

# A comparative analysis of VGG19 and Magenta models for Neural Style Transfer

**Wenhe Xu**

Department Of Computer Science, University of Liverpool, Brownlow Hill,  
Liverpool, L69 7ZB, United Kingdom

sgwxu18@liverpool.ac.uk

**Abstract.** In recent years, artificial intelligence (AI) has witnessed remarkable progress, with deep learning techniques playing a pivotal role in this transformative journey. One prominent facet of this progress is Neural Style Transfer (NST), a domain within computer vision and image processing. NST entails merging the content of one image with the stylistic attributes of another, resulting in a seamless fusion of content and style. This paper presents a comprehensive comparative analysis of two influential style transfer models, VGG19 and Magenta, shedding light on their strengths and limitations. VGG19, a well-established convolutional neural network, excels in feature extraction, dimensionality reduction, and image classification, making it ideal for computer vision tasks. In contrast, Magenta, a creative content generation framework, specializes in producing innovative content, such as music and art, while promoting human-AI collaboration. Evaluation reveals that Magenta's generated images tend to exhibit softer colors but higher clarity, while VGG19 produces vibrant and faithful color reproductions with finer details. The choice between these models should be driven by specific task requirements. VGG19 is suitable for tasks emphasizing color fidelity, while Magenta excels in image clarity. Future research directions may involve enhancing algorithmic capabilities to harness the strengths of both models and improving efficiency in style transfer techniques. This analysis empowers researchers, practitioners, and enthusiasts to make informed decisions when selecting a model aligned with their objectives, contributing to the evolving landscape of AI-assisted creative content generation.

**Keywords:** Neural Style Transfer, Convolutional Neural Network, Computer Vision.

## 1. Introduction

Recent years have witnessed significant strides in the realm of artificial intelligence (AI), with deep learning techniques assuming a pivotal role in this transformative journey [1]. Among these techniques, Neural Style Transfer (NST) has emerged as a prominent and influential domain within computer vision and image processing [2]. NST involves the intricate task of merging the content of one image with the stylistic attributes of another, giving rise to images that exhibit a seamless amalgamation of content and style. Fundamentally, NST harnesses the capabilities of Convolutional Neural Networks (CNNs) for feature extraction while integrating advanced image optimization principles into its framework [3,4].

The applications of style transfer are manifold and extend across creative and practical domains. Primarily, it serves as a powerful tool in art creation and image editing, providing artists and designers

with a means to explore and merge diverse styles creatively. In the ever-evolving entertainment industry, style transfer contributes significantly by enabling the creation of striking special effects and enhancing post-processing techniques in film production and game development. However, the utility of style transfer transcends mere artistic expression, as it holds substantial promise in the realms of medical image analysis, automated image enhancement, and the burgeoning field of virtual reality [5,6].

This paper's central objective is to undertake a comprehensive comparative analysis of two widely recognized style transfer models: VGG19 and Magenta. VGG19, a venerable deep neural network architecture, has established itself as a stalwart in style transfer applications, primarily due to its robust feature extraction capabilities [7]. Conversely, Magenta, a more recent entrant into the field, stands out for its innovative approaches to artistic style transfer and creative image generation [8]. By conducting a thorough analysis of these models, this research aims to clarify their individual strengths and weaknesses, thus providing valuable insights into how they can be applied effectively in various fields.

In the forthcoming sections of this paper shall delve into the technical intricacies of VGG19 and Magenta. The exploration will encompass a meticulous analysis of their architectural nuances, training methodologies, and practical utility. Furthermore, the paper shall present a comparative assessment of their capabilities, delineating the advantages they offer and the constraints models impose within various application scenarios. This exhaustive scrutiny seeks to contribute to the burgeoning body of knowledge surrounding style transfer techniques, empowering researchers, practitioners, and enthusiasts to make informed decisions when selecting a model best suited to their specific requirements. This article will elucidate two distinct models for the purpose of style transfer: VGG19 and Magenta.

## 2. Method

The forthcoming sections will delve into the foundational aspects of VGG19 and Magenta, elucidating their fundamental principles and functionalities.

