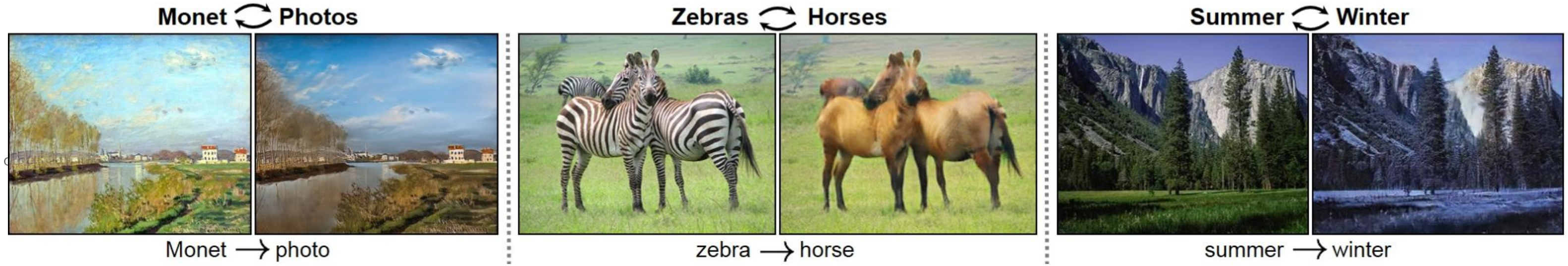
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks



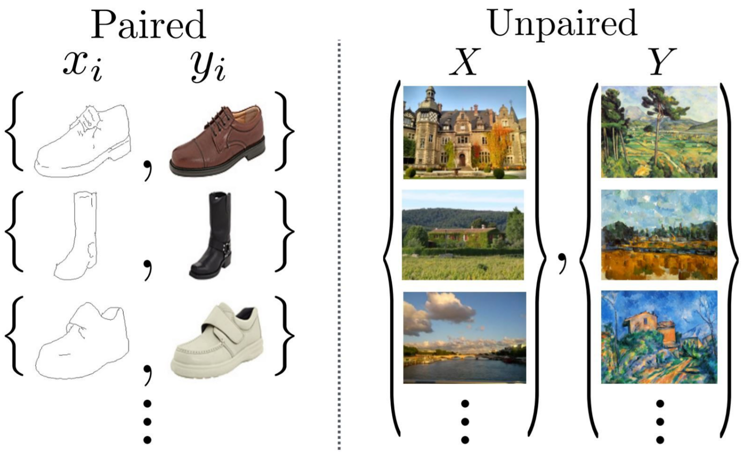
Arushi Arora: aa10350@nyu.edu

Chandana Thimmalapura Jagadeeshaiah: ct3002@nyu.edu

Pallabi Chandra: pc3131@nyu.edu

Introduction

- Paired Images are expensive to annotate for image-to-image translation.



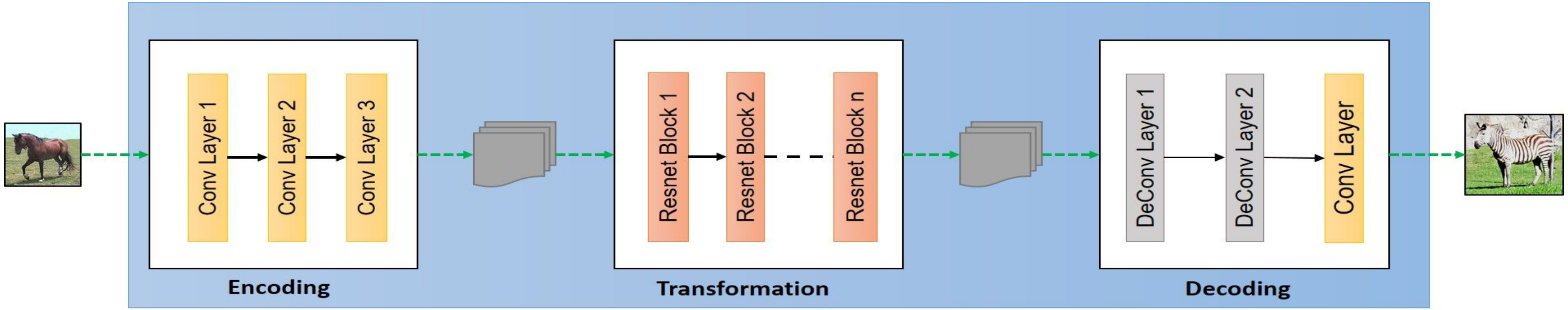
- Our goal is to learn a mapping $G : X \rightarrow Y$ such that the distribution of images from $G(X)$ is indistinguishable from the distribution Y using an adversarial loss.

