

Mirror-3DGS: Incorporating Mirror Reflections into 3D Gaussian Splatting

Jiarui Meng^{1†}, Haijie Li^{1†}, Yanmin Wu¹, Qiankun Gao¹, Shuzhou Yang¹,
 Jian Zhang¹ , Siwei Ma²

¹ School of Electronic and Computer Engineering, Peking University

² School of Computer Science, Peking University

Abstract. 3D Gaussian Splatting (3DGS) has marked a significant breakthrough in the realm of 3D scene reconstruction and novel view synthesis. However, 3DGS, much like its predecessor Neural Radiance Fields (NeRF), struggles to accurately model physical reflections, particularly in mirrors that are ubiquitous in real-world scenes. This oversight mistakenly perceives reflections as separate entities that physically exist, resulting in inaccurate reconstructions and inconsistent reflective properties across varied viewpoints. To address this pivotal challenge, we introduce **Mirror-3DGS**, an innovative rendering framework devised to master the intricacies of mirror geometries and reflections, paving the way for the generation of realistically depicted mirror reflections. By ingeniously incorporating mirror attributes into the 3DGS and leveraging the principle of plane mirror imaging, Mirror-3DGS crafts a mirrored viewpoint to observe from behind the mirror, enriching the realism of scene renderings. Extensive assessments, spanning both synthetic and real-world scenes, showcase our method's ability to render novel views with enhanced fidelity in real-time, surpassing the state-of-the-art Mirror-NeRF specifically within the challenging mirror regions. Our code will be made publicly available for reproducible research.

Keywords: Gaussian Splatting · Mirror Scene · Novel View Synthesis

1 Introduction

The 3D reconstruction and novel view synthesis are crucial in numerous applications, including the movie industry, computer gaming, virtual reality, and autonomous navigation. However, accurately capturing and rendering reflections from mirrors, which are prevalent in real-world scenes and significantly increase the visual complexity, remains a significant challenge for these technologies.

Traditional methods [25, 31, 32, 36, 53] have attempted to simulate surface reflections by decomposing the material properties of objects and lighting conditions, thereby handling scenes with mirrors. However, these methods often face limitations in accurately inferring the surfaces and mirror components, struggling to effectively manage pure mirror reflections. With the advancement of Neural

[†] means equal contribution.

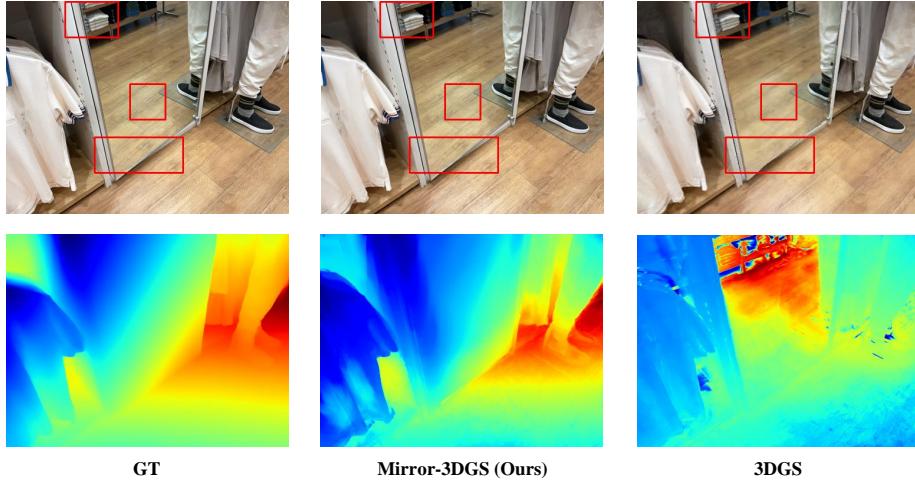


Fig. 1: The synthesized novel views and depth maps. 3DGS confuses mirror reflection, resulting in erroneous mirror depth. Ours properly handles content in the mirror with depth map almost identical to the ground truth (GT).

Radiance Field (NeRF) , some new methods [14, 36, 49] have been proposed. NeRFReN [14] enhances the rendering quality of scenes containing mirrors by employing independent radiance fields to simulate reflective and transmissive parts separately. Mirror-NeRF [49] utilizes Whitted Ray Tracing [41] to achieve physically accurate rendering of mirror reflections, achieving exceptional realism in novel view synthesis. However, the main challenges faced by these NeRF-based methods include long training times and slow rendering speeds, limiting their application in real-time rendering scenarios.

Recently, 3D Gaussian Splatting (3DGS) [18] has emerged as a promising approach for real-time, high-fidelity 3D scene rendering. Instead of the slow volumetric rendering of NeRF, 3DGS explicitly represents scenes using 3D Gaussians and rasterizes 3D Gaussians onto the image plane using splatting techniques, achieving significant improvements in training and rendering speeds. However, 3DGS tends to treat mirror reflections as separate scenes, constructing Gaussian points for them as if they physically exist, leading to inaccurate capture of these virtual images within mirrors during reconstruction and rendering, affecting the realism and visual coherence of the scene, as shown in Fig. 1.

To address these challenges, we propose Mirror-3DGS, a novel rendering framework based on 3DGS, aimed at producing high-fidelity novel views in mirror-containing scenes. *Our Mirror-3DGS is centered on the principle that the virtual image presented on the mirror is equivalent to observing the real world from behind the mirror.* In general, we first identify 3D Gaussians representing the mirror. Then, we observe from both the current viewpoint and its mirrored viewpoint, integrating views from these perspectives to get the final rendered image. To achieve this, we introduce an attribute to signify mirror-like properties to 3D Gaussians, and estimate the plane equation of the mirror based on

the Gaussians with pronounced mirror properties, which allows us to derive the mirror transformation matrix and mirrored camera parameters. Our method involves a two-stage training process: 1) During the first stage, we learn a rough 3D Gaussian representation of the scene without content in the mirror and the mirror plane equation. We replace content within the mirror with a fixed color based on the ground truth mirror mask to better learn mirror attributes and prevent interference from virtual images. 2) In the second stage, we filter Gaussians with high mirror properties for precise estimation of the mirror equation. By integrating views from current and its mirrored viewpoints, we further optimize the rendering quality of the entire scene. To ensure the stability of the rendering process and improve the accuracy of plane equation estimation, we introduce planar consistency constraints and depth constraints through the whole training process. These constraints help achieving more realistic and refined rendering results. Experiments conducted across three synthetic and three real-world scenes demonstrate that our proposed Mirror-3DGS is capable of delivering rendering quality comparable to the state-of-the-art Mirror-NeRF, while significantly reducing training time and achieving real-time rendering. Our contributions can be summarized as follows:

- **Innovative Rendering Framework for Mirror Scenes.** We present Mirror-3DGS, a novel rendering framework that effectively addresses the inherent limitation of 3DGS in handling scenes with mirror, empowering it to achieve precise and realistic synthesis of novel views.
- **Precise Physical Modeling of Mirror Reflection.** We leverage the principle of plane mirror imaging to enhance 3DGS, enabling accurate modeling of physical space and producing high-fidelity rendering results.
- **Extensive Experiments.** We conducted comprehensive experiments across both synthetic and real scenes to validate our proposed method. The results demonstrate our superiority in real-time rendering of high quality novel views, particularly surpassing the state-of-the-arts in mirror regions.

