Geometry-Propagated Gaussian Splatting for Aerial Sparse Novel View Synthesis

Yijing Wang[1,2], Xu Tang[1]*, Jingjing Ma[1], and Xiangrong Zhang[1]

[1] School of Artificial Intelligence, Xidian University, Xi'an, China
[2] University of Extremadura, Cáceres, Spain
yijingwang@stu.xidian.edu.cn, tangxu128@gmail.com

Abstract. 3D Gaussian Splatting achieves impressive novel view synthesis for ground-level scenes. However, its performance degrades significantly in aerial domains due to sparse viewpoint sampling imposed by platform constraints. Existing sparse-view methods typically rely on structure-from-motion reconstruction, which often yields incomplete geometry due to limited feature correspondences under repetitive textures and sparse viewpoint overlap. We present Geometry-Propagated Gaussian Splatting (GeoProp-GS), which overcomes the limitations of insufficient initialization. Our approach introduces two core components: Depth-guided Geometric Initialization (DGI) generates dense point clouds and extends coverage to unobserved regions; Anchor-constrained Gaussian Optimization (AGO) stabilizes under-supervised regions by decomposing proposal Gaussians into reliable anchors and learnable residuals. Extensive experiments demonstrate that GeoProp-GS achieves state-of-the-art performance and serves as an effective plug-and-play module that consistently improves existing methods. Code available at *GeoProp-GS*.

Keywords: Aerial Photogrammetry · Aerial-view Sparse Novel View Synthesis · 3D Gaussian Splatting

1 Introduction

3D Gaussian Splatting (3DGS) has emerged as the dominant paradigm for novel view synthesis (NVS). It achieves real-time photorealistic rendering through explicit 3D Gaussian primitives [1, 5, 12, 13]. Its explicit, scene-specific representation is particularly well-suited for aerial mapping applications such as infrastructure inspection and autonomous navigation [10, 16, 50], where controllable rendering, geometric interpretability, and on-device deployment are essential. However, extending 3DGS to aerial domains [26, 37] exposes a structural tension. The very platform constraints that motivate aerial deployment (limited UAV endurance, satellite orbital windows, weather dependency, and regulatory airspace restrictions) inherently force sparse viewpoint sampling [7, 21]. This sparsity directly conflicts with 3DGS's need for dense multi-view observations, causing incomplete geometry, floater artifacts, and severe rendering degradation in under-observed regions [34, 44, 48].

* Corresponding author.

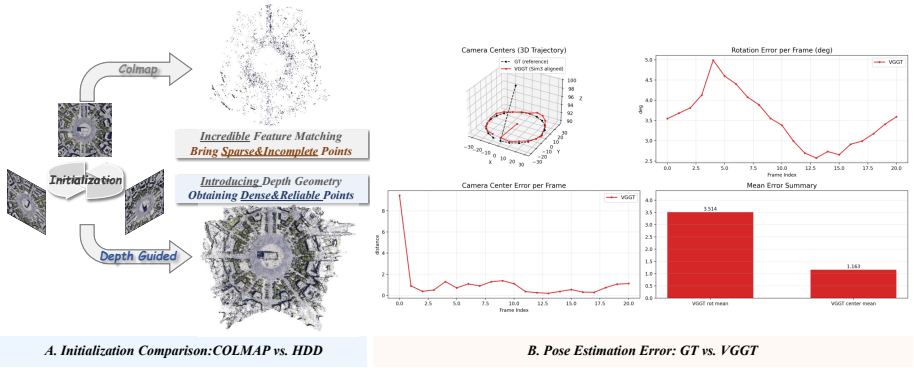

Fig. 1: Analysis of the fundamental bottleneck in sparse-view aerial 3DGS. (a) COLMAP-based initialization often yields sparse and incomplete point clouds under repetitive aerial textures, while our HDD produces dense, metric-consistent point clouds with full scene coverage. (b) Pose estimation comparison on aerial *Building* 0: COLMAP initialized with ground-truth poses and refined via BA achieves near-perfect accuracy, whereas VGGT suffers from large rotation and translation errors without BA. This confirms that the bottleneck lies in feature matching collapse, not the optimization strategy.

Existing sparse-view NVS methods fall into two broad paradigms. Optimization-based methods adopt strategies including depth-based regularization [4, 15, 22, 41, 42, 53, 55], optimization regularization [8, 25, 27, 43, 51], and diffusion model guidance [6, 17, 18, 32, 39, 49, 54], most building upon reliable point clouds produced by SfM pipelines such as COLMAP [24, 29–31] as geometric scaffolds. Feed-forward methods [35, 36] take a different route, directly regressing scene representations from image features.

