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

Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros; ICCV, 2017

Outline

1. **Introduction**
    1. Image-to-image translation
    2. Paired vs unpaired
    3. GANs
2. **CycleGAN**
    1. General framework
    2. Cycle consistency
    3. Results
    4. Limitations
3. **Medical applications**
    1. Cross-modal image synthesis
    2. Segmentation with CycleGAN
    3. Shape consistency
4. **Conclusions and discussion**

I. Introduction

Image-to-image translation

?

Claude Monet, *The Seine river at Argenteuil, 1873*

Realistic scene as seen by the artist

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

Image-to-image translation

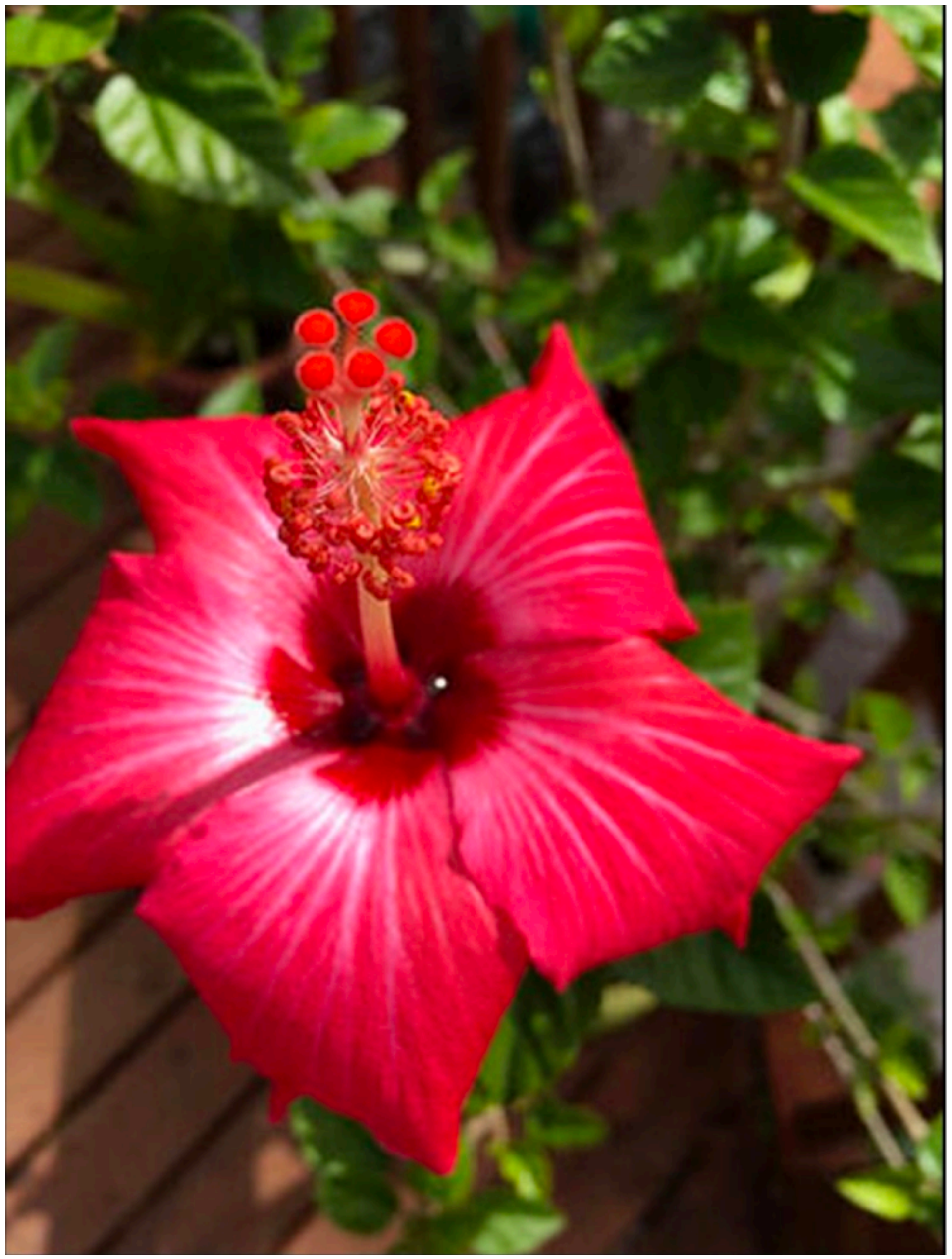
Claude Monet, *The Seine river at Argenteuil, 1873*

CycleGAN, 2017

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

Image-to-image translation

Mapping a given scene from a representation in domain X to another representation in domain Y



$\longrightarrow$ **?**

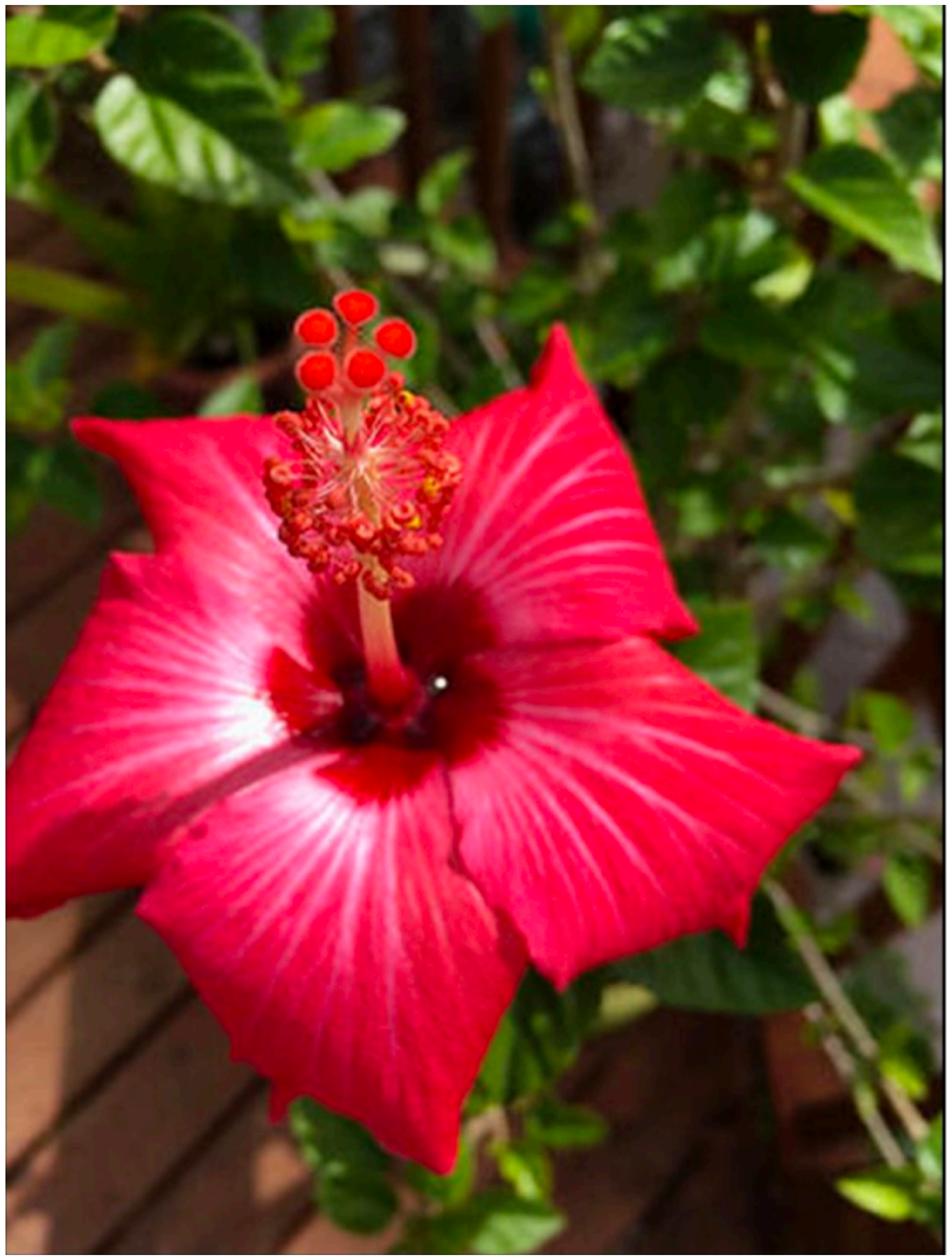
taken on iPhone

taken by DSLR

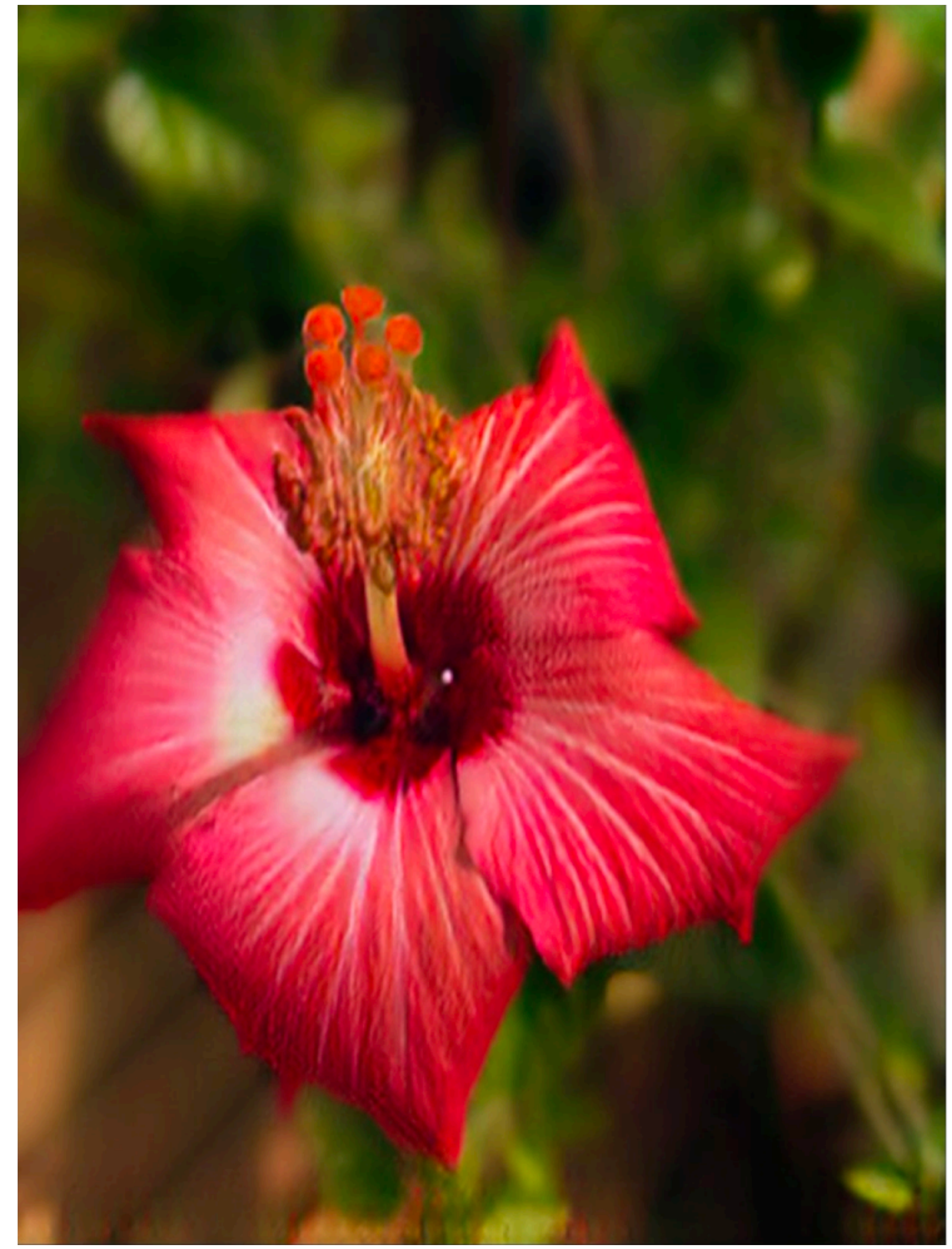
How can we learn such a translation?