- Because this mapping is highly under-constrained, we couple it with an inverse mapping $F : Y \rightarrow X$ and introduce a cycle consistency loss to enforce $F(G(X)) \approx X$ (and vice versa).

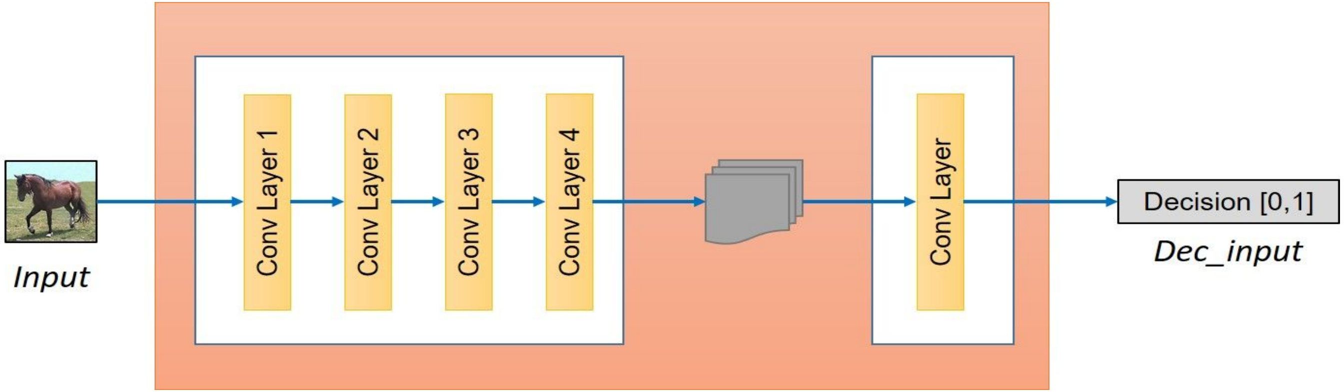
- We then use the method to demonstrate the results on various applications such as style transfer, object transfiguration, season transfer, photo enhancement etc.

Network Architecture

- Generator



- Discriminator



Loss Function

- **Adversarial Loss:**

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

where, G denotes the mapping X → Y and D_y is the discriminator for G(x) and y.

$$L_{GAN}(F, D_X, Y, X) = E_{x \sim p_{data}(x)}[log D_X(x)] + E_{y \sim p_{data}(y)}[log(1 - D_X(F(y))]$$

where, F denotes the mapping Y→ X and D_y is the discriminator for F(y) and x.

- **Cycle Consistency Loss:**

Forward Cycle Consistency: x → G(x) → F (G(x)) ≈ x

Backward Cycle Consistency: y → F (y) → G(F (y)) ≈ y

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G, F) &= \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] \\ &+ \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].\end{aligned}$$

- **Final Loss Function:**

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda \mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

where λ controls the relative importance of the two objectives.

- **Aim to solve:**

$$G^*, F^* = \arg \min_{G,F} \max_{D_x, D_Y} \mathcal{L}(G, F, D_X, D_Y).$$

Experiments

- Experiments on Datasets
  - Summer2Winter Yosemite (1540 Summer Photos & 1200 Winter Photos)
  - Monet2Photo (1193 Monet Paintings & 7038 Natural Photos)
  - Cityscapes (5000 fine annotated images and 20000 coarse annotated images)
  - Facades (506 Building Facades & corresponding Segmentations)

- Hyperparameter tuning on Summer2Winter Yosemite dataset
  - Number of Generator/Discriminator filters
  - Number of ResNetBlocks in Generator
  - **Learning rate**
  - Batch Size
  - **Lambda values**
  - Optimizer
  - **Beta values for the Adam optimizers.**

