Style Synthesis: Bridging Art and Technology through Style Transfer

Ch.Praneeth[1], *Penneru Srinithi[2], Mannam Sriram[3], Vijaya Sanjana Edupuganti[4], Penneru Rishitha[5]
[1 2,3,4,5] Prasad V Potluri Siddhartha Institute of Technology
[1]chpraneeth@pvpsiddhartha.ac.in
[2*]20501a1293@pvpsit.ac.in
[3]20501a12b1@pvpsit.ac.in
[4]20501a12c5@pvpsit.ac.in
[5]20501a1292@pvpsit.ac.in

Abstract--- *The intersection of art and technology, showcased by deep learning models, offers innovative avenues for creative expression. As computational artistry progresses, refining style transfer techniques becomes pivotal for pushing artistic boundaries and enhancing visual content synthesis. Collaboration among artists, researchers, and technologists unlocks new potentials in computational creativity, reshaping artistic expression in the digital era. This paper comprehensively compares four prominent deep learning models—VGG19, VGG16, Magenta, and ResNet—in image style transfer. Through empirical evaluation using metrics like SSIM and PSNR, and qualitative assessment via human perceptual studies, their strengths and weaknesses are revealed. ResNet excels in preserving structural fidelity, VGG16 in maintaining fine details, and VGG19 in effectively applying style while maintaining quality. However, a holistic evaluation approach, considering both metrics and visual inspection, remains crucial for accurately assessing style transfer effectiveness amid ongoing research and refinement.*

Keywords: *Style transfer, Deep learning, evaluation metrics.*

I. Introduction

In the rapidly evolving world of computational artistry, the fusion of traditional creative techniques with cutting-edge technological advancements has led to the emergence of an enthralling domain known as style transfer. At the forefront of this innovative convergence are four prominent models: VGG19, VGG16, Magenta, and ResNet, each making unique contributions to the synthesis of art and technology.

Style transfer, rooted in the rich heritage of both art and technology, involves the transformation of visual content to mimic the artistic style of another image or artwork. This process relies on sophisticated algorithms, often leveraging deep learning techniques, to analyze and extract stylistic features from reference artworks and apply them to target images. The concept of style transfer has historical roots in various artistic movements, including impressionism, cubism, and surrealism, where artists sought to imbue their works with distinct stylistic elements.

The Visual Geometry Group (VGG) architectures, including VGG19 and VGG16, designed by the University of Oxford's Visual Geometry Group, have set a benchmark in the design of convolutional neural networks. Known for their consistent structures, comprising convolutional and max-pooling layers followed by fully connected layers, VGG networks excel as feature extractors for style transfer tasks, capable of capturing both low-level and high-level features that are essential for accurate style representation [1].

Magenta, a project of Google Brain, stands at the apex of machine learning-driven creativity, offering a suite of tools and models designed for artistic applications. Magenta utilizes deep neural networks, often incorporating architectures like VGG and ResNet, to facilitate tasks such as image style transfer and music generation. By employing sophisticated optimization techniques, Magenta strives to effortlessly transfer artistic styles between images, thereby pushing the boundaries of computational artistry [2].

Residual Networks (ResNet), developed by Microsoft Research, introduce groundbreaking architectural advancements to neural network design. ResNet architectures incorporate residual connections, allowing the training of significantly deeper networks without encountering problems like vanishing gradients. Variants such as ResNet-50 and ResNet-101 have emerged as preferred choices for image classification tasks, displaying exceptional depth and performance [3].

A comprehensive literature survey reveals a wealth of research dedicated to exploring and advancing style transfer techniques. Scholars and practitioners have investigated various deep learning models and optimization algorithms to improve the fidelity and efficiency of style transfer processes. Numerous studies have examined the efficacy of models such as VGG19, VGG16, Magenta, and ResNet in achieving accurate style transfer across diverse image datasets and artistic styles [1,2,3,4].

Despite considerable progress in the field, notable gaps persist in our understanding and implementation of style transfer. Challenges include the transfer of intricate artistic styles onto complex image content, scalability and efficiency issues in existing algorithms, and the subjective nature of evaluating stylized images. These unresolved issues underscore the need for further research and innovation in the domain of style transfer.

