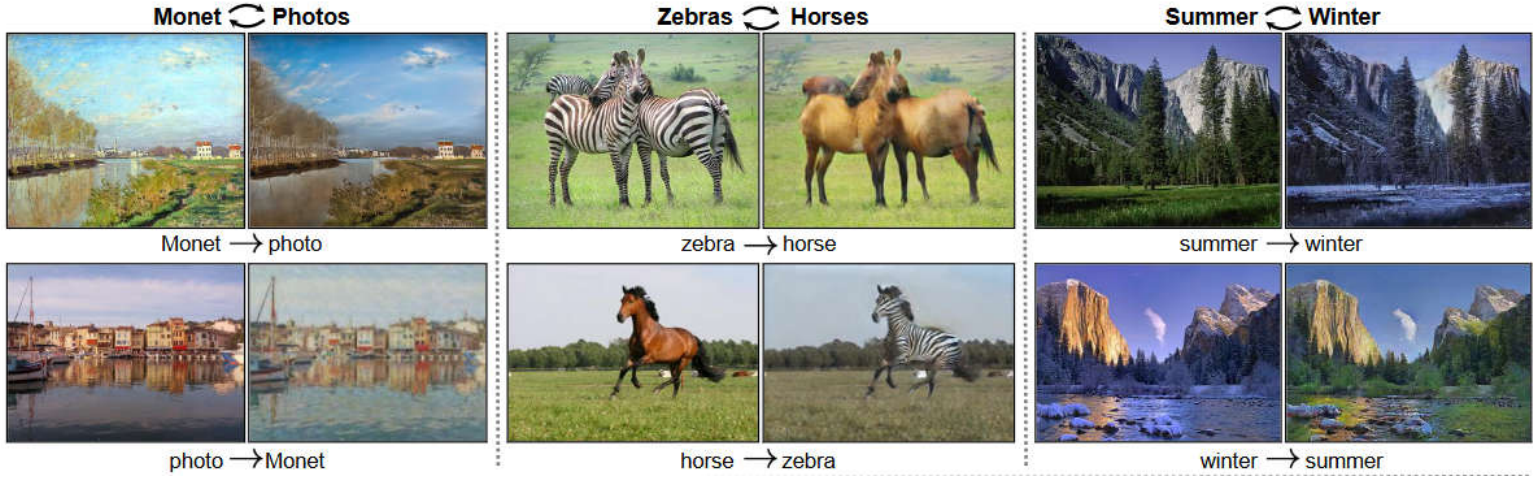
# Image-to-Image Translation

# Background

- Image-to-image translation

  Mapping an input image of a source class to an analogues image of a target class.




- Unsupervised image-to-image translation

  Learning the mapping without pair supervision.

# Background

- Multi-class unsupervised image-to-image translation


- Few-shot multi-class unsupervised image-to-image translation

# CycleGAN

- Training samples $\{x_i\}_{i=1}^{N} \quad \{y_j\}_{j=1}^{M}$

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

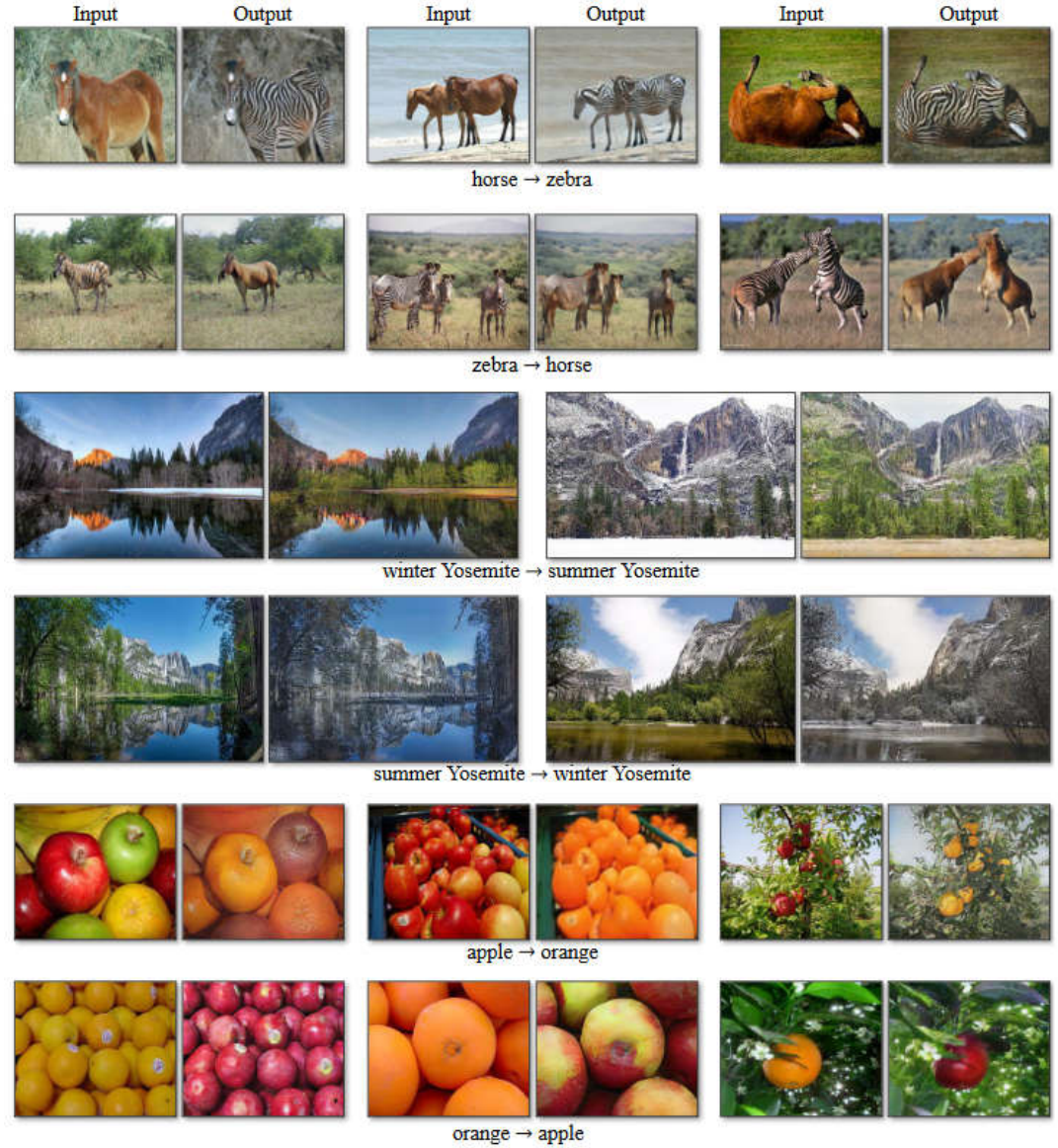
- X -> Y $\min_G \max_{D_Y} \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y)$
- Y -> X $\min_F \max_{D_X} \mathcal{L}_{\text{GAN}}(F, D_X, Y, X)$

