DAPS-AGF: Depth-Aware Perceptual Similarity with Adaptive Gradient Filtering for Enhanced Outdoor Scene Reconstruction

Aqsa Yousaf, Arkajyoti Mitra, Paul Agbaje, Afia Anjum, Habeeb Olufowobi
The University of Texas at Arlington
{aqsa.yousaf, arkajyoti.mitra, pauloluwatowoju.agbaje, habeeb.olufowobi}@uta.edu

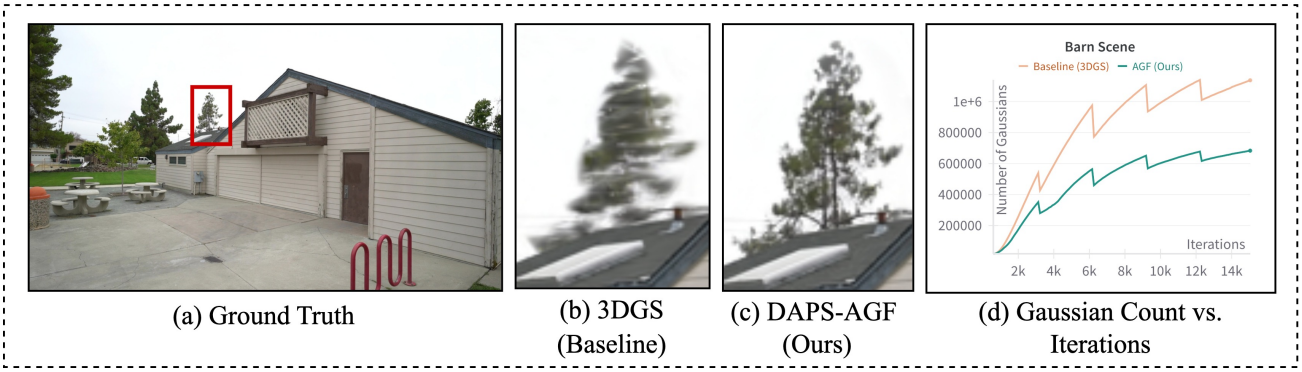

Figure 1. Qualitative and quantitative comparison of reconstruction results. **(a)** Ground Truth image of the Barn scene, highlighting a region of interest in red. **(b)** Reconstruction result from baseline 3D Gaussian Splatting, showing noticeable artifacts and lack of detail. **(c)** Reconstruction results from our proposed method (DAPS-AGF) demonstrate improved reconstruction of distant objects. **(d)** Plot showing the memory efficiency of AGF compared to the baseline 3DGS. It significantly reduces the number of Gaussians required during optimization while maintaining high-quality reconstruction.

Abstract

3D Gaussian Splatting (3DGS) is a recent technique for real-time scene reconstruction. However, in large outdoor scenes, it often fails to reconstruct peripheral or distant regions accurately. These areas appear in fewer views and receive weaker multi-view supervision, leading to lower gradient signals, fewer Gaussians, and ultimately degraded structural quality. To address this limitation, we propose two complementary enhancements to the 3DGS pipeline. Our Depth-Aware Perceptual Similarity (DAPS) module uses monocular depth to strengthen optimization in weakly supervised regions, leading to sharper edges and improved reconstruction quality. Additionally, we introduce Adaptive Gradient Filtering (AGF), a dynamic densification mechanism that selectively clones Gaussians based on gradient statistics, ensuring visually faithful reconstruction without excessive memory growth. We curate a challenging outdoor benchmark by selecting scenes from the Tanks & Temples and Mip-NeRF 360 datasets, designed to test reconstruction quality and memory efficiency under real-world visual complexity. Experimental results show DAPS improves SSIM by 3.05%, PSNR by 3.94%, and reduces LPIPS by 17.78%, while AGF cuts memory usage by 38.5% over baseline 3DGS without sacrificing image quality.

1. Introduction

Scene reconstruction is the cornerstone of numerous AI-driven applications, from creating immersive augmented and virtual reality (AR/VR) environments to the reliable operation of autonomous systems [4]. Among recent advances in scene reconstruction, 3D Gaussian Splatting [12] stands out for its ability to reconstruct scenes with both accuracy and real-time efficiency. 3DGS effectively captures visual and structural complexities using anisotropic Gaussian splats and spherical harmonics for precise color representation.

Recent research on 3D Gaussian Splatting has produced numerous improvements, mainly in memory optimization and quality enhancements. On the memory optimization

side, several works [8, 9, 11, 15, 18–21, 25] aim to reduce the storage footprint (vector quantization) and number of Gaussians (pruning) required to reconstruct a scene. Meanwhile, other works focus on improving 3DGS quality. A prominent direction is incorporating depth guidance to improve reconstruction from limited or imperfect data [3, 5–7, 14, 17, 23, 24].

While recent depth-guided approaches have made notable progress in addressing sparse-view or few-shot settings, they primarily focus on incorporating external priors to compensate for limited data. In contrast, we identify a fundamentally different and underexplored challenge. We noticed that even with access to full training views, 3DGS often fails to adequately reconstruct regions that receive weak gradient supervision, such as peripheral or distant structures. These regions may be visible in the data a few times, yet remain under-optimized due to their low photometric error and limited multi-view constraints, as illustrated in Figure 1. However, attempts to enhance such under-optimized regions can often trigger excessive Gaussian cloning, especially when standard densification mechanisms are used. Although recent works have proposed pruning [20] or compression techniques [25] to reduce the memory footprint, these approaches are largely applied post hoc or uniformly across the scene. They are not designed to regulate memory dynamically during optimization, nor do they account for localized reconstruction needs in weakly supervised areas.

