Unpaired Image-to-Image Translation using Cycle-Generative Adversial Networks and application on a human image dataset

Mohammed Junaid Anwar Qader 21CSB0B36
National Institute of Technology, Warangal
Computer Science and Engineering
Warangal, India
mj21csb0b36@student.nitw.ac.in

***Abstract*—Generative Adversarial Networks has been used in many fields now, and it is particularly essential in the field of computer vision. In respect of image-to-image translation, CycleGAN is an important part.**

***Index Terms*—CycleGAN, style transfer**

I. Introduction

CycleGAN, a deep learning model, has gained substantial attention for its proficiency in unsupervised image-to-image translation. This program implements a CycleGAN architecture, a cornerstone in Generative Adversarial Networks (GANs), facilitating image transformation between two distinct domains. Specifically designed for unpaired image translation, CycleGAN enables the seamless conversion of images from one domain to another without necessitating direct correspondences between datasets.

This program is structured to demonstrate the potential of CycleGAN in transforming images between diverse domains, elucidating its pivotal role in various applications such as style transfer, artistic rendering, and domain adaptation. Through its implementation, this model elucidates the underlying mechanisms of domain transfer and the intricate interplay between generators and discriminators in achieving compelling image translations.

By diving into the workings of CycleGAN, this program aims to show how adaptable and useful it is in real-world situations. It highlights CycleGAN's potential to change how we transform images from one type to another, offering new possibilities in fields like computer vision and image processing. This exploration is set to bring fresh ideas and advancements in how we manipulate and understand images.

II. Related Work

A. Style Transfer

Style transfer is one way of image-to-image translations. It integrates the content of one image with the basic style of the other. For example, a photograph taken from the real world can be transformed to an anime style illustration, keeping the characters and layout of the original image, which means image in domain X is mapped to domain Y in mathematics.

B. CycleGAN

CycleGAN Many sorts of generative adversarial networks are able to realize the image-to-image translation with paired image examples. However, under many circumstances, a large set of paired images are not available easily.

CycleGAN has two generators and discriminators in terms of its overall architecture. As is shown in Fig.1, what the generators do is transforming the source image in domain X (or Y) to the target image in domain Y (or X). Specifically, we train a mapping G : $X \to Y$ (or F : $Y \to X$)

CycleGAN is a variant of the Generative Adversarial Networks (GANs) designed for unsupervised image-to-image translation. It aims to learn a mapping between two different image domains, X and Y, without requiring paired examples for training. The model consists of two generators, $G : X \to Y$ and $F : Y \to X$, along with two discriminators, D_X and D_Y. The generators aim to transform images from one domain to another (G transforms from X to Y, and F transforms from Y to X), while the discriminators aim to differentiate between the generated images and real images from their respective domains.

The cycle-consistency loss is a crucial component in the CycleGAN architecture and is formulated as follows:

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

This loss enforces the idea that transforming an image from domain X to Y and then back to X should result in an image similar to the original. It encourages the generators G and F to produce consistent translations, maintaining the identity of the input images through the cycle consistency loss function.

III. Implementation

A. Network Architecture

The CycleGAN architecture consists of two generators (`G_XtoY` and `G_YtoX`) and two discriminators (`D_X` and `D_Y`). The generators aim to transform images from domain

