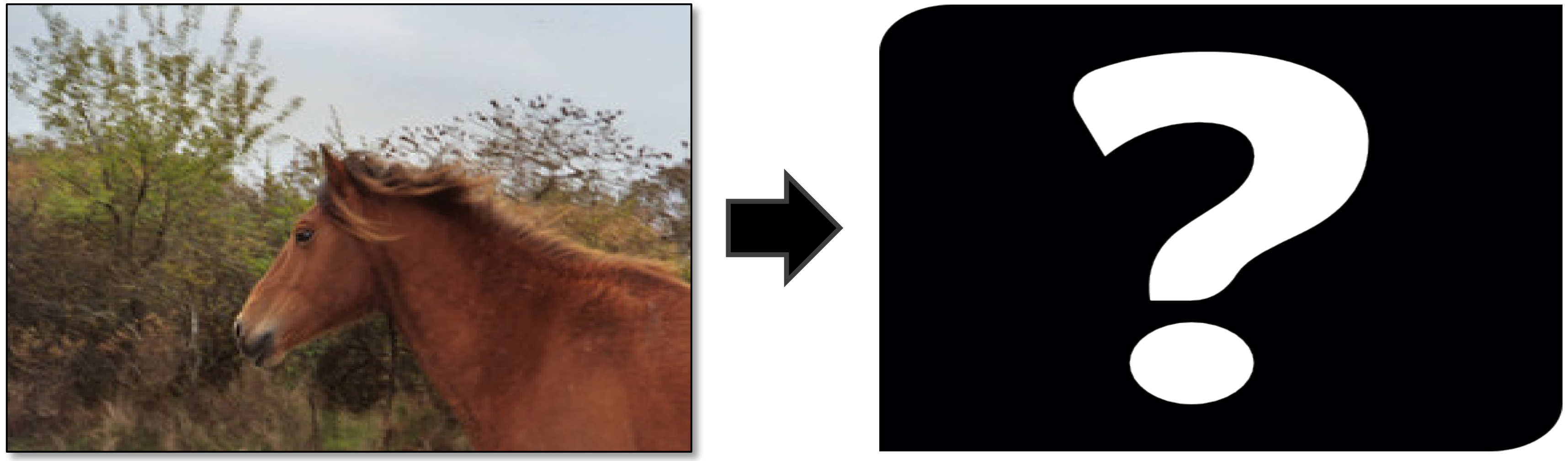


Paired



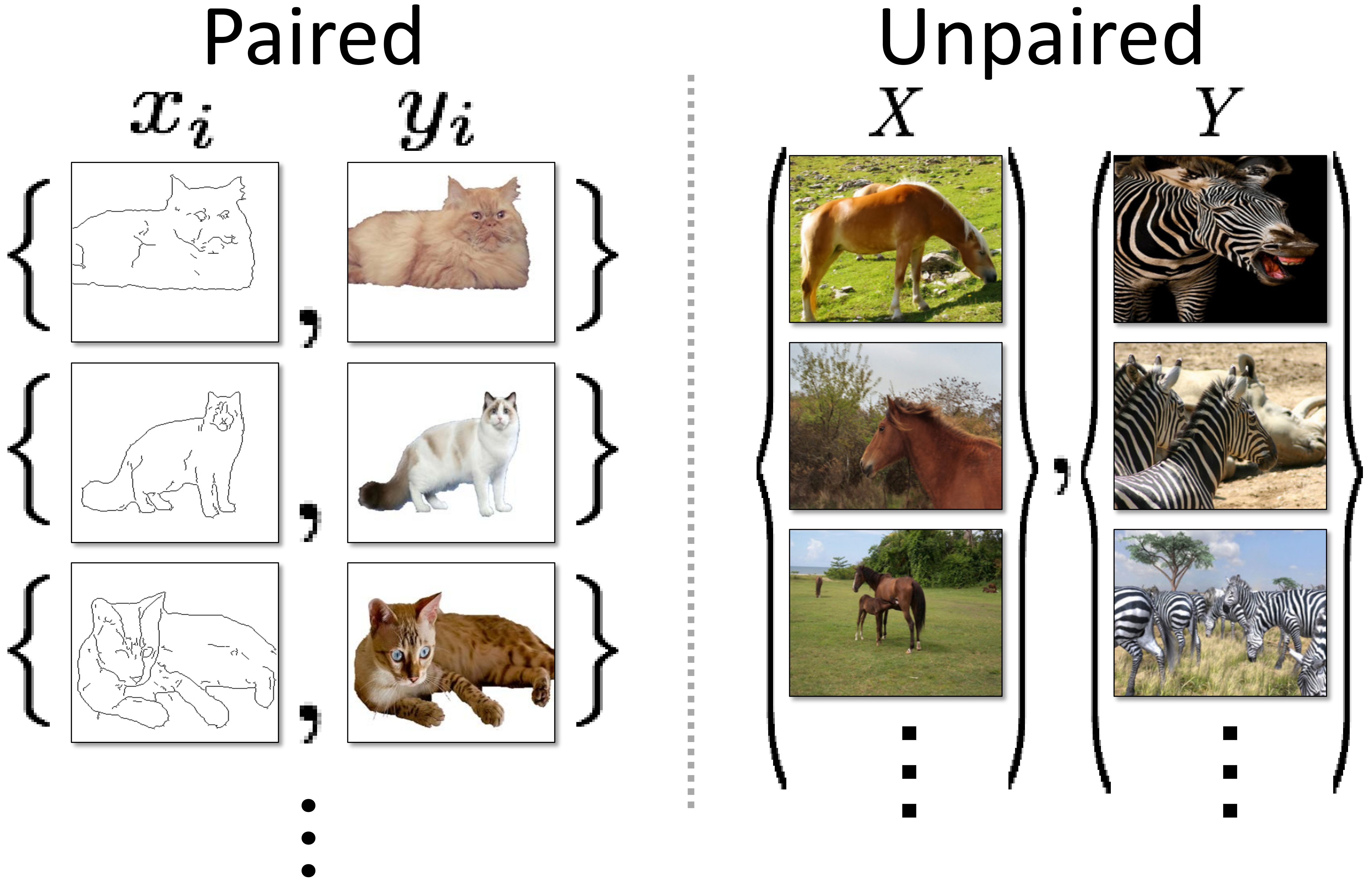


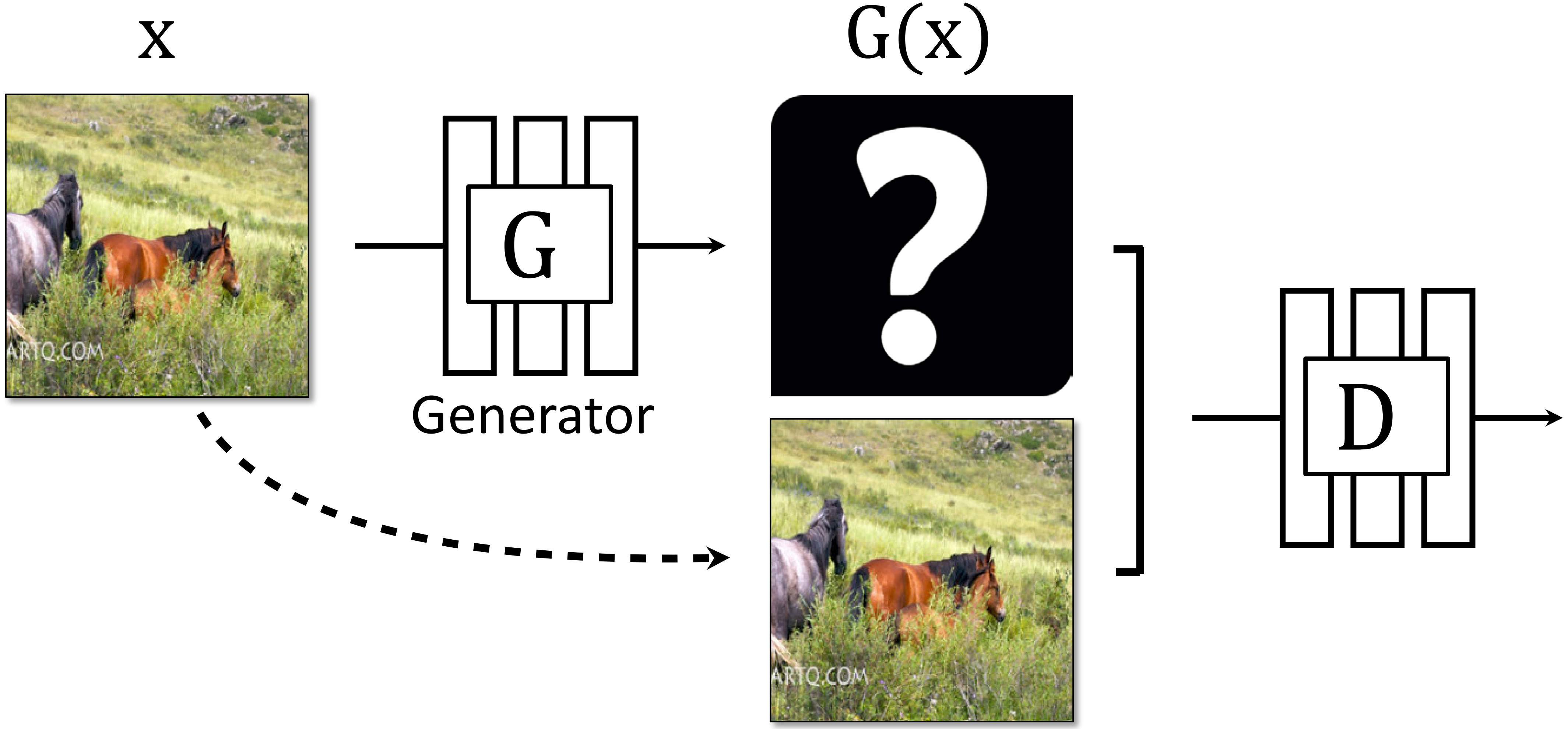
Label ↔ photo: per-pixel labeling



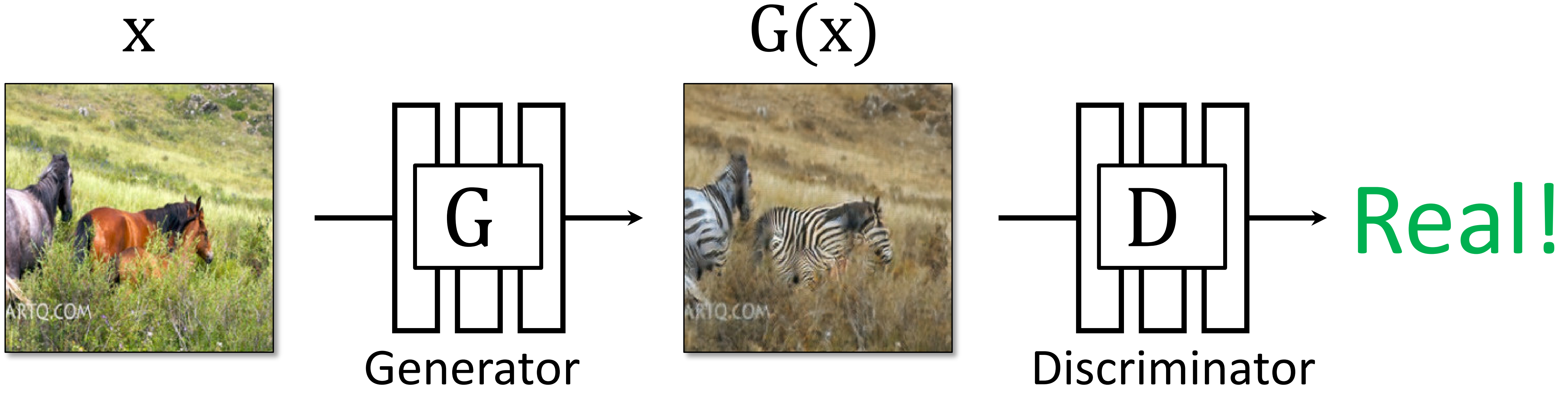
Horse ↔ zebra: how to get zebras?

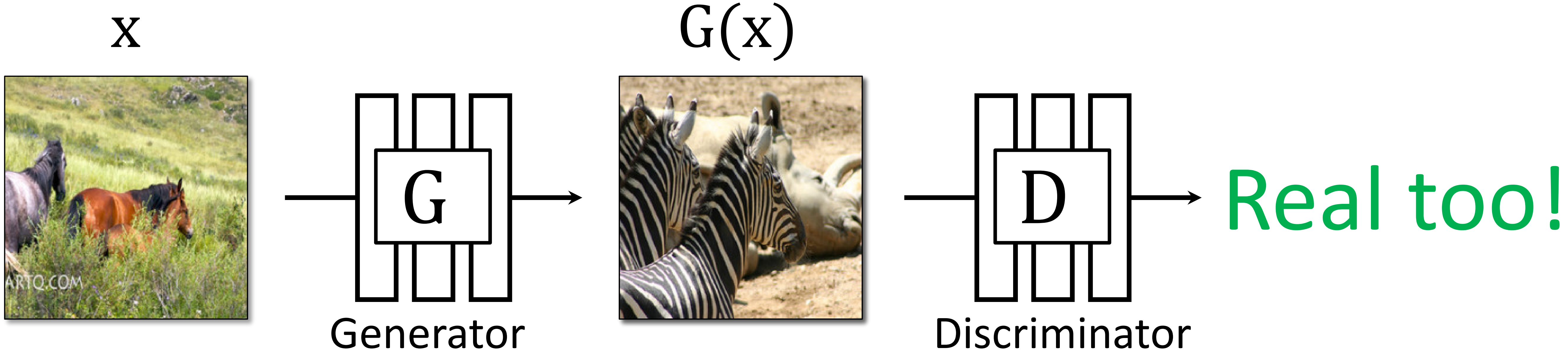
- Expensive to collect pairs.
- Impossible in many scenarios.


Paired
x_i
y_i
Unpaired
X
Y

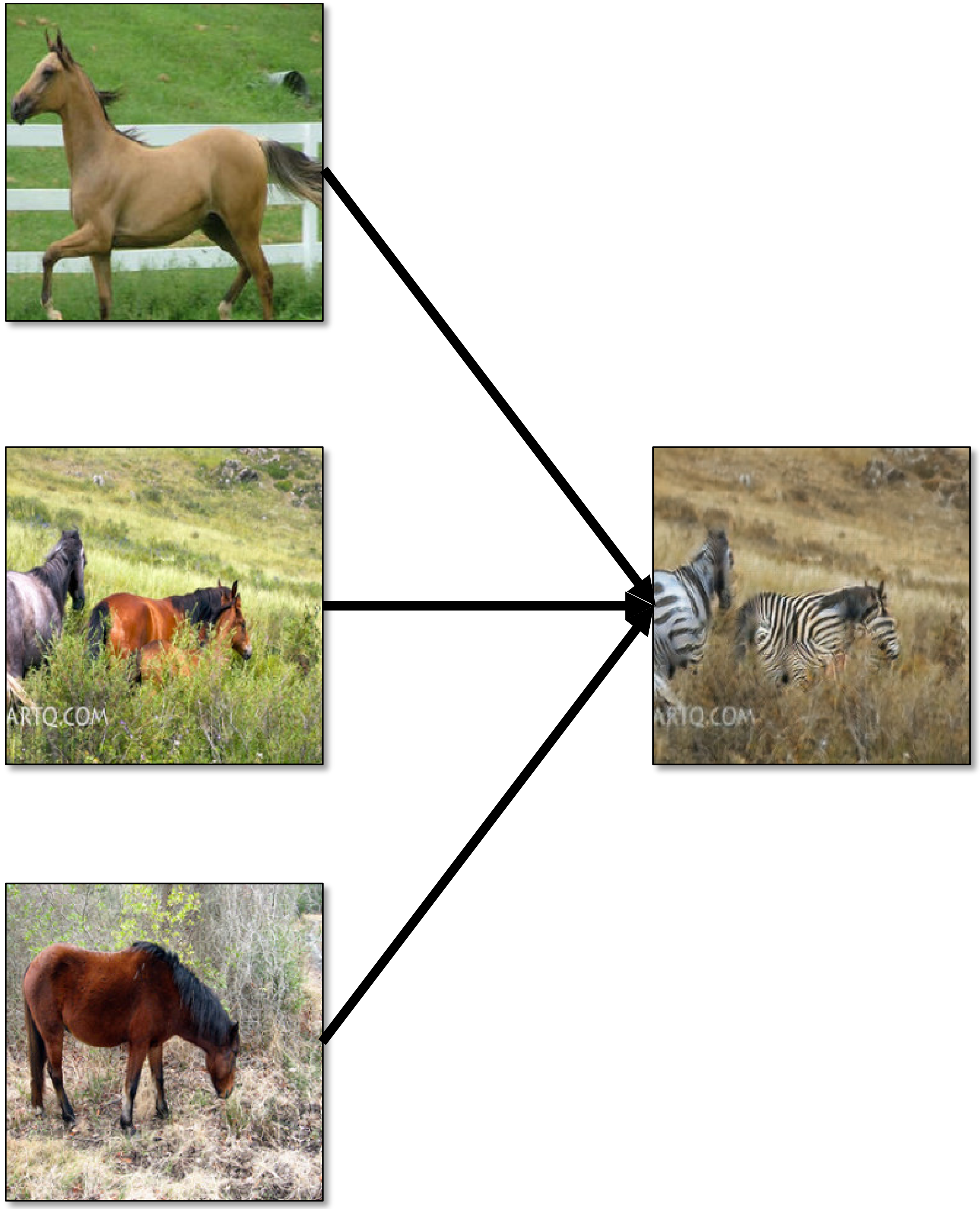


No input-output pairs!


x
G(x)
G
Generator
D
Discriminator
Real!