Aerial-view scenes present a qualitatively different challenge for both paradigms. Repetitive and low-texture patterns, such as flat rooftops, uniform roads, and similar facades (see Fig. 1), yield few discriminative keypoints. Rotational UAV trajectories and wide-baseline satellite passes further limit inter-view overlap. Under these conditions, COLMAP frequently produces severely incomplete point clouds, stripping optimization-based methods of their geometric scaffolds. Feed-forward methods fare no better: homogeneous aerial textures substantially degrade joint estimation of camera poses and scene geometry, yielding degenerate scene representations. We verify this empirically in Sec. 4.3 and 4.4: both COLMAP-based initialization and representative feed-forward methods [35, 36] yield substantially degraded NVS quality under sparse aerial views (Tables 3 and 4), confirming that the bottleneck lies in the collapse of feature matching rather than the choice of optimization strategy.

To address these challenges, we present **Geometry-Propagated Gaussian Splatting (GeoProp-GS)**, a robust sparse-view aerial 3DGS framework built around two core principles: complete geometric initialization and geometry

propagation from observed to unobserved regions. We introduce two core components. First, Depth-guided Geometric Initialization (DGI) ensures complete scene coverage from the initialization stage. For observed regions, Hierarchical Depth-guided Densification (HDD) produces dense, metric-consistent point clouds that replace incomplete SfM outputs. For unobserved regions, Geometry-aware Viewpoint Augmentation (GVA) synthesizes pseudo-views and generates geometric proposals, extending initialization coverage beyond the input viewpoints. Second, Anchor-constrained Gaussian Optimization (AGO) tackles the challenge of optimizing Gaussians in regions that lack direct photometric supervision. It decomposes these Gaussians into reliable anchors and learnable proposals, effectively propagating geometric knowledge from well-observed regions into under-supervised areas. Together, DGI and AGO form a cohesive pipeline where reliable geometry is first established, then propagated, ensuring robust reconstruction across the full scene. Extensive experiments on LEVIR-NVS [40] and 3D-AS [33] demonstrate that GeoProp-GS achieves state-of-the-art NVS quality under sparse aerial viewpoint constraints.

Our main contributions are summarized as follows:

- We identify the fundamental bottleneck of sparse-view aerial 3DGS: the collapse of geometric initialization under repetitive textures and limited viewpoint overlap. We propose GeoProp-GS to address this through complete initialization and geometry propagation across the scene.
- To address the lack of reliable geometric initialization in aerial sparse-view settings, we propose DGI. HDD densifies observed regions into dense, metric-consistent point clouds, while GVA extends coverage to unobserved regions by synthesizing pseudo-views and generating geometric proposals.
- To address the challenge of optimizing Gaussians in regions without direct photometric supervision, we propose AGO. It decomposes proposal Gaussians into geometric anchors and learnable residuals, propagating reliable geometric knowledge from observed regions into under-supervised areas.
- Extensive experiments on LEVIR-NVS and 3D-AS demonstrate state-of-the-art NVS quality under sparse aerial viewpoint constraints. HDD further serves as a plug-and-play module that consistently improves existing 3DGS methods.

2 Related Work

2.1 3DGS under Sparse Views

Sparse-view 3DGS methods fall into two broad paradigms. Optimization-based methods further divide into three strategies. Geometry-guided methods leverage depth-based priors to enforce multi-view consistency [4, 15, 22, 41, 42, 53, 55], effectively resolving scale ambiguities and stabilizing geometry in limited-view scenes. Generative approaches employ diffusion models to synthesize novel viewpoints and inpaint unobserved regions [6, 17, 18, 23, 32, 39, 49, 54], compensating

for incomplete scene coverage. Optimization-centric strategies introduce cross-view consistency constraints and selective pruning to suppress unstable Gaussians without relying on external priors [8, 25, 27, 43, 51]. Almost three strategies build upon reliable point clouds from SfM pipelines such as COLMAP as geometric scaffolds. Feed-forward methods take a different route, directly regressing Gaussian parameters from image features in a single forward pass without per-scene optimization. MVSplat [3] constructs a cost volume via plane sweeping to localize Gaussian centers, learning all Gaussian parameters under photometric supervision alone. NoPoSplat [47] further removes the pose requirement by anchoring one input view as the canonical space and predicting Gaussians for all views within it.

