Dissertation Submitted for the partial fulfillment of the **M.Sc. as a part of M.Sc. (Integrated) Five Years Program AIML/Data Science** degree to the Department of AIML & Data Science.

Project Dissertation

Converting Images to Monet Style Paintings Using CycleGAN and Neural Style Transfer

submitted to

By

Esha Mishra
Semester-VIII

Under the guidance of
“Dr. Ravi Gor and Mrs. Rashmi Pandey”

M.Sc. (Integrated) Five Years Program AIML

Department of AIML & Data Science
School of Emerging Science and Technology
Gujarat University

April 2023

DECLARATION

This is to certify that the research work reported in this dissertation entitled "**Converting Images to Monet Style Paintings Using CycleGAN and Neural Style Transfer**" for the partial fulfillment of M.Sc. as a part of M.Sc. (Integrated) in Artificial Intelligence and Machine Learning/Data Science degree is the result of investigation done by myself.

Place: Ahmedabad Esha Mishra

Date: 24 April 2023

ACKNOWLEDGEMENT

I would like to express my sincere gratitude towards Dr. Ravi Gor, co-ordinator School of Emerging Science and Technology at Gujarat University for giving me the opportunity to work on this project. Invaluable guidance under the supervision of Machine Learning faculties Rashmi Kumari helped me a lot. I am also grateful towards department administration for their cooperation.

Last but not the least, I would like to express my appreciation towards my parents Indra Mishra and A.K Mishra along with my sisters Diksha Mishra and Varsha Mishra and my friends Shruti H. Agarwal, Radhika Sharma and Prableen Sandhu for their unparalleled support throughout this project.

- Esha Mishra

CONTENTS

CHAPTER I ABSTRACT & KEY WORDS

Abstract
Art plays a significant role in human expression, with painting being one of the most prominent forms of visual art. In today's world, the importance of visual elements in content is well-understood. However, creating high-quality graphics can be time-consuming and expensive, and using images from the internet may not always be a feasible option due to copyright issues. This project aims to train two types of models to convert images to monet style paintings. First model is a type of GAN used for unpaired image-to-image translation. While the second model is implementation of an old graphic design technique using a type of deep learning model called CNN. It is called Neural Style Transfer.

Keywords
Monet, cycleGAN, Neural Style Transfer

CHAPTER II INTRODUCTION

2.1 Background

Art plays a significant role in human expression, with painting being one of the most prominent forms of visual art. In today's world, the importance of visual elements in content is well-understood. However, creating high-quality graphics can be time-consuming and expensive, and using images from the internet may not always be a feasible option due to copyright issues. Therefore, a cost-effective and user-friendly solution is needed. Monet paintings are named after French artist Claude Monet, who was one of the most popular Impressionist Era Artists.If we can produce Monet's style paintings, they can be used to create visually appealing content without infringing on copyright laws.

2.2 Problem statement

The objective of this project is to compare the monet style paintings produced by Generative Adversarial Networks (GANs) and Neural Style Transfer.

2.3 Objective

Train a cycleGAN and neural style transfer model to convert photos to Monet Style paintings.

2.4 Motivation and significance

To make it easier to create visually appealing content without infringing on copyright laws.

CHAPTER III BASIC TERMINOLOGY

3. Basic Terminology

3.1 ADAM optimizer - ADAM (Adaptive Moment Estimation) is a stochastic gradient descent optimization algorithm that is commonly used to update the weights and biases of neural network models during the training process. It combines the advantages of both Adagrad and RMSprop optimizers by using adaptive learning rates that are based on the first and second moments of the gradients.

3.2 CNN (Convolutional Neural Network)- is a type of neural network commonly used for image recognition and classification tasks. It consists of several layers of convolutional and pooling operations, followed by fully connected layers. CNNs can automatically learn features from raw image data and are able to achieve state-of-the-art performance on a variety of computer vision tasks.

