ReInGS: Re-Initializing 3D Gaussians against Sparsity Discrepancy in Few-Shot Novel View Synthesis

Junao Shen[1], Tian Feng[2,3⋆], Haojie Dong[4], Jinkang Ji[4], and Tianjia Shao[3]

[1] Hangzhou VIVO Information Technology Co., Ltd., Hangzhou, China
[2] Hangzhou Research Institute of AI and Holographic Technology, Hangzhou, China
[3] State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, China
[4] School of Software Technology, Zhejiang University, Ningbo, China
{jashen, t.feng, donghaojie, jkji, tjshao}@zju.edu.cn

Abstract. Recent advancements in 3D Gaussian Splatting (3DGS) have achieved significant progress in novel view synthesis (NVS), offering computationally efficient and photorealistic rendering, even in the few-shot setting where input training views are sparse. However, several representative 3DGS-based methods for few-shot NVS face a critical issue of sparsity discrepancy, *i.e.*, Gaussian initialization leverages dense fused stereo points from extended views, creating an inconsistency with sparse training views. To resolve this issue, we propose ReInGS, a novel method featuring a two-stage coarse-to-fine 3D Gaussian re-initialization process designed explicitly for few-shot NVS. In the coarse-level initialization stage, an expansion-based hybrid point sampling (EHPS) strategy is deployed to generate dense points and effectively capture the scene's global geometric structure. In the fine-level re-initialization stage, a view-dominant details-aware point sampling (VDPS) strategy, which comprises two sequential sub-strategies on intra-view point sampling and cross-view point sampling, is adopted to reconstruct the scene's local geometric details with high fidelity. Comprehensive experiments on two benchmarking datasets demonstrate that ReInGS achieves state-of-the-art performance across 2-view, 3-view, and 4-view settings, especially delivering PSNR gains of up to 2.92 dB, 1.92 dB, and 1.38 dB, respectively, while synthesizing high-quality novel views without data leakage from extended views.

Keywords: 3D Gaussian Splatting · Novel View Synthesis · Few-Shot

1 Introduction

Few-shot novel view synthesis (NVS) aims to generate images of the target scene from unseen viewpoints given a limited amount of images from known viewpoints, and demonstrates significantly practical values in high-quality rendering

⋆ Corresponding author.

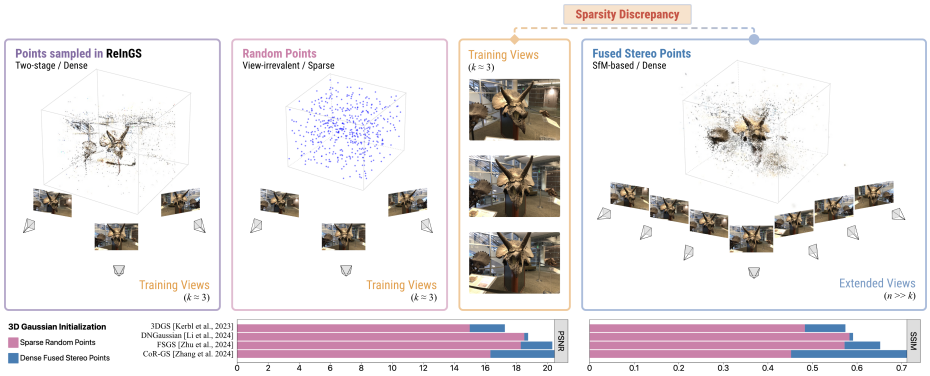

Fig. 1: Recent few-shot NVS methods [21, 36, 38] based on 3DGS [14] leverage dense fused stereo points from extended views rather than sparse training views for 3D Gaussian initialization, causing an issue termed **sparsity discrepancy** in this paper. When initializing Gaussians using random points to resolve sparsity discrepancy, these methods would face significant drops in their performance. To address the issue, we propose ReInGS characterized by a two-stage coarse-to-fine process to re-initialize Gaussians for efficient and high-quality few-shot NVS.

upon data sparsity [38]. To accomplish the task of few-shot NVS, conventional methods based on Neural Radiance Fields (NeRF) [19] generally focus on exploiting prior information [3,11,16,20,26,27,32,35], requiring substantial memory consumption and leading to less satisfactory efficiency. More recent methods based on 3D Gaussian Splatting (3DGS) [14] have achieved considerable success in efficiently rendering the scene's geometric structure and details [8, 9, 17, 21, 36, 38].

According to the vanilla 3DGS [14], however, the initialization of 3D Gaussians relies on fused stereo points, which are usually estimated by a Structure-from-Motion (SfM) method [23] from the input images as *training views*. In the few-shot setting, the quality of such points could drop significantly due to the sparsity of training views and negatively impact on the output images. To meet this challenge, DNGaussian [17] adopts randomly sampled points that are sparse and view-irrelevant to replace the unstable fused stereo points from few-shot views. CoherentGS [21] assigns a single Gaussian to each pixel per view by estimating the depth maps; Nevertheless, it moves Gaussians merely along the rays during the optimization, causing hollows in the unseen views. In contrast, FSGS [38] leverages the dense fused stereo points generated from an extended set of views rather than training views, *i.e.*, all available views in the dataset, and proximity-guided Gaussian unpooling for density; This Gaussian initialization strategy using extended views has also been adopted by several subsequent studies [34, 36]. As shown in Figure 1, such methods have indeed achieved considerably better performance with Gaussians initialized using dense fused stereo points compared to sparse random points. Despite this promise, we argue that using information from data in addition to sparse training views for Gaussian initialization actually causes data leakage and violates the few-shot nature of the task, and identify this issue as **sparsity discrepancy**.

Identifying sparsity discrepancy has motivated us to answer a question: *why the fused stereo points from extended views benefit 3D Gaussian initialization more than sparse random points?* Our response is two-fold: (1) geometric information with higher accuracy, which originally serves scene reconstruction, is available; and (2) fused stereo points are of a notably higher density. As shown in Table 1, using dense random points to initialize Gaussians could lead to better performance for representative methods [17,38] compared to sparse random points. Consequently, it is worth investigating a feasible 3DGS-based solution to few-shot NVS against sparsity discrepancy , and such a solution is expected to exploit dense random points and the information provided only by training views for initialization.

