Project Report

Deep Learning

[CSE4007]

Deep Dream & Neural Style Transfer Experimentation

By

Aanchal Sharma (210354)
Jazlyn Jose (210236)

Department of Computer Science and Engineering
School of Engineering and Technology
BML Munjal University
November 2023

Declaration by the Candidates

We hereby declare that the project entitled "Deep Dream and Neural Style Transfer Experimentation" has been carried out to fulfill the partial requirements for completion of the course Deep Learning offered in the 5 th Semester of the Bachelor of Technology (B.Tech) program in the Department of Computer Science and Engineering during AY-2023-24 (odd semester). This experimental work has been carried out by us and submitted to the course instructor Dr. Soharab Hossain Shaikh. Due acknowledgments have been made in the text of the project to all other materials used. This project has been prepared in full compliance with the requirements and constraints of the prescribed curriculum.

Aanchal Sharma
Jazlyn Jose

Place: BML Munjal University
Date: 18th November 2023

Table of Contents

Introduction

In the ever-evolving landscape of artificial intelligence, the fusion of advanced techniques has opened unprecedented avenues for creative exploration. This project undertakes a thorough experimentation of two powerful image synthesis methodologies: Neural Style Transfer (NST) and DeepDream. Through exploration, the project aims to look at the artistic potential and implications of combining these cutting-edge technologies.

In contrast, DeepDream, a novel deep learning technology developed by Google, enhances the visual attributes of images. This method of creating surreal, dream-like imagery transcends traditional cognitive controls, resulting in art that captures the essence of the neural network's interpretive capabilities.

DeepDream operates by visualizing and enhancing features that activate specific neurons within a neural network. Through an iterative process, the network accentuates patterns triggered by these neurons, leading to the emergence of intricate and fantastical details. The technique involves the repetition of image feeds to the CNN model, starting with the detection of low-level features (e.g., edges and lines) and progressing to high-level features (e.g., faces and trees). These features are then combined to create complex visual effects.

According to Yongcheng Jing et al. in their review, the Deep Dream was the first experiment to produce artistic images by reversing the representation of CNN with the techniques of IOB-IR. This was the base of neural style transfer.

In contrast, Neural Style Transfer(NST) relies on pre-trained Convolutional Neural Networks (CNNs), often leveraging architectures like VGG19 or Inception. These networks serve as feature extractors, capturing hierarchical representations of content and style in different layers. Deeper layers focus on extracting semantic content, while shallower layers capture textural style. This hierarchical feature extraction allows for the disentanglement of content and style information.

Neural Style Transfer operates by combining two input images: a content image and a style image. The output image synthesized by NST blends the content of the former with the stylistic elements of the latter. This technology, situated within the realm of deep learning, has garnered attention for its ability to produce visually appealing and harmonious compositions.

Problem Statement

The project aims to combine Neural Style Transfer and Deep Dream to generate new images that combine the features of deep dream and the artistic style of neural style transfer. The goal is to enhance the visual attributes of images and create images that simulate hallucinations among psychiatric patients and drug addicts. Since it was found that the increase in drug abuse cases and cases of schizophrenia and hallucinations for many people, especially young people, prompted researchers to develop a technique that helps specialists in psychiatric clinics and psychiatrists to see these hallucinations that their patients see to be able to improve the treatment methods used. Secondly, most of the images and art movements nowadays have become computer-dependent. Finally, researchers in artificial intelligence have always been looking for ways and techniques to show what is happening between the hidden layers.

Deep Dream delves into the neural networks' interpretative capacity, generating visually intricate and surreal patterns that emulate the essence of human hallucination. The need for Deep Dream arises from its ability to bypass conventional cognitive controls, allowing for the exploration of artistic realms unattainable through conventional means. As a tool for artists, creators, and researchers, Deep Dream unlocks the potential to reimagine reality, infusing a touch of the surreal into digital art and pushing the boundaries of visual interpretation.

However, in harnessing the hallucinatory power of Deep Dream, several challenges emerge. The inherent opacity of neural networks poses difficulties in understanding how specific features and patterns emerge during the process.

Neural Style Transfer (NST) emerges as a transformative tool within the domain of artificial intelligence, offering a paradigm shift in the synthesis of visual content. Unlike conventional image processing methods, NST harnesses the power of deep neural networks to seamlessly blend the content of one image with the stylistic features of another, creating visually compelling compositions that transcend traditional artistic boundaries. The need for NST arises from its unique ability to infuse artworks with diverse stylistic influences, offering a versatile and dynamic approach to digital creativity. As a tool for artists, designers, and content creators, NST becomes indispensable in the quest for generating visually captivating and stylized imagery.

However, the integration of Neural Style Transfer into creative workflows is not without challenges. One of the primary concerns lies in achieving a harmonious balance between content and style, with optimal parameter configurations that preserve the essence of the content while properly rendering the chosen style.

Literature Review

The table below encapsulates a comprehensive literature review on studies exploring Deep Dream and Neural Style Transfer. Highlighting key research endeavors, the table outlines authors, publication years, objectives, methodologies, findings, and unique features of each study.

Author(s)	Title	Objective	Methodology	Findings	Unique Features
Suzuki et al.	Simulating Psychedelic States with Deep Dream	To simulate the visual hallucinatory aspects of the psychedelic state using Deep Dream.	Clamping different layers of the Deep Convolutional Neural Network (DCNN) to emphasize various features.	Successful simulation of altered states of consciousness similar to those induced by psychedelic drugs.	Clamping DCNN layers for varied effects; combining virtual reality and machine learning.
Al-Khazraji et al.	Artistic Image Generation using NST and Deep Dream	To combine Neural Style Transfer (NST) and Deep Dream for generating artistic images.	Using pre-trained networks VGG-19 for NST and Inception v3 for Deep Dream; incorporating the Gram matrix in NST.	Progressive incorporation of style and hallucinatory features, achieving a balance between content and style.	Integration of NST and Deep Dream; use of Gram matrix for effective style transfer.
Kiran	Enhancing Image Features with Deep Dream	To introduce a Deep Dream algorithm utilizing pretrained models.	Iterative image modification through gradient ascent at each layer of neural networks like ResNet, CNN, and ANN.	Emphasized feature extraction through gradient ascent, enhancing feature representation.	Iterative image modification; utilization of pre-trained models like ResNet, CNN, and ANN.
El-Rahiem et al.	Deep Dream in	To implement a	Using the Inception v3	Creation of secure	Application of Deep Dream in

	Multi-Biometric Cancellable Schemes	multi-biometric cancellable scheme (MBCS) using Deep Dream.	model and applying gradient ascent to maximize the loss function.	fingerprint cancellable patterns by maximizing the loss function for feature extraction.	secure multi-biometric cancellable schemes.
	A Hybrid Artistic Model Using Deepy-Dream Model and Multiple Convolutional Neural Networks Architectures	To create a hybrid artistic model integrating Deep Dream and NST using multiple CNNs.	The model combines features from VGG16, VGG19, Inception v3, Inception-ResNet-v2, and Xception; tested on blurred images.	Successful generation of diverse, high-quality artistic images; useful for simulating visual experiences of schizophrenia and drug addiction.	Hybrid model combining Deep Dream and NST; usage of multiple CNN architectures; effective in simulating hallucinations and enhancing aesthetics.
Emily L. Spratt	Dream Formulations and Deep Neural Networks: Humanistic Themes in the Iconology of the Machine Learned Image	To explore the intersection of art, image recognition, and AI.	Analysis of Google's DeepDream and Grad-CAM programs; comparing AI image recognition with human visual perception.	Highlights the importance of interdisciplinary collaboration in AI and machine learning for art historical research and image analysis.	Discussing DeepDream in the context of art history; advocating collaboration between computer scientists and art historians.

