ART STYLE TRANSFER USING CYCLE GAN

Presentation by
Achintya Gupta, Kannan Rustagi

INTRODUCTION

- Art style transfer is a technique where the style of a reference image (like a painting) is applied to a content image (like a photograph), creating an output that combines the content of the photograph with the artistic style of the reference image.

- The traditional methods depended upon heuristics and handmade rules while deep learning approach harnesses the capabilities of neural networks to autonomously learn complex patterns.

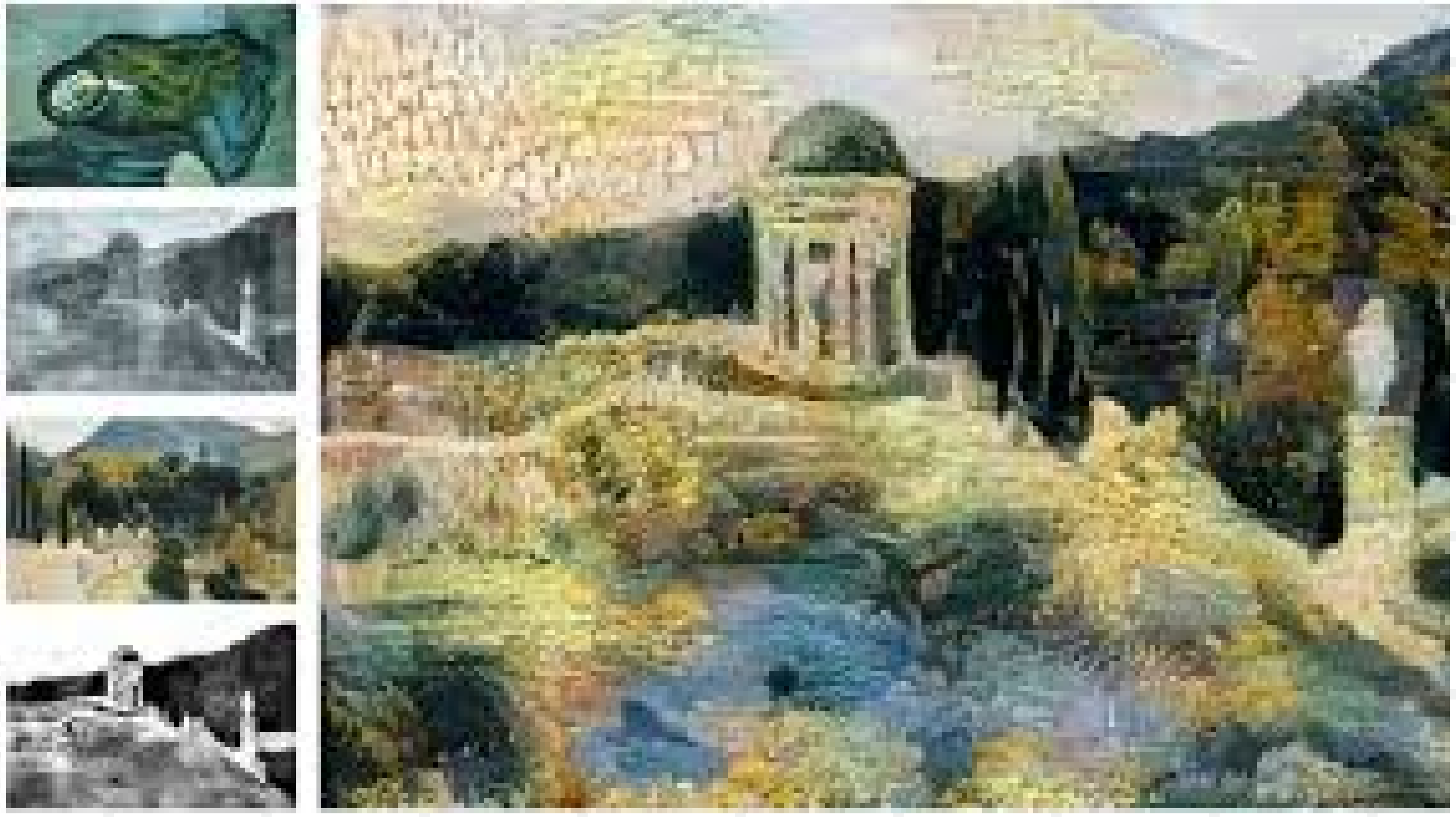
MOTIVATION

- **Overcoming Dependency on Paired Datasets: Traditional style transfer methods necessitate paired datasets, constraining the range of applicable artistic styles. CycleGAN offers a solution by enabling unsupervised learning, thus eliminating the requirement for paired examples and expanding the scope of accessible artistic transformations**
- **Flexibility and Adaptability to Varied Input: The utilization of CycleGAN in neural art style transfer provides flexibility and adaptability to varied input images, enabling the transformation of content across a spectrum of styles without requiring explicit alignment or preprocessing, thus streamlining the creative process.**

APPLICATION/USE CASES

1. **Artistic Exploration and Creativity: Empowering artists to seamlessly blend diverse artistic styles, fostering a new era of creative exploration.**
2. **Content Adaptation for Visual Media: Offering a powerful tool for adapting the visual style of content in filmmaking, advertising, and visual media.**
3. **Virtual Reality and Gaming: Dynamically altering the visual style of environments and characters in virtual reality (VR) and gaming.**
4. **Fashion and Design: Enabling experimentation in fashion and interior design by merging vintage elegance with contemporary trends.**
5. **Preservation and Restoration of Art: Aiding in the preservation and restoration of artworks, providing a digital tool for art historians and conservators.**

Style Transfer in the fashion industry



Neural style transfer reconstructs unseen Picasso painting