Despite their effectiveness in ground-level scenes, both paradigms struggle in aerial settings for distinct reasons. Optimization-based methods depend on SfM pipelines for geometric scaffolds. However, repetitive structures and low-texture surfaces yield few discriminative keypoints, causing COLMAP to collapse. Feed-forward methods face a different obstacle as homogeneous aerial textures prevent reliable estimation of camera poses and scene geometry. Both paradigms therefore leave 3DGS without the geometric foundation needed for robust sparse-view aerial reconstruction.

2.2 Monocular Depth Estimation

The evolution of monocular depth estimation has broadly followed a trajectory from early discriminative regression to large-scale foundation models and, more recently, generative diffusion-based approaches. Early discriminative methods [2, 28] formulated depth estimation as direct dense regression, achieving reasonable in-domain performance while remaining less robust under cross-domain settings and a tendency to produce over-smoothed predictions that lose fine-grained structural details. To address the generalization bottleneck, foundation model approaches [45, 46] adopted large-scale data engines combining millions of labeled and unlabeled images. In detail, Depth Anything V1 leverages over 62M unlabeled images via semi-supervised training, while V2 further improves prediction sharpness and robustness through synthetic-to-real teacher-student distillation. The two models yield strong zero-shot relative depth estimation, though absolute metric scale recovery remains an inherent limitation of purely discriminative pipelines. Diffusion-based methods [11] represent a paradigm shift by reformulating depth estimation as a conditional denoising process: Marigold, derived from Stable Diffusion and fine-tuned solely on synthetic data, harnesses the rich visual priors of generative image models to produce affine-invariant depth maps with exceptional detail and zero-shot generalizability across diverse domains.

For our pipeline, we adopt Depth Anything V2 [46] as our monocular depth prior. This choice is motivated by its robust performance on outdoor and aerial scenes without scene-specific finetuning, and its ability to preserve critical structural boundaries essential for Gaussian initialization. Compared to slower diffusion-

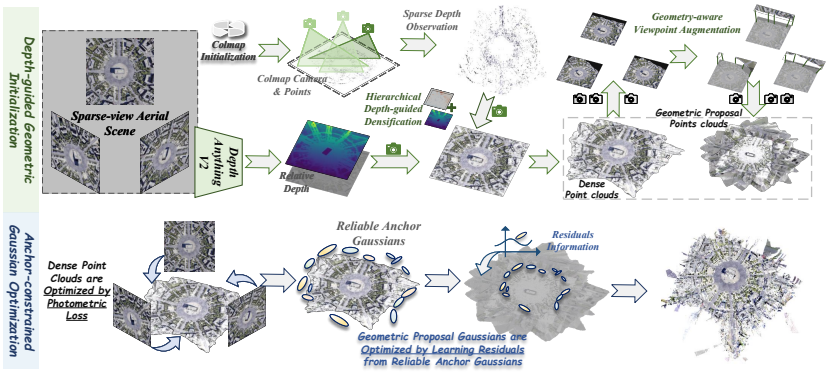

Fig. 2: Overview of GeoProp-GS. DGI converts sparse aerial inputs into a dense initialization $\mathcal{C}_{\text{init}}$ via hierarchical depth-guided densification and geometry-aware viewpoint augmentation, while AGO optimizes reliable Gaussians and anchor-constrained proposal Gaussians for stable NVS in under-constrained regions.

based alternatives [11], it offers a practical balance between quality and efficiency for sparse-view aerial NVS.

3 Methodology

Given a sparse set of aerial-view images, our goal is to reconstruct a photorealistic 3D Gaussian representation for high-quality novel view synthesis under a limited observation setting. As identified in Section 1, sparse-view aerial 3DGS faces two critical challenges. First, both SfM pipelines and feed-forward methods fail to provide reliable geometric initialization under repetitive aerial textures. Second, limited viewpoint coverage leaves large regions of the scene unobserved. To address these challenges, we propose GeoProp-GS as shown in Fig. 2. It establishes complete and reliable scene initialization, then propagates geometric knowledge from observed regions into unobserved areas. It consists of two core components. DGI (Sec. 3.2) initializes observed regions through HDD and extends scene coverage to unobserved regions via GVA. Gaussians initialized from observed regions serve as reliable Gaussians $\mathcal{G}_{\text{reliable}}$, while those generated for unobserved regions form proposal Gaussians $\mathcal{G}_{\text{prop}}$ requiring further stabilization. AGO (Sec. 3.3) stabilizes $\mathcal{G}_{\text{prop}}$ by decomposing them into geometry-constrained anchors and learnable residuals, propagating reliable geometric knowledge from $\mathcal{G}_{\text{reliable}}$.

