EAP-GS: Efficient Augmentation of Pointcloud for 3D Gaussian Splatting in Few-shot Scene Reconstruction

Dongrui Dai[1,2] Yuxiang Xing[1,2]
[1]Department of Engineering Physics, Tsinghua University
[2]Key Laboratory of Particle and Radiation Imaging (Tsinghua University), Ministry of Education
ddr23@mails.tsinghua.edu.cn xingyx@mail.tsinghua.edu.cn

(a) FSGS (b) EAP-GS (Ours)

Figure 1. Comparison of the FSGS [41] and our proposed EAP-GS with 12 training views. With Attentional Pointcloud Augmentation technique, our method generates significantly more valuable points in the initialization stage than the traditional COLMAP method, especially in regions with weakly pronounced texture and peripheral area, which ultimately improves the reconstruction quality effectively.

Abstract

*3D Gaussian splatting (3DGS) has shown impressive performance in 3D scene reconstruction. However, it suffers from severe degradation when the number of training views is limited, resulting in blur and floaters. Many works have been devoted to standardize the optimization process of 3DGS through regularization techniques. However, we identify that inadequate initialization is a critical issue overlooked by current studies. To address this, we propose **EAP-GS**, a method to enhance initialization for fast, accurate, and stable few-shot scene reconstruction. Specifically, we introduce an Attentional Pointcloud Augmentation (APA) technique, which retains two-view tracks as an option for pointcloud generation. Additionally, the scene complexity is used to determine the required density distribution, thereby constructing a better pointcloud. We implemented APA by extending Structure-From-Motion (SFM) to focus on pointcloud generation in regions with complex structure but sparse pointcloud distribution, which significantly increases the number of valuable points and effectively harmonizes the density distribution. A better pointcloud leads*

to more accurate scene geometry and mitigates local overfitting during reconstruction stage. Furthermore, our APA can be framed as a modular augmentation to existing methods with minimal overhead. Experimental results from various indoor and outdoor scenes demonstrate that the proposed EAP-GS achieves outstanding scene reconstruction performance and surpasses state-of-the-art methods.

1. Introduction

In recent years, 3D scene representation has emerged as a cutting-edge research field in computer vision. Techniques such as Neural Radiance Field (NeRF) [22] and 3D Gaussian Splatting (3DGS) [14] propose novel representation methods and utilize rendering equation supervised by 2D images, enabling real-world scene reconstruction. These methods have been widely applied in downstream tasks such as autonomous driving [19, 33, 40], 3D generation [18, 29, 30] and medical imaging [4, 16, 23, 34]. However, due to the slow training and inference speed of NeRF-based methods, 3DGS stands out for its rapid reconstruction speed and real-time rendering capabilities, while still delivering high-quality results.

In practice, a sufficient number of images are often difficult to obtain due to various limitations. Under few-shot conditions, the scarcity of scene information causes the reconstruction process to converge toward local optima, resulting in artifacts like "floaters" and degradation in 3DGS performance. This issue arises because 3DGS is an unstructured representation [26]. With a lack of coherence between Gaussians , their attributes can only be optimized individually via image supervision. In few-shot scenarios, this leads to local overfitting, as the geometric information provided by the views is insufficient. Although some regularizations have been explored [5, 32, 38, 41], the results remain unsatisfactory.

While 3DGS can reconstruct simple objects well with random initialization [14], we found it exhibits a strong dependence on the quality of initialized points for complex scene. In other words, **inadequate initialization** is a key factor contributing to the deficient performance in few-shot optimization. 3DGS relies on pointcloud and camera poses that usually generated by Structure-from-Motion (SfM) [31] method, particularly COLMAP [28]. When constructing the pointcloud, COLMAP selects matched 2D feature points from different views (typically $\geq$ 3 views) and uses triangulation to estimate 3D point coordinates. While this mechanism ensures the position accuracy of 3D points, it filters out significant amounts of information (e.g., points visible in only 2 views) in a few-shot case. Since points serve as carriers for Gaussians, inadequate pointcloud leads to a lot of missing details, often generating blur and floaters in the corresponding regions, thereby degrade the quality of 3D reconstruction, as shown in Fig. 2. To address this, we propose an easy-to-implement attentional pointcloud augmentation technique to improve the accuracy of 3DGS reconstruction.

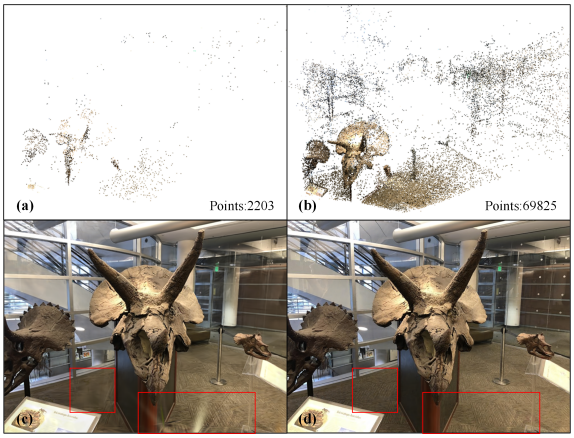

Figure 2. Reconstruction results by original 3DGS with six training views. Left column: unsatisfying reconstruciton (c) from inadequate pointcloud (a). Right column: fine reconstruciton (d) from adequate pointcloud (b).