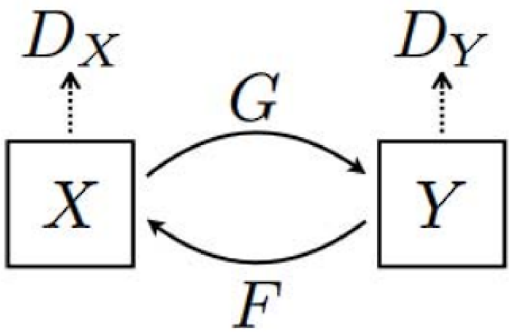

Fig. 1. CycleGAN Architecture

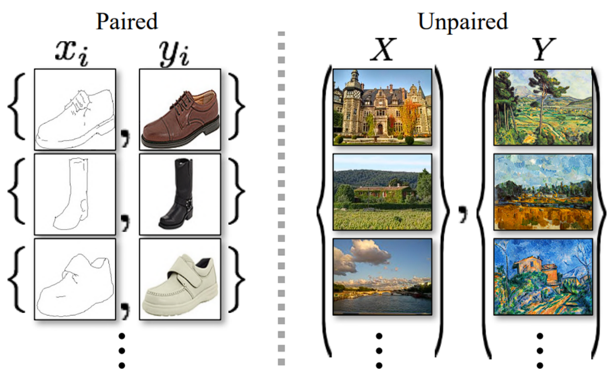

Fig. 2. Paired training data (left) consists of training examples $\{x_i, y_i\}_{i=1}^N$, where the correspondence between x_i and y_i exists [**?**]. We instead consider unpaired training data (right), consisting of a source set $\{x_i\}_{i=1}^N$ ($x_i \in X$) and a target set $\{y_j\}_{j=1}^M$ ($y_j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

X to domain Y and vice versa, while the discriminators aim to distinguish between real and generated images.

During training, the generators are trained to minimize the discrepancy between the generated images and the real images in both domains. Meanwhile, the discriminators are trained to distinguish between real and generated images. The loss functions used for the generators involve adversarial losses (for fooling the discriminators) and cycle consistency losses (to ensure the reconstructed images resemble the originals).

The model iterates through epochs, updating the generators and discriminators in a cycle, aiming to improve the quality of the generated images and the discriminators' ability to distinguish between real and fake images. Over time, the generators become more adept at translating images between the two domains while maintaining their visual fidelity.

B. Notebook

The Notebook[5] is using the kaggle inbuilt interface to run.

C. Dataset

The dataset used in this study is sourced from the 'Selfie2Anime' collection available on Hugging Face Datasets [6]. This dataset offers a diverse range of paired images, consisting of selfies and corresponding anime-style portraits. Each pair within the dataset represents a transformation from a real-world selfie to a stylized anime depiction, facilitating research in image translation, style transfer, and generative modeling. The dataset's comprehensive nature and paired structure make it an ideal resource for training and evaluating image-to-image translation models.

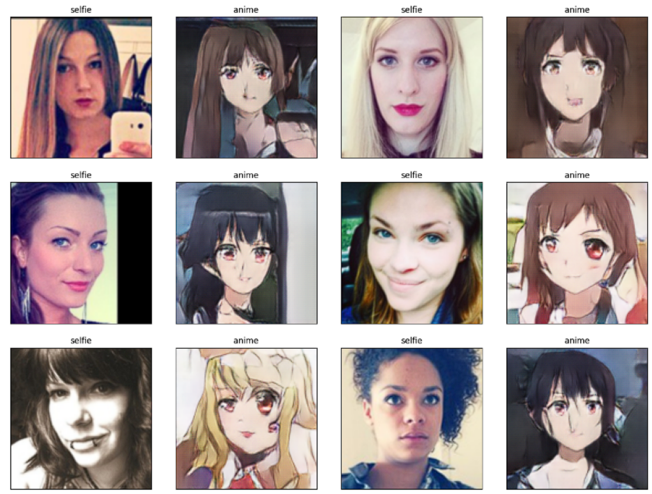

Fig. 3. Output from the CycleGan Implementation

IV. Output

Fig. 3 & 4 showcases a compelling transformation between randomly selected standard selfie images and their corresponding anime-style illustrations. Each pair of images demonstrates a fascinating translation, presenting the original selfie alongside its stylized anime counterpart. This output emphasizes the efficacy of the developed model in seamlessly converting real-world photographs into captivating anime-style renditions. The pairs exhibit diverse transformations, highlighting the versatility of the model in capturing various facial features, expressions, and stylistic nuances characteristic of anime illustrations. This output not only demonstrates the model's proficiency in image translation but also underscores its potential for creative applications in transforming personal photos into vibrant, anime-inspired artworks.

