Interpreting Neural Networks in Artistic Style Transfer

Authors: Srikar Konakanchi, Liam Banahan, Christian Montano, Adisri Rajkumar

Abstract

This project investigates neural style transfer, a machine learning technique that creates artistic images by merging the content of one image with the artistic style of another. Using a pre-trained convolutional neural network (VGG-19), we extract layered visual features and blend them through optimized loss functions, including content, style, and total variation losses. The outcome demonstrates neural networks' capacity to synthesize visually compelling artwork and provides clear insight into their complex internal processes.

Introduction

Neural networks, especially Convolutional Neural Networks (CNNs), have transformed computer vision and digital art through hierarchical image feature extraction, yet their inner workings often remain opaque to newcomers. Understanding how CNNs encode visual information raises an important question: How can we visually explain the complex internal processes of CNNs to make them more accessible and intuitive?

Objectives

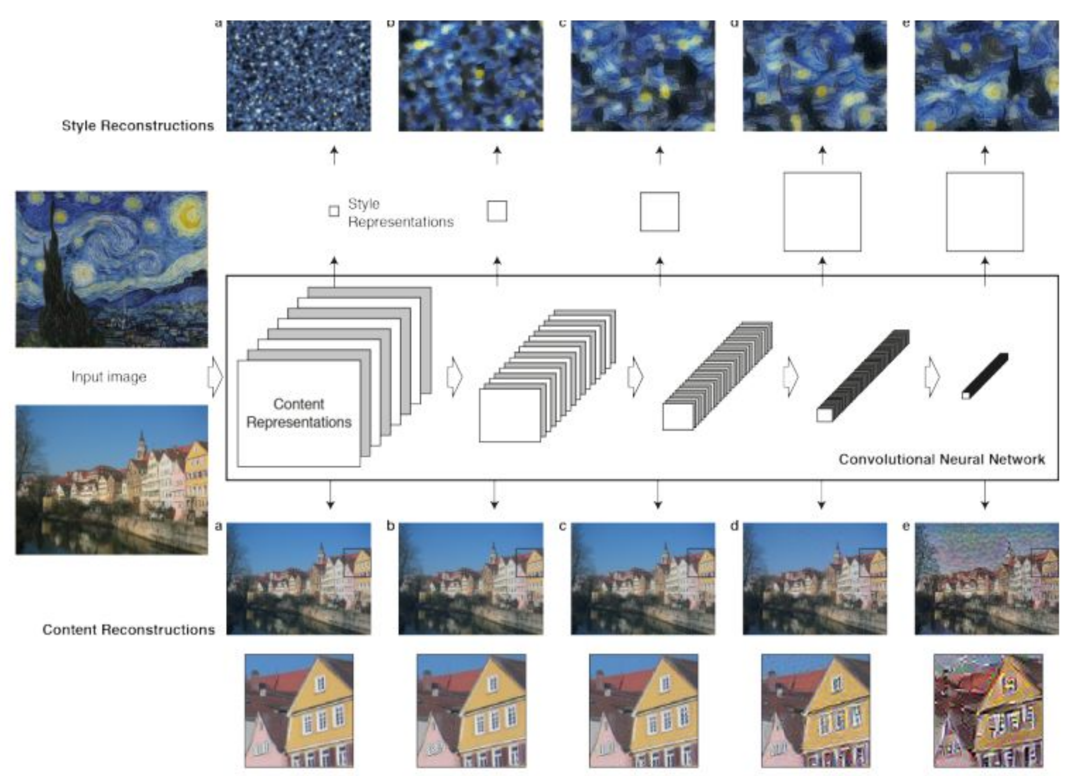
* Use an interactive style transfer setup to visualize how CNNs process and transform images.

* Demonstrate how CNNs separate and represent content (structure) and style (texture, color).

* Explain neural network behavior through loss functions that guide the image generation process.

* Create an engaging educational tool that makes deep learning concepts more accessible

VGG-19 Structure



Theory

- Content details (shapes, structure)
- Style patterns (colors, textures)

Lower layers of VGG-19 capture style (textures, colors), while deeper layers capture content (shapes, structure), enabling precise control in neural style transfer.

Visualizing these layers guides meaningful tuning of loss weights (e.g., α/β, wl), making the process more interpretable and customizable.

Content Loss: How different is it from the photo?