In this work, we propose our 3D Gaussian splatting with efficiently augmented pointcloud, **EAP-GS**: a 3DGS-based method that incorporates optimization in the initialization stage, aiming at improving few-shot scene reconstruction. Specifically, we present a pointcloud augmentation technique, resulting in multiple times increase in the number of initialized points compared to traditional COLMAP method, which greatly enhances the scene reconstruction quality. Additionally, the introduction of the attention mechanism can effectively harmonize the density distribution of pointcloud and avoid overfitting in dense regions. Our experimental results on the LLFF dataset [21] and Mip-NeRF360 dataset [1] demonstrate that our approach achieves greater improvements in few-shot scenarios than other 3DGS-based methods, particularly in regions with weakly pronounced texture and peripheral area. Furthermore, we show that initialization quality has a critical impact on the optimization process of 3DGS for few-shot scene reconstruction.

In summary, our main contributions are as follows:

- A key insight that inadequate initialization can lead to poor performance in few-shot optimization, which is currently not well explored in the field from our knowledge.
- An Attentional Pointcloud Augmentation technique proposed in the initialization stage, that increases the number of valuable points and harmonizes the density distribution of pointcloud.
- A novel SfM-based initialization that can be easily implemented in other 3DGS-based methods, improving scene reconstruction quality in few-shot scenarios.

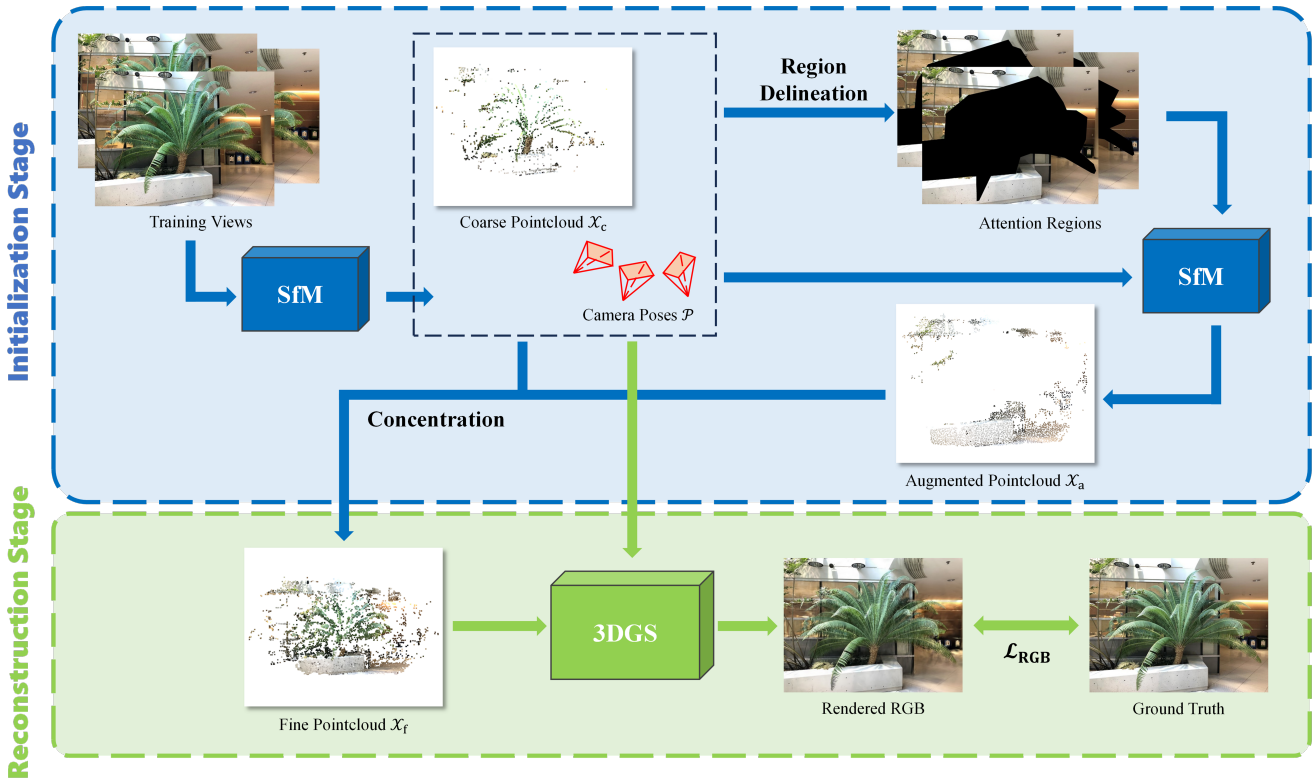

Figure 3. Pipeline of the EAP-GS. we utilize the original 3DGS in the reconstruction stage and it can be easily replaced by other optimization methods. Our main innovations are in the initialization stage, an Attentional Pointcloud Augmentation (APA) technique is adapted by retaining two-view tracks as an option in attention regions for pointcloud generation, which dramatically increases the number of valuable points and effectively harmonizes the density distribution. A better pointcloud leads to more stable and accurate reconstruction.

2. Related Work

2.1. 3D Reconstruction

In recent years, 3D reconstruction has made significant progress within the field of computer vision [4, 16, 29, 30, 33, 34, 40]. Based on the task, it can be broadly classified into object reconstruction [12, 16, 22, 29] and scene reconstruction [20, 24, 33, 40]. Object reconstruction is relatively straightforward and has achieved nice results [12, 29]. Scene reconstruction is more complex and closer to real-world applications, particularly for large, unbounded scenes [1, 33, 40]. Existing methods often struggle to achieve both fast and accurate reconstruction in such complex environments.

In terms of representation, 3D reconstruction methods can be divided into implicit and explicit representation. NeRF [22] is a typical example of Implicit Neural Representation (INR), where scene details are embedded within a neural network, offering high resolution and compactness. On the other hand, explicit representations, such as pointcloud [11, 28], voxels [7], and 3D Gaussians [14], provide direct geometric representations. Among these, 3DGS stands out for its flexibility and low memory requirements. Moreover, 3DGS offers the advantages of real-time rendering and strong editability, making it a leading technology in the field of 3D reconstruction.