Image-to-image translation

Mapping a given scene from a representation in domain X to another representation in domain Y



taken on iPhone



CycleGAN

How can we learn such a translation?

Paired Im-2-Im

- We have matching pairs of training data
- Supervised learning approach
- Training on joint distribution of the two representations

$$\{x_i, y_i\}_{i=1}^{N}$$

Drawbacks:

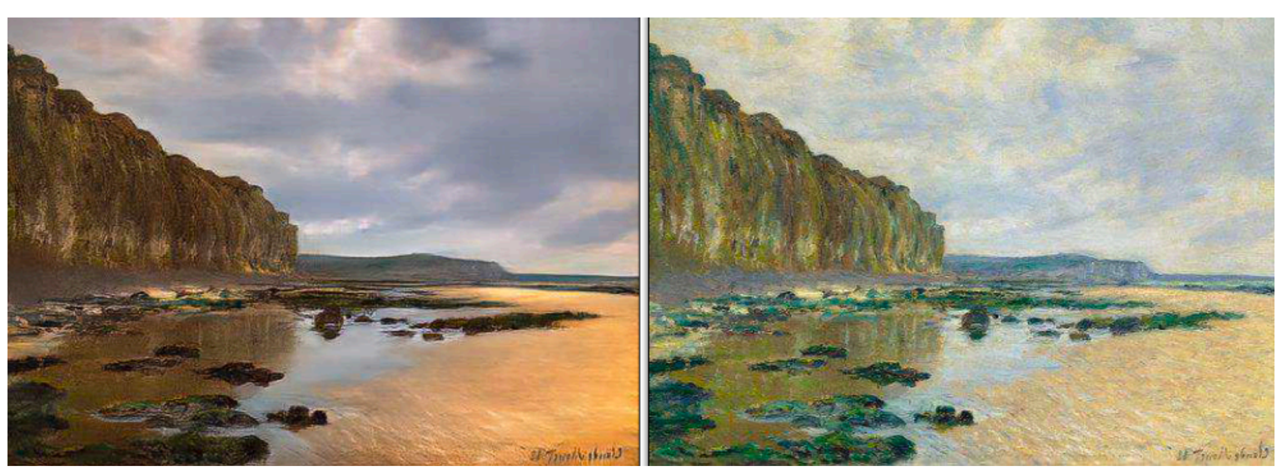
- obtaining paired training data is difficult
- might need expert authoring to create dataset
- some desired outputs might not even be well-defined

Unpaired Im-2-Im

- Training data consists of $x_i \in X$ source set and $y_i \in Y$ target set
- No information provided on matching between X and Y
- Learn joint distribution from marginal distributions

X: landscape photos

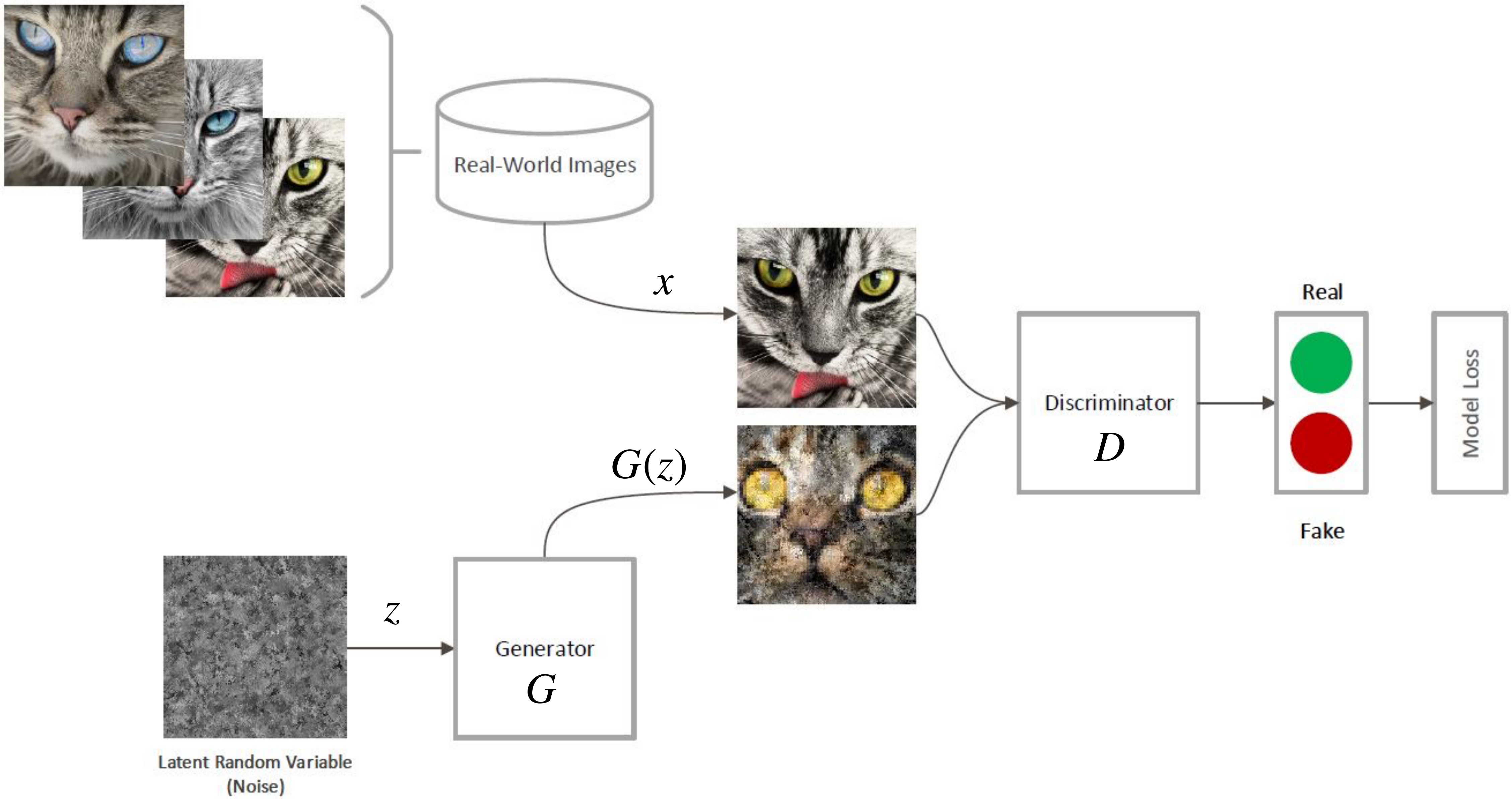
Y: paintings of landscapes



unseen photo generated painting

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

Generative adversarial nets



$$\min_G \max_D \mathbb{E}_{x \sim q_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

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

II. CycleGAN

CycleGAN framework

- We want to learn mapping functions between two domains:

  $G : X \to Y$ maps images from domain X to domain Y

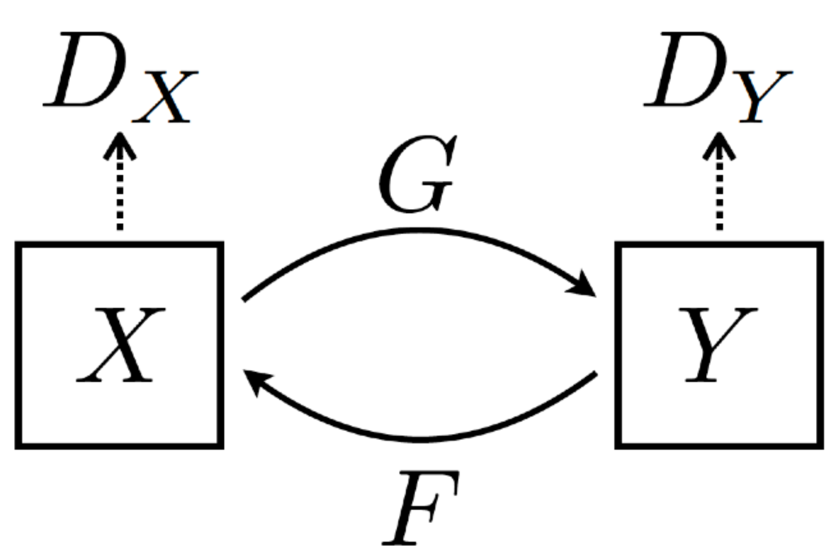
  $F : Y \to X$ maps images from domain Y to domain X

- We use adversarial training framework:

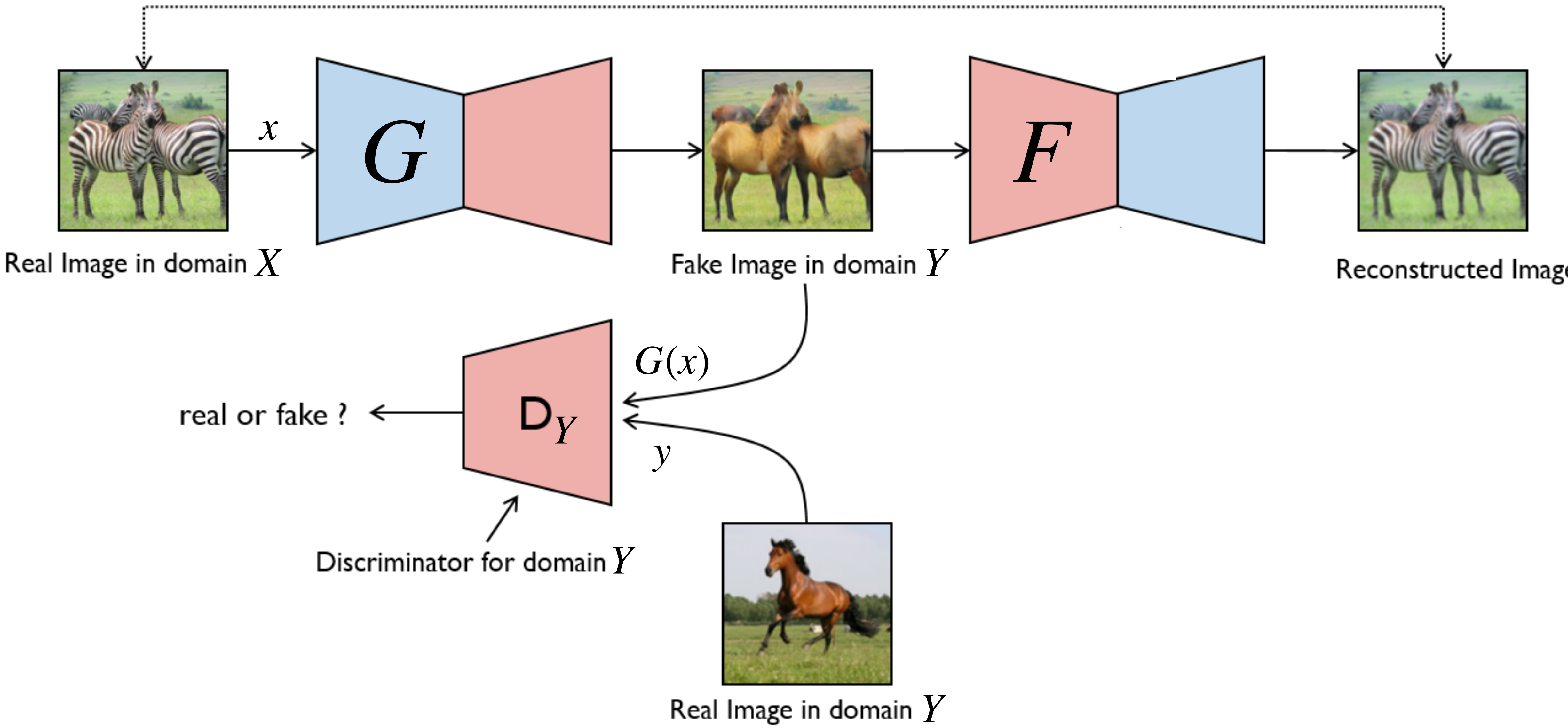
  D_X discriminator trying to distinguish between images from domain X and translated 'fake' images $F(y), \quad y \in Y$

  D_Y discriminator trying to distinguish between images from domain Y and translated 'fake' images $G(x), \quad x \in X$

- Generators (G, F) and discriminators (D_X, D_Y) are parameterized as neural networks



Adversarial loss



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

$$\mathcal{L}_{GAN}(F, D_X, X, Y) = \mathbb{E}_{x \sim p_{data}(x)}\left[\log D_X(x)\right] + \mathbb{E}_{y \sim p_{data}(y)}\left[1 - \log D_X(F(y))\right]$$

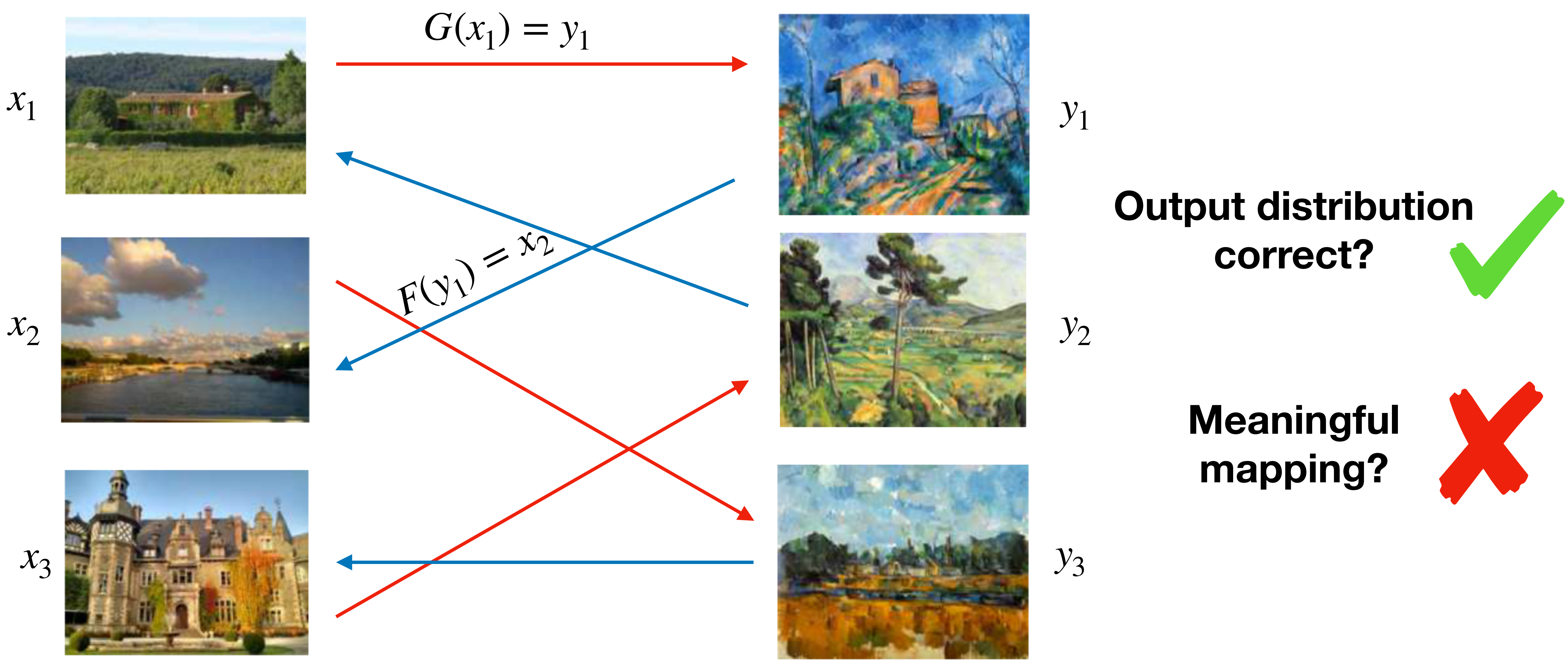
Adversarial loss

Is adversarial loss enough?

- In theory: yes, network can learn mappings such that

$$G(X) \sim Y \quad \text{and} \quad F(Y) \sim X$$

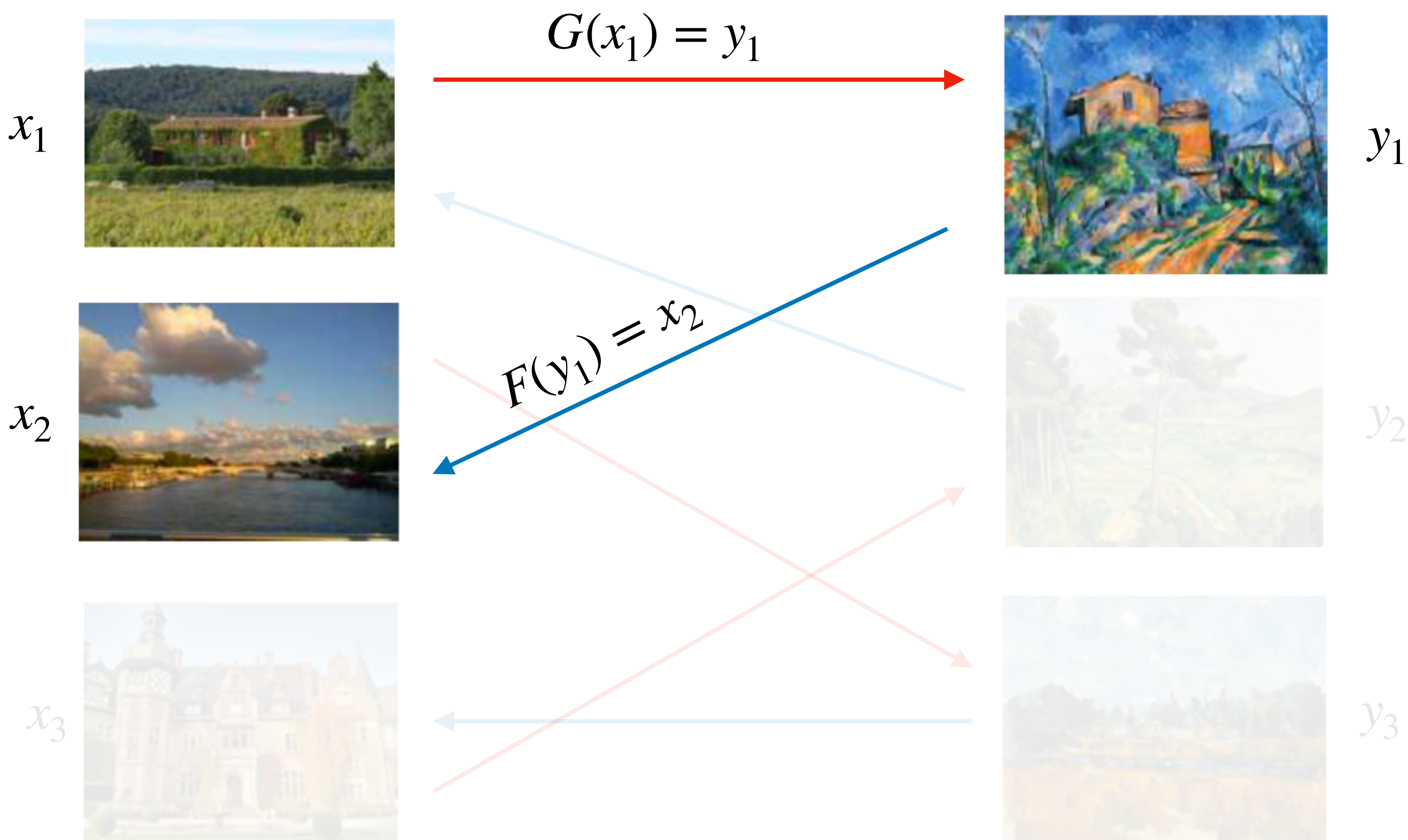
- but with large enough network capacity the generators can memorize some permutation of the target dataset



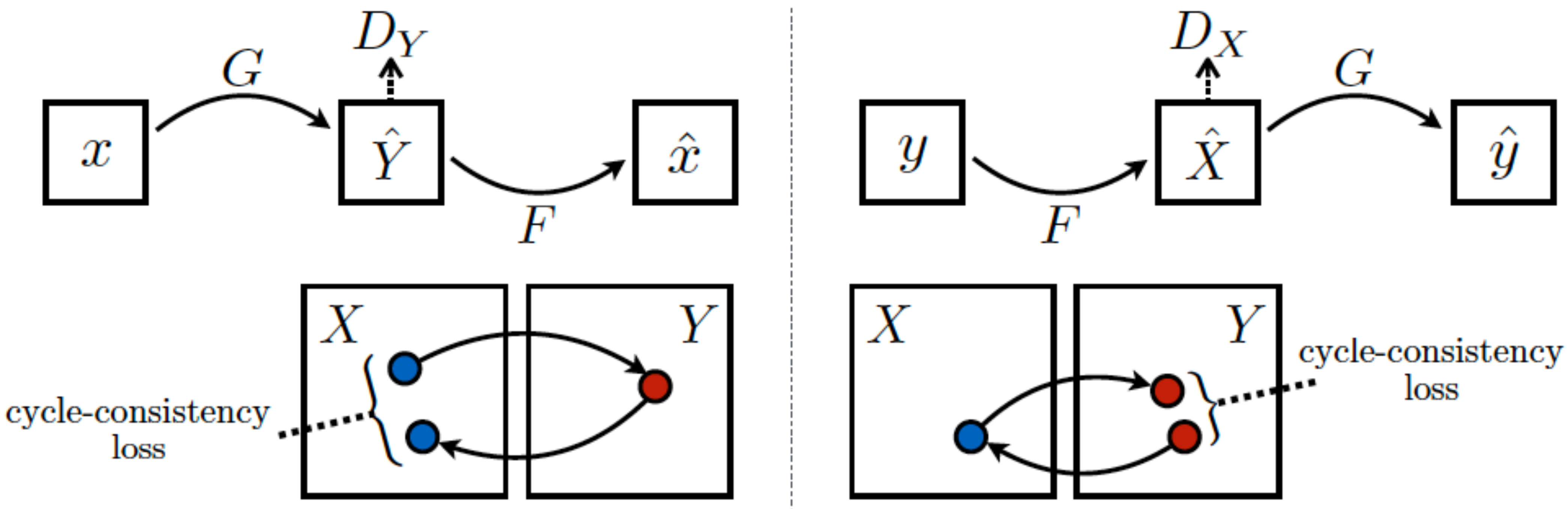
Cycle consistency

- What is wrong with the previous network?