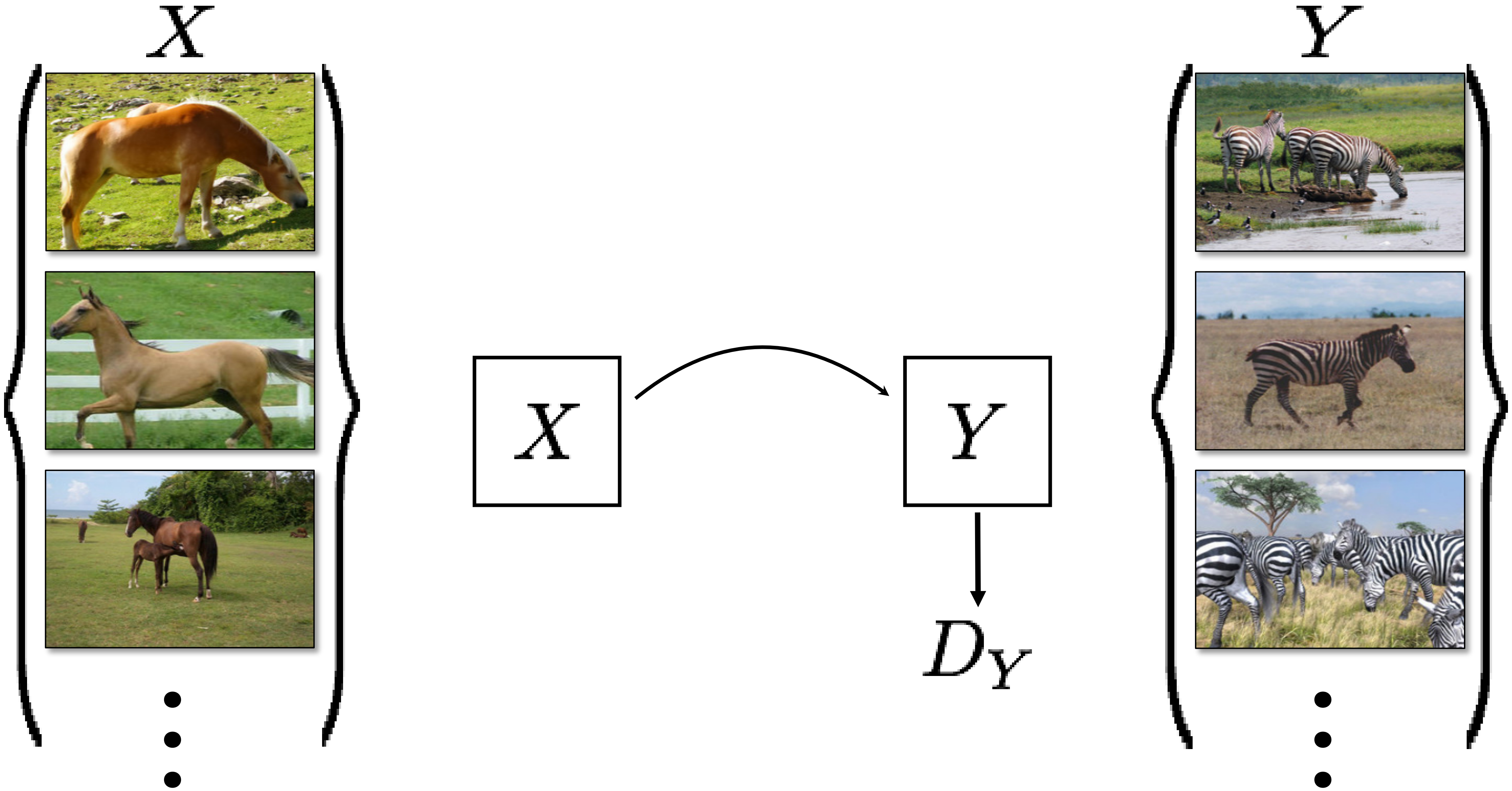


GANs do **not** force output to correspond to input



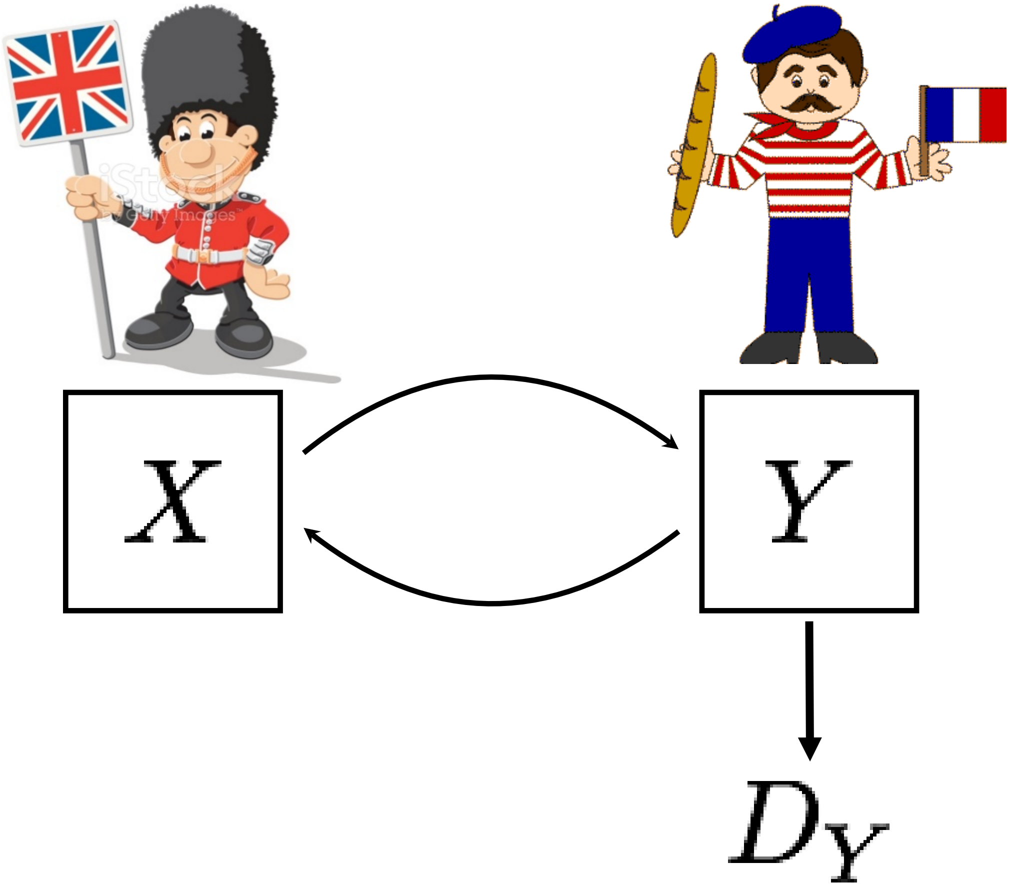
mode collapse!

Cycle-Consistent Adversarial Networks



[Zhu*, Park*, Isola, and Efros, ICCV 2017]

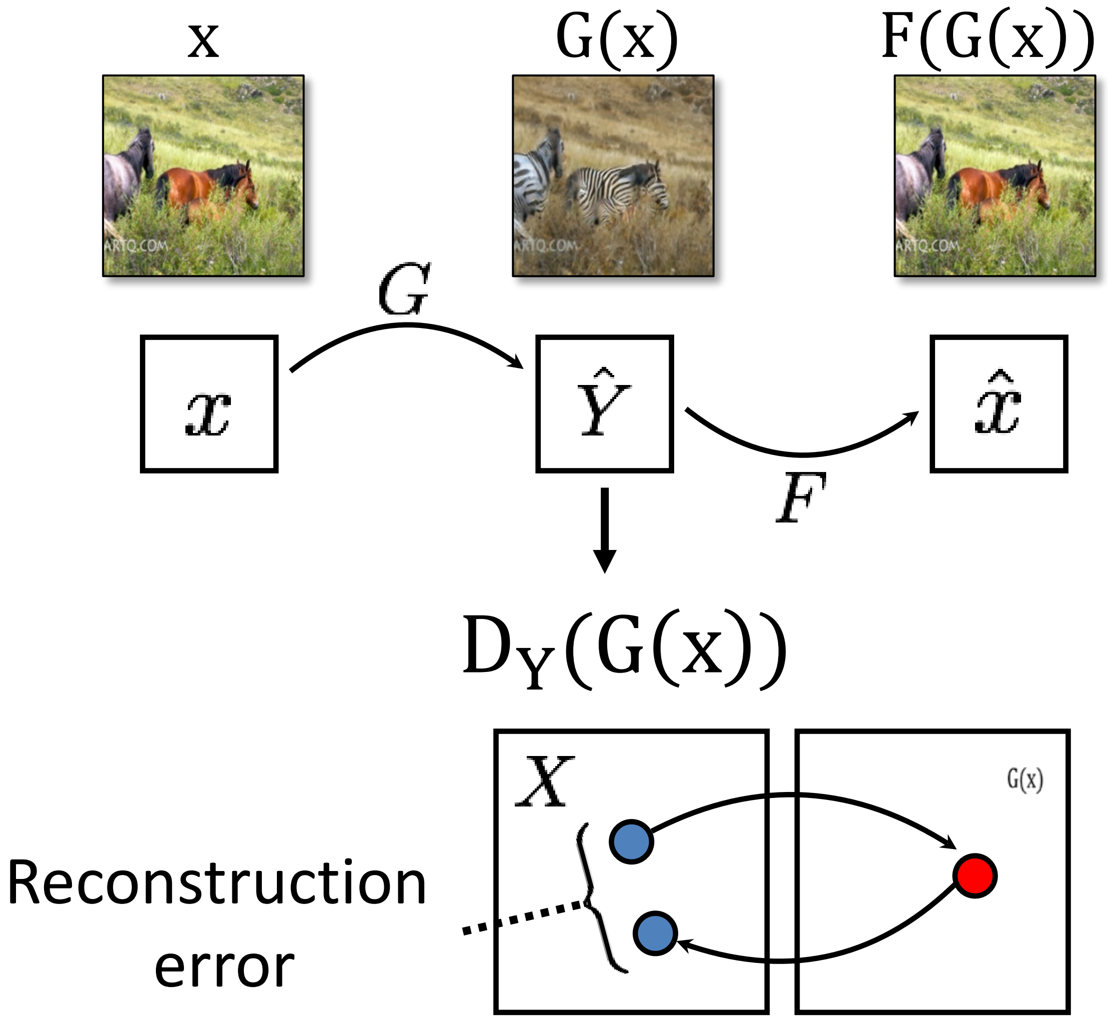
Cycle-Consistent Adversarial Networks



[Mark Twain, 1903]

[Zhu*, Park*, Isola, and Efros, ICCV 2017]

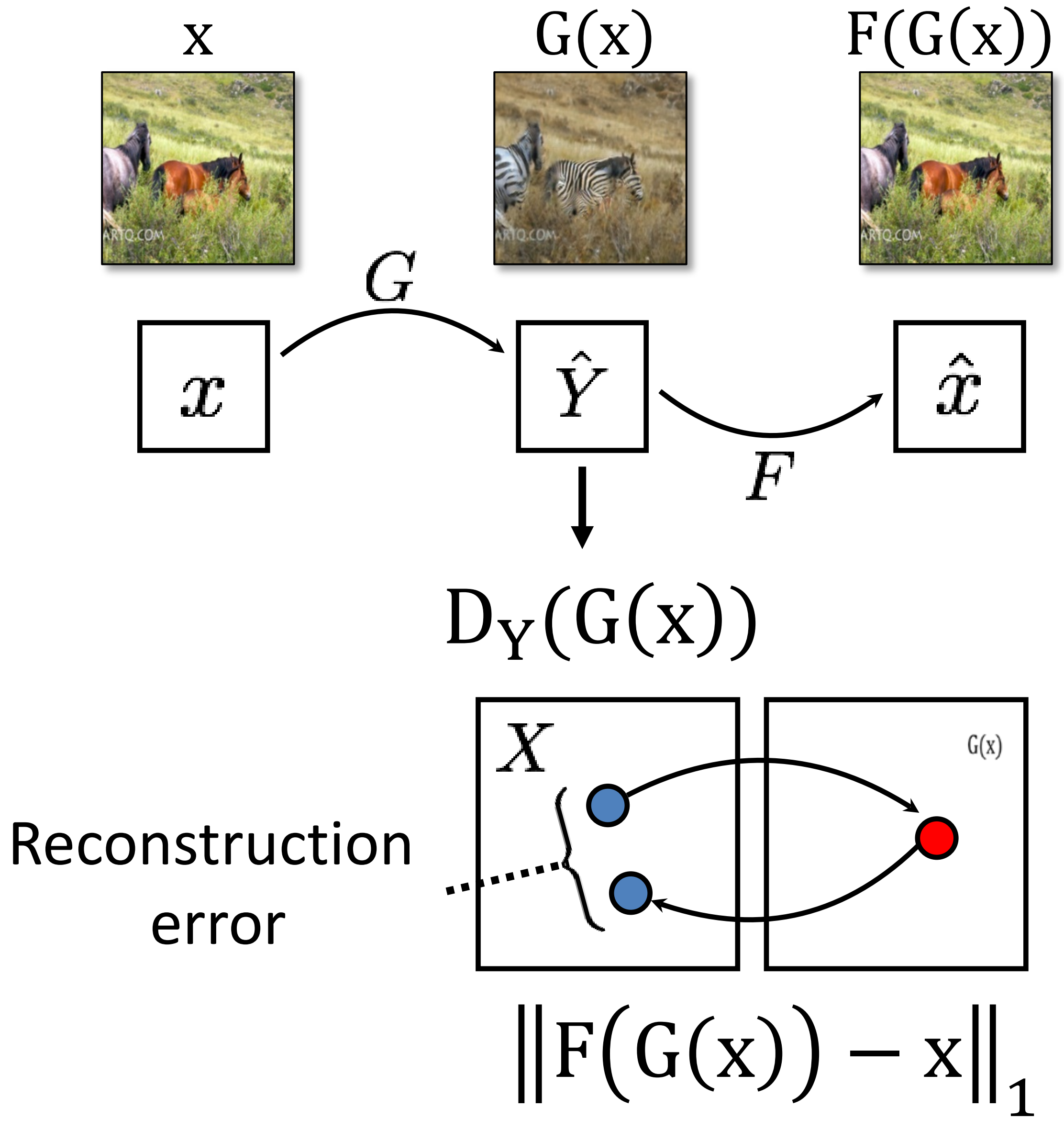
Cycle-Consistent Adversarial Networks

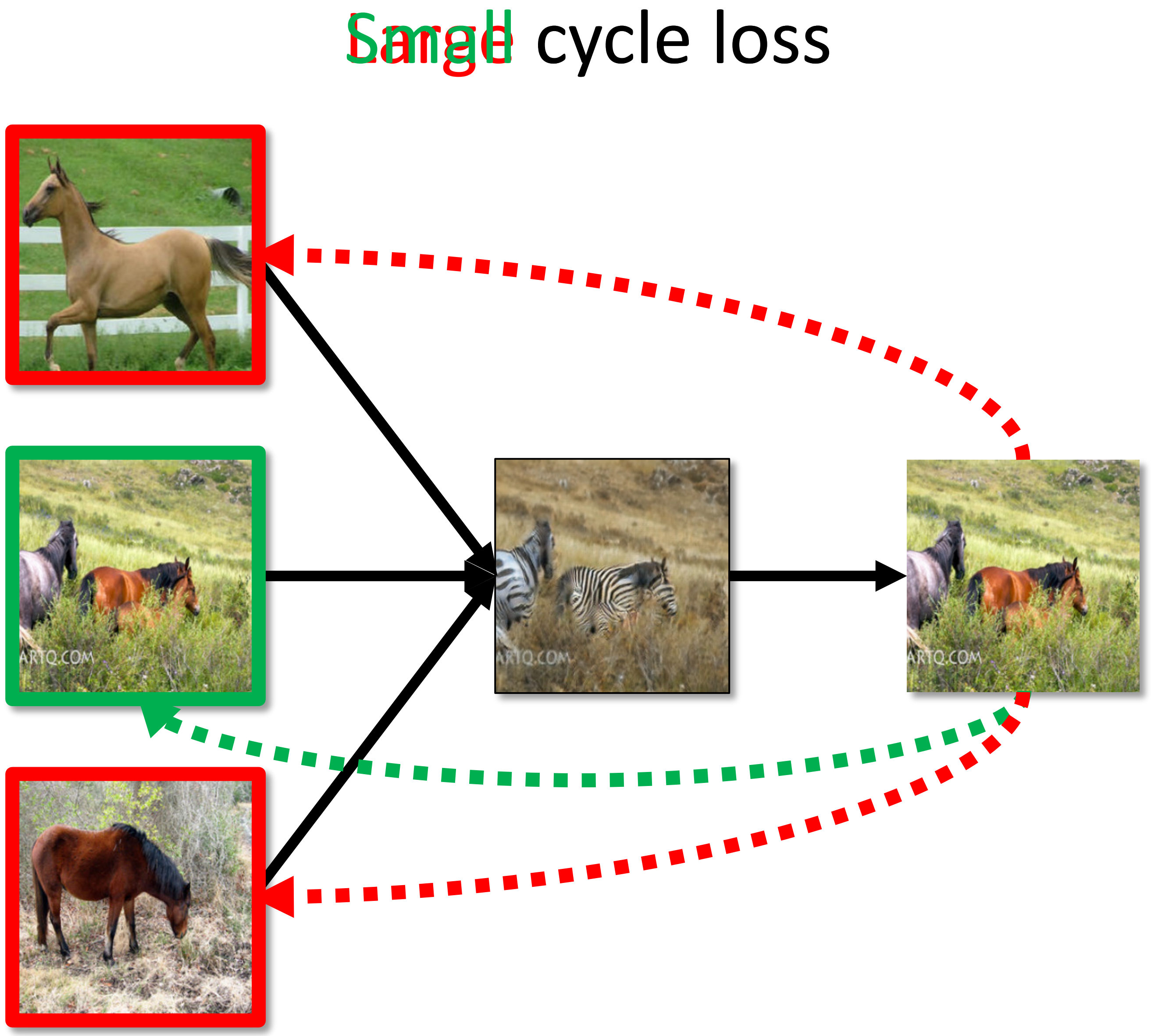


$$\|F(G(x)) - x\|_1$$

[Zhu*, Park*, Isola, and Efros, ICCV 2017]

