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

Ruyue Han

September 14, 2018

1. Introduction

In this paper,author presented a new model called cycleGan for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples.Figure. 1 shows the cycleGan model.CycleGan 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 and F. To further regularize the mappings, author introduces two cycle consistency losses that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started: (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$.

2. Formulation

2.1. Adversarial Loss

Author apply adversarial losses to both mapping functions. For the mapping function G : $X \rightarrow Y$ and its discriminator D_Y , expressing the objective as:

$$\begin{aligned} L_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \backsim p_{data}(y)} [\log D_Y(y)] \\ + \mathbb{E}_{x \backsim p_{data}(x)} [\log (1 - D_Y(G(x)))] \end{aligned} \tag{1}$$

where 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 aims to minimize this objective against an adversary D that tries to maximize it, i.e. $\min_G \max_{D_Y} L_{GAN}(G, D_Y, X, Y)$ .There are also a similar adversarial loss for the mapping function F : $Y \rightarrow X$ and its discriminator D_X as well: i.e.$\min_F \max_{D_X} L_{GAN}(F, D_X, X, Y)$.

2.2. Cycle Consistency Loss

Adversarial losses alone cannot guarantee that the learned function can map an individual input x_i to a desired output y_i . To further reduce the space of possible mapping functions, so author argue that the learned mappingAs shown in Figure. 1(b), for each image x from domain X, the image translation cycle should be able to bring x back to the original image, i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$.They call this *cycle consistency*.Similarly, as illustrated in Figure. 1(c), for each image y from domain Y , G and F should also satisfy backward cycle consistency: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$.Author incentivize this behavior using a cycle consistency loss:

$$\begin{aligned} L_{cyc}(G, F) = \mathbb{E}_{y \backsim p_{data}(y)} [||G(F(y)) - y||_1] \\ + \mathbb{E}_{x \backsim p_{data}(x)} [||F(G(x)) - x||_1] \end{aligned} \tag{2}$$

2.3. Full Objective

$$\begin{aligned} L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) \\ + L_{GAN}(F, D_X, Y, X) \\ + \lambda L_{cyc}(G, F) \end{aligned} \tag{3}$$

also aiming to solve:

$$G^*, F^* = arg \min_{G,F} \max_{D_X, D_Y} L(G, F, D_X, D_Y) \tag{4}$$

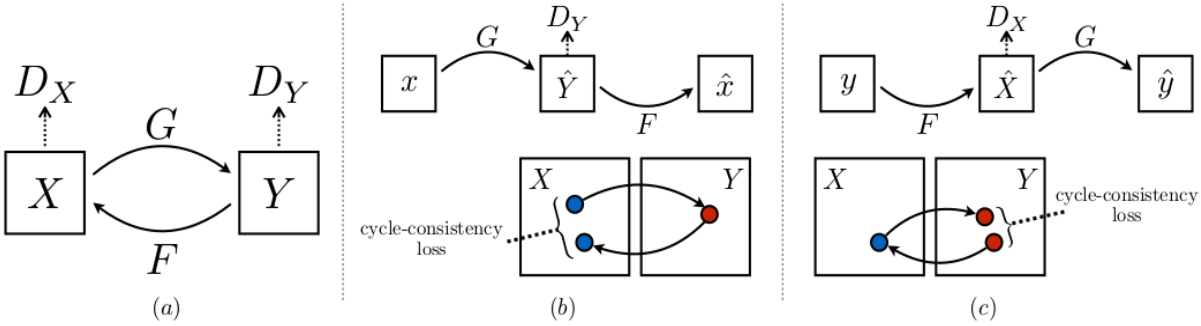

Figure 1. CycleGan model