2 Related Work

2.1 Novel View Synthesis

The Neural Radiance Fields (NeRF) [27] significantly advanced the field of novel view synthesis, enhancing the capabilities of multi-view reconstruction. NeRF catalyzed an exponential surge in subsequent research efforts, focusing on pushing the frontiers of scene reconstruction across diverse domains. These endeavors have been geared towards refining reconstruction quality [2–4, 15, 38, 43], enhancing computational efficiency [13, 28, 30, 33, 37], enabling advanced editing functionalities [1, 23, 40, 44, 45], and progressing dynamic scene representation [8, 10, 12, 22, 29]. As a potent successor, 3D Gaussian Splatting (3DGS) [18] reconstructs and represents 3D scenes through the utilization of millions of 3D Gaussians. In contrast to explicit scene representations such as meshes and voxels, 3DGS excels in capturing complex geometries with an elevated level of rendering fidelity. Furthermore, when compared to prior implicit representations

like NeRF, 3DGS presents unmatched advantages in terms of both training and rendering efficiency, thereby showcasing its utility in a variety of applications ranging from SLAM [17, 26] to 4D reconstruction [21, 24, 42, 46] and 3D content generation [34, 47].

2.2 Reflection-aware Rendering

Extensive research initiatives [5–7, 51, 52] have been dedicated to surmounting the complexities of accurately rendering physical reflections within 3D environments. Notably, RefNeRF [36] splits light into diffuse and specular components, employing a reflection-dependent radiance field for modeling reflections based on the view direction. SpecNeRF [25] introduces a learnable Gaussian directional encoding, enhancing the depiction of view-dependent effects under proximate lighting conditions. NeRFReN [14] segregates reflections from transmissions through distinct radiance fields, markedly improving renderings of scenes inclusive of mirrors. Both TraM-NeRF [35] and Mirror-NeRF [49] meticulously simulate mirror geometries and employ path tracing for rendering, delivering exceptional quality. However, the lingering issue of slow rendering speeds significantly hampers the broader applicability of these NeRF-based methods. The method proposed in this paper, grounded in the principles of plane mirror imaging, innovatively applies to modeling mirror scenes within the 3DGS framework. This strategy not only maintains 3DGS’s fast training and real-time rendering benefits but also adeptly navigates the challenges previously outlined, presenting a robust solution to mirror scene representation.

3 Methodology

3.1 Preliminaries

3DGS is an innovative and state-of-the-art approach in the field of novel view synthesis. Distinguished from implicit representation methods such as NeRF [27], which utilize volume rendering, 3DGS leverages the splatting technique [48] to generate images, achieving remarkable real-time rendering speed. Specifically, 3DGS represents the scene through a set of anisotropic Gaussians, defined with its center position $\boldsymbol{\mu} \in \mathbb{R}^3$, covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{3 \times 3}$ which can be decomposed into scaling factor $\mathbf{s} \in \mathbb{R}^3$ and rotation factor $\mathbf{r} \in \mathbb{R}^4$, color defined by spherical harmonic (SH) coefficients $\mathbf{c} \in \mathbb{R}^{3 \times (k+1)^2}$ (where k represents the order of spherical harmonics), and opacity $\alpha \in \mathbb{R}^1$. And the 3D Gaussian can be queried as follows:

$$G(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad (1)$$

where \mathbf{x} represents the position of the query point. Subsequently, an efficient 3D to 2D Gaussian mapping [54] is employed to project the Gaussian onto the image plane:

$$\boldsymbol{\mu}' = \mathbf{P}\mathbf{W}\boldsymbol{\mu}, \quad (2)$$

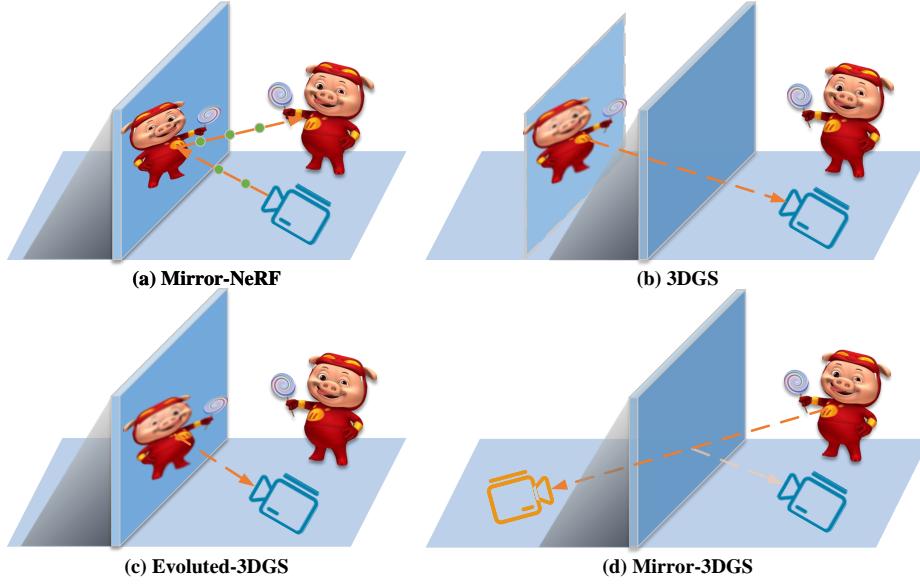


Fig. 2: The underlying principles in handling mirrored contents. **a)** Mirror-NeRF employs a technique involving ray tracing and sampling to achieve mirror reflection. **b)** 3DGS mistakenly considers objects reflected by the mirror to be placed behind it, resulting the floaters behind the mirror. **c)** By employing depth and mirror mask loss on 3DGS, Evolved-3DGS constrains the floaters directly onto the mirror surface. **d)** Our proposed Mirror-3DGS, captures the reflected objects by simulating a novel viewpoint situated behind the mirror.

$$\Sigma' = \mathbf{J} \mathbf{W} \Sigma \mathbf{W}^\top \mathbf{J}^\top, \quad (3)$$

where μ' and Σ' separately represent the 2D mean position and covariance of the projected 3D Gaussian. \mathbf{P} , \mathbf{W} and \mathbf{J} denote the projective transformation, viewing transformation, and Jacobian of the affine approximation of \mathbf{P} , respectively. The color of the pixel on the image plane, denoted by \mathbf{p} , uses a typical neural point-based rendering [19, 20]. Let $\mathbf{C} \in \mathbb{R}^{H \times W \times 3}$ represent the color of rendered image where H and W represents the height and width of images, the rendering process outlined as follows:

$$\mathbf{C}(\mathbf{p}) = \sum_{i=1}^N c_i \sigma_i \prod_{j=1}^{i-1} (1 - \sigma_j), \quad \sigma_i = \alpha_i e^{-\frac{1}{2} (\mathbf{p} - \mu')^\top \Sigma'^{-1} (\mathbf{p} - \mu')}. \quad (4)$$

where N represents the number of sample Gaussians that overlap the pixel \mathbf{p} , c_i and α_i denote the color and opacity of the i -th Gaussian, respectively.

3.2 Mirror-aware 3D Gaussian Representation

In the real world, the color of a point varies when observed from different viewpoints due to its intrinsic material and the environmental lighting. Although

NeRF and 3DGS respectively utilize view-dependent MLPs and higher-order spherical harmonics to simulate this characteristic, the extreme specularity of mirrors amplifies such view-dependent inconsistency. Following, we review existing or potential solutions and introduce our reflection-agnostic Mirror-3DGS.

- **Mirror-NeRF:** Mirror-NeRF achieves mirror modeling by learning a unified probabilistic reflectance field and simulates mirror reflection through physics-based ray tracing, enabling accurate mirror rendering, as illustrated in [Fig. 2\(a\)](#). However, its limitations lie in the intensive computational cost of ray tracing and the prolonged NeRF training time.