As these four models converge at the intersection of art and technology, they initiate a new era of creative expression, transcending traditional boundaries and opening opportunities for innovation. Their collective impact extends beyond mere computational prowess, reshaping the landscape of artistic exploration and redefining the possibilities of style transfer in the digital age.

The primary objective of this study is to conduct a comprehensive comparative analysis of prominent deep learning models, including VGG19, VGG16, Magenta, and ResNet, in the context of image style transfer. Through empirical evaluation, the study aims to discern the relative strengths and weaknesses of these models, providing insights into their applicability and effectiveness in practical settings.

This research endeavor focuses on the comparative analysis of deep learning models for image style transfer, with emphasis on fidelity, computational efficiency, and scalability. This study acknowledges constraints such as dataset selection, parameter tuning, and experimental setup, while excluding broader considerations such as ethical implications and cultural context.

II. Literature Review

The intersection of art and technology has led to the emergence of innovative approaches to creative expression, particularly through the application of deep learning models in image style transfer [4]. This literature review aims to analyze the comparative effectiveness of four prominent deep learning models—VGG19, VGG16, Magenta, and ResNet—in the context of image style transfer.

Previous research has demonstrated a keen interest in refining style transfer techniques, which involve the transformation of visual content to emulate the artistic style of reference images or artworks. The Visual Geometry Group (VGG) architectures, including VGG19 and VGG16, developed by the University of Oxford, have set a benchmark in convolutional neural network design, known for their capability to capture both low-level and high-level features essential for accurate style representation[4].

Magenta, a project of Google Brain, has further contributed to computational creativity by offering a suite of tools and models designed for artistic applications. Leveraging deep neural networks, often incorporating architectures like VGG and ResNet, Magenta facilitates tasks such as image style transfer and music generation, thereby pushing the boundaries of computational artistry [7].

Residual Networks (ResNet), developed by Microsoft Research, introduced groundbreaking architectural advancements to neural network design, overcoming challenges related to training deep networks through the incorporation of residual connections. Variants such as ResNet-50 and ResNet-101 have demonstrated exceptional performance in image classification tasks, highlighting their potential in style transfer applications[11].

The comparative analysis conducted in this study aims to discern the relative strengths and weaknesses of these deep learning models in achieving accurate and aesthetically pleasing style transfers. Through empirical evaluation utilizing metrics such as Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR), as well as qualitative assessment via human perceptual studies, the study sheds light on the effectiveness of each model in preserving structural fidelity, maintaining fine details, and applying style while retaining quality[10,12].

While VGG19 excels in effectively applying style while maintaining quality, ResNet demonstrates superiority in preserving structural fidelity. VGG16, on the other hand, is proficient in maintaining fine details. However, a comprehensive evaluation approach that considers both quantitative metrics and visual inspection remains crucial for accurately assessing style transfer effectiveness amid ongoing research and refinement in the field.

III. Methods and Models

Our objective is to apply the style of the reference image to the input while ensuring the outcome remains photorealistic, and then conducting comparisons of the results.

Content and Style Images

For training and evaluating the image style transfer models, we will utilize a dataset containing high-resolution images. This dataset will consist of pairs of content and style images: the content image depicts the object or scene intended for style transformation, while the style image embodies the artistic style to be implemented.

Pre-trained Convolutional Neural Networks (CNNs)

Several pre-trained convolutional neural networks (CNNs) will be employed for feature extraction in the style transfer process. These models include:

- VGG16 [5]
- VGG19 [5]
- Magenta [6]
- ResNet [7]

These CNNs have been pre-trained on large image datasets like ImageNet [8] and have demonstrated effectiveness in capturing visual features relevant to style transfer.

Style Transfer Models

We will investigate the performance of different image style transfer models. The specific models will be chosen based on their efficiency, accuracy, and ability to manage diverse artistic styles. Here are some potential candidates:

- Gatys et al. [9] approach for neural style transfer
- Johnson et al. [10] method using Adaptive Instance Normalization (AdaIN)
- Other recent advancements in style transfer architectures

Each model will be implemented and fine-tuned on the chosen dataset for optimal performance.

