Big Data Advanced Research Project Report

Exploring Deep Learning in Art

Team Artificial Impressions
Blessy Chinthapalli
Shriya Yegalapati

Table of Contents

Executive Summary

In this research project, we delved into the fascinating realm of style transfer techniques applied to the visual arts, harnessing the power of deep learning methods to inspire and provide innovative solutions for emerging artists. Our exploration began with an extensive literature review, where we delved into the historical context and evolution of AI in art, with a specific focus on generative art and the advancements made in deep learning models, particularly Generative Adversarial Networks (GANs).

One of the primary objectives of our research was to investigate the concept of style transfer and its practical implementation using pre-trained models, such as VGG19. We provided a detailed explanation of the underlying model architecture and outlined the step-by-step process involved in training a style transfer model. This encompassed essential stages, including feature extraction, content and style loss calculations, and optimization techniques.

To demonstrate the efficacy of our approach, we presented a comprehensive showcase of the results obtained from our experiments. These compelling visual outputs illustrated the model's ability to seamlessly blend content and style from diverse images, yielding distinct and visually captivating artistic creations.

We also acknowledged several limitations inherent in the current model. These included issues of inflexibility, challenges in achieving an optimal style-content balance, demanding memory and computational requirements, and the relatively narrow scope of style transfer possibilities. In response to these limitations, we proposed a set of recommendations aimed at enhancing the model's performance and expanding its capabilities. These suggestions encompassed the exploration of advanced model architectures, the integration of perceptual losses to improve visual fidelity, the application of regularization techniques for better generalization, the implementation of progressive or multiscale training for increased flexibility, and the exploration of user-defined style patterns to offer greater artistic control.

The field of deep learning and AI in art is a rapidly evolving domain, brimming with untapped potential for further exploration and innovation. By leveraging the power of deep learning techniques, artists and innovators have the opportunity to push the boundaries of creativity and unlock new possibilities within the visual arts realm. Our research project aspires to contribute to this exciting field by providing a versatile and accessible tool for emerging artists to explore and experiment with style transfer techniques, fostering a more diverse and dynamic landscape of artistic expression.

In conclusion, through this research project, we have delved into the captivating world of style transfer in the visual arts, utilizing deep learning methodologies. We have showcased the possibilities and limitations of the current model while providing recommendations for its improvement. By embracing the potential of deep learning and AI in art, we empower emerging artists to expand their creative horizons and forge new paths of artistic expression.

Introduction and Background

Art and technology have always had a symbiotic relationship, constantly influencing and inspiring each other. In recent years, the field of deep learning has emerged as a powerful tool in the realm of visual arts, enabling artists and innovators to explore new creative possibilities. This research project aims to delve into the application of style transfer techniques in the visual arts, leveraging deep learning methods to provide inspiration and innovative solutions for emerging artists.

Objective

Our primary business objective is to explore and utilize style transfer techniques in the visual arts to provide inspiration for emerging artists and innovators. We will delve into different Deep Learning methods that already exist in our literature review and strive to implement them with a novel approach, tailored uniquely for our purpose. We envision our project being developed as a versatile tool (website and mobile application) applicable across multiple domains, including graphic design, traditional art inspiration, avatar generation for gaming and streaming platforms, printing and more.

Literature Review

To lay the foundation for our research, an extensive literature review was conducted, exploring the evolution of artificial intelligence (AI) in art throughout history. Notable articles, blogs, and research papers were analyzed to gain insights into the progression of deep learning methods in this space over time. By understanding the historical context, we can appreciate the advancements and transformations that have led to the current state of AI in the visual arts.
Here is a brief overview of the timeline of AI in Art :

3000 B.C. – Talking Knots

The ancient Inca used a system called *Quipu*—"talking knots"—to collect data and keep records on everything from census information to military organization. The practice, in use centuries before algebra was born, was both aesthetically intricate and internally logically robust enough that it could be seen as a precursor to computer programming languages.