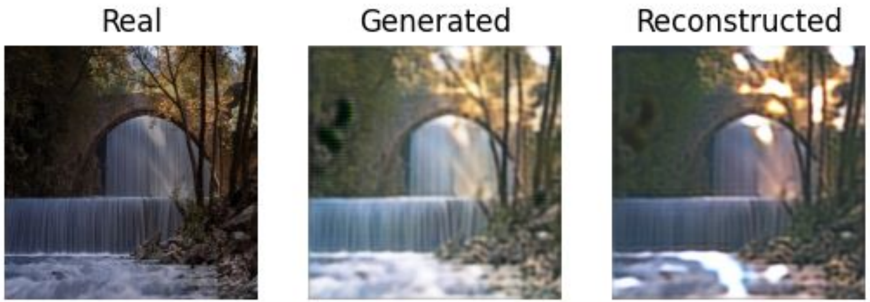
Best Hyperparameters from the experiments:

Hyperparameter	**Value**
Lambda	10
Batch Size	1
Optimizer	Adam
Learning Rate	0.0002
Beta1	0.5
Beta2	0.999

Results

Results on different applications:

Style transfer Monet's paintings (Real) to a photographic style (Generated)



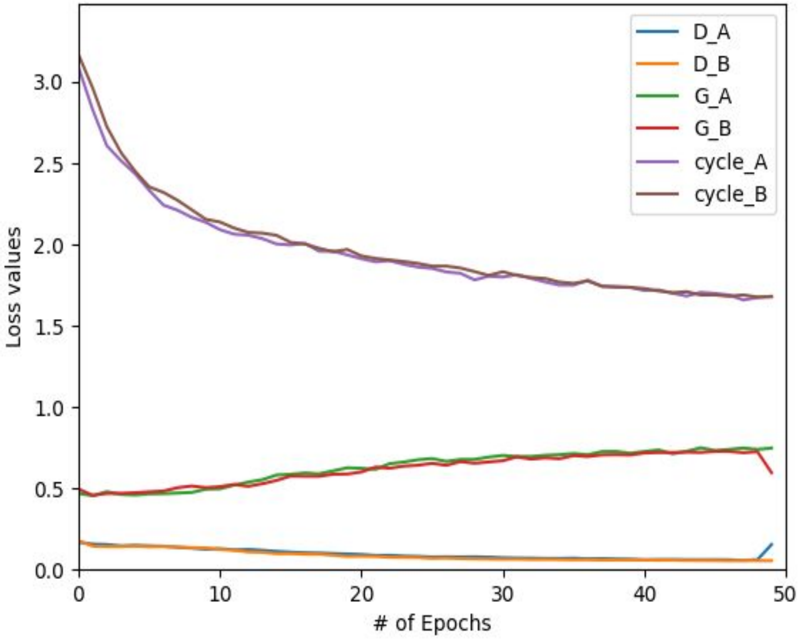
Style transfer from digital images (Real) to a Monet's painting style (Generated)

Season transfer from Summer images of Yosemite to Winter

Season transfer from Winter images of Yosemite to Summer

Transfer from label images to a Photographs (generated)

Loss Plot for Number of Epochs on Summer2Winter Yosmite dataset trained for 50 epochs:



Performance Evaluation

- Evaluation of our generative model is difficult since the image translation is carried out on unpaired images with no pairwise supervision.
- We use two methods for evaluation of models- FCN score on CityScapes dataset (label2photo) and Inception Score on Facades dataset (label2photo).
- FCN score allows us to calculate Per pixel accuracy, Per class accuracy and Class IOU for the semantic segmentation results.
- Inception Score gives us an idea about the quality and diversity of images generated.

Loss	CycleGAN	pix2pix
Per-pixel acc.	0.48	0.71
Per-class acc.	0.16	0.25
Class IOU	0.11	0.18

Table 4: FCN model score summary

Model	Score (Mean, Variance)
CycleGAN	(1.1008, 0.1161)
Pix2Pix	(1.3475, 0.2886)

Table 5: Inception Score Summary

Challenges

- **High Computational Demands**

- **Mode Collapse while training:** A GAN failure in which only a small variety of outputs can be generated. Remedy is to build a diverse dataset with complex patterns in data.

- **Model Evaluation:** Absence of Ground Truth creates challenge to evaluate model performance. Inception Score and FCN score both have certain limitations. FID (Frechet Inception Distance) may be considered to evaluate performance.

Future Scope

- More recently, several models have been released to tackle the problem of unpaired image-to-image translation.

- These include- **UNIT** (Unsupervised Image-to-Image Translation Networks), **MUNIT** (Multimodal Unsupervised Image-to-Image Translation), and **DRIT** (Disentangled Representation for Image-to-Image Translation).

- More work is to be done to evaluate the performance of these models in comparison to CycleGAN and consider each one's strengths and weaknesses.

Thank You!