Synthetic Data 3D Style Transfer

Seunghwan Hyun, Yucheng Huang

April 26, 2024

1 Problem Definition

The challenge of style transfer involves rendering stylized views of a 3D scene with consistency across multiple perspectives. Traditional methods face difficulties in balancing accurate geometry reconstruction with high-quality stylization, especially when adapting to arbitrary new styles. This project explores the application of the recently introduced StyleRF [1](CVPR 2023), which promises to solve these issues through innovative style transformation techniques within the radiance field's feature space.

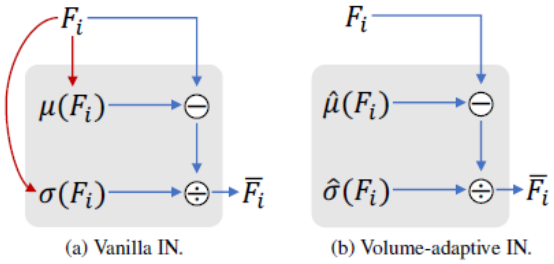

Figure 1: Sampling Invariant Content Transformation Flow

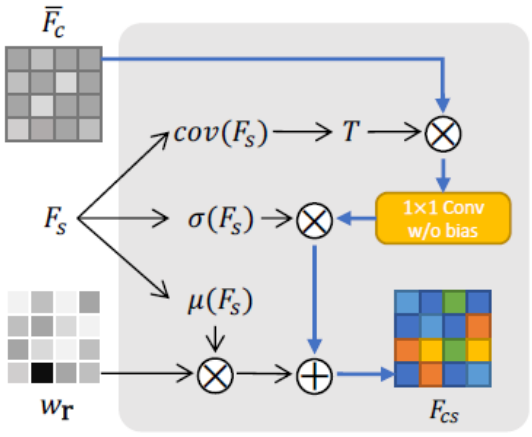

Figure 2: Deferred Style Transfer

In the realm of Neural Radiance Fields, recent innovations have been made in the area of view synthesis and object blending using text prompts for zero-shot generation. Gu et al.'s NerfDiff [2] introduces a novel distillation approach from a 3D-aware conditional diffusion model into NeRF, enhancing details in synthesized views under occlusion. Gordon et al. present Blended-NeRF [3], a framework capable of editing regions within a NeRF scene to blend new objects in a semantically and physically consistent manner using text prompts. These advancements pave the way for integrating style transfer within NeRF by facilitating the synthesis and refinement of virtual views and localized scene editing.

For style transfer, traditional 2D style transfer methods show great performance in flat images but does not apply the depth and perspective shifts in 3D scenes and lead to multi-view inconsistencies and poor stylization quality [4]. StyleRF resolves such issues by applying style transformation on the feature space of a radiance field which allows more precise geometry and more repressive stylization quality [5].

2 Method and Implementation

Inspired by the capabilities of StyleRF as detailed in its seminal paper, our project implemented this to evaluate its efficacy on both real-world and synthetic datasets. StyleRF's approach uses a grid of high-level features for 3D scenes, enabling precise geometry restoration and high-quality stylization:

Feature Grid Representation: StyleRF represents scenes using a feature grid rather than traditional RGB 2D image plane, enhancing the detail and accuracy of the geometric reconstruction

There are two key components lying in Feature Transformations for Style Transfer which are Sampling-Invariant Content Transformation (SICT) and Deferred Style Transformation (DST).

SICT: Maintains consistency across different views by adapting feature transformations to be independent of the sampled 3D points' holistic statistics. In NeRF representation, there are N points along the ray with each point having a feature dimension of C. This transformation maps

the 3D NeRF space into a better representation to apply style on. Let $F_i \in \mathbb{R}^C$ where $i \in \{1, 2, \ldots, N\}$.

