

# INITIALIZE TO GENERALIZE: A STRONGER INITIALIZATION PIPELINE FOR SPARSE-VIEW 3DGS

000  
001  
002  
003  
004  
005 **Anonymous authors**  
006 Paper under double-blind review  
007  
008  
009  
010

## ABSTRACT

011 Sparse-view 3D Gaussian Splatting (3DGS) often overfits to the training views,  
012 leading to artifacts like blurring in novel view rendering. Prior work addresses  
013 it either by enhancing the initialization (*i.e.*, the point cloud from Structure-from-  
014 Motion (SfM)) or by adding training-time constraints (regularization) to the 3DGS  
015 optimization. Yet our controlled ablations reveal that initialization is the decisive  
016 factor: it determines the attainable performance band in sparse-view 3DGS, while  
017 training-time constraints yield only modest within-band improvements at extra  
018 cost. Given initialization’s primacy, we focus our design there. Although SfM  
019 performs poorly under sparse views due to its reliance on feature matching, it still  
020 provides reliable seed points. Thus, building on SfM, our effort aims to supple-  
021 ment the regions it fails to cover as comprehensively as possible. Specifically,  
022 we design: (i) frequency-aware SfM that improves low-texture coverage via low-  
023 frequency view augmentation and relaxed multi-view correspondences; (ii) 3DGS  
024 self-initialization that lifts photometric supervision into additional points, com-  
025 pensating SfM-sparse regions with learned Gaussian centers; and (iii) point-cloud  
026 regularization that enforces multi-view consistency and uniform spatial cover-  
027 age through simple geometric/visibility priors, yielding a clean and reliable point  
028 cloud. Our experiments on LLFF and Mip-NeRF360 demonstrate consistent gains  
029 in sparse-view settings, establishing our approach as a stronger initialization strat-  
030 egy.  
031

## 1 INTRODUCTION

032 Novel view synthesis (Mildenhall et al., 2021; Chan et al., 2023; Liu et al., 2023) plays a cen-  
033 tral role in applications such as virtual and augmented reality (Anthes et al., 2016), free-viewpoint  
034 video (Zhou et al., 2025), and digital content creation (Cao et al., 2024), where the goal is to gener-  
035 ate realistic novel views from captured imagery. To enable such capabilities, recent advances in 3D  
036 neural scene representations (Mildenhall et al., 2021; Deng et al., 2022) have demonstrated remark-  
037 able performance, with 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) emerging as a highly  
038 efficient alternative to neural radiance fields (Mildenhall et al., 2021). In practice, however, data  
039 acquisition is often limited to a small number of viewpoints due to constraints in hardware, cost, or  
040 capture conditions. Under such sparse-view inputs, vanilla 3DGS is prone to severe overfitting to the  
041 training views (Zhu et al., 2024; Dai & Xing, 2025), resulting in noticeable artifacts and persistent  
042 floaters when rendering unseen views.  
043

044 To address this challenge, existing solutions broadly fall into two categories. The first focuses on  
045 enhancing the initial point cloud of 3DGS, which is typically obtained from Structure-from-Motion  
046 (SfM) (Ullman, 1979). For instance, CoMapGS (Jang & Pérez-Pellitero, 2025) leverages a depth es-  
047 timator together with a robust visual encoder to produce a stronger initialization. The second line of  
048 work constrains the growth or optimization of Gaussian primitives, often by introducing regulariza-  
049 tion losses during training. A representative example is CoR-GS (Zhang et al., 2024b), which prunes  
050 outlier primitives by enforcing consistency between two parallel Gaussian fields. While such meth-  
051 ods have shown effectiveness in practice, our evidence suggests that training-time regularization  
052 mainly serves as a patch for poor initialization rather than addressing the fundamental bottleneck in  
053 sparse-view 3DGS.

We begin with ablation studies comparing the effectiveness of two strategies for mitigating overfitting. In Section 3, we evaluate several state-of-the-art training-time constraints across a range of

054 initialization strengths. These initialization strengths are explicitly controlled ideally (rather than  
 055 being determined solely by the input images). The results show that initialization quality sets the  
 056 attainable performance band, whereas these constraints provide only modest within-band gains. The  
 057 findings indicate that initialization is the more influential lever, which motivates focusing on develop-  
 058 ing a stronger initialization pipeline.

059 We start from the vanilla SfM algorithm since it provides reliable initial seed points, though it pro-  
 060 duces only a few in sparse-view scenarios due to its feature-matching nature. Our work is dedicated  
 061 to accurately supplementing those undercovered regions by fully exploiting the visual information  
 062 contained in the available images, step by step. First, inspired by EAP-GS (Dai & Xing, 2025),  
 063 we modify the standard SfM pipeline by relaxing the minimum track matching numbers from three  
 064 to two, yielding a denser initial point set. Moreover, distinct from prior work, we pre-mask high-  
 065 frequency regions and perform SfM on the resulting augmented, doubled image set, which encour-  
 066 ages a more balanced and richer point distribution. Secondly, to further enrich the point set, we pro-  
 067 pose 3DGS Self-Initialization, which lifts per-pixel photometric constraints into additional points  
 068 by leveraging the learning signal of 3DGS. Specifically, after training a first-pass lightweight 3DGS  
 069 on the input views and reusing all primitive centers as a point cloud, we are able to compensate for  
 070 regions with insufficient image features. Finally, we introduce further regularization techniques to  
 071 refine the final point cloud by removing noise and unreliable points: (1) discarding points observed  
 072 from only a single view due to depth ambiguity, (2) applying clustering-based noise reduction, and  
 073 (3) enforcing normal-based consistency filtering.

074 To summarize, our contributions are as follows:

- 075 1. Our controlled ablations with explicitly set initialization strengths show that sparse-view perfor-  
 076 mance in 3DGS is initialization-limited, and training-time constraints provide only modest improve-  
 077 ment.
- 078 2. We design a rich and reliable three-stage initialization: (i) a low-frequency-aware SfM variant  
 079 that relaxes the minimum track length from three to two and pre-masks high-frequency content to  
 080 improve low-texture coverage; (ii) 3DGS Self-Initialization that trains a first-pass 3DGS and reuses  
 081 all primitive centers as the initial point cloud, turning photometric cues into additional points; and  
 082 (iii) point-cloud regularization for removing noise or unreliable points, including removal of single-  
 083 view points, clustering-based denoising, and a simple geometric prior for consistency and coverage.
- 084 3. Experiments demonstrate that our proposed method achieves state-of-the-art performance on  
 085 LLFF and Mip-NeRF360 datasets and offers a superior initialization choice for sparse-view 3DGS.

## 087 2 RELATED WORK

### 090 2.1 NOVEL VIEW SYNTHESIS

091 Novel view synthesis (NWS) aims to generate photorealistic images from novel viewpoints given  
 092 only a finite set of calibrated observations of a typically static scene (Goesele et al., 2007). Among  
 093 the different strategies, optimization-based methods reconstruct a full 3D scene representation, from  
 094 which novel views can be rendered at arbitrary camera poses. Neural scene representations (Tewari  
 095 et al., 2022) have recently emerged as the dominant paradigm: NeRF, for example, encodes view-  
 096 dependent color and density and renders images through volumetric integration (Mildenhall et al.,  
 097 2021). Beyond such implicit fields, explicit formulations like 3D Gaussian Splatting (3DGS) repre-  
 098 sent scenes as anisotropic Gaussians and achieve real-time rendering with competitive quality (Kerbl  
 099 et al., 2023; Lin et al.; Zhang et al., 2024a). In parallel, generative approaches that exploit large-scale  
 100 priors have also been developed (Liu et al., 2023; Wang et al., 2025; Tang et al., 2024; Szymanow-  
 101 icz et al., 2024; Voleti et al., 2024). These methods provide extremely fast inference and reduced  
 102 reliance on precise camera pose annotations, but they remain limited in maintaining multi-view  
 103 consistency and geometric fidelity.