2.2. Few-shot Scene Reconstruction

3D reconstruction algorithms require a large number of images to produce an accurate and plausible scene representations, which hinder their practical applications. To address this, recent researches have focused on few-shot scene reconstruction, particularly for scene-level reconstruction [6, 15, 16, 26, 41]. FSGS [41] is the first method to employ 3DGS for few-shot scene reconstruction from our knowledge, introducing a Gaussian unpooling strategy to adjust the density of Gaussian distribution and enhance scene representation. CoR-GS [36] utilizes the randomness of densification implementation as an unsupervision for predicting reconstruction quality. DNGaussian [15] incorporates depth values estimated by a pretrained monocular depth model [27] to correct scene geometry. Further, DRGS [6] obtains absolute depths by combining relative depths from a monocular depth model [2] and pointcloud depths to each

view. CoherentGS [26] deploys a learnable implicit decoder to impose coherence, enabling sparse and local regularization constraints to propagate across Gaussians.

2.3. 3DGS Initialization Improvement

Most existing works focus on enhancing performance of 3DGS by introducing regularization [35, 38, 41] or integrating pretrained models [6, 17, 39, 42]. Very few attentions have been given to improving the initialization process. Most 3DGS-based optimization methods use SfM to initialize pointcloud. Some approaches like RAIN-GS [13] and InstantSplat [9] focus on jointly optimizing camera poses during the reconstruction and reduce the impact of initialization. RadSplat [25] points out the importance of initialization and adopts NeRF as a prior to establish a better pointcloud, but it significantly increases time cost.

In this work, we start with SfM-based initialization to obtain pointcloud and camera poses. On top of that, we introduce an Attentional Pointcloud Augmentation technique to improve the initialization by significantly increasing the number of 3D points that is properly distributed. The higher-quality pointcloud effectively improves the reconstruction of 3DGS.

3. Method

Based on multi-view pointcloud generation mechanism of most SfM methods [3, 10, 11, 28], which requires 2D feature points from at least 3 views to determine a 3D point. While this mechanism yields accurate pointcloud, it did not explore view information sufficiently, thus not fit the optimization process of 3DGS in few-shot cases. **Our main idea** is to obtain more 3D points by retaining two-view tracks as an option for pointcloud generation, i.e. extracts 3D points that appear in only 2 views. We recognize that these points contain rich structural information although prone to slightly bigger triangulation error than multi-view tracks. This triangulation error is tolerable in the optimization of 3DGS. The pipeline of EAP-GS is shown in Fig. 3.

In Sec. 3.1, we briefly review the basic principles of 3DGS, which lay the foundation for constructing augmented pointcloud. In Sec. 3.2, we present an Attentional Pointcloud Augmentation technique to effectively increase the number of initial points and harmonize the overall pointcloud density distribution of the scene.

3.1. Preliminary

3DGS is a novel paradigm for explicit scene representation. Given a pointcloud generated by SfM, each point is modeled as an anisotropic 3D Gaussian:

$$G(\mathbf{x}) = \alpha e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{1}$$

which is characterized by: the mean value $\boldsymbol{\mu} \in \mathbb{R}^3$ denotes the position of the Gaussian; the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times3}$ denotes its size and shape; the spherical harmonic function coefficients $\mathbf{c} \in \mathbb{R}^k$ denotes the color with k determined by the order of the function; and the $\alpha \in \mathbb{R}$ denotes the opacity.

The optimization process involves splatting 3D Gaussian into the image domain, sorting the $\mathcal{N}$ 2D Gaussians on the pixel by depth, and then calculating the final color $\mathbf{C}$ via α-blending:

$$\mathbf{C} = \sum_{i\in\mathcal{N}} \mathbf{c}_i \alpha'_i \prod_{j=1}^{i-1}(1-\alpha'_j) \tag{2}$$

$$\alpha'_i = \alpha_i e^{-\frac{1}{2}(\mathbf{x}'-\boldsymbol{\mu}'_i)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}_{2\mathrm{D},i}(\mathbf{x}'-\boldsymbol{\mu}'_i)} \tag{3}$$

where $\boldsymbol{\Sigma}_{2\mathrm{D}}$ is the 2D covariance from the projection of the 3D covariance $\boldsymbol{\Sigma}$ to the image domain by:

$$\boldsymbol{\Sigma}_{2\mathrm{D}} = \mathbf{JW}\boldsymbol{\Sigma}\mathbf{W}^{\mathrm{T}}\mathbf{J}^{\mathrm{T}} \tag{4}$$

with $\mathbf{J}$ the Jacobian of the affine approximation of the projective transformation and $\mathbf{W}$ the view transformation matrix.

For each image rendering, the loss function relative to the ground truth (GT) can be computed directly as:

$$\mathcal{L}_{\mathrm{RGB}} = (1-\lambda_{\mathrm{RGB}})\mathcal{L}_1 + \lambda_{\mathrm{RGB}}\mathcal{L}_{\mathrm{D-SSIM}} \tag{5}$$

where λ_{RGB} is a hyperparameter.