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

Style Loss: How different is it from the painting?

$$L_{content}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^{l} - P_{ij}^{l})^2$$

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^{l} - A_{ij}^{l})^2$$

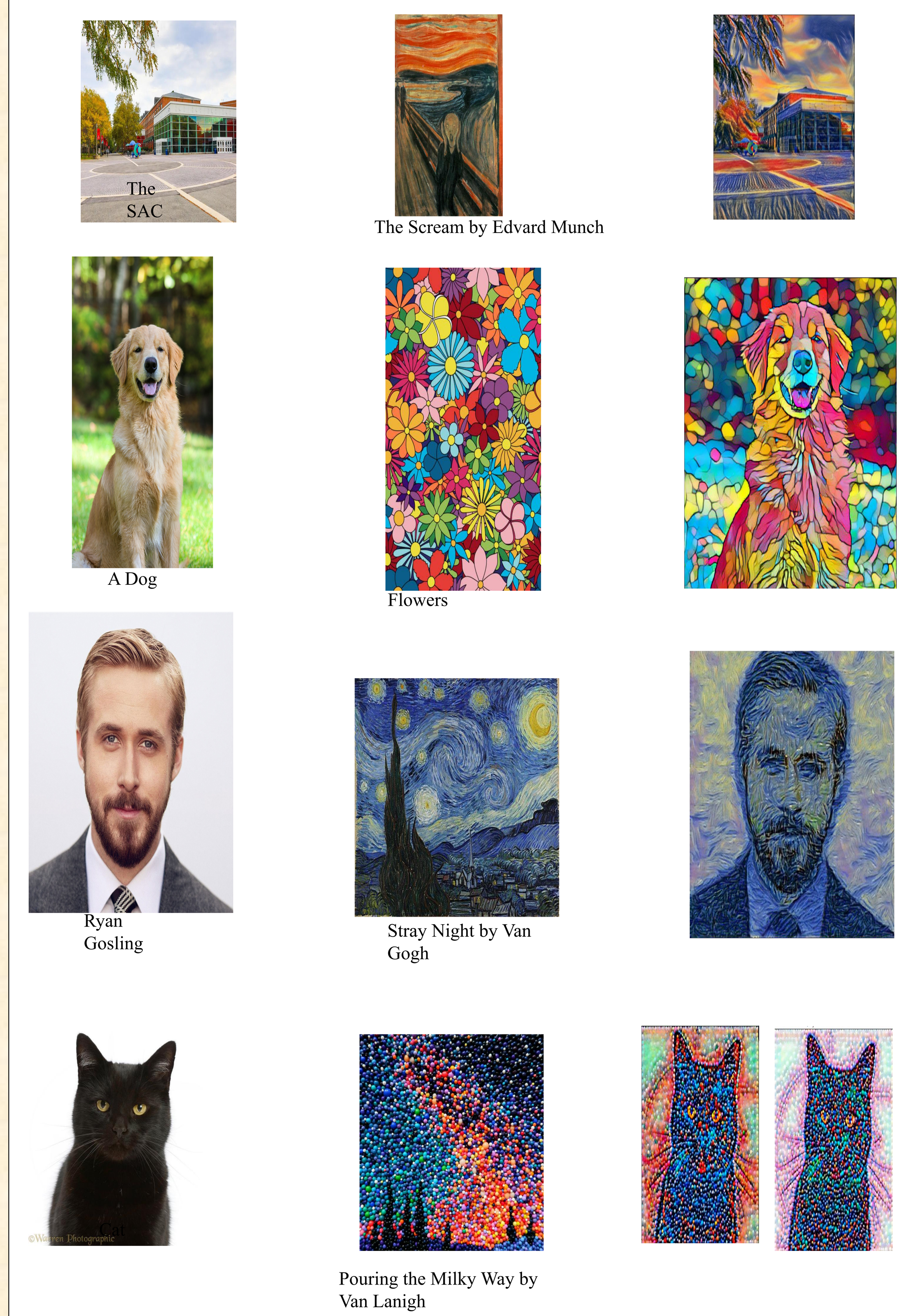
$$L_{style}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l$$