104 **3D Reconstruction with Sparse Views.** Reconstructing 3D scenes from sparse views is inherently  
 105 under-constrained: limited viewpoints lead to incomplete surface coverage, ambiguous depth esti-  
 106 mation, and frequent artifacts such as floaters and texture misalignment. To address these challenges,  
 107 some works incorporate additional geometric and photometric priors, introducing supervision such  
 as depth (Chung et al., 2024; Zheng et al., 2025) to guide the optimization process. Another line

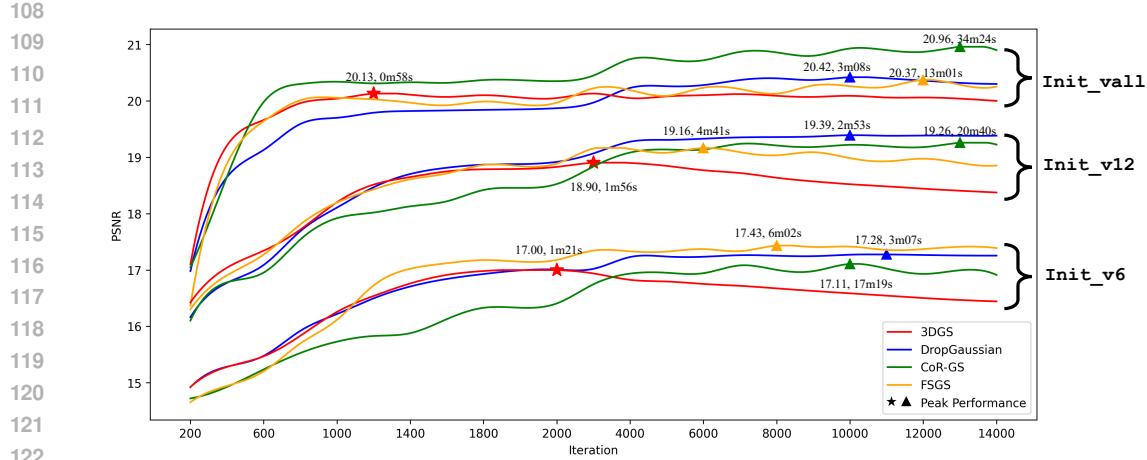


Figure 1: Ablation study on Mip-NeRF360 under the 12-view sparse-view protocol PSNR on average, comparing three training-time constraints (DropGaussian, CoR-GS, and FSGS) against vanilla 3DGS across different initialization strengths (Init\_v6, Init\_v12, Init\_vall). Results show that while regularization methods modestly delay overfitting and improve performance within each initialization band, the final reconstruction quality is primarily governed by initialization strength.

leverages controllable 2D generative models to synthesize novel training views, effectively densifying the supervision and mitigating the scarcity of input observations (Mithun et al., 2025; Wan et al., 2025; Xiong et al., 2023; Chen et al., 2024b;a; Liu et al., 2024).

## 2.2 OVERTFITTING IN SPARSE-VIEW 3DGS

Under sparse views, vanilla 3DGS optimized from photometric supervision tends to overfit the limited inputs, degrading markedly on unseen view rendering. Efforts to mitigate overfitting generally take two forms, the first being to enhance the initialization of the 3DGS point cloud (Bao et al., 2025; Jang & Pérez-Pellitero, 2025; Xu et al., 2025). Intuitively, initialization provides the spatial and geometric bias for subsequent optimization; with a better-distributed seed point cloud, the system is guided toward plausible geometry rather than overfitting to limited photometric cues. For instance, LoopSparseGS (Bao et al., 2025) introduces a progressive Gaussian initialization strategy that iteratively densifies the seed point cloud using pseudo-views, significantly improving coverage and stability under sparse inputs. EAP-GS (Dai & Xing, 2025) augments the SfM point cloud by relaxing track constraints and adaptively densifying underrepresented regions, providing a stronger initialization for few-shot 3DGS.

The second form constrains the optimization or growth of Gaussian primitives during training, typically by introducing regularization losses or pruning strategies to suppress overfitting artifacts (Park et al., 2025; Paliwal et al., 2024; Li et al., 2024; Xu et al., 2025). For instance, CoR-GS (Zhang et al., 2024b) introduces a co-regularization framework that enforces consistency between two parallel Gaussian fields, effectively pruning outliers and improving generalization under sparse views. FSGS (Zhu et al., 2024) imposes frequency-based structural regularization, suppressing high-frequency artifacts and stabilizing training when input views are highly limited. DropGaussian (Park et al., 2025) proposes a dropout-style scheme that randomly removes Gaussians during training, enhancing gradient flow to the remaining primitives and mitigating overfitting without extra priors.

## 3 INITIALIZATION VS. REGULARIZATION: AN EMPIRICAL STUDY

### 3.1 PRELIMINARY AND DEFINITION

**Structure-from-Motion (SfM)** (Ullman, 1979). Given images  $\{I_i\}_{i=1}^N$  with intrinsics  $\{K_i\}$ , SfM estimates camera poses  $P_i = (R_i, t_i)$  and a sparse 3D point set  $\mathcal{X} = \{X_m \in \mathbb{R}^3\}$  from multi-

view feature tracks. Tracks link projections of the same 3D point across views. In standard SfM system (*i.e.*, COLMAP (Schonberger & Frahm, 2016)), tracks observed in at least three images are retained for triangulation and bundle adjustment, yielding a conservative but reliable reconstruction. Candidate points are obtained via two-view triangulation  $T$ :

$$X_m = T(\tilde{x}_{i,m}, \tilde{x}_{j,m}, P_i, P_j), \quad i \neq j, \quad (1)$$

Formally,

$$\{\hat{P}_i\}, \{\hat{X}_m\} = \arg \min_{\{P_i\}, \{X_m\}} \sum_{m: |\mathcal{V}_m| \geq 3} \sum_{i \in \mathcal{V}_m} \rho(\|\pi(K_i, P_i, X_m) - \tilde{x}_{i,m}\|_2^2), \quad (2)$$

where  $\mathcal{V}_m = \{i : \text{point } m \text{ observed in image } i\}$ ,  $\tilde{x}_{i,m} \in \mathbb{R}^2$  is the observed pixel,  $\pi(\cdot)$  denotes the standard perspective projection with intrinsics  $K_i$  and pose  $(R_i, t_i)$ , and  $\rho(\cdot)$  is a robust loss.

**3D Gaussian Splatting (3DGS)** (Kerbl et al., 2023). 3DGS represents a scene with anisotropic 3D Gaussian primitives  $\mathcal{G} = \{g_n\}_{n=1}^M$ , each  $g_n = (\mu_n, \Sigma_n, \alpha_n, \mathbf{c}_n)$  (mean, covariance, opacity, color). In standard practice, primitives are initialized from the SfM point cloud  $\{X_m\}$ , with both their positions and colors initialized from the reconstructed points, and then optimized end-to-end with fixed camera poses. During training, the primitive set is adaptively densified by splitting or cloning Gaussians to grow  $\mathcal{G}$  in regions indicated by image evidence. Rendering produces predictions  $\hat{I}_i$  via differentiable alpha compositing of depth-sorted 2D projections of the Gaussians, and the parameters are learned via a compact photometric objective with a structural-similarity term and regularization:

$$\min_{\mathcal{G}} \sum_{i=1}^N \left[ (1 - \lambda) \sum_{p \in \Omega_i} \|\hat{I}_i(p) - I_i(p)\|_1 + \lambda \text{D-SSIM}(\hat{I}_i, I_i) \right] + \beta \mathcal{R}_{\text{reg}}(\mathcal{G}), \quad (3)$$

where  $\Omega_i$  is the pixel domain,  $\lambda \in [0, 1]$  balances the terms,  $I_i$  and  $\hat{I}_i$  are the  $i$ -th ground-truth and rendered images,  $\text{D-SSIM}(\hat{I}_i, I_i) = 1 - \text{SSIM}(\hat{I}_i, I_i)$ , and  $\mathcal{R}_{\text{reg}}$  collects mild priors on scale-opacity.

**Initialization Strength.** Ideally, a strong initialization should approximate a point cloud in which the point set fully covers all object visible surfaces. Such an ideal seed could, in principle, be obtained by running SfM on a sufficiently large set of views of the scene. Since unlimited viewpoints are not available, we approximate initialization strength per scene by varying the number of input views for SfM.

Formally, let  $\mathcal{V}_{\text{all}}$  denote all available views of a scene, and let the SfM reconstruction from  $\mathcal{V}_{\text{all}}$  serve as a best-available proxy for an ideal initialization, producing a point cloud  $\mathcal{X}_{\text{all}}$ . For a view budget  $n$ , we uniformly sample a subset  $\mathcal{V}_n \subset \mathcal{V}_{\text{all}}$  with  $|\mathcal{V}_n| = n$  in camera pose space, run SfM on  $\mathcal{V}_n$ , and obtain a seed point cloud  $\mathcal{X}_n$ . Larger  $n$  therefore corresponds to higher initialization strength in expectation. On the Mip-NeRF360 (Barron et al., 2022) dataset, we instantiate three levels of initialization strength: `Init_v6`, `Init_v12`, and `Init_vall`, where the subscript denotes the number of input views used by SfM (with `Init_vall` using all available views).