We propose two complementary strategies, 1) Depth-Aware Perceptual Similarity (DAPS) and 2) Adaptive Gradient Filtering (AGF). DAPS improves reconstruction quality in sparsely observed regions by integrating monocular depth information directly into the optimization process. By emphasizing regions with higher depth uncertainty, DAPS guides the reconstruction to target structurally ambiguous areas explicitly. This approach results in more consistent and detailed reconstructions of peripheral scene elements. Complementing DAPS, AGF introduces a dynamic Gaussian densification strategy. Instead of using fixed thresholds, AGF adaptively regulates Gaussian cloning based on the statistical distribution of gradient magnitudes encountered during training. This adaptive approach effectively balances the memory usage and computational efficiency of the reconstruction, achieving detailed reconstructions without incurring excessive computational costs.

The key contributions of this work can be outlined as follows:

- We propose DAPS, a depth-aware optimization strategy that improves reconstruction in regions with weak supervision, such as distant or peripheral areas, by using monocular depth as a proxy to identify where optimization may be insufficient.
- We propose AGF, an adaptive Gaussian densification mechanism that regulates cloning dynamically based on the statistical distribution of gradient magnitudes, achieving a superior balance between reconstruction quality and memory usage.
- We create a new Outdoor Scenes benchmark by selecting challenging outdoor scenes from Mip-NeRF 360 and Tanks & Temples datasets.
- We evaluate our approach on the Outdoor Scenes benchmark, demonstrating consistent improvements in reconstruction quality, particularly in sparsely observed regions, while reducing memory usage compared to baseline densification strategies.

2. Related Work

Scene reconstruction is a critical area of research in computer vision and graphics, with applications in autonomous driving, virtual reality, and robotics [26]. While traditional methods rely on dense point clouds [16] or meshes [10], recent advances such as 3D Gaussian Splatting offer more efficient and realistic representations. In this section, we review the key developments in GS, focusing on the integration of depth information for enhanced reconstruction and strategies for improving memory efficiency.

2.1. Depth-Aware Gaussian Splatting

Several recent works have integrated depth information into 3D Gaussian Splatting (3DGS) to improve reconstruction quality, particularly in sparse-view settings [6], few-shot scenes [7], or SLAM-based pipelines [17]. Chung et al. [7] propose depth-guided geometric losses and smoothness constraints to improve surface consistency under few-shot supervision. This idea is further refined in SAD-GS [14], which introduces shape-aligned depth regularization, leading to sharper reconstructions and improved novel view synthesis.

Other works use multi-view stereo (MVS) cues. MVSplat [6] incorporates cost volumes, plane-sweeping, and cross-view feature matching to produce consistent depth maps in sparse multi-view settings, enhancing geometry quality while maintaining inference efficiency. Similarly, Depth-Splat [23] integrates monocular depth into 3DGS by fusing global context with image features, achieving strong depth estimation with limited views. DASH-Gaussian [5] enhances reconstruction by combining sparse depth and semantic priors to guide splat initialization and improve performance in indoor scenes. CDGS [24] introduces an adaptive confidence-driven depth fusion strategy for handling depth ambiguity across sparse input views.

While these approaches successfully exploit depth priors for geometric refinement and sparse-view compensation, they largely focus on improving depth accuracy or enforcing global consistency. Notably, most operate under sparse or few-shot supervision and do not address the opti-

mization bias present in full-supervision settings. In large-scale scenes, peripheral or distant regions may still suffer from weak gradient supervision due to limited visibility or low photometric error in those regions, leading to under-densification and degraded reconstruction.

In contrast, our proposed DAPS uses monocular depth not as a geometric constraint but as a driver of perceptual supervision. By selectively emphasizing regions with high depth uncertainty, DAPS guides optimization toward under-represented areas. This targeted supervision improves reconstruction detail in peripheral regions.

2.2. Memory Optimized Gaussian Splatting

3DGS has a high memory footprint, requiring millions of Gaussians to represent complex scenes. This substantial memory usage poses significant challenges in storage, rendering efficiency, and scalability for larger scenes and resource-constrained devices. To address these challenges, researchers have explored techniques such as vector quantization and pruning [1].

Vector Quantization. This approach compresses Gaussian attributes by clustering similar parameters. Compression techniques such as CompGS [18] framework use K-means clustering to optimize Gaussian parameters into compact codebooks, achieving up to a 45–50$\times$ reduction in storage and a 3$\times$ speed improvement in rendering. This approach was further refined with sensitivity-aware clustering [19], where parameters contributing minimally to reconstruction were grouped and quantized, achieving up to 31$\times$ compression. Additionally, the EAGLES [11] framework introduces lightweight encoding for attributes such as color and rotation by using quantized embeddings, achieving memory reductions of 10–20$\times$ while preserving the quality of reconstruction.