BRIEF LITERATURE

1. **"Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", Authors:** Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros
2. **"Stable and Controllable Neural Texture Synthesis and Style Transfer Using Histogram Losses", Authors:** Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, Timo Aila
3. **"Image Style Transfer Using Convolutional Neural Networks"** Leon A. Gatys, Alexander S. Ecker, Matthias Bethge

OBJECTIVE

- Image-to-image translation is a class of vision and graphics problems where the goal is to learn the mapping between an input image and an output image using a training set of aligned image pairs. However, for many tasks, paired training data will not be available.
- Use of CycleGAN is an approach for learning to translate an image from a source domain X to a target domain Y in the absence of paired examples. Our goal is to learn a mapping G:X→Y such that the distribution of images from G(X) is indistinguishable from the distribution Y using an adversarial loss.

MODEL ARCHITECTURE

- Discriminator
  - The adversarial discriminators DX and DY in our model are CNNs that see an image and try to classify it as real of fake. Real images are classified with an output close to 1 while fake classifications have an output close to 0. DY encourages and classifies the transformation from domain X to Y (making photos into monet pictures). DX represents transformation classification from domain Y to X.

MODEL ARCHITECTURE

- Discriminator Architecture (D_X):
  - Comprises five convolutional layers, each with increasing output channels (64, 128, 256, 512, 512), utilizing a kernel size of (4, 4), a stride of (2, 2), and padding of (1, 1).
  - Applies instance normalization after the first three convolutional layers and batch normalization after the fourth convolutional layer.
  - Concludes with a convolutional layer with 1 output channel, using a kernel size of (4, 4), a stride of (1, 1), and padding of (1, 1).
- Discriminator Architecture (D_Y):
  - Shares the same architecture as D_X but operates on images in the Y domain.