1953 – Reactive Machines

Cybernetician Gordon Pasks developed his "MusiColour" machine, a reactive machine that responded to sound input from a human performer to drive an array of lights. Around the same time, others were also developing autonomous robots that responded to their environments, such as Grey Walter's Machina Speculatrix Tortoises Elmer and Elsie, and Ross Ashby's adaptive machine, Homeostat.

1968 – Cybernetic Serendipity

Artists in the 1960s were influenced by these "cybernetic" creations, and many created "artificial life" artworks that behaved according to biological analogies, or began to look at systems themselves as artworks. Many examples were included in the 1968 "Cybernetic Serendipity" exhibition at London's

Institute of Contemporary Art.

1973 – An Autonomous Picture Machine

In 1973, artist Harold Cohen developed algorithms that allowed a computer to draw with the irregularity of freehand drawing. Called Aaron, it is one of the earliest examples of a properly autonomous picture creator—rather than creating random abstractions of predecessors, Aaron was programmed to paint specific objects, and Cohen found that some of his instructions generated forms he had not imagined before; that he had set up commands that allowed the machine to make something like artistic decisions.

By the late 20th century, the field began to develop more quickly amid the boom of the personal computer, which allowed people who did not necessarily come from a tech background to play with software and programming.

Early 2000s

The field opened up considerably thanks to resources specifically geared toward helping artists learn how to code, such as artist Casey Reas and Ben Fry's Processing language, and open-source projects accessible on the Github repository. Meanwhile, researchers were creating and making public vast sets of data, such as ImageNet, that could be used to train algorithms to catalog photographs and identify objects. Finally, ready-made computer vision programs like Google DeepDream allowed artists and the public to experiment with visual representations of how computers understand specific images. Amid all these innovations, developments in the field of AI art began branching and overlapping. In this paper, we will focus on Generative Art.

Generative Art

The most commonly associated algorithm with A.I. Art today are Generative Adversarial Networks—or GANs. There are two things to understand about how a GAN works. First, the "generative" part: the programmer trains the algorithm on a specific dataset, such as pictures of flowers, until it has a large enough sample to reliably recognize "flower." Then, based on what it has learned about flowers, they instruct it to "generate" a completely new image of a flower. The second part of the process is the "adversarial" part—these new images are presented to another algorithm that has been trained to distinguish between images produced by humans and those produced by machines (a Turing-like test for artworks) until the discriminator is fooled. Let's dive deeper into the evolution of GANs

2014 – GANs are developed

Researcher Ian Goodfellow coined the term in a 2014 essay theorizing that GANs could be the next step in the evolution of neural networks because, rather than working on pre-existing images like Google DeepDream, they could be used to produce completely new images.

2017 –The Birth of GANism

After Goodfellow's essay about GANs was published in 2014, tech companies open-sourced their raw and untrained GANs, including Google (TensorFlow), Meta (Torch), and the Dutch NPO radio broadcaster (pix2pix). While there were a few early adopters, it took until around 2017 for artists to really begin to experiment with the technology.

2018 – Auction Milestone

Probably the most famous example of a GAN-made artwork in the contemporary art world is a portrait made by the French collective Obvious, which sold at Christie's in 2018 for a whopping $432,000. The trio of artists trained the algorithm on 15,000 portraits from the 14th to 20th century, and then asked it to generate its own portrait, which in a stroke of marketing genius, they attributed to the model.

Present Age

The current age of deep learning and AI in art is marked by significant advancements in the field. Researchers, artists, and technologists continue to explore and push the boundaries of AI-generated art through various algorithms and products/services. Here are some notable developments and models:

StyleGAN: One of the major breakthroughs in generative AI is the StyleGAN algorithm. Introduced by T. Karras et al. in the paper "A Style-Based Generator Architecture for Generative Adversarial Networks," StyleGAN allows for the creation of highly realistic and diverse synthetic images. It enables artists to manipulate and control the visual style of generated images, leading to the production of impressive and novel artworks.

DALL·E: DALL·E, introduced by OpenAI, is a state-of-the-art model that generates images from textual descriptions. It combines techniques from generative models and transformers to produce high-quality and highly creative visual outputs. DALL·E has the ability to imagine and generate images of objects and scenes that don't exist in the real world. This model showcases the potential of AI to create novel and imaginative visual content based on textual prompts.

