REPORT
NEURAL STYLE TRANSFER

SARTHAK SHAURYA

ECE

22116082

Introduction

Neural Style Transfer (NST) is a captivating technique in computer vision that allows for the creation of novel images by combining one image's content with another's artistic style. Imagine transforming a photograph of your family vacation into a masterpiece reminiscent of Van Gogh's vibrant brushstrokes! This project delves into the world of NST, aiming to construct a system capable of achieving such creative image manipulation.

Objectives

This project is centred around three key objectives:

1. **Model Construction:** The project entails building a neural network model specifically designed for NST. This model will leverage deep learning techniques to extract content and style information from input images and subsequently combine them to generate a new image.
2. **Hyperparameter Exploration:** NST involves various parameters that significantly influence the final outcome. The project aims to explore these hyperparameters, such as the number of training iterations (epochs) and the weights assigned to content and style losses. This exploration is crucial for achieving optimal style transfer results.
3. **Image Generation:** The ultimate goal is to generate images that effectively showcase the fusion of content and style. The generated image should faithfully retain the objects and scenes from the content image while being rendered in the artistic style of the chosen artwork.

Approach

To achieve the aforementioned objectives, the following steps were undertaken:

1. **Data Preprocessing:**
    - **Image Selection:** The process begins with selecting two crucial images: a content image (e.g., a photograph) and a style image (e.g., a painting). The choice of these images significantly impacts the final outcome. The content image ideally possesses clear and well-defined objects, while the style image should embody a distinct artistic style.
    - **Resizing and Preprocessing:** Both images are resized to a predetermined size suitable for the neural network architecture. This ensures consistency in the data fed into the model. Additionally, the images are converted from their raw pixel format into tensors. Tensors are numerical representations that can be processed by the neural network.
2. **Model Architecture:**
    - **Pre-trained VGG19 Model:** The model's core is the VGG19 network, a pre-trained convolutional neural network (CNN). VGG19 has been trained on a massive dataset of images (ImageNet) and excels at feature extraction. Convolutional layers in VGG19 act as filters that progressively extract features from the input image. Early layers capture low-level features like edges and textures, while later layers learn high-level features like shapes and objects.
    - **Feature Extraction:** The project leverages VGG19's feature extraction capabilities. However, the final classification layers of VGG19 are not used as the focus is not on image classification. Instead, specific layers within VGG19 are chosen to extract informative features for NST.
    - **Content and Style Representation:** Specific layers in VGG19 are chosen to extract content and style representations. The content representation captures the essential elements of the content image, while the style representation captures the artistic characteristics of the style image.

3. **Loss Functions:**
    - **Content Loss:** This loss function measures the discrepancy between the content features extracted from the content image and those extracted from the generated image. In simpler terms, it ensures that the generated image retains the key objects and scenes from the content image.
    - **Style Loss:** This loss function measures the difference between the style features extracted from the style image and those extracted from the generated image across chosen layers in VGG19. It essentially enforces the generated image to adopt the stylistic elements of the style image, such as brushstrokes, colour palettes, and textures.
    - **Total Loss:** A weighted sum combines the content loss and style loss. This allows for control over the balance between content preservation and style transfer. By adjusting the weights, one can prioritise either the content or the style during the image generation process.
4. **Optimisation:**
    - **Initialisation:** An initial random image is used as the starting point for the iterative optimisation process.
    - **Adam Optimizer:** The Adam optimiser, a popular optimisation algorithm in machine learning, minimises the total loss function. The optimiser iteratively updates the generated image to reduce the total loss.
    - **Loss Minimization:** During training, the optimiser adjusts the pixels of the generated image to bring the content and style features closer to those of the target content and style images, respectively. This effectively minimises the total loss and creates an image that blends content and style.
5. **Training and Evaluation:**
    - **Training Process:** The model is trained for a specified number of epochs. During each epoch, the model processes the content and style images, calculates the content and style losses, and updates the generated image using the Adam optimiser.