Pruning Strategies. Pruning eliminates redundant Gaussians to improve memory and computational efficiency. A compact 3D Gaussian representation framework was introduced by [15] that leverages learnable masks to identify and remove Gaussians with minimal impact on rendering quality. Similarly, LightGaussian [8] employed a global significance score to guide pruning, achieving a 15$\times$ compression rate, while maintaining visual quality and increasing the rendering speeds to over 200 FPS. Pateux et al.[21] introduce a Bayesian optimization framework that selects a minimal subset of Gaussians to balance model size with reconstruction fidelity. Zhang et al.[25] present LP-3DGS, a learnable pruning approach that predicts which Gaussians to retain based on scene visibility and feature importance. Papantonakis et al.[20] introduce a pruning and quantization pipeline that compresses the 3DGS representation by adapting the number of spherical harmonics and reducing redundancy. Farooq et al.[9] propose a coarse-to-fine optimization strategy that incrementally refines Gaussians while maintaining compactness through spatial grouping.

Despite these advancements, current approaches primarily focus on post-optimization strategies, such as compression and pruning, which identify redundancies after Gaussian attributes have been fully optimized. A significant drawback of these strategies is their requirement for high-storage GPUs for execution.

To address this limitation, we propose Adaptive Gradient Filtering (AGF), a novel mechanism that selectively filters Gaussians during the optimization process. This approach effectively prevents unnecessary Gaussian densification at the source, enhancing memory efficiency from the outset and reducing reliance on resource-intensive post-processing techniques.

3. Preliminary

3D Gaussian Splatting [12] introduces a novel approach to scene reconstruction by introducing Gaussian spheres as the fundamental representation. Unlike traditional point clouds, 3DGS encodes each Gaussian with a set of attributes (Gaussian primitives) such as position, color, opacity, and shape, enabling adaptive reconstruction based on input images and camera priors. The core of 3DGS lies in the Gaussian function, defined in eq. 1, where $G(x)$ represents the Gaussian density at a point x. The Σ is the covariance matrix that encodes the spatial structure of the Gaussian.

$$G(x) = e^{-\frac{1}{2}x^T\Sigma^{-1}x} \tag{1}$$

The covariance matrix Σ, expressed in eq. 2, combines a diagonal scaling matrix S representing the anisotropic scaling of the Gaussian along its principal axes, and a rotation matrix R that determines the orientation of the Gaussian.

$$\Sigma = RSS^TR^T \tag{2}$$

The optimization process employs a stochastic gradient descent to align each Gaussian primitive to the underlying scene geometry and appearance.

4. Methodology

This section outlines the proposed DAPS and AGF techniques, highlighting their contributions to enhancing the efficiency and quality of scene reconstruction. Figure 2 provides a high-level overview of their integration within the 3DGS framework.

Depth-Aware Perceptual Similarity (DAPS). Regions that appear less frequently across camera views, such as distant backgrounds or side-facing surfaces, often receive weaker gradient supervision during training. As a result, these under-represented areas are prone to reconstruction errors and lack structural detail. To address this issue, we propose Depth-Aware Perceptual Similarity (DAPS), a loss

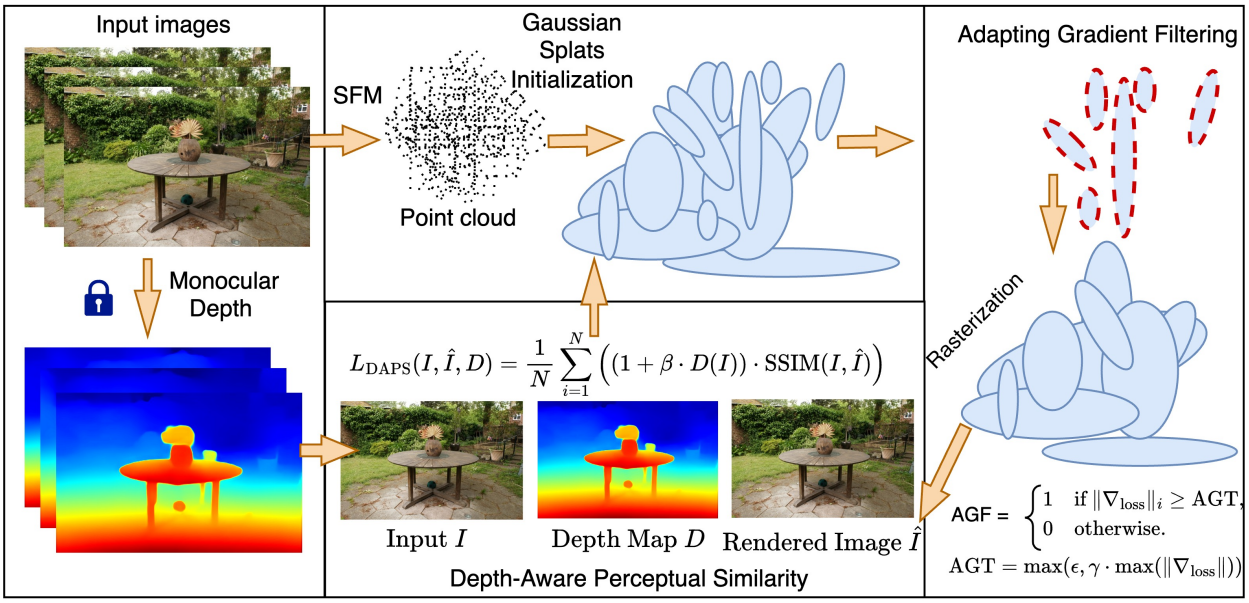

Figure 2. **Overview of the proposed approach.** The pipeline begins with Structure-from-Motion (SfM) to generate a sparse point cloud, followed by Gaussian initialization. Monocular depth maps are extracted from input images to guide the DAPS module, which emphasizes regions farther from the camera during optimization. AGF dynamically adjusts the Gaussian cloning threshold based on the statistical distribution of gradient magnitudes, reducing unnecessary densification. Together, these modules improve reconstruction quality in under-represented regions while maintaining memory efficiency during rendering.

function that incorporates depth information to guide learning toward these challenging regions. Instead of treating all pixels equally, DAPS uses per-pixel depth predictions to assign greater importance to areas that are farther from the camera, encouraging the model to preserve structure and improve perceptual consistency in those regions.