### 3.2 EMPIRICAL ANALYSIS AND FINDINGS

To investigate the causes of sparse-view overfitting, we study two common strategies: (i) improving the SfM initialization cloud and (ii) adding training-time regularization of Gaussian primitives. These strategies are treated as two factors in a controlled ablation, enabling us to isolate their individual and combined effects. Specifically, for training-time constraints, we adopt representative off-the-shelf methods to reflect current practice: FSGS (Zhu et al., 2024), Cor-GS (Zhang et al., 2024b), and DropGaussian (Zhang et al., 2024b). For initialization, rather than comparing heterogeneous approaches, we treat our designed *Initialization Strength* as a controlled variable to isolate its effect. Experiments are conducted on the Mip-NeRF360 dataset, following the standard 12-view sparse-view reconstruction protocol and reporting average PSNR across all scenes.

The performance curves are shown in Figure 1. Across all three initialization levels, different regularization methods show some effectiveness: compared with vanilla 3DGS, they mitigate overfitting by delaying the performance peak and slightly improving reconstruction quality, albeit at the cost of increased computation. Strikingly, the strength of initialization proves decisive for the final outcome. The curves stratify into distinct bands according to initialization level, within which regularization offers only limited gains. This initialization-dominated phenomenon motivates us to design a more effective initialization strategy for sparse-view 3DGS.

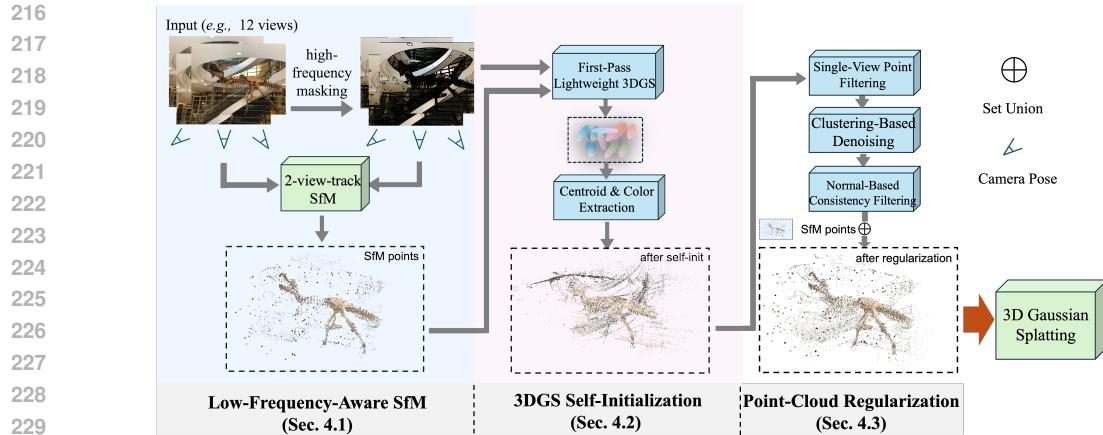


Figure 2: Our initialization pipeline for sparse-view 3DGS. Given sparse multi-view images with camera poses, we first mask high-frequency regions and perform SfM to obtain a raw point cloud, which better captures smooth areas. To compensate for regions lacking distinctive features, we train a lightweight first-pass 3DGS and incorporate its primitive centroids into the initialization. Finally, we apply clustering-based denoising, single-view filtering, and physics-based regularization to rectify the point cloud and suppress noise.

## 4 METHOD

In this section, we aim to develop a rich and reliable initialization pipeline tailored for 3D Gaussian Splatting (3DGS) under sparse-view settings. Building upon vanilla 3DGS, our objective is to compensate for undercovered regions by fully leveraging the visual information available in the input images. The overall framework is illustrated in Figure 2. It comprises three successive components: 1) Low-Frequency-Aware SfM (Sec. 4.1), which augments vanilla SfM with high-frequency masking to improve initialization in smooth regions; 2) 3DGS Self-Initialization (Sec. 4.2), which exploits the pixel-level photometric supervision of 3DGS to supplement points in regions with few distinctive features; and 3) Point-Cloud Regularization (Sec. 4.3), which applies clustering, filtering, and geometry-inspired constraints to refine the point cloud and suppress noise.

### 4.1 LOW-FREQUENCY-AWARE STRUCTURE-FROM-MOTION (SfM)

Under sparse-view capture, limited overlap makes tracks with  $|\mathcal{V}_m| \geq 3$  scarce in vanilla SfM, leading to under-covered and uneven point distributions. To better exploit limited views, EAP-GS (Dai & Xing, 2025) retains two-view tracks in addition to the standard three-view requirement with fixed camera poses  $\hat{P}_i$ .

$$\{\hat{X}'_m\} = \arg \min_{\{\hat{X}'_m\}} \sum_{m: |\mathcal{V}_m| \geq 2} \sum_{i \in \mathcal{V}_m} \rho(\|\pi(K_i, \hat{P}_i, X'_m) - \tilde{x}_{i,m}\|^2). \quad (4)$$

where camera poses  $\hat{P}_i$  can be either provided or estimated from SfM. Unlike EAP-GS, which identifies low-density regions and performs SfM twice merely to enlarge the point set, we adopt a more effective strategy. High-frequency regions naturally produce more track matches in SfM due to their richer image features, leading to an undesired concentration of points. To mitigate this, we pre-mask these regions and feed two image sets into SfM simultaneously, thereby generating a more balanced point cloud.

Formally, given  $N$  scene views  $\{I_i\}_{i=1}^N$  with camera poses  $\{P_i\}_{i=1}^N$ , we derive low-frequency images  $\{I_{mask}\}$  by Gradient-based masking (Sobel et al., 1968). We then build an augmented view set  $\mathcal{I}^{aug} = \{I_i, I_{mask}\}_{i=1}^N$ , extract features on all  $2N$  views, and obtain the initial 3D point set  $\mathcal{P}_0$  by running SfM once with Equation 4.

270 4.2 3D GAUSSIAN SPLATTING SELF-INITIALIZATION  
271272 While SfM reconstructs 3D points by detecting and matching keypoints across views, its reliance on  
273 local features causes failures in weakly textured or repetitive regions. To overcome this bottleneck,  
274 we introduce a novel 3DGS self-initialization method, which elevates pixel-level photometric super-  
275 vision into 3D space and constructs complementary point clouds by repurposing primitive centers  
276 as new points.277 **Light-weight 3DGS.** We train a light-weight first-pass 3DGS  $\mathcal{G}^{(0)}$  on downsampled images seeded  
278 by  $\mathcal{P}_0$ , aiming to convert dense photometric cues into additional 3D points rather than high-fidelity  
279 rendering. We use an economical parameterization and schedule: SH degree 0 (DC color) with  
280 a short optimization on downsampled low-resolution inputs. Training stops when densification  
281 plateaus or a small step budget is reached. We then form a colored point set by taking, for each Gaus-  
282 sian primitive  $g_n$ , the primitive center and its DC color, *i.e.*,  $\mathcal{P}_1 = \{(X_n, C_n)\}_{n=1}^M$  with  $X_n = \mu_n$   
283 and  $C_n = \mathbf{c}_n$ .284 285 4.3 POINT CLOUD REGULARIZATION  
286287 Before feeding the merged points into the final 3DGS optimization, we regularize the initial colored  
288 point cloud  $\mathcal{P}_{\text{init}} = \{(X_k, C_k)\}$  obtained by combining SfM points and light-weight 3DGS points,  
289  $\mathcal{P}_{\text{init}} = \mathcal{P}_0 \cup \mathcal{P}_1$ . While this union improves coverage, it also aggregates errors from both sources:  
290 (i) 3DGS-generated points with only single-view supervision, which lack geometric consistency.  
291 (ii) noisy/duplicated points introduced by 3DGS split/clone densification; and (iii) outliers from  
292 unstable two-view tracks; To obtain a reliable and uniformly distributed point set, we introduce three  
293 complementary procedures: *single-view point filtering*, *clustering-based denoising*, and *normal-  
294 based consistency filtering*. These procedures operate on disjoint criteria, so the order of application  
295 has a negligible effect on the point set. We present the pseudocode in the Appendix Sec A.3.296 **Single-view Point Filtering.** During 3DGS self-initialization, points supervised by a single view  
297 suffer from inherent depth ambiguity and are thus relatively unreliable. Nevertheless, their reliability  
298 is not uniform: among single-view-supported points, those closer to regions with two-view support  
299 are more trustworthy, as they are typically produced by densifying (splitting or cloning) points al-  
300 ready supported by multiple views. In other words, proximity to accurate two-view-supported points  
301 mitigates depth ambiguity. To balance preserving single-view information with reducing noise, we  
302 retain only the top 20% of single-view-supported points with the highest reliability. To formalize  
303 this process simply:304 Starting from the colored point cloud  $\mathcal{P}_{\text{init}} = \{X_k, C_k\}$ , We split it into single-view and multi-view-  
305 supported subsets by camera projection:

306 
$$\mathcal{P}_{\text{init}} = \mathcal{P}_{\text{sv}} \cup \mathcal{P}_{\text{mv}}. \quad (5)$$

307 For each single-view point  $X \in \mathcal{P}_{\text{sv}}$ , we assign a reliability score  $r(X)$  based on its proximity to  
308  $\mathcal{P}_{\text{mv}}$ . Finally, we retain only the top fraction (*i.e.*, 20%) of  $\mathcal{P}_{\text{sv}}$  with the highest  $r(X)$ :

309 
$$\mathcal{P}_{\text{sv}}^* = \text{Top}_{20\%}\{X \in \mathcal{P}_{\text{sv}} \mid r(X)\}. \quad (6)$$

310 The resulting point cloud is then

311 
$$\mathcal{P}_{\text{filter}} = \mathcal{P}_{\text{mv}} \cup \mathcal{P}_{\text{sv}}^*. \quad (7)$$

312 **Clustering-Based Point-Set Denoising.** To reduce the noisy points introduced from unstable 2-  
313 view SfM tracks and duplicated points from densify process in 3DGS self-initialization, we propose  
314 a denoising technique based on the clustering algorithm, which discards 70% of the points. To  
315 formalize:316 Given the single-view filtered cloud  $\mathcal{P}_{\text{filter}} = \{X_k, C_k\}$ , we apply  $K$ -means clustering with  $K =$   
317 1000 clusters. For each cluster  $c$  with centroid  $\mu_c$  and size  $|\mathcal{Q}_c|$ , we retain 30% nearest points to  $\mu_c$ .  
318 The cluster-filtered cloud is

319 
$$\mathcal{P}_{\text{clu}} = \{X_k \in \mathcal{P}_{\text{filter}} ; k \in \text{Top}_{30\%}\{\|X_k - \mu_c\|_2 : X_k \in \mathcal{Q}_c\}, c = 1, \dots, K\}. \quad (8)$$

320 **Normal-based Consistency Filtering.** Finally, we remove geometrically inconsistent points by  
321 enforcing local normal agreement in 3D space. Starting from the cluster-filtered cloud  $\mathcal{P}_{\text{clu}} =$

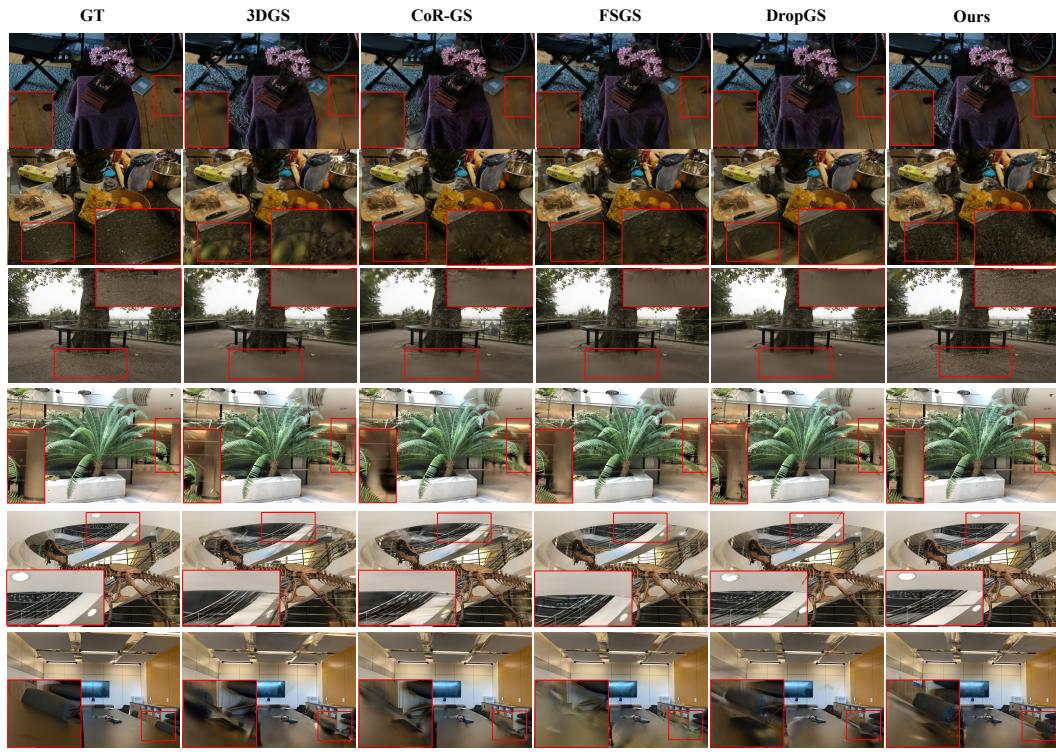


Figure 3: Qualitative comparisons on Mip-NeRF360 dataset and LLFF dataset.

$\{(X_k, C_k)\}$ , we estimate a normal  $n_k$  for each point via PCA on its 10 nearest neighbors in Euclidean space. Let  $\mathcal{N}_k$  denote the neighbor set and assume all normals are unit-length. We compute the mean cosine similarity

$$\bar{c}_k = \frac{1}{|\mathcal{N}_k|} \sum_{j \in \mathcal{N}_k} n_k^\top n_j, \quad (9)$$

and retain a point if  $\bar{c}_k \geq 0.2$ . This rejects isolated outliers and unstable estimates whose normals disagree with the local surface. The final set is

$$\mathcal{P}_{\text{final}} = \{(X_k, C_k) \in \mathcal{P}_{\text{clu}} \mid \bar{c}_k \geq 0.2\}, \quad (10)$$

which yields a clean, geometrically consistent point cloud.

## 5 EXPERIMENTS

### 5.1 EXPERIMENTAL SETTINGS

**Evaluation Datasets.** We evaluate our method on two representative benchmarks, Mip-NeRF360 (Barron et al., 2022) and LLFF (Mildenhall et al., 2019). Following the conventional sparse-view 3DGS setup (Zhu et al., 2024), we use 12 uniformly distributed input views for Mip-NeRF360 and 3 views for LLFF. The camera poses are provided by the datasets. For evaluation, every eighth image is reserved as test views, and input images are downsampled by a factor of  $4 \times$  on both datasets.

**Metrics.** For quantitative comparison, we report three widely-used metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) (Wang et al., 2004), and learned perceptual image patch similarity (LPIPS) (Zhang et al., 2018).

378

379 Table 1: Comparison on Mip-NeRF360 and LLFF datasets. We color each cell as **best**,  
 380 **second best**, and **third best**. †: We reimplemented vanilla 3DGS with Multi-view Stereo (MVS,  
 381 a native dense reconstruction stage in COLMAP following SfM) initialization, which is also de-  
 382 ployed in FSGS, CoR-GS, and DropGS in their official implementations. It brings sizeable per-  
 383 formance gains but increases the initialization cost. Reported time includes both initialization and  
 384 training, averaged over all datasets.

