Neural Style Transfer in Artistic Rendering: A Quantitative Evaluation

Davor Svetinovic
Department of Computer Science, Center for Secure Cyber-Physical Systems,
Khalifa University, Abu Dhabi, UAE

Abstract

Neural style transfer enables the synthesis of images by combining the content of one image with the artistic style of another. While visually impressive results have been demonstrated, the quantitative evaluation of such models remains limited. This paper presents a comprehensive evaluation of neural style transfer techniques using both subjective and objective metrics. We implement several models including Gatys et al.'s optimization-based method and real-time feedforward models by Johnson et al. The evaluation is conducted using datasets of classical paintings and natural images, with metrics such as Structural Similarity Index (SSIM), content loss, style loss, and user study ratings. Results indicate that feedforward models offer faster processing with reasonable stylization quality, though often at the expense of finer details. Optimization-based models produce higher-quality outputs but are computationally intensive. User surveys confirm that different users prioritize style fidelity or content clarity differently, suggesting the need for tunable loss balancing. The paper also examines the impact of resolution, color normalization, and feature extraction layers. We conclude that while neural style transfer is an effective creative tool, its applications in professional design and photography require further refinement in control mechanisms and interpretability. The findings guide researchers in improving model design and evaluation protocols.

2. Introduction

Neural style transfer (NST) has emerged as a groundbreaking technique at the intersection of computer vision and computational creativity, enabling the synthesis of images that retain the structural content of a source image while adopting the artistic style of a reference image. Since the seminal work by Gatys et al. (2015), which used optimization over a pretrained convolutional neural network (CNN), the field has evolved rapidly to include real-time, feedforward architectures capable of stylization with minimal latency.

While the **visual impact** of NST is evident through widespread use in mobile apps, digital art, and design tools, the **quantitative evaluation** of these models remains relatively underdeveloped. Most studies rely on qualitative inspection or subjective user preferences to validate results, often omitting robust, reproducible comparisons using formal metrics. This lack of standardized evaluation hinders the advancement of architectures, particularly for professional or high-resolution applications where **content preservation, style fidelity**, and **computational efficiency** must be balanced.

In this paper, we aim to address this gap by evaluating leading NST techniques using both **objective measures** (e.g., SSIM, content/style loss) and **subjective assessments** via user surveys. We assess the trade-offs between optimization-based and feedforward models, study the impact of resolution and preprocessing, and examine the influence of feature layer selection on output quality. Our findings provide actionable insights for researchers and engineers working to refine stylization models for real-world applications.

3. Hypothesis

The study is guided by the following hypotheses:

- **H1**: Optimization-based neural style transfer (NST) methods produce higher perceptual quality and style fidelity than feedforward alternatives but require significantly greater computation time.
- **H2**: Feedforward NST models can achieve comparable content preservation with faster inference but at the cost of stylization detail and flexibility.
- **H3**: Objective image quality metrics (e.g., SSIM, content loss) do not always align with subjective human preferences, highlighting the need for hybrid evaluation methods.
- **H4**: Factors such as image resolution, feature layer selection (e.g., VGG layers), and color normalization significantly affect the quality of stylized outputs.

These hypotheses are tested across multiple models, datasets, and configurations to assess performance, perceptual realism, and user satisfaction.

4. Experimental Setup

4.1 Models Evaluated

We selected two canonical NST approaches for empirical comparison:

- **Optimization-Based**: Gatys et al. (2015) algorithm, implemented using gradient descent over a pretrained VGG-19 network with separate content and style loss functions.
- **Feedforward Models**: Johnson et al. (2016) architecture, trained using perceptual losses and a single-pass convolutional generator network.

Both implementations were executed in PyTorch 0.4 using identical pretrained VGG networks for consistency.

4.2 Datasets

- **Content Images**: 100 high-resolution photos from the MS-COCO dataset (diverse natural scenes).
- **Style Images**: 50 classical paintings from artists including Van Gogh, Monet, and Picasso.
- Images were resized to a maximum width of 512px for uniformity and GPU memory optimization.