3.3 Cycle consistency loss- Cycle consistency loss is a loss function used in CycleGAN models to ensure that the generated images are consistent with the original images. It measures the difference between the original image and the image that is generated by the generator after the image has been passed through the discriminator and back to the generator again.

3.4 cycleGAN- CycleGAN is a type of generative adversarial network (GAN) used for image-to-image translation tasks. Unlike traditional GANs, which require paired datasets for training, CycleGAN can learn to translate between two different domains without paired examples. This is achieved by introducing a cycle consistency loss that ensures the translated images are consistent with the original images.

3.5 GAN - (Generative Adversarial Network) is a type of neural network used for generating new data samples that are similar to a given dataset. It consists of two networks: a generator network that generates new samples, and a discriminator network that tries to distinguish between the generated samples and the real samples from the dataset. The two networks are trained together in a game-like setup, where the generator tries to fool the discriminator, and the discriminator tries to correctly classify the samples.

3.6 Neural style transfer- Neural style transfer is a technique used to transfer the style of one image onto the content of another image. It involves using a pre-trained CNN to extract features from both the style and content images, and then combining these features to generate a new image that has the content of the original image and the style of the other image.

3.7 Loss- It is a measure of how well a model is performing on a given task. It is calculated by comparing the predicted outputs of the model to the actual outputs (i.e., the ground truth) using a loss function. The goal of training a model is to minimize the loss, which is typically achieved by adjusting the model's weights and biases using an optimization algorithm such as stochastic gradient descent.

3.8 Monet - Paintings style named after impressionist era artist Claude Monet.

CHAPTER IV REVIEW OF LITERATURE

4.1 Understanding Monets

Monet paintings are named after French artist Claude Monet, who was one of the most popular Impressionist Era (1850-1895) Artists. Critics believed that the impressionist artists did not paint with technique, but rather simply smeared paint onto a canvas. Presently, impressionist art is considered one of the finest art works. Claude Monet's work had broad brush stroke, no sharp lines, mostly outdoor scenes and vivid depiction of reflection of light on water.

Image 1: Woman with a Parasol – Madame Monet and Her Son, by Claude Monet, 1875

4.2 Already Existing Tools For AI Image Generation And Editing

Mid-journey is so far the most popular tool in the market for text to image generation. It is free for non-commercial use but paid subscription is required for subscription use. Dalle-2 is an image to text generating model by Openai. They allow commercial use of their tool but result quality may vary based on quality of input.

Adobe illustrator plans to introduce GANs based image editing tools. Nvidia's styleGAN has producedHigh quality results in AI generating tools. Their most recent product is a text-to-video model.

4.3. Understanding cycleGAN

CycleGAN uses a cycle consistency loss to enable training without the need for paired data. In other words, it can translate from one domain to another without a one-to-one mapping between the source and target domain. This is called Image-to-Image Translation of Unpaired Data.

4.4 Understanding Neural Style Transfer

Style transfer is a computer vision technique that takes two images—a content image and a style reference image—and blends them together so that the resulting output image retains the core elements of the content image, but appears to be "painted" in the style of the style reference image.

CHAPTER V METHODOLOGY

5. Methodology

5.1 Data Collection

Data is taken from a kaggle competition called I'm Something of a Painter Myself.

This is a kaggle competition dataset. The aim of the competition is to convert photos to monet style paintings.There are two types of images:

1. Monet Paintings (jpg) - 300 samples (size: 256x256x3)
2. Photo (jpg) - 1000 samples (size: 256x256x3)

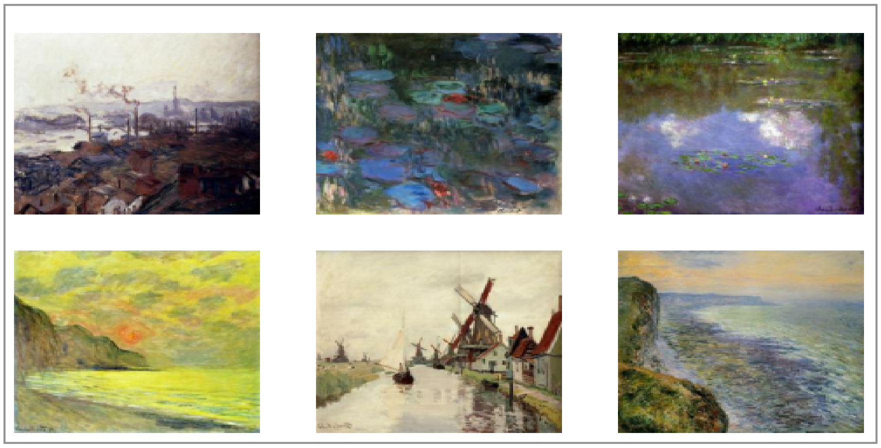

Image 2: Monet Sample

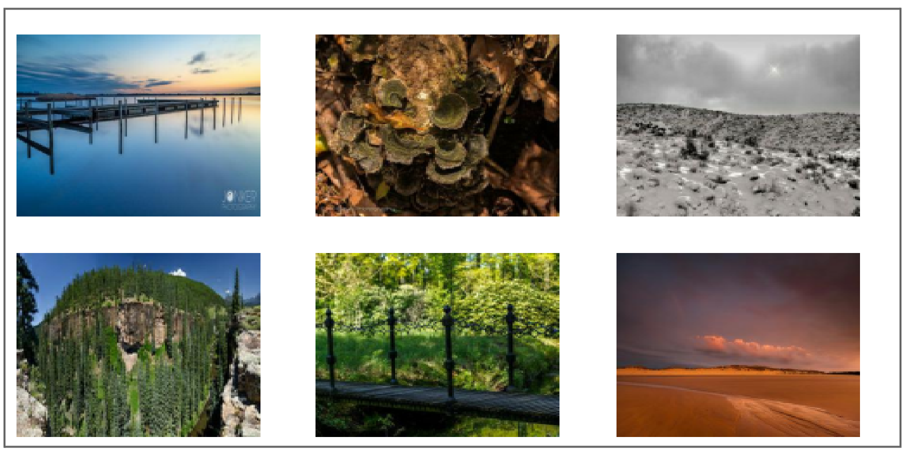

Image 2: Photo Sample

5.2 Data Visualization and Preprocessing

Related file name:Dataset_Visualization.ipynb, Cycle GANs.ipynb

Visualization of sample from the dataset. After splitting the data image into train and test images would be preprocessed. Preprocessing involved adding random noise and then flipping (left-right, right-left) and cropping at random. As seen in the code snippet (image 5)

Image 4: Gray version of image and RGB channels

```python
#Preprocessing funcitons

#1 crop
def random_crop(image):
  cropped_image = tf.image.random_crop(
      image, size=[256, 256, 3], seed = None, name = None)
  return cropped_image

#2 normalizing the images to [-1, 1]
def normalize(image):
  image = tf.cast(image, tf.float32)
  image = (image / 127.5) - 1
  return image

#3 random jitters
def random_jitter(image):
  # resizing to 360 x 360 x 3
  image = tf.image.resize(image, [360, 360],
                          method=tf.image.ResizeMethod.NEAREST_NEIGHBOR)
  image = tf.squeeze(image) # Remove the "None" dimension
  # randomly cropping to 256 x 256 x 3
  image = random_crop(image)
  # random mirroring
  image = tf.image.random_flip_left_right(image)
  return image
```