Methods	Mip-NeRF360 Dataset					LLFF Dataset				
	PSNR	SSIM	LPIPS	Time	Iter	PSNR	SSIM	LPIPS	Time	Iter
3DGS†	19.24	0.5743	0.3660	8m56s	5k	18.95	0.6464	0.1862	4m00s	5k
FSGS	19.25	0.5719	0.4072	12m42s	10k	19.88	0.6120	0.3400	19m35s	10k
CoR-GS	19.52	0.5580	0.4180	39m11s	30k	19.45	0.6520	0.2664	19m45s	10k
EAP-GS	19.21	0.5721	0.3072	4m27s	5k	18.84	0.6358	0.1768	5m38s	5k
DropGS	19.74	0.5770	0.3640	11m25s	10k	19.54	0.6549	0.1856	7m11s	10k
Ours	19.77	0.5892	0.3374	10m48s	5k	19.60	0.6681	0.1852	5m46s	5k
+DropGS	20.07	0.5992	0.3276	10m38s	10k	19.91	0.6835	0.1659	8m33s	10k

395

396

397 Table 2: Ablation study on the Mip-NeRF360 dataset. Each module is incrementally added to vanilla  
 398 3DGS.

Configuration	PSNR↑	SSIM↑	LPIPS↓
<b>vanilla 3DGS</b>	18.52	0.5230	0.4150
+ Low-Frequency-Aware SfM (Sec. 4.1)	19.25 <sup>↑0.73</sup>	0.5758 <sup>↑0.053</sup>	0.3575 <sup>↓0.575</sup>
+ 3DGS Self-Initialization (Sec. 4.2)	19.42 <sup>↑0.17</sup>	0.5924 <sup>↑0.017</sup>	0.3246 <sup>↓0.329</sup>
+ Single-View Point Filtering (Sec. 4.3)	19.49 <sup>↑0.07</sup>	0.5872 <sup>↓0.005</sup>	0.3363 <sup>↑0.011</sup>
+ Clustering-Based Denoising (Sec. 4.3)	19.61 <sup>↑0.12</sup>	0.5910 <sup>↓0.004</sup>	0.3356 <sup>↓0.001</sup>
+ Normal-Based Consistency Filtering (Sec. 4.3)	19.77 <sup>↑0.16</sup>	0.5892 <sup>↓0.002</sup>	0.3374 <sup>↑0.002</sup>

408

409

410

411 **Baselines.** To demonstrate the effectiveness of our initialization strategy, we compare against  
 412 several state-of-the-art 3DGS-based methods, including FSGS (Zhu et al., 2024), CoR-GS (Zhang  
 413 et al., 2024b), EAP-GS (Dai & Xing, 2025), and DropGaussian (Park et al., 2025).

414

415

416

417

418

419

420

421

422

423

424

425 **Implementation Details.** Following the official training set-ups of each baseline, we train FSGS,  
 426 CoR-GS, EAP-GS, and DropGaussian for 10k, 10k, 5k, and 10k iterations, respectively, on the  
 427 Mip-NeRF360 dataset, and for 10k, 30k, 5k, and 10k iterations on the LLFF dataset. For vanilla  
 428 3DGS and our method, we adopt 5k iterations on both benchmarks. COLMAP (Schonberger &  
 429 Frahm, 2016) is configured with the same parameters as FSGS (Zhu et al., 2024) to initialize all  
 430 baselines, except for EAP-GS and our approach, which are initialization-oriented methods. **For each**  
 431 **baseline, we report the better result between our reimplementation and the original reported**  
 432 **performance**, to ensure a fair and representative comparison. All experiments are conducted on the  
 433 same hardware with a single NVIDIA RTX 4090 GPU.

426

427

428

429

430

431

432 **5.2 PERFORMANCE EVALUATION.**

433 **Qualitative Results.** We report quantitative results on the Mip-NeRF360 and LLFF datasets in  
 434 Table 1. The experiments demonstrate that our initialization method alone already achieves state-  
 435 of-the-art overall performance across key metrics and two datasets. Moreover, when combined with  
 436 DropGS, our approach achieves further performance gains, indicating that the proposed initialization  
 437 can effectively raise the upper bound of sparse-view 3DGS. We report the total time, including both  
 438 initialization and training. Our method incurs a comparable time cost to existing baselines. The time  
 439 cost breakdown can be found in Appendix Sec A.2.

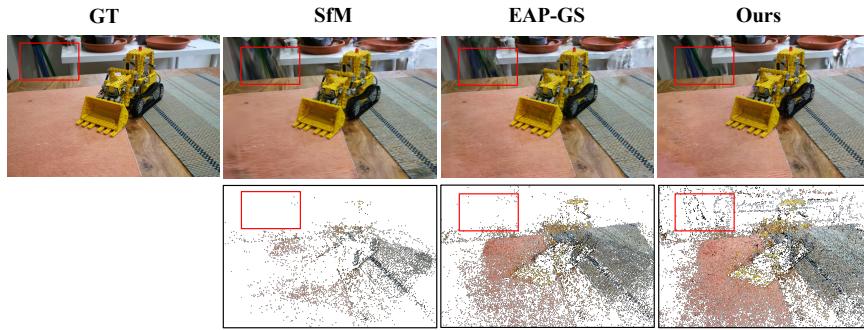


Figure 4: Comparison of initialization quality and its impact on reconstruction performance. Our method provides more accurate initialization points, especially around scene edges, leading to improved final renderings.

We further compare the performance curves of our method against baseline methods on the Mip-NeRF360 dataset, as shown in Figure 5. It demonstrates that our initialization not only achieves superior final performance but also enables faster convergence.

**Qualitative Results.** The qualitative comparisons are presented in Figure 3. Our method demonstrates consistently superior performance: it reconstructs more balanced textures in low-frequency regions (row 1), achieves robust recovery in less-featured areas (rows 2 and 3), and produces sharper object boundaries (rows 3-5) by converting pixel-level variations into reliable seed points through 3DGS self-initialization. More visualization can be found in the Appendix Sec A.2.

**Initialization v.s. Performance.** Figure 4 illustrates the relationship between final reconstruction quality and three initialization strategies: vanilla SfM, EAP-GS, and ours. Our method generates more robust initial points at scene boundaries, leading to consistently superior performance. More visualization can be found in the Appendix Sec A.2.

### 5.3 ABLATION STUDY

We conduct an ablation study to evaluate the contribution of each component in our initialization pipeline to the final performance metrics, as shown in Table 2. The proposed Low-Frequency-Aware SfM (Sec. 4.1) and 3DGS Self-Initialization (Sec. 4.2) yield the largest improvements across all metrics, as they significantly increase the number of reliable initial points, particularly in low-texture and feature-sparse regions. Furthermore, our point cloud regularization techniques (Sec. 4.3) effectively suppress noise and redundancy in the initialization, leading to higher PSNR.

## 6 CONCLUSION

We show that sparse-view 3DGS is fundamentally initialization-limited: while training-time regularization offers only modest gains, the quality of the seed point cloud determines the achievable performance. To address this, we propose a three-stage initialization pipeline, low-frequency-aware SfM, 3DGS self-initialization, and point-cloud regularization that yields cleaner, denser, and more reliable points. Experiments on LLFF and Mip-NeRF360 confirm that our method not only surpasses prior approaches but also synergizes with existing regularization techniques, setting a stronger foundation for sparse-view novel view synthesis.

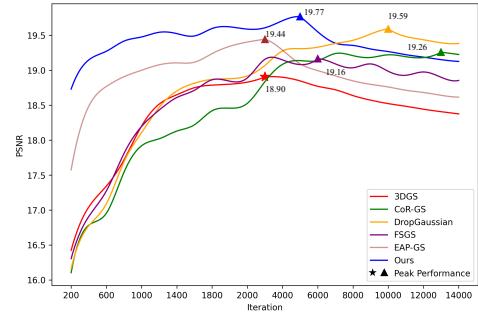


Figure 5: Performance curves between ours and baselines on the Mip-NeRF360 dataset.

486 **Ethics Statement:** This work aims to advance the development of sparse-view 3D Gaussian recon-  
 487 struction. While the method has positive academic value, potential risks include misuse in sensitive  
 488 scenarios, the inheritance and amplification of dataset biases, and environmental impact due to com-  
 489 putational cost. To mitigate these risks, we rely only on publicly available datasets, clearly document  
 490 data sources and preprocessing steps, restrict the method to academic research purposes, and adopt  
 491 efficiency-oriented experimental settings to reduce energy consumption.

492 **Reproducibility Statement:** We have taken concrete measures to ensure reproducibility: all  
 493 datasets are publicly available, preprocessing steps are described in the main text, and all hyper-  
 494 parameter settings are reported in the main text.

## 496 REFERENCES

498 Christoph Anthes, Rubén Jesús García-Hernández, Markus Wiedemann, and Dieter Kranzlmüller.  
 499 State of the art of virtual reality technology. In *2016 IEEE aerospace conference*, pp. 1–19. IEEE,  
 500 2016.

501 Zhenyu Bao, Guibiao Liao, Kaichen Zhou, Kanglin Liu, Qing Li, and Guoping Qiu. Loopsparsegs:  
 502 Loop based sparse-view friendly gaussian splatting. *IEEE Transactions on Image Processing*,  
 503 2025.

504 Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf  
 505 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference*  
 506 *on computer vision and pattern recognition*, pp. 5470–5479, 2022.

508 Pu Cao, Feng Zhou, Qing Song, and Lu Yang. Controllable generation with text-to-image diffusion  
 509 models: A survey. *arXiv preprint arXiv:2403.04279*, 2024.

