

Geometry-Aware Style Transfer in 3D Gaussian Splatting

Min Hyeok Bang^{*1}, Jun Hyeong Kim^{*1}, Seung-Wook Kim^{†2}, and
Se-Ho Lee^{†1}

¹ Department of Computer Science and Artificial Intelligence/Center for Advanced Image Information Technology, Jeonbuk National University, Jeonju-si, South Korea

² Division of Electronic and Communication Engineering, Pukyong National University, Busan, South Korea
{bunsss77,kjun627}@jbnu.ac.kr, swkim@pknu.ac.kr, seholee@jbnu.ac.kr

Abstract. In this paper, we present a novel geometry-aware style transfer framework for 3D Gaussian splatting (3DGS) that simultaneously transfers appearance attributes and geometric structures. Unlike prior works that primarily focus on color-based stylization and often overlook structural adaptation, our method explicitly incorporates geometry adaptation through a decoupled optimization scheme that alternately updates color and geometry parameters. This strategy alleviates potential interference between color and geometry updates, leading to stable and consistent scene-level geometry transformation. The decoupled optimization is enabled by the proposed geometry-aware contrastive feature matching (GCFM). GCFM integrates RGB, depth, and edge cues into a contrastive objective and is employed in both optimization phases to effectively transfer structural characteristics from style images to Gaussian primitives. Extensive experiments show that our approach achieves superior performance in both qualitative fidelity and quantitative metrics, significantly outperforming existing 3DGS-based stylization methods. Our code is available at <https://github.com/oweixx/gast>.

Keywords: 3D Gaussian Splatting · Decoupled Optimization · Geometry-Aware Stylization

1 Introduction

Neural style transfer [5] sparked extensive research on transferring the artistic characteristics of a style image onto a content image by manipulating color, texture, and high-level feature statistics. Several works have improved visual quality and user controllability of style transfer using CNNs [13, 32], GANs [11, 19, 40], and diffusion models [33]. While these approaches achieve impressive

^{*} These authors contributed equally.

[†] Corresponding authors.

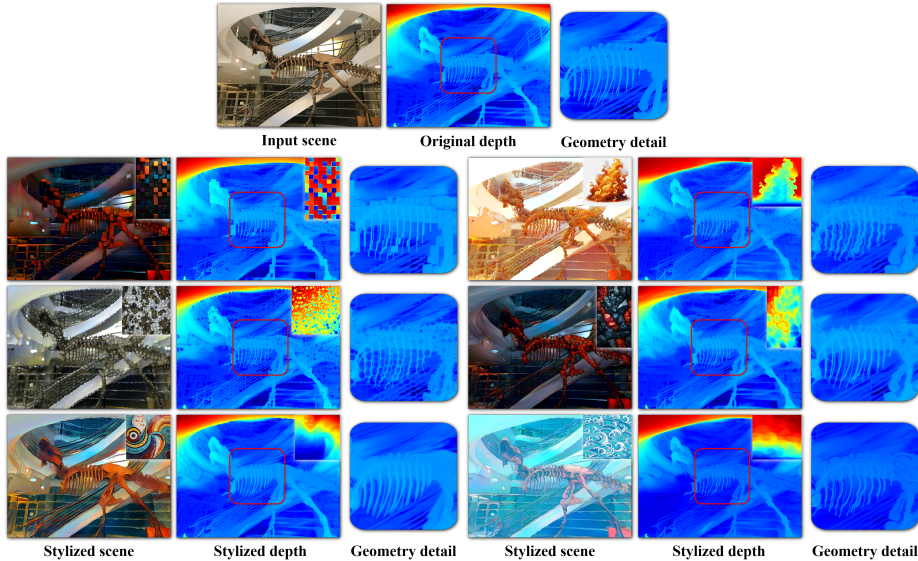


Fig. 1: Examples of geometry-aware style transfer in 3D Gaussian splatting (3DGS). Given an input scene and various style guides, our method stylizes both color and geometry while preserving overall scene structure. Highlighted depth regions show how geometric details adaptively reflect structural cues from the target style.

results on single-view images, they are inherently limited to 2D representations and cannot maintain consistent stylization across different viewpoints. With the advent of neural radiance fields (NeRFs) [24], some researchers began exploring 3D style transfer to render stylized scenes with a consistent appearance across different viewpoints. NeRF-based methods [12, 14, 20, 25, 35, 37, 39] successfully introduced geometric and structural awareness into style transfer, but they suffer from heavy optimization costs and slow rendering, which limit their practical use in interactive applications.

As a powerful alternative for real-time novel view synthesis, 3D Gaussian splatting (3DGS) [15] has been recently proposed. By modeling a scene as a set of anisotropic 3D Gaussians with learnable positions, scales, colors, and opacities, 3DGS achieves both high-quality rendering and remarkable efficiency compared to NeRFs. Its explicit and compact nature has made it attractive for applications beyond photorealistic reconstruction, such as interactive graphics and immersive content creation, leading to active exploration of 3D style transfer [4, 8, 18, 21, 22, 36].

However, applying style transfer to 3DGS often introduces a significant challenge: the rendered output is far more sensitive to changes in geometry than to changes in color. While color modifications mainly affect the surface appearance, direct manipulation of geometric parameters alters the underlying structure of a scene, which means that even minor miscalculations in geometry can drastically degrade visual quality. This difficulty explains why most existing style transfer methods tend to adopt a conservative approach. They impose strong constraints on geometry and focus on adapting color-related attributes for styl-

ization, thereby limiting the full expressive potential of 3D style transfer. Essentially, such a conservative approach is analogous to applying wood-patterned wallpaper onto a piece of furniture: while the paper conforms to the furniture’s overall shape and surface details, it cannot modify the underlying structure to embody the new style. As a result, the outcome would be a superficial stylization that lacks the authentic geometric character of real wood—its grooves, indentations, and structural irregularities. In contrast, our work treats both geometry and appearance as active components in the stylization process. We move beyond merely *painting* a surface and instead *sculpt* the stylistic features by actively shaping both the form and texture.