3.2. Attentional Pointcloud Augmentation

In this section, we elaborate the idea of pointcloud augmentation by retaining two-view tracks as an option. It is important to note that without sufficient supervised views to provide constraints, simply using this 3D feature point generation mechanism may degrade reconstruction results in certain regions of the scene. Usually, complex regions in a scene contain rich high-frequency components, and there is a strong correlation between the required pointcloud density distribution and high-frequency components for fine reconstruction. As illustrated in Fig. 4, denser points are needed for the region with textures than for the smooth region to obtain fine reconstruction. Since the error of points from two-view tracks could have slightly bigger than results from multi-view tracks, introducing these points in areas where the high-frequency components match density distribution will increase the overall reconstruction error. However, in rich high-frequency components but sparse density distribution regions, these points could reduce the overall error to a limited extent. Therefore, based on the density distribution, we attentionally relax the view number of local tracks.

Initially, for n views, after feature matching, an initial set of estimated camera poses $\mathcal{P} = \{\mathbf{P}_i \in \mathbb{R}^{4\times4} | i = 1, \ldots, n\}$ can be obtained. The input to reconstruction stage consists of the n scene views $\mathcal{I} = \{\mathbf{I}_i \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}} | i = 1, \ldots, n\}$ and

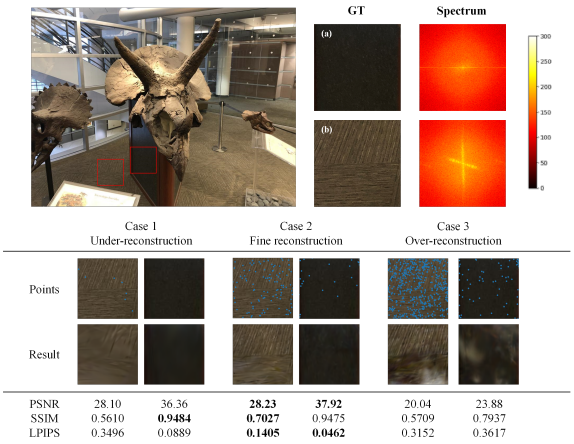

Figure 4. Relationship between scene complexity and density distribution. (a) zoom-in of a smooth region with its 2D spectrum on the right; (b) zoom-in of a region with weakly pronounced texture with its 2D spectrum on the right. The blue points in the second row of the table represent the projection of initialized pointcloud onto the testing view. The testing view results corresponding to over-sparse, properly distributed, over-dense points are demonstrated using the same five training views.

their corresponding estimated camera poses $\mathcal{P}$. SfM generates a coarse pointcloud $\mathcal{X}_{\mathrm{c}} = \{\mathbf{X}_i \in \mathbb{R}^3 | i = 1, \ldots, m\}$ from two different views triangulation $\mathcal{T}$ [10]:

$$\mathbf{X} = \mathcal{T}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{P}_i, \mathbf{P}_j) \tag{6}$$

where $\mathbf{x}_i \in \mathbb{R}^2$ is a 2D feature point of view $\mathbf{I}_i$.

After a new image registration, bundle adjustment is performed to refine the parameters of camera pose $\mathbf{P}_i$ and 3D point $\mathbf{X}$ to minimizes the reprojection error and filter observations with large errors:

$$\mathbf{X}^*, \mathbf{P}_i^* = \underset{\mathbf{X}, \mathbf{P}_i}{\arg\min} \left\{ \sum_i^{i \geq 3} \rho \| \pi(\mathbf{P}_i, \mathbf{X}) - \mathbf{x}_i \|_2^2 \right\} \tag{7}$$

where π is the projection function mapping scene points to the image domain. ρ is the threshold function used to filter observations and mitigate the effect of outliers on the reconstruction. We update pose $\mathbf{P}_i^*$ to $\mathcal{P}$ and add new point $\mathbf{X}^*$ to $\mathcal{X}_{\mathrm{c}}$, followed by re-triangulation to keep track of points that previously failed to triangulate.

Bundle adjustment and triangulation are performed alternately until all images are registered. Ultimately, the optimal camera poses $\mathcal{P}^*$ and coarse pointcloud $\mathcal{X}_{\mathrm{c}}$ are obtained.

Due to the limited view number of tracks, the point density in peripheral areas tends to be low, and regions with weakly pronounced texture may also suffer from low point density because of failing to triangulate in multi-view tracks. We pay more attention to relaxing the view number of local tracks in these regions. Specifically, based on the pointcloud density distribution in the view $\mathbf{I}_i$, we delineate an attention region $\mathbf{M}_i$:

$$\mathbf{M}_i = \mathbf{I}_i \setminus \mathcal{C}(\pi(\mathbf{P}_i, \mathcal{X}_{\mathrm{c}})) \tag{8}$$

where $\mathcal{C}$ is a function to delineate area based on density distribution (in our case, using DBSCAN Clustering [8]).We extract 2D feature points $\mathbf{x}_i'$ only within $\mathbf{M}_i$, i.e. $\mathbf{x}_i' \in \mathbf{M}_i$.

Now, for the attention regions $\mathcal{M} = \{\mathbf{M}_i \in \mathbb{R}^{\mathrm{H} \times \mathrm{W}} | i = 1, \ldots, n\}$, we construct another augmented pointcloud $\mathcal{X}_{\mathrm{a}}$ using the previously estimated camera poses $\mathcal{P}$, while retaining the two-view tracks during the construction:

$$\mathbf{X}' = \mathcal{T}(\mathbf{x}_i', \mathbf{x}_j', \mathbf{P}_i^*, \mathbf{P}_j^*) \tag{9}$$

$$\mathbf{X}'^* = \underset{\mathbf{X}'}{\arg\min} \left\{ \sum_i^{i \geq 2} \rho \| \pi(\mathbf{P}_i^*, \mathbf{X}') - \mathbf{x}_i' \|_2^2 \right\} \tag{10}$$

Adding $\mathbf{X}'^*$ to $\mathcal{X}_{\mathrm{a}}$, similar to the previous process, results in the augmented pointcloud $\mathcal{X}_{\mathrm{a}}$. 3DGS can leverage $\mathcal{X}_a$ to generate preliminary scene geometry in the early stage of reconstruction and correct some error introduced by two-view tracks. Finally, we combine $\mathcal{X}_{\mathrm{c}}$ and $\mathcal{X}_{\mathrm{a}}$ to construct the fine pointcloud $\mathcal{X}_{\mathrm{f}}$:

$$\mathcal{X}_{\mathrm{f}} = \mathcal{X}_{\mathrm{c}} \cup \mathcal{X}_{\mathrm{a}} \tag{11}$$

The resulted pointcloud $\mathcal{X}_{\mathrm{f}}$ contains a higher number of points with a better distribution, making it more beneficial for subsequent reconstruction.