Table 1: Results of a pilot study on the influence of the number of random points for Gaussian initialization in DNGaussian [17] and FSGS [38] on LLFF [18] under 3-view setting. Underline denotes number setting as in DNGaussian.

Method	# of Random Points	PSNR↑	LPIPS↓	SSIM↑
DNGaussian [17]	2000	18.65	0.318	0.578
	5000	18.87	0.313	0.583
	8000	18.94	0.311	0.592
FSGS [38]	2000	18.15	0.321	0.566
	5000	18.28	0.312	0.572
	8000	18.33	0.310	0.577

In this paper, we propose a novel method characterized by a two-stage process to **Re-In**itialize 3D Gaussians for efficient and high-quality few-shot NVS based on 3D**GS** (**ReInGS**). Inspired by the Diffusion model [10] accomplishing generation via progressive denoising, our Gaussian re-initialization process comprises a *coarse-level initialization* stage and a *fine-level re-initialization* stage, each of which involves a Gaussian pre-optimization. In the coarse-level initialization stage, the proposed method deploys an *expansion-based hybrid point sampling* (EHPS) strategy. This strategy generates dense points, both uniformly and randomly, covering the scene's space after expansion and provides positions for Gaussians to be pre-optimized toward the scene's global geometric structure. In the fine-level re-initialization stage, our ReInGS adopts a *view-dominant details-aware point sampling* (VDPS) strategy that exploits the pre-optimized Gaussians from the previous stage to produce points benefiting the reconstruction of the scene's local geometric details, following the idea of image-order rendering and object-order rendering. Specifically, the VDPS strategy comprises sequential *intra-view point sampling* (IVPS) and *cross-view point sampling* (CVPS) sub-strategies to provide positions and colors for Gaussians. The two sub-strategies are connected by a Gaussian pre-optimization. Afterwards, the re-initialized 3D Gaussians are finally optimized regarding appearance and geometry to synthesize images from given viewpoints.

Experimental results on two benchmark datasets demonstrate that our ReInGS achieves the state-of-the-art performance in challenging 2-view, 3-view and 4-view settings. The contributions of this paper can be summarized as follows:

- Identifying for the first time the issue of sparsity discrepancy in representative 3DGS-based NVS methods that violates the few-shot setting;
- A novel and efficient 3DGS-based few-shot NVS method that resolves sparsity discrepancy via re-initializing 3D Gaussians with respect to the scene's

geometry in a coarse-to-fine manner, outperforms representative methods on different datasets in the case without sparsity discrepancy;
– Effective strategies on expansion-based hybrid point sampling and view-dominated details-aware point sampling for 3D Gaussian initialization that focus on the target scene's geometric information.

2 Related Work

3D Gaussian Splatting. Recent advanced unstructured radiance fields [4, 14, 31] employed explicit primitives for scene representation, offering compelling alternatives to traditional volumetric solutions. Among these methods, 3D Gaussian Splatting (3DGS) [14] represents the target scene through a set of 3D Gaussians, replacing the dense neural networks used in NeRF [19] with an efficient and differentiable splatting renderer. This explicit representation enables 3DGS to achieve superior performance in both quality and speed in real-time applications, including SLAM [22], generative 3D content creation [25], and dynamic scene reconstruction [29]. The success of 3DGS stems from its interpretable scene representation and efficient rendering pipeline, which enables direct optimization of Gaussian parameters through differentiable rasterization. However, these advantages primarily manifest in scenarios with dense training views, as the method relies on comprehensive multi-view information to accurately position and optimize its Gaussian primitives. Our work addresses the limitation by investigating 3DGS in challenging few-shot NVS.

Few-shot Novel View Synthesis. Characterized by data sparsity, few-shot NVS remains a significant challenge, as traditional methods often suffer from severe overfitting and degraded reconstruction quality [27, 32]. Recent studies have attempted to meet this challenge, usually based on NeRF and 3DGS. NeRF-based methods involve various regularization strategies to improve reconstruction from sparse views, including geometric and appearance regularization from unseen views [20], depth supervision from sparse 3D points [5], and depth distillation-based regularization [27]. Pre-trained models trained on large-scale datasets are also exploited by such methods to enhance generalization [3, 16, 35] or guide the training objective [11, 30]. However, NeRF-based methods are computationally intensive during training and inference, and the implicit representations make direct integration with discrete scene representations challenging.

More recently, a series of methods have extended 3DGS to address few-shot NVS. Specifically, FSGS [38] and DNGaussian [17] incorporate geometric constraints through depth regularization; CoherentGS [21] employs optical flow-based regularization to enforce consistency among Gaussians representing the same 3D point; CoR-GS [36] introduces joint regularization across multiple Gaussian sets to eliminate mispositioned primitives. While some of these methods have attempted to generate a fine initialization. The promising results of 3D-based methods tend to rely on dense fused stereo points constructed from extended views, which limits their applicability in real few-shot scenarios. Our

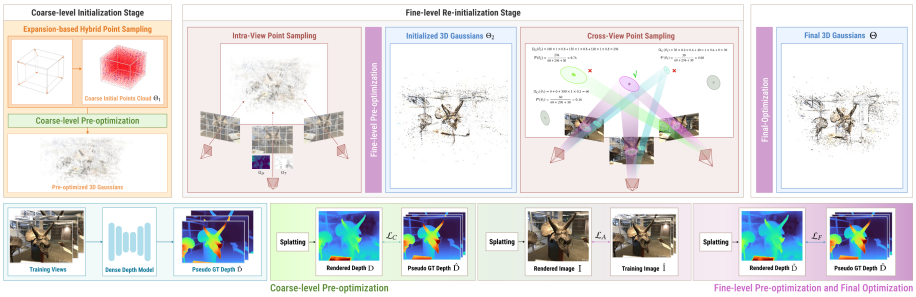

Fig. 2: The pipeline of ReInGS features a two-stage coarse-to-fine 3D Gaussian re-initialization process as its core. The *expansion-based hybrid point sampling* (EHPS) strategy is applied to coarse-level initialization, while the *view-dominant details-aware point sampling* (VDPS) strategy is employed in fine-level re-initialization. The VDPS strategy consists of two sequential sub-strategies: *intra-view point sampling* (IVPS) and *cross-view point sampling* (CVPS).

work focuses on resolving this issue by using dense points sampled with respect to sparse training views for 3D Gaussian initialization.