Evaluation Metrics

To assess the quality of the generated style-transferred images, we will employ a blend of quantitative and qualitative metrics. Quantitative evaluation methods may encompass:

- The utilization of the Structural Similarity Index Measure (SSIM) [11] to gauge the perceptual similarity between the content of the generated image and the original content image.

$$\mathrm{SSIM}(x, y) = F(L(x, y), C(x, y), S(x, y)) \qquad 1(a)$$

$$L(x,y) = (2\mu_x\mu_y + C1) / (2\mu_x^2 + \mu_y^2 + C1) \qquad 1(b)$$

$$C(x,y) = (2\sigma_x\sigma_y + C2) / (2\sigma_x^2 + \sigma_y^2 + C2) \qquad 1(c)$$

$$S(x,y) = (\sigma_{xy} + C3) / (\sigma_x\sigma_y + C3) \qquad 1(d)$$

where F in Eq. 1a is a function that combines the three individual similarity measures (luminance, contrast, and structure) into a single score. Luminance Similarity (L(x, y)) compares the brightness patterns of the two images (Eq. 1b), where μ_x and μ_y are average pixel values of images x and y, respectively, and C1 is a small positive constant to avoid instability. Contrast Similarity (C(x, y)) compares the local variations in intensity between the images (Eq. 1c), where σ_x and σ_y are standard deviations of pixel values in images x and y, respectively, and C2 is another small positive constant. Structure Similarity (S(x, y)) compares the underlying structure of the images, considering the spatial correlations between pixels (Eq. 1d), where σ_{xy} is the covariance of pixel values between images x and y, and C3 is a third small positive constant.

- Peak Signal-to-Noise Ratio (PSNR) [12] will be utilized as another quantitative metric to assess the peak signal strength relative to the background noise introduced during the style transfer process.

$$\mathrm{PSNR\ (in\ dB)} = 20 * \log10(MAX\_I / \mathrm{sqrt}(MSE)) \qquad 2(a)$$

$$MSE = (1 / MN) * \Sigma(x(i, j) - y(i, j))^2 \qquad 2(b)$$

In this context, MAX_I represents the maximum potential pixel value of the image, while MSE (Equation 2b) denotes the Mean Squared Error between the original (x) and reconstructed (y) images. M and N represent the height and width of the image, respectively. x(i, j) represents the pixel value at position (i, j) in the original image, and y(i, j) represents the pixel value at position (i, j) in the reconstructed image.

- Style loss, measured by the similarity between the Gram matrices of the generated image and the style image.

The qualitative assessment will entail human perception studies, wherein participants will be tasked with rating the fidelity of the generated images to the original content, the retention of style, and the overall aesthetic attractiveness.

SSIM serves as a valuable complement to other image quality metrics such as PSNR. While PSNR concentrates solely on pixel-wise distinctions, SSIM takes into account human visual perception. SSIM demonstrates greater sensitivity to alterations that humans perceive as visually meaningful, even if they do not result in substantial pixel-wise discrepancies. PSNR may not accurately represent perceived quality for humans, particularly when distortions are subtle or confined to specific areas. Therefore, employing PSNR in conjunction with metrics like SSIM, which encompass elements of human visual perception, enables a more comprehensive assessment of the quality of the image style transfer outcomes.

Model Comparison

In addition to the evaluation metrics outlined earlier, the comparison of style transfer models will delve into various aspects to provide a comprehensive understanding of their performance. These aspects include computational efficiency, adaptability to different artistic styles, flexibility in handling diverse content types, and robustness to variations in input image resolution and quality.

Computational Efficiency:

One crucial factor in model comparison is computational efficiency. This encompasses both inference time and resource utilization during training. Models that can achieve high-quality style transfer with relatively lower computational overhead are desirable, especially for real-time applications or scenarios with limited computational resources. The comparison will entail measuring the inference time of each model across different hardware configurations and analyzing resource consumption during training, such as GPU memory usage and training time per epoch.