MODEL ARCHITECTURE

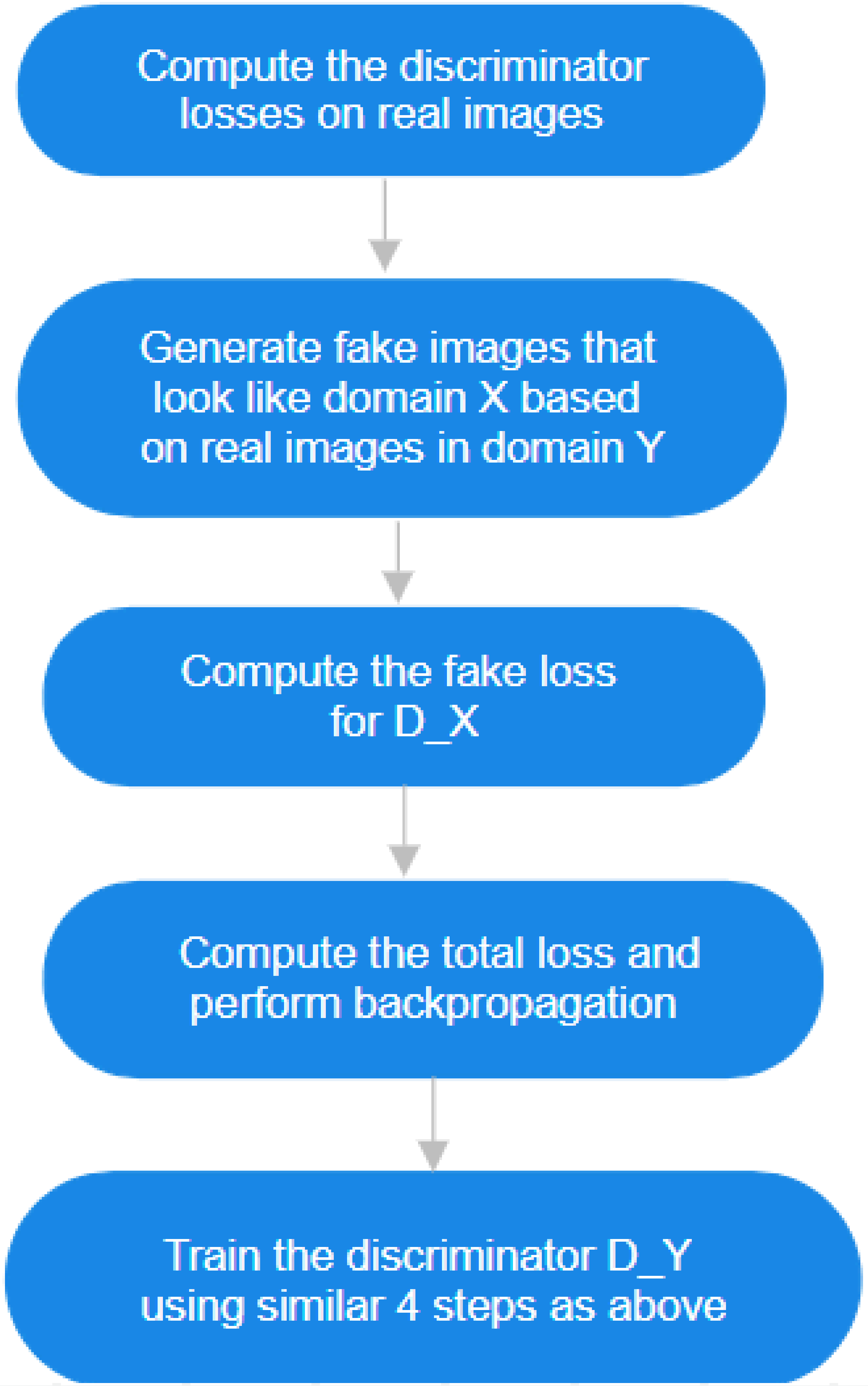
- Generator
  - We have two generators (G_XtoY and G_YtoX). They each consist of
    - Encoder (compressing the image into a smaller feature representation),
    - Residual blocks (connecting the output of one layer with the input of an earlier layer) and
    - Decoder (turning compressed representation into a transformed image).

MODEL ARCHITECTURE

- Generator Architecture (G_XtoY):
  - Starts with a convolutional layer with 3 input channels and 64 output channels, using a kernel size of (4, 4), a stride of (2, 2), and padding of (1, 1).
  - Utilizes three convolutional layers with increasing output channels (64, 128, 256), each followed by instance normalization.
  - Employs six residual blocks, each consisting of two convolutional layers with 256 output channels, using a kernel size of (3, 3) and padding of (1, 1), followed by instance normalization.
  - Applies three deconvolutional layers with decreasing output channels (256, 128, 64), each followed by instance normalization and using a kernel size of (4, 4), a stride of (2, 2), and padding of (1, 1).
  - Concludes with a deconvolutional layer with 3 output channels, using a kernel size of (4, 4), a stride of (2, 2), and padding of (1, 1).
- Generator Architecture (G_YtoX):
  - Follows the same architecture as G_XtoY but operates in the reverse direction.

