Artistic Style Generator Using CycleGAN, VGG16 and ResNet-50

Course CSE472: Machine Learning Sessional

Ayesha Binte Mostofa
1805062
Computer Science & Engineering
Bangladesh University of Engineering & Technology
Dhaka, Bangladesh

Md. Mahmudul Hasan
1805084
Computer Science & Engineering
Bangladesh University of Engineering & Technology
Dhaka, Bangladesh

ABSTRACT

Style Transfer is a technique in computer vision and graphics that involves generating a new image by combining the content of one image with the style of another image. The goal of style transfer is to create an image that preserves the content of the original image while applying the visual style of another image.

KEYWORDS

Computer Vision, Deep Learning, Image Style Transfer, VGGNet, ResNet

1 INTRODUCTION

In generating artificial output images with certain artistic style, we explored three different artistic style generating algorithms: CycleGAN, Neural Style Transfer by VGG16 and Resnet-50. CycleGAN is capable of finding the relationship amongst given unpaired image datasets, further facilitating unpaired image-to-image translation. With its transitivity within the network, we mapped from dataset of painting of some artists to dataset real photo in the respect of artistic style and content. VGG16 is a convolutional neural network that is 16 layers deep. It is widely applied in general use of image classification tasks, as well as the facial recognition tasks in terms of transfer learning. ResNet-50 is a 50-layer convolutional neural network (48 convolutional layers, one MaxPool layer, and one average pool layer). Eventually, our goal is to investigate all three methods in terms of the model architecture, the quality of the generated artworks and the efficiency of the whole generating process to figure out which method can better help us become a "modern but ancient artist".

2 RELATED WORK

In recent years, deep learning techniques have made significant advancements in the field of image generation and transformation. This section provides an overview of key approaches, including CNN GANs and image style transfer models.

2.1 CNN GAN

Convolutional Neural Network Generative Adversarial Networks (CNN GANs) have emerged as a powerful framework for generating high-quality images. The GAN architecture consists of a generator and a discriminator network trained adversarially. CNNs are commonly used in both the generator and discriminator to capture spatial hierarchies of features, enabling the generation of realistic images.

2.2 Image Style Transfer Models

Image style transfer models aim to transfer the artistic style of one image onto another while preserving the content. These models leverage deep learning techniques to achieve impressive results.

2.2.1 Neural Style Transfer. Neural style transfer separates and recombines the content and style of two input images using CNNs. By optimizing the generated image to match the content statistics of the content image and the style statistics of the style image, it produces visually appealing stylized images. VGG16 [1] and RESNET-50 [4] are also used in this type of works.

2.2.2 CycleGAN. CycleGAN [2] is a variant of GAN that can perform style transfer between unpaired images. It learns to translate images from one domain to another without requiring paired examples, making it suitable for various applications such as image-to-image translation and artistic style transfer.

2.2.3 AdaIN. AdaIN (Adaptive Instance Normalization) allows for controlling the style of an image through the mean and standard deviation of its features. This technique is commonly used in style transfer models to adjust the style of generated images dynamically.

2.2.4 Fast Neural Style Transfer. Fast neural style transfer accelerates the style transfer process by training feedforward neural networks to directly apply a stylized look to input images. This approach significantly reduces computational overhead compared to traditional optimization-based methods.

These models represent a subset of the diverse range of techniques in the field of image generation and transformation, each with its strengths and limitations. Ongoing research continues to advance the state-of-the-art in this area.

3 DATASET AND PREPROCESSING

We collected all the paintings of different artists from Kaggle's Best Artworks of All Time [3] (a collection of 8446 artworks of 50 influential artists). For the content images, we collected Landscape images from Kaggle. Any dataset of landscape can be used here.

For data preprocessing, we have checked the quality for both datasets, and there's no duplicated images or broken ones, which is good. For both models, to resolve the problem of having varied resolution across images, we converted our input images into size of 256 x 256 pixels, in JPEG format. In addition, our fast neural style transfer model is built on VGG16 and resnet-50, which is pre-trained on ImageNet.

4 METHODS

4.1 CycleGAN - TensorFlow

In our implementation, we utilize the CycleGAN architecture for unpaired image-to-image translation tasks. CycleGAN is a type of generative adversarial network (GAN) introduced by Zhu et al.

Architecture: CycleGAN consists of two main components: generators and discriminators. It employs two generators (G) and two discriminators (D), where each generator learns to map images from one domain to another, and each discriminator distinguishes between real and generated images.