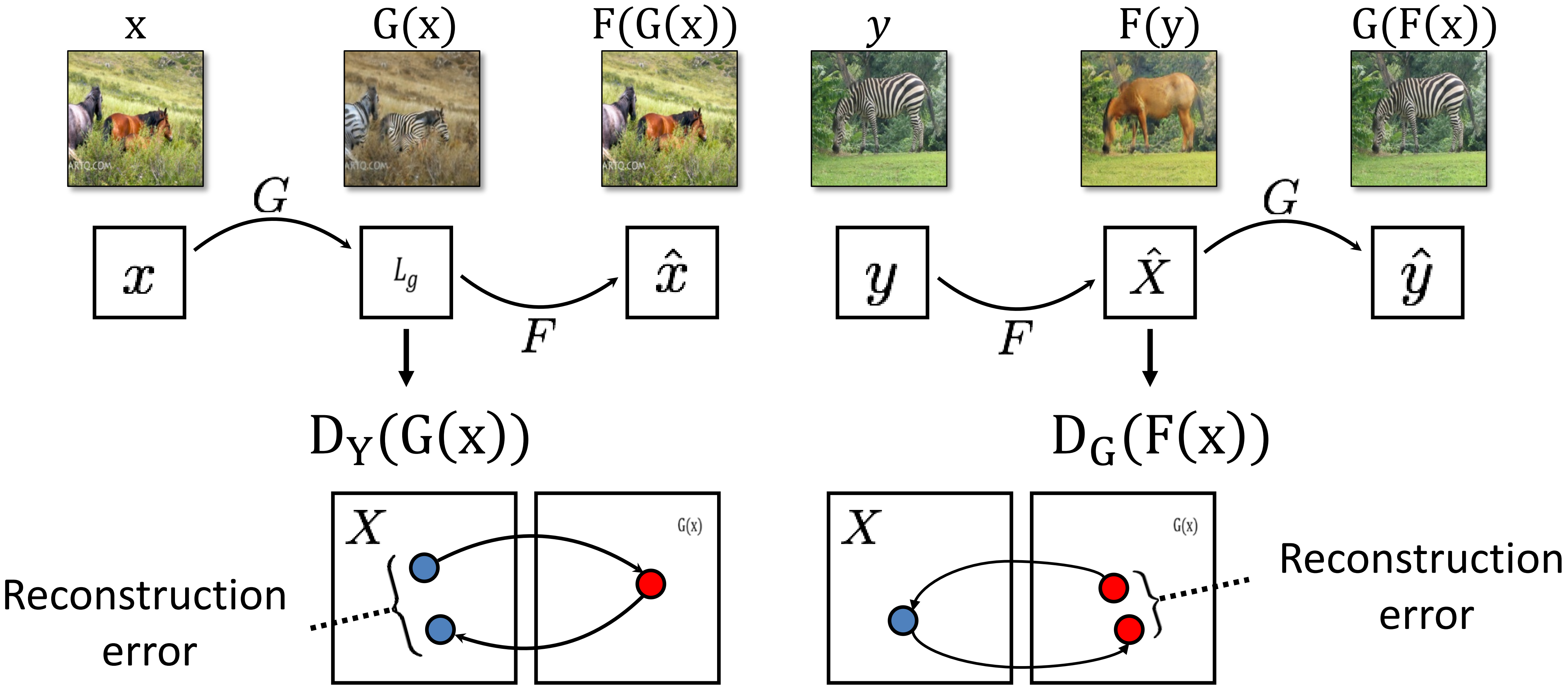
Cycle Consistency Loss





[Zhu*, Park*, Isola, and Efros, ICCV 2017]

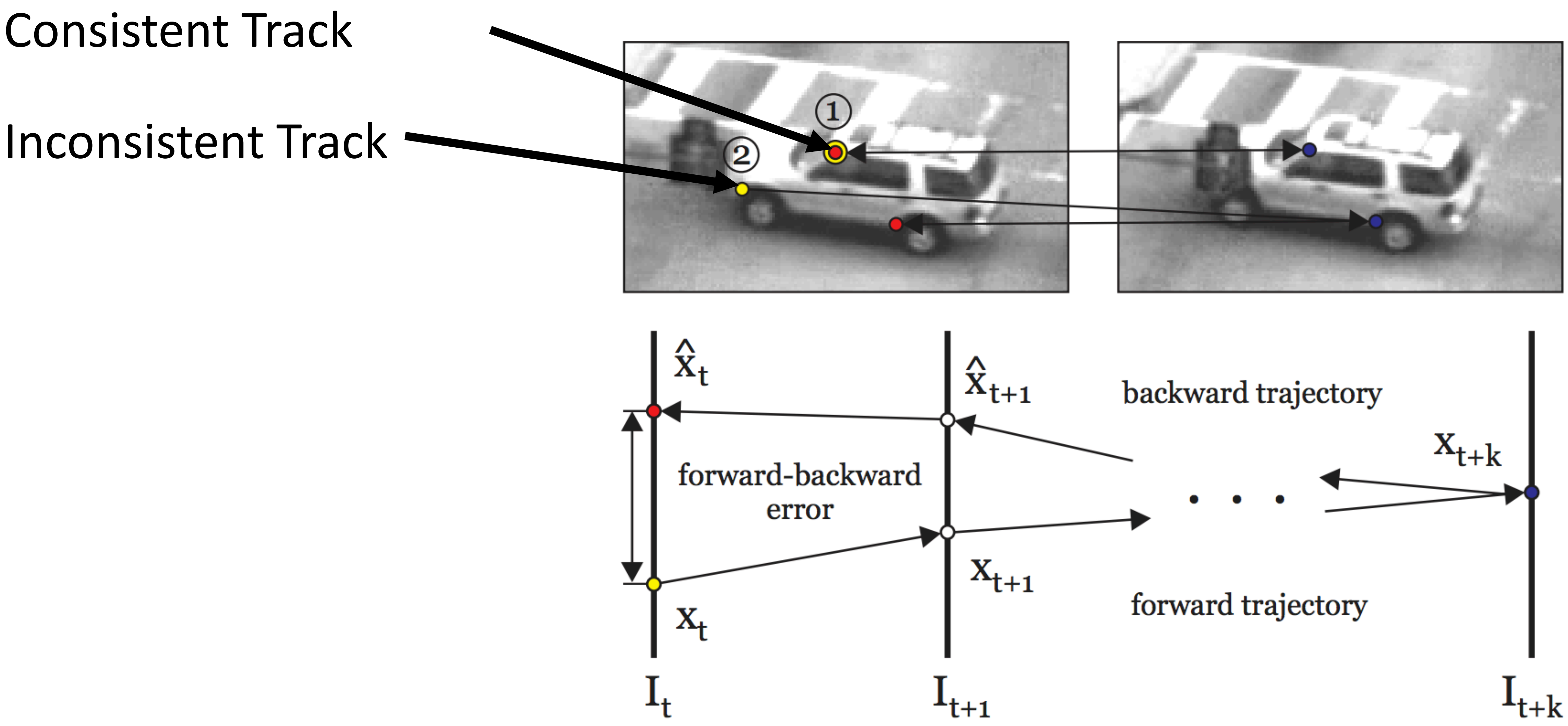
Cycle Consistency Loss



$$\|F(G(x)) - x\|_1 \qquad \|G(F(y)) - y\|_1$$

See similar formulations [Yi et al. 2017], [Kim et al. 2017] [Zhu*, Park*, Isola, and Efros, ICCV 2017]

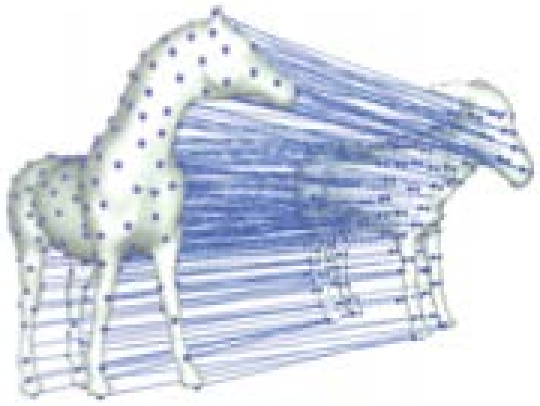
Cycle Consistency in Vision



Forward-Backward Error: Automatic Detection of Tracking Failures. ICPR 10’
Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas.
Also see [Sundaram, Brox, Keutzer, ECCV 10’]

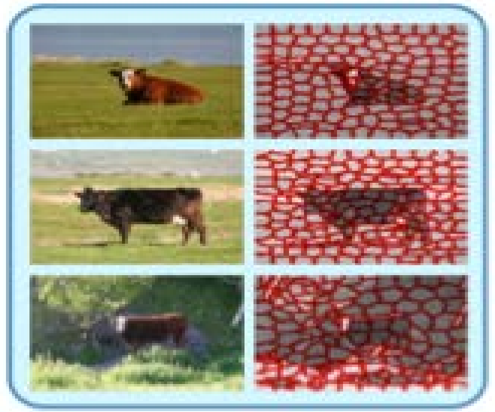
Cycle Consistency in Vision

Shape Matching



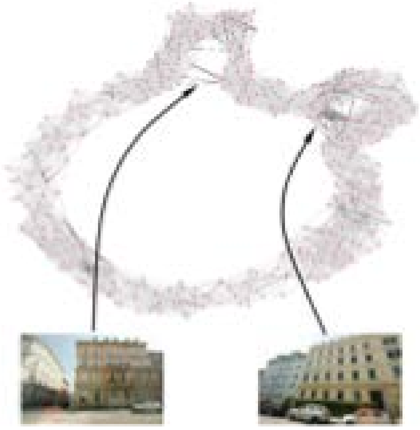
Huang *et al*, SGP'13

Co-segmentation



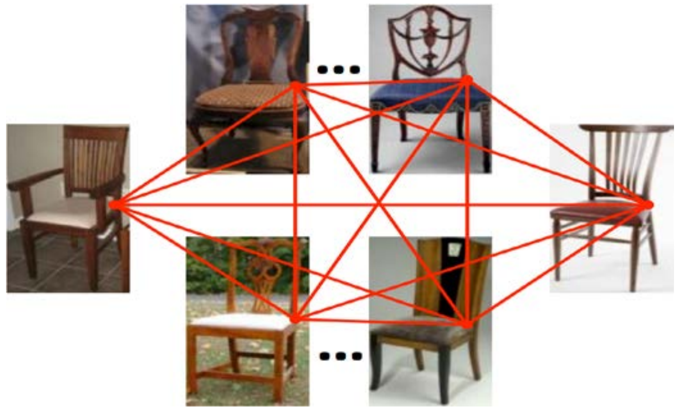
Wang *et al*, ICCV'13

SfM

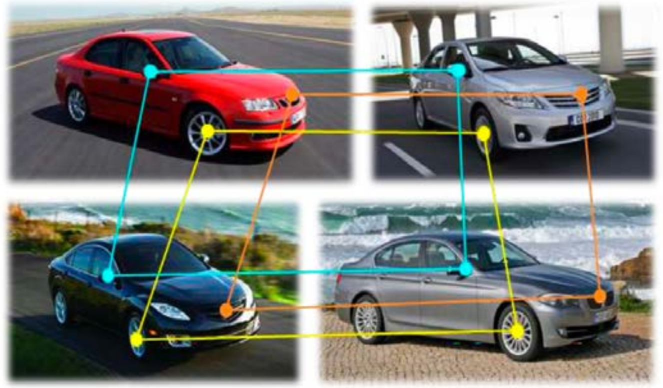


Zach *et al*, CVPR'10

Collection Correspondence



Zhou *et al*, CVPR'15



Zhou *et al*, ICCV'15

slides credit @Tinghui Zhou

Results

Loss	**Map → Photo** % Turkers labeled *real*	**Photo → Map** % Turkers labeled *real*
CoGAN [30]	0.6% ± 0.5%	0.9% ± 0.5%
BiGAN/ALI [8, 6]	2.1% ± 1.0%	1.9% ± 0.9%
SimGAN [45]	0.7% ± 0.5%	2.6% ± 1.1%
Feature loss + GAN	1.2% ± 0.6%	0.3% ± 0.2%
CycleGAN (ours)	**26.8% ± 2.8%**	**23.2% ± 3.4%**

AMT ‘real vs fake’ test on maps ↔ aerial

Loss	**Per-pixel acc.**	**Per-class acc.**	**Class IOU**
CoGAN [30]	0.40	0.10	0.06
BiGAN/ALI [8, 6]	0.19	0.06	0.02
SimGAN [45]	0.20	0.10	0.04
Feature loss + GAN	0.06	0.04	0.01
CycleGAN (ours)	**0.52**	**0.17**	**0.11**

FCN scores on cityscapes labels→ photos

Loss	**Per-pixel acc.**	**Per-class acc.**	**Class IOU**
CoGAN [30]	0.45	0.11	0.08
BiGAN/ALI [8, 6]	0.41	0.13	0.07
SimGAN [45]	0.47	0.11	0.07
Feature loss + GAN	0.50	0.10	0.06
CycleGAN (ours)	**0.58**	**0.22**	**0.16**

Classification performance of photo→labels


Collection Style Transfer

Photograph
@ Alexei Efros

Monet

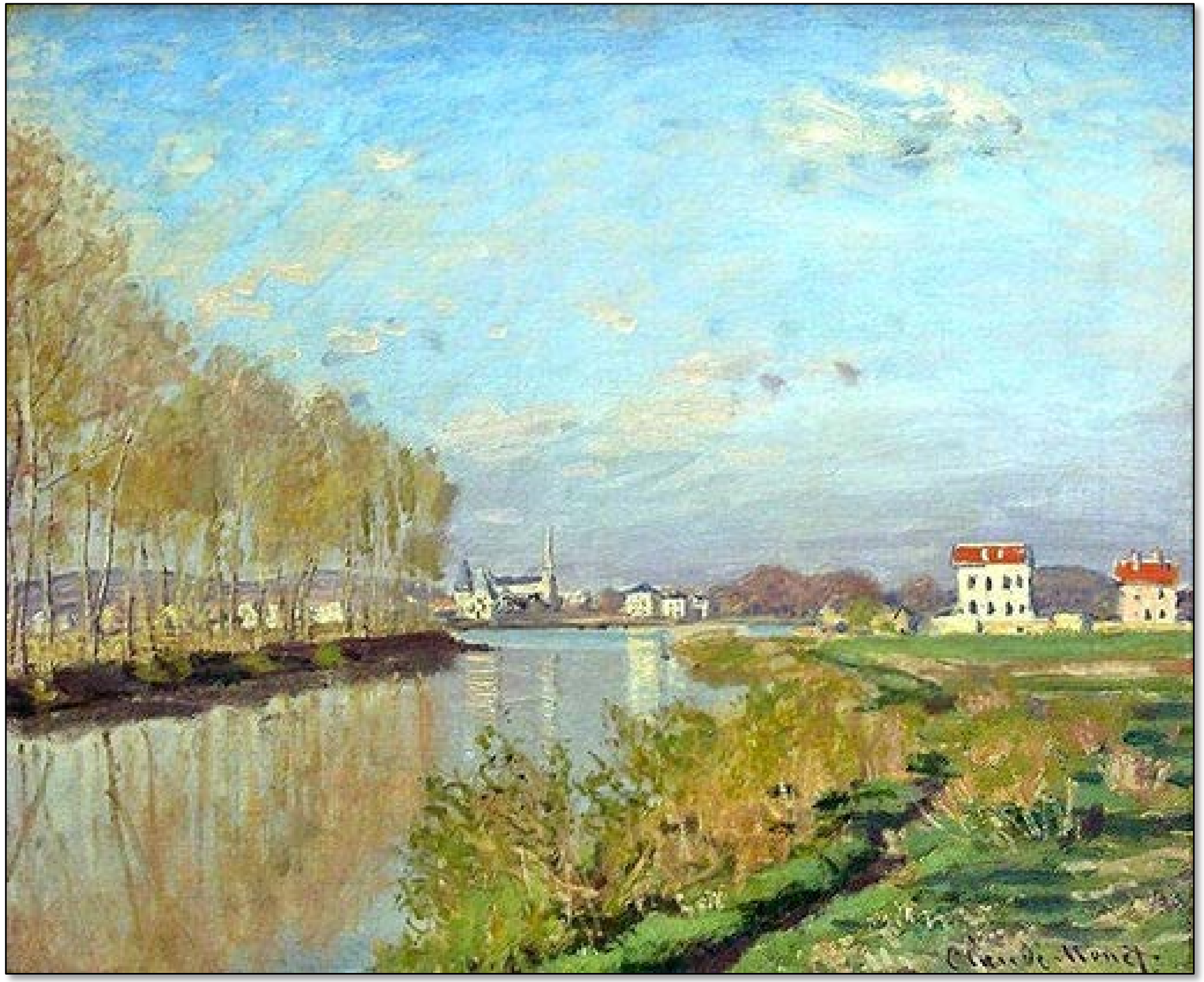
Van Gogh

Cezanne

Ukiyo-e


Input
Monet
Van Gogh
Cezanne
Ukiyo-e

Monet’s paintings → photos



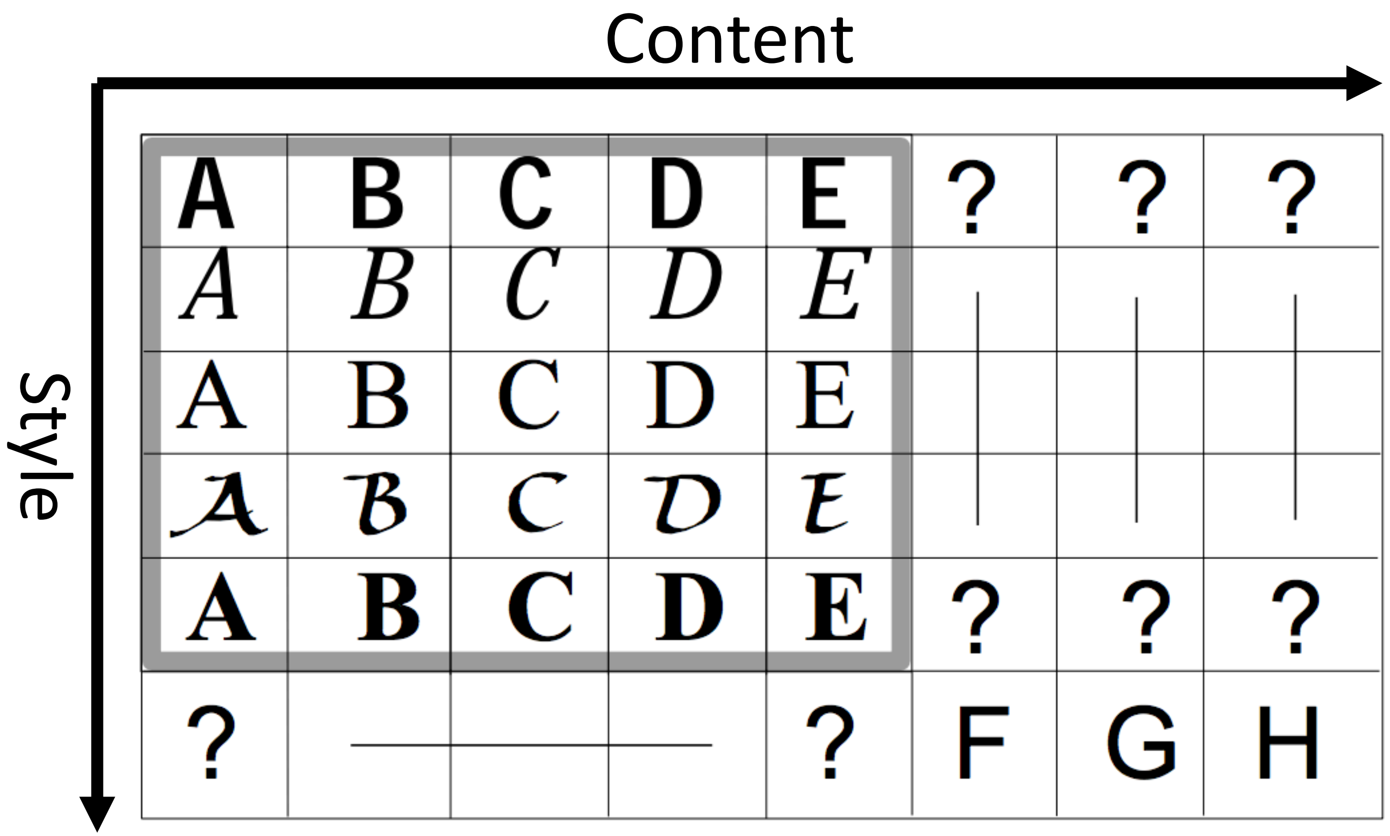
Monet's paintings → photos




Why CycleGAN works

Style and Content Separation

Paired Separation



Separating Style and Content with Bilinear Models
[Tenenbaum and Freeman 2000']

