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

Jun-Yan Zhu et.al. ICCV 2017

Related Work

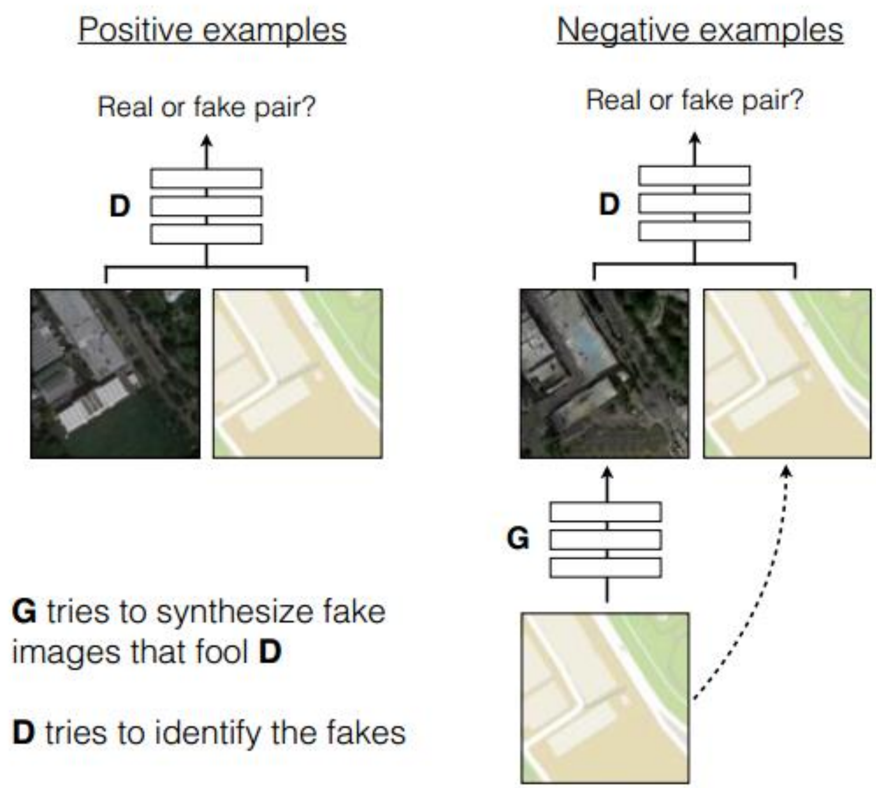


Image-to-Image Translation with Conditional Adversarial Networks. Phillip Isola, et.al. CVPR 2016