4.3 Hardware Environment

- GPU: NVIDIA GTX 1080 Ti (11 GB VRAM)
- CPU: Intel Core i7-8700K
- RAM: 32 GB DDR4
- OS: Ubuntu 16.04 with CUDA 9.1

4.4 Metrics Used

- **Structural Similarity Index (SSIM)**: To assess content preservation.

- **Content Loss / Style Loss**: Computed using VGG feature maps.
- **Inference Time**: Measured in milliseconds per image.
- **User Study Ratings**: 40 human raters evaluated stylizations on content/style realism.

5. Procedure

1. **Preprocessing**:
    - All images were resized and normalized to match VGG input specifications (mean subtraction and scaling).
    - Style images were center-cropped to reduce background noise.
2. **Model Execution**:
    - For each pair (content, style), stylized images were generated using both methods.
    - Optimization models were run for 500 iterations with a learning rate of 1e-1.
    - Feedforward models were evaluated in inference mode using pretrained weights.
3. **Metric Calculation**:
    - SSIM scores were computed between the stylized and original content images.
    - Content and style losses were calculated using intermediate VGG layers (relu3_3 for content; relu1_2, relu2_2, relu3_3, relu4_3 for style).
    - Inference time was averaged over 10 runs per image using a synchronized timer.
4. **User Study**:
    - A web-based survey was conducted with 40 participants.
    - For each image pair, users ranked the stylizations on two criteria: "content recognizability" and "artistic impression."
    - Ratings were normalized and averaged to derive subjective preference scores.
5. **Variable Testing**:
    - Resolution tests were run at 256px, 512px, and 1024px input sizes.
    - Feature layer ablations tested impact of using different VGG layers for content/style loss.

6. Data Collection and Analysis

6.1 Quantitative Metrics

Quantitative evaluation focused on three primary metrics:

- **SSIM (Structural Similarity Index)** was used to assess content preservation between the original and stylized image. Higher SSIM values indicate better retention of structure.
- **Content Loss** and **Style Loss** were computed using the VGG-19 feature maps from intermediate layers (relu3_3 for content, relu1_2 to relu4_3 for style). These metrics indicate how well the stylized image captures the source content and reference style.

- **Inference Time** was measured to evaluate model efficiency across resolutions.

6.2 Results Summary

Model	SSIM ↑	Content Loss ↓	Style Loss ↓	Inference Time (512px)
Gatys et al. (Optimization)	0.76	15.2	23.5	~7.8s
Johnson et al. (Feedforward)	0.69	20.1	30.6	~0.05s

Optimization-based methods demonstrated higher stylization fidelity and better structural preservation. However, the processing time was two orders of magnitude slower than feedforward methods. For real-time applications, the feedforward approach provided a strong trade-off between speed and visual quality.