Various studies have showcased numerous applications of deep dream beyond image and video manipulation and some of them include using deep dream for Semi-supervised Semantic Segmentation, Centralized Intelligent Authentication System, Mental health research and therapy etc.

Hong, Noh, and Han's work introduces the Decoupled Deep Neural Network for Semi-supervised Semantic Segmentation, a notable application in deep learning. Unlike traditional image recognition networks, this innovative architecture goes beyond object recognition, utilizing backpropagation to extend understanding down to the pixel level. Although distinct from deep dreaming, this approach showcases the adaptability of deep learning principles, laying the foundation for detailed pixel-level analysis in supervised segmentation networks.

A study by *El-Rahiem et al. (2022)* aimed to implement a multi-biometric cancellable scheme (MBCS) using Deep Dream. The methodology involved using the Inception v3 model and applying gradient ascent to maximize the loss function. The findings demonstrated the creation of secure fingerprint cancellable patterns by maximizing the loss function for feature extraction. The unique feature includes the application of Deep Dream in secure multi-biometric cancellable schemes.

Similarly a study conducted by *Suzuki et al. (2017)* with the objective of the study being to simulate the visual hallucinatory aspects of the psychedelic state using Deep Dream. The methodology involved clamping different layers of the Deep Convolutional Neural Network (DCNN) to emphasize various features, producing distinct hallucinatory effects. The findings demonstrated the successful simulation of altered states of consciousness induced by psychedelic drugs, generating visually altered versions of original scenes. The unique feature includes clamping DCNN layers for varied effects and combining virtual reality and machine learning.

Another research by *Al-Khazraji et al. (n.d.)*aimed to combine Neural Style Transfer (NST) and Deep Dream for generating artistic images. The study used pre-trained networks VGG-19 for NST and Inception v3 for Deep Dream, incorporating the Gram matrix in NST for effective blending of content and style. The findings revealed a progressive incorporation of style and hallucinatory features through iterations, achieving a balance between content and style in the final images. Unique features include the integration of NST and Deep Dream and the use of the Gram matrix for effective style transfer.

A researcher also introduced a Deep Dream algorithm, utilizing pretrained ResNet, CNN, and ANN models. The methodology involved iterative image modification through gradient ascent at each layer of the neural network. The findings emphasized feature extraction through gradient ascent, enhancing the network's feature representation. The unique feature includes iterative image modification and the utilization of pre-trained models.

These research showcases the versatility of deep dream and neural style transfer in transforming conventional approaches to image processing and perception. From simulating psychedelic experiences to generating artistic imagery, the applications of these techniques extend beyond traditional boundaries, offering innovative solutions for detailed pixel-level analysis in segmentation networks and pushing the frontiers of deep learning in diverse domains

Description of the Dataset

The selection of datasets played a pivotal role in shaping the outcomes of our various experiments. Utilizing pre-trained models like VGG for Neural Style Transfer, VGG16, Inception V3, and EfficientNet for Deep Dream provided valuable insights into how the outputs varied based on the datasets these models were initially trained on.

VGG, InceptionV3, and EfficientNet were specifically trained on the ImageNet dataset. ImageNet is an extensive collection comprising 14,197,122 annotated images organized according to the WordNet hierarchy. Originating in 2010, the dataset serves as a cornerstone in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), a renowned benchmark for image classification and object detection.

The ImageNet dataset consists of a diverse range of images, and its annotations fall into two main categories. The first category involves image-level annotations, providing binary labels indicating the presence or absence of an object class in the image. For example, an image-level annotation could convey the presence of cars but the absence of tigers. The second category encompasses object-level annotations, offering precise information about the location and class label of an object within the image. This includes details such as tight bounding boxes and class labels, providing valuable insights into the spatial distribution and classification of objects.

In summary, our experiments benefit from the richness and diversity of the ImageNet dataset, allowing the pre-trained models to capture a broad spectrum of visual features and nuances.

Methodology

The primary goal of this project is to explore the intersection of two prominent image processing techniques—Deep Dream and Neural Style Transfer—and their collective application on a curated set of images.

We can outline the project as a series of research and implementation stages :

Project Initialization:

- Set up the computational environment, ensuring access to the necessary computational resources, such as a GPU for efficient processing.

Selection of Neural Network Models:

- Choose advanced neural network models that have been proven effective in image processing tasks. This project utilizes variations of the VGG16 and ResNet50 models, which are renowned for their performance in visual recognition tasks.

Algorithm Development:

- Implement the Deep Dream algorithm, which iteratively enhances patterns in images via network layer activations to create dream-like, visually complex images.
- Integrate the Neural Style Transfer technique to apply stylistic elements from one image to the content of another.

Optimization and Experimentation:

- Fine-tune the algorithms by experimenting with different hyperparameters, such as the number of iterations and the intensity of the style features.
- Utilize techniques like image pyramids to apply the Deep Dream effect at various scales and resolutions.

Image Processing and Transformation:

- Apply the developed algorithms to the prepared images, transforming them through the Deep Dream and Style Transfer processes to create new, artistically enhanced images.

Result Evaluation and Refinement:

- Assess the outcomes of the transformation process, examining the efficacy of different layers within the neural networks and their impact on the final images.
- Make iterative adjustments to the algorithms based on visual assessments and qualitative criteria.

Technology Stack

The implementation of the Deep Dream and Neural Style Transfer involved a comprehensive technical stack leveraging various tools, libraries, and frameworks. The technical stack is as follows:

Platform for Development:

The primary platform for development and experimentation was Google Colab, a cloud-based Jupyter notebook environment, providing computational resources and collaboration capabilities

Programming Language and Libraries:

- Python: Primary programming language.
- PyTorch: A deep learning framework for building and training neural networks.
- Torchvision: Provides popular datasets, model architectures, and common image transformations for computer vision.
- OpenCV (cv2): Image processing library used for various image operations.
- Pillow (PIL): Used for opening, manipulating, and saving many different image file formats in Python.
- NumPy: A fundamental package for scientific computing with Python, used for handling and transforming the image data into arrays.
- Matplotlib: A plotting library for creating static, interactive, and animated visualizations in Python. In this context, it is used to display images during the process.

Neural Network Model:

- VGG16: A convolutional neural network model used for experimentation. This model is modified to expose intermediate feature maps from specific layers.
- InceptionV3 : A convolutional neural network model incorporated for its efficiency in scaling and performance.
- EffiecientNetB3 : Convolutional neural network architecture utilized for feature extraction.

Deep Learning Framework:

- TensorFlow 2.x: This serves as the core framework for building and running the neural network models. It provides the necessary tools and libraries to execute high-level neural network tasks.
- Keras API: Integrated within TensorFlow, Keras provides high-level building blocks for developing and training deep learning models. It is used here to customize the VGG19 model and to construct the loss functions.

Utilities and Tools:

- IPython.display: Provides API for displaying tools in Jupyter notebooks.

- tf.GradientTape: A TensorFlow API for automatic differentiation; recording operations for the forward pass of the neural network and computing gradients on the backward pass.

Experimental Results

This section presents the experimental results derived from deep dream and neural style transfer, including the implementation of neural style transfer through deep dream techniques. The findings collectively demonstrate the diverse visual outcomes and creative possibilities that emerge from these advanced image processing approaches.

