

# Tune-Your-Style: Intensity-tunable 3D Style Transfer with Gaussian Splatting

## Appendix

This Appendix contains the following parts:

- **Additional Qualitative Comparison** (Appendix A). We provide the qualitative comparison with NeRF-based methods.
- **Additional Visualization Results** (Appendix B). We provide additional visualization results for the intensity-tunable style injection.
- **Details of Quantitative Evaluation** (Appendix C). We provide details of the quantitative evaluation, including detailed data for calculating the CLIP metrics and evaluation criteria for the user study.
- **Preliminary: 3D Gaussian Splatting** (Appendix D). We introduce the technical details of 3D Gaussian Splatting.
- **Limitations and Social Impact** (Appendix E). We discuss the limitations and social impact of our work.

### A. Additional Qualitative Comparison

We conduct additional qualitative comparison with existing NeRF-based style transfer methods, and the results are presented in Fig. A. Concretely, We compare our method with StylizedNeRF [1], ARF [4], and StyleRF [3] on five scenes from the LLFF dataset: fern, horns, orchids, trex, flower, following their official settings. These NeRF-based methods adhere to a fixed-output paradigm, which in some cases fails to maintain a reasonable balance between content and style, leading to suboptimal stylization results. In contrast, our method not only produces visually appealing results, but also enables the user to adjust the style intensity injected into the scene to achieve a desired content-style balance, significantly enhancing the practicality of 3D style transfer.

### B. Additional Visualization Results

We present more visualization results of intensity-tunable style injection, as presented in Fig. B. Additional results with more reference style images and 3D scenes further demonstrate the effectiveness and distinct advantages of the proposed method.

### C. Details of Quantitative Evaluation

#### C.1. Evaluation Data

We present the detailed data used for the quantitative evaluation of CLIP similarity and CLIP direction similarity in Tab. A. Each data sample contains the scene, the original scene description, the stylized scene description, and the reference style image.

#### C.2. Evaluation Criteria for User Study

We also present the detailed evaluation criteria for the user study. Specifically, we request participants to evaluate from three dimensions: content consistency, style consistency, and visual appeal, using a 5-point scale. The criteria for each rating level are elaborated in Tab. B. We calculate the average score across the three dimensions and provide the 95% confidence interval. The user study results are collected from a total of 20 participants.

### D. Preliminary: 3D Gaussian Splatting

Gaussian Splatting(3DGS) [2] introduces splatting-based rasterization and explicitly models a scene as a collection of 3D Gaussian primitives. Each Gaussian primitive  $\Theta_i$  is parameterized by a center point  $\mathbf{x}$  and a covariance matrix  $\Sigma_i$ , which represents the Gaussian distribution as:

$$\Theta_i(\mathbf{x}) = e^{-\frac{1}{2}\mathbf{x}^T \Sigma_i^{-1} \mathbf{x}}. \quad (\text{A})$$

For the derivation of a physically meaningful covariance matrix, the following equivalent representation is applied:

$$\Sigma_i = \mathbf{R}_i \mathbf{S}_i \mathbf{S}_i^T \mathbf{R}_i^T, \quad (\text{B})$$

where the covariance matrix  $\Sigma_i$  is decomposed into a scaling factor  $\mathbf{S}_i$  and a rotation quaternion  $\mathbf{R}_i$ . Moreover, each primitive is assigned an opacity  $\sigma_i$  and a color  $\mathbf{c}_i$  to represent the appearance of the scene. In summary, 3DGS represents a scene as a set of Gaussian primitives:  $\Theta = \{(\mu_i, \mathbf{S}_i, \mathbf{R}_i, \sigma_i, \mathbf{c}_i)\}_{i=1}^N$ , where  $N$  indicates the number of primitives, and  $\mu_i$  represents the position of the center point. In practice, 3D Gaussians can be rendered in real time to compute the color  $\mathbf{C}$  of each pixel within the camera planes by blending  $N$  ordered Gaussians overlapping the pixel:

$$\mathbf{C} = \sum_{i \in N} \mathbf{c}_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j), \quad (\text{C})$$

where  $\alpha_i$  is calculated by evaluating  $\Theta_i$  with Eq. (A) multiplied by its opacity  $\sigma_i$ .