Unpaired Image-to-Image Translation

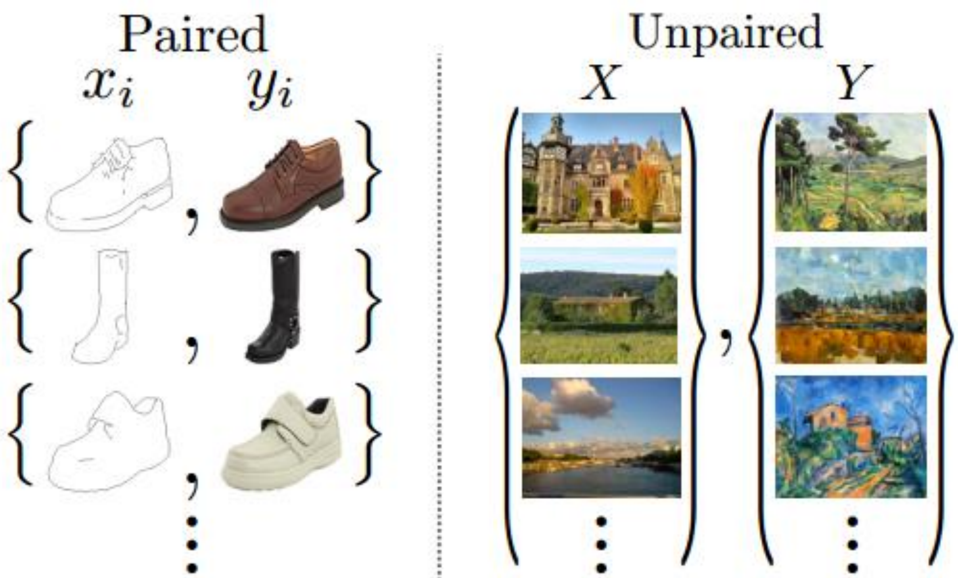


Figure 2: *Paired* training data (left) consists of training examples $\{x_i, y_i\}_{i=1}^N$, where the y_i that corresponds to each x_i is given [18]. We instead consider *unpaired* training data (right), consisting of a source set $\{x_i\}_{i=1}^N \in X$ and a target set $\{y_j\}_{j=1}^M \in Y$, with no information provided as to which x_i matches which y_j.

Method

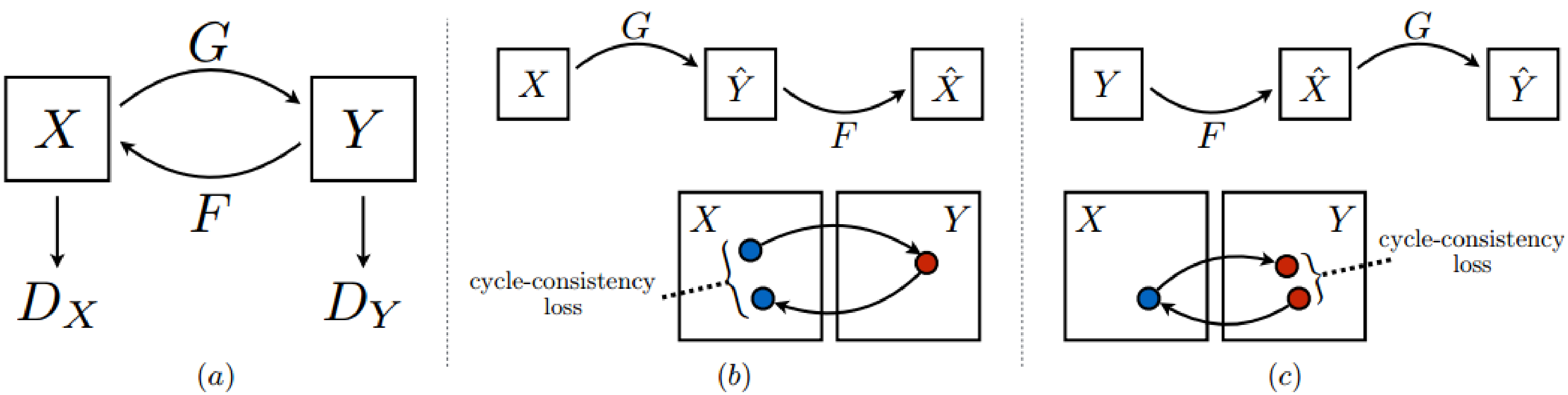


Figure 3: (a) Our model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X, F, and X. To further regularize the mappings, we introduce two "cycle consistency losses" that capture the intuition that if we translate from one domain to the other and back again we should arrive where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

Formulation

- Goal: to learn mapping functions between two domains X and Y given training samples $\{x_i\}_{i=1}^{N} \in X$ and $\{y_j\}_{j=1}^{M} \in Y$

- Two mapping functions: $G: X \rightarrow Y,\ F: Y \rightarrow X$

- Two adversarial discriminators: D_X and D_Y
  D_X aims to distinguish between images $\{x\}$ and translated images $\{F(y)\}$
  D_Y aims to discriminate between $\{y\}$ and $\{G(x)\}$

Adversarial Loss

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

G tries to generate images $G(x)$ that look similar to images from domain Y, while D_Y aims to distinguish between translated samples $G(x)$ and real samples y.