VGG19, a prominent deep learning architecture, encompasses 16 convolutional layers, 3 fully connected layers, and 5 pooling layers, forming a structured design pivotal for feature extraction and classification. At its core, VGG19 utilizes convolutional layers to extract intricate image features, followed by the application of the Rectified Linear Unit (ReLU) activation function to capture complex patterns. The integration of maximum pooling layers reduces dimensionality while preserving spatial features, enhancing computational efficiency and robustness. Subsequently, a fully connected layer connects abstract features crucial for classification and predictions, culminating in a final layer, typically featuring the softmax activation function, which translates outputs into a probability distribution for precise multi-class classification. In essence, VGG19's architecture excels in feature extraction, nonlinearity, dimensionality reduction, abstract feature composition, and classification, rendering it a fundamental component in deep learning across diverse applications.

Magenta's core principle revolves around the generation of creative content, particularly music and art, using deep learning and neural network techniques. It employs models such as recurrent neural networks (RNNs) and CNNs to capture and understand patterns in existing artistic works [9]. These trained models can then generate novel compositions by learning from a vast dataset of artistic examples.

Magenta's unique aspect is its emphasis on human-AI collaboration, encouraging artists and musicians to work alongside the AI models to co-create expressive and innovative content. This collaborative approach aims to enhance the creative process and push the boundaries of what is possible in artistic expression through AI assistance. VGG19 and Magenta exhibit significant differences in their principles, primarily in terms of their tasks, architectures, and application domains. Firstly, VGG19 is a classical convolutional neural network primarily designed for image classification and feature extraction. It utilizes deep layers of convolution, fully connected layers, and pooling layers to learn image features, commonly applied for image recognition and categorization. In contrast, Magenta is a creative content generation framework with tasks encompassing the generation of music, art, and other creative content. Its architecture typically includes recurrent neural networks and convolutional neural networks for generating creative works.

Secondly, these two models differ in their architectures and accepted data types. VGG19 employs a deep convolutional neural network that primarily takes image data as input, focusing on image classification and feature extraction. Typically, VGG19 requires labeled image data for supervised learning. On the other hand, Magenta's architecture is more versatile, capable of accepting various types of input data, including audio or image data, for the generation of music, images, or other creative content. It places a stronger emphasis on generative tasks rather than supervised learning.

Lastly, these models diverge in their application domains. VGG19 finds its primary application in the field of computer vision, including tasks such as image recognition, object detection, and image segmentation. In contrast, Magenta's applications are more diverse, spanning music generation, artistic creation, image generation, and other creative domains. Therefore, the choice of which model to use should be based on the specific task and requirements at hand. VGG19 is suitable for image classification and feature extraction, while Magenta is better suited for generating creative content [10].