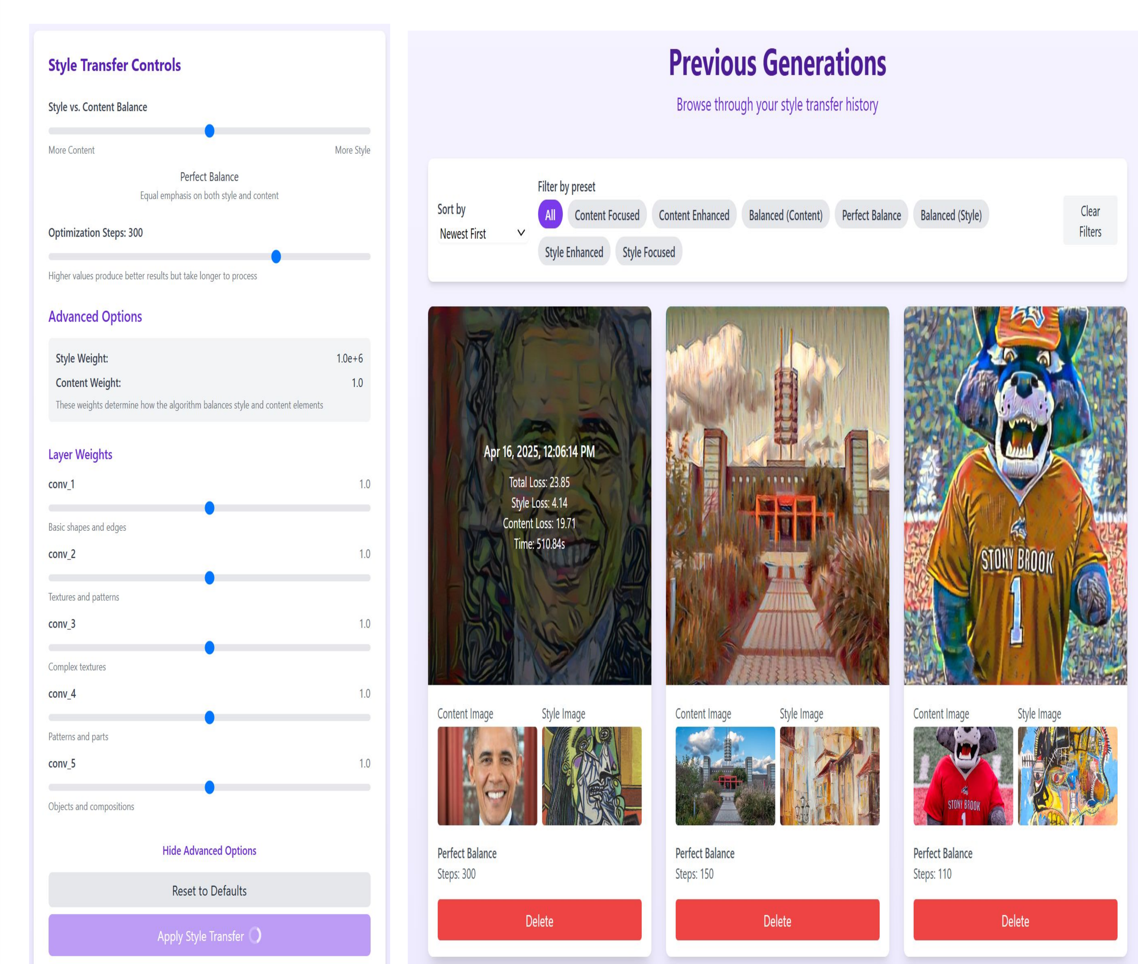
Computed weighted total loss and minimized using L-BFGS Optimization

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

Gatys, 12

Results





Conclusions

Limitations

- Style transfer requires powerful hardware (GPU) and can be slow, especially for high-res images.
- In some cases, the final image may lose key details from the original photo if the style is too strong.
- Traditional neural style transfer takes time and is not suited for live video or interactive applications without heavy optimization.
- If style images come from limited artists or cultures, the model may not generalize well or reflect diverse artistic styles.

Final Thoughts & Discussion

Our work showcases the creative potential of deep learning by applying neural style transfer to video frames. While challenges such as processing time and maintaining style coherence remain, the project bridges art and AI, opening doors for innovation in media production, content creation, and personal expression through technology.

References

Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A Neural Algorithm of Artistic Style." arXiv, 2 Sept. 2015, https://arxiv.org/abs/1508.06576.

Li, Yanghao, et al. “Demystifying Neural Style Transfer.” *arXiv.Org*, Cornell University, 1 July 2017, arxiv.org/abs/1701.01036.

Developers Hutt. (n.d.). Why Gram Matrix in Style Transfer || Quick Explained. https://www.youtube.com/watch?v=Elxnzxk-AUk.