In summary, pointcloud can be directly used for practical tasks or for 3DGS initialization, but the accuracy requirements for the points differ in these two types of applications. Existing methods generate pointcloud primarily for practical tasks. When used for 3DGS initialization, some points with lower matching accuracy are discarded due to stricter accuracy requirements, yet these points often contain valuable information. Discarding them can lead to incomplete exploration of view information, resulting in insufficient initialization, especially in few-shot cases. Therefore, we propose a pointcloud generation method specifically designed for 3DGS initialization, which significantly increases the number of initial points. A denser initialization helps reconstruct high-quality details, rather than forcing sparse Gaussians to stretch and fill the volume, which results in blurrier results.

It should be emphasized that this idea can also be easily realized with any existing SfM method. In this work, we implement our algorithm based on DetectorfreeSfM [11], which leverages a detector-free matcher to enhance feature extraction in texture-poor scenarios. This idea can also be

Figure 5. Qualitative Comparison on LLFF and Mip-NeRF360 datasets. We demonstrate testing view reconstruction by 3DGS [14], DRGS [6], FSGS [41], CoR-GS [36] and our method for comparison. The results of the rest scenes can be viewed in the supplementary material.

easily applied to other 3DGS-based optimization methods with minimal overhead, as it only focuses on improvement in the initialization stage.

4. Experiments

4.1. Dataset and Implementation Details

Dataset. We evaluated our method on all scenes of the LLFF [21] and Mip-NeRF360 dataset [1]. For the LLFF, we selected 3 views for few-shot reconstruction. While for the Mip-NeRF360, we chose 12 views due to the higher complexity of the scenes. The selected scenes span various styles, ranging from indoor to unbounded outdoor scenarios. Aligning with the previous studies [6, 35, 36, 41], we select every eighth image as the testing view, and evenly sample sparse views from the remaining images for training. Considering the computational resources, our experiments were performed at a resolution of 1/4 of the original image size.

Implementation Details. We configured COLMAP [28] with the same parameters as FSGS for the initialization of various baselines. Building upon this pointcloud, we employed DetectorfreeSfM [11] to further refine and obtain a high-quality pointcloud for our method. At reconstruction stage, λ_{RGB} was set to 0.2 and we simplified the iterative process to accommodate the few-shot task. For baseline methods, we maintained their original training strategies and reported both the time cost and the number of Gaussian kernels used for reconstruction. All experiments were conducted on a single NVIDIA TITAN RTX GPU. We used PSNR, SSIM, and LPIPS [37] as quantitative metrics, where LPIPS uses AlexNet to extract features. Further implementation details are provided in the supplementary material.

4.2. Experimental Results

We compare our method, EAP-GS, with the original 3DGS [14] as well as FSGS [41], CoR-GS [36] and DRGS [6] that are both designed for few-shot reconstruction on LLFF and Mip-NeRF360 dataset. For fair comparison, all methods were trained with the same training data and hardware.

The qualitative results are shown in Fig. 5. Across all scenes, our method stands out and produces more accurate geometry (e.g., iron pail in the *Garden*). Moreover, due to the initialization with a higher number of points containing valuable structural information, our method provides

Methods	LLFF Dataset					Mip-NeRF360 Dataset				
	PSNR	SSIM	LPIPS	Time (min)	Number	PSNR	SSIM	LPIPS	Time (min)	Number
3DGS	14.63	0.4374	0.3425	11.98	379k	16.06	0.3997	0.3892	21.25	1440k
DRGS	17.48	0.5347	0.2922	0.95	463k	17.11	0.4406	0.5412	1.20	207k
FSGS	19.10	0.6246	0.1888	25.49	411k	17.93(4)	0.4802	0.4421	6.51	178k
CoR-GS	18.73	0.6317	0.2214	14.96	91k	17.93(2)	0.4892	0.4639	49.07	171k
EAP-GS(Ours)	18.93	0.6399	0.1792	1.75	125k	18.08	0.4998	0.3696	2.06	457k

Table 1. Quantitative results on LLFF and Mip-NeRF360 datasets. Best score and second-best score are in red and orange respectively. The results of each scene can be viewed in the supplementary material.

richer and more detailed reconstruction, particularly in regions with weakly pronounced texture (e.g., tablecloth in the *Bonsai* and carpet in the *Horns*) and peripheral area (e.g., ceiling of the upper right corner in the *Fern*).

The quantitative results are shown in Tab. 1. Our method achieves leading scores across all metrics while using fewer Gaussians and requiring less computation time. The training strategy of original 3DGS is not suitable for few-shot reconstruction, leading to severe overfitting. FSGS and CoR-GS incorporate pseudo-view regularization, necessitating multiple view renderings per iteration. While this approach yields strong results, it significantly increases computational overhead. In contrast, DRGS mitigates training time through an early-stop strategy, but this may lead to insufficient training. Additionally, the sparse pointcloud might have a large error when scale matching the relative depth. In comparison, our augmented pointcloud inherently encodes depth information, providing a good guidance for Gaussian generation.