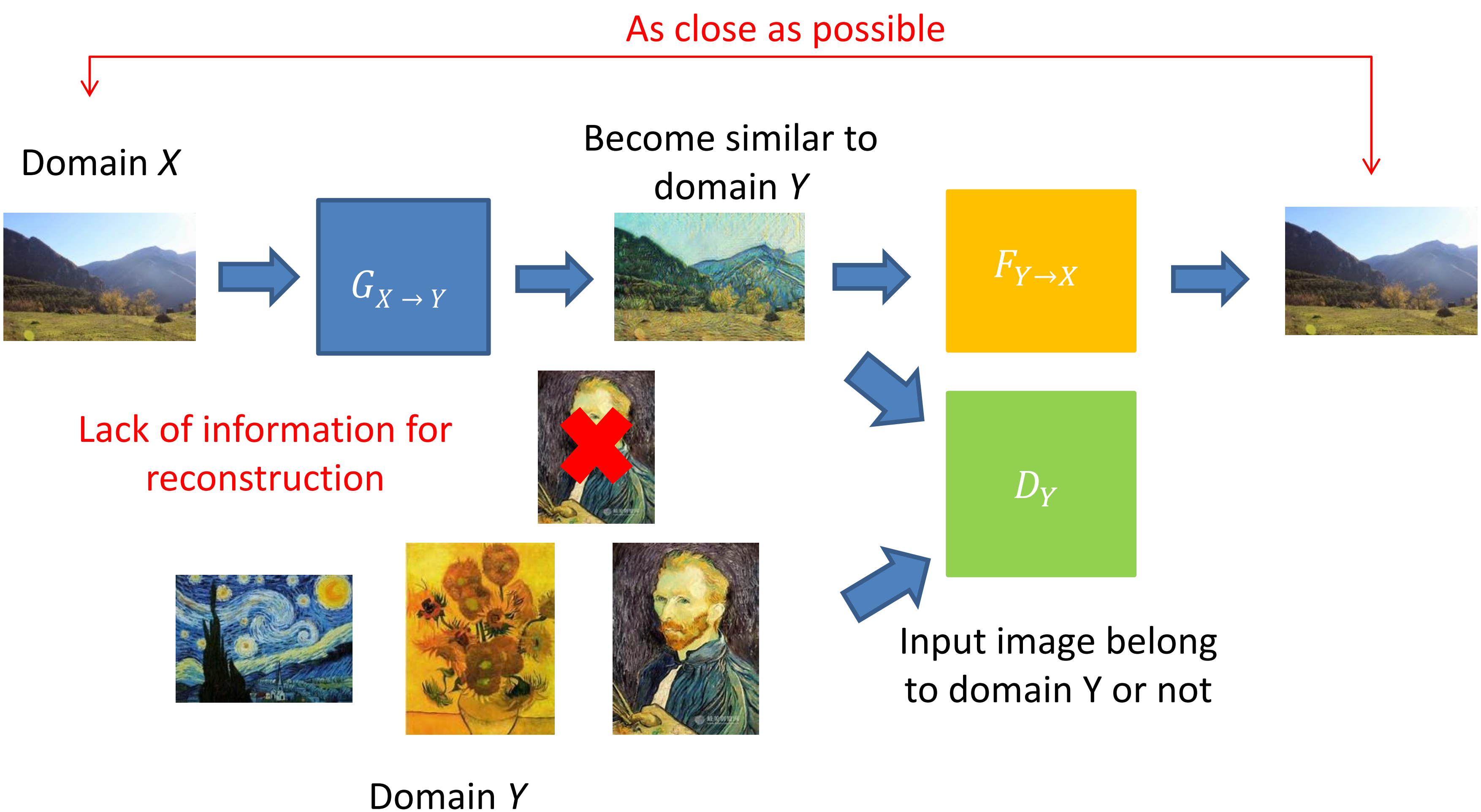
G tries to minimize this objective against an adversarial D that tries to maximize it.