To this end, we propose a novel geometry-aware style transfer framework for 3DGS that transfers both stylistic attributes and geometric structures, as illustrated in Fig. 1. Given a 3DGS representation reconstructed from an input scene, our method applies the target style through a decoupled optimization process that alternately refines geometry and color parameters. The decoupled geometry and color optimization is driven by the proposed geometry-aware contrastive feature matching (GCFM). To ensure a holistic alignment, GCFM leverages joint representations from color, depth, and edge modalities. A contrastive objective then guides the optimization by pulling each rendered feature toward its most similar counterpart in the style image while pushing it away from the most dissimilar one. Experimental results demonstrate that our method achieves superior performance both qualitatively and quantitatively. Our main contributions are summarized as follows.

- We propose a simple yet effective geometry-aware style transfer framework for 3DGS, built upon a decoupled optimization scheme that alternates between color and geometry updates for stable and consistent stylization.
- We introduce GCFM, a geometry-aware contrastive feature matching scheme that aligns the scene’s geometry with the style’s distinctive structures by optimizing a contrastive loss across multi-modal representations incorporating color, depth, and edge cues.
- Experimental results demonstrate that our method produces visually faithful stylization and consistently outperforms existing approaches in both qualitative and quantitative evaluations.

2 Related Work

2.1 Neural Style Transfer

Gatys *et al.* [5] pioneered neural style transfer by using Gram matrices of CNN feature activations to encode image style, capturing correlations between feature channels while discarding spatial information. Following this work, research expanded rapidly in several directions: feed-forward networks for real-time style transfer [13, 32], adaptive instance normalization (AdaIN) [10] for arbitrary style transfer, and GAN-based approaches for unpaired domain-to-domain translation [11, 19, 40]. Recently, diffusion models enable text-driven and highly controllable style transfer with unprecedented fidelity [33]. Despite these advances,

image-based methods remain inherently limited to single-view domains, making them unsuitable for 3D scenarios where strict multi-view consistency is essential.

2.2 3D Style Transfer

3D style transfer aims to achieve consistent stylization across novel views. Early studies explored explicit scene representations such as point clouds [9] and meshes [7]. With the emergence of NeRFs [24], subsequent approaches embedded style information into volumetric representations to achieve multi-view consistency [12, 14, 20, 35, 37]. However, the high-dimensional implicit nature and the slow rendering speed of NeRF posed challenges for real-time applications and direct geometric manipulation.

The advent of 3DGS offers a compelling alternative with real-time rendering capabilities and explicit geometric primitives. In [21], zero-shot style transfer was introduced, where VGG features are embedded into Gaussian primitives to maintain view coherence. Based on [21], Kovács *et al.* [18] pruned and split Gaussians and incorporated multi-scale style losses for enhanced style fidelity. In [36], user-controllable parameters for stylization (scale, region, and color) and losses to avoid artifacts were introduced. Liu *et al.* [22] improved global style consistency by aligning scene content with style features using segmentation masks and regularization. Galerne *et al.* [4] optimized only the constant color component of each Gaussian to match multi-scale style statistics. Recently, Howil *et al.* [8] used CLIP guidance to enable both image- and text-guided style transfer across images, videos, 3D objects, and even 4D dynamic scenes within a unified Gaussian splatting framework.

Despite recent progress, most 3DGS-based style transfer methods remain limited to appearance adaptation. Modifying geometry is highly unstable and can easily distort the underlying scene structure, so most methods adopt a conservative design that preserves geometry. For example, [21] and [4] update only color attributes while keeping geometry fixed; [18] introduces limited geometric adaptation through Gaussian splitting during pre-processing and fine-tuning stages; and [36] and [22] employ depth preservation losses to maintain geometric fidelity rather than to actively incorporate geometry as a stylistic element. [8] enables joint optimization of color and geometry, but it still places primary emphasis on appearance adaptation, leaving geometry-oriented stylization relatively under-explored. While these strategies yield stable results, they fall short of treating geometry as a medium of style. In this work, we address this gap by exploring geometric stylization as a means to achieve more expressive and structurally consistent 3D style transfer.

3 Proposed Method

The goal of this paper is to transfer the artistic style from a 2D style image \mathbf{I}^S to a 3D scene constructed from a set of content images $\mathbb{I}^C = \{\mathbf{I}_i^C\}_{i=1}^M$ with spatial and geometric coherence. Fig. 2 illustrates the overall framework of the proposed style transfer method leveraging 3DGS [15]. As a preliminary step, a

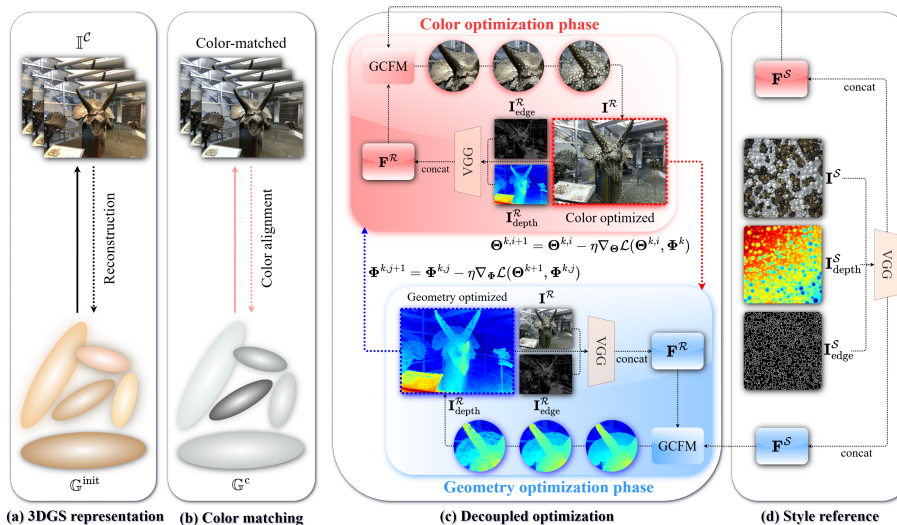


Fig. 2: Overview of the proposed geometry-aware style transfer framework for 3DGS. The pipeline consists of (a) 3DGS representation, (b) color matching, and (c) decoupled optimization between color and geometry optimization phases guided by multi-modal (color, depth, edge) features from (d) the style reference via the proposed GCFM.