3 Method

As illustrated in Figure 2, ReInGS is characterized by a two-stage coarse-to-fine 3D Gaussian re-initialization process. During coarse-level initialization, dense points are sampled both uniformly and randomly in an expanded space for the scene to initialize the first set of Gaussians Θ_1, followed by a pre-optimization for the representation of the scene's global geometric structure. The following fine-level re-initialization begins with selecting pixels based on intra-view importance regarding the scene's local geometric details in the view images rendered using Θ_1. Points mapped in the world coordinate system from these pixels serve as the positions of the second set of Gaussians Θ_2, followed by pre-optimization. Afterward, those Gaussians in Θ_2 with cross-view importance that contribute to rendering the scene's local geometric details are chosen to produce the re-initialized Gaussians Θ. The final optimization adjusts Θ in terms of appearance and geometry to synthesize images from unseen viewpoints. In this section, both stages in the Gaussian re-initialization process are described from a strategy-wise perspective.

3.1 Preliminaries

3D Gaussian Splatting. 3DGS [14] adopts a set of 3D Gaussians Θ for scene representation and achieves a 2D image I with pixel-wise color C from a viewpoint characterized by a camera pose P. The i-th Gaussian is formulated as $\theta_i = \{\mu_i, S_i, R_i, \alpha_i, f_i\}$, where $\mu_i \in \mathbb{R}^3$ is the mean representing the centroid or the position, $S_i \in \mathbb{R}^{3\times3}$ is the scaling matrix, $R_i \in \mathbb{R}^{3\times3}$ is the rotation matrix,

$\alpha_i \in \mathbb{R}$ is the opacity, and $f_i \in \mathbb{R}^K$ is the K-dimensional color feature. The basis function of θ_i on the position x is defined as follows:

$$\mathrm{G}_i(x) = e^{-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)}, \quad \Sigma_i = R_i S_i S_i' R_i'. \tag{1}$$

where covariance matrix $\Sigma_i \in \mathbb{R}^{3\times3}$ is defined with a orthogonal matrix R_i and a diagonal matrix S_i.

During the rasterization process, 3D Gaussians are splatted onto an image plane via the projection. For an image pixel p, its color $\mathrm{C}(p)$ is the result of the α-blending process of N ordered Gaussians intersecting with the camera ray toward p, as follows:

$$\mathrm{C}(p) = \sum_{i\in N} c_i \alpha_i T_i, \quad T_i = \prod_{j=1}^{i-1}(1-\alpha_j). \tag{2}$$

where c_i is the color calculated from the spherical harmonic (SH) coefficients of f_i, and T_i is the corresponding accumulated transmittance at p. Different from the ray sampling strategy in NeRF [19], these Gaussians are hit by a parallelized rasterizer with respect to the pixel p and the camera pose P.

Few-Shot Novel View Synthesis. To accomplish the task of NVS, the vanilla 3DGS usually relies on a set of adequate images as training views along with the camera poses of the corresponding viewpoints, *i.e.*, $\{(\hat{\mathrm{I}}_i, \hat{P}_i)\}, i \in [1, 2, ..., n]$ where n is relatively large. In comparison, few-shot NVS is characterized by a very limited number of training views with the camera poses from dissimilar viewpoints, *i.e.*, $\{(\hat{\mathrm{I}}_i, \hat{P}_i)\}, i \in [1, 2, ..., k]$ where $k \ll n$ is usually set to 2, 3 or 4. In the few-shot setting, 3DGS works with insufficient initialization and yields outputs with less accuracy. In this paper, we focus on three challenging settings regarding the sparsity of training views, *i.e.*, $k \in [2, 3, 4]$.

3.2 Expansion-based Hybrid Point Sampling

As raised in [2, 13], fitting of 3D Gaussians in 3DGS for few-shot NVS demonstrates a close relevance to a Gaussian Mixture Model (GMM), *i.e.*, the parameters of Gaussians are sought to maximize the likelihood given samples for training. During this process, local optima would arise upon transporting sparsely initialized Gaussians to their target positions that are distant. The underlying cause is two-fold: (1) the position μ_i of a Gaussian θ_i is dominantly influenced by a set of limited neighbors, whose positions are within three standard deviations σ_i of θ_i from μ_i, following the empirical rule; (2) since Adaptive Density Control (ADC) [14] relies on cumulating the directionally various gradients, the long-distance paths transporting Gaussians toward the expected positions hardly enable the loss function to monotonically decrease. This phenomenon is reflected in few-shot NVS as overly large Gaussians that attempt to approximate the scene by filling its space, but are not placed where expected, due to the insufficient

information provided by sparse training views. In particular, it has a negative impact on representing the scene's global structure [38]. Therefore, the key to mitigate this impact is to compromise the long-range transportation of Gaussians by shortening the corresponding paths.

As a countermeasure, we design an *expansion-based hybrid point sampling* (EHPS) strategy and deploy it in the coarse-level stage to generate dense points efficiently covering the scene's space for Gaussian initialization. Due to the sparsity of training views, the scene's space usually being reconstructed by a SfM method [23] is likely to have an incomplete extent V compared to the ground truth, leading to the undersampling of points. Consequently, the neighbors within a point's three standard deviations could become even more sparse, and the corresponding Gaussian would be underinfluenced by the inadequate neighbors. Therefore, the EHPS strategy begins with a scaling operation on V as a rectangular cuboid for space expansion along each axis with a scaling coefficient $\gamma > 1$. As a result, the expansion broadens the exploration spaces for point sampling and Gaussian optimization.