3.1 Preliminaries

3DGS [12] represents scenes as collections of 3D Gaussian primitives optimized through differentiable rendering. Each Gaussian primitive is characterized by a center position $\mathbf{x} \in \mathbb{R}^3$, a covariance matrix parameterized by anisotropic scaling

$\mathbf{s} \in \mathbb{R}^3$ and rotation $\mathbf{q} \in \mathbb{R}^4$, opacity $o \in \mathbb{R}$, and view-dependent appearance encoded by spherical harmonics coefficients $\mathbf{c}$. The framework relies on two critical aspects, namely geometric initialization and photometric supervision. For initialization, 3DGS employs SfM pipelines such as COLMAP [30] to generate a sparse 3D point cloud $\{\mathbf{p}_j\}_{j=1}^{M}$ via keypoint matching and triangulation. These SfM points serve as the initial Gaussian centers. During rendering, Gaussian primitives are projected onto the image plane and composited via alpha-blending in depth order, enabling fully differentiable rasterization. For supervision, all Gaussian parameters are optimized by minimizing photometric loss between rendered and training images [12]. Importantly, this supervision is only provided by pixels observed in the training views: gradients propagate solely from pixels with direct photometric evidence. This makes the quality of geometric initialization particularly critical—incomplete or inaccurate initialization can lead to floater artifacts and degraded rendering in under-observed regions.

3.2 Depth-guided Geometric Initialization

To address the dual challenges of unreliable geometric initialization and limited viewpoint coverage, we introduce DGI, a two-stage initialization pipeline that establishes complete scene coverage without relying on feature matching. In the first step, HDD replaces unreliable SfM point clouds with dense, metric-consistent geometry by integrating monocular depth priors with COLMAP reconstructions. In the second step, GVA synthesizes pseudo-views for unobserved regions through geometry-conditioned inpainting, generating proposal points that extend initialization coverage beyond the input viewpoints.

Hierarchical Depth-guided Densification. The goal of HDD is to generate dense and reliable 3D point clouds from observed viewpoints by integrating monocular depth predictions with COLMAP reconstructions. We first perform SfM using COLMAP to extract camera extrinsics $\mathbf{T}_{w\leftarrow c,i} = (\mathbf{R}_{wc,i}, \mathbf{t}_{wc,i}) \in \mathrm{SE}(3)$, intrinsics $\mathbf{K}_i$, and 3D points $\mathcal{P}_{\mathrm{colmap}}$. To establish metric-scale supervision, we project these sparse 3D points into each view's image plane, yielding sparse depth observations $\mathbf{D}_{\mathrm{sparse}}$ with a binary mask $\mathbf{M}_{\mathrm{sparse}}$ indicating valid locations. These sparse depth observations serve as geometric reference points for aligning monocular predictions with the metric scale of the reconstructed scene. Then, relative depth maps $\hat{\mathbf{D}}_i$ are estimated using Depth Anything V2 [46].

These predictions are scale-ambiguous and require metric alignment before lifting to 3D. To this end, we employ a hierarchical correction strategy (HCS) that progressively refines $\hat{\mathbf{D}}_i$ into metrically accurate depth maps consistent with $\mathcal{P}_{\mathrm{colmap}}$. At the global level transformation, we optimize scale factor α and shift β via repeated least-squares alignment [28] with sparse depth observations $\mathbf{D}_{\mathrm{sparse}}$:

$$(\alpha, \beta) = \arg\min_{\alpha,\beta} \left\| \mathbf{M}_{\mathrm{sparse}} \odot \left(\alpha \hat{\mathbf{D}}_i + \beta - \mathbf{D}_{\mathrm{sparse}} \right) \right\|_2^2, \tag{1}$$

yielding the globally aligned depth $\mathbf{D}_{\mathrm{global}} = \alpha \cdot \hat{\mathbf{D}}_i + \beta$, where $\odot$ denotes element-wise multiplication. However, a global level transformation cannot capture spa-

tially varying geometric distortions in monocular predictions. We therefore introduce a local correction field by computing residual depth deviations between $\mathbf{D}_{\text{global}}$ and $\mathbf{D}_{\text{sparse}}$ at sparse SfM point locations. To propagate these corrections to dense pixels while maintaining spatial coherence, distance-weighted nearest-neighbor interpolation is employed, where the correction magnitude decays exponentially with distance $d(u, v)$ to the nearest sparse point. This formulation ensures that reconstructed 3D points near sparse reference points are closely aligned with $\mathcal{P}_{\text{colmap}}$, while regions far from reference points gracefully degrade to the globally aligned solution.