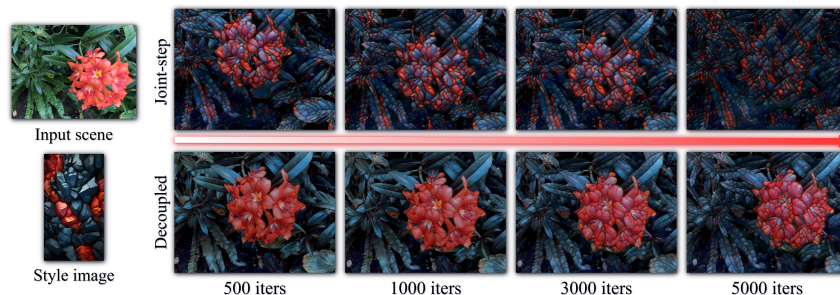


Fig. 3: Comparison between joint-step (top) and decoupled (bottom) optimization. The joint-step optimization updates color and geometry simultaneously, while our decoupled optimization alternates them (both run for the same total iterations).

3DGS representation $\mathbb{G}^{\text{init}} = \{\mathbf{G}_n := (\theta_n, \mu_n, \Sigma_n, \alpha_n)\}_{n=1}^N$ is reconstructed from \mathbb{I}^c , where each Gaussian primitive \mathbf{G}_n is defined by the color representation θ_n , its 3D center position $\mu_n \in \mathbb{R}^3$, its 3D covariance matrix $\Sigma_n \in \mathbb{R}^{3 \times 3}$, and its opacity $\alpha_n \in [0, 1]$. We initialize the 3D style transfer by color matching [37], obtaining \mathbb{G}^c . This process aligns the color statistics of the content images in \mathbb{I}^c with those of the style image \mathbb{I}^S using a linear transformation, ensuring the Gaussian primitives start with a color distribution closer to the target style. Then, we transfer the target style using a decoupled optimization scheme, which alternatively refines the color and geometric parameters. This alternating process is key to stabilizing the transfer process in 3DGS. Both optimization phases are driven by the proposed GCFM, which effectively guides both the color and geometry refinement. In the following sections, we describe our method in detail.

3.1 Decoupled Optimization for 3D Style Transfer

Most existing 3DGS-based style transfer methods [4, 18, 21, 22, 36] concentrate on transferring appearance attributes, whereas geometric stylization has remained a challenging problem. Let the Gaussian representation \mathbb{G} decomposed into color parameters $\Theta = \{\theta_n\}_{n=1}^N$ and geometry parameters $\Phi = \{\phi_n\}_{n=1}^N$. Our ultimate goal is to jointly optimize the color parameters Θ and the geometry parameters Φ to minimize a global style transfer loss $\mathcal{L}(\Theta, \Phi)$. For the color parameters, we optimize only the zeroth-order (DC) component of the spherical harmonics color representation, which corresponds to the base RGB color, while keeping the higher-order coefficients fixed, following [4]. The geometry parameters are defined as $\phi_n = \{\mu_n, \Sigma_n, \alpha_n\}$ for position, covariance, and opacity.

However, a naive simultaneous optimization of both parameter sets often results in interference between the color and geometry updates. This leads to color over-stylization, in which the optimization process sacrifices the essential structure of the scene due to a misguided attempt to embed low-level stylistic patterns directly into the geometry. As illustrated in the first row of Fig. 3, the flower’s structure is progressively corrupted as iterations proceed, with exaggerated color patterns ignoring the underlying geometry and ultimately causing a complete loss of perceptual fidelity.

To overcome these issues and enable stable geometry-aware style transfer, we introduce a *decoupled optimization scheme*. The entire optimization proceeds for K outer cycles. Within each cycle k , we break the global problem into two alternating subproblems as follows:

$$\Theta^{k+1} = \arg \min_{\Theta} \mathcal{L}(\Theta, \Phi^k), \quad (1)$$

$$\Phi^{k+1} = \arg \min_{\Phi} \mathcal{L}(\Theta^{k+1}, \Phi). \quad (2)$$

In this two-phase process, we first update the color parameters Θ using the geometry fixed from the previous cycle, Φ^k . Subsequently, we perform geometric stylization by updating the geometry parameters Φ using the newly refined colors, Θ^{k+1} . In practice, we approximate the solutions to these subproblems in Eqs. (1) and (2) by performing a fixed number of inner gradient descent steps for each phase. Each outer cycle consists of two sequential phases:

Color optimization phase. To solve for Θ^{k+1} , we fix the geometry parameters Φ^k and update the color parameters for N_c inner steps. For $i = 0, \dots, N_c - 1$, the color parameters are updated as

$$\Theta^{k,i+1} = \Theta^{k,i} - \eta \nabla_{\Theta} \mathcal{L}(\Theta^{k,i}, \Phi^k), \quad (3)$$

where η is the learning rate. The final color parameters after N_c steps, denoted as $\Theta^{k+1} = \Theta^{k,N_c}$, are then used in the following phase.

Geometry optimization phase. Subsequently, to solve for Φ^{k+1} , we hold the newly updated color parameters Θ^{k+1} fixed and refine the geometry parameters

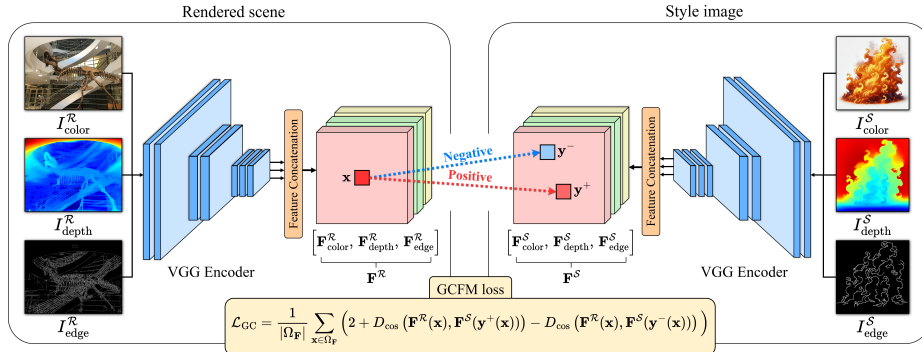


Fig. 4: Conceptual illustration of the GCFM process.

for N_g inner steps. For $j = 0, \dots, N_g - 1$, the geometry parameters are updated as

$$\Phi^{k,j+1} = \Phi^{k,j} - \eta \nabla_{\Phi} \mathcal{L}(\Theta^{k+1}, \Phi^{k,j}). \quad (4)$$

The resulting geometry parameters, $\Phi^{k+1} = \Phi^{k,N_g}$ are then carried forward to the next outer cycle, completing one full round of optimization.