### E. Limitations and Social Impact

#### E.1. Limitations

The style transfer capability of this work is primarily limited by the underlying 2D diffusion model. Compared to 3D style transfer, image style transfer evidently has a higher success rate and more reliable results due to the larger-scale training and lower task difficulty. Therefore, despite being

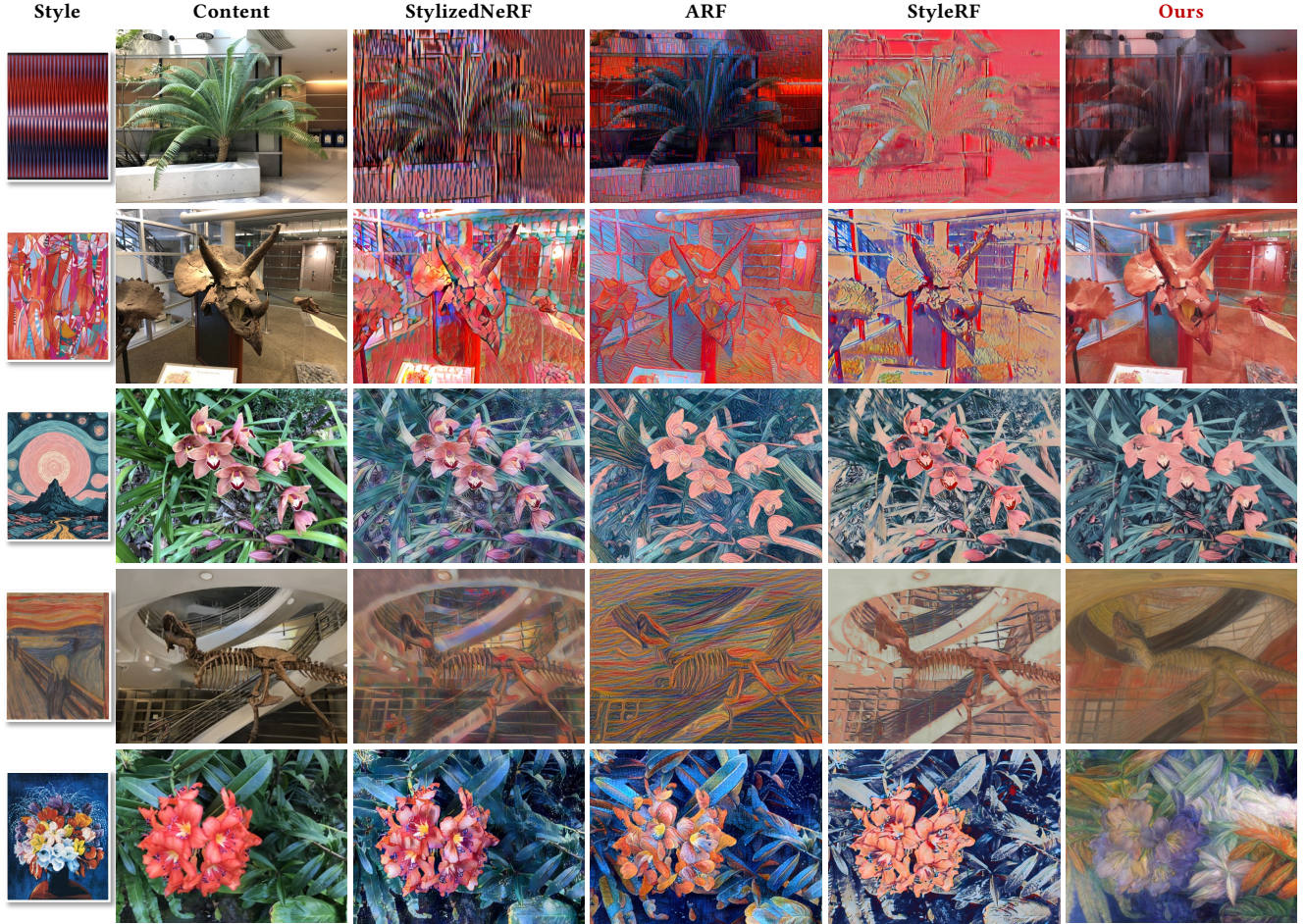


Figure A. **Additional qualitative comparison with NeRF-based style transfer methods.** Best viewed with zoom-in.