511 Eric R Chan, Koki Nagano, Matthew A Chan, Alexander W Bergman, Jeong Joon Park, Axel Levy,  
 512 Miika Aittala, Shalini De Mello, Tero Karras, and Gordon Wetzstein. Generative novel view syn-  
 513 thesis with 3d-aware diffusion models. In *Proceedings of the IEEE/CVF International Conference*  
 514 *on Computer Vision*, pp. 4217–4229, 2023.

515 Yuedong Chen, Haofei Xu, Chuanxia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-  
 516 Jen Cham, and Jianfei Cai. Mvsplat: Efficient 3d gaussian splatting from sparse multi-view  
 517 images. In *European Conference on Computer Vision*, pp. 370–386. Springer, 2024a.

518 Yuedong Chen, Chuanxia Zheng, Haofei Xu, Bohan Zhuang, Andrea Vedaldi, Tat-Jen Cham, and  
 519 Jianfei Cai. Mvsplat360: Feed-forward 360 scene synthesis from sparse views. *Advances in*  
 520 *Neural Information Processing Systems*, 37:107064–107086, 2024b.

522 Jaeyoung Chung, Jeongtaek Oh, and Kyoung Mu Lee. Depth-regularized optimization for 3d gaus-  
 523 sian splatting in few-shot images. In *Proceedings of the IEEE/CVF Conference on Computer*  
 524 *Vision and Pattern Recognition*, pp. 811–820, 2024.

525 Dongrui Dai and Yuxiang Xing. Eap-gs: Efficient augmentation of pointcloud for 3d gaussian  
 526 splatting in few-shot scene reconstruction. In *Proceedings of the Computer Vision and Pattern*  
 527 *Recognition Conference*, pp. 16498–16507, 2025.

529 Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views  
 530 and faster training for free. In *Proceedings of the IEEE/CVF conference on computer vision and*  
 531 *pattern recognition*, pp. 12882–12891, 2022.

532 Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, and Steven M Seitz. Multi-view  
 533 stereo for community photo collections. In *2007 IEEE 11th international conference on computer*  
 534 *vision*, pp. 1–8. IEEE, 2007.

535 Youngkyoon Jang and Eduardo Pérez-Pellitero. Comapgs: Covisibility map-based gaussian splat-  
 536 ting for sparse novel view synthesis. In *Proceedings of the Computer Vision and Pattern Recog-*  
 537 *nition Conference*, pp. 26779–26788, 2025.

539 Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splat-  
 540 ting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

540 Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In *Proceedings*  
 541 *of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20775–20785, 2024.  
 542

543 Xin Lin, Shi Luo, Xiaojun Shan, Xiaoyu Zhou, Chao Ren, Lu Qi, Ming-Hsuan Yang, and Nuno  
 544 Vasconcelos. Hqgs: High-quality novel view synthesis with gaussian splatting in degraded scenes.  
 545 In *The Thirteenth International Conference on Learning Representations*.

546 Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick.  
 547 Zero-1-to-3: Zero-shot one image to 3d object. In *Proceedings of the IEEE/CVF international*  
 548 *conference on computer vision*, pp. 9298–9309, 2023.

549 Tianqi Liu, Guangcong Wang, Shoukang Hu, Liao Shen, Xinyi Ye, Yuhang Zang, Zhiguo Cao,  
 550 Wei Li, and Ziwei Liu. Mvsgaussian: Fast generalizable gaussian splatting reconstruction from  
 551 multi-view stereo. In *European Conference on Computer Vision*, pp. 37–53. Springer, 2024.

552 Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ra-  
 553 mamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with  
 554 prescriptive sampling guidelines. *ACM Transactions on Graphics (ToG)*, 38(4):1–14, 2019.

555 Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and  
 556 Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications*  
 557 *of the ACM*, 65(1):99–106, 2021.

558 Niluthpol Chowdhury Mithun, Tuan Pham, Qiao Wang, Ben Southall, Kshitij Minhas, Bogdan  
 559 Matei, Stephan Mandt, Supun Samarasekera, and Rakesh Kumar. Diffusion-guided gaussian  
 560 splatting for large-scale unconstrained 3d reconstruction and novel view synthesis. *arXiv preprint*  
 561 *arXiv:2504.01960*, 2025.

562 Avinash Paliwal, Wei Ye, Jinhui Xiong, Dmytro Kotovenko, Rakesh Ranjan, Vikas Chandra, and  
 563 Nima Khademi Kalantari. Coherentgts: Sparse novel view synthesis with coherent 3d gaussians.  
 564 In *European Conference on Computer Vision*, pp. 19–37. Springer, 2024.

565 Hyunwoo Park, Gun Ryu, and Wonjun Kim. Dropgaussian: Structural regularization for sparse-view  
 566 gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*,  
 567 pp. 21600–21609, 2025.

568 Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings*  
 569 *of the IEEE conference on computer vision and pattern recognition*, pp. 4104–4113, 2016.

570 Irwin Sobel, Gary Feldman, et al. A 3x3 isotropic gradient operator for image processing. *a talk at*  
 571 *the Stanford Artificial Project in*, 1968:271–272, 1968.

572 Stanislaw Szymanowicz, Chrisitian Rupprecht, and Andrea Vedaldi. Splatter image: Ultra-fast  
 573 single-view 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision*  
 574 *and pattern recognition*, pp. 10208–10217, 2024.

575 Jiaxiang Tang, Zhaoxi Chen, Xiaokang Chen, Tengfei Wang, Gang Zeng, and Ziwei Liu. Lgm:  
 576 Large multi-view gaussian model for high-resolution 3d content creation. In *European Conference*  
 577 *on Computer Vision*, pp. 1–18. Springer, 2024.

578 Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Wang Yifan,  
 579 Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Ad-  
 580 vances in neural rendering. In *Computer Graphics Forum*, volume 41, pp. 703–735. Wiley Online  
 581 Library, 2022.

582 Shimon Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of*  
 583 *London. Series B. Biological Sciences*, 203(1153):405–426, 1979.

584 Vikram Voleti, Chun-Han Yao, Mark Boss, Adam Letts, David Pankratz, Dmitry Tochilkin, Chris-  
 585 tian Laforte, Robin Rombach, and Varun Jampani. Sv3d: Novel multi-view synthesis and 3d  
 586 generation from a single image using latent video diffusion. In *European Conference on Com-*  
 587 *puter Vision*, pp. 439–457. Springer, 2024.

594 Yecong Wan, Mingwen Shao, Yuanshuo Cheng, and Wangmeng Zuo. S2gaussian: Sparse-view  
 595 super-resolution 3d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recog-*  
 596 *nition Conference*, pp. 711–721, 2025.

597  
 598 Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David  
 599 Novotny. Vggt: Visual geometry grounded transformer. In *Proceedings of the Computer Vision*  
 600 *and Pattern Recognition Conference*, pp. 5294–5306, 2025.

601 Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment:  
 602 from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–  
 603 612, 2004.

604  
 605 Haolin Xiong, Sairisheek Muttukuru, Rishi Upadhyay, Pradyumna Chari, and Achuta Kadambi.  
 606 Sparsegs: Real-time 360  $\{\backslash\deg\}$  sparse view synthesis using gaussian splatting. *arXiv preprint*  
 607 *arXiv:2312.00206*, 2023.

608  
 609 Yexing Xu, Longguang Wang, Minglin Chen, Sheng Ao, Li Li, and Yulan Guo. Dropoutgs: Drop-  
 610 ping out gaussians for better sparse-view rendering. In *Proceedings of the Computer Vision and*  
 611 *Pattern Recognition Conference*, pp. 701–710, 2025.

612  
 613 Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, and Eric Xing. Fregs: 3d gaussian splat-  
 614 ting with progressive frequency regularization. In *Proceedings of the IEEE/CVF Conference on*  
 615 *Computer Vision and Pattern Recognition*, pp. 21424–21433, 2024a.

616  
 617 Jiawei Zhang, Jiahe Li, Xiaohan Yu, Lei Huang, Lin Gu, Jin Zheng, and Xiao Bai. Cor-gs: sparse-  
 618 view 3d gaussian splatting via co-regularization. In *European Conference on Computer Vision*,  
 619 pp. 335–352. Springer, 2024b.

620  
 621 Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable  
 622 effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on*  
 623 *computer vision and pattern recognition*, pp. 586–595, 2018.