Unpaired Separation

Adversarial Loss: change the style

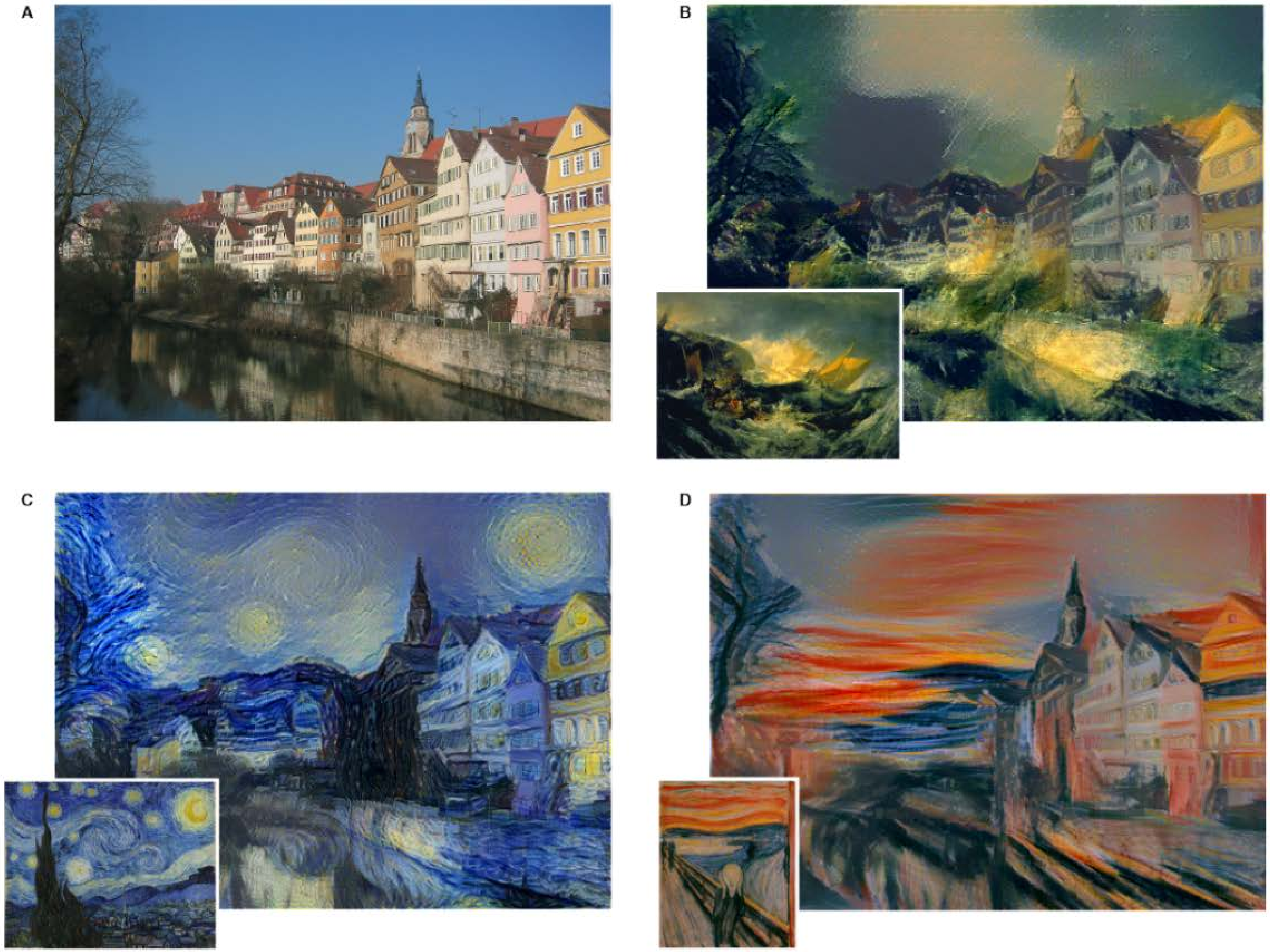
$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = & \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] \\ & + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x)))]\end{aligned}$$

Cycle Consistency Loss: preserve the content

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G, F) = & \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] \\ & + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].\end{aligned}$$

Two empirical assumptions:
- content is easy to keep.
- style is easy to change.

Neural Style Transfer [Gatys et al. 2015]



Style and Content:
- Content: feature difference
- Style: Gram Matrix difference
- Both losses are hard-coded.

PRISMA

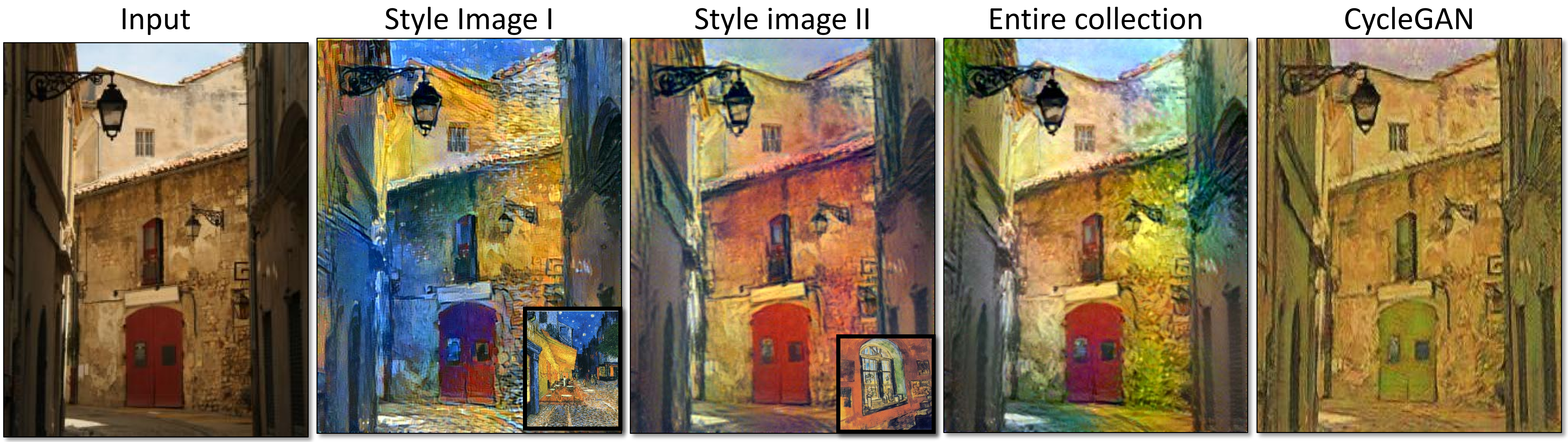
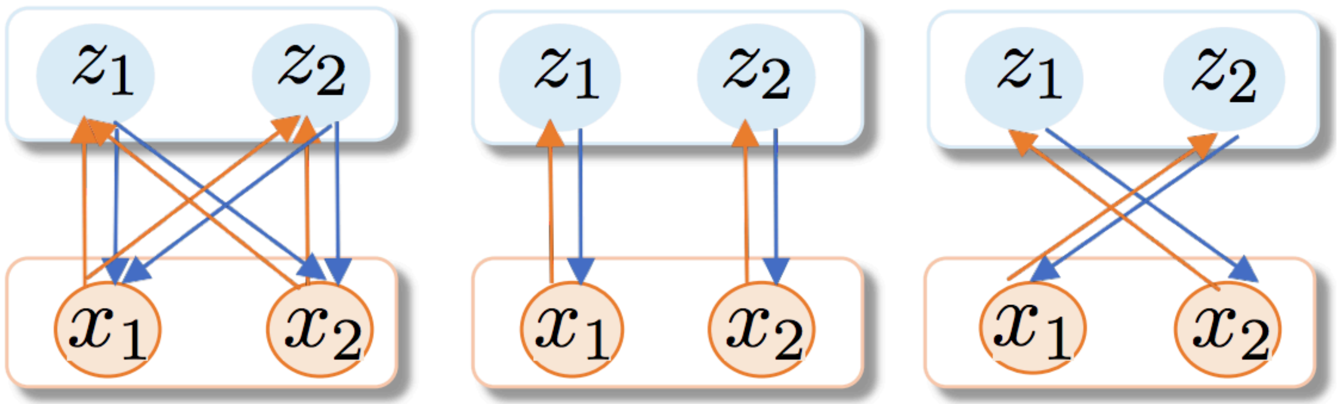


Photo → Van Gogh

horse → zebra

Cycle Loss **upper bounds** Conditional Entropy



	z_1	z_2
x_1	$\delta/2$	$(1-\delta)/2$
x_2	$(1-\delta)/2$	$\delta/2$

	z_1	z_2
x_1	1/2	0
x_2	0	1/2

	z_1	z_2
x_1	0	1/2
x_2	1/2	0

High Conditional Entropy

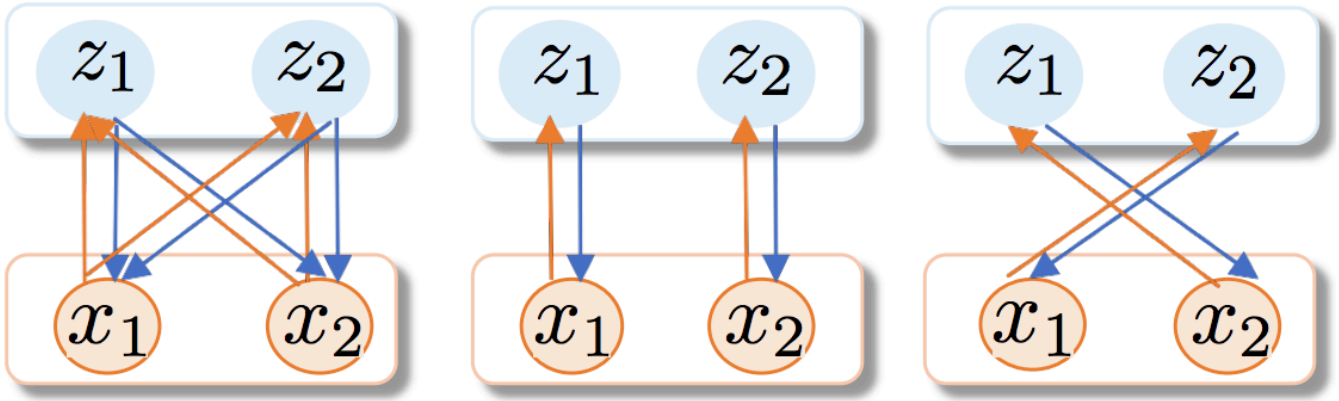
Low Conditional Entropy

Conditional Entropy

$$H^{\pi}(\boldsymbol{x}|\boldsymbol{z}) \triangleq -\mathbb{E}_{\pi(\boldsymbol{x},\boldsymbol{z})}[\log \pi(\boldsymbol{x}|\boldsymbol{z})]$$

“ALICE: Towards Understanding Adversarial Learning for Joint Distribution Matching” [Li et al. NIPS 2017]. Also see [Tiao et al. 2018] “CycleGAN as Approximate Bayesian Inference”

Cycle Loss **upper bounds** Conditional Entropy



	z_1	z_2
x_1	$\delta/2$	$(1-\delta)/2$
x_2	$(1-\delta)/2$	$\delta/2$

	z_1	z_2
x_1	$1/2$	0
x_2	0	$1/2$

	z_1	z_2
x_1	0	$1/2$
x_2	$1/2$	0

Conditional Entropy

$$H^{\pi}(\boldsymbol{x}|\boldsymbol{z}) \triangleq -\mathbb{E}_{\pi(\boldsymbol{x},\boldsymbol{z})}[\log \pi(\boldsymbol{x}|\boldsymbol{z})]$$

Lemma 3 *For joint distributions $p_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z})$ or $q_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{z})$, we have*

$$\begin{aligned}\boxed{H^{q_{\boldsymbol{\phi}}}(\boldsymbol{x}|\boldsymbol{z})} &\triangleq -\mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{z})}[\log q_{\boldsymbol{\phi}}(\boldsymbol{x}|\boldsymbol{z})] = -\mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{z})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})] - \mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{z})}[\mathrm{KL}(q_{\boldsymbol{\phi}}(\boldsymbol{x}|\boldsymbol{z}) \| p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z}))] \\ &\boxed{\leq} -\mathbb{E}_{q_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{z})}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}|\boldsymbol{z})] \triangleq \boxed{\mathcal{L}_{\text{Cycle}}(\boldsymbol{\theta}, \boldsymbol{\phi}).}\end{aligned} \tag{6}$$

“ALICE: Towards Understanding Adversarial Learning for Joint Distribution Matching” [Li et al. NIPS 2017]. Also see [Tiao et al. 2018] “CycleGAN as Approximate Bayesian Inference”

CycleGAN implementations

Torch

pytorch-CycleGAN-and-pix2pix

Image-to-image translation in PyTorch (e.g., horse2zebra, edges2cats, and more)

Python ★ 4.3k 970

PyTorch

CycleGAN

Software that can generate photos from paintings, turn horses into zebras, perform style transfer, and more.

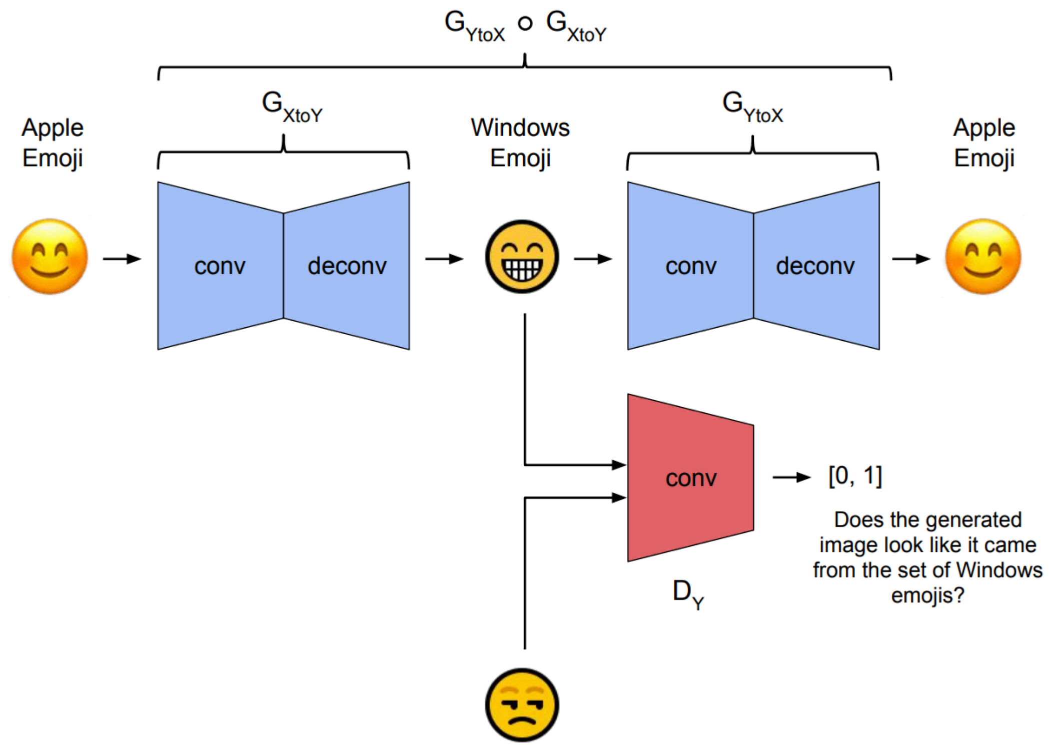
Lua ★ 6.5k 940

20+ implementations by researchers/developers:

- Tensorflow, Chainer, mxnet, Lasagne, Keras…

CycleGAN at School

- Taught at Stanford, UC Berkeley, UoT, Udacity, FastAI, etc.
- Course assignment code and handout designed by Prof. Roger Grosse for CSC321 “Intro to Neural Networks and Machine Learning” at **University of Toronto**.




```
##   FILL THIS IN: CREATE ARCHITECTURE   ##
###########################################

# 1. Define the encoder part of the generator
# self.conv1 = ...
# self.conv2 = ...

# 2. Define the transformation part of the generator
# self.resnet_block = ...

# 3. Define the decoder part of the generator
# self.deconv1 = ...
# self.deconv2 = ...
```

Applications

CG2Real: GTA5 → real streetview

GTA5 CG Input

Output

Inspired by [Johnson et al. 2011]

Real2CG: real streetview → GTA

Cityscape Input

Output

Synthetic Data as Supervision

GTA5 images | Segmentation labels

[Richter*, Vineet* et al. 2016] [Krähenbühl et al. 2018]

Domain Adaptation with CycleGAN

Train on GTA5 data

Test on real images

	meanIOU	Per-pixel accuracy
Oracle (Train and test on Real)	60.3	93.1
Train on CG, test on Real	17.9	54.0

See Judy Hoffman’s talk at 14:30 “Adversarial Domain Adaptation”

Domain Adaptation with CycleGAN

GTA5 data + Domain adaptation

Test on real images

	meanIOU	Per-pixel accuracy
Oracle (Train and test on Real)	60.3	93.1
Train on CG, test on Real	17.9	54.0
FCN in the wild [Previous STOA]	27.1	-

See Judy Hoffman's talk at 14:30 "Adversarial Domain Adaptation"

Domain Adaptation with CycleGAN

Train on CycleGAN data

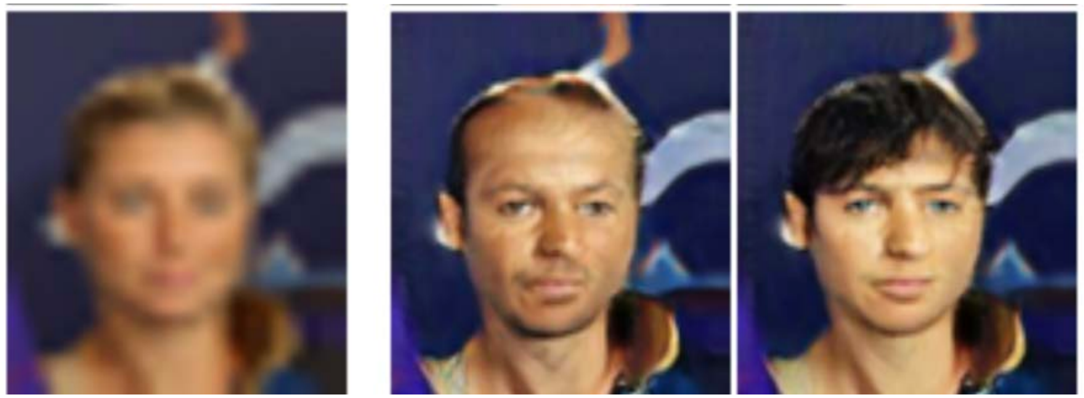
Test on real images

	meanIOU	Per-pixel accuracy
Oracle (Train and test on Real)	60.3	93.1
Train on CG, test on Real	17.9	54.0
FCN in the wild [Previous STOA]	27.1	-
Train on CycleGAN, test on Real	**34.8**	**82.8**

See Judy Hoffman's talk at 14:30 "Adversarial Domain Adaptation"

Applications and Extentions

Attribute Editing [Lu et al.]