Moreover, APA can be easily integrated into other 3DGS-based optimization methods. Tab. 2 and Fig. 6 compares the reconstruction results of various methods with and without APA. The incorporation of APA significantly enhances the performance of these methods. Compared to the original 3DGS, the pointcloud augmentation in the initialization stage and the regularization in the reconstruction stage do not conflict, allowing it to be framed as a modular to existing methods with minimal overhead. In summary, inadequate initialization is a critical aspect neglected in the current research. We address this issue, and the integration of APA into existing methods can enhance their performance.

4.3. Ablation Studies

We conducted ablation studies to assess the impact of our APA technique and the DetectorfreeSfM [11] method. The results are shown in Tab. 3 and Fig. 7. APA significantly improves the overall number and distribution of initial points, resulting in more accurate and reasonable scene geometry. More specifically, it helps generate a sufficient number of points in regions with complex structure but originally sparse pointcloud distribution, such as regions with weakly pronounced texture and peripheral area (e.g., gap in floor tile, black spots on marble), leading to more detailed reconstructions. On the other hand, our results confirm that

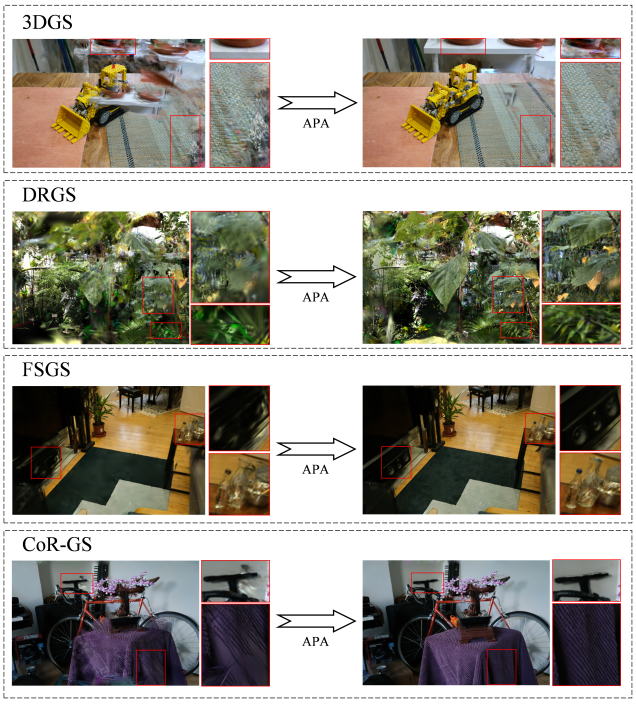

Figure 6. Comparison of different methods. The testing view reconstructions of different 3DGS-based optimization methods without (left column) and with (right column) APA.

	LLFF Dataset				Mip-NeRF360 Dataset			
	PSNR	SSIM	LPIPS	Number	PSNR	SSIM	LPIPS	Number
3DGS	14.63	0.4374	0.3425	379k	16.06	0.3997	0.3892	1440k
APA + 3DGS	14.78	0.4703	0.3239	404k	17.06	0.4525	0.3426	1490k
DRGS	17.48	0.5347	0.2922	463k	17.11	0.4406	0.5412	207k
APA + DRGS	18.12	0.5730	0.2530	759k	17.43	0.4614	0.4752	414k
FSGS	19.10	0.6246	0.1888	411k	17.93	0.4802	0.4421	178k
APA + FSGS	19.76	0.6665	0.1628	199k	18.51	0.5268	0.3922	191k
CoR-GS	18.73	0.6317	0.2214	91k	17.93	0.4892	0.4639	171k
APA + CoR-GS	19.32	0.6638	0.1877	95k	18.76	0.5243	0.4152	202k

Table 2. Comparison of different methods. We compare the reconstruction results for different 3DGS-based optimization methods with and without APA.

	PSNR	SSIM	LPIPS	Number
COLMAP w/o APA	18.05	0.5871	0.2184	102k
DetectorfreeSfM w/o APA	17.43	0.5664	0.2443	85k
COLMAP with APA	18.74	0.6206	0.1936	115k
DetectorfreeSfM with APA	18.93	0.6399	0.1792	125k

Table 3. Ablation study on different initialization. We compare all metrics on the LLFF dataset initialized by COLMAP [28] and DetectorfreeSfM [11] methods with and without APA.

Pointcloud Augmentation	Attention Mechanism	PSNR	SSIM	LPIPS	Number
×	×	17.43	0.5664	0.2443	85k
✓	×	18.78	0.6353	0.1829	126k
✓	✓	18.93	0.6399	0.1792	125k

Table 4. Ablation study on proposed components. We evalute the effect of each component of EAP-GS on the LLFF dataset.

DetectorfreeSfM is more effective at extracting 2D feature points in texture-poor areas (e.g., white marble) compared to COLMAP. This enables us to get more 3D points in texture-poor areas by lowering the matching requirements. This is why *DetectorfreeSfM+APA* generates nearly three times as many initial points as *COLMAP+APA* in the case of Fig. 7. It is also clear evidence showing that APA is a key factor to improve the quality of initialization from the comparison of these two SfM methods.

As shown in Tab. 4, each component of our method improves the reconstruction metrics for the LLFF dataset. Pointcloud Augmentation increases the number of initial points, providing better initialization and enhancing the stability and accuracy of subsequent optimization. Attention mechanism allows the augmentation process to focus more on pointcloud generation in regions with complex structure but sparse pointcloud distribution, enriching reconstruction details, while also preventing to introduce unnecessary error in regions with dense pointcloud.