624  
 625 Yulong Zheng, Zicheng Jiang, Shengfeng He, Yandu Sun, Junyu Dong, Huaidong Zhang, and Yong  
 626 Du. Nexusgs: Sparse view synthesis with epipolar depth priors in 3d gaussian splatting. In *Pro-*  
 627 *ceedings of the Computer Vision and Pattern Recognition Conference*, pp. 26800–26809, 2025.

628  
 629 Boyao Zhou, Shunyuan Zheng, Hanzhang Tu, Ruizhi Shao, Boning Liu, Shengping Zhang, Liqiang  
 630 Nie, and Yebin Liu. Gps-gaussian+: Generalizable pixel-wise 3d gaussian splatting for real-time  
 631 human-scene rendering from sparse views. *IEEE Transactions on Pattern Analysis and Machine*  
 632 *Intelligence*, 2025.

633  
 634 Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. Fsgs: Real-time few-shot view synthe-  
 635 sis using gaussian splatting. In *European conference on computer vision*, pp. 145–163. Springer,  
 636 2024.

637  
 638  
 639  
 640  
 641  
 642  
 643  
 644  
 645  
 646  
 647

648 A APPENDIX  
649650 A.1 THE USE OF LARGE LANGUAGE MODELS  
651652 We used large language models to assist in code refinement and in polishing the writing of this  
653 manuscript.654  
655 A.2 ADDITIONAL COMPARISONS  
656657 **Performance Visualization.** We present additional visualization comparisons on Mip-NeRF360  
658 dataset and LLFF datasets in Figure 6 and Figure 7, respectively.659 **Initialization Performance Visualization.** We present additional initialization performance com-  
660 parisons on Mip-NeRF360 dataset and LLFF datasets in Figure 8 and Figure 9, respectively.  
661662 **Time Cost Breakdown.** The detailed time cost between our method and baselines is presented in  
663 Table 3.664  
665 Table 3: Time Cost Breakdown.  
666

Methods	Mip-NeRF360 Dataset			LLFF Dataset		
	Init.	Training	Total	Init.	Training	Total
3D-GS	5m04s	3m52s	8m56s	1m32s	2m28s	4m00s
FSGS	5m04s	7m38s	12m42s	1m32s	18m03s	19m35s
CoR-GS	5m04s	34m07s	39m11s	1m32s	18m13s	19m45s
EAP-GS	2m15s	4m27s	6m42s	2m23s	3m15s	5m38s
DropGS	5m04s	6m21s	11m25s	1m32s	5m39s	7m11s
Ours	6m06s	4m42s	10m48s	4m16s	1m30s	5m46s
Ours+DropGS	6m06s	4m32s	10m38s	4m16s	4m17s	8m33s

677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701

702 A.3 THE PSEUDOCODE OF SEC. 4.3.  
 703

---

704 **Algorithm 1** Point Cloud Regularization

---

705 1: **Input:** Initial colored point cloud  $\mathcal{P}_{\text{init}} = \mathcal{P}_0 \cup \mathcal{P}_1$   
 706 2: **Output:** Regularized point cloud  $\mathcal{P}_{\text{final}}$   
 707 3: **Step 1: Single-view Point Filtering:**  
 708 4: Split  $\mathcal{P}_{\text{init}}$  into  $\mathcal{P}_{\text{sv}}$  (single-view) and  $\mathcal{P}_{\text{mv}}$  (multi-view)  
 709 5: **for** each  $X \in \mathcal{P}_{\text{sv}}$  **do**  
 710 6:     Compute reliability score  $r(X)$  w.r.t. proximity to  $\mathcal{P}_{\text{mv}}$   
 711 7: **end for**  
 712 8: Retain top-20% of  $\mathcal{P}_{\text{sv}}$  ranked by  $r(X)$ , denoted  $\mathcal{P}_{\text{sv}}^*$   
 713 9:  $\mathcal{P}_{\text{filter}} \leftarrow \mathcal{P}_{\text{mv}} \cup \mathcal{P}_{\text{sv}}^*$   
 714 10: **Step 2: Clustering-based Denoising:**  
 715 11: Apply  $K$ -means clustering ( $K = 1000$ ) on  $\mathcal{P}_{\text{filter}}$   
 716 12: **for** each cluster  $c$  with centroid  $\mu_c$  and points  $\mathcal{Q}_c$  **do**  
 717 13:     Retain nearest-30% points to  $\mu_c$   
 718 14: **end for**  
 719 15:  $\mathcal{P}_{\text{clu}} \leftarrow \bigcup_c$  (retained points from  $\mathcal{Q}_c$ )  
 720 16: **Step 3: Normal-based Consistency Filtering:**  
 721 17: **for** each  $X \in \mathcal{P}_{\text{clu}}$  **do**  
 722 18:     Estimate surface normal  $n(X)$   
 723 19:     **if** angular deviation from neighbors  $> \tau$  **then**  
 724 20:         Discard  $X$   
 725 21:     **end if**  
 726 22: **end for**  
 23:  $\mathcal{P}_{\text{final}} \leftarrow$  remaining points

---

727  
 728  
 729  
 730  
 731  
 732  
 733  
 734  
 735  
 736  
 737  
 738  
 739  
 740  
 741  
 742  
 743  
 744  
 745  
 746  
 747  
 748  
 749  
 750  
 751  
 752  
 753  
 754  
 755

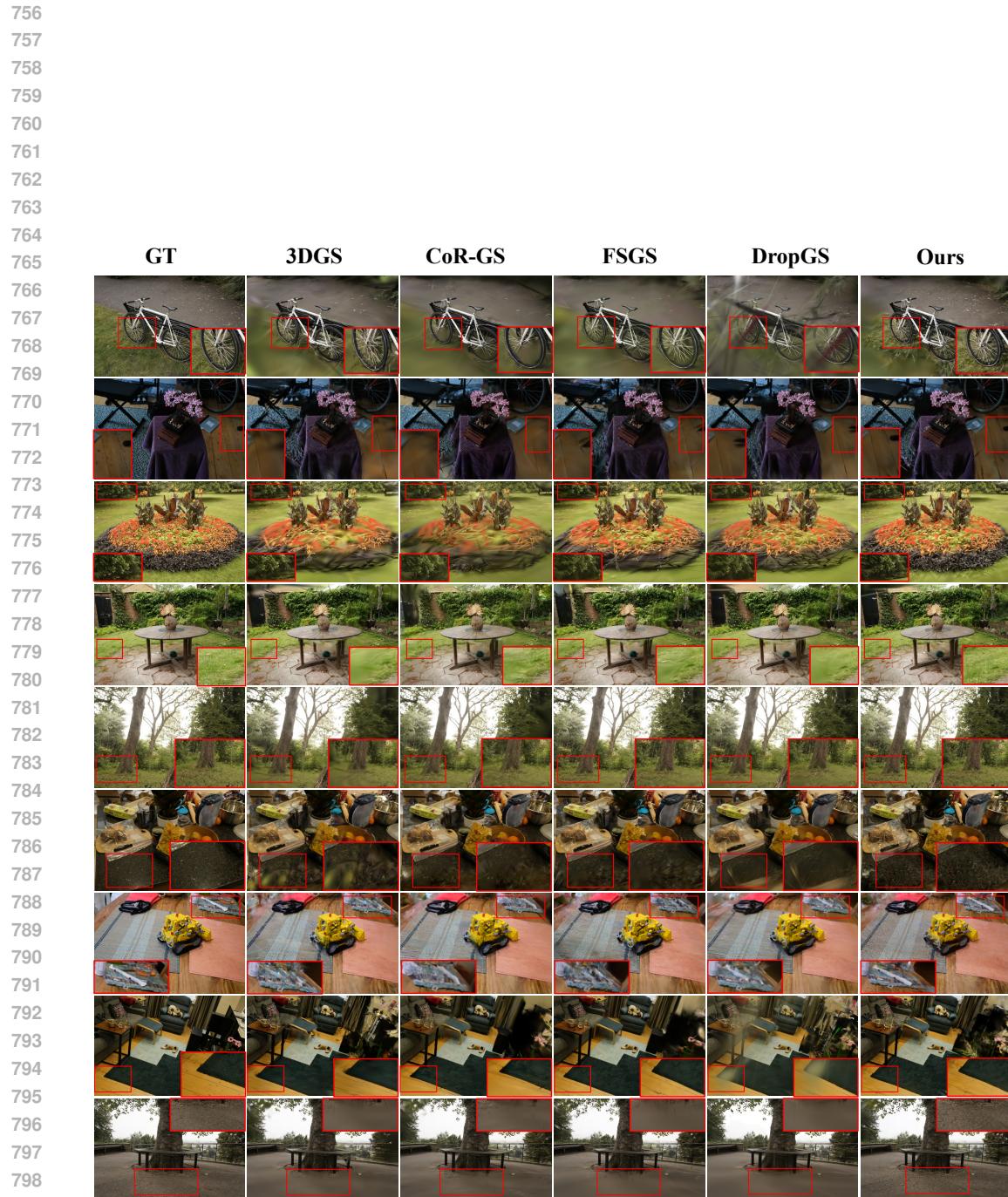


Figure 6: Qualitative comparisons on Mip-NeRF360 dataset.

