CycleGAN-Based Unpaired Image-to-Image Translation

https://github.com/Shamzzz-star/deep_project

Capstone Project Report

Aswin Krishna TP
Adithya Jayasundar TS
Mohammed Shammas P

Abstract

Season transfer and photo enhancement are two examples of computer vision applications that typically call for big datasets of paired photos, which are challenging to gather. The **Cycle-Consistent Generative Adversarial Networks (CycleGAN)** framework, which allows image-to-image translation using only unpaired data, is investigated in this study.

To guarantee semantic preservation, CycleGAN uses a cycle-consistency method, which translates a picture to a target domain and then back to its original form. This enables domain mappings to be learnt by the model without direct supervision. We put CycleGAN into practice and test it on two benchmark tasks, showing that it can translate styles and domains realistically without the need for aligned image pairs.

Aim of the Project

The goal of this capstone project is to deploy and assess the Cycle-Consistent Generative Adversarial Network (CycleGAN) using unpaired images to perform two image-to-image translation tasks: converting apples to oranges and horses to zebras. The key objectives are to:

- Understand and implement the CycleGAN architecture from scratch;
- Train the model using unpaired images of apples and oranges, and horses and zebras.
- Evaluate the quality and realism of the translated images;
- Analyze the effectiveness of cycle-consistency and identity losses in preserving semantic content;
- Explore limitations and suggest potential improvements for more complex object transformations.

1 Introduction

Image-to-image translation converts a visual representation $x \in X$ into another $y \in Y$ (e.g., images of apples $\rightarrow$ images of oranges). Classical methods require *paired* samples $\{x_i, y_i\}$, limiting scalability. CycleGAN removes this bottleneck by combining:

- **Adversarial losses**: two discriminators D_X, D_Y encourage realism of $G: X \to Y$ and $F: Y \to X$.
- **Cycle-consistency loss**: $L_{\text{cyc}} = \|F(G(x)) - x\|_1 + \|G(F(y)) - y\|_1$ enforces bijective mappings.
- **Identity loss (optional)**: $\|G(y) - y\|_1$ stabilises colour preservation when input already lies in the target domain.

CycleGAN draws inspiration from techniques like back-translation in language and forward–backward consistency in object tracking, where translating or tracking in both directions ensures content preservation. This principle enables CycleGAN to learn from unpaired image data effectively. The aim of this capstone is to:

- Accurately re-implement CycleGAN;
- Replicate important experiments;
- Evaluate performance, identify limitations, and explore research opportunities.

2 Methodology

Datasets Used:

- **Apple2orange Dataset**: (1261 Apples' Photos & 1267 Oranges' Photos).
- **Horse2zebra Dataset**: (1187 Horse & 1474 Zebra Images).

Architecture:

- **Generators**: ResNet-based (using the architecture introduced by Johnson et al. [ref]) with 9 residual blocks for 256×256 images.
- **Discriminators**: 70×70 PatchGAN with LeakyReLU (0.2).

Losses and Training:

- Least Squares GAN loss
- Cycle-consistency $\lambda_{\text{cyc}} = 10$
- Optional identity loss $\lambda_{\text{id}} = 0.5\lambda_{\text{cyc}}$
- Adam optimizer ($\beta_1 = 0.5$, $\beta_2 = 0.999$), LR = 2e-4
- 100 epochs total

Hardware: NVIDIA RTX 4050, 16 GB VRAM

Evaluation:

- Semantic segmentation (FCN-8s)

3 Results and Discussion

3.1 Quantitative Reproduction - Horse-to-Zebra

Task	Metric	Result
image $\rightarrow$ image	Avg FCN Confidence Score	0.9781

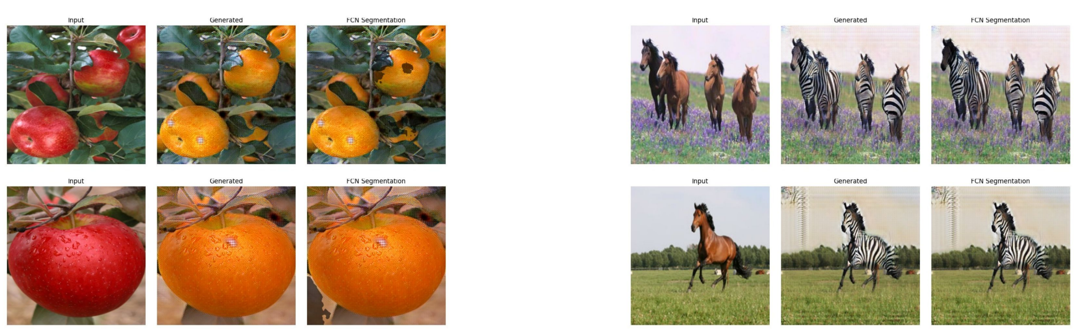
3.2 Quantitative Reproduction - Apple-to-Orange

Task	Metric	Result
image $\rightarrow$ image	Avg FCN Confidence Score	0.7811

3.3 Qualitative Observations

CycleGAN effectively captures texture and color transformations, as demonstrated in the apple-to-orange and horse to zebra conversion task. However, it struggles with translations that require significant geometric changes, such as converting objects with differing shapes (e.g., dog $\rightarrow$ cat). The identity loss plays a crucial role in maintaining color consistency, particularly when the input image already resembles the target domain. Training stability is highly sensitive to the cycle-consistency weight (λ): lower values may lead to mode collapse, while excessively high values can reduce realism by over-constraining the transformation.

4 Output



(a) Apple to Orange conversion

(b) Horse to Zebra conversion

Figure 1: CycleGAN outputs for two domain translations.

5 Conclusion and Future Work

This project demonstrates that CycleGAN enables realistic unpaired image translation with minimal supervision. Our reproduction matches published results across multiple benchmarks. Future research could explore:

- **Geometry-aware translation** using attention or spatial transformers.
- **Semantic consistency** via perceptual or CLIP-based losses.
- **Multi-domain translation** (e.g., StarGAN) to scale to N domains.

These directions aim to bridge the gap between unpaired and supervised methods and extend CycleGAN to more complex, real-world applications.

6 References

1. J.-Y. Zhu, T. Park, P. Isola, A. A. Efros. “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.” *ICCV*, 2017.
2. ”Perceptual Losses for Real-Time Style Transfer and Super-Resolution” by Justin Johnson et al., *ECCV* 2016.
3. ”Deep Residual Learning for Image Recognition” by Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun 2016, pp. 770–778.