5. Discussion

Although EAP-GS provides a better initialized pointcloud, it requires an additional step, which slightly increases the time cost in initialization, normally adds a few seconds and are almost ignorable for few-shot reconstruction tasks.

Moreover, EAP-GS can extract more view information than traditional methods in the initialization stage, but the improvement of reconstruction quality is still limited in few-shot cases, particularly in extrapolation testing views. This issue is primarily due to data incompleteness, and a potential approach to further enhance performance would be to incorporate prior knowledge or generative models for inference, which will be a focus of our future work.

Although the error introduced by retaining two-view tracks is not very large in our current study, there remains the possibility that this mechanism may introduce subtle errors in special cases, thus affecting the reconstruction quality. Lacking a method to limit the error may be a limitation of this approach, and further research is needed.

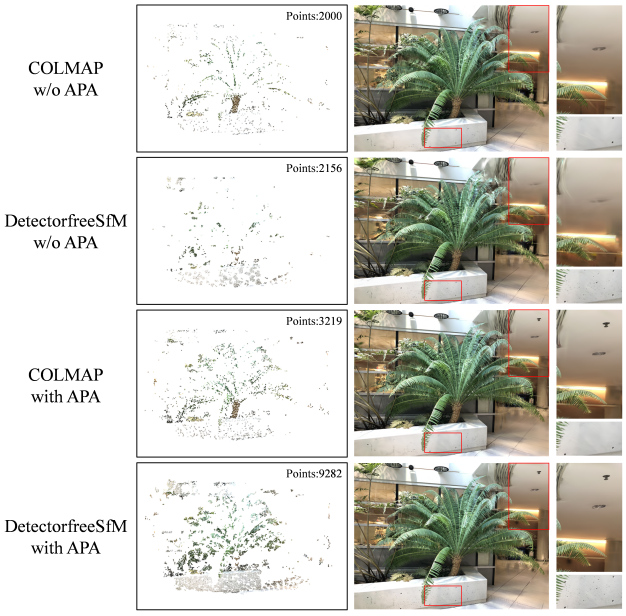

Figure 7. Ablation study on different initialization. The initialized pointcloud generated by COLMAP [28] and DetectorfreeSfM [11] methods with and without APA (left column), corresponding testing view reconstructions (middle column), and zoom-ins of two representative regions (right column). Obvious advantage of *DetectorfreeSfM with APA* can be observed.

6. Conclusion

In this work, we find and confirm that inadequate initialization is a critical factor contributing to scene degradation in few-shot cases. Using traditional SfM methods are not sufficient to exploit the information across all views. Based on this finding, we propose EAP-GS, a novel framework for few-shot scene reconstruction. Specifically, we design an Attentional Pointcloud Augmentation technique, which retains two-view tracks as an option in regions with complex structure but sparse pointcloud distribution, so that it dramatically increases the number of initial points and effectively balances the density distribution of pointcloud, leads to more stable and accurate reconstruction. Additionally, our approach can be easily integrated into 3DGS-based optimization methods to further improve their performance with minimal overhead. As demonstrated by experiments conducted on the LLFF and Mip-NeRF360 datasets from various indoor and outdoor scenes, EAP-GS achieves outstanding scene reconstruction quality and outperforms the state-of-the-art methods.

References

[1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5470–5479, 2022. 2, 3, 6

[2] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*, 2023. 3

[3] Eric Brachmann, Jamie Wynn, Shuai Chen, Tommaso Cavallari, Áron Monszpart, Daniyar Turmukhambetov, and Victor Adrian Prisacariu. Scene coordinate reconstruction: Posing of image collections via incremental learning of a relocalizer. *arXiv preprint arXiv:2404.14351*, 2024. 4

[4] Yuanhao Cai, Yixun Liang, Jiahao Wang, Angtian Wang, Yulun Zhang, Xiaokang Yang, Zongwei Zhou, and Alan Yuille. Radiative gaussian splatting for efficient x-ray novel view synthesis. In *European Conference on Computer Vision*, pages 283–299. Springer, 2025. 2, 3

[5] Guikun Chen and Wenguan Wang. A survey on 3d gaussian splatting. *arXiv preprint arXiv:2401.03890*, 2024. 2

[6] Jaeyoung Chung, Jeongtaek Oh, and Kyoung Mu Lee. Depth-regularized optimization for 3d gaussian splatting in few-shot images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 811–820, 2024. 3, 4, 6

[7] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27, 2014. 3

[8] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, pages 226–231, 1996. 5

[9] Zhiwen Fan, Wenyan Cong, Kairun Wen, Kevin Wang, Jian Zhang, Xinghao Ding, Danfei Xu, Boris Ivanovic, Marco Pavone, Georgios Pavlakos, et al. Instantsplat: Unbounded sparse-view pose-free gaussian splatting in 40 seconds. *arXiv preprint arXiv:2403.20309*, 2024. 4

[10] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 4, 5

[11] Xingyi He, Jiaming Sun, Yifan Wang, Sida Peng, Qixing Huang, Hujun Bao, and Xiaowei Zhou. Detector-free structure from motion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21594–21603, 2024. 3, 4, 5, 6, 7, 8

[12] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5885–5894, 2021. 3

[13] Jaewoo Jung, Jisang Han, Honggyu An, Jiwon Kang, Seonghoon Park, and Seungryong Kim. Relaxing accurate initialization constraint for 3d gaussian splatting. *arXiv preprint arXiv:2403.09413*, 2024. 4