Given an input image $I \in \mathbb{R}^{3\times H\times W}$ and its reconstructed counterpart $\hat{I}$, we first compute the SSIM map using local window statistics:

$$\text{SSIM}_{\text{map}}(I,\hat{I})_i = \frac{(2\mu_I\mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \quad (3)$$

where μ_I, $\mu_{\hat{I}}$, σ_I^2, $\sigma_{\hat{I}}^2$, and $\sigma_{I\hat{I}}$ denote the local means, variances, and covariance within a fixed-size window, and C_1, C_2 are constants for numerical stability.

To modulate SSIM with scene geometry, we estimate a per-pixel depth map $D \in \mathbb{R}^{H\times W}$ from the ground truth image I using the MiDaS monocular depth estimator [22] denoted as $\mathcal{M}$. The depth map is normalized to the $[0, 1]$ range.

$$D = \mathcal{M}(I), \quad (4)$$

The DAPS loss is then defined as the weighted average of the SSIM map and depth map as:

$$L_{\text{DAPS}}(I,\hat{I},D) = \frac{1}{N}\sum_{i=1}^{N}(1+\beta\cdot D_i)\cdot \text{SSIM}_{\text{map}}(I,\hat{I})_i \quad (5)$$

where N is the total number of pixels, β is a scaling parameter controlling depth influence, D_i is the i-th pixel from D, and $\text{SSIM}_{\text{map}}(I,\hat{I})_i$ is the value corresponding to the i-th pixel in $\text{SSIM}_{\text{map}}(I,\hat{I})$. To ensure both perceptual similarity and pixel-level accuracy, we combine L_{DAPS} with the $L1$ norm, L_{L1}, which measures absolute differences between the predicted and ground-truth images, expressed as:

$$L_{\text{L1}}(I,\hat{I}) = \frac{1}{N}\sum_{i=1}^{N}|I_i - \hat{I}_i| \quad (6)$$

resulting in the final loss:

$$L(I,\hat{I},D) = \alpha\cdot L_{\text{L1}}(I,\hat{I}) + (1-\alpha)\cdot L_{\text{DAPS}}(I,\hat{I},D) \quad (7)$$

where $\alpha \in [0, 1]$ determines the balance between the two components. The composite loss ensures both accurate reconstruction and improved perceptual quality in under-optimized regions. This approach reduces artifacts and enhances the completeness of the reconstructed scene.

Adaptive Gradient Filtering (AGF). While DAPS significantly improves reconstruction quality by adding Gaussians

to under-represented regions, this increased supervision can also lead to higher memory usage. Moreover, 3D Gaussian Splatting inherently requires substantial memory due to its dense representation and frequent Gaussian splitting and cloning during optimization.

While both cloning and splitting contribute to Gaussian growth, cloning is the dominant source of redundancy in 3DGS. Cloning creates identical copies of Gaussians and does not alter their attributes, often oversaturating already well-optimized areas. In contrast, splitting introduces variations and is applied more sparingly. Therefore, AGF focuses on controlling cloning, which has a more significant impact on memory.

AGF introduces an adaptive gradient threshold (AGT) mechanism to dynamically adjusts the densification threshold based on the global gradient distribution, allowing for selective and efficient cloning. This adaptive strategy ensures cloning is applied where it is most beneficial, minimizing redundant memory usage with minimal degradation in performance. Mathematically, AGF operates as a binary mask that determines which Gaussians undergo cloning and is defined as:

$$\text{AGF}_i = \begin{cases} 1, & \|\nabla_{\text{loss}}\|_i \geq \text{AGT} \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $\|\nabla_{\text{loss}}\|_i$ represents the gradient magnitude of the loss function for Gaussian i, and AGT is computed as:

$$\text{AGT} = \max(\epsilon, \gamma \cdot \max(\|\nabla_{\text{loss}}\|)) \tag{9}$$

with ϵ setting a lower bound to prevent excessively small thresholds in low-gradient regions, γ scaling the threshold proportionally to the maximum gradient magnitude, and $\max(\|\nabla_{\text{loss}}\|)$ being the global maximum gradient value that changes dynamically.

Cloning is applied to the selected Gaussians after masking. Given an original set $G = \{g_1, g_2, \ldots, g_N\}$ and a subset $G_c = \{g_i \in G \mid \text{AGF}_i = 1\}$ selected for cloning, the operation is expressed as:

$$G' = G \cup \{g'_i \mid g_i \in G_c\} \tag{10}$$

where, G' is the updated set after cloning, and g'_i is a cloned Gaussian derived from g_i. This adaptive strategy ensures that cloning targets regions where it is most beneficial, reducing memory overhead while maintaining reconstruction.

By integrating DAPS and AGF into the 3DGS pipeline, our method adaptively enhances the reconstruction quality of under-represented regions while simultaneously managing memory usage. DAPS prioritizes optimization in peripheral and background areas, ensuring a more balanced scene representation. Meanwhile, AGF dynamically adjusts the Gaussian cloning threshold based on the gradient distribution, effectively reducing unnecessary densification and promoting efficient scene representation without compromising visual quality.