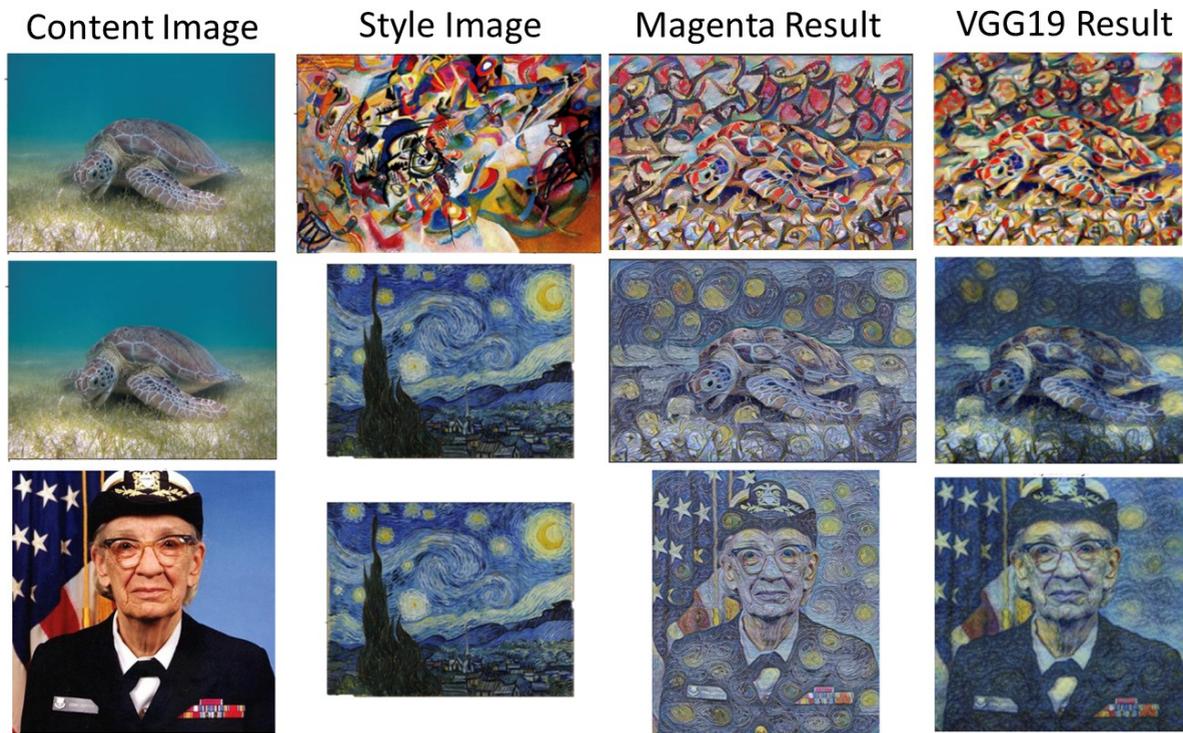
### 3. Result

The following sections will present the results obtained from both models. The generated images are demonstrated in Figure 1.

#### 3.1. Comparison of Image Quality

Magenta's generated images are characterized by a subdued color palette, resulting in a softer visual presentation. Despite the muted color tones, these images exhibit a notable level of clarity and sharpness, contributing to a crisp overall aesthetic. Magenta's image quality is commendable, warranting a rating of 7/10.

In contrast, VGG19 consistently yields images with vibrant and faithful color representation that closely adheres to the original artistic style. These images showcase finer details, resulting in a visually captivating quality, earning VGG19 a commendable image quality rating of 9/10.



**Figure 1.** Comparison of NST results between Magenta and VGG19 (Figure Credits: Original).

### *3.2. Comparison of Transfer Speed*

VGG19 enjoys a significant advantage in terms of speed, boasting an average image generation time of approximately 5 seconds. This swiftness positions VGG19 as a favorable choice for real-time applications or scenarios demanding high-throughput image generation.

Magenta, while proficient in image generation, registers a relatively slower average image generation time of around 20 seconds, making it better suited for less time-sensitive tasks.

### *3.3. Comparison of Style Fidelity*

VGG19 consistently excels in preserving the fidelity of the input artistic style, resulting in images that closely mirror the intended artistic genre.

Magenta, while generally proficient, may occasionally display minor deviations from the desired style, suggesting slightly less stringent style fidelity.

### *3.4. Comparison of Clarity and Details*

Outputs from VGG19 consistently exhibit sharper details and more clearly defined contours, contributing to an elevated overall visual experience.

Magenta occasionally introduces subtle blurring, resulting in images with somewhat less distinct characteristics.

### *3.5. Comparison of Versatility*

A notable attribute of Magenta lies in its versatility, accommodating various content types, including music and art. This versatility positions Magenta as an ideal choice for a diverse range of creative applications.

While exceptional in the domain of image generation, VGG19 remains specialized for visual content and does not extend its versatility to other creative domains.

### *3.6. Comparison of Training Time and Resource Efficiency*

VGG19's training process demonstrates notable expeditiousness and demands fewer computational resources compared to Magenta. This streamlined training translates into reduced time and cost implications for model development. VGG19's resource-efficient operation further allows deployment across a broader spectrum of hardware platforms.

Magenta, in certain cases, necessitates more extensive computational resources and training time.

## **4. Discussion**

In summary, comprehensive evaluations consistently favored VGG19, which showcased superior performance in aspects such as image quality, speed, and style fidelity. However, it is crucial to acknowledge that Magenta's primary strength lies in its versatility, rendering it well-suited for a broader spectrum of creative content types. The choice between these models should be guided by specific task requirements and priorities, with VGG19 excelling in delivering visually captivating results and Magenta offering versatility across diverse creative domains. Based on the analysis of the experimental results, it is essential to highlight the limitations and areas for improvement for both models:

### *4.1. Discussion on Magenta*

**Color Fidelity:** Magenta's propensity to generate lighter and less saturated colors, although resulting in a softer appearance, may not be suitable for tasks that require faithful color reproduction or a vibrant visual style. Improvements in color management algorithms could enhance its color fidelity.

**Speed:** Magenta's image generation speed, while acceptable for certain applications, could benefit from optimization to compete with real-time requirements effectively. Streamlining the underlying neural network architecture or exploring parallel processing techniques might lead to faster performance.

**Style Fidelity:** Magenta, occasionally exhibiting minor deviations from the intended artistic style, can further improve its style fidelity. Refinements in the neural network's ability to precisely capture and reproduce intricate stylistic details are warranted.