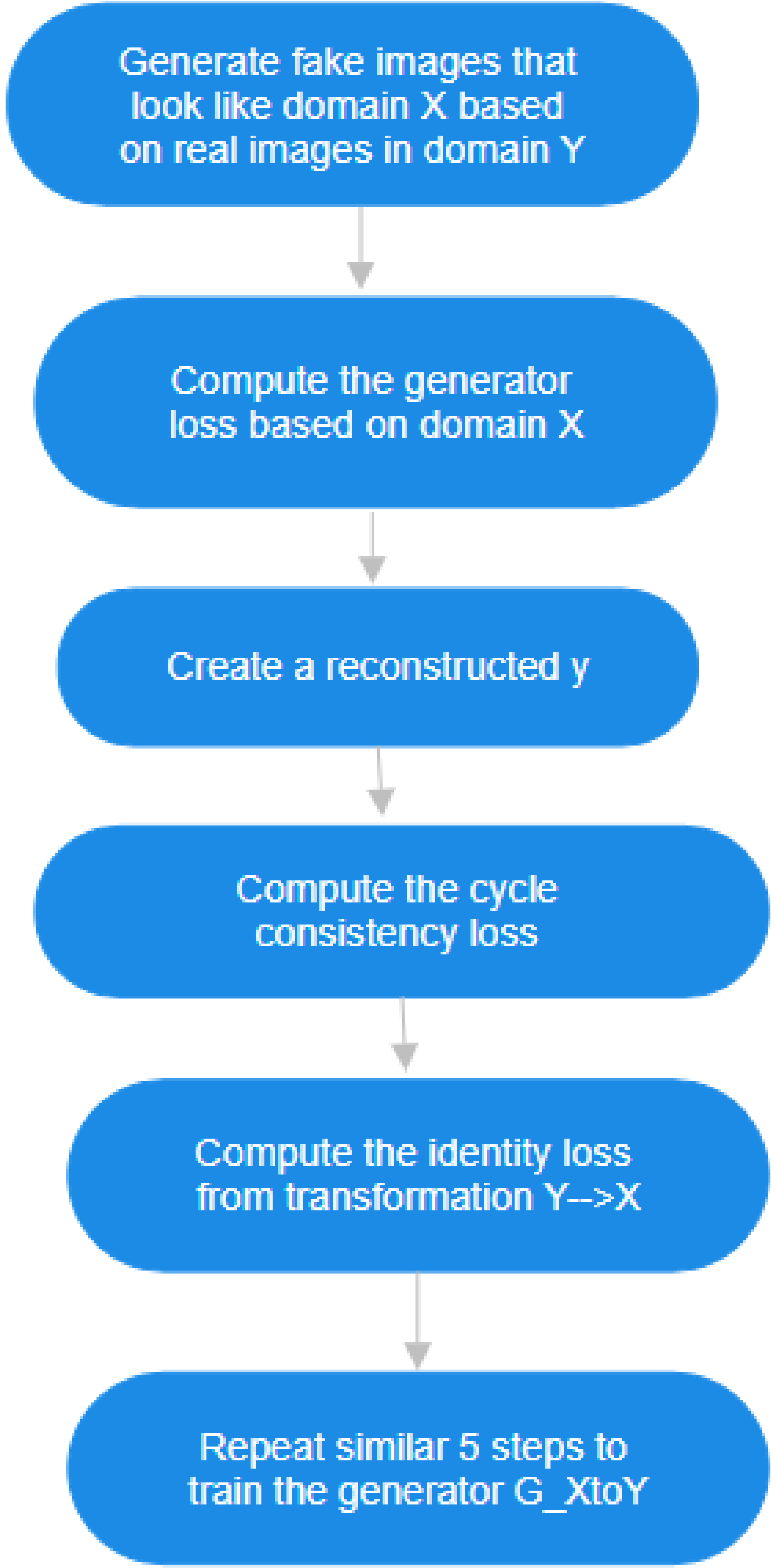
METHODOLOGY

Training both Discriminators



METHODOLOGY

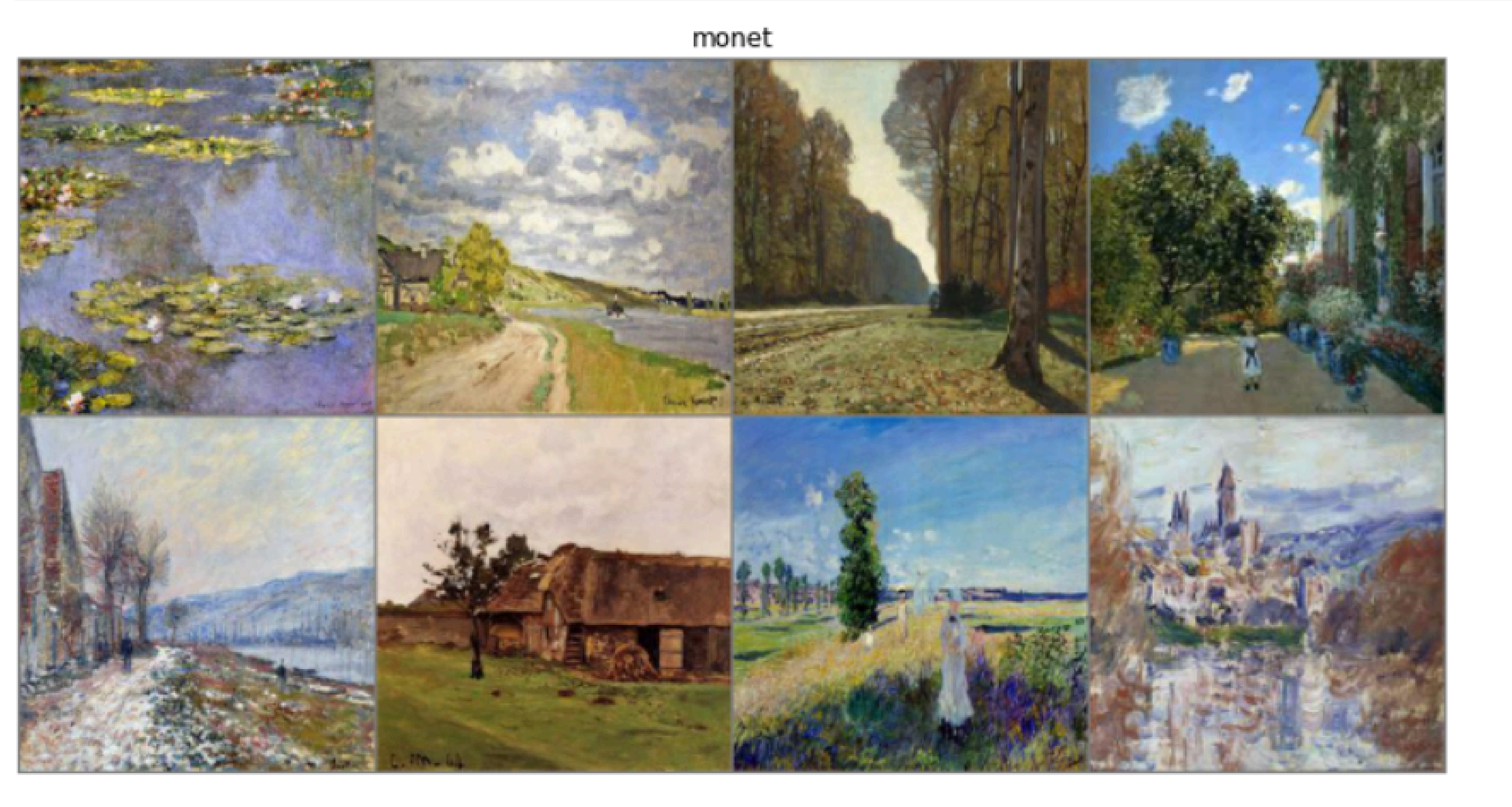
Training both Generators



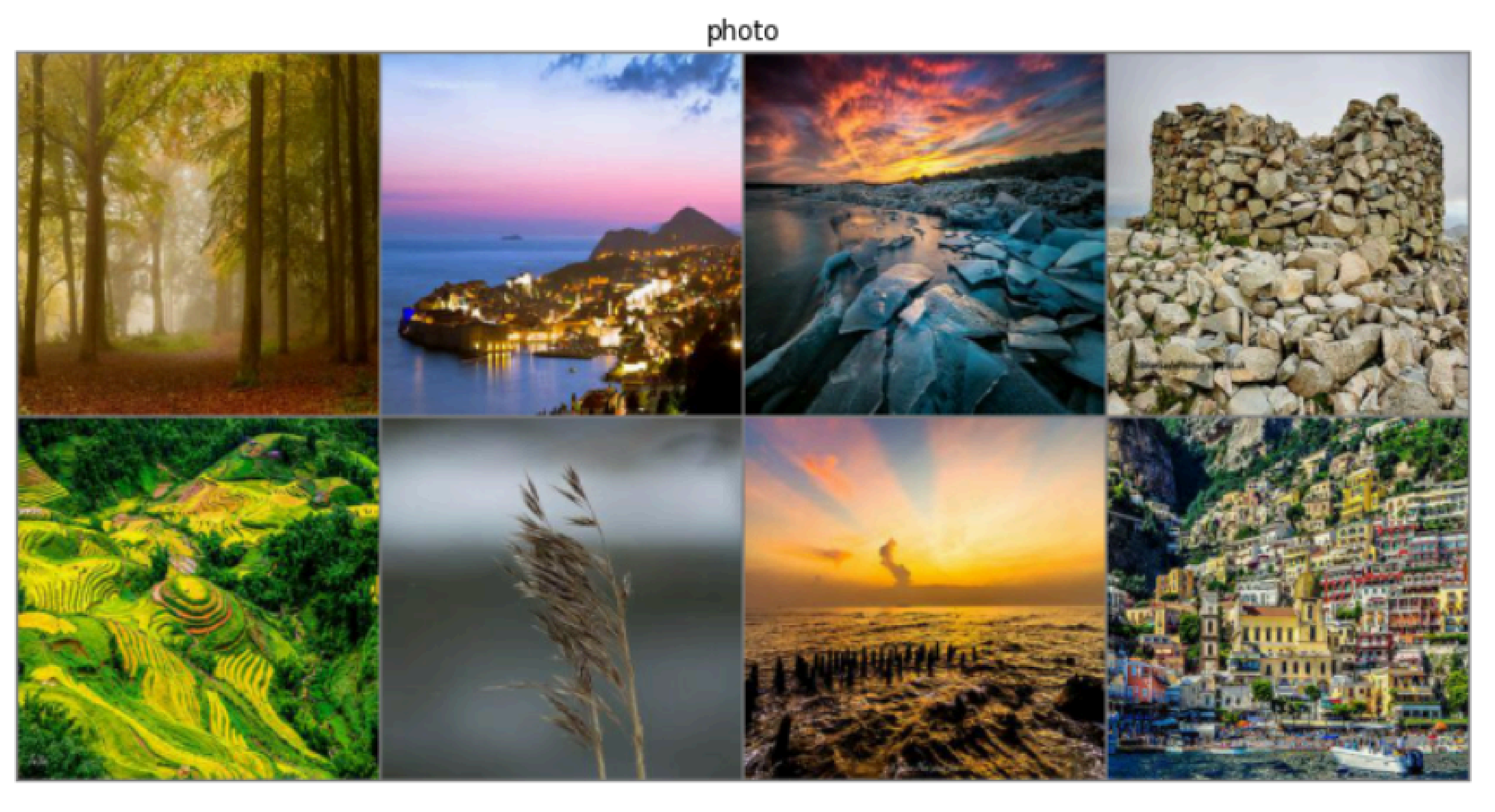
DATA

- The **dataset** contains images for training on Monet images for style and photographs for the content.
  - **monet_jpg** - 300 Monet paintings sized 256x256 in JPEG format
  - **photo_jpg** - 7028 photos sized 256x256 in JPEG format
- **Instance normalization** is a normalization method that seems to work particularly well for style transfer, which is why we use it in our implementation.