CLIP: CLIP (Contrastive Language-Image Pretraining) is another remarkable model from OpenAI. It leverages a large-scale dataset of image-text pairs to learn a joint representation of images and their corresponding textual descriptions. By training on this diverse dataset, CLIP can understand the semantics and context of images and text. It enables a wide range of applications, including image classification, text-based image retrieval, and even generating textual descriptions of images.

MiDaS: MiDaS (Mixed Depth-Attentional Stereo) is a depth estimation model developed by researchers at Facebook AI. This model can take a 2D image as input and predict the corresponding depth map, providing information about the distance of objects in the scene. MiDaS utilizes deep neural networks and attention mechanisms to infer depth information accurately. It has proven to be valuable for various applications, such as virtual reality, robotics, and image editing.

One of the major applications of GANs is Neural Style Transfer, which we will try to implement in further sections and discuss in depth.

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.

Traditional supervised learning approaches for style transfer rely on image pairs consisting of an original image and its artistic representation. However, these pairs are rarely available, making this approach impractical. Neural style transfer (NST) has emerged as a more viable alternative, leveraging deep neural networks to perform transformations. NST utilizes neural networks to extract statistical features related to content and style, eliminating the need for explicit image pairs. By providing a single style reference image, the neural network can apply the desired style to original content images. To address the inefficiency of early NST methods, fast neural style transfer was introduced. This approach trains a standalone model that can transform any image in a single pass, significantly reducing the computational burden. State-of-the-art models even enable multiple style imprints through a single model, allowing creative editing of a single input content image.

Style Transfer Model Architecture Overview

Single style per model: Johnson et al. (2016) pioneered an independent neural network approach using large pre-trained VGG16 models for feature extraction and a small encoder-decoder network for style transfer. Each desired style requires a separate transfer network.

Multiple styles per model: Google researchers extended fast style transfer (2017) to enable a single network to produce images in multiple styles and blend them. Conditional instance normalization layers condition stylized images on additional inputs, allowing users to mix and match styles using a style vector.

Arbitrary styles per model: A common characteristic of both single and multi-style transfer models is that they can only produce images in styles that they've seen during training. A model trained on van Gogh's work cannot produce images like Picasso without retraining the entire network. Huang et al. introduced arbitrary style transfer, enabling a model to extract and apply any style to an image in a single pass. It takes a content image and a style image as input.

Optimizations and extensions: Stable style transfer addresses temporal inconsistencies in video frames, ensuring consistency in stylization. Color preservation techniques maintain the original color palette while transferring artistic styles. Photorealistic style transfer modifies encoder-decoder networks to minimize artifacting caused by deep convolution and upsampling layers.

Overall, all these algorithms enable the models to learn complex patterns and relationships in data, resulting in impressive and realistic outputs. It's worth noting that the field of deep learning and AI in art is evolving rapidly, and new models and algorithms are continuously being developed. These advancements continue to push the boundaries of what AI can achieve in creative endeavors, opening up new avenues for artistic expression and human-AI collaboration.

Methods

Training a style transfer model involves two networks: a pre-trained feature extractor and a transfer network. The feature extractor captures image content and style information from pre-trained layers. The transfer network, an image translation network with an encode-decoder architecture, generates the stylized image.