In the context of Gaussian transportation, our finding (see Section 1 and Table 1) can be interpreted as a *point-path relationship*: if points are with higher density, Gaussians can be initialized probably closer to their expected positions, enabling optimization paths to become less distant. Accordingly, our EHPS strategy continues to sample dense points within the expanded space V'. Since overly dense points boost computational costs, points are sampled both uniformly and randomly in a hybrid manner. This design is based on that uniformly sampled points can benefit the performance-efficiency in 3DGS-based NVS [1], and randomly sampled points have been found to provide the scene's geometric details in **Appendix** Section E (Technical Appendix is included in Supplementary Materials). In addition, using uniformly sampled points enables Gaussians to better capture the scene's global geometric structure, especially when different objects or objects overlap on the same positions during the projection than only randomly sampled points. Our hybrid sampling is conducted using sampling coefficients β_1, β_2 and side lengths $[\Delta'_x, \Delta'_y, \Delta'_z]$ of V'. Please refer to Section A in **Appendix** for details.

Dense points sampled by the EHPS strategy provide the initial positions for the first set of 3D Gaussians Θ_1. These Gaussians are updated by a coarse-level pre-optimization in Section 3.4.

3.3 View-dominant Details-aware Point Sampling

As illustrated in Figure 3, the image rendered using the pre-optimized Θ_1 reflects high-quality global geometric structure but unsatisfactory local geometric details of the scene, *i.e.*, the red box in (a). Meanwhile, Gaussian importance map (highest contribution for rendering) in (b) demonstrates that large-scale Gaussians dominate Θ_1 and focus on the global geometric structure rather than details. This can be attributed to that coarse-level initialization is view-irrelevant and the corresponding pre-optimization handles Gaussians indiscriminately. Conse-

quently, those regions in the scene that are important to the rendering of local geometric details are underestimated, which causes unignorable floating artifacts.

To remedy this shortcoming, it is straightforward to identify such regions from rendered images and training views for point sampling with awareness of local geometric details during Gaussian re-initialization. Nevertheless, we notice a dilemma in few-shot NVS: exploiting each training view for rendering is necessary, whereas scene reconstruction relies heavily on multi-view consistency that could be undermined by overfitting to training views. Hence, we design a *view-dominant details-aware point sampling* (VDPS) strategy and deploy it in the fine-level stage for Gaussian re-initialization. The VDPS strategy is characterized by two sequential sub-strategies from intra-view and cross-view perspectives.

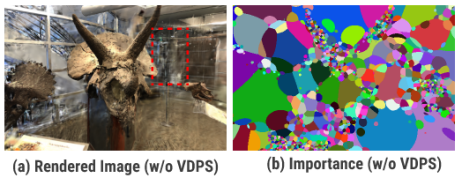

Fig. 3: An example image synthesized using 3D Gaussians initialized without the VDPS strategy and the corresponding importance map.

Intra-View Point Sampling. Given each training viewpoint, those 2D image pixels representing the scene's geometric information can be exploited to sample 3D scene points that are aware of local geometric details. In addition, such points should be associated with the scene regions that are yet to be reconstructed but matter to the improvement of rendering quality. Following the idea of image-order rendering [7], we design an *intra-view point sampling* (IVPS) sub-strategy based on quantifying the importance of an image pixel p to the awareness of local geometric details and the improvement of rendering quality, as follows:

$$\Omega_P(p) = \delta\Omega_D(p) + (1-\delta)\Omega_T(p), \tag{3}$$

where $\Omega_D(p)$ and $\Omega_T(p)$ represent two sub-scores for p regarding depth and accumulated transmittance, and $\delta \in [0,1]$ is a weighting coefficient.

Depth information [27,38] is an effective means to mitigate floating artifacts. In particular, a rendered image satisfactorily demonstrating the scene's geometric structure, which is regarded as the image's foreground, should correspond to a depth map close to the ground truth. Therefore, we formulate the depth sub-score for pixel p as follows:

$$\begin{aligned} \Omega_D(p) &= 1 - \texttt{norm}(\|\mathrm{D}_\mathrm{g}(p) - \hat{\mathrm{D}}_\mathrm{g}(p)\|_2), \\ \mathrm{D}_\mathrm{g}(p) &= \frac{\mathrm{D}(p) - \mu(\mathrm{D})}{\sigma(\mathrm{D})}, \quad \hat{\mathrm{D}}_\mathrm{g}(p) = \frac{\hat{\mathrm{D}}(p) - \mu(\hat{\mathrm{D}})}{\sigma(\hat{\mathrm{D}})}, \end{aligned} \tag{4}$$

where $\hat{\mathrm{D}}$ and D denote the ground truth depth map and the rendered depth map, respectively, and $\hat{\mathrm{D}}(p)$ and $\mathrm{D}(p)$ represent the corresponding depth values at p with image coordinates $[u_p, v_p]$. It is noteworthy that the ground truth

depth map $\hat{\mathrm{D}}$ is generated by a pre-trained monocular depth estimation [33]. Considering the challenge of pixel-point mapping, $\mathrm{D}(p)$ is calculated using the view's camera position o, extrinsic parameters $[W, t] \in \mathbb{R}^{3\times4}$, and N ordered Gaussians intersecting with the ray for p, as follows:

$$\mathrm{D}(p) = \sum_{i\in N} \alpha_i' \prod_{j=1}^{i-1} (1-\alpha_j')(W\|\mu_i - o\|_2 + t), \tag{5}$$

where $\alpha_i' = \alpha_j' = 0.99$ represents high opacity to avoid floating artifacts. As a result, a triple $[u_p, v_p, D(p)]$ is obtained to map p to a corresponding scene point.

At the same time, the α-blending's mechanism suggests that scene regions with low accumulated transmittance are likely to be under reconstruction and need attention for high-quality rendering. Hence, we define the accumulated transmittance sub-score for pixel p as follows:

$$\Omega_T(p) = 1 - \sum_{i\in N} \alpha_i T_i. \tag{6}$$

Using the pixel-wise importance score Ω_P in Eq. 3 as importance-based sampling probability, our IVPS sub-strategy randomly selects approximately 10% of pixels from the rendered image for each training viewpoint. These pixels are mapped to scene points in the world coordinate system using their triples and the view's camera extrinsic parameters [14]. Afterward, the second set of 3D Gaussians Θ_2 are initialized using these points as positions and the corresponding pixel values as colors, followed by a fine-level pre-optimization as described in Section 3.4.