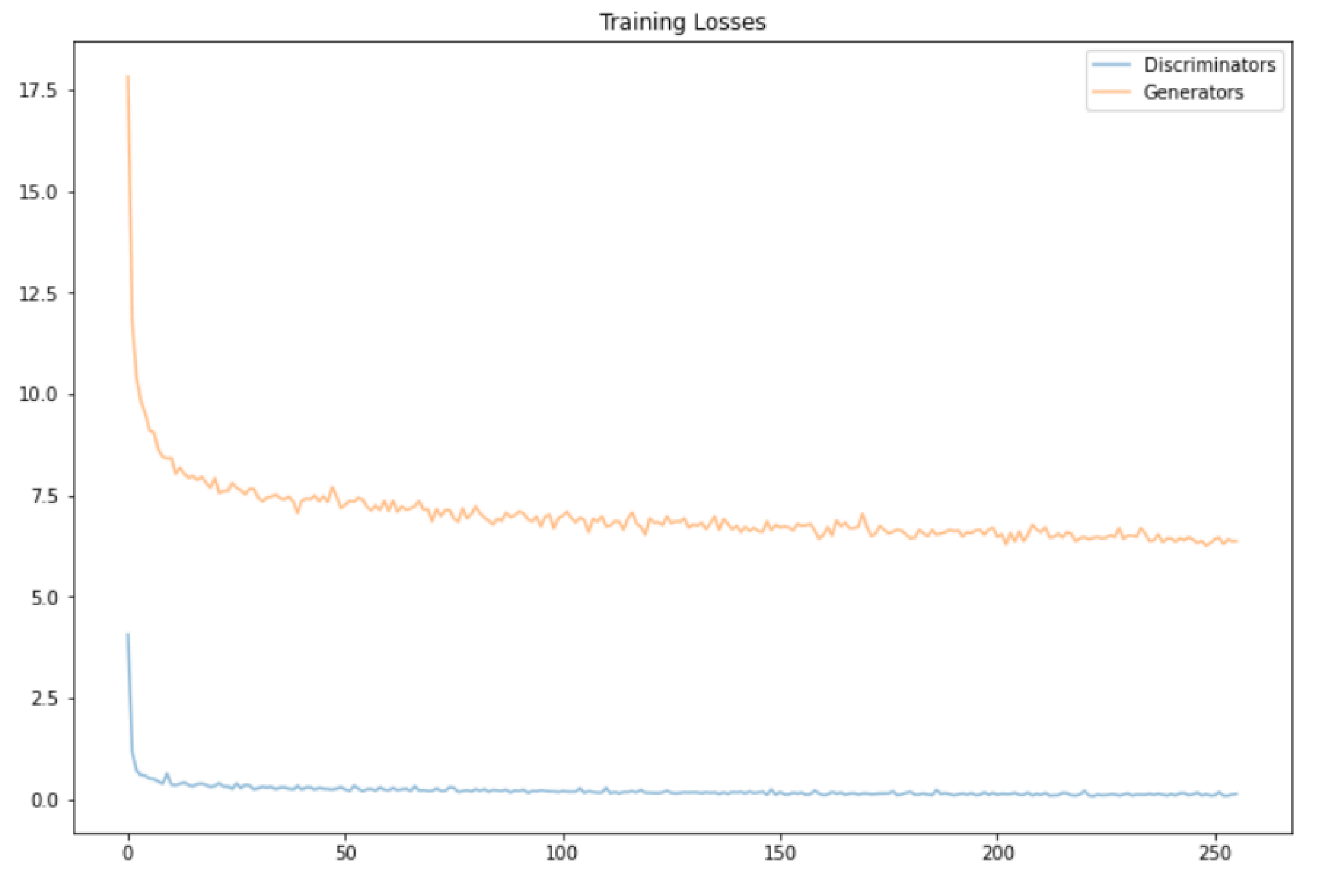
DATA



DATA



RESULTS

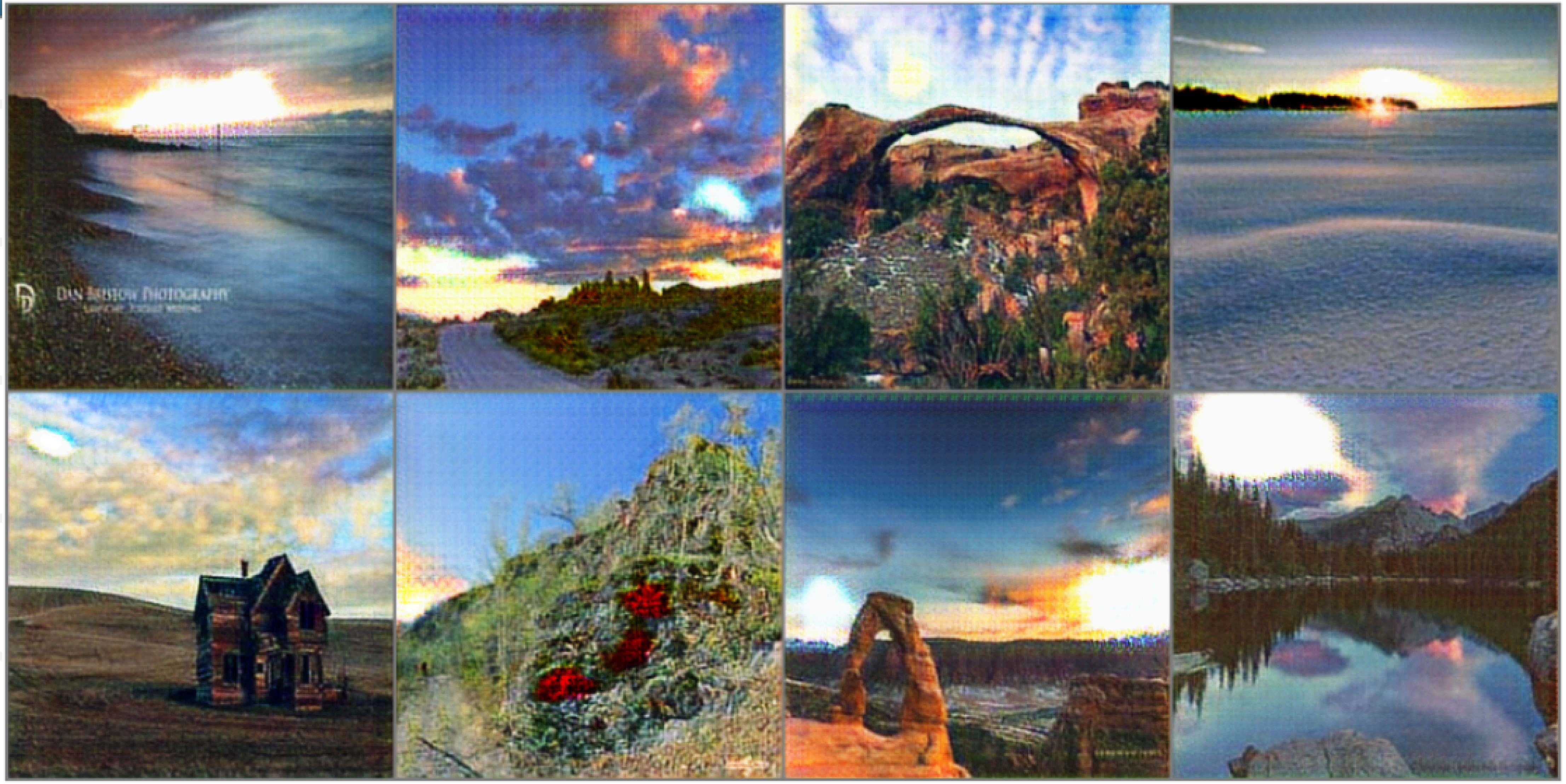

Training Losses
Discriminators
Generators
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
50
100
150
200
250

RESULTS

X

Domain X to Y Translation

Fake Y, Monet-esque



RESULTS

Domain Y to X translation

SUMMARY

This project delves into unpaired image-to-image translation, focusing on transforming ordinary photographs (domain X) into Monet-inspired artworks (domain Y) using CycleGANs. By harnessing this cutting-edge deep learning architecture, the project aims to seamlessly transfer Monet's impressionistic style onto everyday scenes, bridging the gap between photography and art without the need for paired training data. The CycleGAN framework enables the learning of mappings while enforcing cycle consistency, ensuring that the translated images retain original content while adopting Monet's vivid colors and textures.

THANK YOU

Presentation by
Achintya Gupta, Kannan Rustagi