V. Training Losses

During the training process, the model's progress is vividly reflected in the plotted training losses (Fig. 5). The graph illustrates the convergence and divergence of losses for the discriminator networks concerning domains X and Y, represented by the alpha-blended lines of differing colors. The fluctuations in the discriminator losses signify the network's learning dynamics in discerning between real and generated images from both domains. Simultaneously, the losses for the generators, portrayed by another distinct line, demonstrate the optimization journey for creating compelling translations

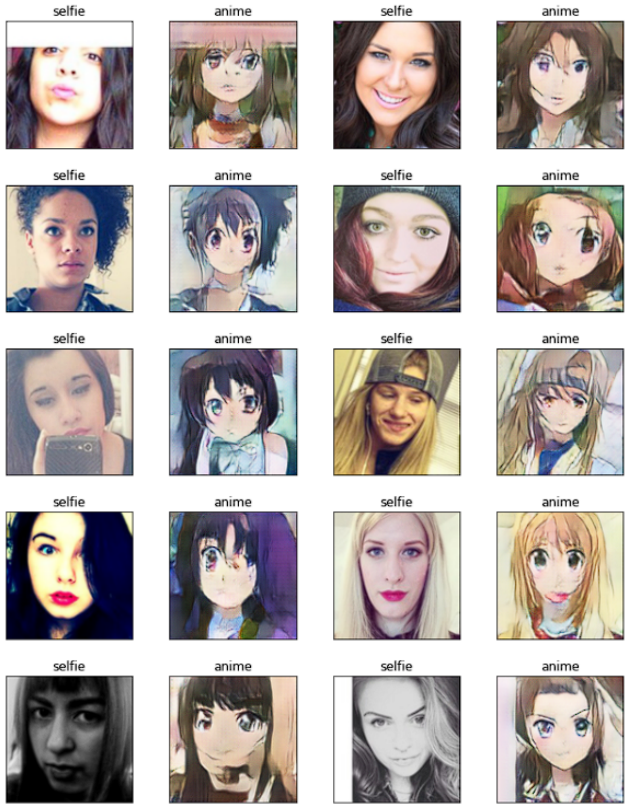

Fig. 4. Output for more images

Fig. 5. The training Losses

between the domains. The graph showcases the interplay between the adversarial networks, with fluctuations indicating the evolution of the model's capability in image translation. The plotted training losses provide valuable insights into the learning process, highlighting the intricate balance achieved between the discriminator and generator networks throughout the training iterations.

ACKNOWLEDGMENT

I extend my sincere gratitude to Professor I. Earnest Paul for the guidance and insightful feedback, which have been pivotal. I am deeply appreciative of the opportunity to work under his guidance.

REFERENCES

[1] Min Jin Chong, David Forsyth "GANs N' Roses: Stable, Controllable, Diverse Image to Image Translation" 11 Jun 2021.
[2] Hao Tang, Hong Liu, Dan Xu, Philip H.S. Torr, Nicu Sebe "AttentionGAN: Unpaired Image-to-Image Translation using Attention-Guided Generative Adversarial Networks" 16 Aug 2021.
[3] Jun-Yan Zhu, Taesung Park, Phillip Isola Alexei A. Efros "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" 24 Aug 2020.
[4] Xulu Wang "Loss functions for Style Transfer with CycleGAN" 2022
[5] notebook : kaggle.com/code/june24/selfie-to-anime/notebook
[6] Dataset : Link to dataset
[7] Junho Kim1, Minjae Kim2, Hyeonwoo Kang2, Kwanghee Lee3 "UGAT-IT: UNSUPERVISED GENERATIVE ATTENTION NETWORKS WITH ADAPTIVE LAYER INSTANCE NORMALIZATION FOR IMAGE-TO-IMAGE TRANSLATION" 8 Apr 2020.