	SSIM (↑)		PSNR (↑)		LPIPS (↓)	
Scene	3DGS	$3DGS_{DAPS}$ (Ours)	3DGS	$3DGS_{DAPS}$ (Ours)	3DGS	$3DGS_{DAPS}$ (Ours)
Family	0.925	**0.952**	28.39	**29.88**	0.112	**0.063**
Francis	0.920	**0.940**	32.19	**34.04**	0.153	**0.110**
Train	0.815	**0.838**	**22.05**	21.80	0.206	**0.159**
Ignatius	0.928	**0.954**	30.17	**31.98**	0.084	**0.046**
Barn	0.951	**0.968**	30.35	**33.53**	0.073	**0.039**
Bicycle	0.663	**0.718**	**23.56**	23.50	0.335	**0.249**
Flowers	0.528	**0.572**	20.55	**20.63**	0.415	**0.360**
Garden	0.828	**0.849**	26.20	**26.22**	0.163	**0.123**
Stump	0.722	**0.760**	25.62	**26.05**	0.294	**0.229**
Treehill	0.587	**0.610**	**22.07**	21.52	0.417	**0.343**
Average	0.787	**0.816**	26.12	**26.92**	0.225	**0.172**

Table 1. Performance evaluation of **$3DGS_{DAPS}$ on the outdoor scenes dataset.** $3DGS_{DAPS}$ demonstrates consistent improvements in SSIM, PSNR, and LPIPS over the baseline 3DGS across both datasets.

5. Experiments

5.1. Evaluation Setup

We adopt 3D Gaussian Splatting (3DGS) [12] as our baseline $3DGS_{Base}$ and evaluate our method by progressively incorporating the two proposed modules. To measure the combined effect of both modules, we evaluate the full model $3DGS_{DAPS+AGF}$, where both DAPS and AGF are integrated into the pipeline. To analyze the individual impact of each component, we further evaluate two intermediate variants. $3DGS_{DAPS}$, which includes only the DAPS module, and $3DGS_{AGF}$, which includes only the AGF module. This staged evaluation setup enables a detailed comparison across all variants.

5.2. Datasets

We construct a custom evaluation dataset by combining outdoor scenes from two widely recognized benchmarks for 3D scene reconstruction, MIP-NeRF 360 [2] and Tanks & Temples [13]. Our selected scenes include Train, Ignatius, Barn, Family, Francis, Bicycle, Flowers, Garden, Stump, and Treethill. This curated subset, referred to as our Outdoor-Scenes Benchmark, is used consistently across all experiments and ablation studies.

5.3. Implementation Details

Our experiments were conducted on an NVIDIA RTX 6000 Ada Generation GPU with 48 GB of VRAM. We use the

Outdoor Scenes	SSIM (↑)		PSNR (↑)		LPIPS (↓)		Memory GB (↓)	
	3DGS	$3DGS_{(DAPS + AGF)}$ (Ours)	3DGS	$3DGS_{(DAPS + AGF)}$ (Ours)	3DGS	$3DGS_{(DAPS + AGF)}$ (Ours)	$3DGS_{DAPS}$	$3DGS_{(DAPS + AGF)}$ (Ours)
Family	0.925	**0.954**	28.39	**30.39**	0.112	**0.062**	0.64	**0.49**
Francis	0.920	**0.941**	32.19	**34.56**	0.153	**0.109**	0.52	**0.36**
Train	0.815	**0.837**	22.05	**22.98**	0.206	**0.173**	0.96	**0.50**
Ignatius	0.928	**0.956**	30.17	**32.38**	0.084	**0.045**	1.09	**0.75**
Barn	0.951	**0.968**	30.35	**33.46**	0.073	**0.041**	0.84	**0.53**
Bicycle	0.663	**0.695**	**23.56**	23.45	0.335	**0.288**	3.9	**1.3**
Flowers	0.528	**0.560**	20.55	20.55	0.415	**0.376**	2.7	**1.6**
Garden	0.828	**0.841**	**26.20**	26.07	0.163	**0.142**	4.0	**2.0**
Stump	0.722	**0.754**	25.62	**25.96**	0.294	**0.245**	4.2	**2.8**
Treehill	0.587	**0.604**	**22.07**	21.72	0.417	**0.366**	3.0	**1.6**
Average	0.787	**0.811**	26.12	**27.15**	0.225	**0.185**	2.19	**1.19**

Table 2. Performance and memory usage evaluation of **$3DGS_{(DAPS + AGF)}$** on outdoor scenes dataset. The combined approach maintains higher SSIM, PSNR, and LPIPS scores than the baseline 3DGS while efficiently controlling the number of Gaussians.