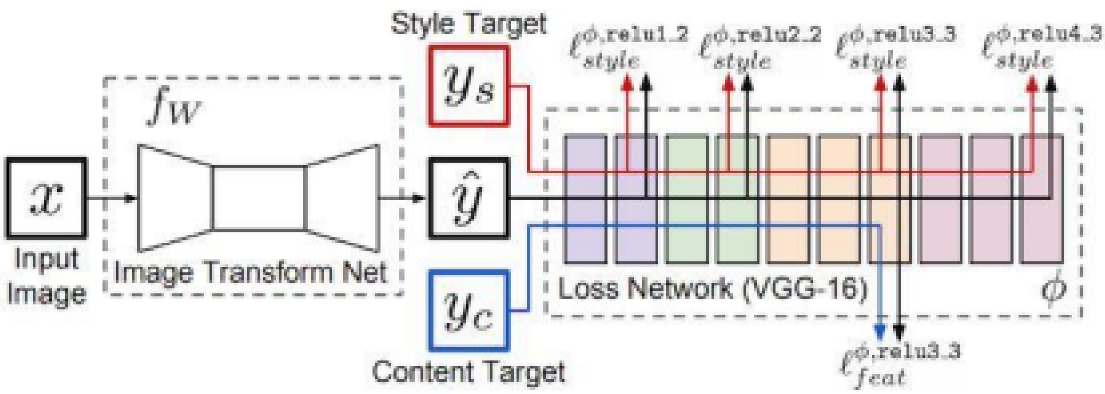

Fig. 1 : Style Transfer Method

During training, style images pass through the feature extractor to save style layer outputs. Content images are processed through the feature extractor and transfer network to obtain stylized images, whose content and style features are also saved. A custom loss function evaluates the stylized image quality based on content and style comparisons. Only the transfer network is updated while the feature extractor weights remain fixed. Weight adjustments in the loss function allow control over the level of stylization in the output images.

Data Description

ImageNet is a large-scale dataset containing millions of labeled images. It is designed for training and evaluating computer vision algorithms, particularly for image classification tasks. ImageNet has played a significant role in advancing the field of deep learning and has been instrumental in the development of deep neural networks. Our implementation is based on the VGG19 pretrained model for Keras., which is a convolutional neural network that is 19 layers deep. An open source, pre-trained version of the network trained on more than a million images from the ImageNet database is available freely for download. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224.

Model Implementation

To develop our model , we employed the following steps on Google Colab Pro using Python.

Step 1: Import the necessary libraries and load the content and style images.
Step 2: Defined a function to load the pre-trained VGG19 model, which will be used as the feature extractor.
Step 3: Extracted content and style features from the loaded images using the VGG19 model.

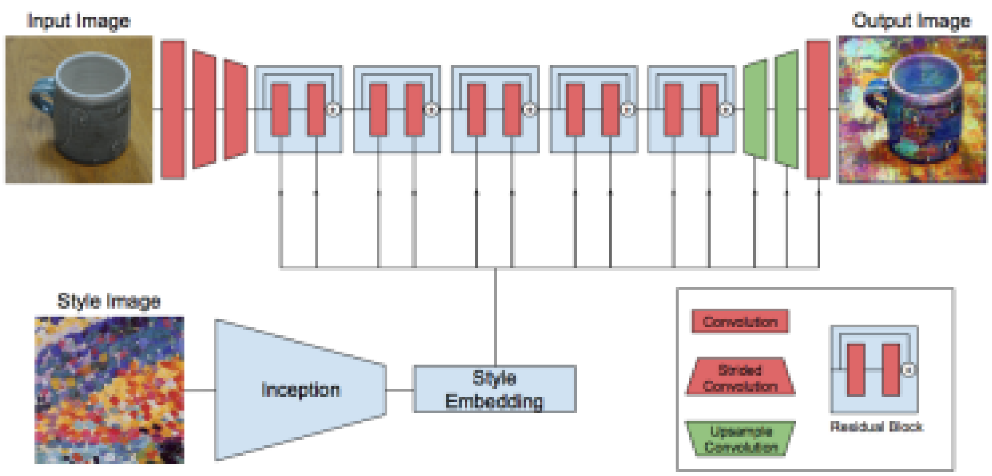

Fig. 2 : Input and Output Process

Step 4: Define the content loss by comparing the content features of the generated image and the content image.
Step 5: Define the style loss by comparing the style features of the generated image and the style image.
Step 6: Compute the total variation loss to encourage spatial smoothness in the generated image. Step 7: Define the total loss as a weighted combination of the content loss, style loss, and total variation loss.
Step 8: Use an optimizer to minimize the total loss and update the generated image. Step 9: Iterate the optimization process for a certain number of iterations to refine the generated image. Step 10: Save the final stylized image.