constrained by the image style transfer model, our method still features a higher upper bound than those 3D style transfer methods that do not leverage diffusion-based stylization priors. Additionally, our method is incapable of independently adjusting the intensity of texture style and geometric style within the stylized scene. We will further explore intensity-tunable 3D style transfer with decoupled texture and geometry in the future.

## E.2. Social Impact

In this work, all the 3D scenes and reference style images we utilized are publicly available. However, the potential outcomes of our method may lead to negative societal impacts if used for harmful purposes, including intellectual property concerns for protected works and misuse of the deceptive results generated. Consequently, we strongly advocate that users exercise responsibility when utilizing our method and strictly observe relevant ethical guidelines and legal regulations.

## References

- [1] Yi-Hua Huang, Yue He, Yu-Jie Yuan, Yu-Kun Lai, and Lin Gao. Stylizednerf: consistent 3d scene stylization as stylized nerf via 2d-3d mutual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18342–18352, 2022. 1
- [2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1
- [3] 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. 1
- [4] Kai Zhang, Nick Kolkin, Sai Bi, Fujun Luan, Zexiang Xu, Eli Shechtman, and Noah Snavely. Arf: Artistic radiance fields. In *European Conference on Computer Vision*, pages 717–733. Springer, 2022. 1



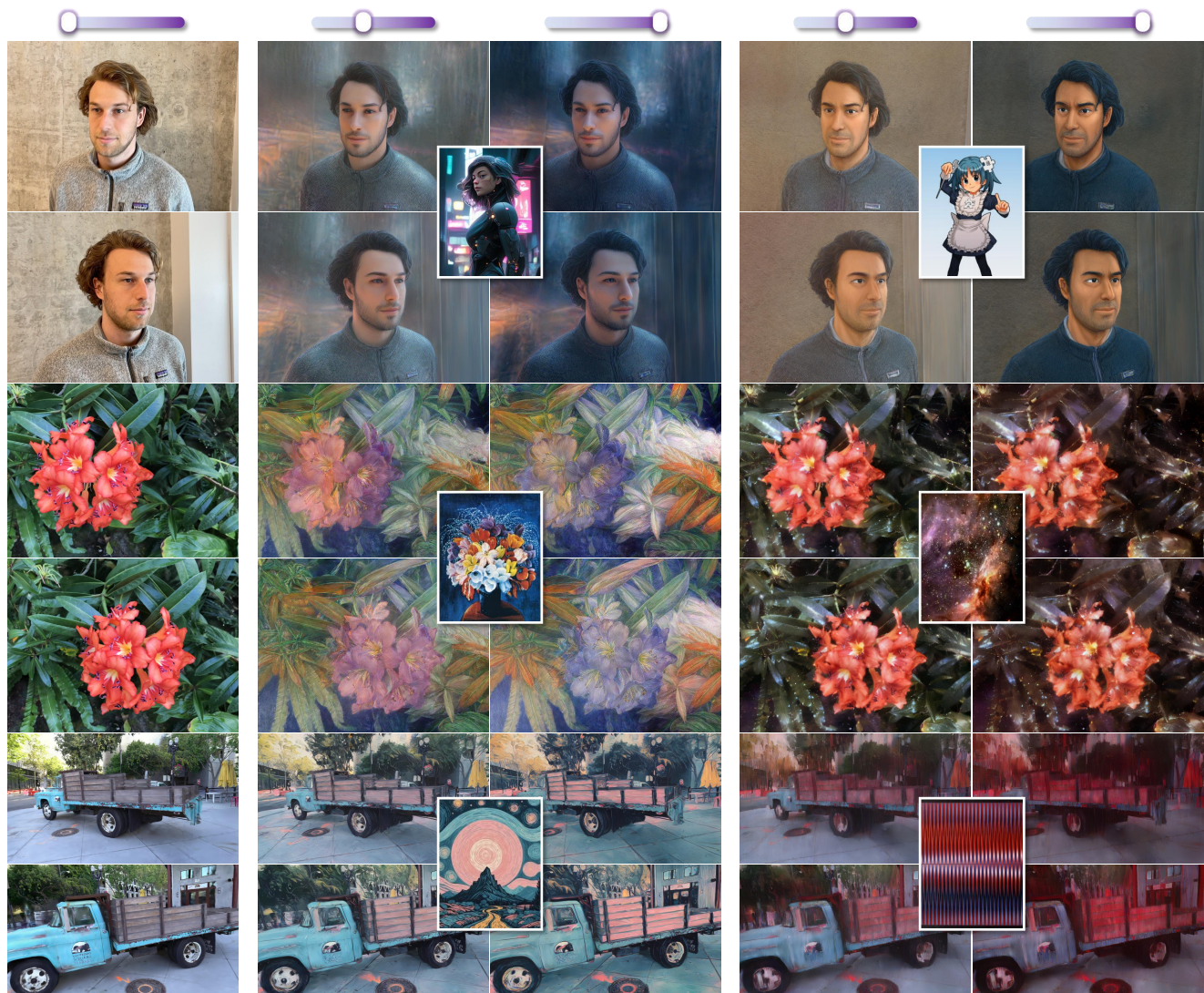


Figure B. More visualization results of intensity-tunable style injection.






Scene	Original Scene Description	Stylized Scene Description	Reference Style Image
Face	“A man with curly hair in a grey jacket”	“a Mona Lisa-style portrait of a man with curly hair”	
Truck	“A vintage blue pickup truck with a wooden cabin.”	“A vintage pickup truck with a wooden cabin featuring vertical dark red stripes.”	