By alternating color and geometry updates in this way, the scheme mitigates optimization conflicts between the two optimization processes and establishes a mutual guidance loop: color updates provide a stable reference for geometry refinement, and refined geometry supports consistent color adaptation. As a result, the optimization converges more stably and achieves reliable geometry-aware style transfer in 3DGS, as illustrated in the second row of Fig. 3.

3.2 Geometry-Aware Contrastive Feature Matching (GCFM)

Feature matching-based style transfer [37] has been widely adopted in many 3D style transfer approaches [14, 22, 35]. In this work, we utilize a multi-modal representation for feature matching, incorporating color, depth, and edge cues to enhance geometry awareness. Depth maps can provide coarse spatial information, but often suffer from noise and missing details. Edge maps, on the other hand, highlight distinct structural boundaries and facilitate the alignment of shapes at the boundary level between content and style images. By jointly leveraging color, depth, and edge modalities, the proposed GCFM offers complementary guidance that enforces consistent 3D style transfer and alleviates the limitations of each individual cue. Specifically, the depth map of the rendered image is obtained directly from the 3DGS rendering process, while the depth map of the style image is estimated using [34]. Since complex textures and fine patterns, which are common in style images, tend to produce noisy edges in the color domain, we extract more structurally reliable edges from the depth map of the style image. The final edge maps are thus generated using the standard Canny detector [2] from two different sources: the rendered image and the depth map of the style image.

Fig. 4 presents an overview of the GCFM process. For a given source image \mathbf{I}^s , where $s \in \{\mathcal{R}, \mathcal{S}\}$ denotes the rendered and style images, respectively, we use

a pretrained VGG network [29] to extract deep features from the three modalities: color ($\mathbf{F}_{\text{color}}^s$), depth ($\mathbf{F}_{\text{depth}}^s$), and edge ($\mathbf{F}_{\text{edge}}^s$). Matching these features independently, however, can lead to inconsistencies, where appearance, shape, and edge patterns fail to align coherently. Therefore, to ensure holistic alignment, we employ a joint feature matching strategy by concatenating the features from all three modalities. For a given source s , a joint feature map is defined as $\mathbf{F}^s = [\mathbf{F}_{\text{color}}^s, \mathbf{F}_{\text{depth}}^s, \mathbf{F}_{\text{edge}}^s]$, where $[\cdot, \cdot, \cdot]$ indicates channel-wise concatenation.

With the joint feature maps defined, it is required to align the structural features between the rendered and style images. Existing methods define positive pairs, *i.e.*, structurally similar features that are pulled closer in the embedding space [14, 37]. However, we argue that for effective 3D style transfer, it is insufficient to only reflect such similarities. To capture more discriminative style representations, the learning process should also explicitly contrast the rendered features against dissimilar features from the style source. Therefore, we additionally define negative pairs, representing dissimilar features that should be pushed apart, enabling the learning process to capture more discriminative style representations. We identify these pairs for each spatial location $\mathbf{x} = (u, v)$ in the rendered feature map $\mathbf{F}^{\mathcal{R}}$. The positive sample is its nearest neighbor in the style feature map $\mathbf{F}^{\mathcal{S}}$ while the negative sample is the farthest one. We find the spatial location of these samples, $\mathbf{y}^+(\mathbf{x})$ and $\mathbf{y}^-(\mathbf{x})$, by respectively minimizing and maximizing a feature distance as follows:

$$\mathbf{y}^+(\mathbf{x}) = \arg \min_{\mathbf{y}} D_{\cos}(\mathbf{F}^{\mathcal{R}}(\mathbf{x}), \mathbf{F}^{\mathcal{S}}(\mathbf{y})), \quad (5)$$

$$\mathbf{y}^-(\mathbf{x}) = \arg \max_{\mathbf{y}} D_{\cos}(\mathbf{F}^{\mathcal{R}}(\mathbf{x}), \mathbf{F}^{\mathcal{S}}(\mathbf{y})), \quad (6)$$

where $D_{\cos}(\cdot, \cdot) = 1 - \cos(\cdot, \cdot)$ represents the cosine distance. Based on the matched positive and negative indices, we formulate a single contrastive objective, called GCFM loss. The core principle is to treat each feature $\mathbf{F}^{\mathcal{R}}(\mathbf{x})$ from the rendered image as an anchor, pulling it toward its positive style counterpart $\mathbf{F}^{\mathcal{S}}(\mathbf{y}^+(\mathbf{x}))$ while pushing it away from its negative counterpart $\mathbf{F}^{\mathcal{S}}(\mathbf{y}^-(\mathbf{x}))$ during optimization. The final GCFM loss is thus defined as follows:

$$\mathcal{L}_{GC} = \frac{1}{|\Omega_{\mathbf{F}}|} \sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} \left(2 + D_{\cos}(\mathbf{F}^{\mathcal{R}}(\mathbf{x}), \mathbf{F}^{\mathcal{S}}(\mathbf{y}^+(\mathbf{x}))) - D_{\cos}(\mathbf{F}^{\mathcal{R}}(\mathbf{x}), \mathbf{F}^{\mathcal{S}}(\mathbf{y}^-(\mathbf{x}))) \right), \quad (7)$$

where $\Omega_{\mathbf{F}}$ is a set of all spatial locations of the feature map.