The 3D space are transformed into Q,K,V tensors as follows:

$$Q = q(\text{Norm}(F_i)), \tag{1}$$
$$K = k(\text{Norm}(F_i)), \tag{2}$$
$$V = v(\text{Norm}(F_i)). \tag{3}$$

Then, the feature $\hat{F}_i$ can be constructed by:

$$\hat{F}_i = V \cdot \text{Softmax}(\text{fcov}(Q, K)). \tag{4}$$

To further elevate its stability across different rays, it uses the learned mean and variance for computing the ground truth $\hat{F}_i$ as show in Figure 1(b).

DST: To maintain the consistency in different views, it imposes the style in the 3D space.

From VGG, we obtain the style feature f_s and compute feature covariance matrix T from $\text{conv}(f_s)$. With SICT feature F_{cs}, the style is then applied in the 3D space using the function, visualized in Figure 2:

$$F_{cs} = \sum_{i=1}^{N} w_i \left(\text{conv}\left(T \otimes F_i\right) \times \sigma(F_s) + \mu(F_s)\right),$$

where conv denotes a convolution operation, and $\cdot$ denotes matrix multiplication or dot product, depending on context.

3 Experiments

We evaluated the quality of style transfer on synthetic datasets, where traditional methods often struggle to maintain style consistency and detail. Synthetic data, known for its controlled environment, enables precise manipulations and provides consistent, repeatable testing conditions.

For this analysis, we applied various style images to synthetic data and captured 200 images of the 3D grid from multiple angles. This approach allowed us to comprehensively assess how well StyleRF maintains style consistency and detail across different views.

Given the mixed performance on synthetic data, we expanded our testing to include a broader range of style images. This additional phase was designed to investigate how different characteristics of style images influence the quality of style transfers. Our aim was to identify specific features or attributes within the style images that significantly affect the effectiveness and visual appeal of StyleRF's outputs. This exploration is essential for improving our understanding of the interaction between style elements and the 3D rendering process, which will inform future enhancements in style transfer applications.

3.1 Evaluation Metrics

In our study, we employed two critical metrics, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), to provide both quantitative and qualitative evaluations of the fidelity and perceptual quality of the style transfers conducted by StyleRF:

PSNR is widely recognized for its ability to measure the fidelity of an image after processing. It is effective in assessing how much the noise introduced by style transfer deviates from the original content structure. This metric allowed us to determine how effectively StyleRF preserved the original details and textures of the original 3D object, providing a clear numerical measure of the degradation or preservation of image quality following the application of various styles.

To complement the quantitative insights of PSNR, we utilized SSIM to evaluate the perceptual quality of the stylized outputs. This metric is crucial for understanding how structural changes induced by style application affect the viewer's perception of the images, especially in terms of consistency and detail across different views. SSIM helps in identifying variations in texture, contrast, and structure, offering a more viewer-centric analysis of image quality that aligns closely with human visual perception.

4 Results

We calculated the PSNR and SSIM for 200 screenshots captured from various angles around the 3D outputs and compared them with the 200 screenshots of the original object from the same angle.

4.1 PSNR

Style Image	PSNR Values
Camille Pissarro	32.7714
Tiger	32.7776
Ghibli	32.8170
HSV	32.9136
Marsh	32.8283

Table 1: PSNR values for different style images

PSNR values for 5 different style images were rather consistently high. All 5 values were around

between 32 and 33 dB which is considered good for most applications with minor artifacts. If PSNR is above 40 dB, the image degradation is imperceptible to the human eye. The differences in PSNR values, being quite small, imply that the model consistently maintains a similar level of quality across different styles regarding fidelity.

4.2 SSIM

Style Image	SSIM Values
Camille Pissarro	0.05683
Tiger	0.05855
Ghibli	0.07330
HSV	0.08520
Marsh	0.07755

Table 2: SSIM values for different style images

SSIM values for 5 different style images were also consistent but very low. All 5 values were below 0.1 which is considered poor quality and this indicates that, despite the geometric and textural fidelity suggested by the PSNR scores, the perceptual quality and the visual appearance of the images vary greatly from the original. This could be due to the style transfer significantly altering visual elements that SSIM is sensitive to, such as texture and contrast, even if the overall spatial structure remains intact.