spaceship $\xrightarrow[G]{\textit{en-hun}}$ űrhajó $\xrightarrow[F]{\textit{hun-en}}$ spaceship



Cycle consistency loss



$$\mathcal{L}_{cyc}(G,F) = \mathbb{E}_{x \sim p_{data}(x)}\left[\|F(G(x)) - x\|_1\right] + \mathbb{E}_{y \sim p_{data}(y)}\left[\|G(F(y)) - y\|_1\right]$$

Full objective function

- Training objective:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, X, Y) + \lambda \mathcal{L}_{cyc}(G, F)$$

- Formulation:

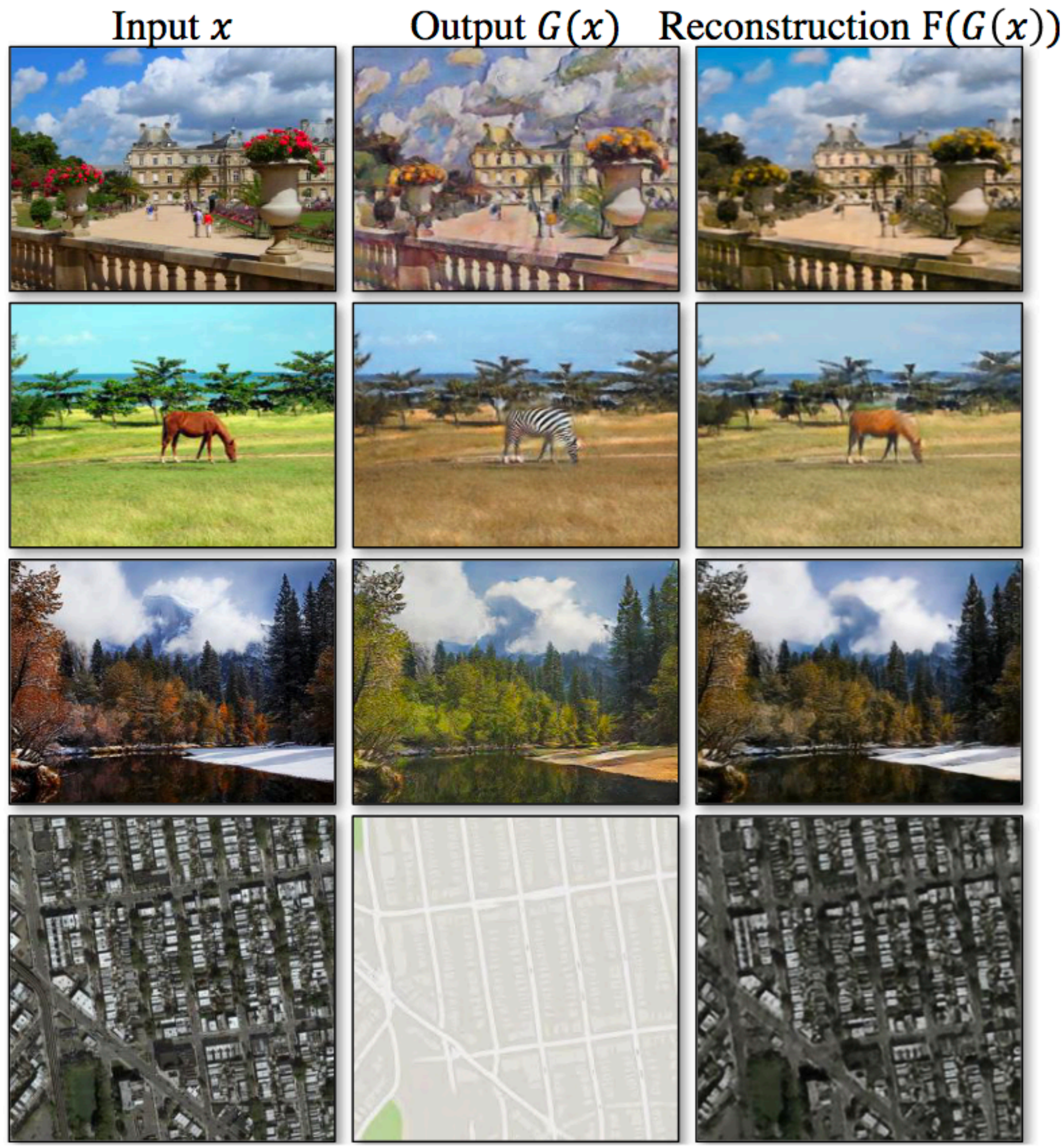
$$G^*, F^* = \arg \min_{G,F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y)$$