3.3 Loss Function

In this section, we present the loss function $\mathcal{L}(\Theta, \Phi)$ used for the proposed method to optimize the color parameters Θ and the geometry parameters Φ . For the sake of brevity, we simply denote the loss function as \mathcal{L} , which is given by

$$\mathcal{L} = \lambda_{GC} \mathcal{L}_{GC} + \lambda_{\text{cont}} \mathcal{L}_{\text{cont}} + \lambda_{\text{TV}} \mathcal{L}_{\text{TV}} + \delta_{\text{reg}} \mathcal{L}_{\text{reg}}, \quad (8)$$

where λ_{GC} , λ_{cont} , and λ_{TV} are empirically chosen hyper-parameters that control the relative contribution of each loss term. Here, $\delta_{reg} = 1$ during the geometry optimization phase, and $\delta_{reg} = 0$ during the color optimization phase. \mathcal{L}_{cont} , \mathcal{L}_{TV} , and \mathcal{L}_{reg} denote the content preservation loss, total variation (TV) loss, and geometry regularization loss, respectively. Each term is detailed as follows:

Content preservation loss. To ensure that the resulting stylized image preserves the structural fidelity of the original content, we adopt a perceptual loss based on VGG feature embeddings [29]. To this end, we minimize the distance between the VGG features of the rendered image $\mathbf{I}^{\mathcal{R}}$ and those of the content image \mathbf{I}^c as follows:

$$\mathcal{L}_{cont} = \|\psi(\mathbf{I}^{\mathcal{R}}) - \psi(\mathbf{I}^c)\|_2^2, \quad (9)$$

where $\psi(\cdot)$ denotes the feature representation extracted from a pre-trained VGG-16 network. The content preservation loss maintains the original structure of the underlying scene while allowing visual adaptation to the style image.

TV loss. To further improve the visual quality of the stylized results, we impose a TV regularization that encourages local spatial smoothness in the rendered image. Minimizing the TV loss effectively suppresses high-frequency noise and reduces artifacts along object boundaries. The TV loss is given by

$$\mathcal{L}_{TV} = \frac{1}{|\Omega_{\mathbf{I}}|} \sum_{\mathbf{x} \in \Omega_{\mathbf{I}}} [\|\mathbf{I}^{\mathcal{R}}(\mathbf{x}) - \mathbf{I}^{\mathcal{R}}(\mathbf{x} + \mathbf{e}_h)\|^2 + \|\mathbf{I}^{\mathcal{R}}(\mathbf{x}) - \mathbf{I}^{\mathcal{R}}(\mathbf{x} + \mathbf{e}_v)\|^2], \quad (10)$$

where $\Omega_{\mathbf{I}}$ is a set of all spatial locations of the rendered image, and $\mathbf{e}_h = (0, 1)$ and $\mathbf{e}_v = (1, 0)$ are the unit offset vectors for the horizontal and vertical neighbors, respectively.

Geometry regularization loss. Unconstrained geometry optimization may cause Gaussian primitives to degenerate into undesirable configurations, such as excessive expansion that produces global blurring or collapse into elongated shapes. To prevent such artifacts, we adopt a geometry regularization loss that penalizes deviations in opacity, scale, and rotation from their initial color-matched states, and additionally enforces depth consistency:

$$\mathcal{L}_{reg} = \lambda_{\alpha} \|\alpha - \alpha^c\|_2 + \lambda_s \|\mathbf{s} - \mathbf{s}^c\|_2 + \lambda_r \|\mathbf{r} - \mathbf{r}^c\|_2 + \lambda_d \|\mathbf{D}^{\mathcal{R}} - \mathbf{D}^c\|, \quad (11)$$

where the weighting coefficients λ_{α} , λ_s , λ_r , and λ_d are chosen empirically. Here, α , \mathbf{s} , and \mathbf{r} denote the opacity, scale, and rotation parameters of all Gaussian primitives, respectively, where the superscript c indicates their values after color matching with \mathbb{G}^c . $\mathbf{D}^{\mathcal{R}}$ denotes the rendered depth map from the current 3DGS model, and \mathbf{D}^c represents the reference depth map obtained from \mathbb{G}^c , respectively. Overall, this regularization anchors the stylization process to the original scene structure, enabling effective style-driven adaptations while preventing instability in geometry.

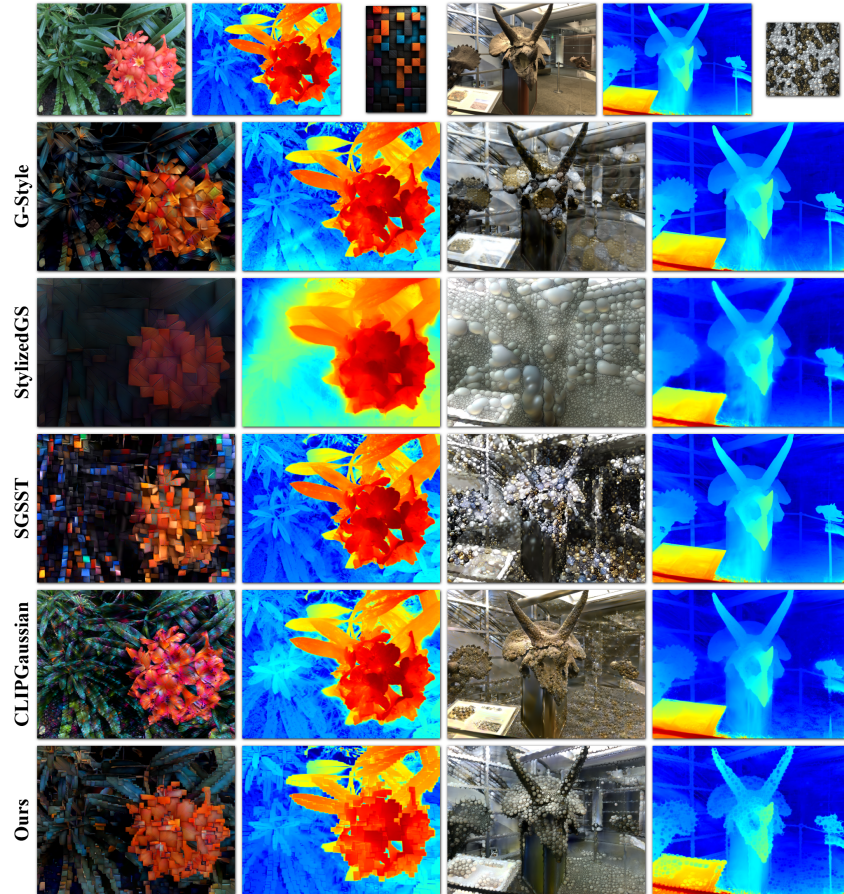


Fig. 5: Qualitative comparison of the *Flower* and *Horns* scenes. The top row shows the input scenes, rendered depth maps, and the corresponding style image for each scene. The subsequent rows present stylized results produced by different methods, including G-Style [18], StylizedGS [36], SGSST [4], CLIPGaussian [8], and Ours.