Deep Dream

This section delves into the exploration of diverse pretrained models, including VGG16, InceptionV3, and EfficientNetB3, to understand their unique behaviors. Through meticulous hyperparameter adjustments and deliberate variations in configurations, the deep dream algorithm is applied, offering insights into nuanced visual transformations. The EfficientNetB3 model is specifically investigated for its potential in artistic style transfer. Additionally, Inception V3 is employed to observe the evolving response to a noise image. The iterative application of the deep dream algorithm on Inception V3 and VGG16 provides a focused exploration of model responses and visual patterns, setting the stage for a comprehensive analysis of artistic and transformative capacities in subsequent sections.

VGG16 - Experimentation

The model architecture Vgg16, is designed to expose specific layers for experimentation. The key elements of the implementation include the model definition, preprocessing and postprocessing functions, and the application of Deep Dream to static images. It is a convolution neural network (CNN) architecture where instead of having a large number of hyper-parameters, VGG16 uses convolution layers with a 3x3 filter and a stride 1 that are in the same padding and maxpool layer of 2x2 filter of stride 2. It follows this arrangement of convolution and max pool layers consistently throughout the whole architecture. In the end it has two fully connected layers, followed by a softmax for output. The 16 in VGG16 refers to it having 16 layers that have weights. This network is a large network with about 138 million (approx) parameters.

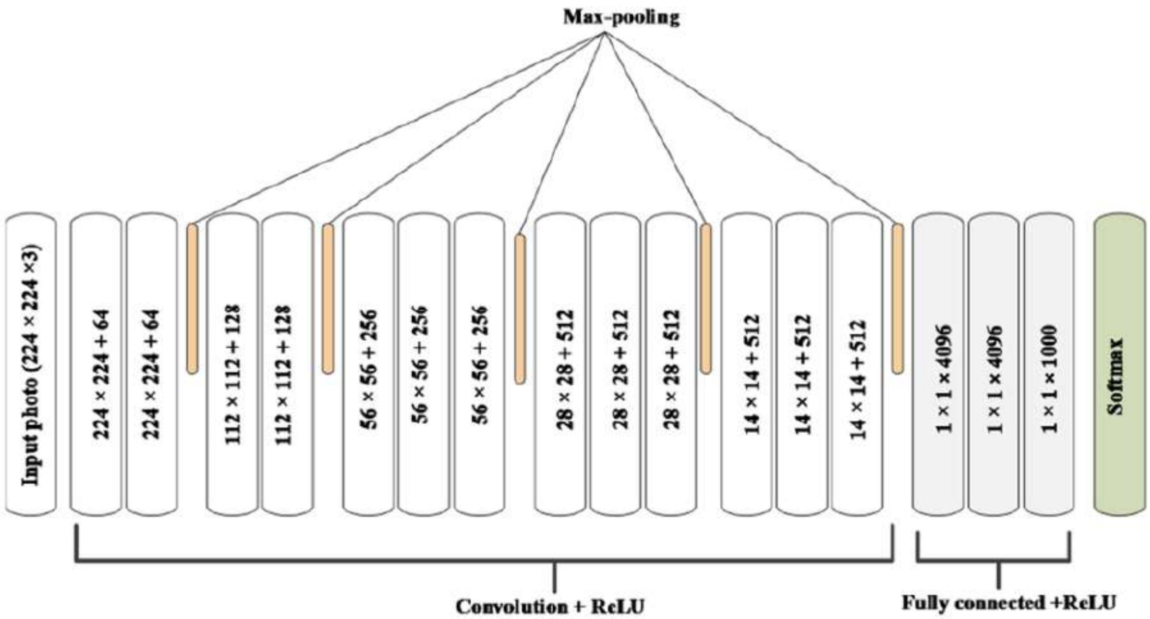


Fig. VGG16 Architecture

1. Model Architecture

The Vgg16 Experimental model is based on the VGG16 architecture, a widely used deep learning model for image classification. The key layers of interest, denoted as relu3_3, relu4_1, relu4_2, relu4_3, relu5_1, relu5_2, relu5_3, and mp5, have been selectively exposed for experimentation. These layers capture hierarchical features at different levels of abstraction in the neural network.

2. Deep Dream Implementation

2.1. Image Preprocessing

Images are preprocessed using standard normalization techniques with mean and standard deviation values obtained from the ImageNet dataset. Subsequently, image preprocessing and postprocessing were executed, accompanied by the introduction of random circular spatial shifts to infuse stochasticity into the algorithm to explore different configurations leading to a wider range of visual outcomes and allowing for the generation of diverse visual patterns.

2.2. Deep Dream Algorithm

The Deep Dream algorithm involves iteratively applying gradient ascent to an input image in the feature space represented by the exposed layers. The process includes introducing stochasticity through spatial shifts, which enhances diversity in the generated results. The application of the algorithm is performed over multiple iterations, and at each iteration, the gradients are collected to update the input image.

3. Experimental Configurations

3.1. Layer-wise Analysis

To understand the influence of different layers on the Deep Dream results, we conducted experiments for each exposed layer individually. The provided images showcase the visualizations achieved through the DeepDream algorithm, with an emphasis on the exposure of distinct layers within the neural network. Specifically, the layers exposed for visualization include 'relu3_3', 'relu4_1', 'relu4_2', 'relu4_3', and 'relu5_1.' Each of these layers represents a different level of abstraction in the network's hierarchy. The alteration of exposed layers introduces a dynamic aspect to the visual outcomes, revealing how the algorithm responds to various levels of features and complexities. Layers closer to the input capture low-level features such as edges and textures, while deeper layers capture increasingly complex and abstract representations. When a lower layer, such as 'relu3_3,' is exposed, the image tends to emphasize simpler patterns and details like edges and textures. As the exposed layer progresses to higher levels, such as 'relu4_3' or 'relu5_1,' the generated image evolves to highlight more intricate and abstract features, often resembling hallucinatory and complex visual patterns.

Fig. original image

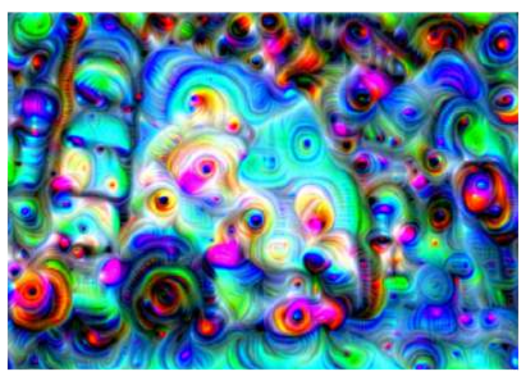


Fig. Exposed layer : relu3_3

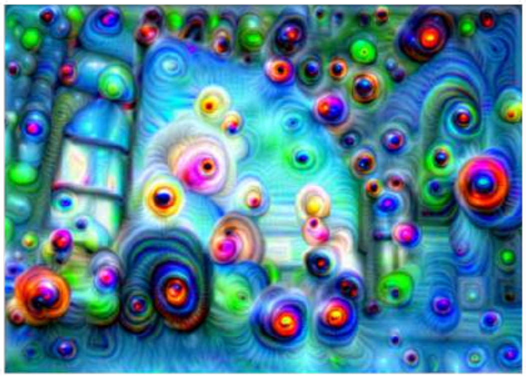


Fig. Exposed layer : relu4_3

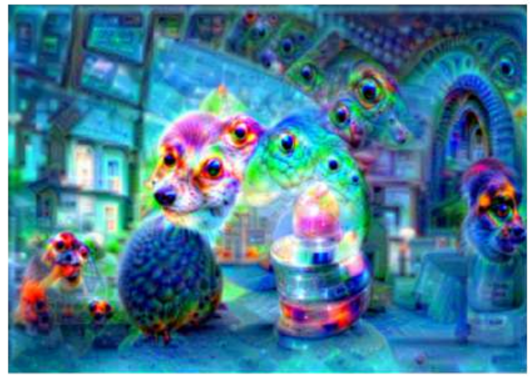


Fig. Exposed layer : relu5_1

3.2. Noise and Pyramid Size Variation

Additional experiments were conducted to explore the impact of noise and pyramid size on the generated images. The introduction of noise adds stochasticity to the algorithm, leading to more diverse results. Varying the pyramid size allows the algorithm to operate at different resolutions, influencing the scale of features captured during the process. The pyramid size was varied with the values 1, 3 and 5, the "pyramid size" refers to the scale at which the input image undergoes processing. It entails generating DeepDream effects on multiple copies of the original image at different scales. A larger pyramid size, like 5, processes the image at more scales, capturing varied details for intricate visual patterns. Conversely, a smaller size, such as 1, emphasizes prominent features at a single scale. This choice influences the diversity and detail in the DeepDream output—larger sizes yield intricate patterns, while smaller sizes highlight broader features.