- Typical solution: alternating optimization

1. Sample from X and Y
2. Fix discriminator, update generator
3. Fix generator, update discriminator

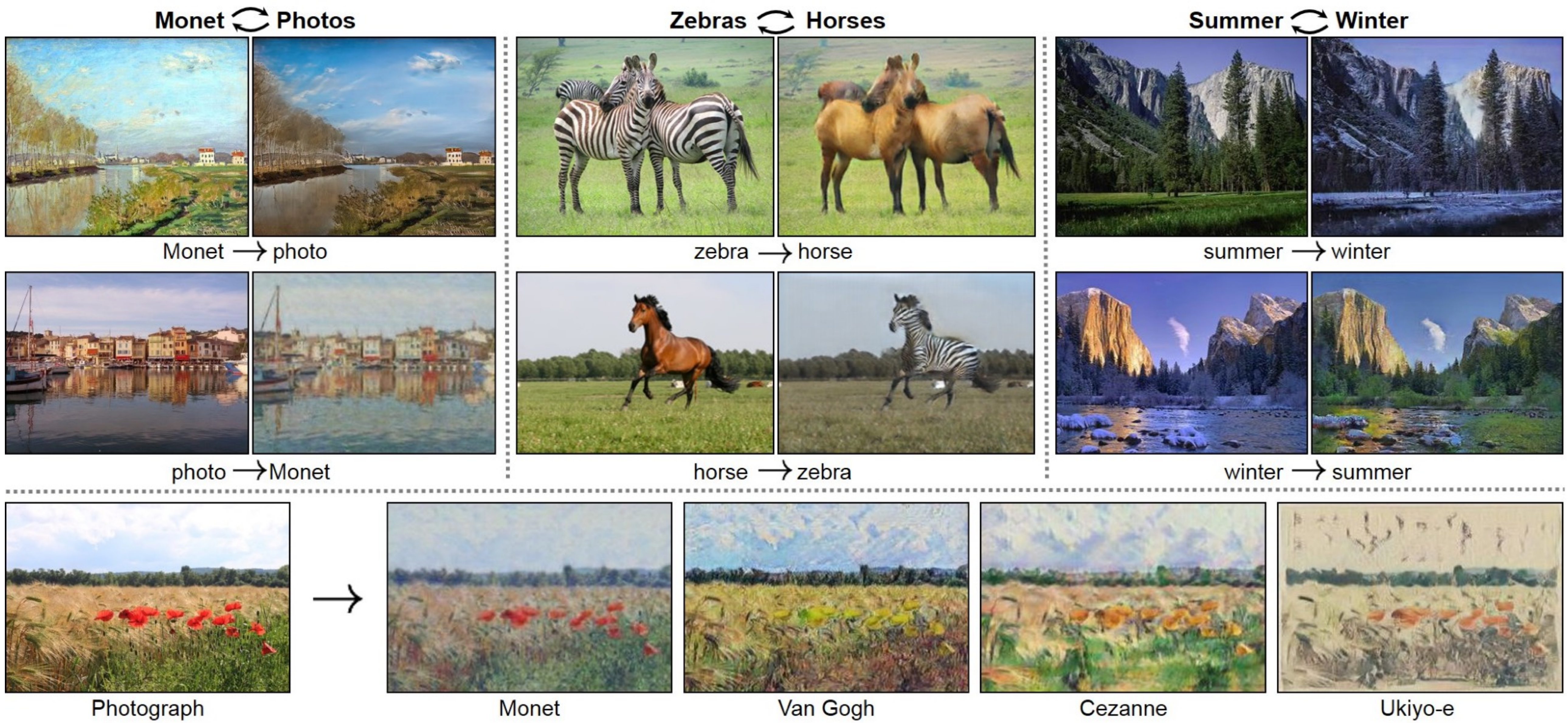
Results

Cycle consistency on reconstructions



Results

Where it works…



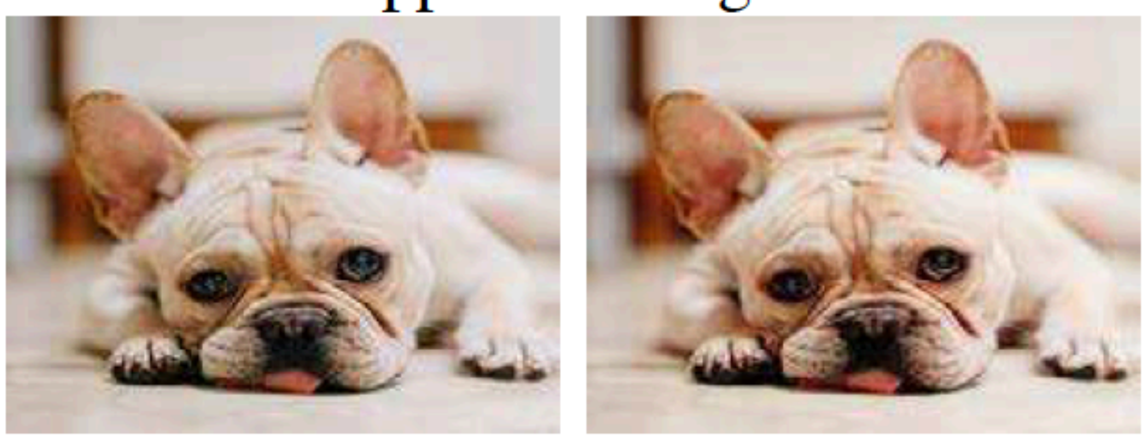
Changes in texture ✓

Changes in color ✓

Limitations

and where it doesn't…

apple → orange



dog → cat

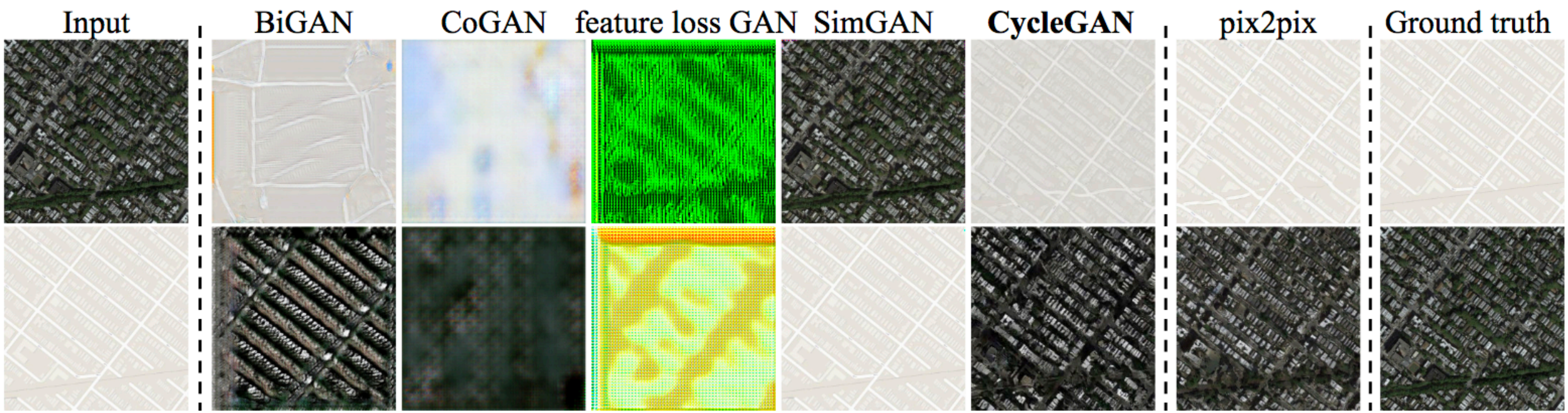
horse → zebra

geometric changes*

unexpected objects

**UNIT is better at geometric transformation: http://papers.nips.cc/paper/6672-unsupervised-image-to-image-translation-network*

Comparison



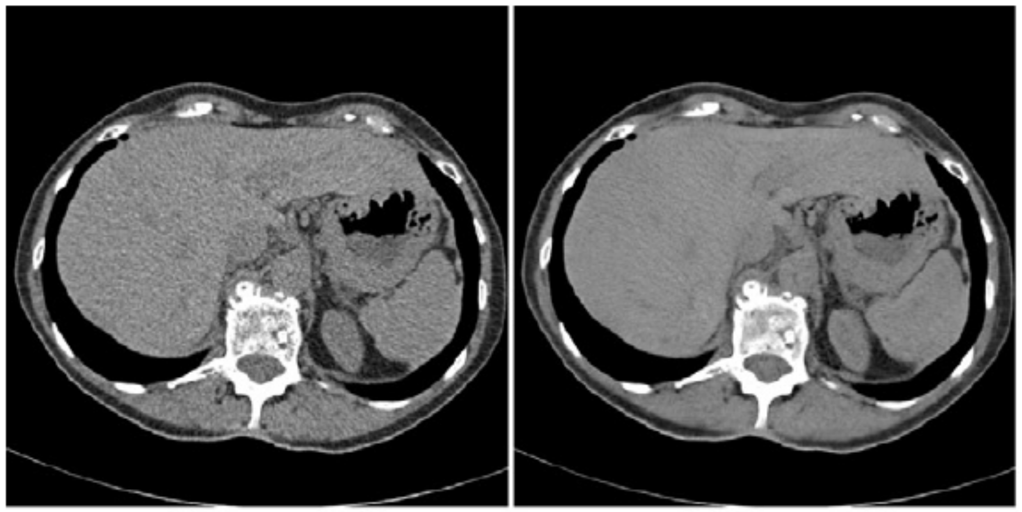
aerial photos ⟷ *maps*

Loss	**Map → Photo** % Turkers labeled *real*	**Photo → Map** % Turkers labeled *real*
CoGAN	0.6% ± 0.5%	0.9% ± 0.5%
BiGAN/ALI	2.1% ± 1.0%	1.9% ± 0.9%
SimGAN	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%**

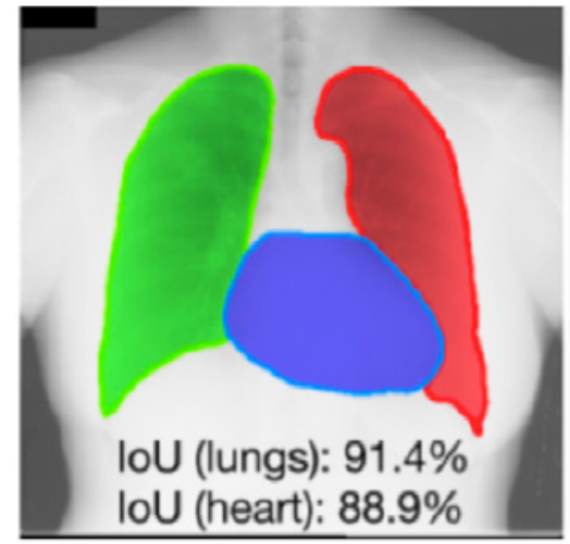
III. Medical applications

Pulications	Method	Loss	Dataset	Measures	Remarks
MR → CT					
Nie et al. (2017, 2018)	Cascade GAN	L1, 2, 4	D16	M11, 12	[√]Brain; Pelvis
Emami et al. (2018)	cGAN	L1, 2	–	M11, 12, 13	[√]Brain
CT → MR					
Jin et al. (2019)	CycleGAN	L1, 2, 3	–	M11, 12	[✗] Brain
Jiang et al. (2018)	CycleGAN*	L1, 2, 3, 7, 8	D8	M32	[✗] Lung
MR ↔ CT					
Chartsias et al. (2017)	CycleGAN	L1, 3	D9	M32	[✗] Heart
Zhang et al. (2018d)	CycleGAN*	L1, 3, 7	–	M32	[✗][3D] Heart
Huo et al. (2018)	CycleGAN*	L1, 3, 7	–	M32	[✗] Spleen
Chartsias et al. (2017)	CycleGAN	L1, 3	–	M32	[✗] Heart
Hiasa et al. (2018)	CycleGAN*	L1, 3, 4	–	M19, 32	[✗] Musculoskeletal
Wolterink et al. (2017a)	CycleGAN	L1, 3	–	M11, 12	[✗] Brain
Huo et al. (2018b)	CycleGAN	L1, 3, 7	–	M32	[✗] Abdomen
Yang et al. (2018b)	CycleGAN*	L1, 2, 3, 10	–	M11, 12, 13	[✗] Brain
Maspero et al. (2018)	pix2pix	L1, 2	–	M11, 22	[√] Pelvis
CT → PET					
Bi et al. (2017)	cGAN	L1, 2	–	M11, 12	[√] Chest
Ben-Cohen et al. (2018)	FCN+cGAN	L1, 2	–	M11, 12, 31	[√] Liver
PET → CT					
Armanious et al. (2018c)	cGAN*	L1, 2, 8, 11	–	M11, 12, 13, 14, 15, 18	[√] Brain
MR → PET					
Wei et al. (2018)	cascade cGAN	L1, 2	–	M29	[√] Brain
Pan et al. (2018)	3D CycleGAN	L1, 2, 3	D16	M30	[√] Brain
PET → MR					
Choi and Lee (2017)	pix2pix	L1, 2	D16	M13, 29	[√] Brain
Synthetic → Real					
Hou et al. (2017)	synthesizer+cGAN	L1, 2, 7	D35, 36	M1, 32	[√] Histopathology
Real → Synthetic					
Mahmood et al. (2018)	cGAN	L1, 12	–	M34	[✗] Endocsocpy
Zhang et al. (2018c)	CycleGAN*	L1, 3, 7	–	M32	[✗] X-ray
Domain adaption					
Chen et al. (2018a)	CycleGAN*	L1, 3, 7	D32, 33	M32	[✗] X-ray
T1 ↔ T2 MR					
Dar et al. (2019)	CycleGAN	L1, 3	D11, 19, 22	M12, 13	[✗] Brain
Yang et al. (2018c)	cGAN	L1, 2	D19	M11, 12, 19, 32, 33	[✗] Brain
Welander et al. (2018)	CycleGAN, UNIT	L1, 2, 3	D24	M11, 12, 19	[✗] Brain
Liu (2018)	CycleGAN	L1, 2, 3	D14	M32	[✗] Knee

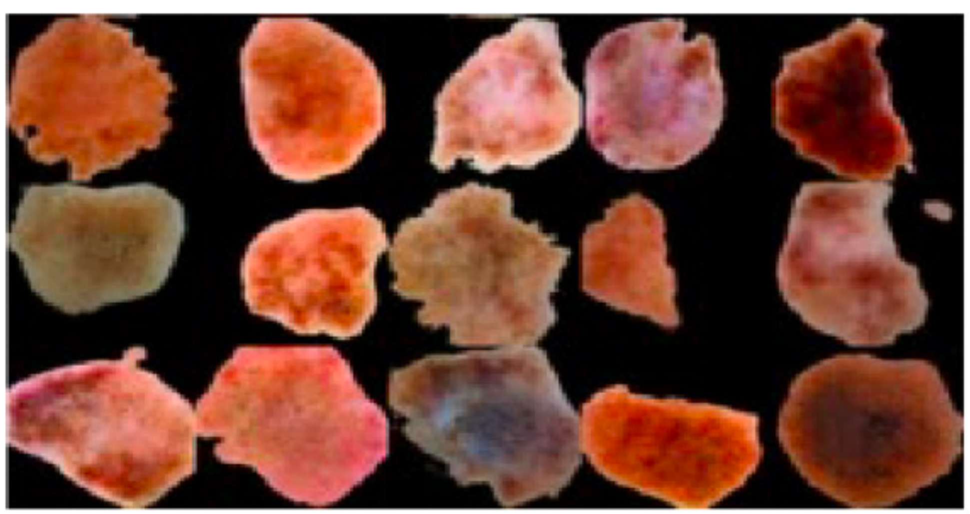
Deep learning in medicine



Denoising: low-field MRI, low dose CT (above)