4 Experiments

4.1 Datasets

We evaluate our 3DGS-based style transfer method across a total of 72 scene–style combinations ($8 \text{ scenes} \times 9 \text{ styles}$).

Scenes. We use three representative multi-view scene datasets: LLFF [23] (4 forward-facing scenes), Tanks & Temples (T&T) [17] (2 complex indoor/outdoor 360° scenes), and MipNeRF-360 [1] (2 large-scale unbounded scenes). Together, these eight scenes provide a wide range of geometric structures and camera trajectories, from small bounded scenes to large-scale 360° environments, allowing robust evaluation of geometry-aware stylization.

Table 1: Quantitative comparison of style transfer quality (SIFID) and multi-view consistency. **Bold** and underline indicate the best and the second-best results.

Methods	SIFID↓	Short-Range Consistency		Long-Range Consistency	
		LPIPS↓	RMSE↓	LPIPS↓	RMSE↓
StyleGaussian [21]	3.3405	0.050	0.053	0.141	0.114
G-Style [18]	<u>1.2955</u>	0.045	0.035	0.118	0.085
StylizedGS [36]	2.2439	0.028	0.021	0.072	0.062
SGSST [4]	1.3936	0.048	0.048	0.129	0.109
CLIPGaussian [8]	3.3685	0.048	0.046	0.128	0.108
Ours	1.1736	<u>0.042</u>	<u>0.034</u>	<u>0.114</u>	<u>0.084</u>

Styles. We use a set of nine style exemplars gathered from multiple sources—four from [14], three from WikiArt [30], one from [8], and one curated from publicly available online art collections [3]. To ensure stylistic diversity and generalization, these exemplars are deliberately drawn from heterogeneous sources, covering both classical and modern art movements as well as 3D-aware styles. The selection emphasizes spatially structured patterns to evaluate the model’s ability to capture geometric as well as appearance-based stylistic characteristics.

4.2 Implementation Details

Our implementation is built upon the public PyTorch [27] codebase of 3DGS [15]. We adopt the proposed decoupled optimization scheme for geometry-aware style transfer, where color and geometry parameters are alternately updated. Algorithm A in the supplementary material summarizes the overall optimization procedure, illustrating how the color and geometry parameters are updated in an alternating manner under the proposed framework. Unless otherwise specified, the loss weights are set as $\lambda_{GC} = 2.0$, $\lambda_{cont} = 1 \times 10^{-3}$, $\lambda_{TV} = 0.02$, $\lambda_{\alpha} = 1.0$, $\lambda_s = 1.0$, $\lambda_r = 1.0$, and $\lambda_d = 10.0$. We perform the stylization process for $K = 30$ outer cycles using the Adam optimizer [16], with $N_c = 10$ color updates and $N_g = 90$ geometry updates per cycle, totaling 3,000 iterations. We set the learning rate $\eta = 1 \times 10^{-4}$ for all experiments, except for the T&T dataset [17], where it is reduced to 2.5×10^{-5} for stable convergence, following [18] which adopted a similar adjustment for scenes with higher geometric complexity. All experiments are conducted on a single NVIDIA RTX A6000 GPU with 48 GB of memory.

4.3 Comparison with State-of-the-Arts

Qualitative comparison. Fig. 5 compares our method with state-of-the-art 3DGS-based style transfer approaches [4, 8, 18, 36]. Unlike previous methods that primarily focus on color-based appearance transfer, our approach jointly adapts both appearance and geometric structure, producing visually richer and structurally coherent 3D stylization across all evaluated scenes. For instance, in the *Flower* scene on the left, our method successfully transforms the object’s shape to reflect the block-like structural patterns of the target style, rather than merely applying color changes. In the *Horns* scene on the right, our method transforms the original surfaces into particle-like structures while preserving fine geometric details and clear depth variations around the horn. This enhanced structural expressiveness leads to more immersive and visually compelling results overall.

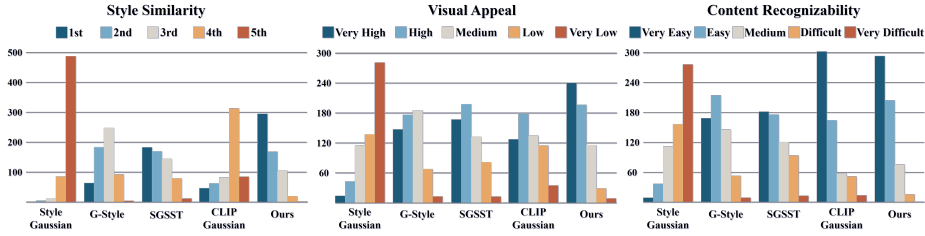


Fig. 6: Results of the user study comparing StyleGaussian [21], G-Style [18], SGSST [4], CLIPGaussian [8], and Ours.

Quantitative comparison. Table 1 summarizes the quantitative evaluation using the Single Image Fréchet Inception Distance (SIFID) [28], which measures the perceptual distance between each stylized image and its target style. Unlike FID [6], which requires large image sets, SIFID is suitable for evaluating single-scene stylization results. Each result is rendered from multiple viewpoints and averaged to ensure robustness. A lower SIFID indicates closer alignment with the target style distribution. As shown in the table, our method achieves the lowest SIFID score (1.1736), demonstrating superior style fidelity and perceptual alignment with the target style compared to previous methods.