$$G^* = \arg\min_G \max_{D_Y} \mathcal{L}_{GAN}(G, D_Y, X, Y)$$

Similarly

$$F^* = \arg\min_F \max_{D_X} \mathcal{L}_{GAN}(F, D_X, Y, X)$$

Cycle Consistency Loss



Cycle Consistency Loss

For each image x from domain X, the image translation cycle should be able to bring x back to the original image

$x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ *forward cycle consistency*

$y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ *backward cycle consistency*

Cycle consistency loss:

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

Full Objective

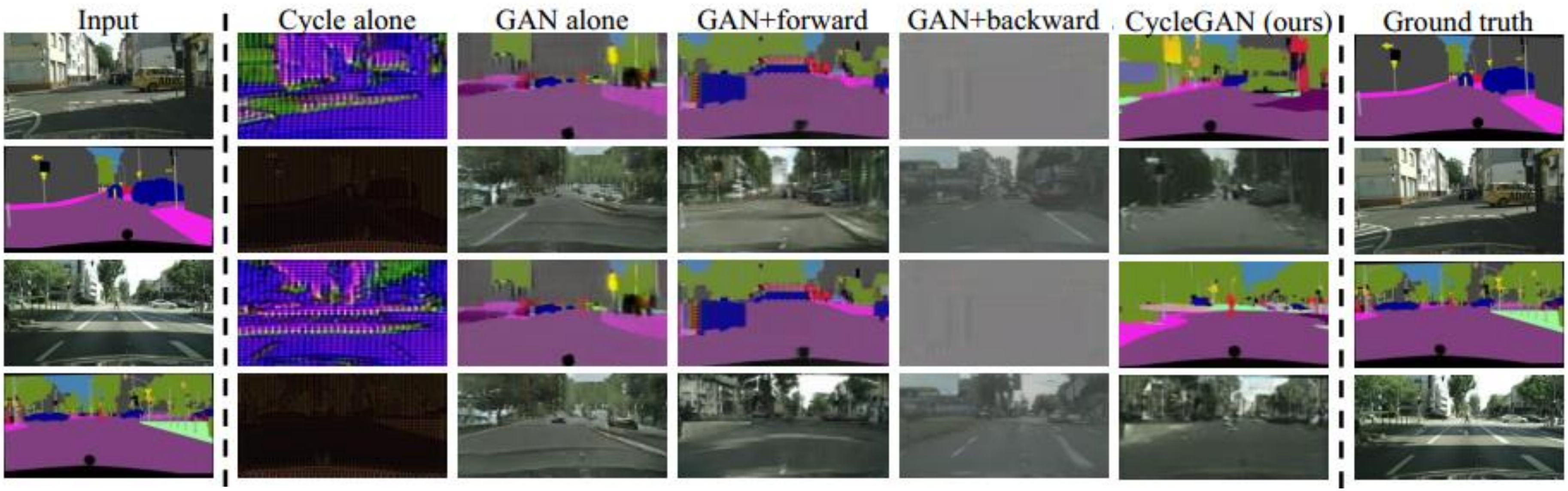
$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda \mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

where λ controls the relative importance of the two objectives

We aim to solve:

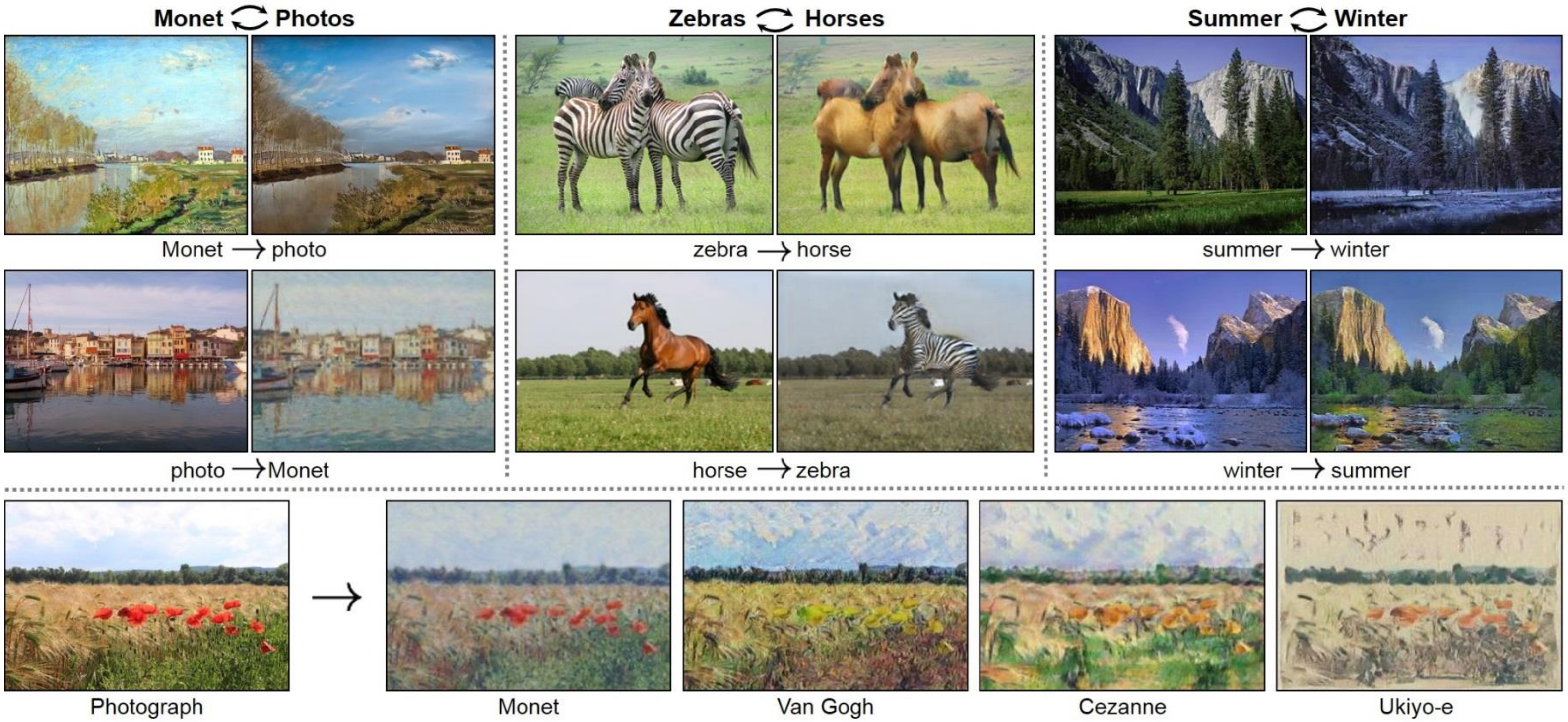
$$G^*, F^* = \arg \min_{F,G} \max_{D_x, D_Y} \mathcal{L}(G, F, D_X, D_Y).$$

Analysis of the loss function



- Both *Cycle alone* and *GAN + backward* fail to produce images similar to the target domain.
- *GAN alone* and *GAN + forward* suffer from mode collapse, producing identical label maps regardless of the input photo.

Experiment



Monet ⇄ Photos
Zebras ⇄ Horses
Summer ⇄ Winter
Monet → photo
zebra → horse
summer → winter
photo → Monet
horse → zebra
winter → summer
Photograph
→
Monet
Van Gogh
Cezanne
Ukiyo-e

Experiment

Turning a horse video as a zebra video.

Failure Case

Our model does not work well when a test image looks unusual compared to training images, as shown in the above figure.