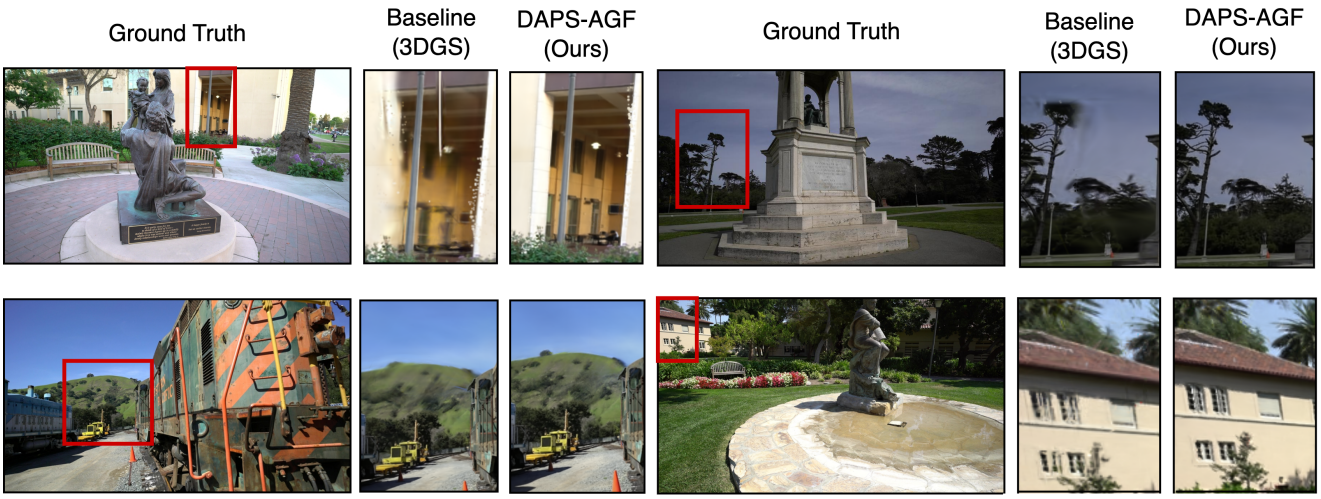

Figure 3. Qualitative comparison between baseline 3DGS and our proposed DAPS-AGF method across various outdoor scenes. Red boxes highlight under-optimized regions that are typically difficult to reconstruct, such as background structures and distant objects. Compared to the baseline, our method produces sharper edges, more consistent geometry, and improved detail preservation in such areas.

following hyperparameters for the experiments: $\alpha = 0.5$, $\beta = 1$, $\gamma = 0.03$, and $\epsilon = 3 \times 10^{-4}$. These values were selected to ensure balanced contributions from the DAPS and AGF modules during optimization. Scenes from Tanks & Temples are optimized for 30k iterations, while MIP-NeRF 360 scenes are optimized for 7k iterations at 4× image resolution.

6. Results and Comparisons

3DGS with DAPS + AGF. Table 2 presents the comparison between the original 3DGS baseline and our enhanced method ($3DGS_{DAPS + AGF}$) on the outdoor scenes datasets. Across all scenes, our approach consistently yields higher SSIM and PSNR scores, along with significantly lower LPIPS, indicating improved structural fidelity and perceptual similarity. These gains are most evident in scenes with complex geometry or peripheral content, where traditional 3DGS tends to underperform (Figure 3).

While DAPS improves reconstruction quality by enhancing under-represented regions, it also increases memory usage due to additional Gaussian proliferation. By introducing AGF, we reduce this overhead by 45.6% relative to DAPS, without sacrificing visual quality. This demonstrates

Scene	SSIM (↑)		PSNR (↑)		LPIPS (↓)		Memory GB (↓)	
	3DGS	$3DGS_{AGF}$ (ours)	3DGS	$3DGS_{AGF}$ (ours)	3DGS	$3DGS_{AGF}$ (ours)	3DGS	$3DGS_{AGF}$ (ours)
Family	0.925	0.923	28.39	28.57	0.112	0.117	0.21	**0.15**
Francis	0.920	0.919	32.19	32.41	0.153	0.155	0.14	**0.09**
Train	0.815	0.806	22.05	21.93	0.206	0.221	0.26	**0.16**
Ignatius	0.928	0.925	30.17	30.22	0.084	0.089	0.33	**0.20**
Barn	0.951	0.952	30.35	31.98	0.073	0.072	0.27	**0.16**
Bicycle	0.663	0.637	23.56	23.39	0.335	0.368	0.85	**0.51**
Flowers	0.528	0.501	20.55	20.22	0.415	0.444	0.60	**0.35**
Garden	0.828	0.809	26.20	25.99	0.163	0.201	1.10	**0.59**
Stump	0.722	0.701	25.62	25.37	0.294	0.327	0.86	**0.66**
Treehill	0.587	0.572	22.07	22.04	0.417	0.442	0.56	**0.33**
Average	0.787	0.775	26.12	26.21	0.225	0.243	**0.52**	**0.32**

Table 3. Performance and memory usage evaluation of $\mathbf{3DGF_{AGF}}$ on outdoor scenes dataset. The table demonstrates AGF's ability to achieve substantial memory savings while maintaining competitive reconstruction quality.

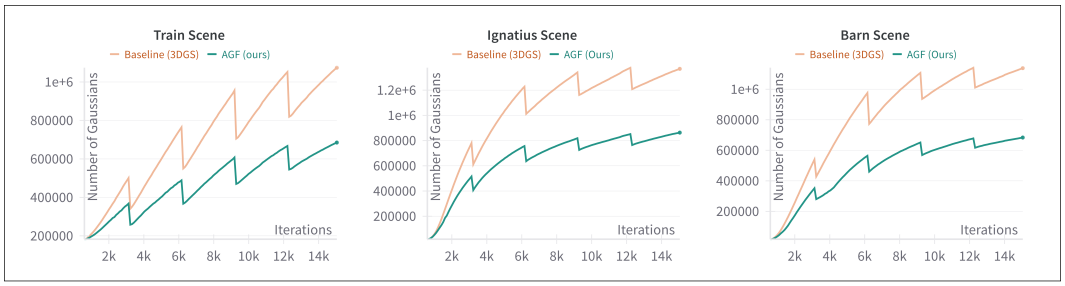

Figure 4. Comparison of the **number of Gaussians** over optimization iterations between the baseline 3DGS and our AGF module on outdoor scenes, demonstrating significant memory efficiency with AGF across different scenes.

that AGF effectively controls memory growth, enabling high-quality reconstructions with improved efficiency. The combined effect of DAPS and AGF leads to more accurate reconstructions with fewer artifacts and enhanced efficiency, validating the strength of our proposed enhancements in both reconstruction quality and resource management.