Cross-View Point Sampling. Given the pre-optimized Θ_2 after intra-view point sampling, challenges still occur in preventing multi-view consistency from being undermined. Following the idea of object-order rendering [6,7], we design a *cross-view point sampling* (CVPS) sub-strategy leveraging intersections between Gaussians and camera rays across all training views. Specifically, we define a Gaussian-wise importance score for each $\theta_i \in \Theta_2$ to measure its contribution to rendering by calculating the amount of pixels covered by its projection, as follows:

$$\begin{aligned} \Omega_G(\theta_i) &= \sum_{\mathrm{I}\in\mathcal{I}} \sum_{p\in\mathrm{I}} \Pi(\theta_i, p)\alpha_i T_i, \\ \Pi(\theta_i, p) &= \begin{cases} 1, & \text{if the projection of } \theta_i \text{ covers } p, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{7}$$

where $\mathcal{I}$ represents the set of rendered images for all training viewpoints. It is noteworthy that our CVPS sub-strategy adjusts this importance score with Gaussian's opacity and transmittance at each pixel, reflecting its engagement in α-rendering.

As Gaussians vary in size, large-scale ones could yield higher importance scores Ω_G for probably covering more image pixels. Consequently, directly using

Table 2: Quantitative comparison of our ReInGS and other state-of-the-art methods regarding the performance of few-shot NVS on the LLFF dataset under 2-view, 3-view, and 4-view settings. Best, second-best, and third-best scores are highlighted. −, ✓and ✗denote N/A, with, and without sparsity discrepancy. ∗ denotes our reproduction after removing the evaluation trick in [21].

Methods	Sparsity Discrepancy	PSNR↑			SSIM↑			LPIPS↓			AVGE↓		
		2	3	4	2	3	4	2	3	4	2	3	4
RegNeRF [20]	-	16.55	19.41	21.49	0.468	0.627	0.713	0.417	0.306	0.257	0.207	0.149	0.105
FlipNeRF [24]	-	16.57	19.74	21.55	0.485	0.668	0.721	0.407	0.282	0.260	0.188	0.123	0.098
FreeNeRF [32]	-	17.07	19.97	21.80	0.513	0.652	0.713	0.376	0.280	0.259	0.187	0.134	0.105
SparseNeRF [27]	-	17.74	20.33	21.90	0.513	0.657	0.720	0.386	0.302	0.260	0.171	0.127	0.101
3DGS [14]	✓	15.45	17.25	18.81	0.522	0.574	0.669	0.383	0.300	0.257	0.196	0.155	0.124
FSGS [38]	✓	18.45	20.31	21.11	0.617	0.652	0.703	0.325	0.288	0.220	0.137	0.104	0.098
DNGaussian [17]	✓	17.72	19.12	20.76	0.537	0.591	0.717	0.357	0.294	0.236	0.160	0.132	0.102
CoR-GS [36]	✓	18.73	20.45	21.70	0.636	0.712	0.769	0.247	0.196	0.169	0.125	0.101	0.082
3DGS [14]	✗	12.83	14.99	17.31	0.311	0.483	0.584	0.470	0.362	0.297	0.273	0.202	0.153
FSGS [38]	✗	15.58	18.28	20.30	0.461	0.572	0.673	0.463	0.312	0.273	0.217	0.147	0.110
DNGaussian [17]	✗	16.47	18.87	20.86	0.450	0.583	0.690	0.408	0.313	0.266	0.192	0.141	0.107
CoR-GS [36]	✗	14.96	16.32	18.03	0.363	0.452	0.561	0.429	0.354	0.295	0.225	0.188	0.151
CoherentGS* [21]	✗	15.04	16.98	18.49	0.571	0.671	0.724	0.345	0.281	0.253	0.183	0.141	0.126
FewViewGS [34]	✗	-	18.96	-	-	0.585	-	-	0.307	-	-	0.135	-
ReInGS (Ours)	✗	19.39	20.88	22.24	0.599	0.687	0.748	0.296	0.256	0.239	0.132	0.108	0.092

Ω_G for deterministic selection, *i.e.*, Top-K selection, would overshadow small-scale ones that may also maintain the scene's local geometric details. Therefore, an importance-based sampling probability is calculated based on normalizing Ω_G, as follows:

$$\mathcal{P}(\theta_i) = \frac{\Omega_G(\theta_i)}{\sum_{j \in N} \Omega_G(\theta_j)}. \tag{8}$$

Our CVPS sub-strategy adopts this sampling probability to randomly select approximately 40% Gaussians from Θ_2. Their positions and colors are used to initialize the last set of Gaussians Θ, *i.e.*, the Gaussians that have been re-initialized for final optimization.

3.4 Gaussian Optimization

As discussed, all three optimization processes aim to inject information about the scene's appearance and geometry into 3D Gaussians. In terms of appearance, our ReInGS adopts the photometric loss [14, 17, 38] as $\mathcal{L}_A$ between the rendered image I and the corresponding training view $\hat{\text{I}}$.

For geometry, different losses are applied as follows: We use hard depth regularization $\mathcal{L}_C$ [17] in *Coarse-level Pre-optimization* to optimize near surface Gaussians for reducing floating artifacts at the beginning. In *Fine-level Pre-optimization and Final Optimization*, we adopt soft depth regularization $\mathcal{L}_F(p)$ [17] to extract reasonable local geometric details and less hollowness in an unseen view.

