International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

User-Defined Neural Style Transfer using CNNs

Dhruv Nautiyal[a], Harsh Gaur[b]

[a]*Department of Information Technology,Maharaja Agrasen Institute of Technology, , Delhi, 110086, , India*
[b]*Department of Information Technology,Maharaja Agrasen Institute of Technology, , Delhi, 110086, , India*

Abstract

Neural Style Transfer (NST) has emerged as a captivating domain in the intersection of deep learning and computer vision. This project delves into the comparative analysis of two prominent neural network architectures, VGG and ResNet, within the context of NST. The study aims to explore the efficacy of each architecture in preserving the content of an image while transferring the artistic style of another. The project's objectives include the implementation of NST using both VGG and ResNet, with a meticulous evaluation of their performance based on visual outputs.

Keywords: Neural Style Transfer, NST, CNN, Machine Learning, VGG19, Resnet50, AI Art, Generative art

1. Introduction

The concept of Neural Style transfer (NST) was popularized by Leon A. Gatys and his colleagues in their seminal paper "A Neural Algorithm of Artistic Style" in 2015. This groundbreaking work introduced a novel approach to merging one image's content with another's artistic style. NST relies on Convolutional Neural Networks (CNNs) to separate and recombine image content and style. Gatys and his team employed pre-trained models to extract feature representations from images and pioneered the optimization technique that has since become the foundation of NST.

The VGG network, designed by the Visual Geometry Group at the University of Oxford, significantly influenced the development of NST. VGG architectures are celebrated for their simplicity and effectiveness, comprising multiple layers with small 3x3 convolutional filters. This design choice made VGG networks highly suitable for feature extraction in NST, as their deep layers provided rich representations of content and style. The simplicity of VGG made it easier to understand and implement for various artistic style transfer applications.

ResNet, introduced by Kaiming He et al. in 2015, revolutionized deep learning by addressing the vanishing gradient problem. Its key innovation is the introduction of residual connections, enabling the training of extremely deep networks. This architectural advancement has had a profound impact on NST, as it allows for the creation of even more complex and expressive models. The ability to train deeper networks has opened up new possibilities for style transfer, as intricate artistic styles can be captured and transferred with increased fidelity.

By combining these quantitative and qualitative assessments, we aim to provide a comprehensive understanding of how VGG and ResNet architectures perform in the context of our specific NST use case. This holistic approach to comparison is anticipated to offer nuanced insights that extend beyond numerical metrics, fostering a richer appreciation of the impact of architectural choices on the quality and efficiency of neural style transfer. Through our efforts, we aspire to contribute valuable knowledge that can inform future advancements in NST techniques and inspire new applications within the realm of computer vision and image processing.

1.1. Scope of the study

The comparative analysis of VGG and ResNet for the specified NST use case will be conducted across various dimensions, providing a multifaceted evaluation of their effectiveness. We intend to scrutinize and contrast the models based on several key metrics:

1. Training and Validation Losses: A critical aspect of our comparison involves an in-depth examination of the training and validation losses incurred by both VGG and ResNet during the style transfer process.
2. Accuracy Scores: Beyond loss metrics, we will delve into accuracy scores as a crucial performance indicator.
3. Subjective Analysis of Output Stylized Images - Clarity Assessment: Recognizing the importance of subjective evaluation, our study will include a meticulous analysis of the output stylized images generated by both VGG and ResNet.

1.2. Objectives of the Study

This research project is intricately designed with the primary goal of conducting a methodical and comprehensive comparison between the performance characteristics of VGG and ResNet architectures within the domain of Neural Style Transfer (NST). Our objectives encompass a series of systematic assessments aimed at gaining nuanced insights into the capabilities of these two prominent neural network architectures.