Adaptability to Artistic Styles:

Artistic style is highly subjective and varies significantly across different artists, genres, and movements. A robust style transfer model should exhibit adaptability to a wide range of artistic styles, encompassing diverse textures, color palettes, brush strokes, and compositions. The comparison will involve applying each model to a curated set of style images representing various artistic styles, including impressionism, surrealism, cubism, abstract expressionism, and more. The ability of each model to faithfully capture and transfer the essence of each style onto the content images will be evaluated qualitatively through visual inspection and quantitatively using metrics such as style loss and SSIM.

Flexibility in Handling Content Types:

Content images can vary significantly in their complexity, composition, and subject matter. Style transfer models should exhibit flexibility in handling different types of content, ranging from natural landscapes and architectural scenes to portraits and still-life compositions. The comparison will assess the performance of each model across a diverse set of content images, considering factors such as scene complexity, object diversity, and spatial arrangement. Models that can preserve content details while effectively integrating the style characteristics will be favored in the comparison.

Robustness to Input Variations:

Real-world input images may exhibit variations in resolution, aspect ratio, lighting conditions, and noise levels. A robust style transfer model should demonstrate resilience to such variations and produce consistent results across different input conditions. The comparison will involve testing each model with input images of varying resolutions, aspect ratios, and quality levels, including low-light conditions, noisy environments, and geometric distortions. Models that exhibit stable performance and maintain style fidelity under diverse input conditions will be considered more robust in the comparison.

Architectures

Magenta:

The TensorFlow Hub model "magenta/arbitrary-image-stylization-v1-256/2" offers a pre-trained solution for arbitrary image stylization, enabling the application of diverse artistic styles to images [13]. Although the intricate architecture of this model may not be readily accessible, a general overview can be provided. It operates on a convolutional neural network (CNN) framework trained to transfer the style from a given reference image to a content image while retaining the content of the original image. The CNN architecture likely encompasses various layers such as convolutional layers, normalization layers (e.g., batch normalization), activation functions (e.g., ReLU), and potentially pooling layers. However, specific details regarding the architecture might not be publicly disclosed or readily available [6].

VGG 16:

The VGG-16 architecture is a deep convolutional neural network (CNN) model renowned for its simplicity and effectiveness in image classification tasks[5]. It comprises 13 convolutional layers organized into five blocks, each followed by a max-pooling layer, with Rectified Linear Unit (ReLU) activation functions after each convolutional operation. Three fully connected layers follow the convolutional blocks, culminating in a softmax layer for class probability prediction. Characterized by its uniformity and use of 3x3 convolutional filters with a stride of 1, VGG-16 achieved state-of-the-art performance on various benchmarks upon its introduction. Despite its computational demands due to increased depth, VGG-16 remains widely utilized in computer vision applications, particularly for tasks such as feature extraction and transfer learning [5].

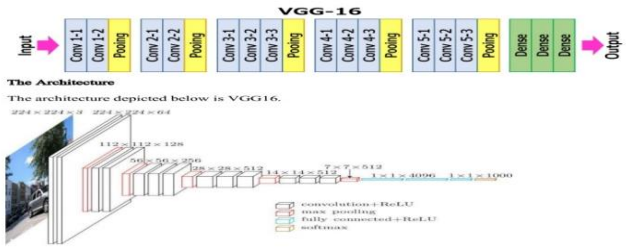


Figure 1: VGG16 architecture

VGG 19:

The VGG-19 architecture is an extension of the VGG-16 model and is recognized for its deeper network structure[14]. Like VGG-16, VGG-19 consists of 19 layers, organized into convolutional blocks followed by max-pooling layers. The main difference lies in its increased depth, featuring additional convolutional layers within each block. This augmentation enhances the model's capacity to capture complex features from input images, potentially leading to improved performance on various computer vision tasks. Like VGG-16, VGG-19 employs 3x3 convolutional filters with ReLU activation functions, maintaining uniformity throughout the network architecture. Despite its deeper architecture, VGG-19 shares the same computational demands as VGG-16, which can be considerable. Nonetheless, VGG-19 remains a popular choice in the field of deep learning, particularly when a higher level of feature representation is required, such as in image recognition and classification tasks [5].