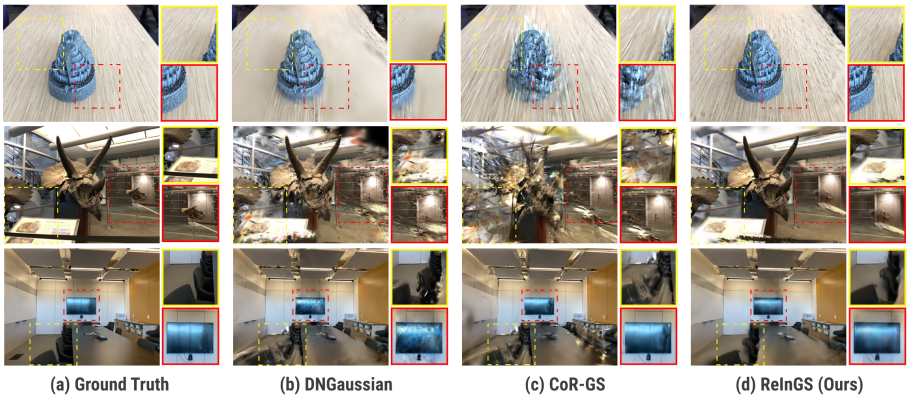

Fig. 4: Qualitative comparisons about our ReInGS, DNGaussian [17], and CoR-GS [36] on the LLFF dataset. ReInGS successfully captured the scene's global geometric structure using the EHPS strategy and eliminated blurring and floating artifacts in local details through the VDPS strategy.

The training loss for Gaussian optimization in rendered image set $\mathcal{I}$ is formulated as follows:

$$\mathcal{L} = \begin{cases} \sum_{\mathrm{I}\in\mathcal{I}} \sum_{p\in\mathrm{I}} \mathcal{L}_C(p) + \sum_{\mathrm{I}\in\mathcal{I}} \mathcal{L}_A(\mathrm{I}), & \text{if being coarse-level,} \\ \sum_{\mathrm{I}\in\mathcal{I}} \sum_{p\in\mathrm{I}} \mathcal{L}_F(p) + \sum_{\mathrm{I}\in\mathcal{I}} \mathcal{L}_A(\mathrm{I}), & \text{otherwise,} \end{cases} \tag{9}$$

Please refer to Section B in the **Appendix** for details.

4 Experiments

We implemented ReInGS in PyTorch using Adam optimizer [15] on an NVIDIA A6000 (48GB) GPU. On each dataset, training lasted for 6000 iterations with parameters $\delta = 0.03$, $\beta_1 = 200$, $\beta_2 = 50$, and $\gamma = 1.4$. For Gaussian initialization in most of 3DGS-based methods for comparison, we evaluated two settings: sparse random points following DNGaussian [17], *i.e.* without sparsity discrepancy, and dense fused stereo points using SfM [23] from extended views following FSGS [38], *i.e.* with sparsity discrepancy. We avoided using fused stereo points from few-shot training views alone due to their inherent instability, as mentioned in Section 1.

4.1 Comparisons

Targeting the challenging few-shot settings using 2, 3 and 4 training views, we conducted the experiments on two benchmarking datasets: LLFF [18] and DTU [12]. Performance was measured in common metrics: PSNR, SSIM [28], LPIPS [37] and AVGE [20]. Please refer to Section C in **Appendix** for details. We compared ReInGS with representative few-shot NVS methods.

Table 3: Quantitative comparison on the DTU [12] under 3-view setting. SD denotes sparsity discrepancy.

Methods	SD	PSNR↑	SSIM↑	LPIPS↓	AVGE↓
FreeNeRF [32]	-	19.92	0.787	0.182	0.098
FSGS [38]	✓	17.14	0.818	0.162	0.110
DNGaussian [17]	✓	18.91	0.790	0.176	0.102
CoR-GS [36]	✓	19.21	0.853	0.119	0.087
3DGS [14]	✗	12.77	0.681	0.278	0.167
FSGS [38]	✗	12.93	0.675	0.296	0.209
DNGaussian [17]	✗	18.85	0.763	0.227	0.116
CoR-GS [36]	✗	15.00	0.685	0.241	0.162
FewViewGS [34]	✗	19.13	0.792	0.186	0.110
ReInGS (Ours)	✗	21.30	0.832	0.170	0.081

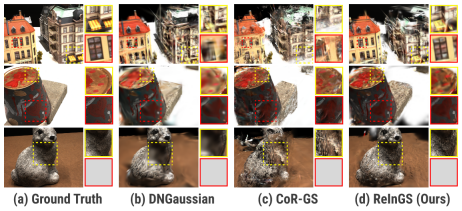

Fig. 5: Qualitative comparisons on the DTU under 3-view setting. Our ReInGS effectively reduced blurring and floating artifacts, producing high-quality images through efficient coarse-to-fine re-initialization.

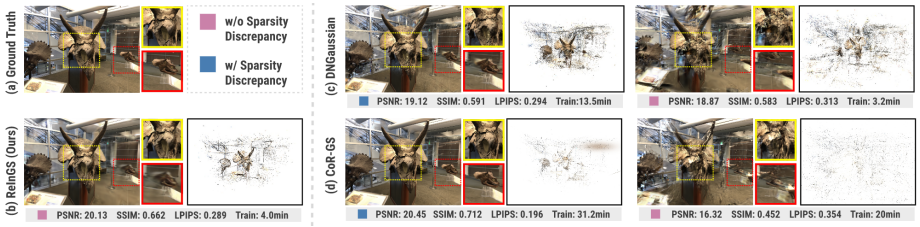

Fig. 6: Visualization of example images and corresponding 3D Gaussians synthesized using our ReInGS, DNGaussian [17], and CoR-GS [36] on LLFF (*Horns* scene) under 3-view setting. In the absence of sparsity discrepancy, DNGaussian (c) and CoR-GS (d) struggled to capture satisfactory results with the massive points cloud, whereas our ReInGS achieved high-quality synthesis with both accurate global geometric structure and refined local geometric details.