Image 5: Preprocessing code for cycle GAN

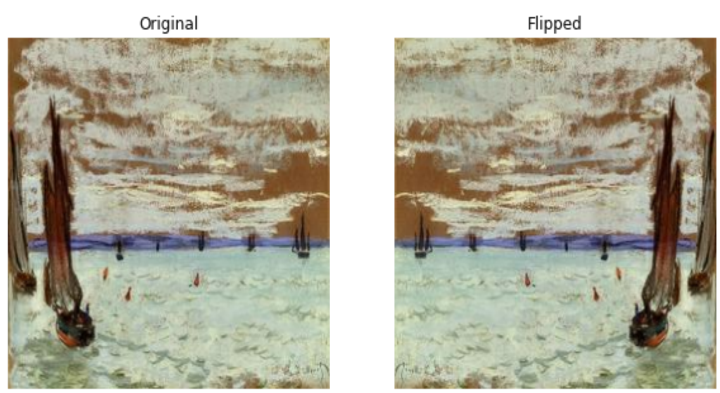

Image 6: Image after flipping

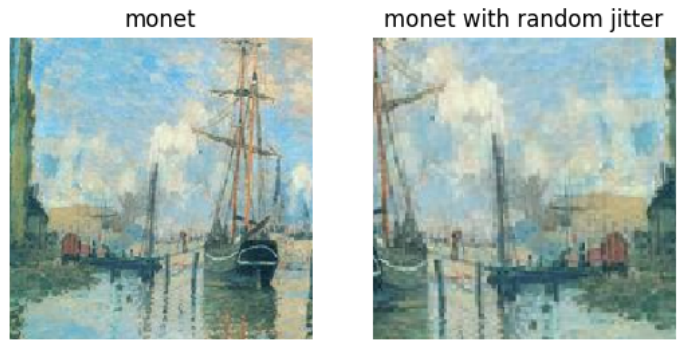

Image 7: Image after flipping and adding random noise

Neural style transfer takes only two images and requires resizing if needed.

5.3 Model

5.3.1 cycleGAN

Importing and using Pix2pix model from tensorflow examples. Import the generator and the discriminator used in Pix2Pix via the installed tensorflow_examples package.

CycleGAN uses a cycle consistency loss to enable training without the need for paired data. In other words, it can translate from one domain to another without a one-to-one mapping between the source and target domain. It is used for Image-to-Image Translation of Unpaired Data.

The model architecture used in this tutorial is very similar to what was used in pix2pix. Some of the differences are:

Cyclegan uses instance normalization instead of batch normalization.
The CycleGAN paper uses a modified resnet based generator. This project is using a modified unet generator for simplicity.

- There are 2 generators (G and F) and 2 discriminators (X and Y) being trained here.

- `Generator G learns to transform image X to image Y. (G:X->Y)`
- `Generator F learns to transform image Y to image X. (F:Y->X)`
- Discriminator D_X learns to differentiate between image X and generated image X (F(Y)).
- Discriminator D_Y learns to differentiate between image Y and generated image Y (G(X)).

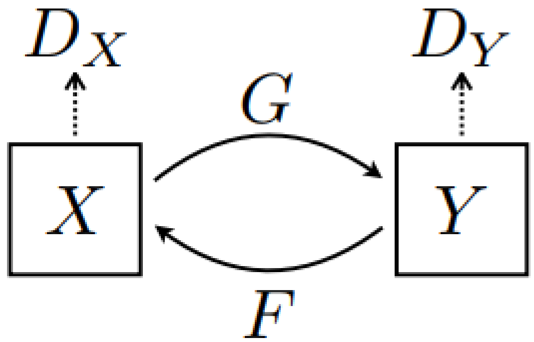

Image 8 cycleGAN

5.3.1 Neural Style Transfer

5.4 Evaluation Metrics

5.4.1 Frechet Inception Distance score

The Frechet Inception Distance score, or FID for short, is a metric that calculates the distance between feature vectors calculated for real and generated images.
The score summarizes how similar the two groups are in terms of statistics on computer vision features of the raw images calculated using the inception v3 model used for image classification. Lower scores indicate the two groups of images are more similar, or have more similar statistics, with a perfect score being 0. Zero indicates that the two groups of images are identical.

$$\mathrm{FID} = ||\mu_r - \mu_g||^2 + \mathrm{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$$

The FID score is used to evaluate the quality of images generated by generative adversarial networks, and lower scores have been shown to correlate well with higher quality images.