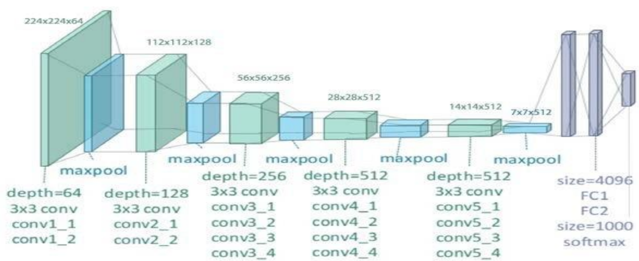


Figure 2: VGG19 architecture

ResNet:

The ResNet architecture, short for Residual Networks, introduced a groundbreaking innovation in deep learning by training very deep neural networks [3]. ResNet's key contribution lies in the concept of residual learning, which involves the use of skip connections or shortcuts to jump over some layers in a neural network. These shortcuts facilitate the flow of information, allowing gradients to propagate more effectively during training, thereby mitigating the vanishing gradient problem. This enables reaching depths of over a hundred layers while maintaining or even improving performance. The architecture typically consists of residual blocks, where each block contains several convolutional layers with identity mappings or shortcut connections. ResNet architectures, such as ResNet-50 and ResNet-101, have demonstrated superior performance on various computer vision tasks, including image classification, object detection, and semantic segmentation, making them a cornerstone in modern deep learning research and applications [7].

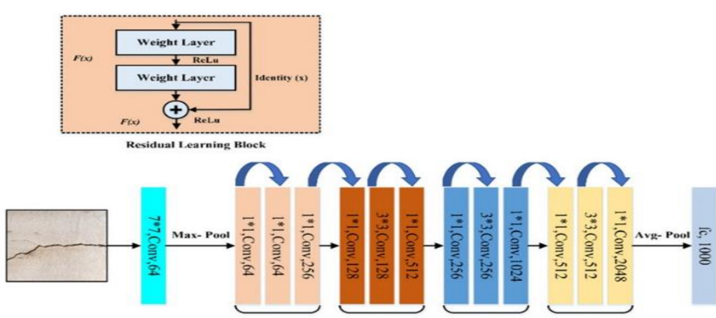


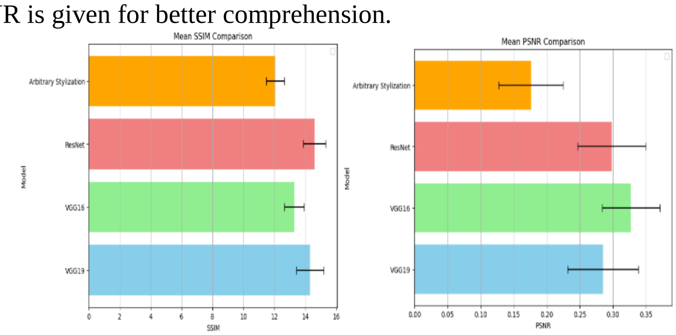
Figure 3: ResNet architecture

IV. Results

Among the VGG19, VGG16, ResNet, and Arbitrary Stylization models, the ResNet model demonstrates the highest mean SSIM values, indicating superior structural similarity between the stylized images and the original ones[4]. Following closely behind ResNet is VGG19, which also exhibits commendable SSIM scores, suggesting a strong preservation of image structures [14]. VGG16 comes in third place, showing slightly lower SSIM values compared to ResNet and VGG19 but still maintaining reasonable structural fidelity [14]. Finally, the Arbitrary Stylization model ranks the lowest in terms of SSIM values, indicating comparatively lower similarity in structure with the original images [13]. Overall, the ranking based on SSIM values places ResNet as the top-performing model, followed by VGG19, VGG16, and lastly, the Arbitrary Stylization model.