4.3 Conclusion

The consistency in PSNR values is encouraging, indicating good fidelity across styles, but the low SSIM values raise concerns about perceptual quality variations that could impact the applicability or acceptance of the stylized outputs. These metrics together suggest a need for further refinement of the style transfer process to enhance both the objective and subjective quality of the results.

PSNR and SSIM values demonstrate that using different style images from natural photo, oil painting, computer graphic and even HSV colormap plays minimal role in the quality of style transfer. For PSNR, consistently high values mean that the model is robust and guarantees outstanding performance. However, it was interesting to find that SSIM did not improve regardless of different colors, lightings, texture and structure could not enhance the quality of stylized outputs.

5 Discussion

Our findings confirm that while StyleRF significantly advances 3D style transfer, especially in handling real-world data, its application on synthetic datasets still requires refinement. The disparity in performance between real-world and synthetic datasets highlights the need for further refinement in how style transfer models handle different types of data. The current limitations may stem from the underlying algorithms' sensitivity to the structured, often repetitive patterns found in synthetic datasets, which differ significantly from the more organic qualities of real-world data.

6 Future Work

Future research could aim to adapt the StyleRF approach to better accommodate the unique properties of synthetic data. This will involve tweaking the model to enhance its understanding and processing of synthetic textures and compositions. Additionally, we plan to explore broader applications of this technology, including virtual reality and augmented reality, where realistic and aesthetically pleasing 3D models are crucial.

7 Acknowledgement

We send thanks to the developers of StyleRF for their pioneering work and profound insights. We also express our gratitude to ChatGPT for its invaluable assistance in revising and refining the text of this document.

Additionally, ChatGPT helped us designing the outline for our user interface, utilizing the 'Gradio' package, which contributed to the implementation of our project.