Fig. original image

Fig. Pyramid size = 1

Fig.Pyramid size = 3

Fig. Pyramid size = 5

3.3. Iteration Sensitivity

The number of gradient ascent iterations was varied to observe its impact on the generated images. By exploring different iteration counts, we aimed to understand the trade-off between computational efficiency and the visual quality of the Deep Dream results. The gradient ascent iterations were varied from 2, 5 and 20, which resulted in a more detailed pattern. The lower numbers 2 and 5 resulted in a smoother appearance while the in the 20th iteration the features in the image looked more sharper

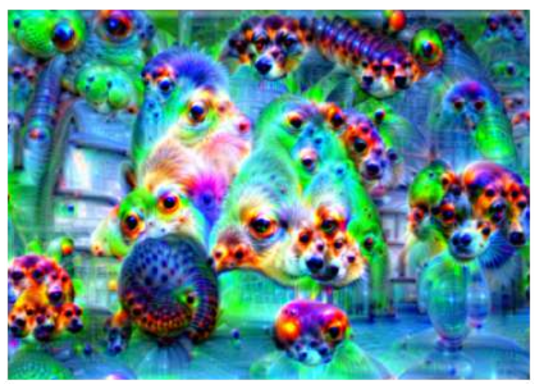


Fig. number of gradient ascent iterations = 2

Fig. number of gradient ascent iterations = 20

4. Visual Results

The following image was generated using the pretrained model VGG16. The selection of the intermediate layer (relu4_3) specifies the depth of feature extraction, impacting the complexity and abstraction of the generated image. Pyramid size (4) and pyramid ratio (1.8) determine the scale and arrangement of features, playing a pivotal role in defining the structure and patterns within the image. The number of iterations (10) influences the duration and depth of the optimization process, directly affecting the level of detail and refinement in the final output. The learning rate (0.09) regulates the step size during the gradient ascent, influencing the convergence speed and stability of the algorithm. Spatial shift (32) introduces randomness and spatial variability, adding stochasticity to the generation process. Finally, smoothness (0.5) contributes to the overall aesthetic appeal by controlling the blending and transitions between features.

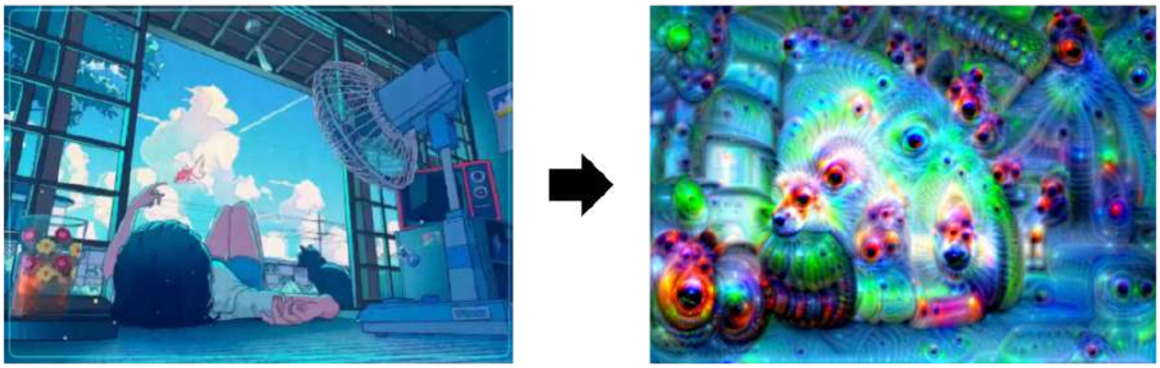


Fig. deep dream image generated for an animated image

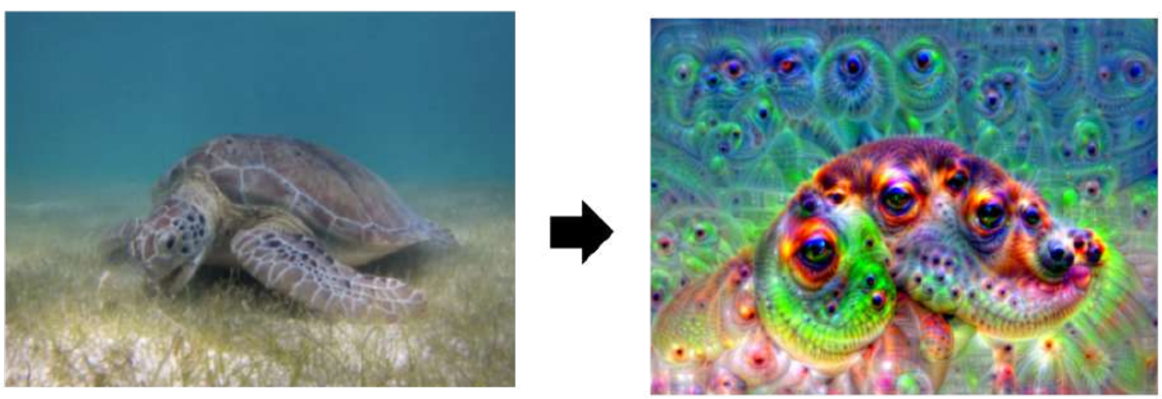


Fig. deep dream image generated for a turtle image

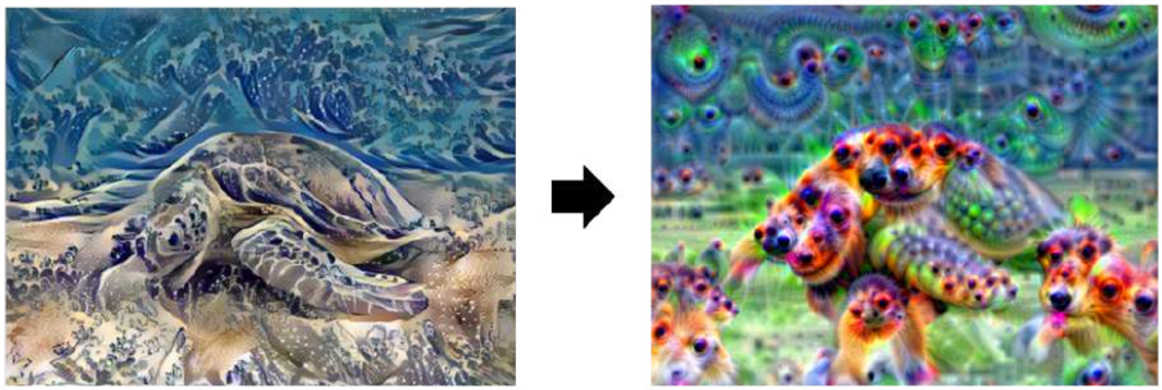


Fig. Deep dream image resulting from neural style transfer between a turtle and a Monet painting.