Quantitative Comparison. As shown in Table 2 and 3, the proposed method consistently outperformed other 3DGS-based methods across all metrics and settings in cases without sparsity discrepancy. On the LLFF dataset, compared to the second-best methods, ReInGS achieved improvements of 2.92 dB / 1.92 dB / 1.38 dB in PSNR, 0.028 / 0.016 / 0.024 in SSIM, 0.049 / 0.025 / 0.014 in LPIPS, and 0.051 / 0.027 / 0.015 in AVGE under 2-view / 3-view / 4-view settings, respectively, without sparsity discrepancy. On the DTU dataset, ReInGS also outperformed with increases of 2.17 dB in PSNR, 0.040 in SSIM, 0.016 in LPIPS, and 0.029 in AVGE for the 3-view settings, compared to the second-best methods, in the case without sparsity discrepancy. For complete results in the 2-view and 4-view settings, please refer to Table 5 in the **Appendix**.

Furthermore, the performance of other 3DGS-based methods dropped significantly when switched from the sparsity discrepancy case to the case using random points. This echoed our findings discussed in Section 1. In contrast, ReInGS delivered results competitive even with counterparts that used data leakage, which demonstrated the effectiveness of our coarse-to-fine re-initialization.

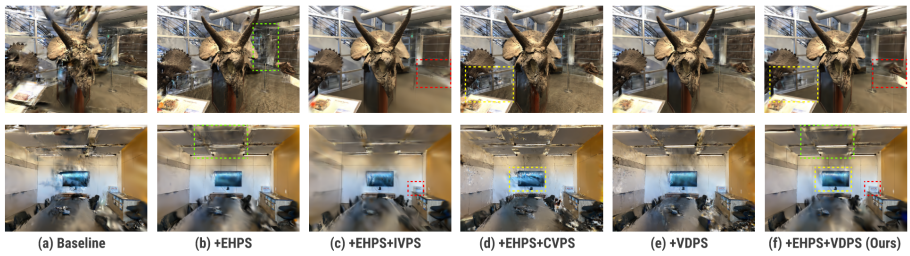

Fig. 7: Visualization of examples in the ablation studies on LLFF under 3-view setting.

Qualitative Comparison. We visualized rendering examples in the case without sparsity discrepancy using our ReInGS, alongside DNGaussian and CoR-GS, on both datasets. As shown in Figure 4 As evidenced by the confused Gaussians within specific regions, *e.g.*, Example 1, and 2, Figure 4 (c) and Figure 5 (c). It remains a challenge to achieve accurate geometry with random Gaussian initialization, as demonstrated in Figure 6 (d). This challenge also explains the significant performance gap between CoR-GS and DNGaussian. As for our ReInGS, the coarse-level EHPS strategy captured the global geometric structure, particularly visible in the *Horns* scene's upper-right corner (Example 2, Figure 4). The fine-level VDPS enhanced local geometric details and eliminated artifacts, improving surface consistency in the *Room* scene (Example 3, Figure 4), and in the other highlighted regions in Figure 4 and Figure 5. These improvements demonstrate the effectiveness of our coarse-to-fine re-initialization process.

4.2 Ablation Studies

To evaluate the contribution of each strategy in our ReInGS, we conducted ablation studies on the LLFF dataset under the 3-view setting, using a reproduced DNGaussian [17] as the baseline.

As shown in Table 4, the results demonstrate the significant positive impact of each strategy, with the combination of all strategies achieving the best performance. Specifically, the progressive adoption of the EHPS strategy in coarse-level initialization stage, along with IVPS and CVPS sub-strategies in the fine-level re-initialization stage, resulted in improvements in PSNR by 0.49 dB, 0.41 dB, and 1.09 dB, respectively.

Employing the IVPS resulted in a significant enhancement in LPIPS and SSIM, highlighting its focus on local geometric details and aligning with our assumption of exploiting geometric importance. In comparison, adopting the CVPS significantly increased PSNR, indicating that preserving multi-view consistency enhances texture details. The sequential

Table 4: Ablation studies of strategy-wise contribution on LLFF under 3-view setting.

Coarse-level	Fine-level		PSNR↑	LPIPS↓	SSIM↑
EHPS	IVPS	CVPS			
			18.87	0.313	0.583
✓			19.38	0.348	0.589
✓	✓		19.79	0.260	0.645
✓		✓	20.05	0.314	0.611
	✓	✓	19.66	0.285	0.638
✓	✓	✓	20.88	0.256	0.687

Fig. 8: Visualization of the limitations of our ReInGS: hollows and cracks.

combination of these two sub-strategies within the VDPS strategy underscores the importance of incorporating both intra-view and multi-view considerations for few-shot NVS. As shown in Rows 5 and 6, adopting the VDPS strategy alone was effective but yielded limited performance gains. However, incorporating the EHPS strategy in the coarse-level initialization stage further improved PSNR by up to 1.12 dB. This significantly enhanced the final rendering results, demonstrating the consistency and effectiveness of our coarse-to-fine design.

We visualized rendering examples from the ablation studies in Figure 7. The EHPS strategy produced superior global geometric structure, while the VDPS strategy significantly reduced the floating artifacts toward high-quality local geometric details. Please refer to Section E in **Appendix** for detailed analysis and more ablation studies.

5 Conclusions

In this paper, we identified the issue of sparsity discrepancy in existent 3DGS-based methods for few-shot NVS, and proposed a novel 3DGS-based method (ReInGS) characterized by a coarse-to-fine process to re-initialize 3D Gaussians against this issue. The proposed method integrates an expansion-based hybrid point sampling in the coarse-level initialization stage, and a view-dominant details-aware point sampling strategy in the fine-level re-initialization stage. Extensive experiments demonstrated that ReInGS achieved the state-of-the-art performance while effectively resolving sparsity discrepancy in few-shot NVS.

6 Limitations

Despite the significant improvements in synthesis quality with the coarse-to-fine process to re-initialize 3D Gaussians against the sparsity discrepancy for few-shot NVS, the proposed method is expected to resolve the following limitations in the future.

– **Computational Efficiency.** As discussed in Section D.2 of the **Appendix**, our ReInGS was not primarily designed with computational efficiency as a main objective, yet it achieves competitive performance across all evaluation metrics. The dense uniform and random point sampling adopted in the coarse stage, together with the subsequent sampling strategies, result in increased training time and memory consumption. Although these computational costs

are acceptable considering the performance gains, we plan to further optimize the sampling strategies in future work to improve efficiency.
– **Hollows and Cracks.** Since the Gaussians could not overlap all pixels, any empty region between neighboring Gaussians would cause hollows and cracks when the camera pose was altered. As illustrated in the highlighted regions of Figure 8, this issue has been regarded as a common limitation in the 3DGS-based methods for few-shot NVS.