Low-res Bald Bangs

arXiv:1705.09966

Object Editing [Liang et al.]

Mask Input Output

arXiv:1708.00315

Front/Character Transfer [Ignatov et al.]

Input output

arXiv: 1801.08624

Data generation [Wang et al.]

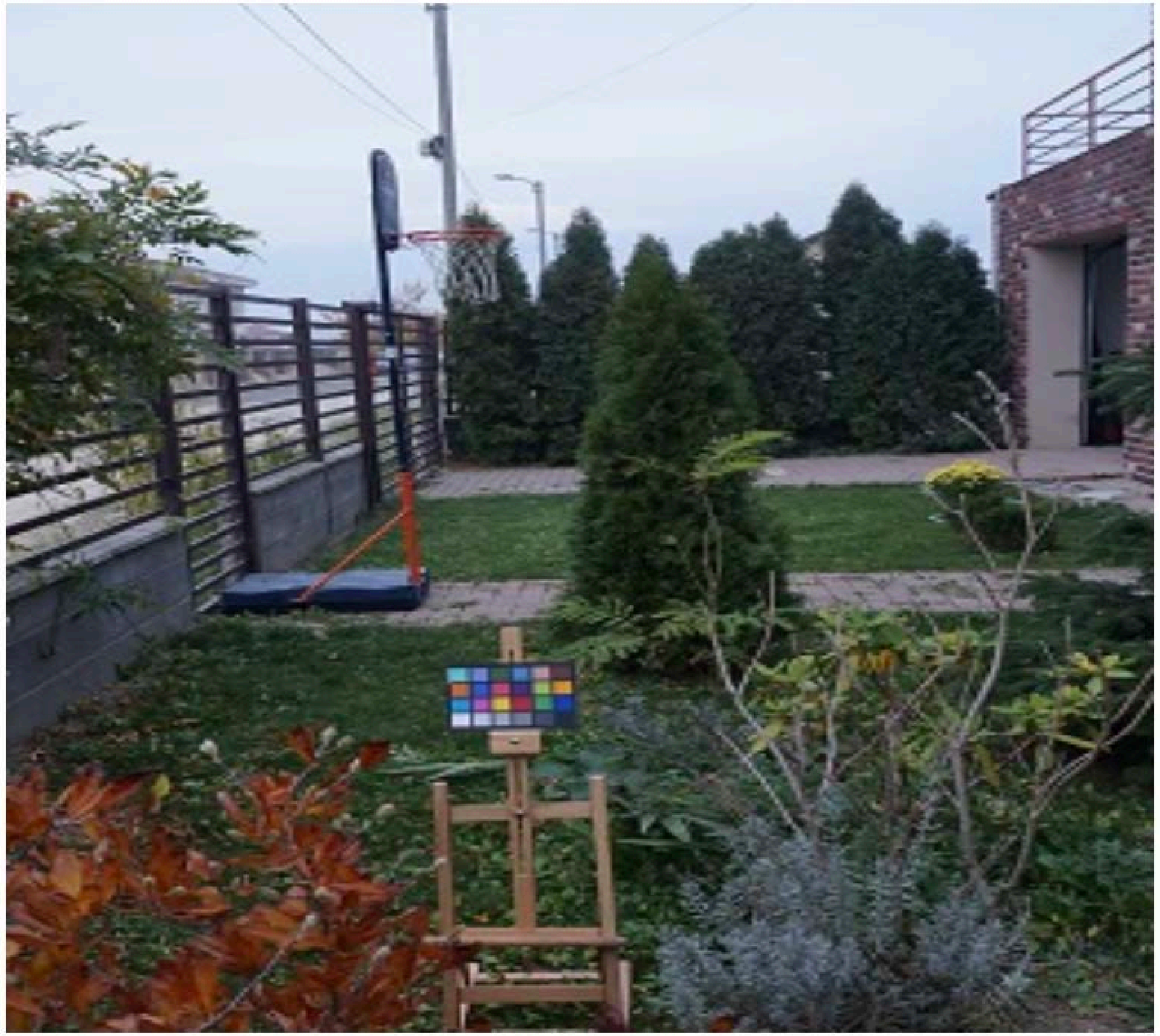
samples by CycleWGAN

arXiv:1707.03124

Photo Enhancement

WESPE: Weakly Supervised Photo Enhancer for Digital Cameras. arxiv 1709.01118
Andrey Ignatov, Nikolay Kobyshev, Kenneth Vanhoey, Radu Timofte, Luc Van Gool

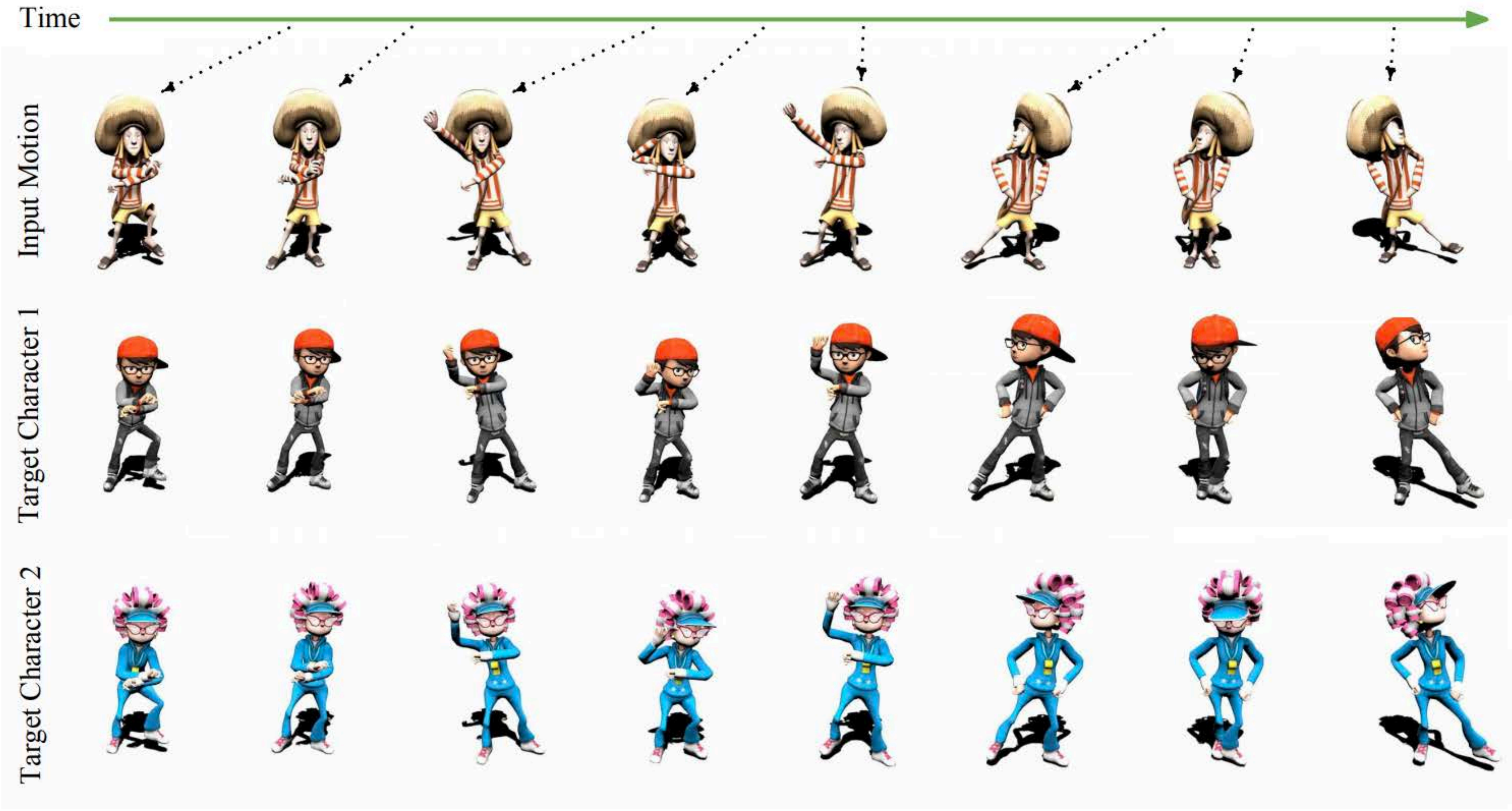
Image Dehazing



Cycle-Dehaze: Enhanced CycleGAN for Single Image Dehazing. CVPRW 2018

Deniz Engin∗ Anıl Genc∗, Hazım Kemal Ekenel

Unsupervised Motion Retargeting



Neural Kinematic Networks for Unsupervised Motion Retargetting. CVPR 2018 (oral)
Ruben Villegas, Jimei Yang, Duygu Ceylan, Honglak Lee

Neural Kinematic Networks for Unsupervised Motion Retargetting. CVPR 2018 (oral)

Ruben Villegas, Jimei Yang, Duygu Ceylan, Honglak Lee

Applications Beyond Computer Vision

- 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.

…

Deep MR to CT Synthesis using Unpaired Data

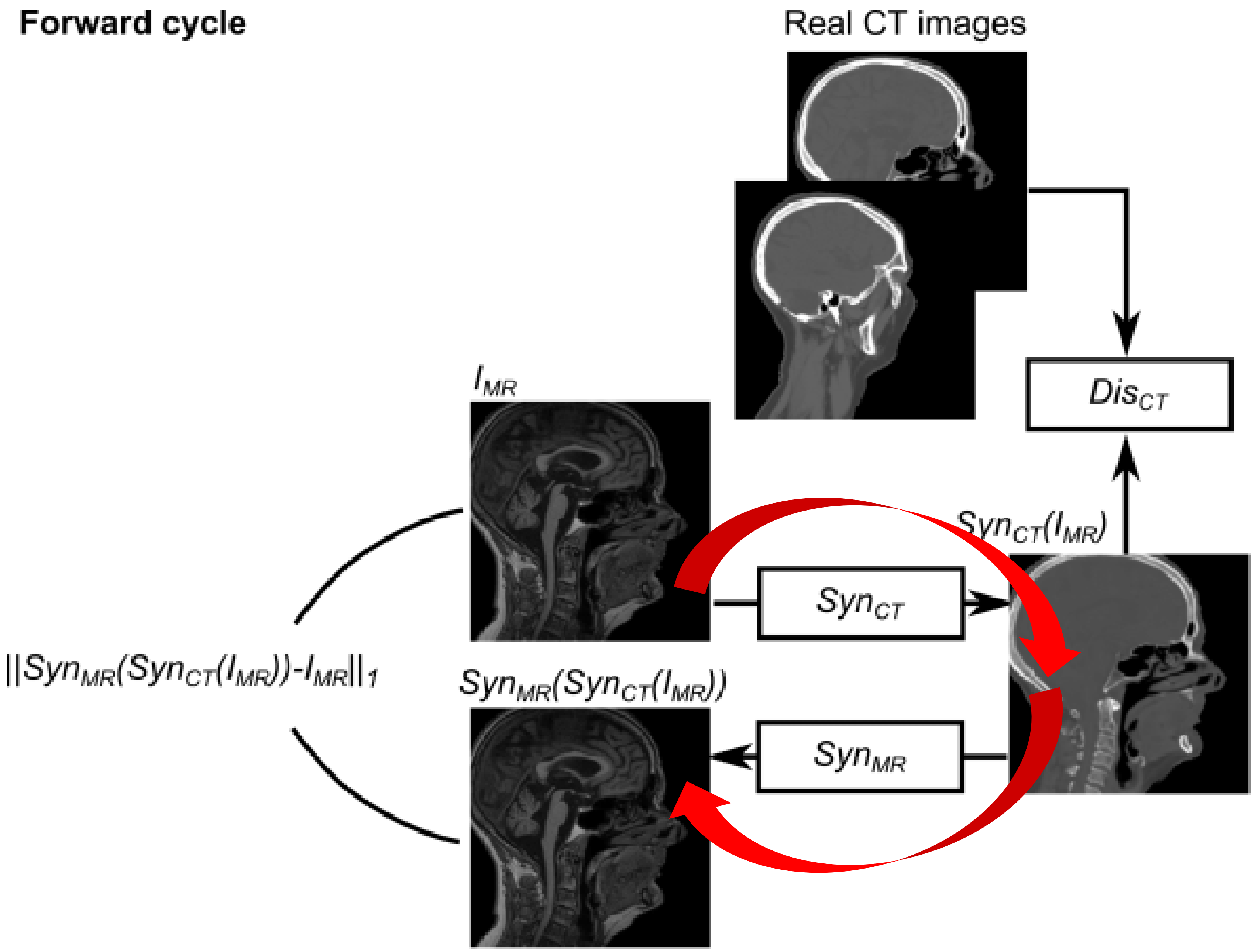
Jelmer M. Wolterink[1]✉, Anna M. Dinkla[2], Mark H.F. Savenije[2],
Peter R. Seevinck[1], Cornelis A.T. van den Berg[2], Ivana Išgum[1]

[1] Image Sciences Institute, University Medical Center Utrecht, The Netherlands
j.m.wolterink@umcutrecht.nl
[2] Department of Radiotherapy, University Medical Center Utrecht, The Netherlands

