IP - CycleGAN Report

Aajay Devaraj - 2021001, Dhvanil Sheth - 2021040

December 2023

1 Introduction

This report presents an in-depth exploration of CycleGAN, a groundbreaking approach in image-to-image translation using unpaired datasets. CycleGAN, standing for Cycle-Consistent Generative Adversarial Networks, has revolutionized the way we approach tasks such as style transfer, photo enhancement, and domain adaptation in the absence of paired training data. This report delves into the technical foundations of CycleGAN, discussing its unique architecture, underlying mathematical principles, and innovative training methodology. Furthermore, we will explore various applications where CycleGAN has shown remarkable results, assess its limitations, and discuss potential future developments in this area of research.

2 Methodology

CycleGAN aims to learn mapping functions between two unpaired domains, X and Y, utilizing two generators, $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and two adversarial discriminators, D_X and D_Y. The discriminators aim to distinguish between actual images and those translated by F and G, respectively.

2.0.1 Adversarial Loss

Applied to both mapping functions, the adversarial loss ensures that images generated by G and F are indistinguishable from actual images in their respective target domains.

For G, the loss is defined as $L_{GAN}(G, D_Y, X, Y)$, where G tries to minimize this objective against D_Y, which aims to maximize it.

2.0.2 Cycle Consistency Loss

This loss ensures that the mappings G and F are cycle-consistent. That means an image x from domain X after being translated to domain Y by G and then back to X by F, should be close to the original image x (forward cycle consistency). The same applies in the reverse direction for domain Y (backward cycle consistency). The cycle consistency loss is $L_{cyc}(G, F)$.

2.0.3 Full Objective

The overall objective combines adversarial and cycle consistency losses with a control parameter λ to balance the importance of these terms. The aim is to optimize G and F to minimize this combined loss.

3 Model Architecture

1. **Mapping Functions**: CycleGAN consists of two generators - $G : X \to Y$ and $F : Y \to X$. G translates images from domain X to Y, and F does the reverse.

2. **Discriminators**: Two discriminators, D_X and D_Y, are used. D_X differentiates between images $\{x\}$ and translated images $\{F(y)\}$, while D_Y differentiates between $\{y\}$ and $\{G(x)\}$.

4 Mathematical Formulation

1. **Adversarial Loss**: An adversarial loss is used for each mapping function (G and F). For G, it's defined as:

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

   This loss encourages G to generate images similar to images from domain Y, fooling the discriminator D_Y.

2. **Cycle Consistency Loss**: To ensure that each image translated from one domain to the other and back again remains consistent, a cycle consistency loss is introduced:

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

   This loss penalizes the model if $G(F(y))$ and $F(G(x))$ deviate from y and x, respectively.

3. **Full Objective**: The full objective combines these losses:

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

   Here, λ is a hyperparameter that controls the relative importance of the cycle consistency loss.

5 Training and Implementation

CycleGAN's training and implementation involve several key steps and techniques to ensure effective learning and stable performance. The model's architecture is designed to handle unpaired image-to-image translation tasks, leveraging adversarial networks and cycle consistency.

5.1 Network Architecture

The work of Johnson et al. inspires the architecture of CycleGAN and includes:

- Generative networks with three convolutional layers, several residual blocks, and fractionally-strided convolutions.
- Instance normalization is utilized for its effectiveness in style transfer tasks.
- Discriminator networks employ 70×70 PatchGANs, which classify whether overlapping image patches are real or fake, allowing the model to work on images of varying sizes in a fully convolutional manner.

5.2 Training Process

The training process involves the following steps:

- **Objective Optimization**: The model is trained to minimize the full objective function, which combines adversarial losses for matching the distribution of generated images to the target domain and cycle consistency losses to ensure the mappings do not contradict each other.
- **Adversarial Training**: For each mapping function $G : X \rightarrow Y$ and $F : Y \rightarrow X$, adversarial training is applied against their respective discriminators D_Y and D_X.
- **Cycle Consistency**: The cycle consistency loss is enforced to ensure that each image is translated from one domain to the other and remains consistent with the original image.

5.3 Stabilization Techniques

To stabilize the training process, the following techniques are employed:

- **Least-Squares Loss**: The negative log-likelihood objective in the GAN loss is replaced with a least-squares loss, which has been found to be more stable during training and generates higher quality results.
- **History of Generated Images**: To reduce model oscillation, a history of generated images is maintained and used to update the discriminators, rather than solely relying on the latest generated images.

5.4 Hyperparameters and Optimization

- The hyperparameter λ is set to control the relative importance of the cycle consistency loss in the full objective.
- The Adam optimizer is used with a batch size of 1, and the networks are trained from scratch.

- The learning rate is initially set to 0.0002 and is then linearly decayed to zero over the course of training.

5.5 Training Data

- The model is trained on various unpaired datasets, including landscape photographs, artistic styles, and seasonal images.

6 Applications and Results

CycleGAN has been successfully applied to various tasks like style transfer, object transfiguration, and season transfer. It demonstrates superior performance compared to several prior methods in unpaired settings.

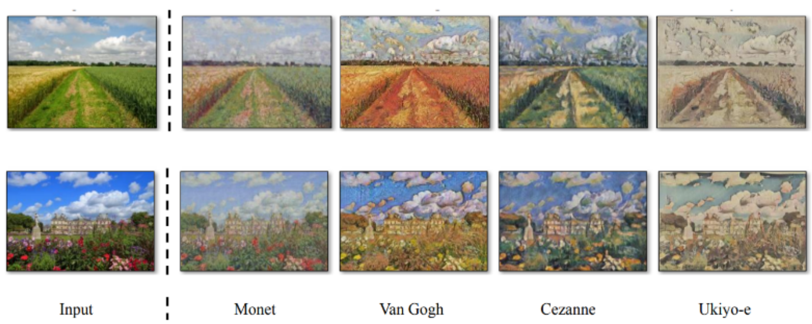

Figure 1: Sample output for CycleGAN

7 Limitations and Future Directions

While effective in many scenarios, CycleGAN struggles with tasks requiring geometric changes or a detailed understanding of object relationships. Future work could explore integrating weak or semi-supervised data to enhance its capabilities.

8 Conclusion

CycleGAN represents a significant advancement in image-to-image translation, particularly in scenarios lacking paired training data. Its innovative combination of adversarial training and cycle consistency enables effective learning of translation functions between diverse image domains.