A Appendix

A.1 Real Data 3D style transfer

Figure 3: T-Rex

Figure 4: T-Rex with style transfer

A.2 Style Images

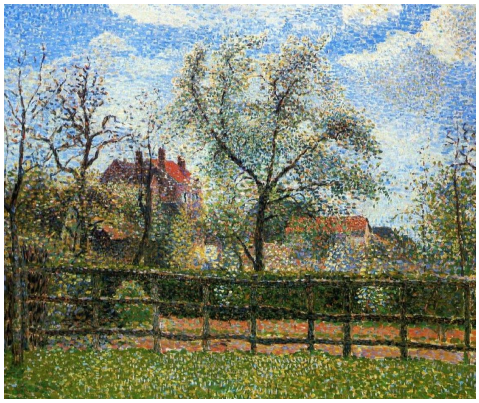

Figure 5: Camille Pissarro

Figure 6: Marsh

Figure 7: Tiger

Figure 8: Ghibli

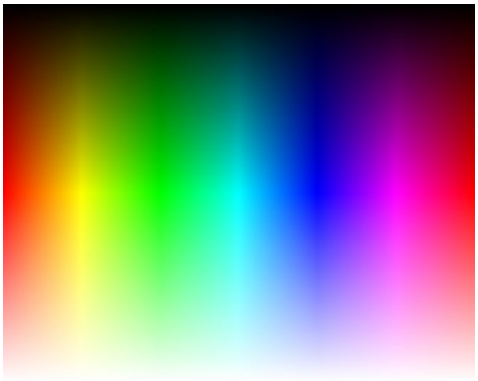

Figure 9: HSV Colormap

A.3 Output Screenshots

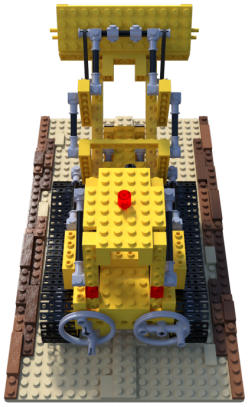

Figure 10: Original Lego

Figure 11: Camille Pissarro Lego

Figure 12: Marsh Lego

Figure 13: Tiger Lego

Figure 14: Ghibli Lego

A.4 User Interface

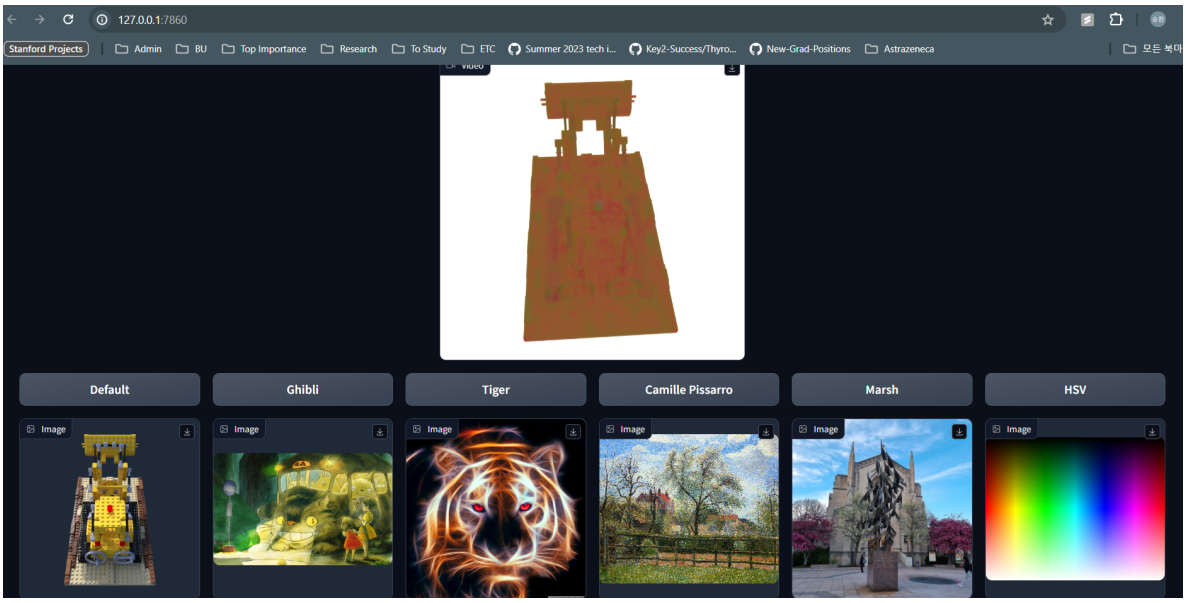

Figure 15: Screenshot of User Interface

References

[1] Kunhao Liu, Fangneng Zhan, Yiwen Chen, Jiahui Zhang, Yingchen Yu, Abdulmotaleb El Saddik, Shijian Lu, and Eric P Xing. Stylerf: Zero-shot 3d style transfer of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8338–8348, 2023.

[2] Jiatao Gu, Alex Trevithick, Kai-En Lin, Josh Susskind, Christian Theobalt, Lingjie Liu, and Ravi Ramamoorthi. Nerfdiff: Single-image view synthesis with nerf-guided distillation from 3d-aware diffusion, 2023.

[3] Ori Gordon, Omri Avrahami, and Dani Lischinski. Blended-nerf: Zero-shot object generation and blending in existing neural radiance fields, 2023.

[4] Hsin-Ping Huang, Hung-Yu Tseng, Saurabh Saini, Maneesh Singh, and Ming-Hsuan Yang. Learning to stylize novel views, 2021.

[5] Pei-Ze Chiang, Meng-Shiun Tsai, Hung-Yu Tseng, Wei sheng Lai, and Wei-Chen Chiu. Stylizing 3d scene via implicit representation and hypernetwork, 2022.