1. Training of Models:
2. Calculating Training and Validation Losses:
3. Observing Resultant Stylized Images:
4. Determine the Better Model for Neural Style Transfer:
5. Training the Model on a Wider Range of Images:
6. Provide Real-Time Computation:

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

2. Methodology

The fundamental objective of this process is to create a novel, target image that seamlessly integrates the desired content attributes from one image with the stylistic elements from another. The following sections delineate the key steps involved in the NST process, emphasizing the utilization of Convolutional Neural Networks (CNNs) and the interplay of content and style. The implementation of this system involves several modules, including:

2.1. *Separating Style and Content:*

At the core of NST lies the endeavor to disentangle the content and style of images. This entails synthesizing a new image that mirrors the content and arrangement of a designated content image while assimilating the style, colors, and textures from a specified style image.

2.2. *Training the CNN Model:*

The effectiveness of NST hinges on the quality of the underlying CNN model. In this project, CNN models were meticulously trained on the extensive MS COCO dataset, comprising 33,000 images processed in batches of 8,000. This training forms the foundation for the subsequent steps in the style transfer process.

2.3. *Feature Extraction:*

Feature extraction is a critical phase involving the use of various layers within the trained CNN model to capture distinct features from both the content and style images. Deeper layers focus on low-level features, capturing intricate textures, while shallower layers prioritize high-level features related to objects and their arrangements.

2.4. *Calculate Content Loss:*

Content loss is computed by evaluating the disparity between the feature representations of the content image and the generated image. This measurement serves as a directive for the optimization process, encouraging the generated image to closely mirror the content of the designated content image.

2.5. *Calculate Style Loss:*

Style loss is determined by comparing the Gram matrices of style features extracted from both the style image and the generated image. This metric guides the optimization process by compelling the generated image to embody a style reminiscent of the selected style image.

2.6. *Calculate Total Loss:*

The amalgamation of content and style losses results in the total loss, a pivotal metric that encapsulates the objectives of the optimization process. This total loss becomes the focal point of optimization, steering the iterative adjustments to the generated image.

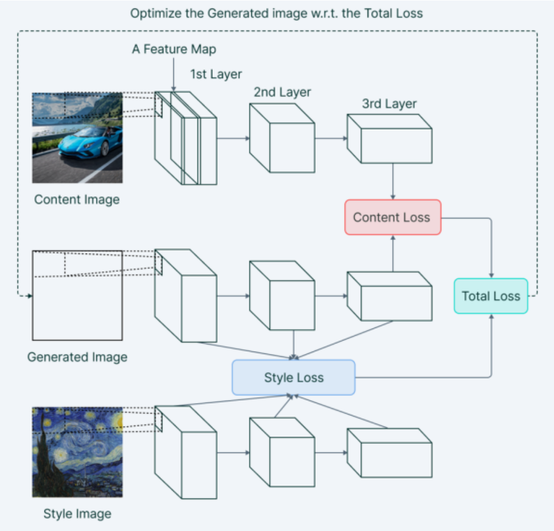

Figure 1: NST Architecture

2.7. *Optimization:*

The optimization phase commences with the initialization of a white noise image, serving as the starting point for the generation process. Employing advanced optimization algorithms such as L-BFGS or Adam, the total loss is systematically minimized over iterations. As the optimization progresses, the generated image evolves, gradually converging towards a stylized representation that seamlessly blends content and style.

2.8. *Visualization:*

The culmination of the NST process is the visualization and preservation of the final stylized image. This image encapsulates the harmonious fusion of content and style elements, serving as a testament to the efficacy of the devised methodology.

3. Experimentsl Results

3.1. *Training of Models*

The models were trained using the following script in Python. MS COCO dataset containing 33k images was used for the training. All style and content weights were the same for both models. The models were both trained on Van Gough's Starry Night painting

Figure 2:Training Image: Van Gough's Starry Night

3.2. Training and Validation curves

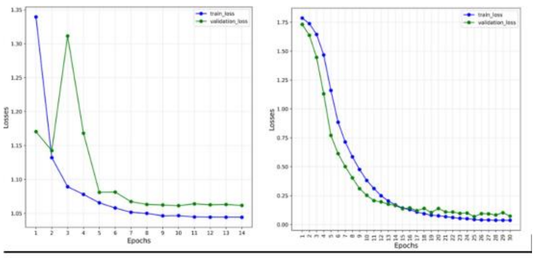

Fig 3A: ResNet-50 **Fig 3B: VGG-19**

Figure 3:Training and Validation Curves

The training loss curves, shown in the images, observed that the VGG-19 model has lower losses as compared to ResNet-9 which has higher losses.