[14] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2, 3, 6, 4

[15] 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 of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20775–20785, 2024. 3

[16] Yingtai Li, Xueming Fu, Shang Zhao, Ruiyang Jin, and S Kevin Zhou. Sparse-view ct reconstruction with 3d gaussian volumetric representation. *arXiv preprint arXiv:2312.15676*, 2023. 2, 3

[17] Guibiao Liao, Jiankun Li, Zhenyu Bao, Xiaoqing Ye, Jingdong Wang, Qing Li, and Kanglin Liu. Clip-gs: Clip-informed gaussian splatting for real-time and view-consistent 3d semantic understanding. *arXiv preprint arXiv:2404.14249*, 2024. 4

[18] Jian Liu, Xiaoshui Huang, Tianyu Huang, Lu Chen, Yuenan Hou, Shixiang Tang, Ziwei Liu, Wanli Ouyang, Wangmeng Zuo, Junjun Jiang, et al. A comprehensive survey on 3d content generation. *arXiv preprint arXiv:2402.01166*, 2024. 2

[19] Jiageng Mao, Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. 3d object detection for autonomous driving: A comprehensive survey. *International Journal of Computer Vision*, 131(8):1909–1963, 2023. 2

[20] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7210–7219, 2021. 3

[21] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (ToG)*, 38(4):1–14, 2019. 2, 6

[22] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2, 3

[23] Amirali Molaei, Amirhossein Aminimehr, Armin Tavakoli, Amirhossein Kazerouni, Bobby Azad, Reza Azad, and Dorit Merhof. Implicit neural representation in medical imaging: A comparative survey. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2381–2391, 2023. 2

[24] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 3

[25] Michael Niemeyer, Fabian Manhardt, Marie-Julie Rakotosaona, Michael Oechsle, Daniel Duckworth, Rama Gosula, Keisuke Tateno, John Bates, Dominik Kaeser, and Federico Tombari. Radsplat: Radiance field-informed gaussian splat-

ting for robust real-time rendering with 900+ fps. *arXiv preprint arXiv:2403.13806*, 2024. 4

[26] Avinash Paliwal, Wei Ye, Jinhui Xiong, Dmytro Kotovenko, Rakesh Ranjan, Vikas Chandra, and Nima Khademi Kalantari. Coherentgs: Sparse novel view synthesis with coherent 3d gaussians. *arXiv preprint arXiv:2403.19495*, 2, 2024. 2, 3, 4

[27] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12179–12188, 2021. 3

[28] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 2, 3, 4, 6, 8

[29] Jiaxiang Tang, Jiawei Ren, Hang Zhou, Ziwei Liu, and Gang Zeng. Dreamgaussian: Generative gaussian splatting for efficient 3d content creation. *arXiv preprint arXiv:2309.16653*, 2023. 2, 3

[30] Jiaxiang Tang, Zhaoxi Chen, Xiaokang Chen, Tengfei Wang, Gang Zeng, and Ziwei Liu. Lgm: Large multi-view gaussian model for high-resolution 3d content creation. In *European Conference on Computer Vision*, pages 1–18. Springer, 2025. 2, 3

[31] Shimon Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405–426, 1979. 2

[32] Tong Wu, Yu-Jie Yuan, Ling-Xiao Zhang, Jie Yang, Yan-Pei Cao, Ling-Qi Yan, and Lin Gao. Recent advances in 3d gaussian splatting. *Computational Visual Media*, 10(4):613–642, 2024. 2

[33] Zhongrui Yu, Haoran Wang, Jinze Yang, Hanzhang Wang, Zeke Xie, Yunfeng Cai, Jiale Cao, Zhong Ji, and Mingming Sun. Sgd: Street view synthesis with gaussian splatting and diffusion prior. *arXiv preprint arXiv:2403.20079*, 2024. 2, 3

[34] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1960–1970, 2021. 2, 3

[35] Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, and Eric Xing. Fregs: 3d gaussian splatting with progressive frequency regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21424–21433, 2024. 4, 6

[36] Jiawei Zhang, Jiahe Li, Xiaohan Yu, Lei Huang, Lin Gu, Jin Zheng, and Xiao Bai. Cor-gs: sparse-view 3d gaussian splatting via co-regularization. In *European Conference on Computer Vision*, pages 335–352. Springer, 2025. 3, 6, 4

[37] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6

[38] Zheng Zhang, Wenbo Hu, Yixing Lao, Tong He, and Hengshuang Zhao. Pixel-gs: Density control with pixel-aware gradient for 3d gaussian splatting. *arXiv preprint arXiv:2403.15530*, 2024. 2, 4

[39] Shijie Zhou, Haoran Chang, Sicheng Jiang, Zhiwen Fan, Zehao Zhu, Dejia Xu, Pradyumna Chari, Suya You, Zhangyang Wang, and Achuta Kadambi. Feature 3dgs: Supercharging 3d gaussian splatting to enable distilled feature fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21676–21685, 2024. 4

[40] Xiaoyu Zhou, Zhiwei Lin, Xiaojun Shan, Yongtao Wang, Deqing Sun, and Ming-Hsuan Yang. Drivinggaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21634–21643, 2024. 2, 3

[41] Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. Fsgs: Real-time few-shot view synthesis using gaussian splatting. In *European Conference on Computer Vision*, pages 145–163. Springer, 2025. 1, 2, 3, 4, 6

[42] Xingxing Zuo, Pouya Samangouei, Yunwen Zhou, Yan Di, and Mingyang Li. Fmgs: Foundation model embedded 3d gaussian splatting for holistic 3d scene understanding. *International Journal of Computer Vision*, pages 1–17, 2024. 4