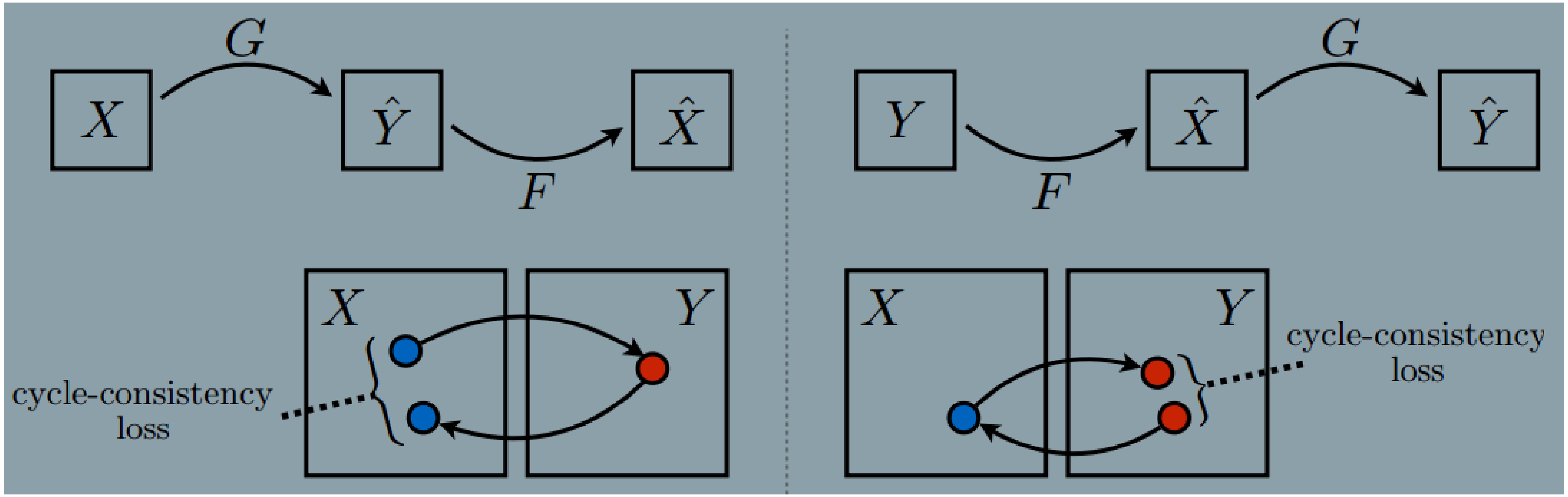
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

published:2017-3

authors:Berkeley AI Research Laboratory(BAIR)

The main contribution of this paper is the authr **present an approach for learning to translate an image from a source domain X to a target domain Y in the abesnce of paired examples.**. The goal of this paper is to learn a mapping $G : X \rightarrow Y$ such that the distribution of images from $G(X)$ is indistinguishable from the distribution Y using an adversarial loss. The main idea of this paper is to capture the special characteristics of one image collection and figuring out how these characteristics could be transloated into the other images collection.

The produce is illustrate in the following figure:



> Formulation
> Our goal is to learn mapping functions between two domains X and Y givens training smaples $x_i \in X$ and $y_j \in Y$. As illustrated in above Figure, our model includes two mapping $G: X \rightarrow Y$ and $F: Y \rightarrow X$. In addition, we introduce two adversarial discriminators D_x and D_y, where D_x aims to distinguish between images {x} and translated images {F(y)}; in the same way, D_y aims to discriminate between {y} and {G(x)}.

Loss Function

1. Adversarial Loss

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[\log D_Y(y)] + E_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))]$$

2. Cycle Consistency Loss

$$L_{cyc}(G,F) = E_{x\sim p_{data}(x)}[\| F(G(x)) - x \|_1] + E_{y\sim p_{data}(y)}[\| G(F(y)) - y \|_1]$$

3. Full Objective

$$L(G,F,D_x,D_y) = L_{GAN}(G,D_Y,X,Y) + L_{GAN}(F,D_X,Y,X) + \lambda L_{cyc}(G,F)$$

4. Target

$$G^*, F^* = arg \min_{F,G} \max_{D_x,D_y} L(G,F,D_x,D_y)$$

Network Architecture

> **G network** This network contains two stride-2 convolutions, several residual blocks, and two fractionally strided convolutions with stride 1/2. We use 6 blocks for 128 × 128 images, and 9 blocks for 256×256 and higherresolution training images.
> **D network** we use 70x70 PatchGANs, which try to classify whether 70*70 voerlapping imagepatches are real or fake.

Training details

1. For L_{GAN}, we replace the negative log likelihood objective be a least square loss.

$$L_{LSGAN}(G,D_Y,X,Y) = E_{y\sim p_{data}(y)}[(D_Y(y)-1)^2] + E_{x\sim p_{data}(x)}[D_Y(G(x))^2]$$

2. To reduce model oscillation, We update discriminators D_x and D_y using a history of generated images rather than

the ones produced by the latest generative networks. We keep an image buffer that stores the 50 previously generated images.