Acknowledgments. This research was supported by the National Natural Science Foundation of China (Grant No. U23A20311).

References

1. Cai, Y., Liang, Y., Wang, J., Wang, A., Zhang, Y., Yang, X., Zhou, Z., Yuille, A.: Radiative gaussian splatting for efficient x-ray novel view synthesis. In: Computer Vision – ECCV 2024. p. 283–299. Springer Nature Switzerland (Sep 2024). https://doi.org/10.1007/978-3-031-73232-4_16, http://dx.doi.org/10.1007/978-3-031-73232-4_16
2. Charatan, D., Li, S., Tagliasacchi, A., Sitzmann, V.: pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In: arXiv (2023)
3. Chen, A., Xu, Z., Zhao, F., Zhang, X., Xiang, F., Yu, J., Su, H.: Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 14124–14133 (2021)
4. Chen, Z., Li, Z., Song, L., Chen, L., Yu, J., Yuan, J., Xu, Y.: Neurbf: A neural fields representation with adaptive radial basis functions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4182–4194 (2023)
5. Deng, K., Liu, A., Zhu, J.Y., Ramanan, D.: Depth-supervised nerf: Fewer views and faster training for free. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12882–12891 (2022)
6. Ding, J., He, Y., Yuan, B., Yuan, Z., Zhou, P., Yu, J., Lou, X.: Ray reordering for hardware-accelerated neural volume rendering. IEEE Transactions on Circuits and Systems for Video Technology (2024)
7. Drebin, R.A., Carpenter, L., Hanrahan, P.: Volume rendering. ACM Siggraph Computer Graphics **22**(4), 65–74 (1988)
8. Fan, Z., Wen, K., Cong, W., Wang, K., Zhang, J., Ding, X., Xu, D., Ivanovic, B., Pavone, M., Pavlakos, G., Wang, Z., Wang, Y.: Instantsplat: Sparse-view gaussian splatting in seconds (2024)
9. Han, L., Zhou, J., Liu, Y.S., Han, Z.: Binocular-guided 3d gaussian splatting with view consistency for sparse view synthesis. In: Advances in Neural Information Processing Systems (NeurIPS) (2024)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
11. Jain, A., Tancik, M., Abbeel, P.: Putting nerf on a diet: Semantically consistent few-shot view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5885–5894 (2021)
12. Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanæs, H.: Large scale multi-view stereopsis evaluation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 406–413 (2014)
13. Jung, J., Han, J., An, H., Kang, J., Park, S., Kim, S.: Relaxing accurate initialization constraint for 3d gaussian splatting. arXiv preprint arXiv:2403.09413 (2024)
14. 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), https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization 3rd international conference on learning representations. In: ICLR 2015-Conference Track Proceedings. vol. 1 (2015)
16. Kulhánek, J., Derner, E., Sattler, T., Babuška, R.: Viewformer: Nerf-free neural rendering from few images using transformers. In: European Conference on Computer Vision. pp. 198–216. Springer (2022)

17. Li, J., Zhang, J., Bai, X., Zheng, J., Ning, X., Zhou, J., Gu, L.: Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20775–20785 (2024)
18. 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 Transactions on Graphics (ToG) **38**(4), 1–14 (2019)
19. 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: ECCV (2020)
20. Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S., Geiger, A., Radwan, N.: Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5480–5490 (2022)
21. Paliwal, A., Ye, W., Xiong, J., Kotovenko, D., Ranjan, R., Chandra, V., Kalantari, N.K.: Coherentgs: Sparse novel view synthesis with coherent 3d gaussians. arXiv preprint arXiv:2403.19495 (2024)
22. Peng, Z., Shao, T., Liu, Y., Zhou, J., Yang, Y., Wang, J., Zhou, K.: Rtg-slam: Real-time 3d reconstruction at scale using gaussian splatting. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–11 (2024)
23. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
24. Seo, S., Chang, Y., Kwak, N.: Flipnerf: Flipped reflection rays for few-shot novel view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 22883–22893 (2023)
25. 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)
26. Truong, P., Rakotosaona, M.J., Manhardt, F., Tombari, F.: Sparf: Neural radiance fields from sparse and noisy poses. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4190–4200 (2023)
27. Wang, G., Chen, Z., Loy, C.C., Liu, Z.: Sparsenerf: Distilling depth ranking for few-shot novel view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9065–9076 (2023)
28. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
29. 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. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20310–20320 (2024)
30. Wynn, J., Turmukhambetov, D.: Diffusionerf: Regularizing neural radiance fields with denoising diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4180–4189 (2023)
31. 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. pp. 5438–5448 (2022)
32. Yang, J., Pavone, M., Wang, Y.: Freenerf: Improving few-shot neural rendering with free frequency regularization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8254–8263 (2023)

33. Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H.: Depth anything: Unleashing the power of large-scale unlabeled data. In: CVPR (2024)
34. Yin, R., Yugay, V., Li, Y., Karaoglu, S., Gevers, T.: Fewviewgs: Gaussian splatting with few view matching and multi-stage training (2024), https://arxiv.org/abs/2411.02229
35. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4578–4587 (2021)
36. Zhang, J., Li, J., Yu, X., Huang, L., Gu, L., Zheng, J., Bai, X.: Cor-gs: Sparse-view 3d gaussian splatting via co-regularization. arXiv preprint arXiv:2405.12110 (2024)
37. 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 Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
38. Zhu, Z., Fan, Z., Jiang, Y., Wang, Z.: Fsgs: Real-time few-shot view synthesis using gaussian splatting. In: Computer Vision – ECCV 2024. p. 145–163. Springer Nature Switzerland (Oct 2024). https://doi.org/10.1007/978-3-031-72933-1_9, http://dx.doi.org/10.1007/978-3-031-72933-1_9