The visual results showcase the diversity and expressive capabilities of the Deep Dream algorithm. Images generated at different layers, with varying amounts of noise and pyramid sizes, provide insights into the model's feature extraction and synthesis capabilities.

In summary, our experimental results demonstrate the effectiveness of the customized Vgg16 Experimental model for Deep Dream generation. The layer-wise analysis, coupled with variations in noise, pyramid size, and iteration count, provides a comprehensive understanding of the algorithm's behavior and its ability to create visually compelling and abstract images.

InceptionV3 - Experimentation

Notably, InceptionV3 was employed to visualize the evolution of a noise image as activations increased during each iteration. The experimental setup aimed at viewing the intricacies of the model's response to varying input configurations. The deepDream algorithm was also iteratively applied progressively refining the input image to emphasize the patterns and features that activate specific layers in the inceptionV3 model. The evolving dream image is visualized at regular intervals throughout the process, providing insights into the intricate interplay between the model's layers and the input image.

1. <u>Visualizing activation patterns on a noise image</u>

In this section, an iterative process is undertaken to visualize the activation patterns within a deep neural network, specifically focusing on the layer 'mixed4d_3x3_bottleneck_pre_relu' and a selected feature channel (channel 139). The algorithm employs a gradient ascent technique on a randomly generated noise image, iteratively modifying it over 20 steps. As the optimization objective (score) at each iteration increases the level of distinct patterns in the noise image also increases. The model utilizes the inception graph of tensor flow' as the underlying architecture for feature extraction. Hyperparameters such as the iteration number, step size, and the layer and channel of interest contribute to the dynamic evolution of the visualized patterns.

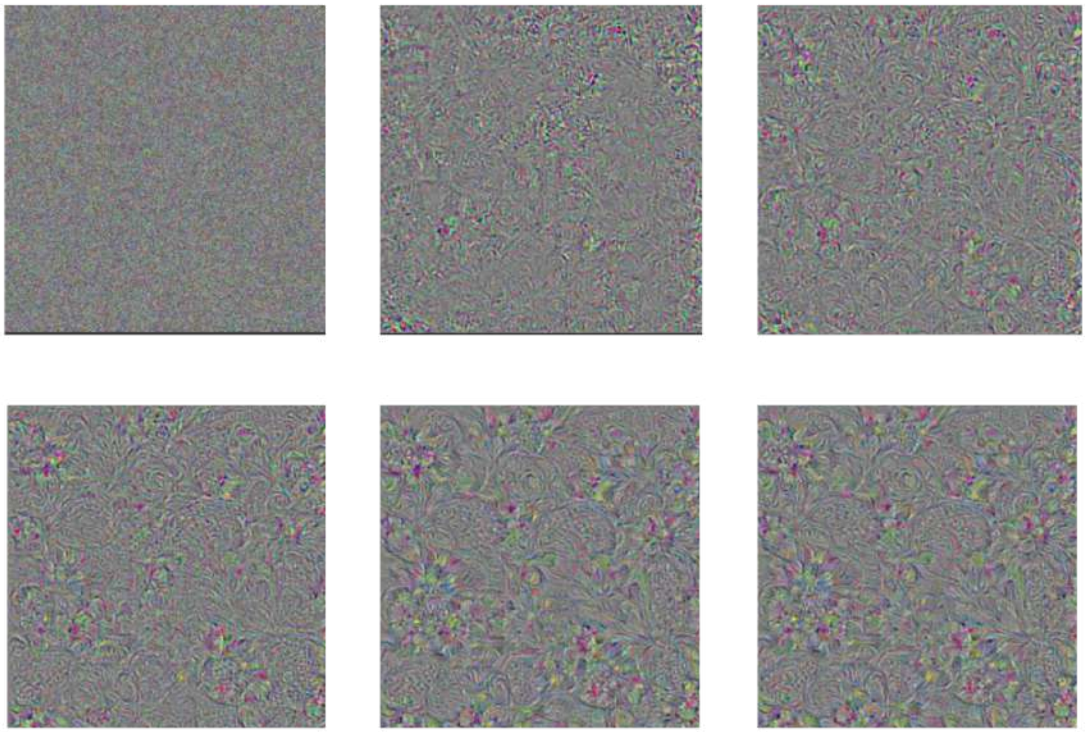


Fig. Evolution of Activation Patterns

The image showcases iteration, 0 with optimization -19.8, iteration 4 with optimisation 146.0, iteration 8 with optimisation 345.5.0, iteration 15 with optimisation 667560, iteration 19 with optimisation 735.33

2. Applying Deep Dream on blended images

Blending Two Images with PIL : The method of image blending is commonly used in various image processing applications, such as creating composite images, watermarking, or in artistic endeavors like creating double exposure effects. The ability to control the degree of blending with alpha parameters offers flexibility in achieving the desired visual result.

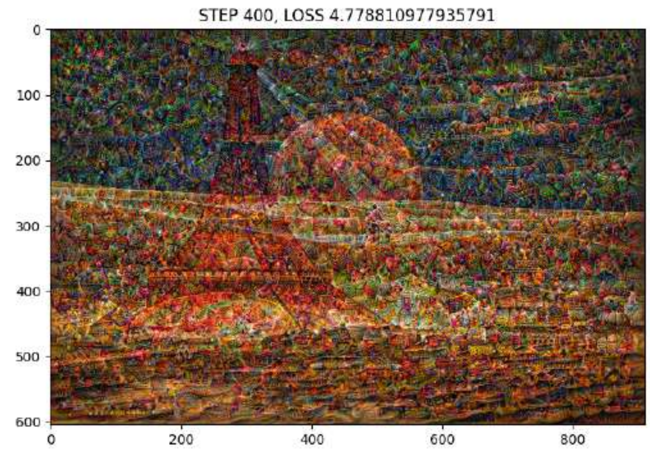
By blending, we prepare a composite image that possesses features from both sources. This composite becomes the input for the subsequent Deep Dream processing, ensuring that the Deep Dream algorithm works on an already stylistically enriched base. Blending allows for creative expression, creating a new image that neither replicates one of the originals nor is completely distinct. It's a way to explore new aesthetics by fusing different visual elements

Fig. image

Fig. image 2

Fig. Blend of image 1 and image 2

After combining two images through blending, we initiate the Deep Dream process. First, a deep network is trained, and specific layers are chosen for retention. Activations from these layers are calculated, and the gradient and loss of these activations are determined. The input image is then modified by amplifying these activations, enhancing the visual patterns. The modified image is fed back into the network, and this iterative process is repeated multiple times. This unique approach not only enriches the blended base with dreamlike features but also offers a creative avenue for exploring novel aesthetics through the fusion of distinct visual elements.



The images show a sequence of results from applying the DeepDream algorithm to an initial image, with each subsequent image representing a further iteration in the process. Here's a breakdown of what is happening across these iterations:

STEP 0: This is the original image before any DeepDream processing. The loss is relatively low because there have been no modifications to increase the activation of the network's layers.

STEP 100 to STEP 900: These images show the transformation of the original image through successive iterations of the DeepDream process. With each step, the loss value increases, which indicates that the image is being modified to more strongly activate the features the neural network is looking for.

The neural network starts to recognize patterns and textures that it has learned during its training. These could be shapes that resemble animal eyes, scales, feathers, or other recognizable features from the dataset the network was trained on. As the steps increase, these features become more pronounced and surreal.

The progression shows a gradual intensification of colors, patterns, and textures. This is consistent with how DeepDream works; by amplifying patterns the network sees in the image, the original image details are transformed into intricate, sometimes bizarre patterns that can resemble psychedelic or dream-like visuals.