The final corrected depth map $\mathbf{D}_i$ combines the global alignment with smoothed local corrections, while strictly preserving the original sparse depth values at observed locations where $\mathbf{M}_{\text{sparse}}(u, v) = 1$. We further apply median filtering at image boundaries to reduce extrapolation artifacts. Each pixel (u, v) with corrected depth $\mathbf{D}_i(u, v)$ is then lifted to 3D space through standard perspective unprojection using camera intrinsics $\mathbf{K}_i$, followed by transformation to world coordinates via the camera extrinsics $\mathbf{T}_{w\leftarrow c,i}$. This generates dense point clouds for each view that are geometrically consistent with SfM reconstruction, which are merged to obtain the initialized set $\mathcal{C}_{\text{dense}}$.

Geometry-aware Viewpoint Augmentation. Limited viewpoint coverage leaves large unobserved regions that sparse-view initialization cannot reach. GVA addresses this by synthesizing pseudo-observations through geometry-conditioned inpainting. Given a pseudo pose $\mathbf{P}_h$, $\mathcal{C}_{\text{dense}}$ is projected into a partial RGB image $\mathbf{I}_{\text{partial}}$ and an occlusion mask $\mathbf{M}$ that reflects the underlying scene geometry. Missing regions are then completed via diffusion-based inpainting [14]: $\mathbf{I}_{\text{complete}} = \text{Inpaint}\big(\mathbf{I}_{\text{partial}}, \mathbf{M}\big)$.

Inpainted RGB appearance may be imprecise, as diffusion models hallucinate plausible textures rather than reconstruct ground-truth appearance. Fortunately, monocular depth estimators can infer 3D structure from visual context rather than precise texture, and thus remain geometrically consistent when applied to $\mathbf{I}_{\text{complete}}$. The same HCS is therefore applied to the Depth Anything V2 depth predicted from $\mathbf{I}_{\text{complete}}$, yielding a corrected pseudo-depth map, from which proposal points $\mathcal{C}_{\text{proposal}}$ are generated for unobserved regions. Overlaps with $\mathcal{C}_{\text{dense}}$ are removed, yielding the final initialization $\mathcal{C}_{\text{init}} = \mathcal{C}_{\text{dense}} \cup \mathcal{C}_{\text{proposal}}$.

Discussion: Sparse-view initialization typically suffers from two complementary issues: insufficient geometric density within observed regions and limited coverage beyond them. HDD addresses the first by replacing sparse SfM points with dense, metrically consistent geometry, while GVA resolves the second by extending initialization into previously unobserved regions. The resulting $\mathcal{C}_{\text{init}}$ therefore provides both geometric fidelity and spatial completeness for stable 3DGS optimization. Compared with the recent method [23] that integrates generative repair or inpainting within iterative optimization pipelines, our design focuses on constructing a complete geometric scaffold at initialization. Depth information is used to produce metrically aligned geometry rather than to supervise training, enabling reliable structure to be established before optimization begins and reducing the risk of persistent floaters or geometric gaps.

3.3 Anchor-constrained Gaussian Optimization

Our initialization yields two types of Gaussian points with fundamentally different supervision characteristics. As discussed in Sec. 3.2, points from HDD ($\mathcal{C}_{\text{dense}}$) correspond to real observed views with ground-truth RGB supervision, forming *reliable Gaussians* $\mathcal{G}_{\text{reliable}}$ that can be directly optimized via photometric losses. In contrast, points from GVA ($\mathcal{C}_{\text{proposal}}$) originate from synthesized pseudo-views lacking reliable appearance information, yielding *proposal Gaussians* $\mathcal{G}_{\text{prop}}$ that are unsuitable for unconstrained conventional photometric supervision. For reliable Gaussians, standard 3DGS optimization [12] works well as they receive dense photometric gradients from training views. Proposal Gaussians occupy regions that are almost entirely unobserved by training views, resulting in sparse supervision and thus a high risk of geometric drift or degenerate configurations.

To stabilize these under-supervised regions without explicit geometric regularization losses, we adopt an anchor–residual parameterization that embeds geometric priors directly in the representation. Specifically, each proposal Gaussian $g_k \in \mathcal{G}_{\text{prop}}$ is expressed as a learnable deviation from its nearest reliable *anchor* $g_a \in \mathcal{G}_{\text{reliable}}$:

$$g_k = g_a + \Delta g_k, \quad a = \arg\min_{j \in \mathcal{G}_{\text{reliable}}} \|\mathbf{x}_k - \mathbf{x}_j\|_2. \tag{2}$$

