ArIES Open Project
NEURAL STYLE TRANSFER

Anirudha Thakur - 23112016

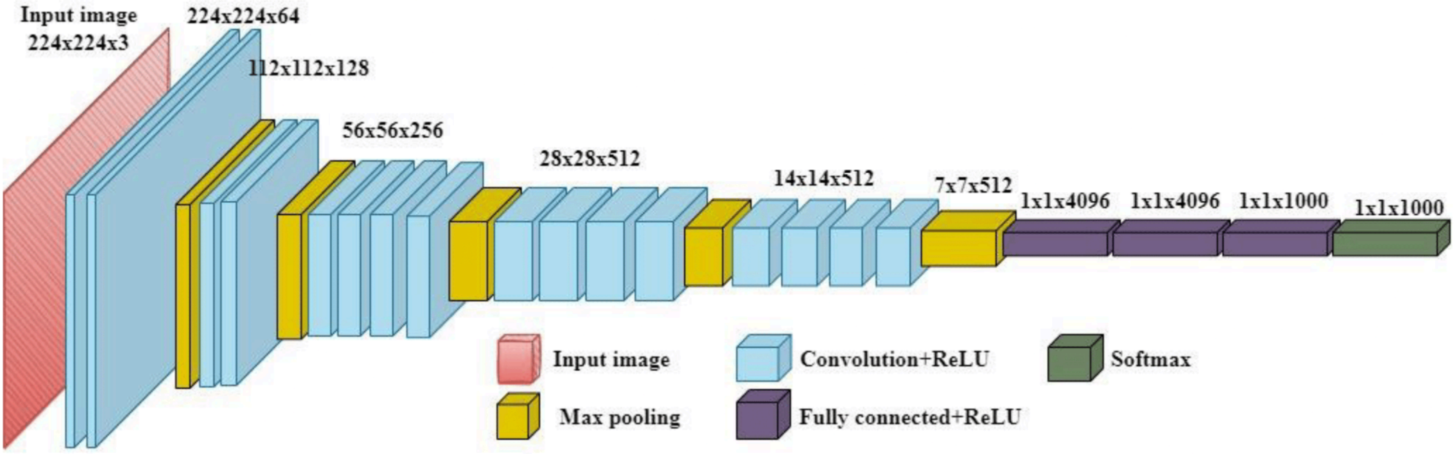
1. Introduction

Neural Style Transfer (NST) is a captivating application of convolutional neural networks (CNNs) that synthesises a new image by combining the content of one image with the artistic style of another. The primary goal of NST is to create an image that maintains the semantic content from a content image while adopting the stylistic features of a style image. This project aims to implement NST using a modified VGG19 network to extract necessary features and optimise a generated image through iterative training.

2. Methodology

- **Network Architecture**

The **VGG19** model, known for its depth and performance in feature extraction tasks, was selected for this NST project. The model was pre-trained on the ImageNet dataset, providing a robust starting point for extracting high-level features. For the purpose of NST, the VGG19 network was truncated after the 30th convolutional layer to obtain intermediate features essential for content and style representation.



- **Image Preprocessing**

Images were preprocessed to fit the input requirements of the VGG19 model. This preprocessing involved resizing the images to a uniform size, normalising pixel values to match the distribution the network was trained on, and converting the images to tensor format for compatibility with PyTorch.

- **Feature Extraction**

Feature extraction was carried out by passing the images through the truncated VGG19 network. Different layers of the network capture varying levels of abstraction, with initial layers focusing on low-level features like edges and textures, while deeper layers capture high-level semantic content. For content representation, features were extracted from one of the deeper layers, and for style representation, features were extracted from multiple layers to capture a comprehensive set of stylistic elements.

- **Loss Functions**

Two primary loss functions were used to guide the optimization process:

1. **Content Loss**: This measures the difference between the feature representations of the content image and the generated image. The mean squared error was used to quantify this difference, ensuring that the generated image retains the semantic content of the original content image. This is mathematically represented as:

$$\mathcal{L}_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} \left(F^l_{ij} - P^l_{ij}\right)^2$$

2. **Style Loss**: This captures the difference in style between the style image and the generated image using Gram matrices, which represent the correlations between different feature maps. By comparing Gram matrices of the style and generated images, the style loss ensures that the generated image adopts the stylistic features of the style image.

The Gram matrix is represented by:

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l}.$$

Total Loss function: To blend the content of a photograph with the style of a painting, the total loss is a combination of the content and style losses:

$$\mathcal{L}_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{content}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{style}(\vec{a}, \vec{x})$$

where α and β are the weighting factors for content and style reconstruction respectively.

- **Optimization Process**

The generated image was initialised as a copy of the content image and iteratively optimised using the Adam optimizer. The total loss, combining content and style losses, was minimised through gradient descent. The weights for content and style losses were adjusted to balance the contribution of each to the final image.

3. Results

- **Training Progress**

The optimization process was carried out over 4000 epochs. The total loss, comprising content and style losses, showed a decreasing trend, indicating successful adaptation of the generated image towards the desired content and style.

- **Visual Results**

Figures below display the original content and style images alongside the generated image at different stages of training. The generated image progressively transitions from resembling the content image to incorporating stylistic elements of the style image.

Original Content Image:

Original Style Image:

Generated Image After 1000 Epochs:

Generated Image After 2000 Epochs:

Generated Image After 3000 Epochs:

Generated Image After 4000 Epochs:

4. Discussion

The NST implementation demonstrated the capability to blend content and style effectively. Key observations include:

- **Content Retention**: The generated image maintained the structural and semantic integrity of the content image throughout the optimization process.
- **Style Adoption**: The stylistic features from the style image were progressively transferred to the generated image, with varying degrees of success across different layers.
- **Optimization Challenges**: Balancing the content and style weights proved critical in achieving a visually pleasing result. Higher style weights led to better stylistic adaptation but sometimes at the expense of content clarity.

5. Conclusion

This NST project successfully implemented a neural style transfer algorithm using a modified VGG19 network. The generated images effectively combined the content of one image with the style of another, demonstrating the potential of CNNs in creative applications. Future work could explore alternative architectures, different optimization strategies, and real-time NST applications.

6. References

- https://arxiv.org/pdf/1508.06576