Also in Table 1, we further evaluate the multi-view consistency of different methods under both short- and long-range settings. Following [21], one view is warped to another using optical flow [31] with softmax splatting [26], and stylization consistency is measured by masked RMSE and LPIPS [38]. While StylizedGS [36] reports the best consistency scores, this may stem from its overly smooth and desaturated renderings, which inherently favor consistency metrics rather than reflecting strong stylization effects.³ Excluding StylizedGS, our method achieves the best overall performance, ranking first in both LPIPS and RMSE across short- and long-range consistency settings. These results indicate that our framework achieves a better balance between stylization quality and multi-view consistency compared to other 3DGS-based methods.

4.4 User Study

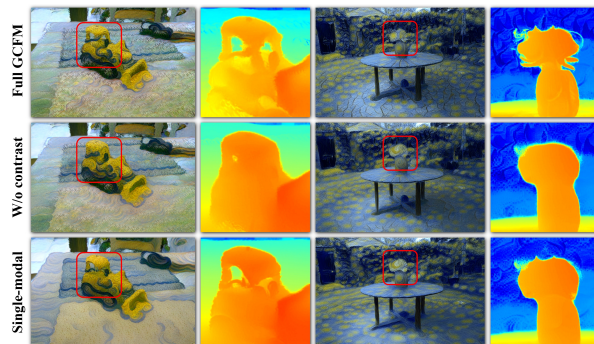
To assess the perceptual quality of the stylized renderings, we conducted an online user study involving 37 participants. Our method was compared against four representative 3DGS-based style transfer methods [4, 8, 18, 21]. We used four distinct 3D scenes and four style reference images, each containing spatially consistent texture and geometric cues—such as blocky patterns and particle-like elements—two of which are shown in Fig. 5. For each result, we asked participants to rate their (1) style similarity to the reference style image, (2) overall visual appeal, and (3) content recognizability of the original scene.

The analyzed outcomes of the user study are shown in Fig. 6. Participants consistently rated our approach as the best across all three evaluation criteria,

³ A detailed comparison with StylizedGS, including qualitative analysis, is provided in Section D of the supplementary material.

Table 2: Comparison of average stylization times across various 3DGS-based methods.

Methods	Stylization Time
StyleGaussian [21]	410m 13s
G-Style [18]	29m 05s
StylizedGS [36]	3m 32s
SGSST [4]	58m 52s
CLIPGaussian [8]	22m 46s
Ours	<u>14m 14s</u>

**Fig. 7:** Ablation study on GCFM. The 1st and 3rd columns show the stylized rendered views, while the 2nd and 4th columns display the corresponding enlarged depth maps.

demonstrating its effectiveness in producing high-quality stylizations that are faithful to the target style while preserving scene content.

4.5 Complexity Analysis

Table 2 compares the average stylization time of representative 3DGS-based methods. While StyleGaussian [21] requires several hours of optimization for a single style, our method completes the stylization in about fourteen minutes, showing a substantial improvement in efficiency. Although StylizedGS [36] achieves the fastest runtime, it often compromises geometric consistency and style fidelity, as discussed in Section D of the supplementary material. Overall, our approach maintains a favorable trade-off between computational speed and visual quality, providing a practical and scalable solution for 3D scene stylization. Detailed per-scene analysis on stylization time and resource usage is provided in Section A of the supplementary material.

4.6 Ablation Studies

Ablation study on GCFM. Fig. 7 presents the ablation results of GCFM, showing how each component contributes to geometry-aware stylization. To capture geometric cues beyond appearance, GCFM integrates both depth and edge modalities into the feature representation. Additionally, a contrastive objective is introduced to enforce stronger feature alignment, enhancing the distinctiveness of geometry-related patterns. As shown in the figure, removing the contrastive objective (w/o contrast) weakens feature alignment, leading to a loss of structural distinctiveness in the depth maps. Similarly, using a single-modal representation (single-modal) results in oversimplified geometric structures, which

Table 3: Ablation study on optimization strategies.

Methods	SIFID↓	Short-Range Consistency		Long-Range Consistency	
		LPIPS↓	RMSE↓	LPIPS↓	RMSE↓
Only geometry	2.5496	0.045	0.040	0.113	0.096
Only color	1.5392	0.042	0.037	0.115	<u>0.086</u>
Joint-step	<u>1.4220</u>	0.042	<u>0.036</u>	0.122	0.087
Decoupled (Ours)	1.1736	0.042	0.034	<u>0.114</u>	0.084

indicates that multi-modal integration is essential for capturing fine geometric contours. Overall, the full GCFM yields sharper contours and faithfully reproduces the underlying geometric structures, confirming that both the contrastive loss and multi-modal feature integration are essential for capturing and utilizing complex scene geometry effectively.

Ablation study on optimization strategies. Table 3 compares different optimization strategies for geometry-aware stylization. When updating only geometry, the model primarily modifies structural parameters while keeping color parameters unchanged. This results in poor stylization fidelity and the highest SIFID due to insufficient adaptation to the target style, and also causes noticeable cross-view inconsistency from unstable geometric updates. In contrast, optimizing only color achieves lower SIFID by transferring color appearance more effectively, but it still fails to capture geometry-dependent style cues, leading to weak structural adaptation and reduced consistency across views. The joint-step approach, which updates color and geometry simultaneously, achieves moderate improvement in SIFID but suffers from color over-stylization and unstable convergence, ultimately leading to inferior performance. Finally, our decoupled optimization scheme achieves the best trade-off between stylization fidelity and geometric stability, showing the lowest SIFID and consistent improvements in both short- and long-range LPIPS and RMSE. These results show that separating the color and geometry updates stabilizes optimization and enables harmonious integration of stylistic appearance and geometric adaptation.

Additional results and analyses are provided in Sections B–E of the supplementary material, including user study details, ablation studies, and extended qualitative results that further validate the effectiveness of the proposed method.

5 Conclusions

In this work, we presented a geometry-aware style transfer framework for 3DGS that jointly stylizes both appearance and geometry, achieving stable optimization and consistent stylization results. To address the optimization instability caused by simultaneous updates of color and geometry, we introduced a decoupled optimization scheme that alternates between the two, enabling mutual guidance between appearance and structure. During this process, we employed the proposed GCFM loss, which integrates RGB, depth, and edge cues into a unified contrastive objective to effectively capture geometric and structural characteristics of the target style. Extensive experiments across various 3D scenes demonstrate that our method achieves superior performance in both stylization quality and multi-view consistency compared to recent 3DGS-based methods.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grants funded by the Korean government(MSIT)(RS-2026-25473686 and RS-2026-25489106) and by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation) in 2026(2022-0-01067).