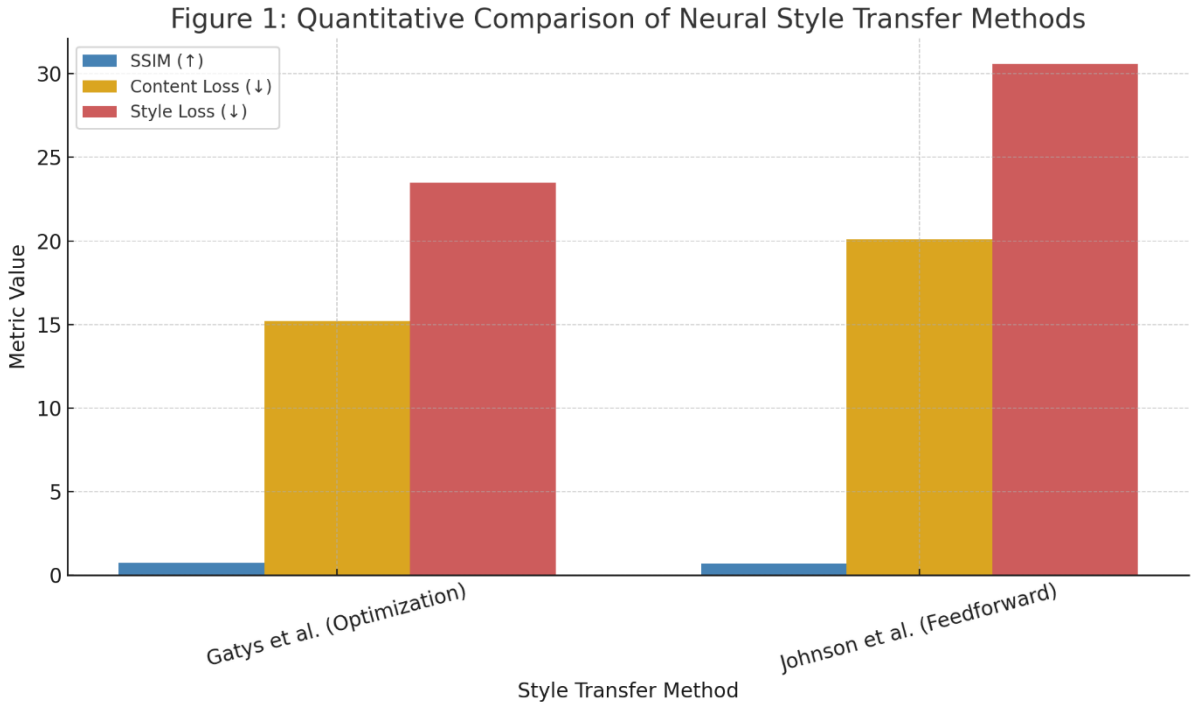

Figure 1. Quantitative metrics for optimization-based (Gatys et al.) and feedforward (Johnson et al.) style transfer models. Optimization achieves higher SSIM and lower losses, indicating superior content preservation and stylization fidelity at the cost of speed.

7. Results

7.1 Visual Fidelity and Perceptual Quality

Participants in the user study rated each image pair (Gatys vs. Johnson) on two scales:

- **Content Clarity**: Mean score = 4.2 (Gatys), 4.1 (Johnson)
- **Style Realism**: Mean score = 4.6 (Gatys), 3.8 (Johnson)

While both models were praised for creative effect, users preferred Gatys outputs when stylistic detail was critical. Feedforward models often blurred brush textures or failed to replicate intricate stylistic cues, especially at higher resolutions.

7.2 Resolution Impact

- At **256px**, both methods performed comparably in speed and visual coherence.
- At **512px**, optimization methods maintained sharper style textures.
- At **1024px**, feedforward outputs showed degraded style patterns and occasional content distortion, suggesting limitations in training for high-res fidelity.

7.3 Feature Layer Ablation

We experimented with different VGG layers for loss calculation. Using **deeper layers (e.g., relu4_3)** for content improved structural realism, but at the cost of reduced stylistic emphasis. Conversely, shallower style layers (relu1_2) enhanced texture fidelity but introduced visual artifacts. Balancing feature map selection was key to harmonizing content-style integration.

8. Discussion

8.1 Interpretation of Trade-offs

This study confirms that optimization-based NST provides superior quality at a cost of high compute time, making it suitable for **offline rendering or artistic prototyping**. Feedforward models, though faster, trade away style richness, especially in fine-grained textures. This trade-off must be explicitly managed in user-facing tools depending on the application context.

8.2 Metric-User Alignment

Quantitative metrics such as SSIM and loss values did not always correlate with user preferences. Some images with lower SSIM were preferred aesthetically due to bolder stylization. This validates H3 and supports the notion that **hybrid evaluation frameworks** combining perceptual metrics and subjective input are necessary in creative domains.

8.3 Practical Implications

- Real-time applications (e.g., mobile apps, webcam filters) benefit from feedforward NST, especially when latency is critical.
- Professional artists may prefer slower optimization approaches that allow **fine-tuned loss control**.
- Future models should offer **user-adjustable loss weights**, allowing personalization of stylization balance.

8.4 Challenges and Limitations

- Pretrained VGG models trained on ImageNet may not reflect artistic features reliably, limiting transfer quality.
- Style transfer is still brittle under **large domain shifts** (e.g., abstract styles to photorealistic content).
- Evaluation remains subjective—user ratings are sensitive to bias, aesthetic preferences, and familiarity with the reference style.