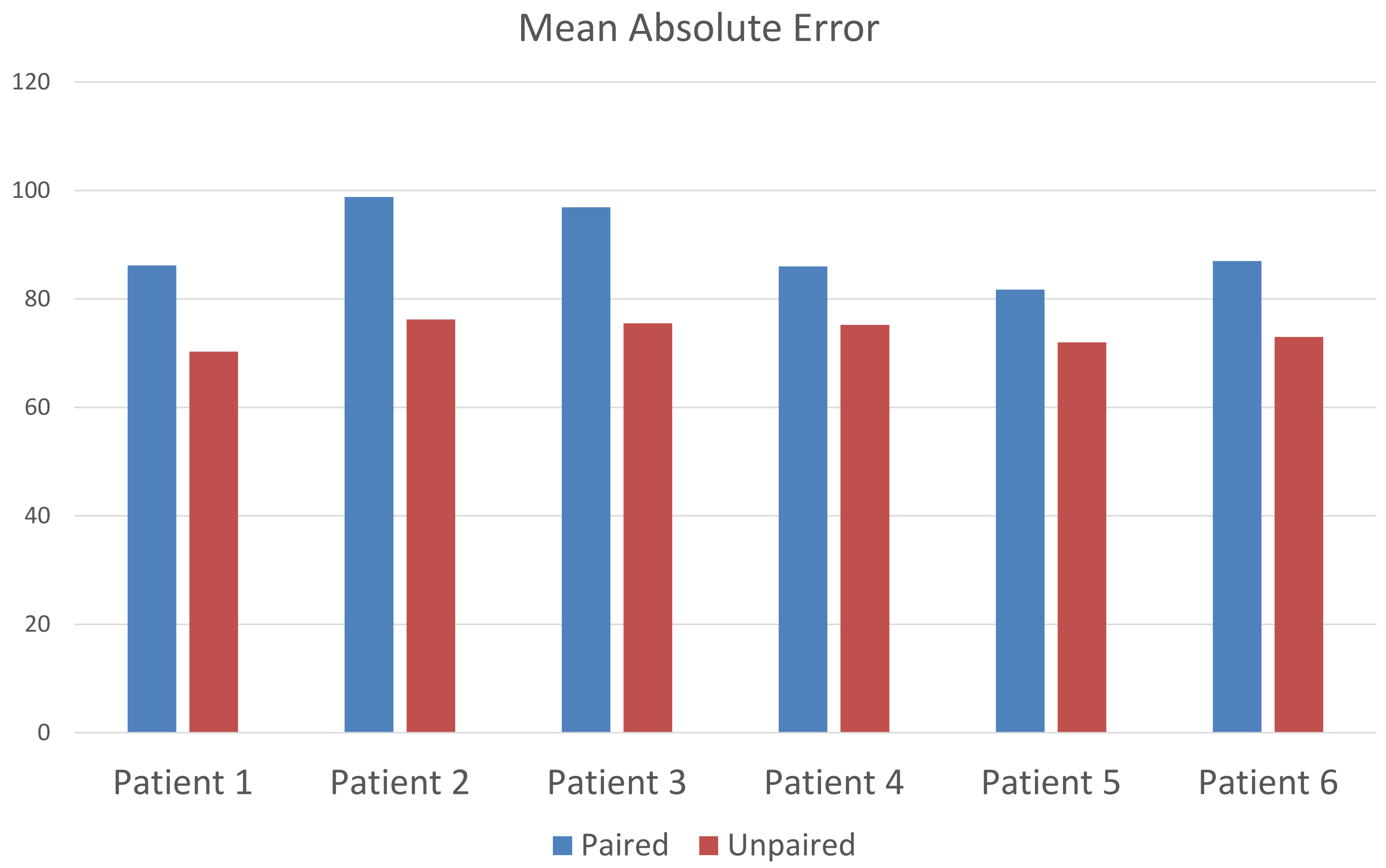
Input MR　　Generated CT　　Ground truth CT

Problem with the paired MR-CT data

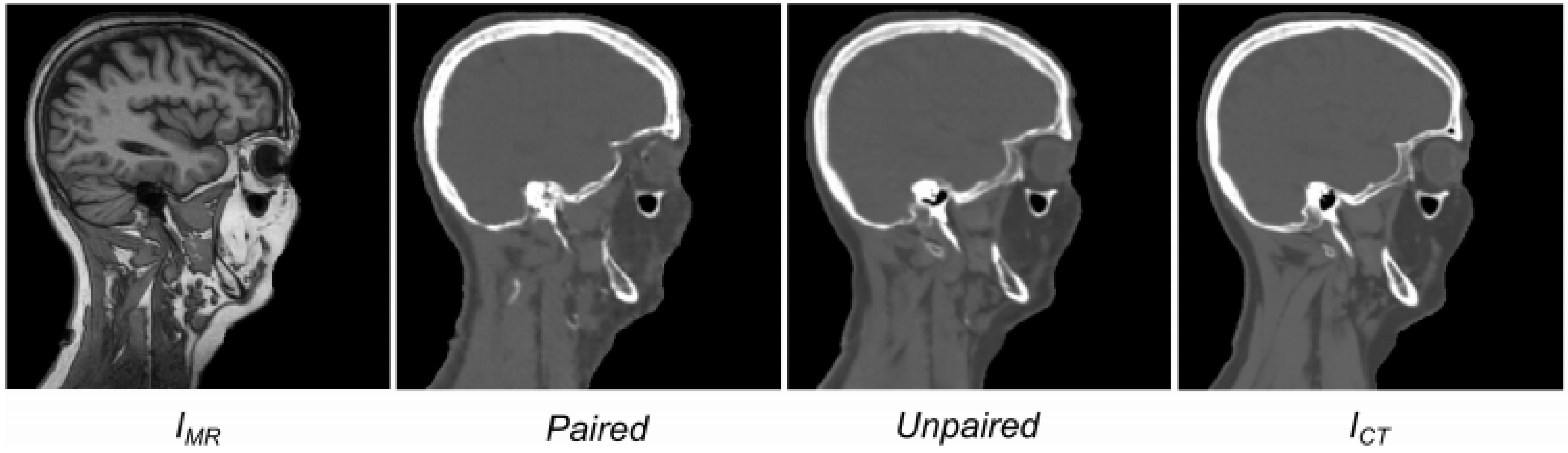
- Images are not perfectly aligned
- There are usually more unpaired data



Forward cycle
Real CT images
Dis_{CT}
I_{MR}
$Syn_{CT}(I_{MR})$
Syn_{CT}
Syn_{MR}
$Syn_{MR}(Syn_{CT}(I_{MR}))$
$||Syn_{MR}(Syn_{CT}(I_{MR}))-I_{MR}||_1$



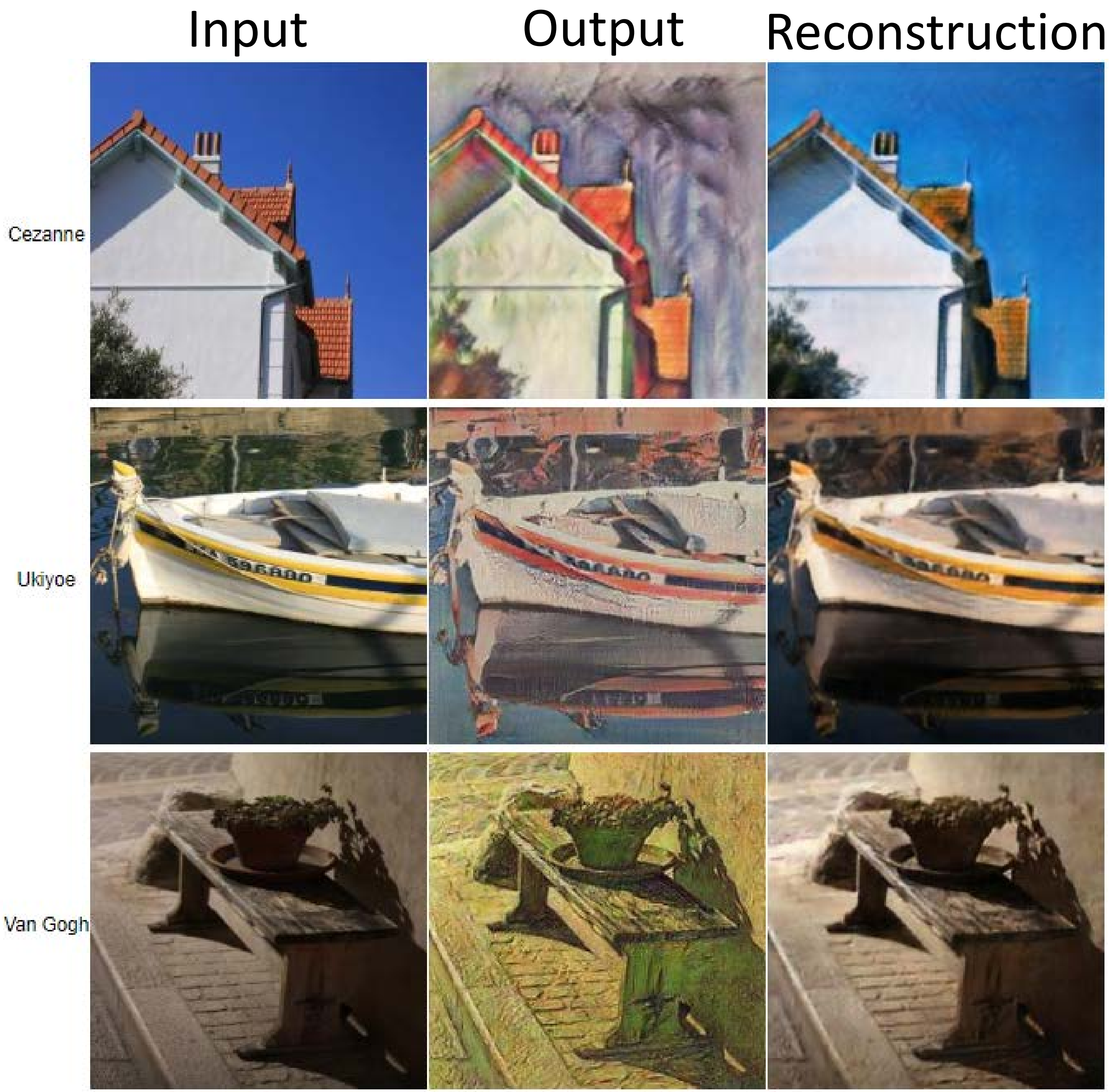
Mean Absolute Error
120
100
80
60
40
20
0
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Paired
Unpaired



I_{MR}
Paired
Unpaired
I_{CT}

Reconstucted Images

Reconstructed Images

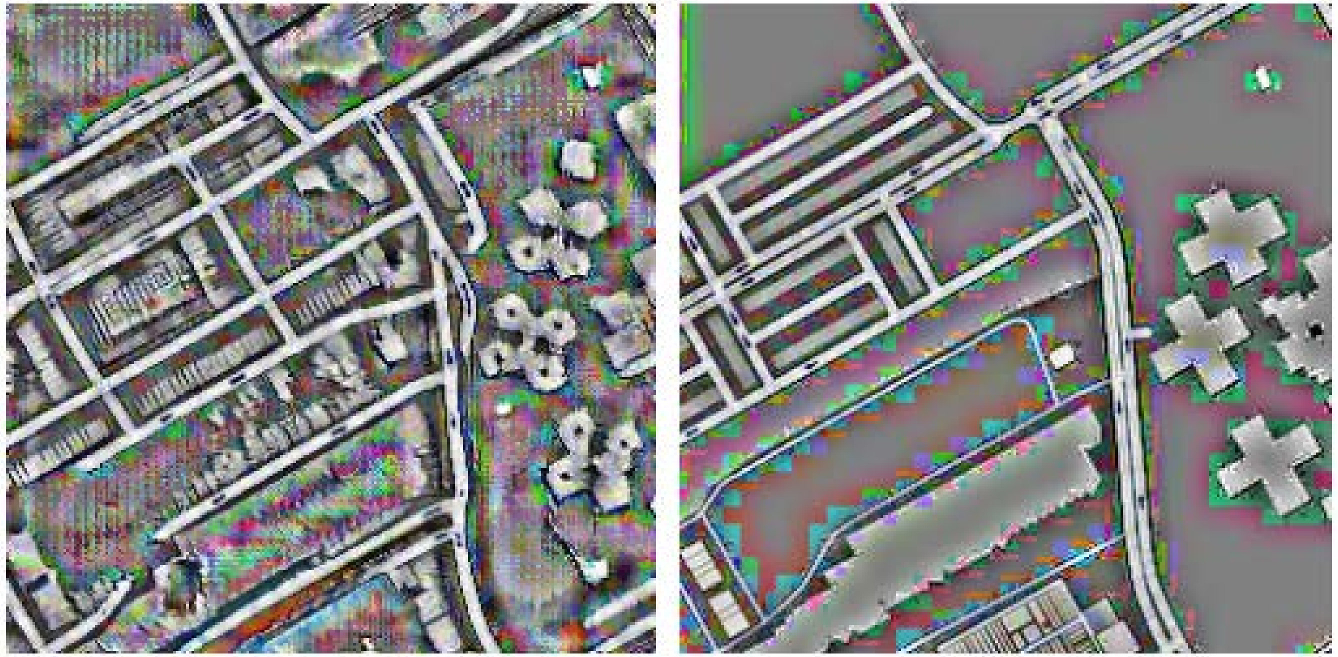


Reconstructed Images (Chu, Zhmoginov and Sandler, NIPSW 2017)

Input Output Reconstruction

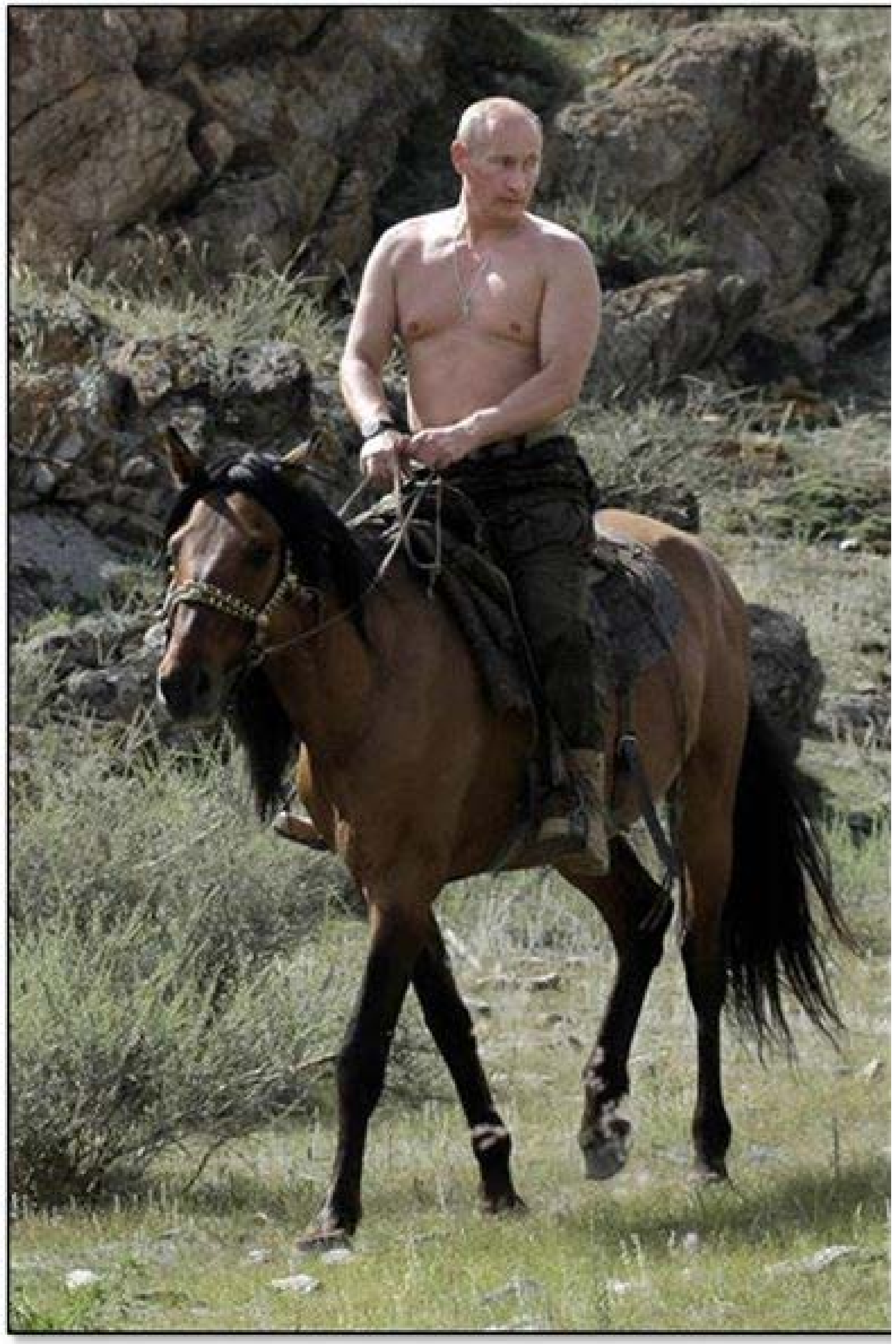
Reconstructed Images (Chu, Zhmoginov and Sandler, NIPSW 2017)