3DGS with DAPS. To assess the contribution of our Depth-Aware Perceptual Similarity module in isolation, we compare $3DGS_{DAPS}$ against the original 3DGS baseline across all outdoor scenes (Table 1). The results show that integrating DAPS alone leads to consistent improvements in structural similarity (SSIM), peak signal-to-noise ratio (PSNR), and perceptual quality (LPIPS).

These gains are particularly pronounced in scenes with larger depth variation and sparse multi-view coverage, where under-optimized regions are more common. By modulating perceptual similarity based on depth cues, DAPS encourages the network to better supervise such regions during optimization.

3DGS with AGF. To evaluate the impact of our Adaptive Gradient Filtering (AGF) module, we compare $3DGS_{AGF}$ against the baseline 3DGS across all benchmark scenes (Table 3). Results show that $3DGS_{AGF}$ reduces memory consumption by more than 38% on average, while maintaining comparable SSIM, PSNR, and LPIPS scores to the baseline. This demonstrates that AGF effectively controls memory growth without compromising reconstruction quality, making it a compelling choice for large-scale or resource-constrained scenarios.

To further illustrate this behavior, Figure 4 presents the number of Gaussians over optimization iterations. The baseline model exhibits uncontrolled growth due to fixed densification thresholds, leading to excessive memory usage. In contrast, AGF produces a smoother, more compact growth curve by suppressing redundant Gaussian replication in well-optimized regions.

6.1. Ablation Studies

To thoroughly evaluate our proposed components, we conducted a series of ablation studies that analyze the individ-

DAPS	AGF	SSIM	PSNR	LPIPS
✗	✗	0.928	30.17	0.084
✓	✗	0.954	31.98	0.046
✗	✓	0.925	30.22	0.089
✓	✓	**0.956**	**32.38**	**0.045**

Table 4. Ablation results on the Ignatius scene using different combinations of our proposed DAPS and AGF modules.

Approach	Memory (GB)
3DGS (Baseline)	0.33
Ours (AGF)	**0.20**

Table 5. Memory consumption comparison on the Ignatius scene.

Scene	PSNR (↑)		SSIM (↑)	
	3DGS	Ours	3DGS	Ours
Family	28.39	**30.39**	0.925	**0.954**
Treehill	**22.07**	21.72	0.587	**0.604**

Table 6. PSNR and SSIM comparison between baseline (3DGS) and our method (DAPS+AGF) on Family and Treehill scenes.

(a) Family Scene (b) Treehill Scene

Figure 5. Example images from the Family and Treehill scenes.

ual contributions of DAPS and AGF on reconstruction quality and memory efficiency.

Impact of DAPS and AGF Modules. Table 4 summarizes the performance impact of different configurations using the Ignatius scene from the outdoor scenes dataset. Integrating only DAPS significantly boosts reconstruction metrics (SSIM, PSNR, and LPIPS), reflecting improved perceptual and structural details. On the other hand, adding only AGF maintains comparable quality to the baseline, confirming that the memory optimization through adaptive gradient filtering does not degrade visual performance. Combining both modules (DAPS + AGF) achieves the best overall performance, indicating that our proposed methods are complementary.

Memory Efficiency Analysis. Table 5 highlights the effectiveness of AGF in controlling memory usage, showing a substantial reduction (approximately 39%) compared to the 3DGS baseline without sacrificing reconstruction quality. This demonstrates that AGF successfully constrains the number of Gaussians during optimization, ensuring efficient memory allocation.

Perceptual vs. Pixel-Level Response to Depth-Aware Supervision. Table 6 shows that while both Family and Treehill scenes gain perceptual improvements (higher SSIM) with DAPS, only Family sees a PSNR boost. This is because the Family scene (Figure. 5 (a)) has distant regions that are composed of well-defined, rigid structures with reliable depth estimation, allowing DAPS to enhance both structural and pixel-level reconstruction.

In contrast, Treehill scene (Figure. 5 (b)) is dominated by natural foliage that often lacks sharp depth boundaries. In such regions, monocular depth estimation becomes less reliable. Consequently, while DAPS improves perceptual structure (SSIM), it may introduce pixel-level inconsistencies, reducing PSNR. This highlights a limitation of applying depth-based weighting in depth-ambiguous scenes.

7. Conclusion

In this work, we identified a key limitation in 3D Gaussian Splatting, that is, poor reconstruction quality in under-supervised regions such as scene backgrounds and peripheral structures, often caused by limited viewpoint coverage. To address this, we proposed two complementary modules. Depth-Aware Perceptual Similarity (DAPS) improves optimization in these regions by using the depth cues to guide the optimization process, while Adaptive Gradient Filtering (AGF) reduces memory overhead by selectively regulating Gaussian cloning via dynamic gradient-based thresholds. Together, DAPS and AGF enhance both accuracy and efficiency. Our method improves SSIM by 3.05%, PSNR by 3.94%, and reduces LPIPS by 17.78% over the 3DGS baseline while maintaining the memory usage. AGF alone cuts memory usage by 38.5% over the baseline. These gains are achieved without compromising visual quality, demonstrating our approach's strength in reconstructing complex outdoor scenes. Extending these benefits to indoor environments with limited depth variation remains a promising direction for future work.