Key idea:

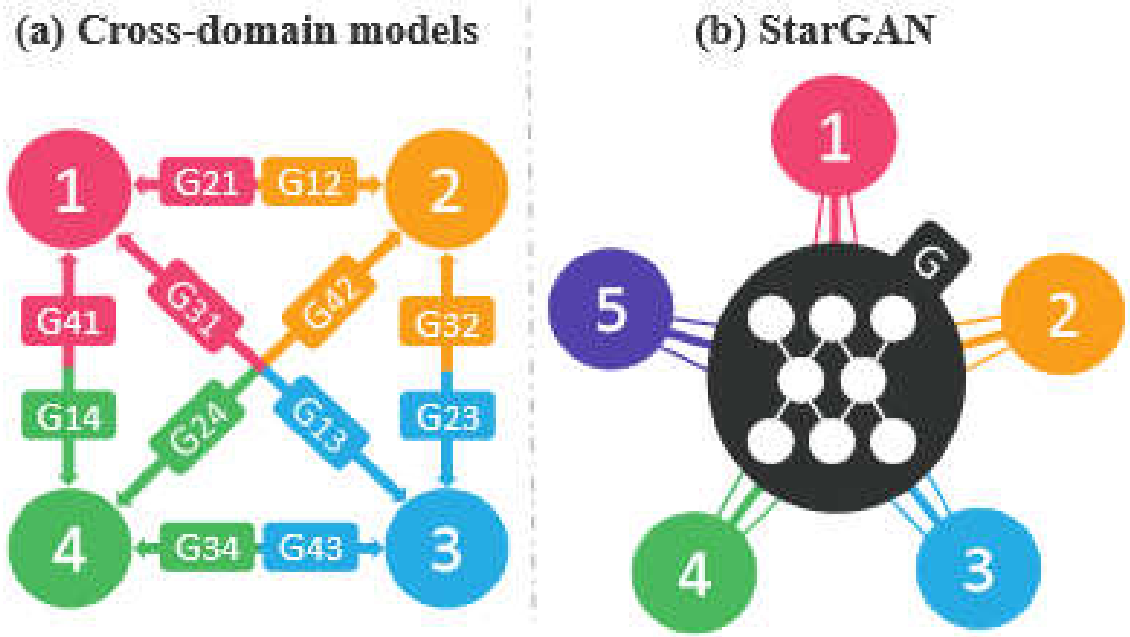
for each image x from domain X, the image translation cycle should be able to bring x back to the original image, i.e.

$$x \rightarrow G(x) \rightarrow F(G(x)) \approx x.$$


Input
Monet
Van Gogh
Cezanne
Ukiyo-e


Input
Output
Input
Output
Input
Output
horse → zebra
zebra → horse
winter Yosemite → summer Yosemite
summer Yosemite → winter Yosemite
apple → orange
orange → apple

# StarGAN


(a) Cross-domain models
(b) StarGAN
1
2
3
4
5
G21
G12
G41
G31
G42
G32
G14
G24
G13
G23
G34
G43
G

# StarGAN

GAN loss:

$$\mathcal{L}_{adv} = \mathbb{E}_x [\log D_{src}(x)] + \mathbb{E}_{x,c}[\log (1 - D_{src}(G(x, c)))],$$

Domain Classification Loss:

$$\mathcal{L}^r_{cls} = \mathbb{E}_{x,c'}[-\log D_{cls}(c'|x)],$$

Reconstruction Loss:

$$\mathcal{L}_{rec} = \mathbb{E}_{x,c,c'}[||x - G(G(x, c), c')||_1],$$