Concretely, we decompose all Gaussian attributes into anchor values and learnable residuals. Each Gaussian is characterized by position $\mathbf{x} \in \mathbb{R}^3$, spherical harmonics coefficients $\mathbf{c}$ for view-dependent appearance, covariance parameters (scale $\mathbf{s} \in \mathbb{R}^3$ and rotation $\mathbf{q} \in \mathbb{R}^4$), and opacity $o \in \mathbb{R}$. For each proposal Gaussian k,

$$\begin{aligned} \mathbf{x}_k = \mathbf{x}_a + \Delta\mathbf{x}_k, \quad \mathbf{c}_k = \mathbf{c}_a + \Delta\mathbf{c}_k, \\ \mathbf{s}_k = \mathbf{s}_a + \Delta\mathbf{s}_k, \quad \mathbf{q}_k = \mathbf{q}_a + \Delta\mathbf{q}_k, \quad \mathbf{o}_k = \mathbf{o}_a + \Delta\mathbf{o}_k, \end{aligned} \tag{3}$$

where the subscript a denotes the anchor attributes and $\Delta(\cdot)$ are the learnable residuals. For clarity of notation, the activation functions applied to certain parameters are omitted in the formulation. In implementation, the additive decomposition is applied in pre-activation parameter space, and standard activations are then used to keep attributes within valid ranges.

During optimization, proposal parameters are reconstructed on-the-fly by composing anchor values with residuals, so each proposal inherits the geometric and appearance priors of its nearest reliable anchor. Proposal Gaussians receive less photometric supervision, as they predominantly occupy unobserved regions. Anchor parameters are primarily optimized by reliable photometric gradients from observed views. This induces implicit regularization: proposals with insufficient photometric support are naturally constrained to remain close to their anchors, while proposals with occasional valid photometric support accumulate residuals to fit the observed appearance. Additionally, we apply a proposal color statistical regularizer that aligns the mean and variance of proposal Gaussians' DC color coefficients with those of all Gaussians, reducing color drift in weakly

supervised regions. Together, these mechanisms ensure proposals remain stably anchored to reliable geometry throughout optimization.

Discussion: The anchor–residual parameterization imposes geometric constraints architecturally rather than through additional loss terms. Each proposal is initialized at its original position, with residuals set to capture the initial offset from its nearest anchor. During optimization, reliable photometric supervision from observed views primarily refines the anchor-residual parameterization representation, providing stable geometric references for associated proposals. This adaptive behavior emerges without dedicated proposal-photometric specific supervision terms, improving robustness to varying scene coverage.

4 Experiments

Due to space limitations, additional experimental settings, experimental analyses, and discussions of limitations are provided in the supplementary material.

4.1 Experimental Setup

Datasets and Evaluation Metrics. We conducted experiments on two aerial datasets, with NVS quality measured using PSNR, SSIM [38], and LPIPS [52]. LEVIR-NVS [40] captures 16 diverse scenes from nadir (vertically downward) perspectives. Following [7], 3 training views per scene were used for extreme sparse-view NVS evaluation. To assess generalization across different viewing conditions, we further evaluated on 3D-AS [33], which provides 9 scenes captured from oblique orbital perspectives with challenging geometric complexity and wide viewing angles, spanning "*City*", "*Country*", and "*Port*" environments. For this dataset, 7 training views per scene were employed.

Experiment Settings. Our architecture and optimization followed FSGS [55] without additional depth supervision. Pruning and densification were disabled. The COLMAP points used in our baseline corresponded to the dense point clouds adopted in FSGS. For comparative experiments (Sec. 4.2), we followed the original experimental protocols of the baseline methods to ensure fair comparisons. For ablation and analysis experiments (Sec. 4.3-Sec. 4.4), we conducted all experiments on LEVIR-NVS, as it represents a more challenging sparse-view scenario with only three training views. Additional implementation details are provided in the supplementary material.

4.2 Comparative Results Analysis

To evaluate our method, we conducted comparisons with baseline approaches spanning five categories: (1) fundamental methods (3DGS [12]); (2) sparse-view NeRF techniques (DietNeRF [9], RegNeRF [20], FreeNeRF [44], PixelNeRF [48]); (3) sparse-view 3DGS methods (FSGS [55], DNGaussian [15], DropoutGS [43], Guidedvd-3DGS [54]); (4) feed-forward 3DGS methods evaluated in both zero-shot and fine-tuned settings (NoPoSplat [47], MVSplat [3]); and (5) the aerial-specific MPNeRF [7]. Feed-forward methods were evaluated on LEVIR-NVS

Table 1: Quantitative comparison with baseline methods on LEVIR-NVS dataset (3 training views). The best, second-best, and third-best entries are marked in , , and , respectively.