- **Monitoring Progress:** To monitor the training progress and assess the evolving style transfer, the following techniques are employed:

    - **Loss Visualization:** The total, content, and style loss are tracked and visualised over the training epochs. This visualisation provides insights into how the model balances content preservation and style transfer. A steady decrease in the total loss indicates successful training, while fluctuations or plateaus might suggest the need for hyperparameter adjustments.
    - **Image Generation at Intervals:** During training, the generated image is periodically saved at specific intervals. This allows for qualitative evaluation of the progress of style transfer. One can observe how the content and style information progressively merge by visually inspecting the generated images at different training stages.
- **Evaluation Criteria:** Once training is complete, the final generated image is evaluated based on the following criteria:
    - **Content Preservation:** The generated image should retain the essential elements and scene structure from the content image. Key objects and their spatial relationships should be recognisable.
    - **Style Transfer:** The generated image should exhibit the artistic style of the style image. This can be manifested in brushstroke patterns, colour palettes, and overall texture.
    - **Image Quality:** The generated image should be visually appealing and free of artefacts or noise introduced during the style transfer process.

Failed Approaches

During the project, several approaches were attempted but ultimately abandoned due to shortcomings:

- **Smaller Image Size:** Using a smaller image size for the content and style images resulted in blurry and poorly defined style transfer effects. Smaller images lack the necessary detail to capture intricate stylistic elements from the style image.
- **Higher Learning Rate:** Experimenting with a higher learning rate for the Adam optimiser led to unstable training. A high learning rate can cause the model to

overshoot the optimal solution, resulting in significant fluctuations in the loss function and potentially generating nonsensical images.

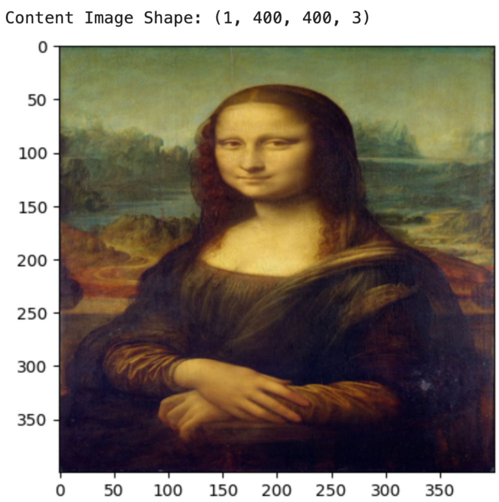
- **Excessive Style Layers:** Including many layers in the style loss calculation can overwhelm the content information. When too many style layers are incorporated, the generated image might become overly dominated by the style image's style, losing resemblance to the content image.

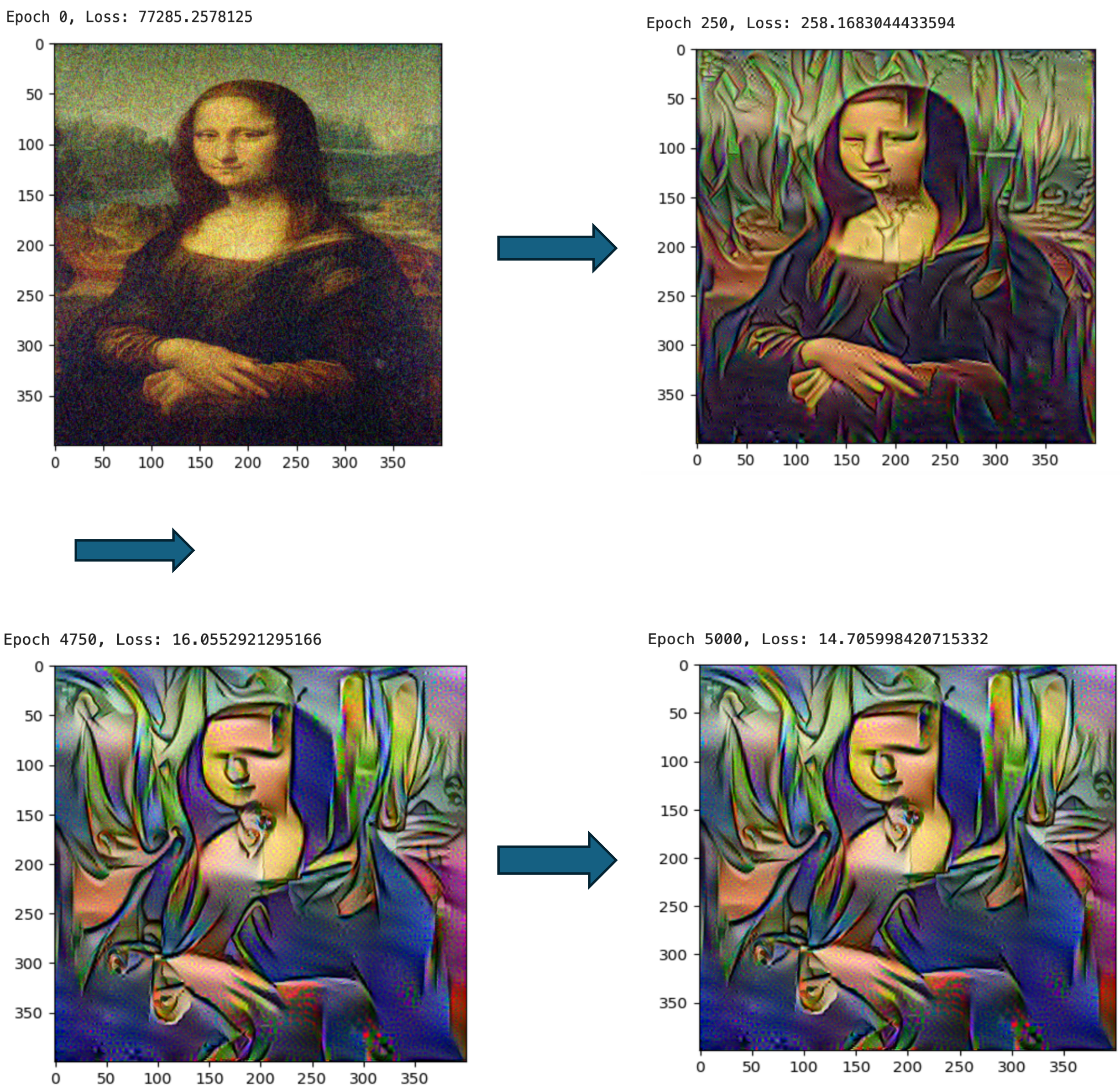
These failed approaches highlight the importance of careful hyperparameter tuning and the need to balance capturing style information and preserving content details.