References

[1] Yanqi Bao, Tianyu Ding, Jing Huo, Yaoli Liu, Yuxin Li, Wenbin Li, Yang Gao, and Jiebo Luo. 3d gaussian splatting: Survey, technologies, challenges, and opportunities. *arXiv preprint arXiv:2407.17418*, 2024. 3

[2] 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. 5

[3] Zehao Cao, Zongji Wang, Yuanben Zhang, Cheng Jin, Weinan Cai, Zhihong Zeng, and Junyi Liu. Enhancing 3d gaussian splatting for low-quality images: semantically guided training and unsupervised quality assessment. *The Visual Computer*, pages 1–19, 2024. 2

[4] Guikun Chen and Wenguan Wang. A survey on 3d gaussian splatting. *arXiv preprint arXiv:2401.03890*, 2024. 1

[5] Youyu Chen, Junjun Jiang, Kui Jiang, Xiao Tang, Zhihao Li, Xianming Liu, and Yinyu Nie. Dashgaussian: Optimizing 3d gaussian splatting in 200 seconds. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 11146–11155, 2025. 2

[6] Yuedong Chen, Haofei Xu, Chuanxia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-Jen Cham, and Jianfei Cai. Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images. In *European Conference on Computer Vision*, pages 370–386. Springer, 2025. 2

[7] 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. 2

[8] Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps. *arXiv preprint arXiv:2311.17245*, 2023. 2, 3

[9] Umar Farooq, Jean-Yves Guillemaut, Adrian Hilton, and Marco Volino. Optimized 3d gaussian splatting using coarse-to-fine image frequency modulation. *arXiv preprint arXiv:2503.14475*, 2025. 2, 3

[10] Jun Gao, Wenzheng Chen, Tommy Xiang, Alec Jacobson, Morgan McGuire, and Sanja Fidler. Learning deformable tetrahedral meshes for 3d reconstruction. *Advances in neural information processing systems*, 33:9936–9947, 2020. 2

[11] Sharath Girish, Kamal Gupta, and Abhinav Shrivastava. Eagles: Efficient accelerated 3d gaussians with lightweight encodings. In *European Conference on Computer Vision*, pages 54–71. Springer, 2025. 2, 3

[12] 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. 1, 3, 5

[13] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36 (4):1–13, 2017. 5

[14] Pou-Chun Kung, Seth Isaacson, Ram Vasudevan, and Katherine A Skinner. Sad-gs: Shape-aligned depth-supervised gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2842–2851, 2024. 2

[15] Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3d gaussian representation for radiance field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21719–21728, 2024. 2, 3

[16] Priyanka Mandikal and Venkatesh Babu Radhakrishnan. Dense 3d point cloud reconstruction using a deep pyramid network. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019. 2

[17] Hidenobu Matsuki, Riku Murai, Paul HJ Kelly, and Andrew J Davison. Gaussian splatting slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18039–18048, 2024. 2

[18] KL Navaneet, Kossar Pourahmadi Meibodi, Soroush Abbasi Koohpayegani, and Hamed Pirsiavash. Compgs: Smaller and faster gaussian splatting with vector quantization. In *European Conference on Computer Vision*, pages 330–349. Springer, 2025. 2, 3

[19] Simon Niedermayr, Josef Stumpfegger, and Rüdiger Westermann. Compressed 3d gaussian splatting for accelerated novel view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10349–10358, 2024. 3

[20] Panagiotis Papantonakis, Georgios Kopanas, Bernhard Kerbl, Alexandre Lanvin, and George Drettakis. Reducing the memory footprint of 3d gaussian splatting. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1):1–17, 2024. 2, 3

[21] Stéphane Pateux, Matthieu Gendrin, Luce Morin, Théo Ladune, and Xiaoran Jiang. Bogauss: Better optimized gaussian splatting. *arXiv preprint arXiv:2504.01844*, 2025. 2, 3

[22] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence*, 44(3):1623–1637, 2020. 4

[23] Haofei Xu, Songyou Peng, Fangjinhua Wang, Hermann Blum, Daniel Barath, Andreas Geiger, and Marc Pollefeys. Depthsplat: Connecting gaussian splatting and depth. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 16453–16463, 2025. 2

[24] Qilin Zhang, Olaf Wysocki, Steffen Urban, and Boris Jutzi. Cdgs: Confidence-aware depth regularization for 3d gaussian splatting. *arXiv preprint arXiv:2502.14684*, 2025. 2

[25] Zhaoliang Zhang, Tianchen Song, Yongjae Lee, Li Yang, Cheng Peng, Rama Chellappa, and Deliang Fan. Lp-3dgs: Learning to prune 3d gaussian splatting. *arXiv preprint arXiv:2405.18784*, 2024. 2, 3

[26] Huixin Zhu, Zhili Zhang, Junyang Zhao, Hui Duan, Yao Ding, Xiongwu Xiao, and Junsong Yuan. Scene reconstruction techniques for autonomous driving: a review of 3d gaussian splatting. *Artificial Intelligence Review*, 58:30, 2024. 2