Metrics	Methods	Bld.0	Chr.0	Col.0	Mtn.0	Mtn.1	Obs.0	Bld.1	Twn.0	Std.0	Twn.1	Mtn.2	Twn.2	Fct.0	Prk.0	Sch.0	Dwn.0	**Mean**
PSNR	DietNeRF [9]	13.44	12.20	14.86	19.34	18.67	15.27	13.73	15.78	16.68	14.21	20.66	14.73	16.73	16.55	14.97	13.61	15.71
	PixelNeRF ft [48]	12.44	11.76	11.74	17.42	17.15	14.44	11.18	15.86	19.65	16.09	23.99	15.24	13.02	15.70	13.86	13.86	15.21
	RegNeRF [20]	12.07	10.79	12.60	16.39	17.36	13.04	11.92	12.94	13.49	11.59	19.37	12.21	12.66	13.98	12.71	11.21	13.40
	FreeNeRF [44]	13.56	11.01	13.93	20.03	19.74	15.29	13.00	16.29	15.47	13.21	21.59	13.15	18.91	18.23	13.35	11.87	15.54
	MPNeRF [7]	18.81	17.93	20.71	25.50	24.92	19.56	18.64	21.59	22.08	21.20	28.57	21.41	22.61	23.57	20.71	19.73	21.72
	3DGS [12]	16.22	16.13	18.73	23.84	22.54	17.56	16.69	20.23	21.37	20.89	27.50	20.78	21.95	22.30	19.18	17.35	20.20
	FSGS [55]	17.94	17.59	20.00	24.29	24.13	18.12	18.11	21.06	22.12	19.84	25.67	20.22	22.99	23.82	19.64	19.30	20.93
	DNGaussian [15]	15.89	13.78	16.74	19.41	19.02	15.62	16.82	18.21	17.68	16.95	20.55	17.37	20.46	17.70	17.25	16.24	17.48
	DropoutGS [43]	17.54	15.37	18.44	22.11	19.96	15.62	18.17	20.55	19.86	19.88	22.55	18.18	21.69	22.83	18.45	17.84	19.31
	Guidedvd-3DGS [54]	15.47	14.31	16.66	23.97	21.61	12.87	14.26	17.48	17.29	14.99	23.88	18.64	15.22	16.95	15.47	13.77	17.05
	NoPoSplat [47]	13.99	13.49	15.91	19.21	20.28	16.57	15.39	15.93	19.22	13.24	19.35	14.04	14.38	15.27	15.52	13.37	15.95
	NoPoSplat ft [47]	15.72	16.84	20.10	22.27	21.99	17.86	16.94	19.91	21.21	18.72	25.99	19.19	22.09	20.70	18.58	17.96	19.75
	MVSplat [3]	6.74	7.17	8.27	12.84	11.39	7.67	7.05	8.34	6.92	7.09	13.49	8.04	6.39	7.90	5.55	7.06	8.24
	MVSplat ft [3]	14.36	13.95	16.22	20.07	20.00	15.54	14.65	17.91	18.51	16.54	21.80	18.37	18.97	18.48	16.74	14.63	17.30
	Ours	19.99	19.85	22.44	27.33	26.85	21.52	20.41	24.90	25.43	24.65	31.78	24.75	27.86	27.45	23.75	20.71	24.35
SSIM	DietNeRF [9]	0.16	0.15	0.23	0.34	0.24	0.21	0.17	0.18	0.28	0.21	0.35	0.26	0.36	0.26	0.28	0.19	0.24
	PixelNeRF ft [48]	0.14	0.20	0.16	0.29	0.30	0.29	0.16	0.30	0.51	0.47	0.48	0.34	0.25	0.26	0.29	0.25	0.29
	RegNeRF [20]	0.12	0.11	0.16	0.28	0.27	0.14	0.13	0.13	0.20	0.13	0.34	0.16	0.20	0.16	0.21	0.11	0.18
	FreeNeRF [44]	0.24	0.11	0.21	0.37	0.33	0.26	0.17	0.28	0.27	0.23	0.38	0.21	0.51	0.42	0.24	0.14	0.27
	MPNeRF [7]	0.73	0.72	0.79	0.82	0.81	0.73	0.71	0.81	0.80	0.84	0.89	0.84	0.86	0.85	0.79	0.76	0.80
	3DGS [12]	0.69	0.57	0.58	0.79	0.82	0.59	0.64	0.75	0.78	0.76	0.84	0.71	0.86	0.87	0.64	0.64	0.72
	FSGS [55]	0.69	0.62	0.63	0.78	0.78	0.59	0.65	0.73	0.78	0.72	0.77	0.69	0.83	0.84	0.65	0.70	0.71