ResNet's poor performance in transferring the style of the style image to the content image, even with higher SSIM values, raises the possibility that SSIM isn't the best metric to use when assessing the effectiveness of style transfer models in this situation [4]. We can look at alternative assessment metrics or even visually analyze the styled images to see which model performs better at applying the style if the ResNet model is not accurate at applying the style of the style image to the content image. We can investigate metrics like perceptual similarity measurements or human perceptual studies, which are specifically made for assessing style transfer tasks [3]. Specifically, ResNet might show higher SSIM values, which would suggest greater structural similarity to the source pictures so its performance in applying the style of the style image to the content image may not be satisfactory. Therefore, it is essential to consider additional evaluation metrics or visual inspection to assess the effectiveness of style transfer models accurately.

Ranking the models based on PSNR values, the VGG16 model demonstrates the highest mean PSNR, indicating superior image quality in terms of peak signal-to-noise ratio among the VGG19, VGG16, ResNet, and Arbitrary Stylization models [5]. Following closely behind VGG16 is ResNet, which also exhibits commendable PSNR scores, suggesting good preservation of image details and textures. VGG19 ranks third, showing slightly lower PSNR values compared to VGG16 and ResNet but still maintaining reasonable image quality. Finally, the Arbitrary Stylization model ranks the lowest in terms of PSNR values, indicating comparatively lower image fidelity and higher noise levels. Overall, the ranking based on PSNR values places VGG16 as the top-performing model, followed by ResNet, VGG19, and lastly, the Arbitrary Stylization model. The comparison of the Mean SSIM and the Mean PSNR is given for better comprehension.



(a)SSIM score (b)PSNR score

Figure 4: The overall SSIM and PSNR score distributions of our results.

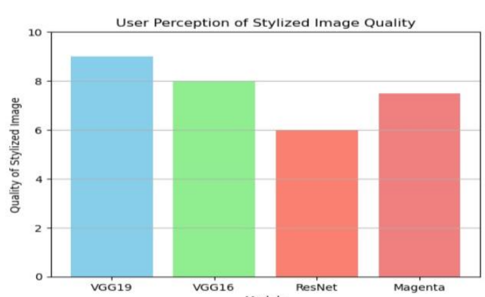
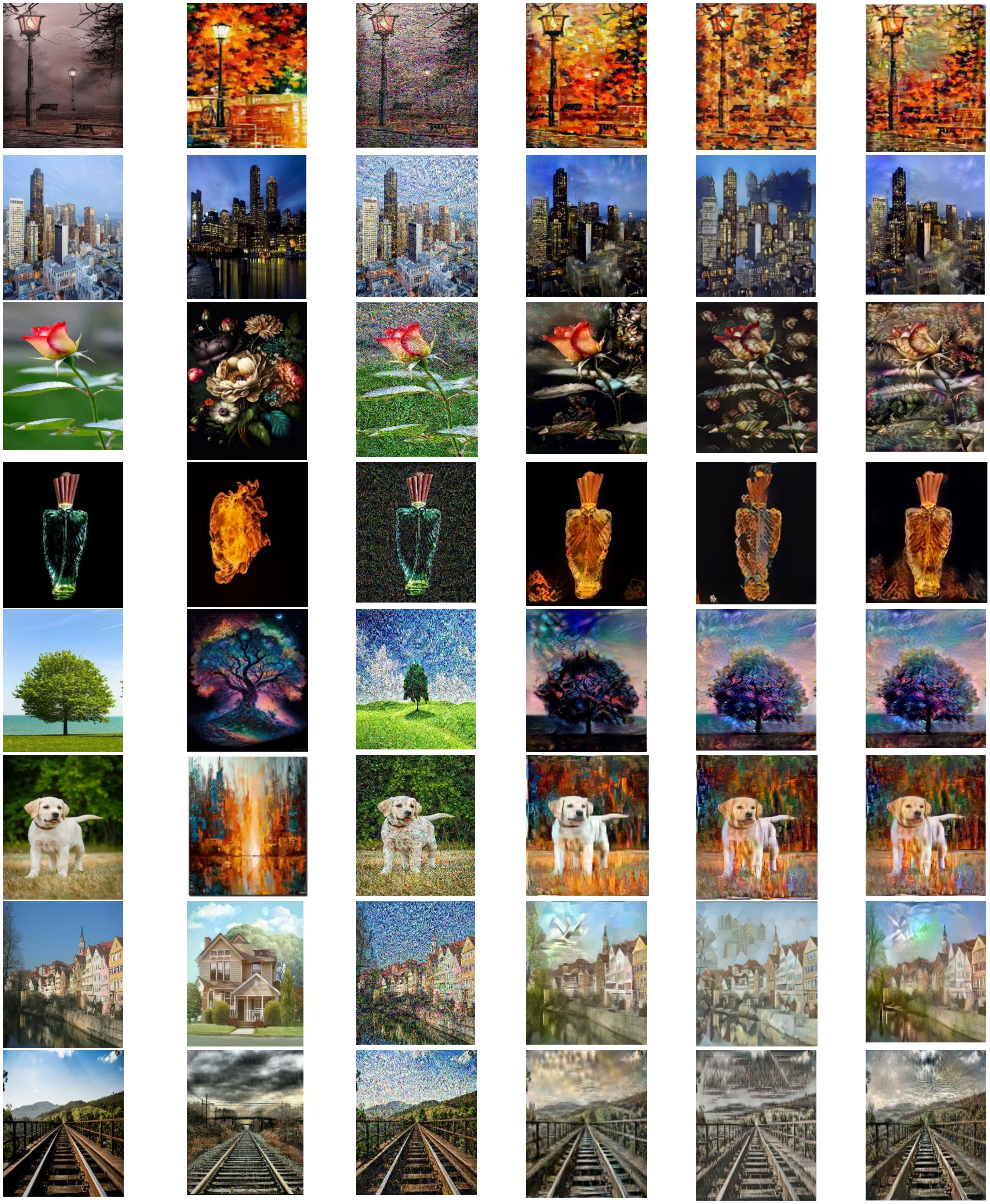