- **3DGS:** 3DGS is based on explicit gaussian spheres representation and alpha-blending along ray directions, tending to generate points behind mirrors as shown in [Fig. 2\(b\)](#). This violates physical laws and undermines the rendering quality of mirror back-side scenes.

- **Evolved-3DGS:** We add mirror mask and depth constraints to the original 3DGS to avoid generating points behind mirrors, as depicted in [Fig. 2\(c\)](#). However, the limited mirror Gaussian spheres struggle to represent abundant scene reflection content, restricting the rendering quality of novel views.

- **Mirror-3DGS (Ours):** We innovatively introduce a mirror imaging method – “**moving behind the mirror for a view of the world**”, where the virtual image presented on the mirror is equivalent to observing the real world from behind the mirror, as illustrated in [Fig. 2\(d\)](#). Therefore, we avoid the time-consuming ray tracing required by Mirror-NeRF and also circumvent the insufficient expressiveness of Evolved-3DGS’s mirror Gaussian spheres.

More specifically, **1)** we render an image from the original viewpoint, *i.e.* observing from in front of the mirror, and retain the parts outside the mirror; **2)** we mirror the original viewpoint to behind the mirror, rendering an image from the virtual viewpoint, and only keep the parts of the mirror (see [Sec. 3.3](#)); **3)** finally, we merge these two images to form the final result (see [Sec. 3.3](#)).

To realize the first step above, we need to enable mirror detection. We add a learnable mirror attribute $m \in [0, 1]$ for each Gaussian in the original 3DGS, representing the probability of it being a mirror. The rendering process is similar to color rendering:

$$\mathbf{M}(\mathbf{p}) = \sum_{i=0}^N m_i \sigma_i \prod_{j=1}^{i-1} (1 - \sigma_j), \quad (5)$$

where $\mathbf{M} \in \mathbb{R}^{H \times W \times 3}$ represents the rendered mirror mask, and the rest of the symbols are defined as in [Eq. \(4\)](#). We utilize the GT mirror mask \mathbf{M}_{gt} provided by the dataset to supervise the learning of the mirror attribute:

$$\mathcal{L}_{mask} = \mathcal{L}_1(\mathbf{M}, \mathbf{M}_{gt}). \quad (6)$$

3.3 Virtual Mirrored Viewpoint Construction and Image Fusion

Mirror Parameterization. In the previous section, we acquired the mirror attribute for each Gaussian. Then, based on the mirror attribute and opacity, we filter out the Gaussians belonging to the mirror for constructing a plane in the

3D space. We adopt the parameterization of an infinite plane [16]: $\pi = (\mathbf{n}_\pi^\top, d)$, where $\mathbf{n}_\pi = (a, b, c)^\top$ represents the normal of the plane, and $\frac{d}{\|\mathbf{n}_\pi\|}$ represents the distance from the origin of the plane to the world coordinate origin. Thereby, any point $\mathbf{p} = (x, y, z)^\top$ on the plane satisfies:

$$\mathbf{n}_\pi^\top \mathbf{p} + d = 0. \quad (7)$$

The optimization of the plane π is presented in Sec. 3.4.

Virtual Mirrored Viewpoint Construction. Based on the above plane equation, given an original observation viewpoint $\mathbf{P}_o \in \mathbb{R}^{4 \times 4}$, our goal is to obtain the mirrored viewpoint $\mathbf{P}_m \in \mathbb{R}^{4 \times 4}$ of this viewpoint relative to the plane. This transformation can be formulated as:

$$\mathbf{P}_m = \mathbf{T}_m \mathbf{P}_o, \quad (8)$$

where we define \mathbf{T}_m as the mirror transformation matrix, represented as:

$$\mathbf{T}_m = \begin{bmatrix} 1 - 2a^2 & -2ab & -2ac & -2ad \\ -2ab & 1 - 2b^2 & -2bc & -2bd \\ -2ac & -2bc & 1 - 2c^2 & -2cd \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (9)$$

The derivation of this matrix is detailed in the supplementary material.

Image Fusion. The image from the mirror viewpoint is obtained using the same rendering equation. Thus, we acquire the rendered image $\mathbf{C}_{\mathbf{P}_o} \in \mathbb{R}^{H \times W \times 3}$ from the current original viewpoint, the rendered image $\mathbf{C}_{\mathbf{P}_m} \in \mathbb{R}^{H \times W \times 3}$ from the mirror virtual viewpoint, and the mirror mask \mathbf{M} for the current viewpoint. The final composite image are fused as follows:

$$\mathbf{C}_{fuse} = \mathbf{C}_{\mathbf{P}_o} \odot (\mathbb{1} - \mathbf{M}) + \mathbf{C}_{\mathbf{P}_m} \odot \mathbf{M}, \quad (10)$$

where \odot means element-wise multiplication.

3.4 Two-Stage Training Strategy

We employ a two-stage training strategy, as illustrated in Fig. 3 .

Stage 1: The training objective of the first stage is to obtain a rough Gaussian representation without mirror content and to optimize the mirror equation. We use the GT mirror mask provided by the dataset to erase the mirror content in the GT image \mathbf{C}_{gt} and fill it with red, denoted as \mathbf{C}_{wo}^{mirror} . This is done to optimize the mirror attribute of Gaussians and to avoid generating false-Gaussians inside the mirror. The loss for rendering RGB images is consistent with 3DGS [18], *i.e.*, L1 loss and D-SSIM term:

$$\mathcal{L}_{rgb}^1 = (1 - \gamma) \mathcal{L}_1(\mathbf{C}_{\mathbf{P}_o}, \mathbf{C}_{wo}^{mirror}) + \gamma \mathcal{L}_{D-SSIM}(\mathbf{C}_{\mathbf{P}_o}, \mathbf{C}_{wo}^{mirror}), \quad (11)$$

where γ is the weight, and $\mathbf{C}_{\mathbf{P}_o}$ is the rendered image from the original viewpoint, as defined in Sec. 3.3.

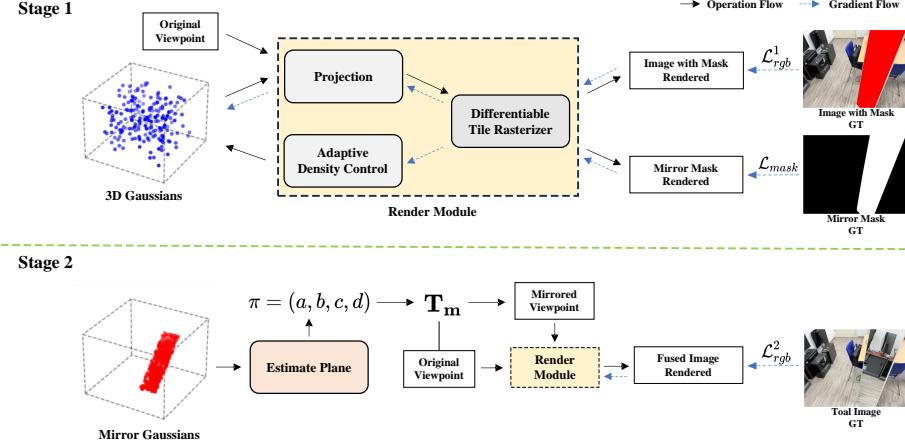


Fig. 3: The illustration of our two-stage training pipeline. In the **Stage 1**, supervision involves the mirror mask and the view image where mirror content is replaced with red color, focusing on learning the mirror plane and a coarse 3D Gaussian representation of the scene. In the **Stage 2**, based on the estimated mirror plane equation, views observed from the original viewpoint and its mirrored viewpoint are fused to form the final rendered result, further optimizing 3D Gaussians of the scene.