The images follow the typical pattern of DeepDream iterations:

- Early iterations (e.g., STEP 100) show subtle enhancements to the image's existing features.

- Middle iterations (e.g., STEP 300-500) show the patterns becoming more intricate and pervasive, often overlapping much of the image.
- Later iterations (e.g., STEP 700-900) often result in highly abstract images where the original content is heavily obscured by the complex, dream-like textures.

This sequence allows us to see the continuous development of the DeepDream process, where the neural network increasingly imposes its learned features onto the original image. The loss function guides this process, seeking to maximize the activation of certain layers within the network, thereby 'dreaming' more intensely about the learned features. As such, the 'loss' can be seen as a measure of how far the image has been transformed from its original state into the dream-like state that the network is creating.

Neural Style Transfer With Deep Dream Technique Using EffiecientNetB3

The combination of Neural Style Transfer with the Deep Dream technique brings a unique dimension to artistic image generation. Instead of a single static stylized output, this approach introduces an iterative process that progressively refines and amplifies stylistic features by maximizing the activations of specific blocks within the EfficientNetB3 network. Deep Dream is employed to enhance the stylization process, adding intricate details and textures to the final image. The experimentation involved the implementation of a deep dream algorithm utilizing an EfficientNetB3 model for style transfer.

EffiecientNetB3 Architecture

EfficientNetB3 represents a sophisticated convolutional neural network architecture designed to achieve a delicate balance between model efficiency and high performance. The architecture is characterized by Mobile Inverted Bottleneck Convolution (MBConv) blocks, which efficiently combine depthwise separable convolutions, linear bottlenecks, and inverted residuals which allows the model to efficiently capture and represent features across various scales in the input images. EfficientNetB3 features a design that includes global average pooling and fully connected layers to distill learned features into final predictions.

Working Mechanism

Feature Extraction using EfficientNet: The deep neural network is used for feature extraction. In this case, EfficientNetB3 was utilized, and features from the last three blocks were extracted, namely block 5e, block 6f, and block 7b. These blocks were chosen for their ability to capture complex and abstract features within the input image.

Gradient Ascent for Style Enhancement: The input image was processed through the EfficientNet model obtaining feature maps from the selected blocks, and then a gradient ascent approach was employed, where the image is iteratively adjusted to maximize the activation of specific neurons within the chosen blocks. To ensure stability and prevent excessively large updates, the gradients were normalized. This normalization step helped in controlling the magnitude of adjustments made to the input image.

The input image was then modified by adding the normalized gradients scaled by a learning rate. This modification amplified the features that corresponded to the chosen style, enhancing the stylization effect. The iterative process was repeated for a predetermined number of iterations (in this case, 150). Each iteration refined the image to increasingly emphasize the desired stylistic features.

Deprocessing: At the end of each iteration, a deprocess function is applied to the generated image. This function reverses the preprocessing steps that were initially applied to the input image. It includes operations such as mean addition, standard deviation scaling, and pixel value clipping. The deprocessed image is then visualized to provide a clear and interpretable representation of the stylized output.

<u>Implementation</u>

The image below illustrates the implementation of style transfer using deep dream techniques. The Learning rate was set to 0.01, determining the step size during gradient ascent, 150 optimization steps were done to maximize the feature activations.Experimental iterations were conducted to visualize the effect of maximizing the output of blocks 5, 6 and 7 in the EfficientNetB3 model. It is notable that maximization of each block resulted in distinct and varied stylized outcomes.

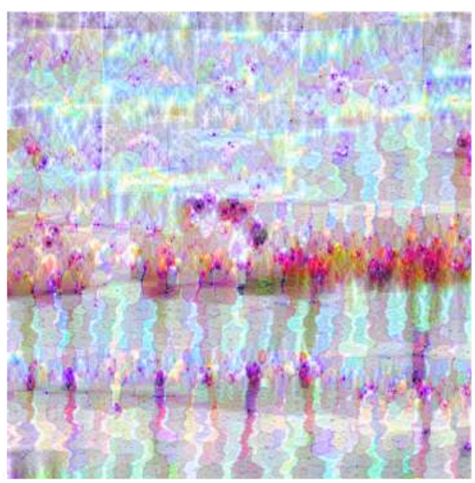


Fig. original image fig. Maximizing the output of block 5

fig. Maximizing the output of block 6 fig. Maximizing the output of block 7

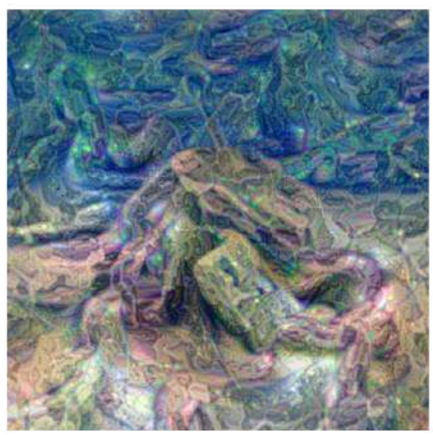


Fig. original image fig. Maximizing the output of block 5

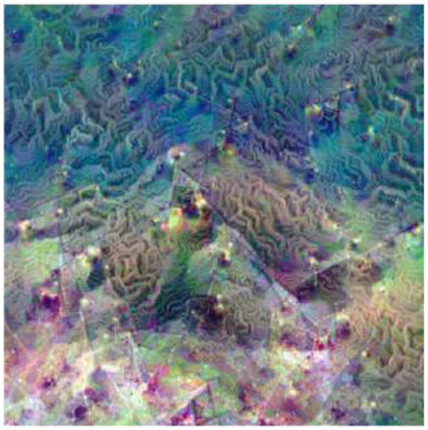


fig. Maximizing the output of block 6

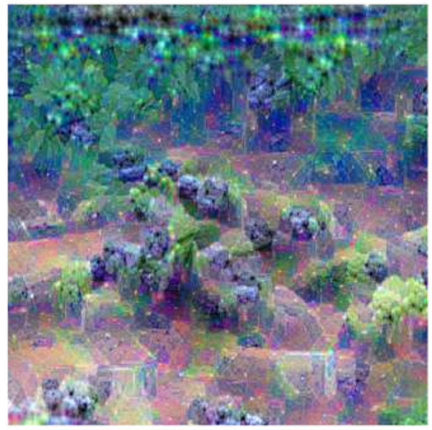


fig. Maximizing the output of block 7

This process encompassed the application of the deep dream technique, wherein stylized images were generated by maximizing the activations of specific blocks within the EfficientNetB3 network. The objective was to explore the model's capacity for artistic style transfer and observe the unique visual outcomes produced by manipulating the activations in targeted network blocks. This experimentation contributes to the broader understanding of the EfficientNetB3 model's artistic rendering capabilities through the lens of the deep dream algorithm.

<u>Neural Style Transfer</u>

The working of Neural Style Transfer can be found in the mechanism figure below and the steps include:

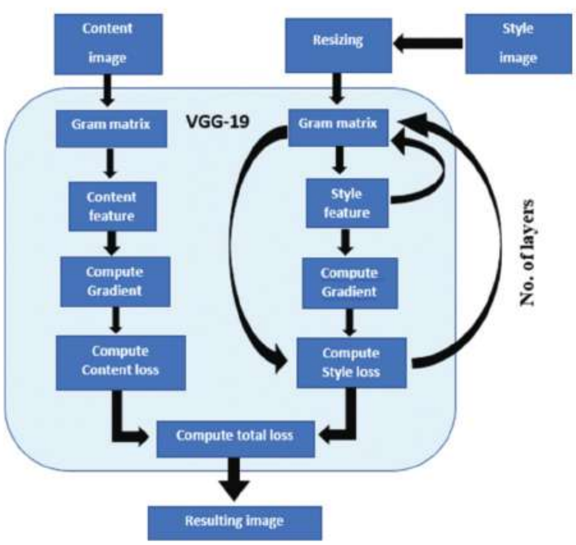


Fig. neural style transfer mechanism

1. Define the Content and Style Images: The choice of the content and style pictures to be used for the style transfer is the first stage. The content image is the one whose content we want to keep, but the style image is the one whose style we wish to borrow.