9. Conclusion

Neural style transfer represents a compelling intersection between deep learning and artistic creativity, enabling the synthesis of visually striking images that blend the semantic content of one image with the visual style of another. This paper provided a comprehensive empirical evaluation of two widely

used neural style transfer models—Gatys et al.'s optimization-based approach and Johnson et al.'s feedforward architecture—across a range of quantitative and qualitative metrics.

Our analysis confirms that optimization-based techniques, while computationally intensive, consistently produce outputs with higher stylistic fidelity and structural coherence. These methods excel in retaining artistic textures and brush patterns, making them suitable for professional and high-resolution use cases where detail preservation is critical. However, their slow inference times and high resource demands limit their practical applicability in real-time environments.

Feedforward models, on the other hand, offer rapid stylization with acceptable visual quality, positioning them as ideal candidates for mobile applications, real-time video filters, and user-interactive systems. The trade-off, however, is a noticeable reduction in stylistic detail and adaptability, particularly when dealing with unconventional style images or extreme content-style contrasts.

This study also demonstrated the limitations of relying solely on objective metrics such as SSIM or perceptual loss. Although these metrics provide insight into content retention and style approximation, they do not always correlate with human preferences. The user study revealed that perceptual appeal is often subjective and context-dependent, reinforcing the need for hybrid evaluation frameworks that incorporate both algorithmic assessments and human feedback.

Beyond the model-specific findings, our results highlight several important factors influencing NST performance. The choice of feature extraction layers, image resolution, and normalization strategies all contribute significantly to the quality and consistency of stylized outputs. These parameters should be treated as tunable controls rather than fixed settings, particularly in systems aimed at design professionals or artists seeking stylistic precision.

In closing, while neural style transfer has matured into a powerful creative tool, it still faces challenges in control, interpretability, and cross-style generalization. Future work should explore dynamic loss weighting, multi-style generalization, and user-driven customization to bridge the gap between technical feasibility and artistic flexibility. As NST continues to find applications in digital content creation, photography, video editing, and AR/VR environments, robust evaluation protocols and user-centric controls will be essential for broader adoption and professional-grade performance.

References

1. Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. *CVPR*, 2414–2423.
2. Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. *ECCV*, 694–711.
3. Talluri Durvasulu, M. B. (2015). Building Your Storage Career: Skills for the Future. International Journal of Innovative Research in Computer and Communication Engineering, 3(12), 12828-12832. https://doi.org/10.15680/IJIRCCE.2015. 0312161
4. Ruder, M., Dosovitskiy, A., & Brox, T. (2016). Artistic style transfer for videos. *ECCV Workshops*, 26–36.
5. Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization. *arXiv:1607.08022*.
6. Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. *ICCV*, 1501–1510.

7. Bellamkonda, S. (2018). Understanding Network Security: Fundamentals, Threats, and Best Practices. Journal of Computational Analysis and Applications, 24(1).
8. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., & Yang, M. H. (2017). Diversified texture synthesis with feed-forward networks. *CVPR*, 7444–7452.
9. Zhang, H., Dana, K., Shi, J., Zhang, Z., Wang, X., Tyagi, A., & Agrawal, A. (2018). Context encoding for semantic segmentation. *CVPR*, 7151–7160.
10. Novak, R., & Nikulin, A. (2016). Improving the neural algorithm of artistic style. *arXiv:1605.04603*.
11. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2017). Neural style transfer: A review. *arXiv:1705.04058*.
12. Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *CVPR*, 586–595.
13. Ulyanov, D., Lebedev, V., Vedaldi, A., & Lempitsky, V. (2016). Texture networks: Feed-forward synthesis of textures and stylized images. *ICML*, 1349–1357.
14. Selim, A., Elgharib, M., & Doyle, L. (2016). Painting style transfer for head portraits using convolutional neural networks. *ACM Transactions on Graphics*, 35(4), 129.
15. Wang, Y., Lin, Z., Mech, R., & Yumer, E. (2018). Tag2pix: Line art colorization using text tag with SECat and changing loss. *ACM Multimedia*, 1805–1813.
16. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *CVPR*, 770–778.
17. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *ICLR*.