Additionally, to improve the quality of rendering and geometric structure, we add depth supervision. First, a pretrained depth estimation model [9] customized for mirror scenes is utilized to extract the depth map \mathbf{D}_{gt} from image \mathbf{C}_{wo}^{mirror} . Then a modified rasterization library [34] is used to render the predicted depth map \mathbf{D}_{pred} , supervised with a similar loss:

$$\mathcal{L}_{depth} = (1 - \gamma)\mathcal{L}_1(\mathbf{D}_{pred}, \mathbf{D}_{gt}) + \gamma\mathcal{L}_{D-SSIM}(\mathbf{D}_{pred}, \mathbf{D}_{gt}). \quad (12)$$

The core of the first stage is to optimize the mirror equation $\pi = (\mathbf{n}_\pi^\top, d)$ to ensure accurate generation of virtual mirrored viewpoints. We filter the mirror attribute by thresholding to obtain Gaussians with high mirror probability, and remove noise Gaussians through the improved RANSAC algorithm [11]. Then we supervise that the remaining Gaussians fall on the mirror plane:

$$\mathcal{L}_{plane} = \frac{1}{N_f} \sum_{i=1}^{N_f} |\mathbf{n}_\pi^\top \mathbf{p}_i + d|, \quad (13)$$

where \mathbf{p}_i is the coordinate of the Gaussian to be optimized, N_f is the number of filtered mirror Gaussians.

Therefore, the overall loss for the first stage is:

$$\mathcal{L}_{stage}^1 = \lambda_{mask}\mathcal{L}_{mask} + \mathcal{L}_{rgb}^1 + \lambda_{depth}\mathcal{L}_{depth} + \mathcal{L}_{plane}, \quad (14)$$

where $\lambda_{mask}, \lambda_{depth}$ are the weight of mask loss and depth loss.

Table 1: Quantitative comparison on the **complete test set**. The first three are NeRF-based methods and Evoluted-3DGS is our baseline described in Sec. 3.2, best results **bolded** and second best underlined. Methods marked with an asterisk ‘*’ are evaluated solely on the mirror regions.

Method	Synthetic				Real			
	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑
InstantNGP [28]	23.54	0.71	0.42	34.89	10.51	0.20	0.71	34.36
DVGO [33]	28.05	0.82	0.29	4.35	22.18	0.67	0.33	4.46
Mirror-NeRF [49]	38.07	0.99	0.01	0.75	25.04	0.86	0.06	0.07
3DGS [18]	37.00	<u>0.97</u>	0.04	341.13	22.55	0.73	0.24	209.62
Evoluted-3DGS	35.52	0.96	0.01	<u>250.73</u>	21.61	0.74	0.11	<u>96.94</u>
Mirror-3DGS (Ours)	<u>37.89</u>	<u>0.97</u>	0.01	171.98	<u>23.05</u>	<u>0.77</u>	<u>0.08</u>	95.97
3DGS* [18]	28.35	-	-	-	21.99	-	-	-
Mirror-NeRF* [49]	29.47	-	-	-	22.39	-	-	-
Mirror-3DGS* (Ours)	32.40	-	-	-	22.51	-	-	-

Stage 2: The objective of the second stage is to refine the rendering quality of the fused images, during which the mirror equation is fixed. As described in Sec. 3.3, we fuse images from the current viewpoint and the virtual mirrored viewpoint, supervising with the complete, unmasked GT image \mathbf{C}_{gt} :

$$\mathcal{L}_{rgb}^2 = (1 - \gamma)\mathcal{L}_1(\mathbf{C}_{fuse}, \mathbf{C}_{gt}) + \gamma\mathcal{L}_{D-SSIM}(\mathbf{C}_{fuse}, \mathbf{C}_{gt}), \quad (15)$$

where \mathbf{C}_{fuse} is the fused rendered image (see Eq. (10)). The overall loss of the second stage is expressed as:

$$\mathcal{L}_{stage}^2 = \lambda_{mask}\mathcal{L}_{mask} + \mathcal{L}_{rgb}^2 + \lambda_{depth}\mathcal{L}_{depth}. \quad (16)$$

4 Experiments

4.1 Experimental Setup

Dataset. We utilized the publicly released dataset by Mirror-NeRF [49], encompassing common scenarios with prevalent mirrors. Specifically, the dataset includes three synthetic scenes (living room, office, and washroom) and three real-world captured scenes (discussion room, lounge, and market). The number of training views per scene ranges from 200 to 300, each accompanied by a mirror mask. Each scene is equipped with a complete test set featuring a mix of views with and without mirrors, alongside a challenging test set exclusively composed of mirror-containing images.

Table 2: Quantitative comparison on the **challenging test set** with all images containing mirrors. The first three are NeRF-based methods and Evolved-3DGS is our baseline described in Sec. 3.2, best results **bolded** and second best underlined. Methods marked with an asterisk '*' are evaluated solely on the mirror regions.

Method	Synthetic				Real			
	PSNR↑	SSIM↑	LPIPS↓	FPS↑	PSNR↑	SSIM↑	LPIPS↓	FPS↑
InstantNGP [28]	14.26	0.44	0.65	9.22	10.54	0.20	0.72	13.00
DVGO [33]	17.65	0.63	0.49	4.34	22.28	0.68	0.32	4.41
Mirror-NeRF [49]	27.33	0.94	0.05	0.55	25.25	0.86	0.06	0.55
3DGS [18]	22.15	0.84	0.17	275.73	22.13	0.72	0.24	208.23
Evolved-3DGS	21.82	0.82	0.09	<u>195.36</u>	21.59	0.75	0.11	<u>98.49</u>
Mirror-3DGS (Ours)	<u>24.73</u>	<u>0.85</u>	<u>0.06</u>	120.61	<u>23.01</u>	<u>0.77</u>	<u>0.09</u>	95.22
3DGS* [18]	16.42	-	-	-	21.27	-	-	-
Mirror-NeRF* [49]	22.14	-	-	-	22.18	-	-	-
Mirror-3DGS* (Ours)	21.60	-	-	-	22.94	-	-	-

Implementation. Our Mirror-3DGS was developed based on the codebase of 3DGS [18]. Training and testing are conducted on $4\times$ downsampled resolution (*i.e.*, 480×360) for all three real-world scenes, and $2\times$ downsampled resolution (*i.e.*, 400×400 for living room and 400×300 for the others) for synthetic scenes. The training process is divided into two stages: in the first stage, we train for 5000 steps, followed by 65000 steps in the second stage. The loss weights λ_{mask} and λ_{depth} are set to 1 and 0.1 in all experiments, respectively.

Evaluation. The rendering quality of all methods is evaluated and compared using the widely recognized PSNR, SSIM [39], and LPIPS [50] metrics. Additionally, we compare the rendering speed using FPS (Frames Per Second). All experiments are conducted on a single RTX 3090 GPU for fair comparison.

4.2 Experimental Results

Quantitative Comparisons. We conduct quantitative comparisons between our method and several competitive counterparts, including InstantNGP [28], DVGO [33], the state-of-the-art Mirror-NeRF [49], vanilla 3DGS [18] and our baseline Evolved-3DGS. The comparison results across the complete and the challenging test sets are detailed in Tab. 1 and Tab. 2, respectively.