Generator: The CycleGAN generator follows an encoder-decoder structure, processing input images through downsampling layers to a bottleneck layer. Skip connections are established within the downsampling layers, and the generator produces an output image based on the processed input.

Discriminator: The CycleGAN discriminator is a Convolutional Neural Network responsible for image classification. It assesses whether an input image is real or fake.

Key Features:

- CycleGAN employs two generators and two discriminators instead of one for each.
- During training, one generator receives additional feedback from the other generator to ensure cycle consistency.
- This feedback loop enhances the generators' performance by ensuring that applying both generators consecutively on an image yields a similar image.
- The discriminators evaluate whether images produced by the generators are realistic or fake.

4.2 Training

We used a learning rate of 0.0002, and the training ran for 10 epoch. It took almost 1 hour 30 minutes to run.

Generator Optimization: The generator is optimized using L1 loss to measure the mean absolute error between input components and output components. The resulting Generator Loss is a weighted combination of the validation loss, reconstruction loss, and identity loss, capturing how well the generator creates stylized images.

Discriminator Optimization: For our discriminators, we utilized binary cross-entropy loss to measure the difference between the predicted probability distribution and the true distribution of real and fake images.

4.3 VGG16 - TensorFlow

In our implementation, we employ the VGG16 architecture for fast style transfer. VGG16 is a convolutional neural network proposed by Simonyan

and Zisserman, known for its effectiveness in image classification tasks.

Architecture: VGG16 consists of 16 weight layers, including 13 convolutional layers and 3 fully connected layers. It is characterized by its simplicity and uniformity, with mostly 3x3 convolutional filters and max-pooling layers for spatial downsampling.

4.4 Training

Training is done by using perceptual loss. The output from the decoder is passed to VGG from which we extract features and calculate style loss and content loss. Then we calculate perceptual loss from the weighted sum of Style loss and content loss. For content loss, higher layers are used. For style loss, lower layers of networks are used. In defining our loss functions according to our content and style image shapes, we utilized the Gram Matrix by which we could obtain the MSE loss between content and style images. With the default number of epochs of 10, content image weight of 20, and style image weight of 100 and total variation weight = .004, total time spent on training is around 2 hours each time with a style image. We used a learning rate of 0.0002 with 15 epoch.

4.5 ResNet-50 for Style Transfer

In our implementation, we leverage the ResNet-50 architecture for fast style transfer tasks. While ResNet-50 is primarily designed for image classification, it can also be repurposed for style transfer by modifying its final layers and loss functions.

Architecture: In the context of style transfer, the ResNet-50 architecture is adapted by replacing its final fully connected layers with convolutional layers to preserve spatial information. This allows the network to transform input images into feature representations that capture both content and style.

4.6 Training

The training objective for style transfer with ResNet-50 typically involves minimizing a combination of content loss and style loss. Content loss measures the similarity between the feature representations of the input image and the content image, while style loss captures the differences in style between the input image and the style image.With the default number of epochs of 1000, content image weight of 1, and style image weight of $1e6$, total time spent on an image is around 30 minutes each time with a style image. We used a learning rate of 0.0002 with 1000 epoch.

5 RESULTS

Paintings by Monet (Giverny in springtime), Pissarro (Chestnut trees, Louveciennes, Spring - 1870), and Van Gogh(Starry Night) are used for image style transfer.