5.5 Model Training

Cycle was trained on collab TPUs. It took about 60-70 minutes to train 1 epoch. The cycleGAN was trained for 12 epochs. It is a time and computation intensive process. While Neural style transfer was trained for 15 minutes on collab TPU.

CHAPTER VI RESULT & DISCUSSION

6.1 RESULTS

6.1.2 Visual Output

Image 9 Output pix2pix model

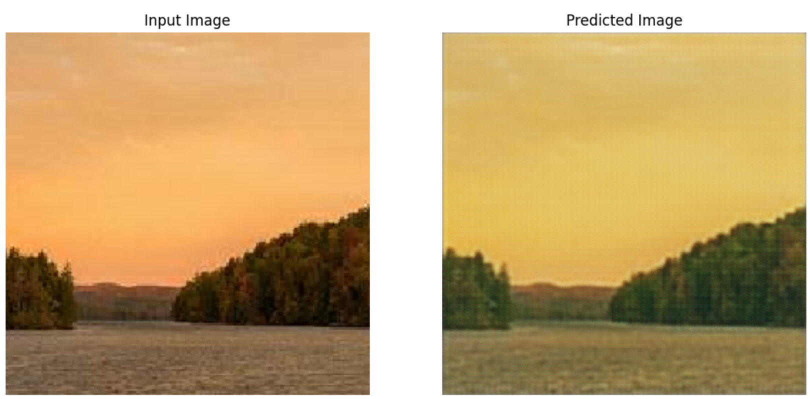

Image 10: Output cyleGAN

Image 11: Output Neural style Transfer

6.1.3 FID Score

Related file name: calculate_fid.ipynb

Sample of 190 images was taken to generate this score.

Model	FID score
pix2pix	19,17,342.13
cycleGAN	11,99,486.18
neural style transfer	2,08,174.26

6.2 Discussion

Visual comparison of output clearly shows that Neural style transfer performs best. It is less complex and has very less training time. This model works best in specific use cases but is difficult to scale in terms of output production as for every new input it requires retraining. If we have a better train GANs it can easily produce thousands of high quality monets with diverse output. Neural style transfer can be used to produce any style of images, not just Monet paintings.

Other than cycleGAN there are other available GANs like DCGAN for unpaired image-to-image translation. We can also try styleGAN by Nvidia as well.

CHAPTER VII CONCLUSION

7. CONCLUSION

Visual output is much better for Neural style transfer than cycleGAN. Neural style transfer is better for a very specific style but not for scaling output. The cycleGAn is does not produce the best result thus we can try using other types of GAN for this problem.

CHAPTER VIII
BIBLIOGRAPHY

8. Bibliography

Brownlee, J. (2019, August 30). *How to Implement the Frechet Inception Distance (FID) for Evaluating GANs - MachineLearningMastery.com*. Machine Learning Mastery. Retrieved April 23, 2023, from https://machinelearningmastery.com/how-to-implement-the-frechet-inception-distance-fid-from-scratch/

CycleGAN. (n.d.). TensorFlow. https://www.tensorflow.org/tutorials/generative/cyclegan

CycleGAN. (n.d.). TensorFlow. Retrieved April 23, 2023, from https://www.tensorflow.org/tutorials/generative/cyclegan

Deep Learning 46: Unpaired Image to Image translation Network (Cycle GAN) and DiscoGAN. (2019, August 24). YouTube. https://www.youtube.com/watch?v=nB8uVGbesZ4

Generative Adversarial Networks (GANs). (2019, April 8). YouTube. https://www.youtube.com/playlist?list=PLdxQ7SoCLQAMGgQAIAcyRevM8VvygTpCu

Home. (n.d.). YouTube. Retrieved April 23, 2023, from https://github.com/emish8/converting-images-to-monet-style-paintings-using-cycleGAN%20References

I'm Something of a Painter Myself. (n.d.). I'm Something of a Painter Myself | Kaggle. https://www.kaggle.com/competitions/gan-getting-started/overview