Among these competitors, Mirror-NeRF consistently achieves the highest PSNR, SSIM, and LPIPS scores on both synthetic and real scenes. However, its rendering speed, *i.e.*, FPS, is significantly lower due to the inherent inefficiencies of NeRF’s volumetric rendering mechanism. Even the fastest NeRF-based

Our code will be made publicly available for reproducible research.

method among these, InstantNGP, is orders of magnitude slower compared to 3DGS and our approach in real scenes. As such, these NeRF-based methods are challenging to apply in scenarios requiring real-time rendering.

Our method yields performance slightly better than 3DGS within the complete test set, which includes a minimal number of views with mirrors, yet it significantly surpasses vanilla 3DGS in the challenging test set, where mirror views are predominant. *Specifically, our Mirror-3DGS demonstrates an average PSNR improvement of 1.58 and 0.8 dB over vanilla 3DGS on the challenging test set of synthetic and real scenes, respectively. Remarkably, our method even surpasses Mirror-NeRF in washroom scene, achieving a PSNR gain of 4.88 dB.* Additionally, our SSIM and LPIPS metrics far exceed those of vanilla 3DGS, indicating that our synthesized views are significantly more realistic. These improvements are mainly attributed to our method’s specialized handling of mirrors and their reflections in the scene, which inevitably introduces some additional computation overhead. Therefore, while our rendering speed is slightly lower than that of vanilla 3DGS, it is significantly faster than NeRF-based methods, making it suitable for real-time rendering applications.

It is important to highlight that *our Mirror-3DGS outperforms both Mirror-NeRF and vanilla 3DGS in terms of the PSNR metric when focusing solely on the mirror regions*, as shown in the last three rows of Tab. 1 and Tab. 2, demonstrating our method’s enhanced capability in handling mirror reflections.

Qualitative Comparisons. The qualitative analysis presented in Fig. 4, focusing on the detailed rendering of mirrors, contrasts our Mirror-3DGS method with three competitive alternatives across six diverse scenes.

The grid-based DVGO method often results in blurry images of mirror content across all scenes, due to limitations in grid resolution and inaccuracies in physical modeling that cause distortions and artifacts on mirror surfaces. The 3DGS method also struggles with accurately rendering mirrors. Its main problem is the incorrect placement of Gaussians for content within the mirrors, positioned behind the mirror, which distorts the scene’s depth information and leads to inaccurate renderings. Mirror-NeRF, utilizing ray tracing to render content within mirrors, more accurately captures physical imaging principles, achieving the highest overall rendering quality. However, its method for implicitly modeling the probabilities of mirror reflections and normals can sometimes lead to inaccuracies in depicting the reflective content of mirrors. Furthermore, integrating ray tracing with the 3DGS is challenging.

Our Mirror-3DGS, by incorporating the principles of plane mirror imaging and accurately modeling spatial geometric relationships, enables high-fidelity rendering of mirror contents. Across all six scenes, our Mirror-3DGS achieves rendering quality comparable to Mirror-NeRF, demonstrating its effectiveness in handling complex mirror reflection.

Training Time and Convergence. A significant advantage of our method is the real-time rendering capability derived from 3DGS. In parallel, our method also boasts a substantial advantage in training efficiency compared to NeRF-based counterparts, particularly against the current state-of-the-art method,

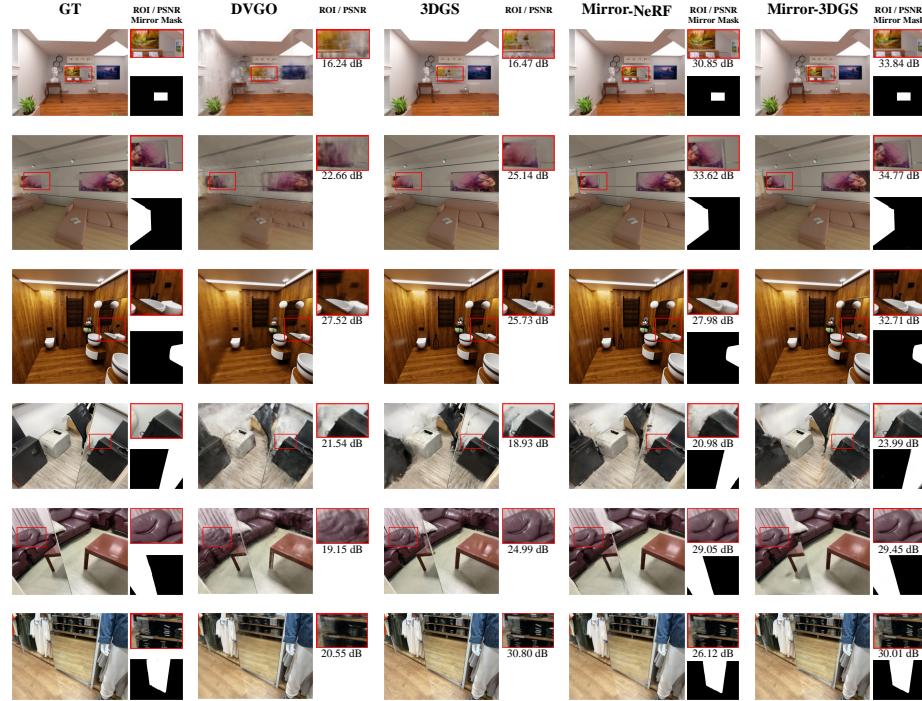


Fig. 4: Qualitative comparison of novel view synthesis. Rows 1-6 respectively showcase rendering results of living room, office, washroom, discussion room, lounge and market scenes, with magnified displays of mirror regions in the top-right and predicted mirror mask in the bottom-right of the rendered views. The PSNR is computed solely on the mirror region. Mirror-NeRF and Mirror-3DGS can accurately predict the mirror masks and our Mirror-3DGS achieves the best performance on mirror regions.

Mirror-NeRF. Fig. 5 and Fig. 6 showcase the time-evolving rendering effects on the washroom scene. It is evident that our method requires only 1 minute of training to achieve satisfactory rendering of the contents within the mirror, whereas Mirror-NeRF starts to render the mirror’s contents only after 5 hours of training. Ultimately, our method converges to its best performance after 15 minutes of training, while Mirror-NeRF takes 8 hours to converge. Beyond the advantage in convergence speed, our method closely matches the visual quality of Mirror-NeRF and even surpasses it in some scenes.

4.3 Ablation Study

Our Mirror-3DGS method incorporates several key designs: 1) \mathcal{L}_{mask} , 2) \mathcal{L}_{depth} , and 3) \mathcal{L}_{plane} . Through ablation results displayed in Tab. 3 and Tab. 4, we confirm their significance to our approach.