Train	“A green Western Pacific train on the tracks.”	“An oil painting of a Western Pacific train on the tracks.”	
Lego	“A Lego model of a yellow bulldozer.”	“A Lego model of a bulldozer with an abstract style in shades of pink.”	
Garden	“A garden scene with a round table holding a vase.”	“A cosmic-style depiction of a garden with a round table and vase.”	

Table A. Detailed data for calculating CLIP similarity and CLIP directional similarity.

Dimension	#Point	Description
Content Consistency	1	Very poor, the content is entirely unrecognizable, with severe distortion of texture and geometry.
	2	Rather poor, the content is almost unrecognizable, with obvious distortion of texture and geometry.
	3	Acceptable, the content is basically recognizable, with some distortion of texture and geometry.
	4	Fairly good, the content is clearly recognizable, with minor distortion of texture and geometry.
	5	Very good, the content is perfectly recognizable, with almost no distortion of texture and geometry.
Style Consistency	1	Very poor, the color and artistic style are completely inconsistent with the reference image.
	2	Rather poor, the color and artistic style are significantly inconsistent with the reference image.
	3	Acceptable, the color and artistic style are basically consistent with the reference image.
	4	Fairly good, the color and artistic style are highly consistent with the reference image.
	5	Very good, the color and artistic style are completely consistent with the reference image.
Visual Appeal	1	Very poor, severe artifacts, very blurry textures, overall visual effect is very poor.
	2	Rather poor, numerous artifacts, relatively blurry textures, overall visual effect is not satisfactory.
	3	Acceptable, moderate artifacts, basically clear textures, overall visual effect is basically satisfactory.
	4	Fairly good, fewer artifacts, relatively clear textures, overall visual effect is satisfactory.
	5	Very good, minimal artifacts, clear textures, overall visual effect is very satisfactory.

Table B. Detailed evaluation criteria for the user study.