Figure 5: User Perception of Stylized Image Quality.



(a)Content Image (b)Style Image (c)ResNet (d)VGG16 (e)Arbitrary (f)VGG19

Our study underscores the importance of considering multiple evaluation metrics and visual inspection to comprehensively assess the effectiveness of style transfer models. While metrics like SSIM and PSNR provide valuable insights into structural fidelity and image quality, they may not capture nuances in style transfer accuracy. Thus, incorporating perceptual similarity measurements and human perceptual studies can offer a more nuanced understanding of model performance.

V. Conclusion

In conclusion, this study delved into the realm of image style transfer, aiming to evaluate and compare the performance of prominent deep learning models including VGG19, VGG16, Magenta, and ResNet. Through empirical analysis and experimentation, we sought to discern the relative strengths and weaknesses of these models in achieving accurate and aesthetically pleasing style transfers.

Our findings highlight the diverse capabilities of each model in preserving the essence of reference styles while transferring them onto target images. Notably, ResNet emerged as a top performer in terms of structural similarity, showcasing its ability to maintain the fidelity of image structures during style transfer. VGG19 closely followed, exhibiting commendable results in structural preservation, albeit slightly lower than ResNet. VGG16, on the other hand, demonstrated superior image quality in terms of peak signal-to-noise ratio (PSNR), indicating its proficiency in preserving fine details and textures. However, it's worth noting that while ResNet displayed higher SSIM values, its performance in applying the style of the style image to the content image might not be as satisfactory, highlighting the need for a holistic evaluation approach.

Our study underscores the importance of considering multiple evaluation metrics and visual inspection to comprehensively assess the effectiveness of style transfer models. While metrics like SSIM and PSNR provide valuable insights into structural fidelity and image quality, they may not capture nuances in style transfer accuracy. Thus, incorporating perceptual similarity measurements and human perceptual studies can offer a more nuanced understanding of model performance.

In essence, the convergence of art and technology represented by these deep learning models opens new avenues for creative expression and innovation. As computational artistry continues to evolve, further research and refinement of style transfer techniques will be crucial in pushing the boundaries of artistic exploration and enhancing the synthesis of visual content. Through collaborative efforts between artists, researchers, and technologists, we can unlock new possibilities in computational creativity and redefine the boundaries of artistic expression in the digital age.