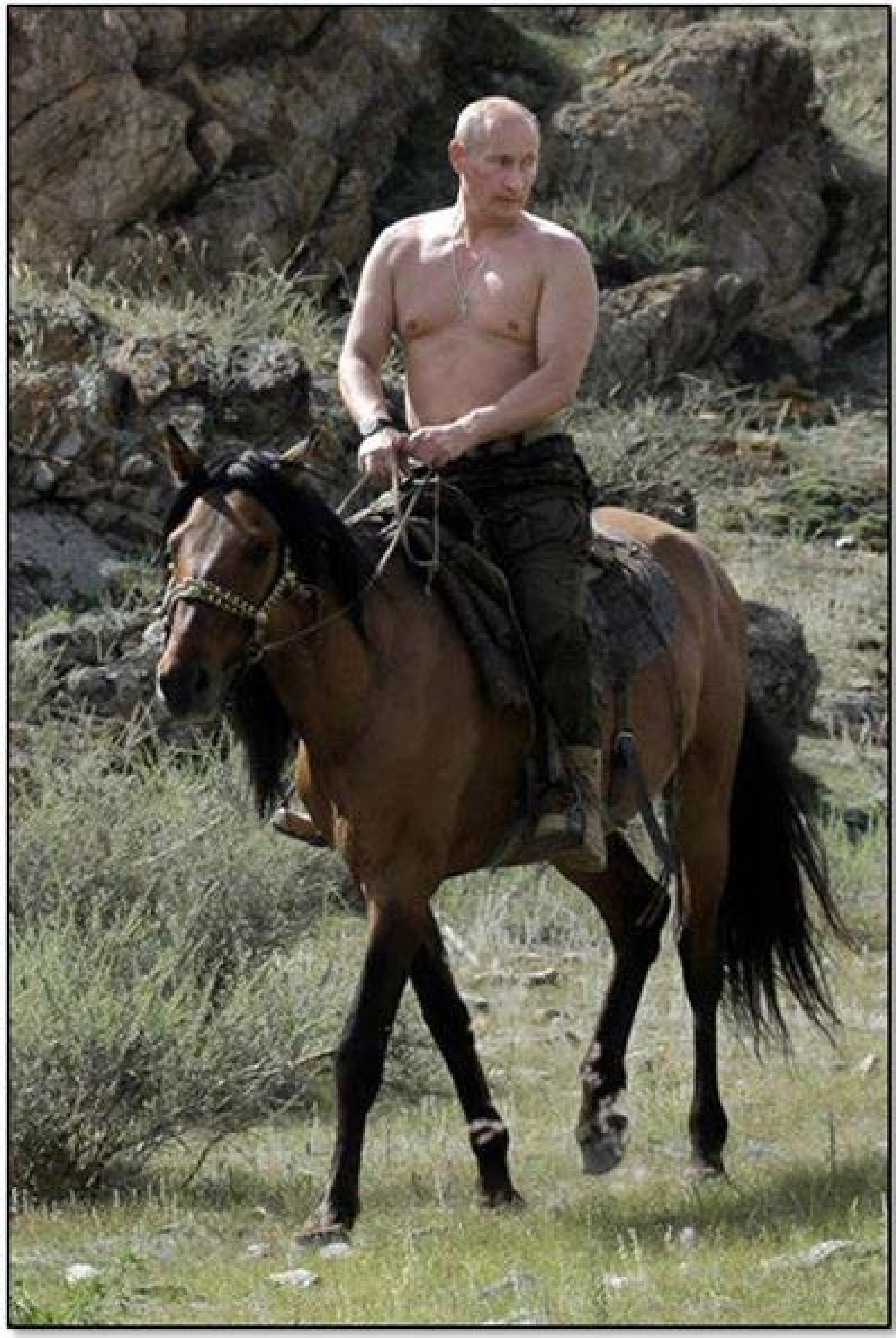
Input Output Reconstruction

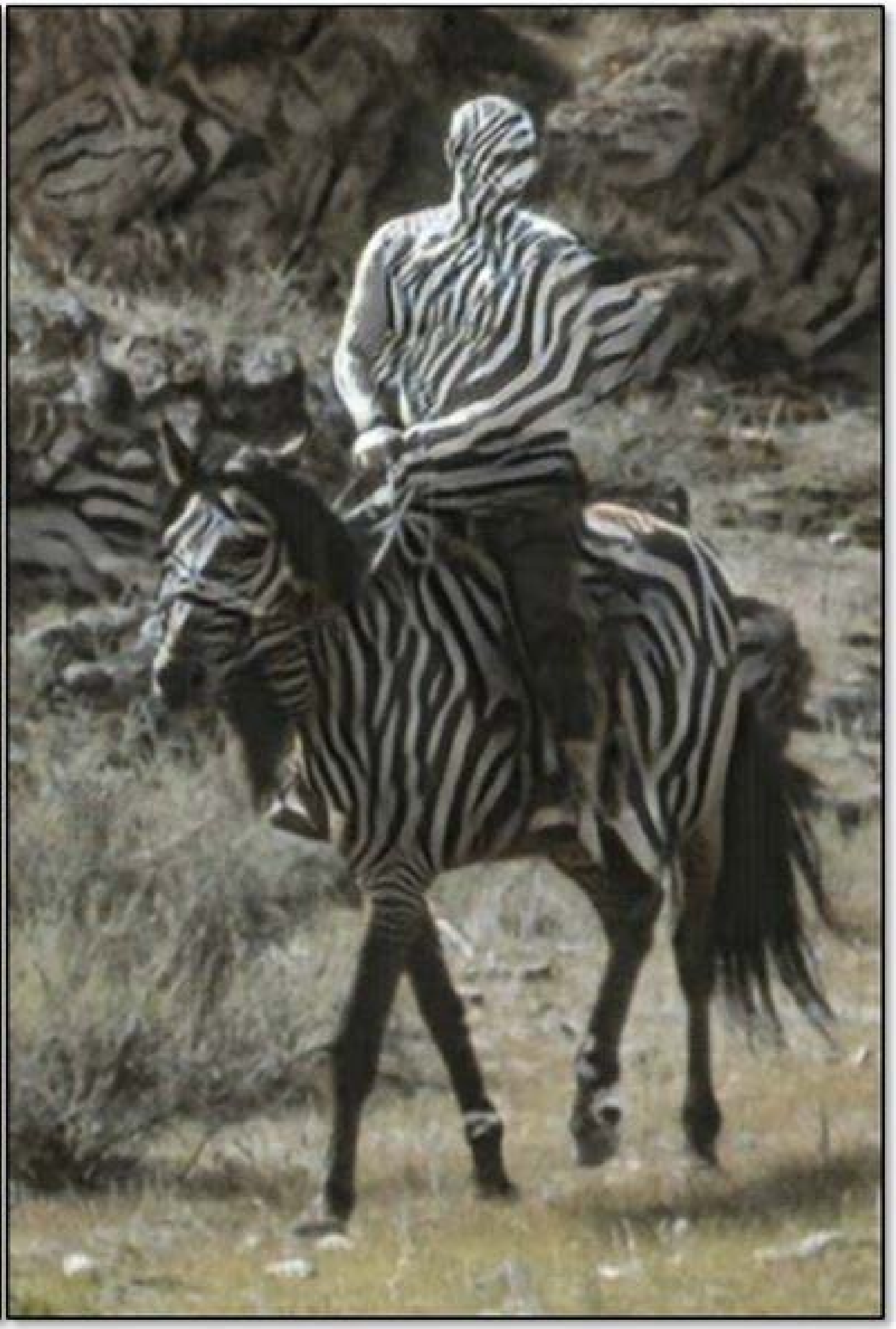


Output Variance Real Map Variance

Failure cases







ImageNet
“Wild horse”


Courtesy of the New York Apple Commission


Beyond CycleGAN

Multi-modality

Style control

More than two domains

Beyond CycleGAN

Multi-modality

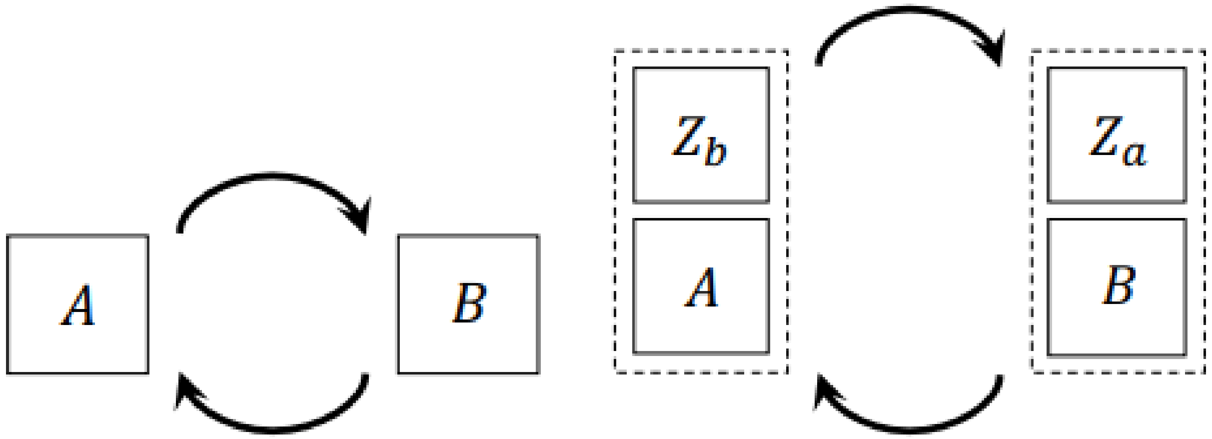
Style control

More than two domains

Beyond CycleGAN – Multi-modality

Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data

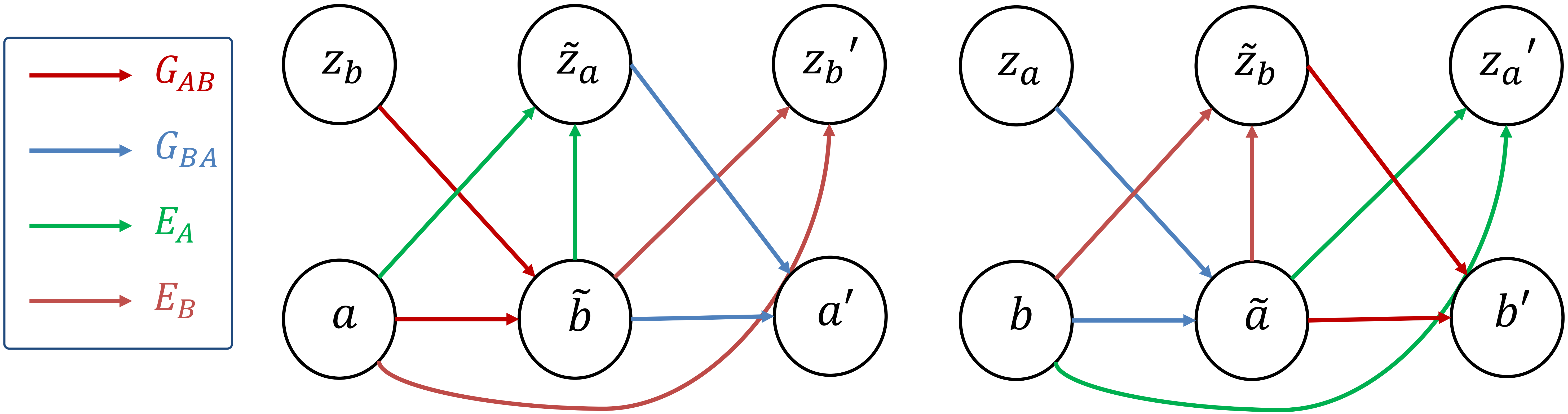
Amjad Almahairi [1 †] **Sai Rajeswar** [1] **Alessandro Sordoni** [2] **Philip Bachman** [2] **Aaron Courville** [1 3]



(a) CycleGAN (b) Augmented CycleGAN

Almahairi et al., ICML 2018

Beyond CycleGAN – Multi-modality



Almahairi et al., ICML 2018

Beyond CycleGAN – Multi-modality

Almahairi et al., ICML 2018

Beyond CycleGAN

Multi-modality

Style control

More than two domains

Beyond CycleGAN – Style Control

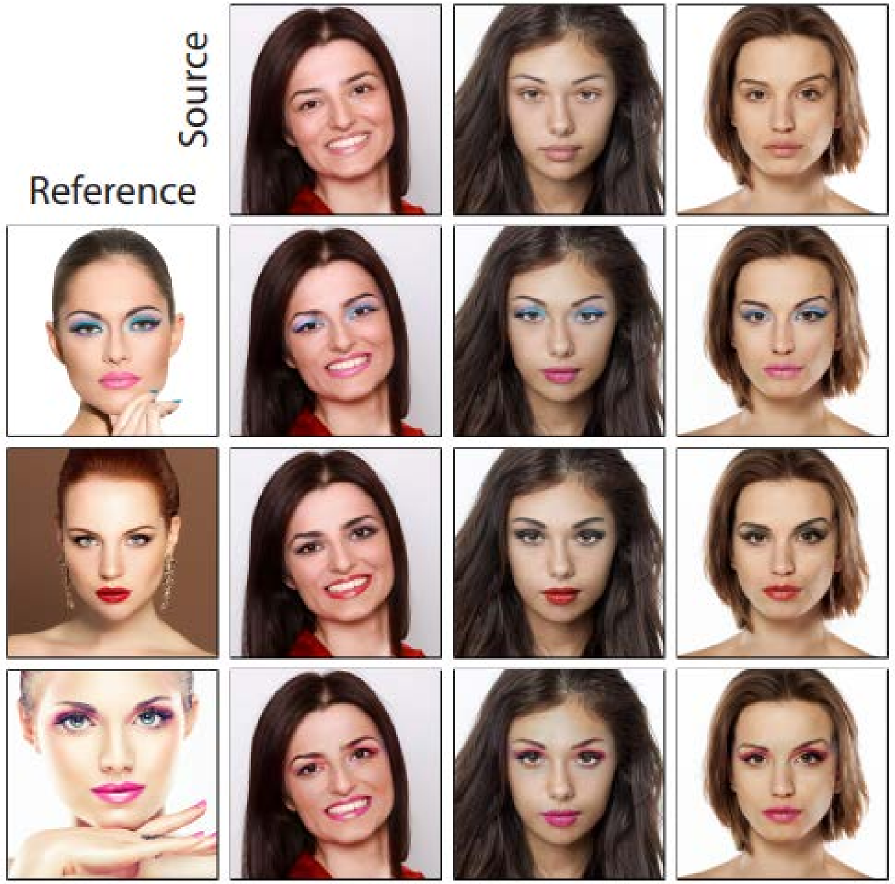
PairedCycleGAN: Asymmetric Style Transfer for Applying and Removing Makeup