Results

We experimented with different images and iterations, and obtained the following results:

4000 Iterations:

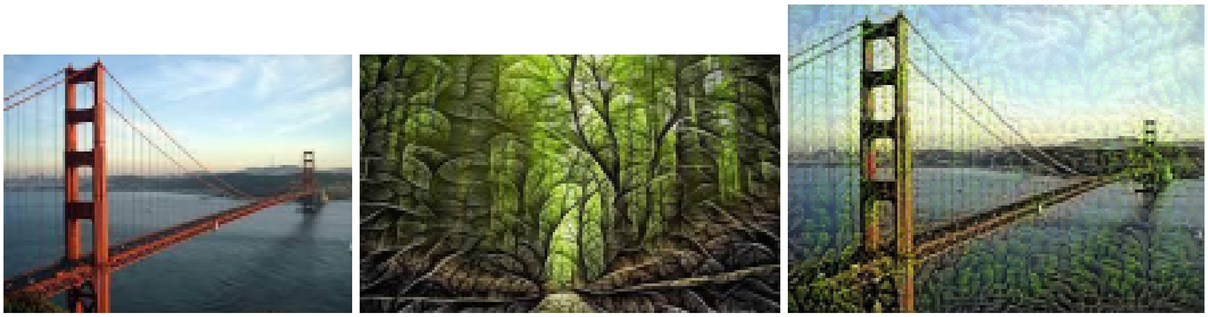

Image Style Result

10,000 Iterations:

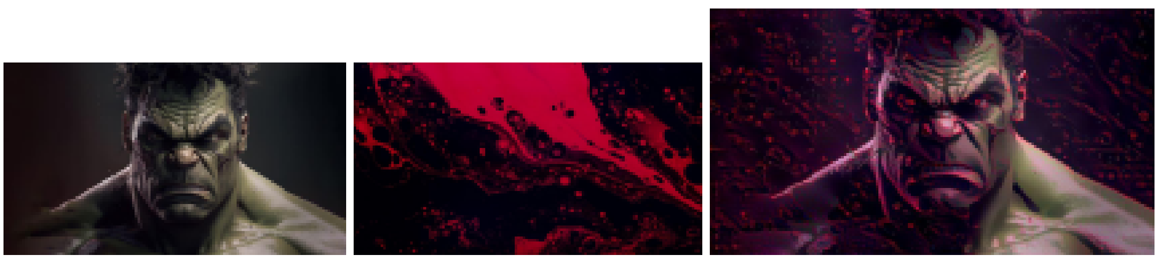

Image Style Result

15,000 Iterations:

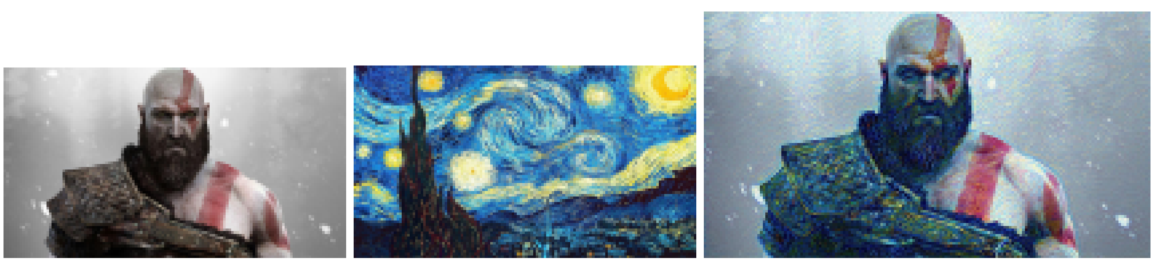

Image Style Result

Image Style Result

Limitations of the Model

1. Lack of Flexibility: The current model utilizes VGG19 as the feature extractor and may not capture all artistic styles effectively. Using different pretrained models or custom architectures can provide more flexibility.
2. Style-Content Balance: The style weight and content weight parameters determine the balance between style and content in the generated image. Fine-tuning these weights for specific style transfer tasks can be challenging and may require trial and error.
3. Memory and Computational Requirements: Neural style transfer can be memory-intensive and computationally expensive, especially for large images or complex styles. Processing high-resolution images may require significant computational resources and time.
4. Limited Style Transfer Scope: The model is limited to combining the style of a single reference image with the content of another image. Expanding the model to support multiple style references or user-defined style patterns can enhance its creative potential.