References

1. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In: CVPR. pp. 5470–5479 (2022)
2. Canny, J.: A computational approach to edge detection. *IEEE TPAMI PAMI-8*(6), 679–698 (1986)
3. Freepik: Seamless pattern with water waves and splashes. https://www.freepik.com/premium-vector/seamless-pattern-with-water-waves-splashes_9413868.htm
4. Galerne, B., Wang, J., Raad, L., Morel, J.M.: SGSST: Scaling Gaussian splatting style transfer. In: CVPR. pp. 26535–26544 (June 2025)
5. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: CVPR. pp. 2414–2423 (2016)
6. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *NeurIPS*. vol. 30. Curran Associates, Inc. (2017)
7. Höllein, L., Johnson, J., Nießner, M.: StyleMesh: Style transfer for indoor 3D scene reconstructions. In: CVPR. pp. 6198–6208 (2022)
8. Howil, K., Waczyńska, J., Borycki, P., Dziarmaga, T., Mazur, M., Spurek, P.: CLIP-Gaussian: Universal and multimodal style transfer based on Gaussian splatting. In: *NeurIPS* (2025)
9. Huang, H.P., Tseng, H.Y., Saini, S., Singh, M., Yang, M.H.: Learning to stylize novel views. In: ICCV. pp. 13869–13878 (October 2021)
10. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: ICCV. pp. 1501–1510 (2017)
11. Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: ECCV. pp. 172–189 (2018)
12. Huang, Y.H., He, Y., Yuan, Y.J., Lai, Y.K., Gao, L.: StylizedNeRF: Consistent 3D scene stylization as stylized NeRF via 2D-3D mutual learning. In: CVPR. pp. 18342–18352 (June 2022)
13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV. pp. 694–711 (2016)
14. Jung, H., Nam, S., Sarafianos, N., Yoo, S., Sorkine-Hornung, A., Ranjan, R.: Geometry transfer for stylizing radiance fields. In: CVPR. pp. 8565–8575 (June 2024)
15. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* **42**(4) (July 2023)

16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2015)
17. Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V.: Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM TOG* **36**(4), 1–13 (2017)
18. Kovács, Á.S., Hermosilla, P., Raidou, R.G.: G-Style: Stylized Gaussian splatting. In: Computer Graphics Forum. vol. 43, p. e15259. Wiley Online Library (2024)
19. Lee, H.Y., Tseng, H.Y., Huang, J.B., Singh, M., Yang, M.H.: Diverse image-to-image translation via disentangled representations. In: ECCV. pp. 35–51 (2018)
20. Liu, K., Zhan, F., Chen, Y., Zhang, J., Yu, Y., El Saddik, A., Lu, S., Xing, E.P.: StyleRF: Zero-shot 3D style transfer of neural radiance fields. In: CVPR. pp. 8338–8348 (June 2023)
21. Liu, K., Zhan, F., Xu, M., Theobalt, C., Shao, L., Lu, S.: StyleGaussian: Instant 3D style transfer with Gaussian splatting. In: SIGGRAPH Asia (2024)
22. Liu, W., Liu, Z., Yang, X., Sha, M., Li, Y.: ABC-GS: Alignment-based controllable style transfer for 3D Gaussian splatting. In: ICME (2025)
23. Mildenhall, B., Srinivasan, P.P., Ortiz-Cayon, R., Kalantari, N.K., Ramamoorthi, R., Ng, R., Kar, A.: Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM TOG* **38**(4) (2019)
24. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) ECCV. pp. 405–421 (2020)
25. Nguyen-Phuoc, T., Liu, F., Xiao, L.: SNeRF: Stylized neural implicit representations for 3D scenes. *ACM Trans. Graph.* **41**(4) (Jul 2022)
26. Niklaus, S., Liu, F.: Softmax splatting for video frame interpolation. In: CVPR. pp. 5437–5446 (June 2020)
27. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: PyTorch: An imperative style, high-performance deep learning library. In: NeurIPS (2019)
28. Shaham, T.R., Dekel, T., Michaeli, T.: SinGAN: Learning a generative model from a single natural image. In: ICCV. pp. 4570–4580 (October 2019)
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)
30. Tan, W.R., Chan, C.S., Aguirre, H., Tanaka, K.: Improved ArtGAN for conditional synthesis of natural image and artwork. *IEEE TIP* **28**(1), 394–409 (2019), <https://doi.org/10.1109/TIP.2018.2866698>
31. Teed, Z., Deng, J.: RAFT: Recurrent all-pairs field transforms for optical flow. In: ECCV. p. 402–419 (2020)
32. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Instance normalization: The missing ingredient for fast stylization. *CoRR* **abs/1607.08022** (2016)
33. Wang, Z., Zhao, L., Xing, W.: StyleDiffusion: Controllable disentangled style transfer via diffusion models. In: ICCV. pp. 7677–7689 (October 2023)
34. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything V2. *CoRR* **abs/2406.09414** (2024)
35. Zhang, D., Fernandez-Labrador, C., Schroers, C.: CoARF: Controllable 3D artistic style transfer for radiance fields. In: 2024 International Conference on 3D Vision (3DV). pp. 612–622 (2024)
36. Zhang, D., Yuan, Y.J., Chen, Z., Zhang, F.L., He, Z., Shan, S., Gao, L.: StylizedGS: Controllable stylization for 3D Gaussian splatting. *IEEE TPAMI* pp. 1–13 (2025)
37. Zhang, K., Kolkin, N., Bi, S., Luan, F., Xu, Z., Shechtman, E., Snavely, N.: ARF: Artistic radiance fields. In: ECCV. pp. 717–733 (2022)

38. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR. pp. 586–595 (2018)
39. Zhang, Y., He, Z., Xing, J., Yao, X., Jia, J.: Ref-NPR: Reference-based non-photorealistic radiance fields for controllable scene stylization. In: CVPR. pp. 4242–4251 (June 2023)
40. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: ICCV. pp. 2242–2251 (2017)