Huiwen Chang
Princeton University

Jingwan Lu
Adobe Research

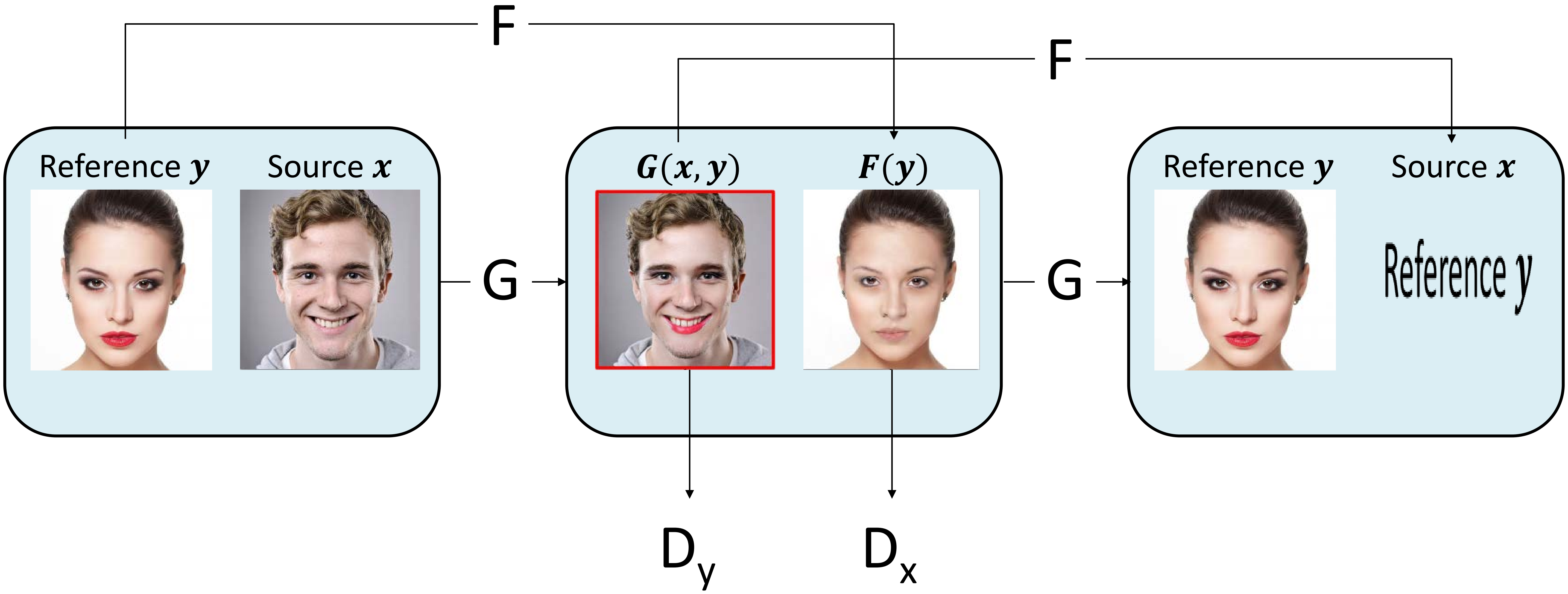
Fisher Yu
UC Berkeley

Adam Finkelstein
Princeton University



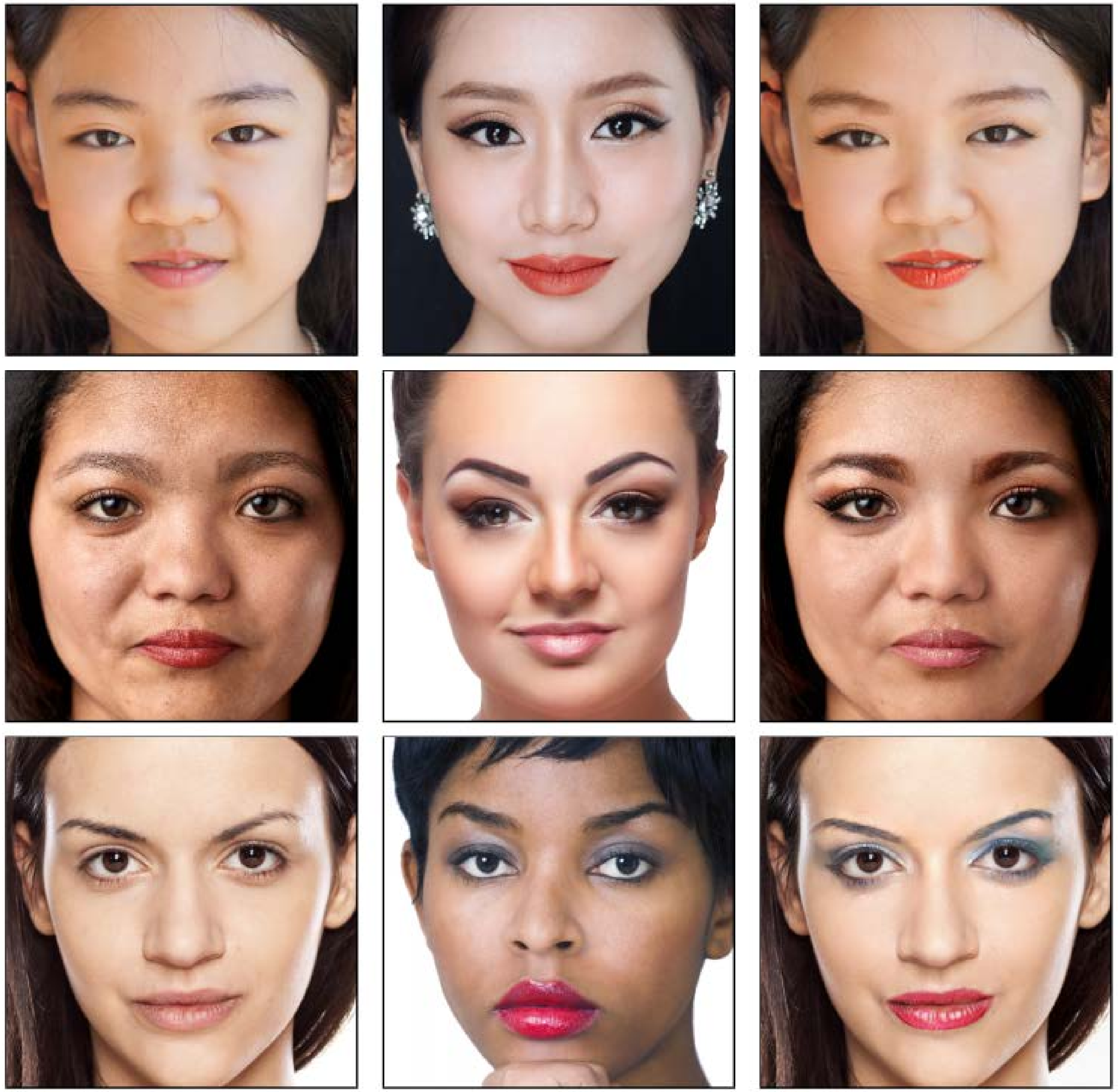
Chang et al., CVPR 2018

Beyond CycleGAN – Style Control



Chang et al., CVPR 2018

Beyond CycleGAN – Style Control



Source Reference Our result

Chang et al., CVPR 2018

Beyond CycleGAN

Multi-modality

Style control

More than two domains

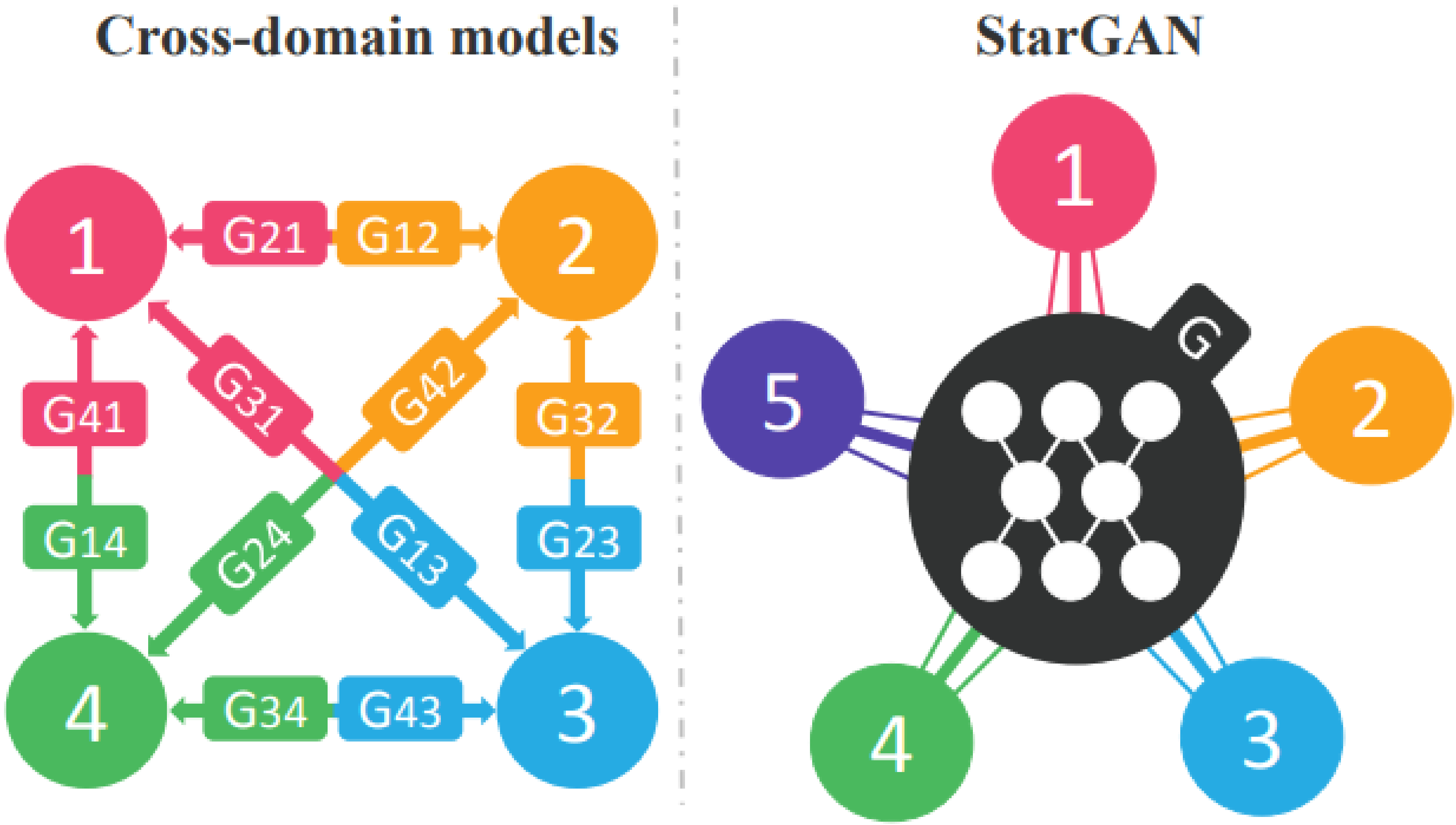
Beyond CycleGAN – More than 2 domains

StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation

Yunjey Choi[1,2] Minje Choi[1,2] Munyoung Kim[2,3] Jung-Woo Ha[2] Sunghun Kim[2,4] Jaegul Choo[1,2]
[1] Korea University [2] Clova AI Research, NAVER
[3] The College of New Jersey [4] Hong Kong University of Science & Technology

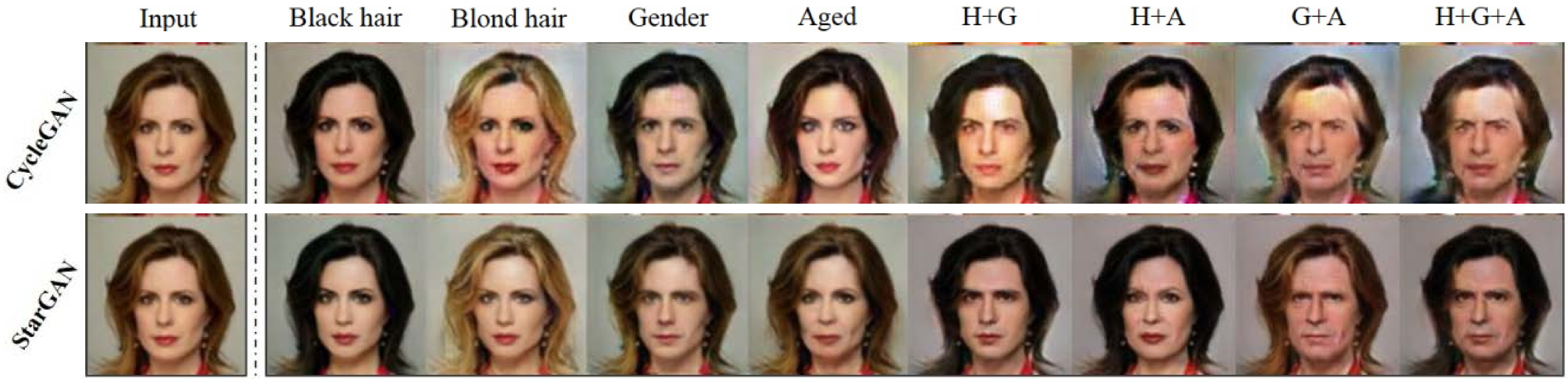
Choi et al., CVPR 2018

Beyond CycleGAN – More than 2 domains



Choi et al., CVPR 2018

Beyond CycleGAN – More than 2 domains



Choi et al., CVPR 2018

Beyond CycleGAN – More than 2 domains

ComboGAN:
Unrestrained Scalability for Image Domain Translation

Asha Anoosheh
Computer Vision Lab
ETH Zürich
ashaa@ethz.ch

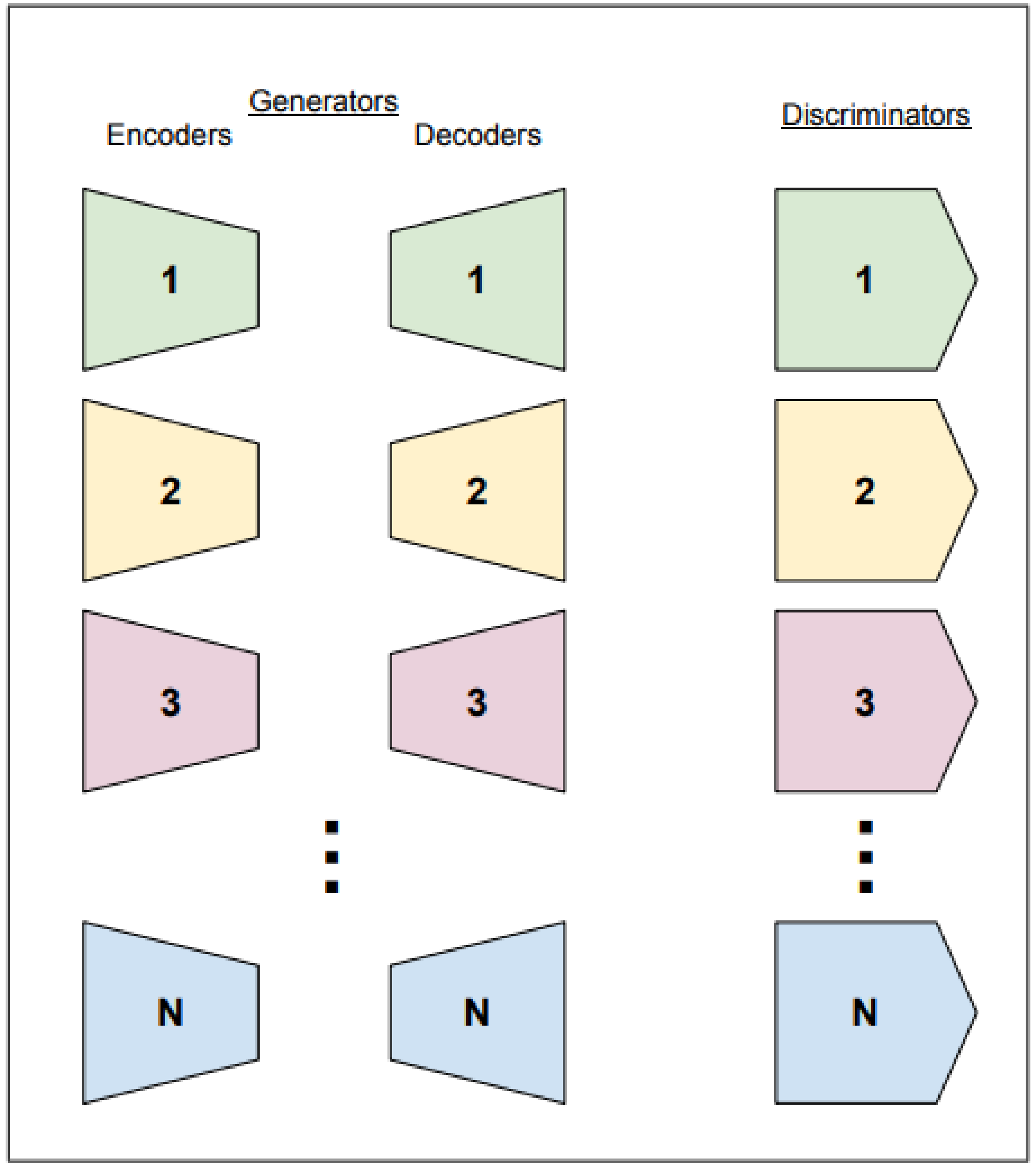
Eirikur Agustsson
Computer Vision Lab
ETH Zürich
aeirikur@ethz.ch

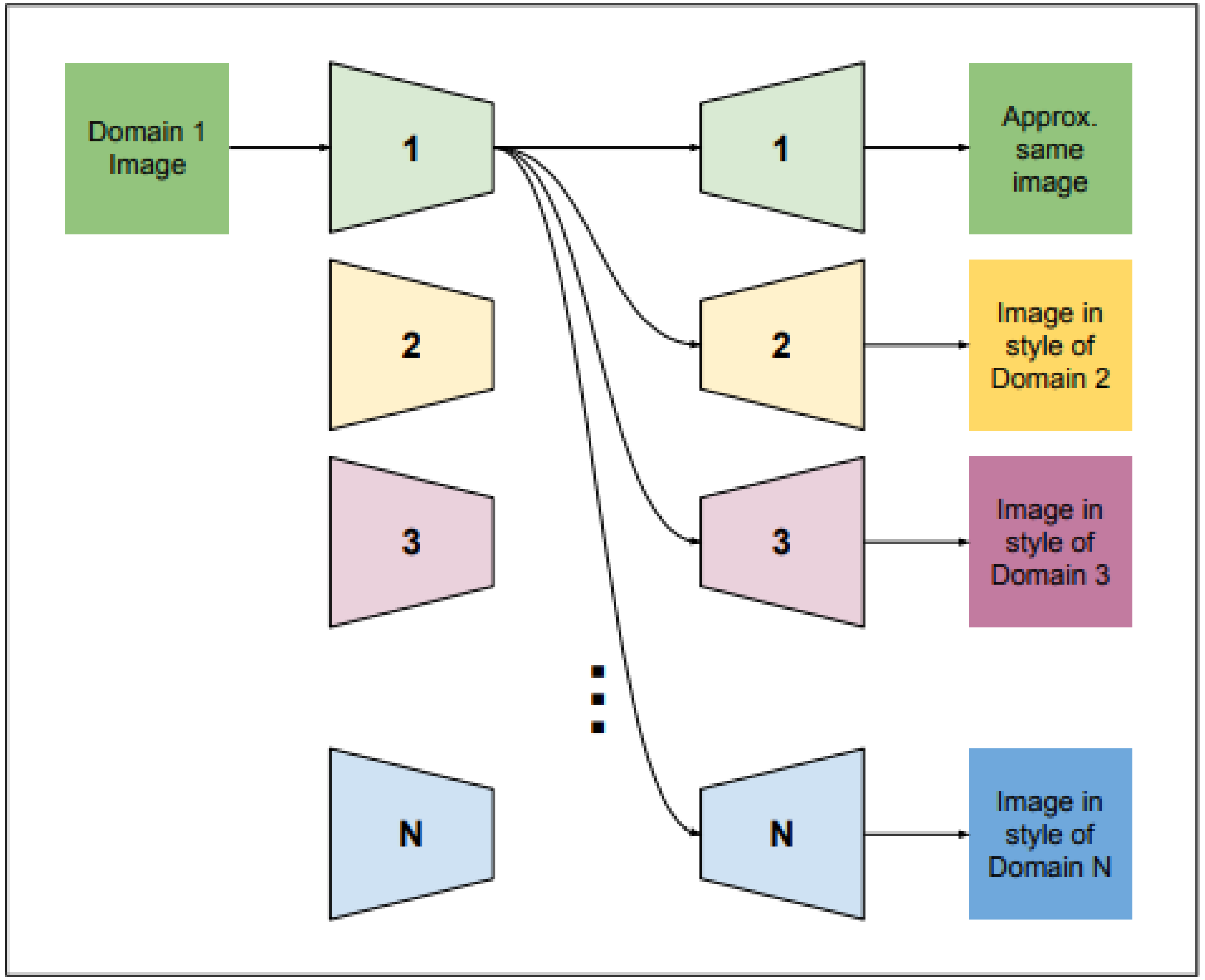
Radu Timofte
ETH Zürich
Merantix GmbH
timofter@ethz.ch

Luc Van Gool
ETH Zürich
KU Leuven
vangool@ethz.ch

Anoosheh et al., CVPR 2018

Beyond CycleGAN – More than 2 domains





Anoosheh et al., CVPR 2018

#CycleGAN at Twitter

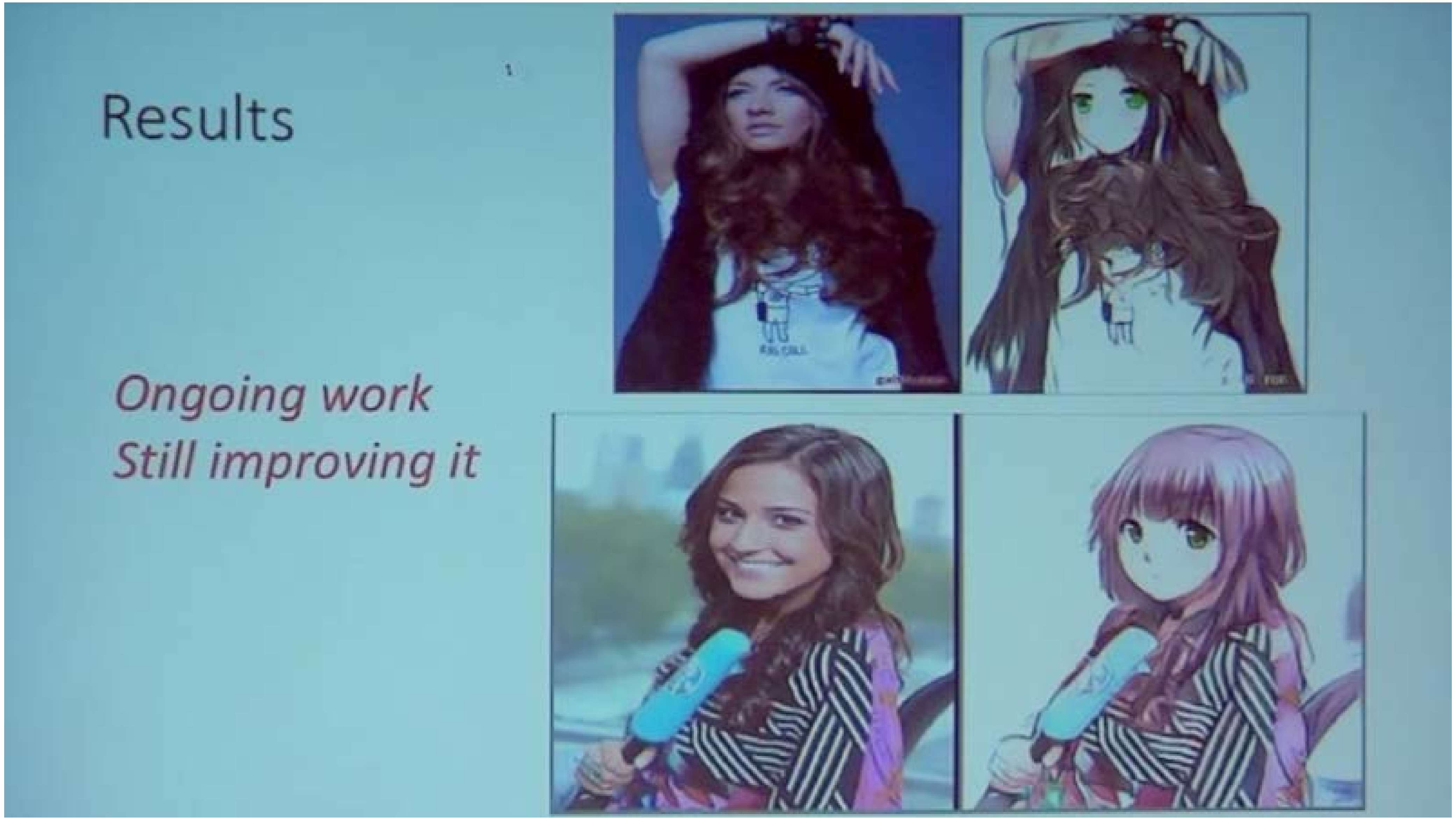
Monet → Thomas Kinkade @David Fouhey

Resurrecting Ancient Cities @ Jack Clark

Birds @Matt Powell

Bear → Panda @Matt Powell

Turn Real People Into Anime Art



@minjunli (Minjun Li), @Aixile (Yanghua JIN), @alanyttian (Yingtao Tian)

Latest from #CycleGAN

CycleGAN with architectural modifications, by itok_msi

https://qiita.com/itok_msi/items/b6b615bc28b1a720afd7

Thank You!

Code and data: junyanz.github.io/CycleGAN/