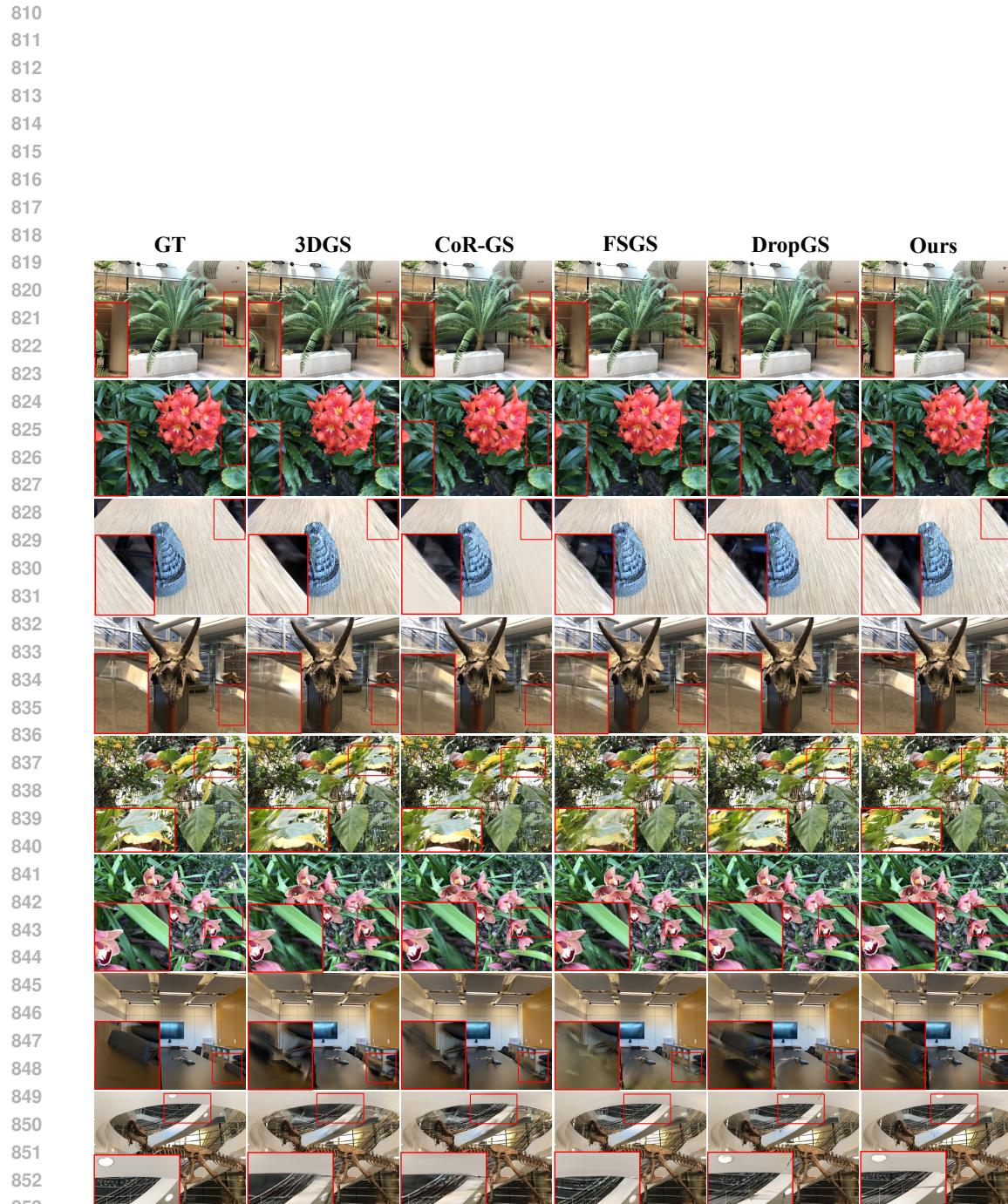


Figure 7: Qualitative comparisons on LLFF dataset.

854  
855  
856  
857  
858  
859  
860  
861  
862  
863

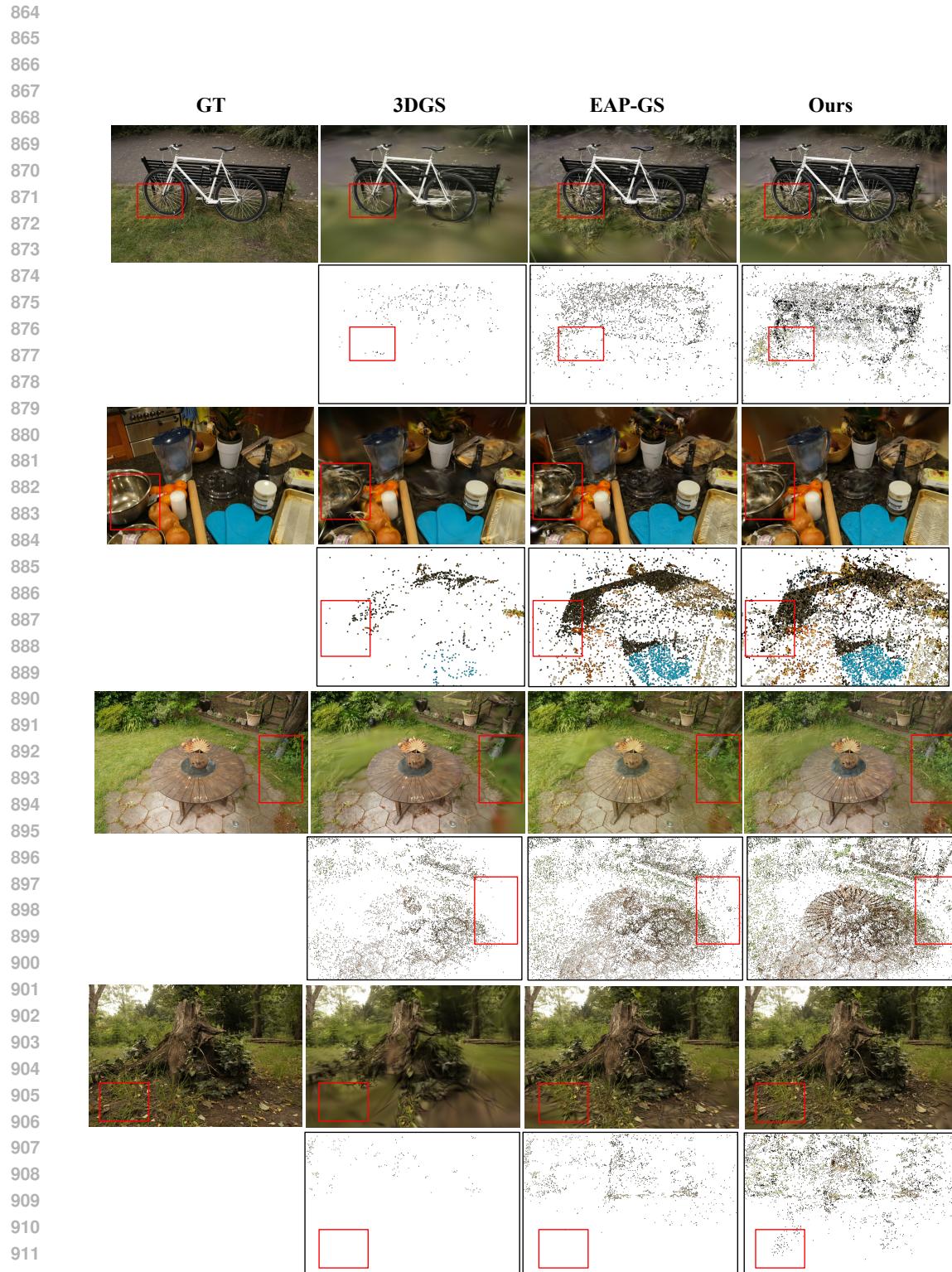


Figure 8: Initialization comparisons on Mip-NeRF360 dataset.

913  
914  
915  
916  
917