The style of an image is captured by the correlations between the different filter responses. The feature correlations are given by the Gram Matrix **Gl** ∈ **R^(Nl×Nl)** , where Gl ij is the inner product between the vectorized feature maps i and j at layer l:

$$l: G_{l,ij} = \sum_k F_{l,ik} F_{l,jk}$$

By including feature correlations of multiple layers, we capture the style information of the image while ignoring the content information(e.g. objects present in scenes, global arrangements of objects). Let oˆ and gˆ be the original image and the image that is generated respectively, let Ol and Gl be their style representations in layer l respectively, the contribution of layer l to the total style loss is given by :

$$S_l(\hat{o}, \hat{g}) = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (O_{l,ij} - G_{l,ij})^2,$$

where sl is a hyperparameter that specifies the weighting factor of the contribution of each layer(note that some sls could be 0, indicating that we don’t use the filter response of that layer

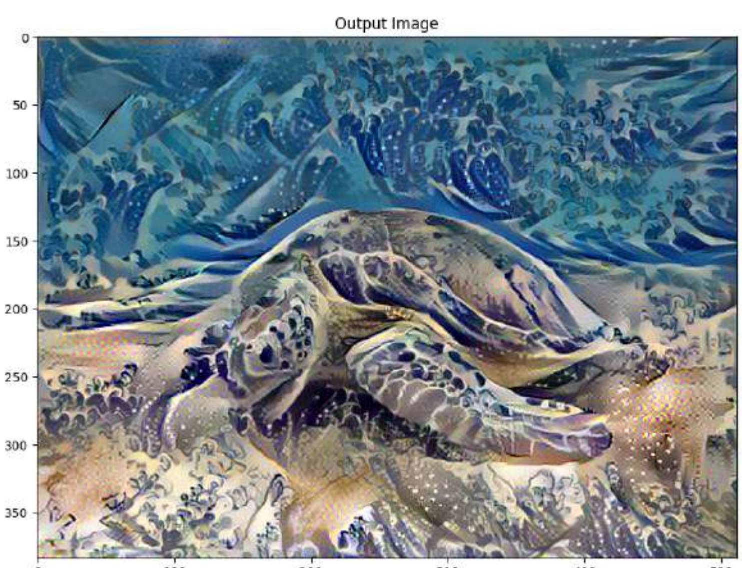
2. Define a Pre-Trained CNN: The following step is to create a pre-trained convolutional neural network (CNN), such as VGG19 or ResNet50 (we chose VGG19), that has been trained on a sizable dataset. The content and stylistic elements from the input photographs will be extracted using this CNN.

3. Extract Content and Style Features: The content and style characteristics are extracted from the content and style photos using CNN. The style features are collected from the network's intermediate levels, whereas the content features are extracted from one of the network's later layers.

4. Define a Loss Function: The output picture will be optimized using a loss function, which must be defined next. The loss function typically consists of two parts: a content loss that calculates the difference between the input image's content features and the output image's content features, and a style loss that computes the difference between the input image's style features and the output image's style features.

5. Optimize the Output Image: The generated picture must then be optimized using a loss function-minimizing technique, such as gradient descent.The optimisation technique modifies the output image's pixel values to reduce content and style loss.

The sequence of images depict a progression of the style transfer output at different stages of the iteration process. As the iterations increase, we see the style becoming more pronounced and the content details starting to blend more seamlessly with the stylistic elements, resulting in an image that combines the two in a visually pleasing manner.



In this final image, we can see how the texture and color patterns of the style have been applied to the sea turtle and its surroundings, creating a harmonious blend that retains the recognizability of the turtle while adopting the new artistic style. The process is guided by a loss function that is minimized during the iterations, resulting in this visually transformed image.

Combining NST and deep dream

In this study, our model combines deep dream and neural style transfer (NST) to produce a new image that combines the two technologies. VGG-19 and Inception v3 pre-trained networks are used for NST and deep dream, respectively. Gram matrix is a vital process for style transfer. The loss is minimized in style transfer while maximized in a deep dream using gradient descent for the first case and gradient ascent for the second. We found that different images produce different loss values depending on the degree of clarity of those images. Distorted images have higher loss values in NST and lower loss values with deep dreams. The opposite happened for the clear images.

Fig. Content image

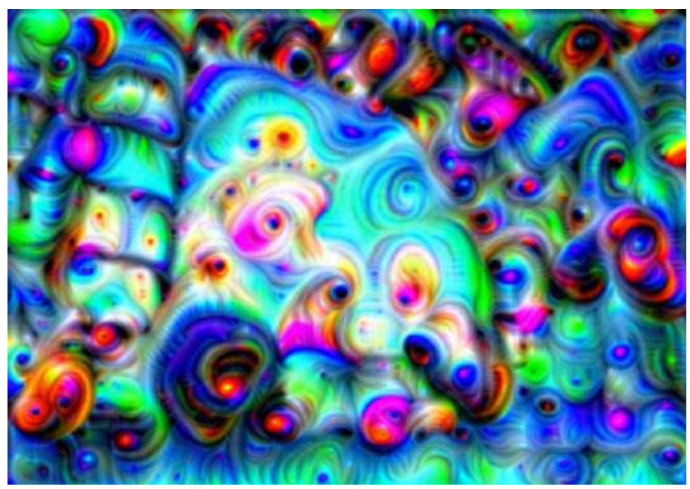
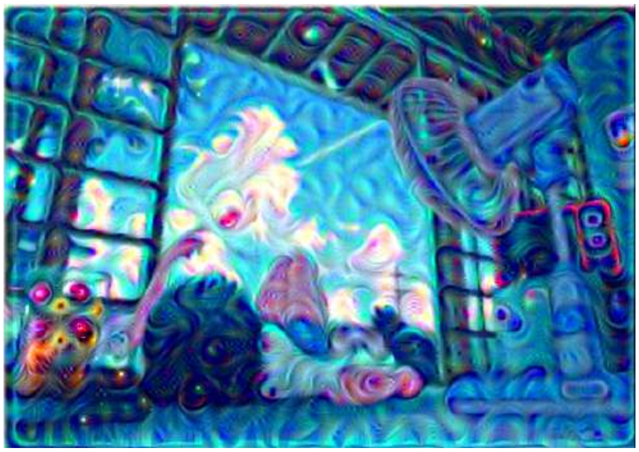


Fig. Style image

These are content and style images. We have taken the image generated from the deep dream algorithm as the style image. The below images showcase the output after 100 iterations with loss 7981 and 4000 iterations with loss5192



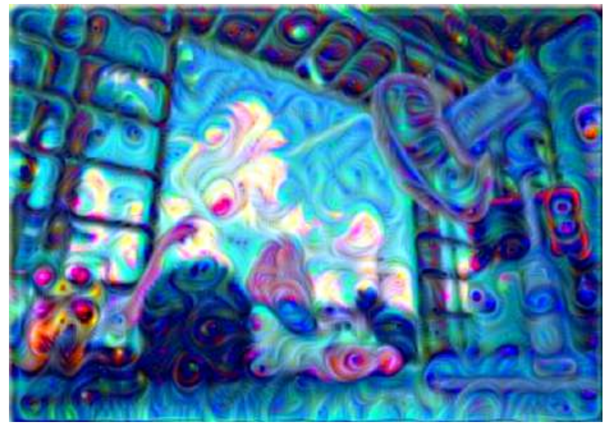


Fig. after 100 iterations (loss=7981.70)

Fig. after 4000 iterations (loss=5192.88)

The number of iterations refers to how many times the neural network adjusts its parameters in an attempt to minimize the loss. The loss is a mathematical representation of how far the current image is from the desired outcome — a successful blend of the style and content images.