Recommendations for Improving the Model

1. Use Advanced Architectures: Consider using more advanced architectures, such as generative adversarial networks (GANs) or transformer networks, which have shown promising results in style transfer tasks.
2. Incorporate Perceptual Losses: Integrate perceptual loss functions, such as the ones used in perceptual loss networks (e.g., VGG16), to capture higher-level features and improve the visual quality of the generated images.
3. Include Regularization Techniques: Apply regularization techniques like total variation regularization or style consistency regularization to reduce artifacts and enhance the overall visual quality of the generated images.
4. Implement Progressive or Multiscale Training: Train the style transfer model in a progressive or multiscale manner, starting from low-resolution images and gradually increasing the resolution. This can improve convergence and the ability to capture fine details.
5. Incorporate User Interaction: Enable user interaction during the style transfer process, allowing users to provide feedback and guide the optimization to achieve desired artistic effects.
6. Parallelize Computation: Utilize parallel processing techniques, such as distributing the optimization process across multiple GPUs or utilizing cloud-based platforms, to expedite the style transfer procedure for large-scale or time-sensitive applications.
7. Implement Post-Processing Techniques: Apply post-processing techniques, such as color correction, smoothing filters, or other image enhancement methods, to refine the generated images and improve their visual appeal.
8. Develop a User-Friendly Interface: Create a user-friendly interface or application that simplifies the style transfer process, making it accessible to a wider range of users and enabling customization of various parameters and settings.

Conclusion

Artists and AI have collaborated for decades to provide us with methods and deep learning frameworks that resonate with art and lead to the creation of artworks. Utilizing the style transfer technique using the vast majority of technology can lead to the creation of impressive aesthetic artworks. While running the model does provide images that are appealing to the user, it is important to note that the model is limited to the current scope of style transfer mechanisms and frameworks available. The advent of deep learning and art is poised to expand rapidly and provide applications that tend to user interest, thinking and further experimentation.

Future Scope

In addition to empowering people all around the world to experiment with their own creativity, we see the importance of style transfer playing out in the commercial art world. And with the continued improvement of AI-accelerated hardware, both in the cloud and on the edge, style transfer can now be applied to captured and live video. This new capability opens up endless doors in design, content generation, and the development of creativity tools. Given this evolution, we can see how style transfer can be applied in a number of ways, whether it's photo/video editing, artist-community engagement, commercial art, gaming or virtual reality.

Citations

1. T. Karras et al. "A Style-Based Generator Architecture for Generative Adversarial Networks." arXiv preprint arXiv:1812.04948 (2018).
2. T. Nguyen-Phuoc et al. "GANPaint Studio: Semantic Photo Manipulation with a Generative Image Prior." ACM Transactions on Graphics (TOG) 37.6 (2018): 228.
3. "AI Music Composition | Jukedeck." Jukedeck, www.jukedeck.com/.
4. R. Radford et al. "DALL·E: Creating Images from Text." OpenAI Blog, openai.com/blog/dall-e/. 5. A. Radford et al. "Learning Transferable Visual Models from Natural Language Supervision." arXiv preprint arXiv:2103.00020 (2021).
6. R. B. Garg et al. "Guided Image-to-Depth Translation with Generative Adversarial Networks." IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2019.
7. "Facebook AI Research." MiDaS, research.fb.com/publications/midas-monocular-depth-estimation-in-real-time/.
8. A. Dosovitskiy et al. "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv preprint arXiv:2010.11929 (2020).
9.http://foundation.org/openaccess/content_cvpr_2016/papers/Gatys_Image_Style_Transfer_CVPR_2016_paper.pdf
10. Pre-Trained Model: https://www.tensorflow.org/api_docs/python/tf/keras/applications/vgg19/VGG19
11. Karen Simonyan, Andrew Zisserman [1409.1556] Very Deep Convolutional Networks for Large-Scale Image Recognition
12. ImageNet Dataset : https://www.image-net.org/