Segmentation: organ (above), tumor



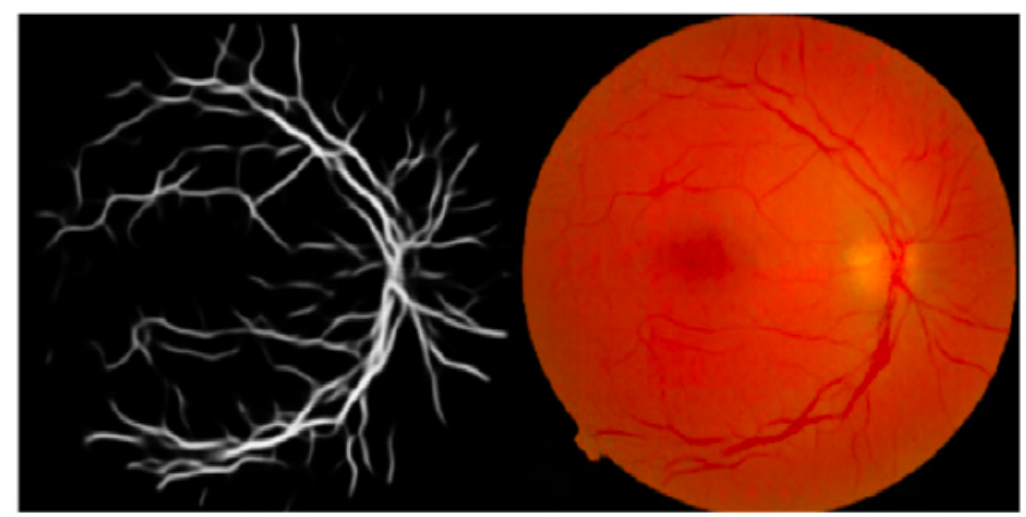
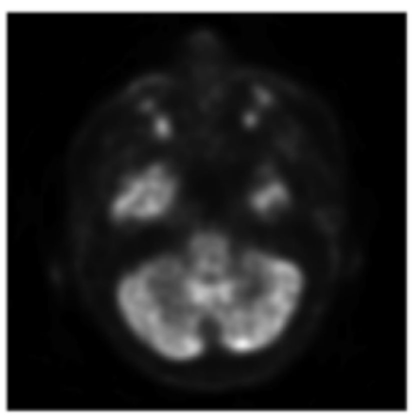


Image synthesis: skin lesions (left), retinal images (right)

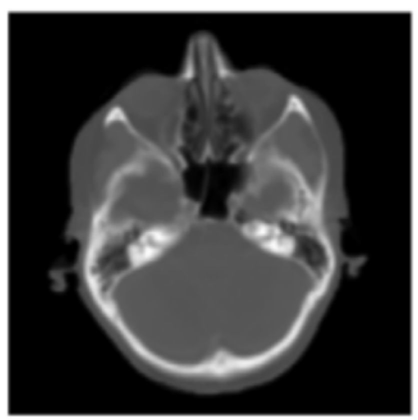
Yi et al., Generative adversarial network in medical imaging: A review, *Medical image analysis*, 2019

Cross-modal image synthesis

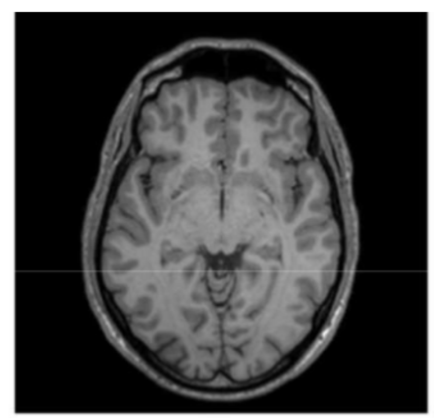
- Deep learning models are extremely data-hungry
- Data collection for medical tasks is challenging:
  - expensive instruments (MR scanner)
  - radiation exposure (CT, PET)
  - expert knowledge (doctors) needed
  - patient confidentiality guidelines
  - lack of medical data standards (compatibility)
- Often we have some labeled medical data but in different modalities



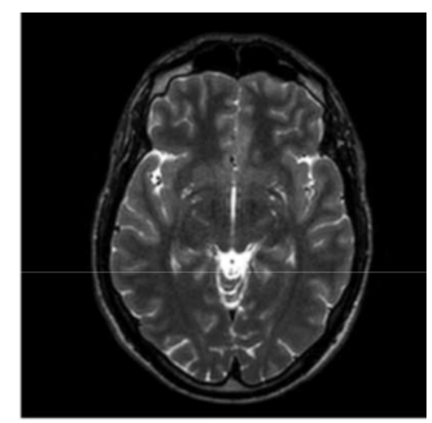
PET



CT



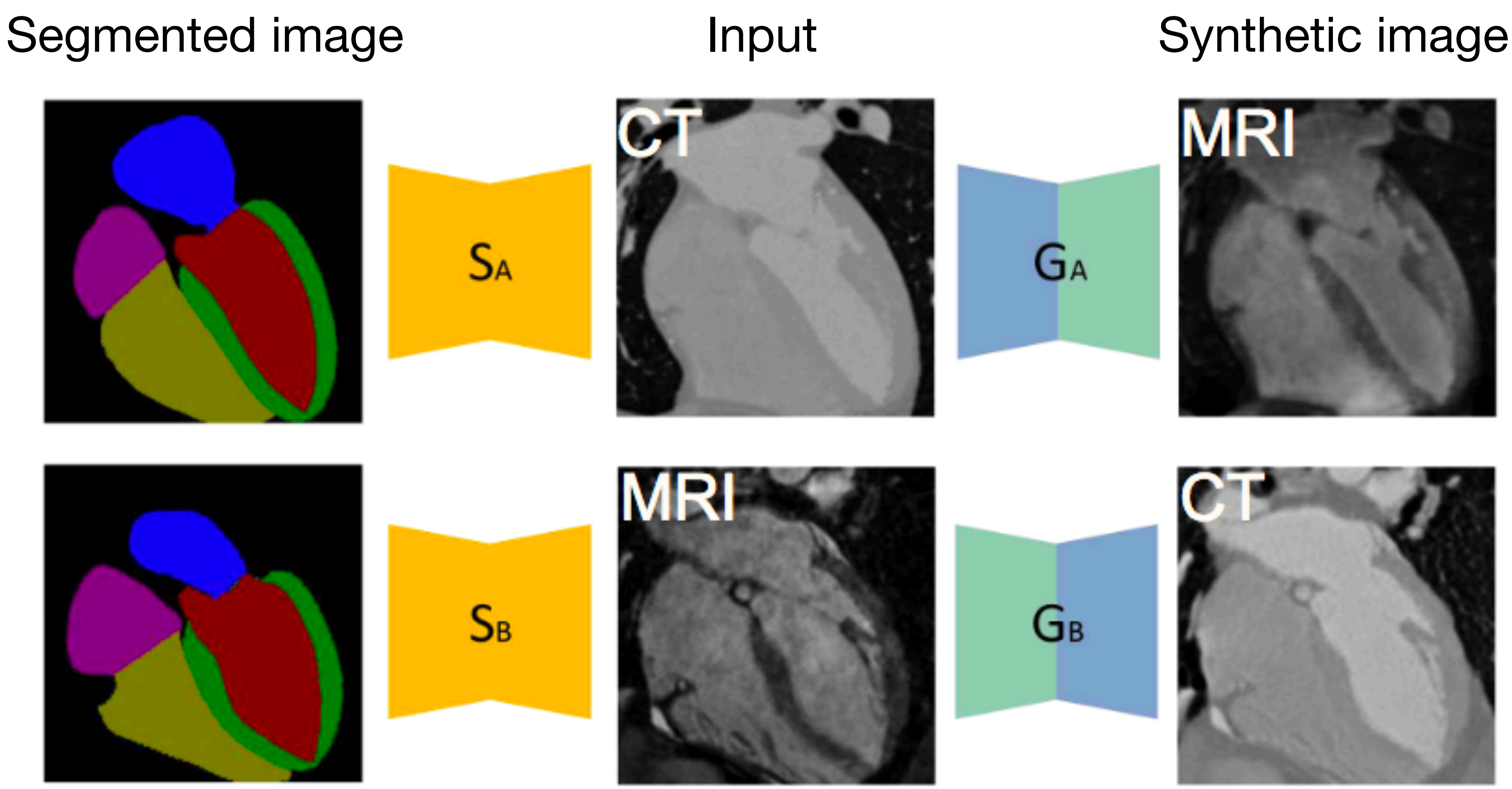
MRI, T1



MRI, T2

- Idea: translate all available data to the same modality!

Segmentation with CycleGAN



*Zhang et al., Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network, *CVPR 2018*

Shape consistency

- Intrinsic ambiguity of cycle consistency to geometric transformations
- Assume G_A and G_B are cycle consistent:

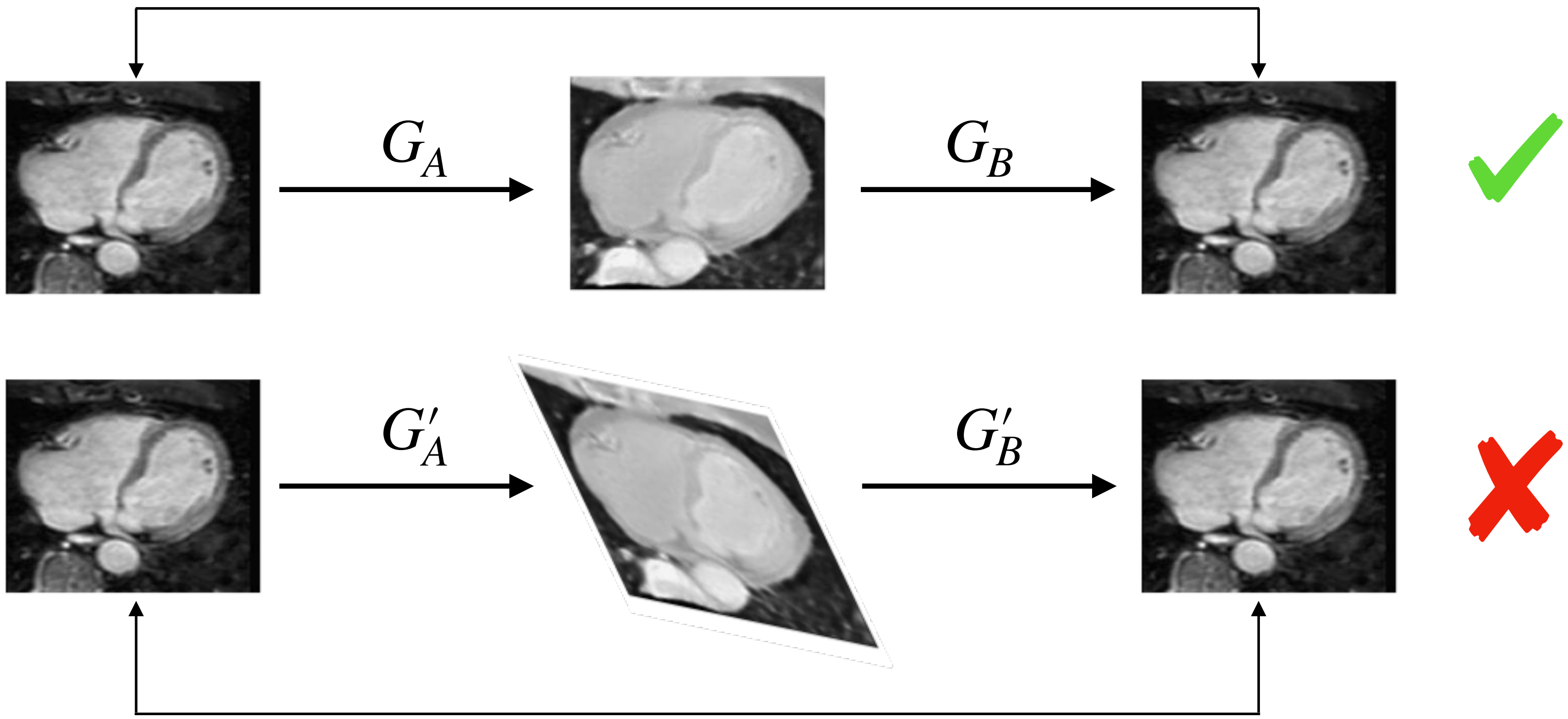
$$G_A(G_B(A)) = A$$

- Let T be a bijective geometric transformation with inverse T^{-1} and

$$G'_A = G_A \circ T$$

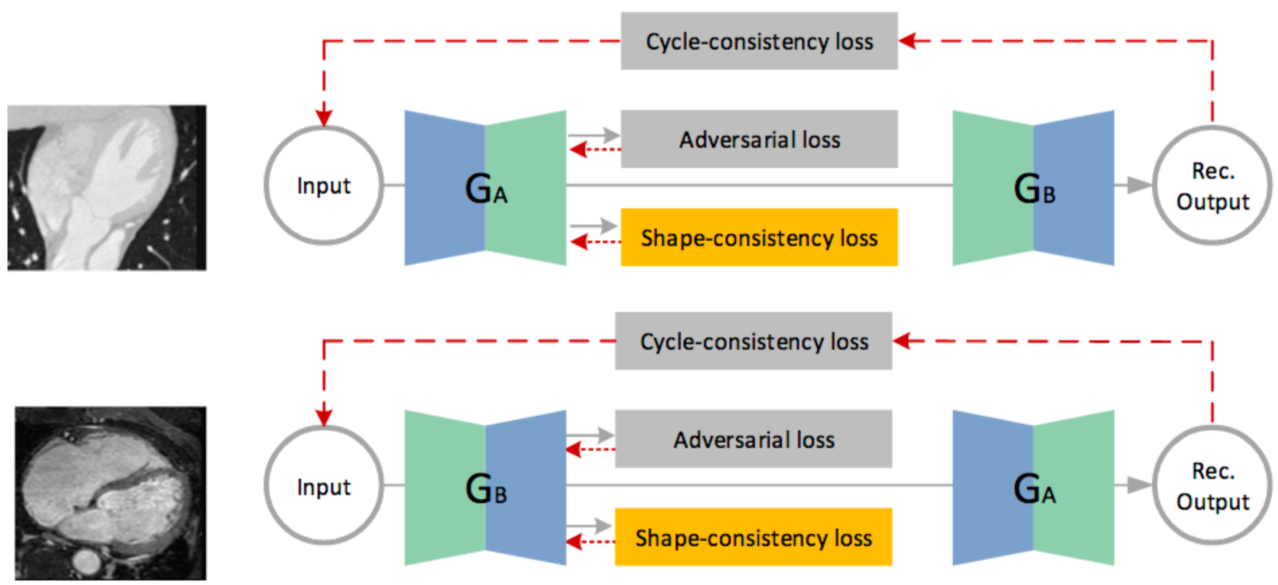
$$G'_B = G_B \circ T^{-1}$$

- Then G'_A and G'_B are also cycle consistent!

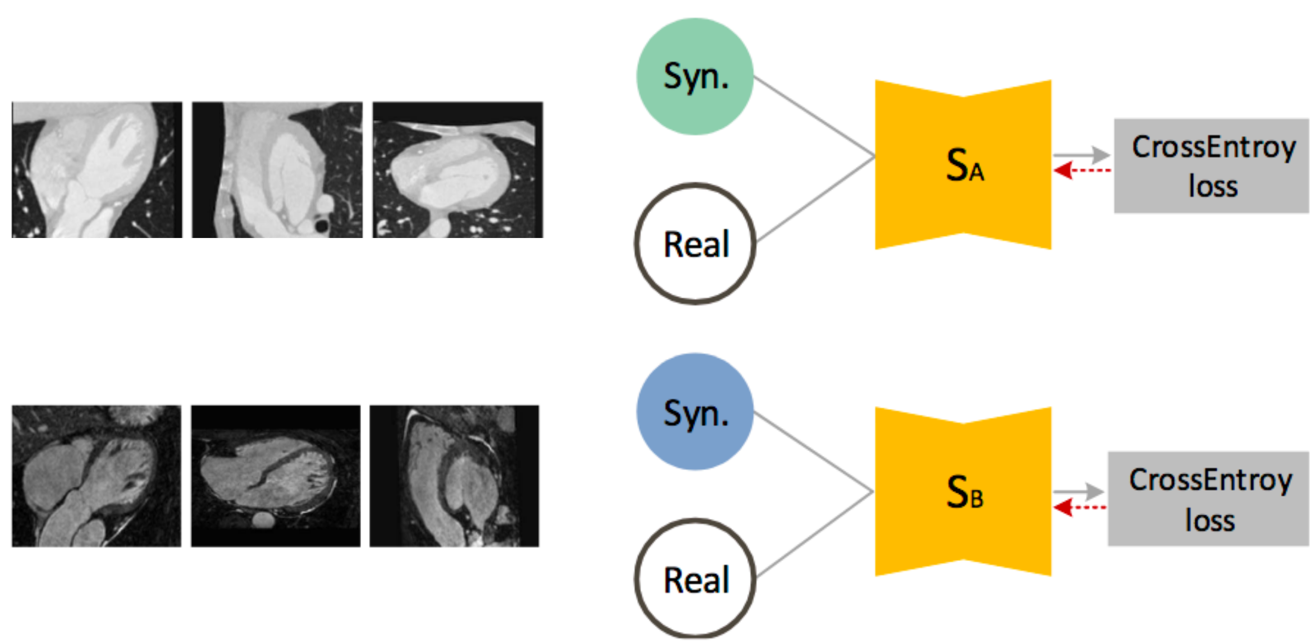


geometric distortion impacts medical diagnosis!