From the results:

- After 100 iterations with a loss of 7891.70: This result suggests the early stages of the process. The neural network has begun adjusting its parameters to minimize the loss, but because the number of iterations is relatively low, the style may not be fully applied to the content image. The image looks less refined, with more evident features of the original content image and less influence from the style image.
- After 4000 iterations with a loss of 5192.88: This indicates a more advanced stage. With more iterations, the network has had more opportunities to refine the style transfer, leading to a lower loss value. The style is more cohesively blended with the content, and the details from the style image are more pronounced and harmonious with the content's structure.

The differences between these two stages illustrate the progression of learning and adjustment that the network undergoes. The final image after 4000 iterations shows a more seamless integration of style onto the content, with smoother transitions, more detailed style features, and a better overall aesthetic that combines both source images.

Conclusion

The document offers a comprehensive analysis of both the DeepDream algorithm and neural style transfer techniques, underscoring the potential and versatility of convolutional neural networks (CNNs) in advanced image processing. While DeepDream focuses on amplifying patterns and features inherent in the images, neural style
transfer emphasizes on applying the stylistic elements of one image to the content of another. Both techniques underscore the importance of layer selection in CNNs. Deep layers contribute to complex visualizations in DeepDream and more abstract style features in neural style transfer, while earlier layers retain more of the original content structure.

The experiments illustrate that both techniques are highly sensitive to parameter tuning, including learning rate, iteration numbers, and style-content weight ratios. This reinforces the need for careful experimentation to achieve the desired balance between style and content in neural style transfer, and between pattern complexity and recognizability in DeepDream.

The study not only shows the potential for creating artistic images but also suggests broader implications in areas like feature visualization, model interpretability and even in understanding human perception of art and patterns. The combined exploration of these two techniques underscores the growing intersection of art and AI, highlighting the creative possibilities that emerge when advanced machine learning techniques are applied to visual arts.

References

1.Lafta R. Al-Khazraji, Ayad R. Abbas, Abeer S. Jamil. A Hybrid Artistic Model Using Deepy-Dream Model and Multiple Convolutional Neural Networks Architectures. Retrieved from https://ieeexplore.ieee.org/abstract/document/10232970

2. Joaquim Silvestre, Yasushi Ikeda and Francois Guena. ARTIFICIAL IMAGINATION OF ARCHITECTURE WITH DEEP CONVOLUTIONAL NEURAL NETWORK. Retrieved from https://papers.cumincad.org/data/works/att/caadria2016_881.pdf

3. Emily L Spratt(2017). Dream Formulations and Deep Neural Networks: Humanistic Themes in the Iconology of the Machine Learned Image. Retrieved from https://www.academia.edu/42799744/Dream_Formulations_and_Deep_Neural_Networks_Humanistic_Themes_in_the_Iconology_of_the_Machine_Learned_Image

4. Keisuke Suzuki,Warrick Roseboom, David J. Schwartzman & Anil K. SethYear (2017). A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology. Retrieved from https://www.nature.com/articles/s41598-017-16316-2

5. LINGLING WANG, XINGGUANG DONG (2022). Image Style Conversion using Deep Convolutional Neural Network. Retrieved from https://wseas.com/journals/bae/2023/a265107-003(2023).pdf

6. Hongxu Yin; Pavlo Molchanov; Jose M. Alvarez; Zhizhong Li; Arun Mallya; Derek Hoiem (2020). Dreaming to Distill: Data-Free Knowledge Transfer via DeepInversion. Retrieved from https://ieeexplore.ieee.org/document/9156864

7. Xueyuan Liu (2021). Research on Neural Style Transfer. Retrieved from https://www.researchgate.net/publication/356022464_Research_on_Neural_Style_Transfer

8. Haochen Li (2020). A Literature Review of Neural Style Transfer. Retrieved from https://www.cs.princeton.edu/courses/archive/spring18/cos598B/public/projects/LiteratureReview/COS598B_spr2018_NeuralStyleTransfer.pdf

9. Keisuke Suzuki, Warrick Roseboom, David J. Schwartzman & Anil K. Seth (2017). A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology. Retrieved from https://www.nature.com/articles/s41598-017-16316-2

10. Lafta R. Al-Khazraji, Ayad R. Abbas , and Abeer S. Jamil. Employing Neural Style Transfer for Generating Deep Dream Images. Retrieved from https://www.semanticscholar.org/reader/778d5ac2447a7792d6674aa5976e49fba207f10e

11. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg & Li Fei-Fei (2015). ImageNet large scale visual recognition challenge. Retrieved from https://link.springer.com/article/10.1007/s11263-015-0816-y

12. Abdelhakim Zeroual, Mohamed Amroune, Makhlouf Derdour, Atef Bentahar (2022). Lightweight deep learning model to secure authentication in Mobile Cloud Computing. Retrieved from https://www.sciencedirect.com/science/article/pii/S1319157821002676

13. Berov, L., & Kühnberger, K. U. (2016). Visual hallucination for computational creation. In Proceedings of the Seventh International Conference on Computational Creativity (pp. 107–114).

14. Briot, J.-P., Hadjeres, G., Pachet, F.-D., & Al., E. (2017). Deep Learning Techniques for Music Generation - A Survey. arXiv preprint arXiv:1709.01620.

15. Halac, D., C., F. N., & Matiás. (2021). DREAMSOUND : DEEP ACTIVATION LAYER SONIFICATION. The 26th International Conference on Auditory Display (ICAD 2021), doi: doi.org/10.21785/icad2021.032

16. Jente Insing, Bonny Vloet (2018). Combining the Deep Dream and style transfer algorithm to create art. Retrieved from https://www.insingdesign.nl/pdf/deep-dream.pdf

17. Owan Evans (2019). Sensory Optimization: Neural Networks as a Model for Understanding and Creating Art. Retrieved from https://www.semanticscholar.org/reader/7f6be3fa062f68b589c0e28b2a7f24074da94fcd

18. Lafta R. Al-Khazraji, Ayad R. Abbas , and Abeer S. Jamil. Employing Neural Style Transfer for Generating Deep Dream Images. Retrieved from https://www.semanticscholar.org/reader/778d5ac2447a7792d6674aa5976e49fba207f10e

19. Authors Unknown. (Year Unknown). The Effect of Changing Targeted Layers of the Deep Dream Technique Using VGG-16 Model. Retrieved from https://online-journals.org/index.php/i-joe/article/view/37235/12845

20. Lafta R. Al-Khazraji1,2(), Ayad R. Abbas1, Abeer S. Jamil (Year Unknown). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions.Retrieved from https://www.semanticscholar.org/reader/778d5ac2447a7792d6674aa5976e49fba207f10e

21. Centralized Intelligent Authentication System Using Deep Learning with Deep Dream Image Algorithm.

22. Yue, D., & Samson, F. (2021). SleepTalk : Textual DeepDream for NLP Model Interpretability. Stanford CS224N Natural Language Processing with Deep Learning. Retrieved from https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1214/reports/final_reports/report041.pdf

Appendix

Code :
https://colab.research.google.com/drive/1y5MqqAZt4ZIetDy_dIj51K8lw37Y5IFW?usp=sharing