Result

The project built a neural network using a pre-trained VGG19 model to achieve neural style transfer. By training on content and style image pairs, the model successfully generated images that retained the content while incorporating the style image's artistic style. Hyperparameter tuning allowed for control over the style transfer process, resulting in visually appealing images that balanced content preservation and artistic transformation. This project demonstrates the effectiveness of NST and opens doors for further exploration in creative image generation.

Example





<u>**Discussion**</u>

The project successfully implemented neural style transfer, demonstrating the effectiveness of this technique in fusing content and style information from different images. The selection of content and style images significantly impacts the final outcome. Using high-quality images with distinct styles leads to more compelling results. Adjusting hyperparameters, such as the number of epochs and style layer weights, allows for control over the style transfer process.

Future Improvements

While the project achieved successful style transfer, there is room for further exploration and improvement:

- **Spatial Control:** Current NST techniques often apply style transfer uniformly across the entire image. Future work could explore methods for incorporating spatial control, allowing for selective style transfer in specific image regions. This would create images where certain elements retain their original style while others are transformed with a different artistic style.
- **High-Resolution Images:** The project primarily focused on images of a specific size. Future exploration could involve investigating techniques for handling high-resolution images. This would enable the generation of high-quality style-transferred images suitable for printing or large displays.
- **Alternative Architectures:** The project utilised the VGG19 model for feature extraction. Future work could explore alternative pre-trained models like ResNet or InceptionV3. These models might offer advantages in capturing different style aspects or potentially lead to more nuanced style transfer effects.

Conclusion

This project successfully implemented neural style transfer using a pre-trained VGG19 model and carefully designed loss functions to combine content and style information. The results demonstrate the potential of NST for artistic image manipulation. Further exploration of hyperparameters, model architectures, and advanced techniques can lead to even more creative and expressive applications of neural style transfer.

References

- Coursera Deep Learning Course- Module on CNN.
- Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). A neural algorithm of artistic style transfer. arXiv preprint arXiv:1605.08807. https://arxiv.org/abs/1605.08807

- Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for image counter-fitting. arXiv preprint arXiv:1603.08175. [http://arxiv.org/abs/1603.08155](http://arxiv.org/abs/1603.08155
- Huang, X., & Belongie, S. (2017). Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization. https://arxiv.org/pdf/1703.06868.pdf
- Ghiasi, G. , Lee, H., Kudlur, M. et. al. (2017). Exploring the structure of a real-time, arbitrary neural artistic stylization network. https://arxiv.org/pdf/1705.06830.pdf