Firstly, the design of \mathcal{L}_{mask} is crucial for effectively learning mirror properties, enabling precise identification of mirror regions during rendering. Removing

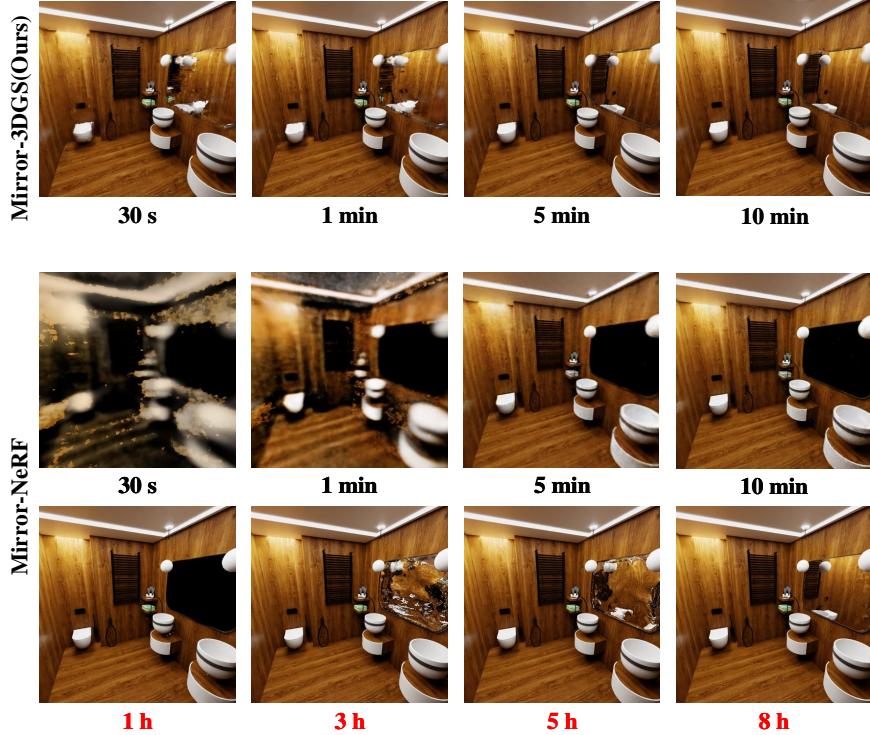


Fig. 5: Evolution of rendering results over training time on the washroom scene. Our Mirror-3DGS rapidly learns to accurately render contents in mirrors, while Mirror-NeRF takes more than five hours to reach similar levels of detail.

\mathcal{L}_{mask} leads to the significant performance drop, attributed to incorrect estimation of mirror areas, which not only affects the rendering of content within mirrors but also results in the erroneous processing of non-mirror regions as if they were mirrors, thereby producing incorrect rendering results.

Secondly, \mathcal{L}_{depth} ensures consistency between the rendered depth in the learning process and the depth extracted by the depth model. Given that the pre-trained depth model is relatively accurate in handling the depth of mirror regions, it effectively prevents the incorrect generation of Gaussians for content within mirrors as if they were physically present. The absence of this supervision noticeably reduces our method’s performance.

Finally, \mathcal{L}_{plane} aids in learning more accurate mirror planes. If the learned mirror plane is rough, applying the transformation in Eq. (8) leads to incorrect imaging results in the mirror. We observed that this loss has a minor impact on performance on real scenes, while its influence is more pronounced in synthetic scenes.

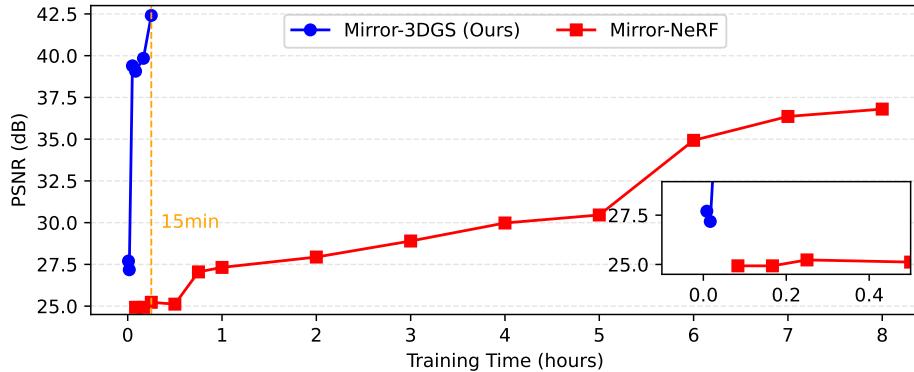


Fig. 6: Convergence speed comparison on the washroom scene. Our method achieves optimal rendering results in approximately 15 minutes, significantly outpacing Mirror-NeRF, which requires over 8 hours to reach a comparable level of performance.

Table 3: Ablation results on the challenging test set of real market scene. The values in red parentheses indicate metric changes relative to our full method.

	PSNR↑	SSIM↑	LPIPS↓
w/o \mathcal{L}_{mask}	25.68 ($\downarrow 1.50$)	0.83 ($\downarrow 0.01$)	0.05 ($\uparrow 0.01$)
w/o \mathcal{L}_{depth}	25.39 ($\downarrow 1.79$)	0.82 ($\downarrow 0.02$)	0.05 ($\uparrow 0.01$)
w/o \mathcal{L}_{plane}	26.84 ($\downarrow 0.34$)	0.84	0.04
Our full method	27.18	0.84	0.04

Table 4: Ablation results on the challenging test set of synthetic washroom scene. The values in red parentheses indicate metric changes relative to our full method.

	PSNR↑	SSIM↑	LPIPS↓
w/o \mathcal{L}_{mask}	17.27 ($\downarrow 3.03$)	0.69 ($\downarrow 0.06$)	0.13 ($\uparrow 0.04$)
w/o \mathcal{L}_{depth}	18.61 ($\downarrow 1.69$)	0.72 ($\downarrow 0.03$)	0.11 ($\uparrow 0.02$)
w/o \mathcal{L}_{plane}	16.29 ($\downarrow 4.01$)	0.67 ($\downarrow 0.08$)	0.15 ($\uparrow 0.06$)
Our full method	20.30	0.75	0.09

5 Conclusion

In this paper, we introduced Mirror-3DGS, a novel rendering framework that addresses the challenge of accurately capturing and rendering physical reflections in mirror-containing scenes. Our Mirror-3DGS employs a two-stage training process that first filters mirror Gaussians and then uses this information to accurately estimate the mirror plane and derive mirrored camera parameters. The effectiveness of Mirror-3DGS in delivering high-quality renderings of scenes with mirrors is demonstrated by experiments on three synthetic and three real scenes, closely matching the state-of-the-art Mirror-NeRF method, while achieving substantial

improvements in training and rendering speed. However, our Mirror-3DGS still faces limitations, notably in the accuracy of mirror estimations and its dependence on dataset-provided mirror masks. Currently, our method is tailored for single flat mirrors, leaving the adaptation to complex shapes and multiple mirrors as an objective for future exploration.

References

1. Bao, C., Zhang, Y., Yang, B., Fan, T., Yang, Z., Bao, H., Zhang, G., Cui, Z.: Sine: Semantic-driven image-based nerf editing with prior-guided editing field. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
2. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
3. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
4. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Zip-nerf: Anti-aliased grid-based neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
5. Bi, S., Xu, Z., Srinivasan, P., Mildenhall, B., Sunkavalli, K., Hašan, M., Hold-Geoffroy, Y., Kriegman, D., Ramamoorthi, R.: Neural reflectance fields for appearance acquisition. arXiv preprint arXiv:2008.03824 (2020)
6. Boss, M., Engelhardt, A., Kar, A., Li, Y., Sun, D., Barron, J., Lensch, H., Jampani, V.: Samurai: Shape and material from unconstrained real-world arbitrary image collections. Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2022)
7. Boss, M., Jampani, V., Braun, R., Liu, C., Barron, J., Lensch, H.: Neural-pil: Neural pre-integrated lighting for reflectance decomposition. Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2021)
8. Cao, A., Johnson, J.: Hexplane: A fast representation for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
9. Costanzino, A., Ramirez, P.Z., Poggi, M., Tosi, F., Mattoccia, S., Di Stefano, L.: Learning depth estimation for transparent and mirror surfaces. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
10. Fang, J., Yi, T., Wang, X., Xie, L., Zhang, X., Liu, W., Nießner, M., Tian, Q.: Fast dynamic radiance fields with time-aware neural voxels. In: SIGGRAPH Asia (2022)
11. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM (1981)
12. Fridovich-Keil, S., Meanti, G., Warburg, F.R., Recht, B., Kanazawa, A.: K-planes: Explicit radiance fields in space, time, and appearance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)

13. Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxtels: Radiance fields without neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
14. Guo, Y.C., Kang, D., Bao, L., He, Y., Zhang, S.H.: Nerfren: Neural radiance fields with reflections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
15. Huang, X., Li, W., Hu, J., Chen, H., Wang, Y.: Refsr-nerf: Towards high fidelity and super resolution view synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
16. Kaess, M.: Simultaneous localization and mapping with infinite planes. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) (2015)
17. Keetha, N., Karhade, J., Jatavallabhula, K.M., Yang, G., Scherer, S., Ramanan, D., Luiten, J.: Splatam: Splat, track & map 3d gaussians for dense rgb-d slam. arXiv preprint arXiv:2312.02126 (2023)
18. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (ToG) (2023)
19. Kopanas, G., Leimkühler, T., Rainer, G., Jambon, C., Drettakis, G.: Neural point catacaustics for novel-view synthesis of reflections. ACM Transactions on Graphics (TOG) (2022)
20. Kopanas, G., Philip, J., Leimkühler, T., Drettakis, G.: Point-based neural rendering with per-view optimization. Proceedings of the Eurographics Symposium on Rendering (EGSR) (2021)
21. Li, Z., Chen, Z., Li, Z., Xu, Y.: Spacetime gaussian feature splatting for real-time dynamic view synthesis. arXiv preprint arXiv:2312.16812 (2023)
22. Liu, J.W., Cao, Y.P., Mao, W., Zhang, W., Zhang, D.J., Keppo, J., Shan, Y., Qie, X., Shou, M.Z.: Devrf: Fast deformable voxel radiance fields for dynamic scenes. Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2022)
23. Liu, S., Zhang, X., Zhang, Z., Zhang, R., Zhu, J.Y., Russell, B.: Editing conditional radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
24. Luiten, J., Kopanas, G., Leibe, B., Ramanan, D.: Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. arXiv preprint arXiv:2308.09713 (2023)
25. Ma, L., Agrawal, V., Turki, H., Kim, C., Gao, C., Sander, P., Zollhöfer, M., Richardt, C.: Specnerf: Gaussian directional encoding for specular reflections. arXiv preprint arXiv:2312.13102 (2023)
26. Matsuki, H., Murai, R., Kelly, P.H., Davison, A.J.: Gaussian splatting slam. arXiv preprint arXiv:2312.06741 (2023)
27. 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: Proceedings of the European Conference on Computer Vision (ECCV) (2020)
28. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (ToG) (2022)
29. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-nerf: Neural radiance fields for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
30. Reiser, C., Szeliski, R., Verbin, D., Srinivasan, P., Mildenhall, B., Geiger, A., Barron, J., Hedman, P.: Merf: Memory-efficient radiance fields for real-time view synthesis in unbounded scenes. ACM Transactions on Graphics (TOG) (2023)

31. Shi, Y., Wu, Y., Wu, C., Liu, X., Zhao, C., Feng, H., Liu, J., Zhang, L., Zhang, J., Zhou, B., et al.: Gir: 3d gaussian inverse rendering for relightable scene factorization. arXiv preprint arXiv:2312.05133 (2023)
32. Srinivasan, P.P., Deng, B., Zhang, X., Tancik, M., Mildenhall, B., Barron, J.T.: Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
33. Sun, C., Sun, M., Chen, H.T.: Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
34. Tang, J., Ren, J., Zhou, H., Liu, Z., Zeng, G.: Dreamgaussian: Generative gaussian splatting for efficient 3d content creation. arXiv preprint arXiv:2309.16653 (2023)
35. Van Holland, L., Bliersbach, R., Müller, J.U., Stotko, P., Klein, R.: Tram-nerf: Tracing mirror and near-perfect specular reflections through neural radiance fields. arXiv preprint arXiv:2310.10650 (2023)
36. Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P.: Ref-nerf: Structured view-dependent appearance for neural radiance fields. in 2022 ieee. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
37. Wadhwan, K., Kojima, T.: Squeezenerf: Further factorized fastnerf for memory-efficient inference. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
38. Wang, C., Wu, X., Guo, Y.C., Zhang, S.H., Tai, Y.W., Hu, S.M.: Nerf-sr: High quality neural radiance fields using supersampling. In: Proceedings of the ACM International Conference on Multimedia (ACM MM) (2022)
39. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing (TIP) (2004)
40. Weder, S., Garcia-Hernando, G., Monszpart, A., Pollefeys, M., Brostow, G.J., Firman, M., Vicente, S.: Removing objects from neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
41. Whitted, T.: An improved illumination model for shaded display. In: ACM Siggraph 2005 Courses (2005)
42. Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X.: 4d gaussian splatting for real-time dynamic scene rendering. arXiv preprint arXiv:2310.08528 (2023)
43. Xu, Q., Xu, Z., Philip, J., Bi, S., Shu, Z., Sunkavalli, K., Neumann, U.: Point-nerf: Point-based neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
44. Xu, T., Harada, T.: Deforming radiance fields with cages. In: Proceedings of the European Conference on Computer Vision (ECCV) (2022)
45. Yang, S., Mou, C., Yu, J., Wang, Y., Meng, X., Zhang, J.: Neural video fields editing. arXiv preprint arXiv:2312.08882 (2023)
46. Yang, Z., Gao, X., Zhou, W., Jiao, S., Zhang, Y., Jin, X.: Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
47. Yi, T., Fang, J., Wang, J., Wu, G., Xie, L., Zhang, X., Liu, W., Tian, Q., Wang, X.: Gaussiandreamer: Fast generation from text to 3d gaussians by bridging 2d and

- 3d diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
- 48. Yifan, W., Serena, F., Wu, S., Öztireli, C., Sorkine-Hornung, O.: Differentiable surface splatting for point-based geometry processing. ACM Transactions on Graphics (TOG) (2019)
 - 49. Zeng, J., Bao, C., Chen, R., Dong, Z., Zhang, G., Bao, H., Cui, Z.: Mirror-nerf: Learning neural radiance fields for mirrors with whitted-style ray tracing. In: Proceedings of the ACM International Conference on Multimedia (ACM MM) (2023)
 - 50. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
 - 51. Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T.: Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics (ToG) (2021)
 - 52. Zhang, Y., Xu, T., Yu, J., Ye, Y., Jing, Y., Wang, J., Yu, J., Yang, W.: Nemf: Inverse volume rendering with neural microflake field. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
 - 53. Zhang, Y., Sun, J., He, X., Fu, H., Jia, R., Zhou, X.: Modeling indirect illumination for inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
 - 54. Zwicker, M., Pfister, H., Van Baar, J., Gross, M.: Ewa volume splatting. In: IEEE Conference on Visualization (VIS) (2001)

Supplementary Material

This supplementary material provides additional details of our proposed method, including formula derivation in Sec. A, and method details in Sec. B.

A Mirror Transformation Matrix

The relationship between the viewpoints \mathbf{P}_o and \mathbf{P}_m is derived here, corresponding to main paper Eq. (8). As described in main paper Eq. (7), an infinite plane can be parameterized as:

$$\mathbf{n}_\pi^\top \mathbf{p} + d = 0. \quad (17)$$

where the unit vector $\mathbf{n}_\pi = (a, b, c)^\top$ represents the normal of the plane and $\mathbf{p} = (X, Y, Z)^\top$ is a point upon the plane. The Eq. (17) can also be written as:

$$aX + bY + cZ + d = 0. \quad (18)$$

Let $(x, y, z)^\top$ be a point outside the mirror plane, then the projection of this point onto the plane is given by:

$$\begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \lambda \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad (19)$$

where $\lambda = ax + by + cz + d$.

