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

https://arxiv.org/pdf/1703.10593.pdf

Paper Summary By:

Ankit Dhankhar

Tuesday 1st January, 2019

INTRODUCTION

- Paper introduces image-to-image translation from a source domain X to target domain Y in absence of paired data.
- They learned a mapping $G : X \rightarrow Y$ such that the distribution image from G(X) is indistinguishable from Y using an adversarial loss preserving the **content** of X.

SUMMARY

- Model have two adversarial discriminators D_X and D_Y, where D_X aims to distinguish between images x and translated image $F(y)$; in same way, D_Y aims to discriminate between y and $G(x)$. So objective contain two terms:
  - **Adversarial Loss**

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

    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 function against adversary D that tries to maximize it, i.e. $\min_G \max_{D_Y} \mathbb{L}_{GAN}(G, D_Y, X, Y)$. Similarly for mapping function $F : Y \rightarrow X$ and its discrimator as well: i.e. $\min_F \max_{D_X} \mathbb{L}_{GAN}(G, D_X, Y, X)$.
  - **Cycle Consistency Loss**
    Learned mapping G and F can produce outputs identically distributed as target

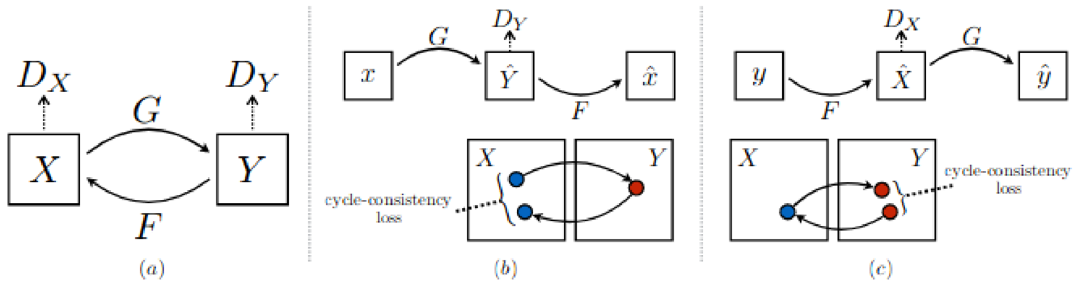

Figure 1: (a) Model contains two mapping function $G: X \rightarrow Y$ and $F: Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguisable from domain Y, and vice versa for D_X and F. To further regularize mappings, they introduced two consistency losses to capture intuitions that if they translate from one domain to other and back again then one should arrive at where the started. (b) forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ and (c) backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

domains Y and X respectively.. Authors have shown that without cycle consistency loss model don't have any incentive to map a particular image from domain X to Y or vice-versa. Instead it can map to random permutation in target domain thus making it a highly under-constrained without cycle consistency loss.

$$\mathbb{L}(G,F) = \mathbb{E}_{x \sim p_{data}}(x)[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}}(y)[\|G(F(y)) - y\|_1] \tag{2}$$

They tried replacing $\mathbb{L}_1$ norm in loss with an adversarial loss between $F(G(x))$ and x, and between $G(F(y))$ and y, but between $G(F(y))$ and y, but did not observe improved performance.

- Thus there Cycle-GAN can be summarized with full objective:

$$\mathbb{L}(G,F,D_X,D_Y) = \mathbb{L}_{GAN}(G,D_Y,X,Y) + \mathbb{L}_{GAN}(F,D_X,Y,X) + \lambda \mathbb{L}_{cyc}(G,F) \tag{3}$$

where λ controls the relative importance of the objective. With an aim to solve:

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