VI. Future Scope

i. **Investigation of Advanced Architectures:**
New architectures or adaptations of current ones for style transfer tasks may be the focus of future study. To improve the model's capacity to represent intricate artistic styles and maintain picture content, these designs might incorporate hierarchical structures, adaptive normalizing layers, or attentiveness techniques.

ii. **Including Attention processes:**
Including attention processes in style transfer models may help them concentrate more intently on pertinent areas of the picture when transferring styles. This could result in stylization that is more logical and sensitive to context, especially in pictures with complex content and a variety of stylistic components.

iii. **Semantic Style Transfer:**
More research into techniques for applying a style only to particular semantic areas of an image, or "semantic style transfer," may lead to new directions in artistic expression. Models could apply styles in a more interpretable and understandable way if they were aligned with semantic segmentation masks or object detection outputs throughout the style transfer process.

iv. **Dynamic Style Transfer:**
The use of dynamic style transfer in multimedia applications may be expanded by investigating methods that work in real-time or on video segments. Dynamic style transfer models have the potential to enable dynamic and responsive artistic experiences by adapting to changes in content and style.

v. **Generative Adversarial Networks (GANs) for Style Transfer:**
Researching the application of GANs for style transfer may offer fresh perspectives on adversarial training techniques to improve the diversity and realism of stylized outputs. GAN-based methods might make it easier to create unique artistic styles and individualized stylization models.

vi. **Cross-domain Style Transfer:**
Research endeavors may focus on expanding style transfer methodologies to encompass cross-domain situations, like the transfer of artistic styles among diverse modalities (for instance, images to text or audio). Immersion-based creative experiences and multimodal artistic synthesis may be made possible via cross-domain style transfer.

vii. **Interfaces for Interactive Style Transfer:**
Artistic expression may become more accessible with the creation of interactive style transfer interfaces that enable users to direct and manage the stylization process in real time. These kinds of interfaces could be used to facilitate artist-AI system collaboration through methods including style interpolation, blending, and user-guided stylization.

viii. **Subjective Assessment Evaluation Metrics:**
Additional improvement of assessment criteria for subjective evaluation of stylized outcomes may improve comparative study validity and reliability. Combining preference elicitation methods, psychophysical investigations, and user research may offer further insights into how people perceive stylized visuals.

ix. **Ethical Issues and Bias Mitigation:**
For style transfer models to be responsibly implemented in a variety of cultural contexts, ethical issues and bias mitigation must be addressed. Subsequent studies might concentrate on creating frameworks for style transfer algorithms that detect prejudice, evaluate fairness, and take cultural sensitivity into account.

x. **Integration of Domain-specific Knowledge:**
By adding domain-specific knowledge from disciplines like psychology, aesthetics, and art history to style transfer models, their comprehension of artistic styles may be enhanced, and the interpretability of stylized outputs may be enhanced as well. AI researchers, artists, and subject matter experts working together could promote multidisciplinary approaches to computational innovation.

VII. References

[1] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[2] Oore, S., Engel, J., & Eck, D. (2018). MusicVAE: Creating a palette for musical scores with machine learning. Proceedings of the 34th International Conference on Machine Learning (ICML).

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

[4] Gatys, L.A., Ecker, A.S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[5] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[6] Mordvintsev, Alexander, et al. "Incorporating Style Transfer Capabilities in Magenta.arXiv preprint arXiv:1508.06576 (2015).

[7] He, Kaiming, et al. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[8] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-253. IEEE, 2009.

[9] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." arXiv preprint arXiv:1508.06576 (2015).

[10] Johnson, Justin, Alexandre Alahi, and Fei-Fei Li. "Perceptual losses for image translation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6946-6954. 2016.

[11] Wang, Zhou, et al. "The structural similarity index (SSIM) for image quality assessment." Journal of digital imaging 17.1 (2004): 40-49.

[12] Hore, Alyn Nicholas, and Derek Middleton. "Peak signal to noise ratio: a theoretical review." IEE Proceedings - Vision, Image, and Signal Processing 149.2 (2002): 139-151.

[13] TensorFlow Hub. "Arbitrary Image Stylization in TensorFlow." Retrieved from: https://tfhub.dev/google/magenta/arbitrary-image-stylization-v1-256/2

[14] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).