**Resource Efficiency:** Reducing Magenta's resource requirements, particularly in terms of computational power and memory, would broaden its accessibility and utility across a wider range of hardware platforms.

#### 4.2. Discussion on VGG19

**Clarity and Detail:** While VGG19 consistently generates visually captivating images, it occasionally exhibits slightly softer details compared to Magenta. Enhancements in the model's architecture or post-processing techniques could further sharpen image details.

**Style Customization:** VGG19's specialization in visual content limits its ability to adapt to non-visual creative tasks. Exploring methods to expand VGG19's capabilities to accommodate different content types, similar to Magenta's versatility, could be a valuable avenue for improvement.

**Interactivity:** Both models could benefit from enhanced interactivity features, allowing users to fine-tune the generated content or provide real-time feedback during the generation process. Incorporating user preferences into the generation process would increase the models' practicality in creative applications.

**Data Efficiency:** Improving the models' ability to generate high-quality content with fewer training examples could make them more accessible to users with limited data resources. Techniques such as transfer learning or data augmentation can contribute to this objective.

**Realism:** While both models excel in artistic stylization, further research into making generated content more realistic, when required, would extend their applicability to areas like computer graphics and design.

In conclusion, the analysis of the experimental results reveals several areas of improvement for both Magenta and VGG19. These include enhancements in color fidelity, speed, style fidelity, resource efficiency, clarity, versatility, interactivity, data efficiency, and realism. Addressing these limitations and exploring these improvement spaces will contribute to the continued advancement and utility of these creative AI models.

## 5. Conclusion

In summary, this paper has conducted a comprehensive comparative analysis of two prominent style transfer models: VGG19 and Magenta. Through a meticulous examination of their architectural nuances, functionalities, and practical utility, valuable insights into their respective strengths and limitations have been gleaned. VGG19, a well-established convolutional neural network architecture, excels in feature extraction, nonlinearity, dimensionality reduction, abstract feature composition, and classification. It is ideally suited for tasks involving image classification and feature extraction, making it a reliable choice in the field of computer vision. In contrast, Magenta is a creative content generation framework that harnesses the power of recurrent neural networks and convolutional neural networks to produce innovative content, including music and art. Its unique emphasis on human-AI collaboration opens up new avenues for artistic expression and creative exploration.

In the comparative evaluation, it was observed that Magenta's generated images tend to exhibit lighter colors, while VGG19 produces images with more vibrant and faithful color reproduction. However, Magenta's images are characterized by higher clarity and sharpness, making them ideal for tasks where image details are crucial. The choice between VGG19 and Magenta should be driven by the specific requirements of the given task. If precise color fidelity and vibrant styles are paramount, VGG19 may be the preferred option. Conversely, if image clarity and sharpness are critical, Magenta excels in producing clear and well-defined images.

Looking ahead, future research directions in the field of style transfer could involve algorithmic enhancements aimed at combining the strengths of both models, exploring methods to enhance the

performance and efficiency of style transfer techniques, and further integrating AI-assisted creative content generation across various domains.

In conclusion, this comparative analysis contributes to the understanding of the roles of VGG19 and Magenta in the field of style transfer, assisting researchers, practitioners, and enthusiasts in making informed decisions when selecting a model aligned with their specific objectives and creative aspirations.

## References

- [1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- [2] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2019). Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11), 3365-3385.
- [3] Singh, A., Jaiswal, V., Joshi, G., Sanjeeve, A., Gite, S., & Kotecha, K. (2021). Neural style transfer: A critical review. *IEEE Access*, 9, 131583-131613.
- [4] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern recognition*, 77, 354-377.
- [5] Cai, Q., Ma, M., Wang, C., & Li, H. (2023). Image neural style transfer: A review. *Computers and Electrical Engineering*, 108, 108723.
- [6] Bhangale, K. B., Desai, P., Banne, S., & Rajput, U. (2022). Neural Style Transfer: Reliving art through Artificial Intelligence. In *2022 3rd International Conference for Emerging Technology*, 1-6.
- [7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [8] Magenta: Music and Art Generation with Machine Intelligence, URL: <https://github.com/magenta/magenta/tree/main>. Last Accessed: 2023/09/24.
- [9] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation*, 31(7), 1235-1270.
- [10] Wang, P., Li, Y., & Vasconcelos, N. (2021). Rethinking and improving the robustness of image style transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 124-133.