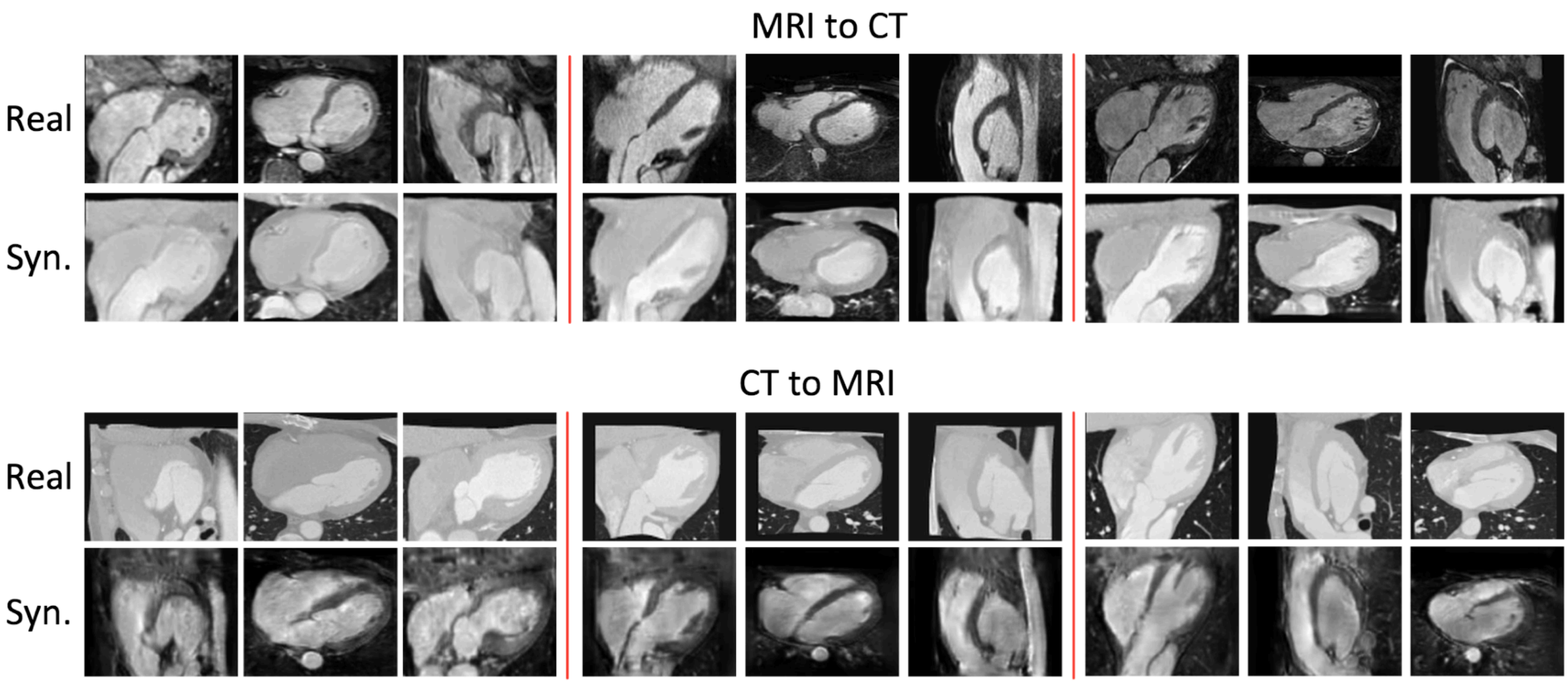
Segmentation with CycleGAN



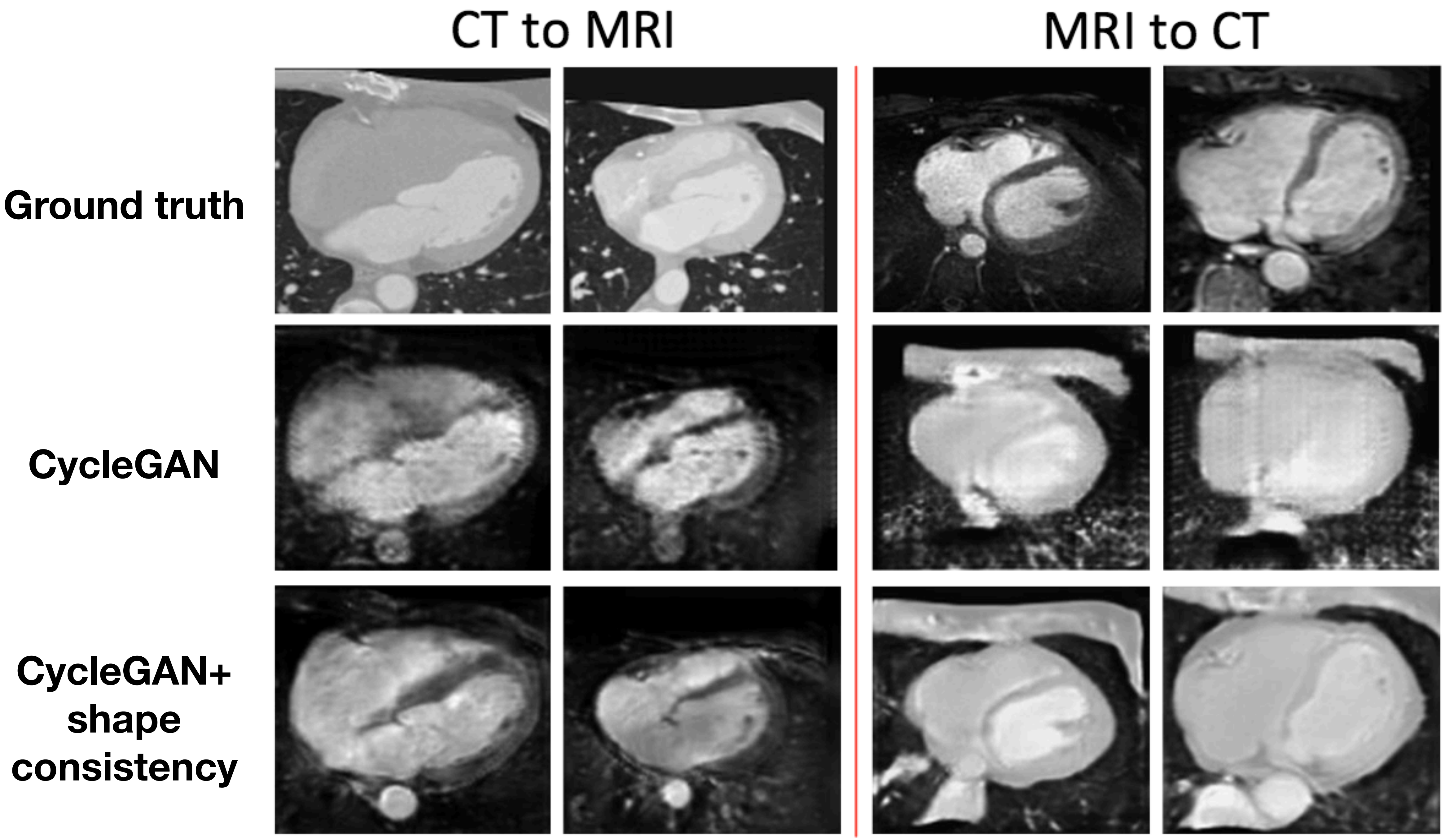
$$\mathcal{L}_{shape}(S_A, S_B, G_A, G_B) = \mathbb{E}_{x_B \sim p_d(x_B)}[-\frac{1}{N}\sum_i y_B^i log(S_A(G_A(x_B))_i)] + \mathbb{E}_{x_A \sim p_d(x_A)}[-\frac{1}{N}\sum_i y_A^i log(S_B(G_B(x_A))_i)]$$



Translation results



Translation results

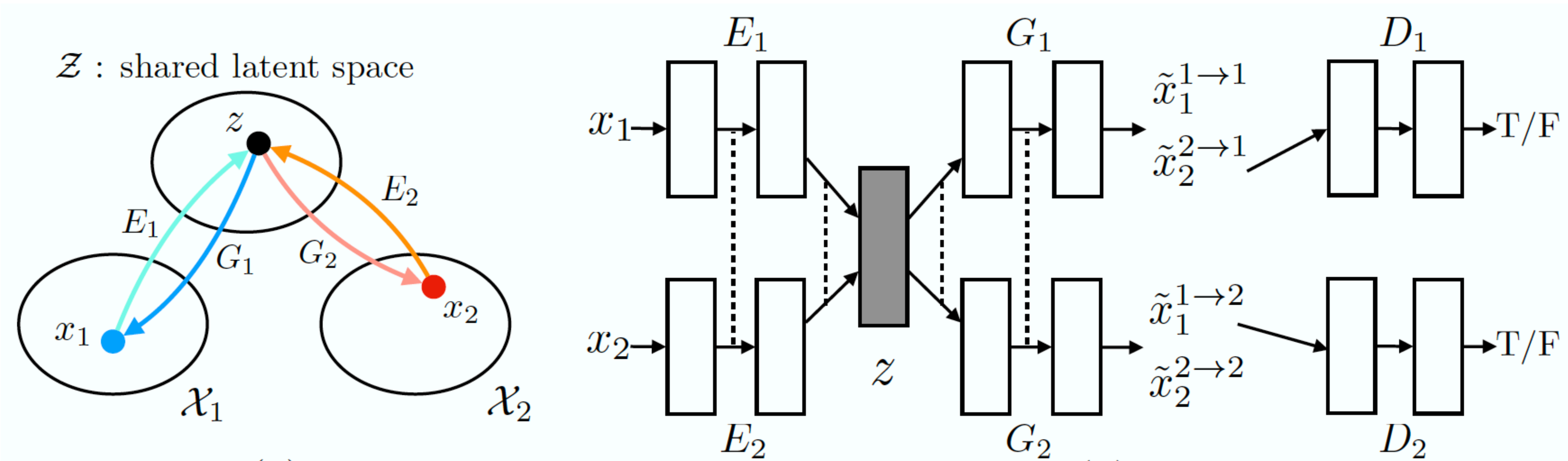


Conclusion

- In image-to-image translation we want to learn a **meaningful** mapping from one image domain to another.
- Generative adversarial models are powerful tools for such problems
- But we need extra regularization on top of adversarial loss
- Cycle consistency narrows down the space of desirable mappings by ensuring that translating an image forward and backward results in the original image
- Applications range from style transfer and photo enhancement to medical image synthesis

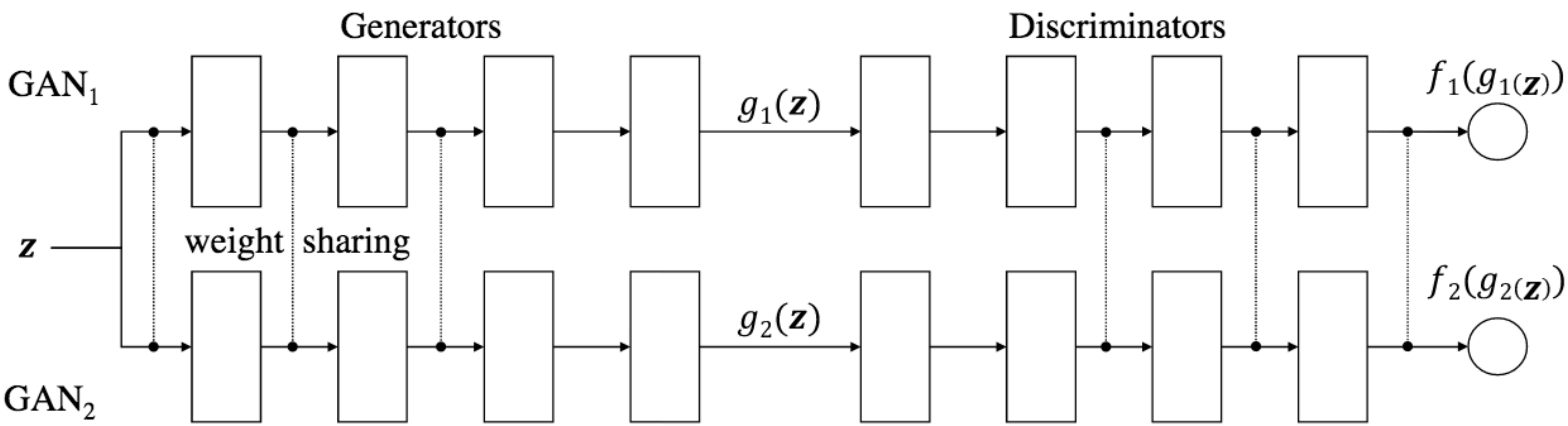
Q/A

UNIT



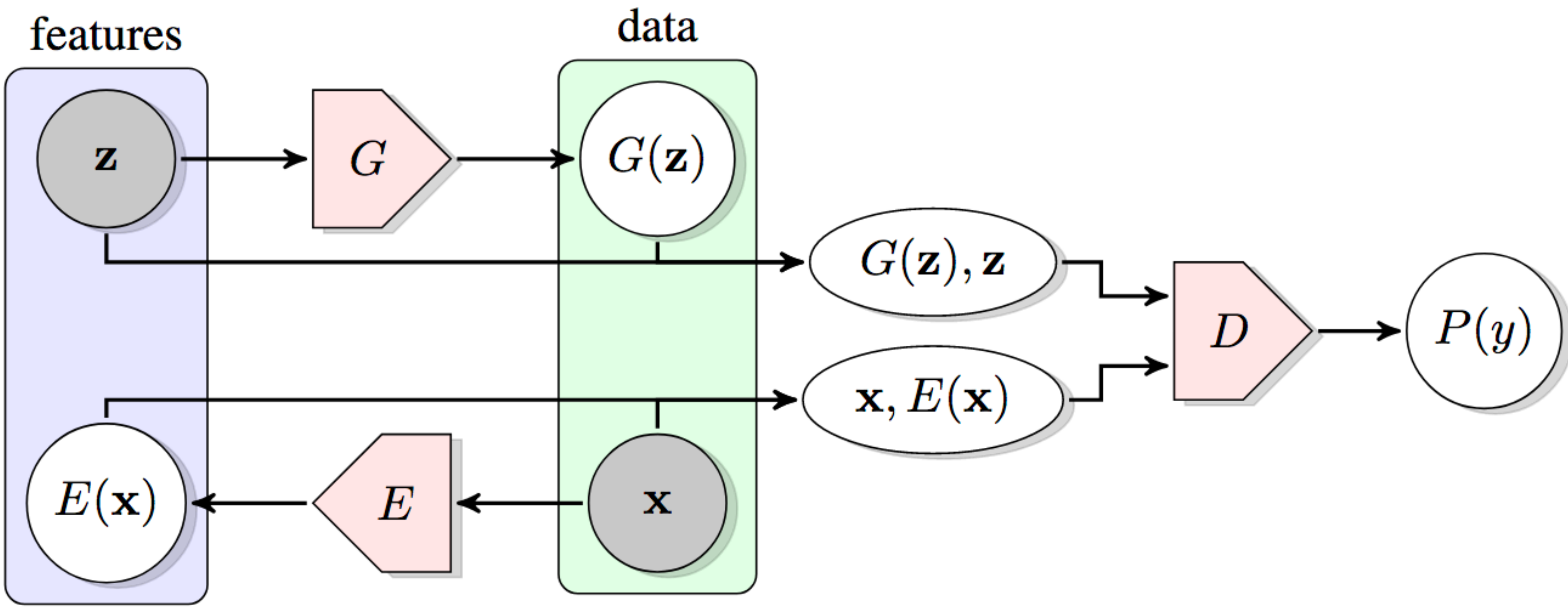
$\mathcal{Z}$: shared latent space
z
E_1
G_1
G_2
E_2
x_1
x_2
$\mathcal{X}_1$
$\mathcal{X}_2$
E_1
G_1
D_1
x_1
x_2
$\tilde{x}_1^{1\to1}$
$\tilde{x}_2^{2\to1}$
$\tilde{x}_1^{1\to2}$
$\tilde{x}_2^{2\to2}$
T/F
T/F
z
E_2
G_2
D_2

CoGAN



Generators
Discriminators
GAN$_1$
GAN$_2$
z
weight sharing
$g_1(\mathbf{z})$
$g_2(\mathbf{z})$
$f_1(g_{1(\mathbf{z})})$
$f_2(g_{2(\mathbf{z})})$

BiGAN



pix2pix