According to the principle of mirror reflection, the coordinates of the mirror-symmetrical point $(x_r, y_r, z_r)^\top$ can be calculated as:

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} - \lambda \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} - 2\lambda \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad (20)$$

The corresponding homogeneous coordinate representation is:

$$\begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} - 2\lambda \begin{bmatrix} a \\ b \\ c \\ 0 \end{bmatrix} = \begin{bmatrix} 1 - a^2 & -2ab & -2ac & -2ad \\ -2ab & 1 - 2b^2 & -2bc & -2bd \\ -2ac & -2bc & 1 - 2c^2 & -2cd \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (21)$$

We define \mathbf{T}_m as:

$$\mathbf{T}_m = \begin{bmatrix} 1 - a^2 & -2ab & -2ac & -2ad \\ -2ab & 1 - 2b^2 & -2bc & -2bd \\ -2ac & -2bc & 1 - 2c^2 & -2cd \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (22)$$

corresponding to main paper Eq. (9).

Similarly, the relation between $(x, y, z)^\top$ and $(x_r, y_r, z_r)^\top$ can be written as:

$$\mathbf{Q}_o \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \mathbf{Q}_m \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}, \quad (23)$$

where \mathbf{Q}_o and \mathbf{Q}_m are original and mirrored viewpoint respectively, representing the transformation from world coordinates to camera coordinates. Then,

$$\mathbf{Q}_o \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \mathbf{Q}_m \mathbf{T}_m \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \quad (24)$$

Since $(x, y, z)^\top$ represents any arbitrary point, we can deduce that:

$$\mathbf{Q}_o = \mathbf{Q}_m \mathbf{T}_m. \quad (25)$$

Finally, the relation between their inverse matrices \mathbf{P}_o and \mathbf{P}_m , as illustrated in Eq. (8) of our main paper, can be written as:

$$\mathbf{P}_m = \mathbf{T}_m \mathbf{P}_o. \quad (26)$$

B Method Details

B.1 The detail of rendering

In the Section 3.3 of the main paper, we elaborate on the virtual mirrored viewpoint construction and image fusion. Here, we represent the detailed algorithmic procedure as pseudocode, which is displayed in Algorithm 1.

Algorithm 1 Reflective Rendering

Input: original viewpoint $\mathbf{P}_o \in \mathbb{R}^{4 \times 4}$, gaussians $\mathbf{G} \in \mathbb{R}^{N \times C}$, threshold τ .
Output: the fused view image $\mathbf{C}_{fuse} \in \mathbb{R}^{H \times W \times 3}$.

```
# C_P_o, D_pred, M are rendered image, depth map, and mirror mask, respectively
C_P_o, D_pred, M ← Render(P_o, G)
G_mirror ← MirrorFilter(G, τ) # Gaussian points represent mirror plane
a, b, c, d ← RANSAC(G_mirror) # mirror plane parameters described in Eq. (18)
C_P_m ← Render(P_m, G) # Render image of mirrored viewpoint
C_fuse = C_P_o ⊙ (1 - M) + C_P_m ⊙ M # Image fusion, Eq. (10) in the main paper

return C_fuse
```

B.2 Mirror Plane Estimation

To learn the accurate parameters of the mirror plane, we introduce \mathcal{L}_{plane} to constrain the Gaussians constituting the mirror surface.

In main text, we employ the RANSAC [11] method for plane selection and use the average distance from points to the mirror surface as \mathcal{L}_{plane} . However, for scenes where the mirror is distant from the camera center, significant distortion occurs, which RANSAC cannot process well.

In order to handle the distortion, we develop another way to choose our plane. By selecting the median normal vector and median coordinates of the mirror points as the normal of and a point upon the plane, we can determine the coordinates of the mirror. For above reasons, we need to constrain the normal vectors.

According to [31], we consider the short-axis direction of the Gaussian as its normal vector. Constraints to the normal vectors are appended to ensure them to be parallel. Specifically, we randomly select three mirror Gaussians and require that their normal vectors, denoted as $\mathbf{n}_a, \mathbf{n}_b, \mathbf{n}_c$, are parallel to each other:

$$\mathcal{L}_{plane} = \frac{1}{N_p} \sum_{i=1}^{N_p} |\mathbf{n}_a \times \mathbf{n}_b \cdot \mathbf{n}_c|, \quad (27)$$

where N_p denotes the number of the 3-points sets randomly selected from mirror plane, \times represents the cross product, and \cdot represents the dot product.

The rest of the approach remains the same as mentioned in the main text. We recommend using the RANSAC to determine the plane in scene discussion room, lounge and market, while using the approach mentioned above to determine the rest scene.

B.3 Activation Function

We use the **SINE** activation function instead of the activation function **Sigmoid** in 3DGS. The expression of SINE is as follows:

$$SINE(x) = \frac{\sin(x) + 1}{2}. \quad (28)$$

We apply SINE and Sigmoid activation function to normalize the opacity and mirror properties respectively, both of which are bounded within the range of $[0, 1]$, while keeping all other experimental variables constant.

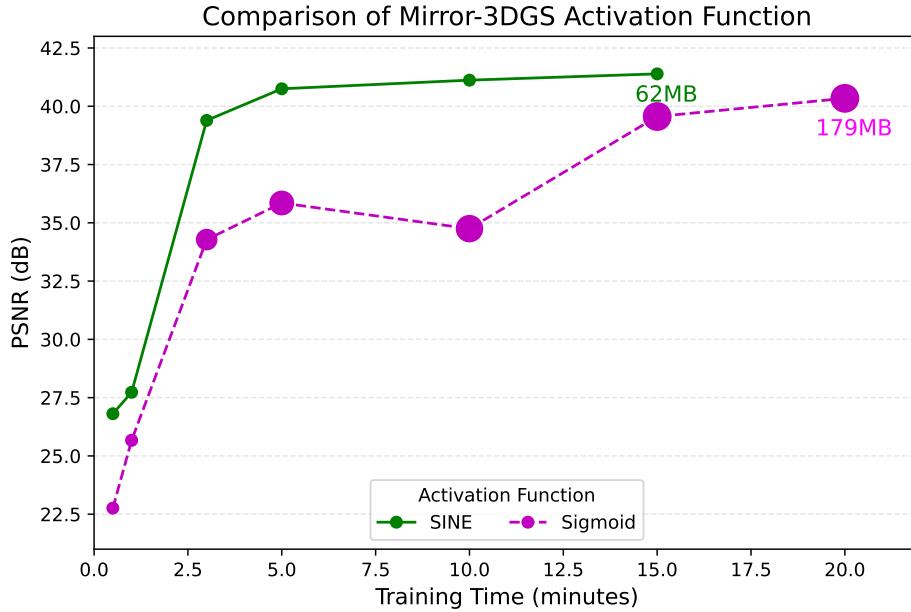


Fig. 7: Illustration of Convergence Speed and Storage Size for the Washroom Scene. The radius of points on the line indicates the storage size of the scene. The SINE activation function accelerates convergence and significantly reduces the number of Gaussians, *i.e.*, lowering storage cost.

As depicted in Fig. 7, the empirical findings indicate that SINE activation function significantly enhances the model's convergence rate and diminishes its complexity. This improvement is attributed to the SINE function's circumvention of intricate exponential computations. Moreover, in contrast to Sigmoid function, SINE function mitigates the vanishing gradient issue and moderates the localization gradient, thereby curtailing the superfluous densification process.