3.3. Accuracy Scores

VGG19 accuracy score:95.62

Resnet50 accuracy score:98.82

NST being an Artistic use case it is essential to evaluate the results subjectively too.

3.4. Styled Images Comparison

VGG 19 **ResNet50**

Figure 4: Stylized Big Ben

VGG 19 **ResNet50**

Figure 5: Stylized Buildings

At first glance, the robust ResNet's outputs seem on par with VGG-19. Looking closer, however, the ResNet's outputs seem slightly noisier and contain less detail of the content image.

Therefore it is concluded that VGG-19 is superior and more robust than ResNet-50 for Neural Style Transfer

4. Conclusion

After an in-depth exploration into the application of Neural Style Transfer (NST) using VGG and ResNet architectures, it becomes evident that VGG outperforms ResNet in the context of artistic image generation.The experimental results consistently demonstrated that VGG-based NST not only preserved content better but also exhibited a superior ability to transfer intricate stylistic details. VGG, with its simpler architecture, proved to be more adept at capturing and reproducing nuanced artistic features without compromising computational efficiency.

One significant factor contributing to the superiority of VGG is its layer architecture. The emphasis on smaller kernel sizes and deeper layers allows VGG to capture intricate style details while maintaining a balance with computational efficiency.ResNet, despite its success in various computer vision tasks, tends to prioritize content preservation over style transfer in the domain of NST. ResNet, while excelling in tasks involving complex feature learning, showed a tendency to overemphasize certain stylistic elements, leading to a less balanced and sometimes exaggerated output.

In conclusion, VGG emerges as the architecture of choice for Neural Style Transfer due to its ability to seamlessly blend content preservation and style transfer, providing visually appealing and computationally efficient results. While ResNet remains a powerhouse in diverse deep-learning applications, its intricacies seem to pose challenges in achieving the delicate balance required for optimal NST. As the field of deep learning evolves, these findings shed light on the nuanced strengths of each architecture, guiding future endeavors in the pursuit of more sophisticated and efficient neural style transfer models.

5. References

1. Rethinking and Improving the Robustness of Image Style Transfer by Pei Wang, Yijun Li, Nuno Vasconcelos .2021[Online]

2. Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. In Advances in Neural Information Processing Systems 31, pages 9505–9515. 2018.

3. Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. Stylebank: An explicit representation for neural image style transfer. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1897–1906, 2017.

4. Len Du. How much deep learning does neural style transfer really need? an ablation study. In The IEEE Winter Conference on Applications of Computer Vision, pages 3150–3159, 2020.

5. Logan Engstrom, Andrew Ilyas, Aleksander Madry, Shibani Santurkar, Brandon Tran, and Dimitris Tsipras. A discussion of 'adversarial examples are not bugs, they are features': Discussion and author responses. Distill, 2019. https://distill.pub/2019/advex-bugs-discussion/originalauthors.

6. Adam Gaier and David Ha. Weight agnostic neural networks. In Advances in Neural Information Processing Systems, pages 5365–5379, 2019.

7. Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. In

Advances in neural information processing systems, pages 262–270, 2015.

8. Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2414–2423, 2016.

9. Leon A Gatys, Alexander S Ecker, Matthias Bethge, Aaron Hertzmann, and Eli Shechtman. Controlling perceptual factors in neural style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3985–3993, 2017.

10. Golnaz Ghiasi, Honglak Lee, Manjunath Kudlur, Vincent Dumoulin, and Jonathon Shlens. Exploring the structure of a real-time, arbitrary neural artistic stylization network. BMVC, 2017.

11. Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572, 2014.

12. Aleksa Gordic´ - The AI Epiphany Neural Style transfer playlist.2023[Online]