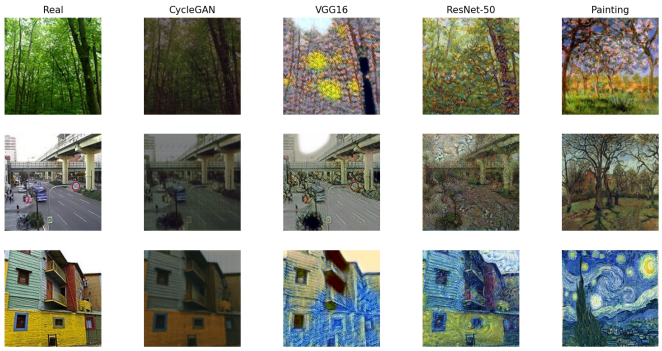

Figure 1: Images generated by different models from many painting

6 LOSS PLOTS

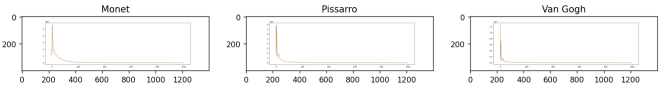

Figure 2: Resnet-50 Loss for three paintings

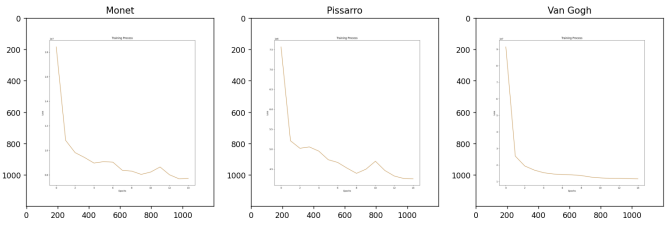

Figure 3: VGG16 Loss for three paintings

7 SCORES

The Fréchet Inception Distance (FID) compares the feature representations of images extracted from a pre-trained deep convolutional neural network, typically Inception-v3, trained on a large dataset. It considers both the mean and covariance of the feature representations to capture the distributional differences between real and generated images. A lower FID score indicates better similarity between the distributions and thus higher quality of generated images.

Painter's Name	FID_CYCLEGAN	FID_VGG	FID_Resnet
Claude Monet	2.2427e+89	2.2e+99	3.899e+97
Camille Pissarro	1.52e+86	5.57e+88	4.2e+95
Vincent van Gogh	1.15e+92	4.5e+98	1.28e+89

Table 1: Comparison of FID scores for different models with the real image

The PSNR metric is expressed in decibels (dB) and is computed using the mean squared error (MSE) between the original and reconstructed images. A higher PSNR value indicates lower distortion and better image quality, while a lower PSNR value suggests higher distortion and poorer image quality.

Painter's Name	PSNR_CYCLEGAN	PSNR_VGG	PSNR_Resnet	PSNR_Painting
Claude Monet	.45	.33	0.42	0.07
Camille Pissarro	.49	.41	.28	.1
Vincent van Gogh	0.259	0.30	0.283	0.05

Table 2: Comparison of Peak signal-to-noise ratio for different models with the real image

SSIM operates by comparing local patterns of pixel intensities in the reference and distorted images. The resultant SSIM index is a decimal value between -1 and 1, where 1 indicates perfect similarity, 0 indicates no similarity, and -1 indicates perfect anti-correlation.

Painter's Name	SSIM_CYCLEGAN	SSIM_VGG	SSIM_Resnet	SSIM_Painting
Claude Monet	16.6	13.5	15.5	9.72
Camille Pissarro	15.9	10.2	12.1	9.3
Vincent van Gogh	10.28	10.9	13.5	8.42

Table 3: Comparison of structural similarity index measure for different models with the real image

8 DISCUSSION

ResNet-50 outperforms VGG16 and CycleGAN in our tasks due to its deeper architecture, skip connections, better preservation of texture details, integration of global and local information, stability in training, faster convergence, flexibility in handling various image sizes, and adaptability to different input resolutions. However, the effectiveness of ResNet relative to other architectures may vary depending on specific dataset characteristics, task requirements, and optimization strategies employed during training.

9 CONCLUSION

In conclusion, our project has shown that Resnet-50 outperforms both the CycleGAN and VGG16 model in image style transfer. VGG16 also performs better than CycleGAN. So, VGG models and RESNET models are better for image style transferring.

10 FUTURE WORK

In our future work, we'd like to explore more different computer vision architectures, probably some other Generative Adversarial Network. We would like to apply the image style of some Bengali painters in near future. We'd be like to apply calligraphy style in bengali sentences. Finally, our evaluation metrics might be further improved by introducing more mathematical and statistical components.

11 ACKNOWLEDGMENTS

We extend our sincere gratitude to our project supervisor, Sheikh Azizul Hakim Sir, for his guidance and support throughout the research and preparation phases of our deep learning project.

REFERENCES

[1] Sanyam Bhutani. YYYY. Neural Style Transfer using VGG Model. https://towardsdatascience.com/neural-style-transfer-using-vgg-model-ff0f9757aafc. *Towards Data Science* (YYYY).
[2] Sandipan Dey. 2023. Monet Style Transfer with GANs (CycleGAN). https://sandipanweb.wordpress.com/2023/03/31/monet-style-transfer-with-gans-cyclegan/.
[3] I. Kuznetsov. 2020. Best Artworks of All Time. https://www.kaggle.com/datasets/ikarus777/best-artworks-of-all-time.
[4] Leo Luo. YYYY. Neural Style Transfer using ResNet50 with tf.keras. https://github.com/Leo8216/Neural-Style-Transfer-using-ResNet50-with-tf.keras-.