	DNGaussian [15]	0.54	0.30	0.44	0.48	0.44	0.38	0.53	0.53	0.49	0.47	0.45	0.53	0.70	0.52	0.53	0.51	0.49
	DropoutGS [43]	0.57	0.43	0.55	0.55	0.41	0.20	0.58	0.60	0.58	0.65	0.51	0.57	0.73	0.74	0.61	0.61	0.56
	Guidedvd-3DGS [54]	0.44	0.38	0.49	0.69	0.56	0.20	0.33	0.53	0.53	0.44	0.61	0.59	0.51	0.46	0.48	0.28	0.47
	NoPoSplat [47]	0.17	0.17	0.26	0.33	0.36	0.30	0.22	0.20	0.41	0.17	0.34	0.23	0.25	0.19	0.31	0.13	0.25
	NoPoSplat ft [47]	0.47	0.53	0.60	0.48	0.55	0.52	0.51	0.58	0.70	0.57	0.64	0.54	0.80	0.63	0.51	0.57	0.57
	MVSplat [3]	0.02	0.02	0.03	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.07	0.04	0.03	0.03	0.03	0.03	0.03
	MVSplat ft [3]	0.34	0.27	0.30	0.36	0.40	0.27	0.25	0.42	0.48	0.39	0.36	0.47	0.63	0.51	0.39	0.33	0.39
	Ours	0.79	0.74	0.78	0.87	0.86	0.76	0.77	0.87	0.88	0.87	0.90	0.84	0.93	0.92	0.82	0.76	0.83
LPIPS	DietNeRF [9]	0.59	0.61	0.59	0.56	0.59	0.56	0.61	0.58	0.52	0.56	0.56	0.58	0.48	0.54	0.57	0.59	0.57
	PixelNeRF ft [48]	0.70	0.61	0.72	0.59	0.59	0.58	0.67	0.59	0.48	0.52	0.56	0.61	0.67	0.65	0.67	0.58	0.61
	RegNeRF [20]	0.65	0.65	0.67	0.61	0.62	0.63	0.63	0.62	0.63	0.64	0.59	0.65	0.58	0.62	0.64	0.66	0.63
	FreeNeRF [44]	0.61	0.67	0.64	0.55	0.58	0.58	0.63	0.58	0.61	0.61	0.57	0.64	0.48	0.54	0.61	0.65	0.60
	MPNeRF [7]	0.21	0.24	0.18	0.20	0.18	0.25	0.24	0.18	0.20	0.16	0.12	0.17	0.12	0.14	0.20	0.19	0.19
	3DGS [12]	0.24	0.33	0.33	0.20	0.17	0.30	0.28	0.21	0.19	0.21	0.17	0.25	0.12	0.13	0.30	0.30	0.23
	FSGS [55]	0.27	0.30	0.33	0.27	0.29	0.33	0.29	0.24	0.22	0.26	0.32	0.28	0.20	0.19	0.32	0.27	0.27
	DNGaussian [15]	0.39	0.51	0.47	0.52	0.54	0.54	0.39	0.37	0.42	0.44	0.56	0.41	0.30	0.41	0.42	0.41	0.44
	DropoutGS [43]	0.42	0.45	0.43	0.56	0.46	0.66	0.38	0.36	0.42	0.36	0.62	0.41	0.31	0.35	0.41	0.37	0.43
	Guidedvd-3DGS [54]	0.49	0.53	0.47	0.28	0.42	0.67	0.61	0.39	0.49	0.54	0.36	0.39	0.49	0.51	0.54	0.62	0.49
	NoPoSplat [47]	0.67	0.57	0.55	0.53	0.44	0.46	0.50	0.60	0.31	0.70	0.64	0.70	0.57	0.72	0.52	0.64	0.57
	NoPoSplat ft [47]	0.30	0.28	0.24	0.28	0.24	0.28	0.27	0.22	0.16	0.23	0.20	0.24	0.13	0.19	0.25	0.26	0.24
	MVSplat [3]	0.78	0.76	0.80	0.78	0.80	0.79	0.77	0.78	0.77	0.78	0.79	0.77	0.79	0.79	0.80	0.76	0.78
	MVSplat ft [3]	0.46	0.48	0.48	0.48	0.45	0.49	0.49	0.37	0.34	0.43	0.50	0.38	0.30	0.35	0.44	0.47	0.43
	Ours	0.17	0.22	0.19	0.13	0.14	0.19	0.19	0.12	0.11	0.13	0.11	0.14	0.07	0.08	0.17	0.22	0.15

only, as their architectures are designed for a small number of input views. All other methods were evaluated on both datasets for comprehensive comparison.

Tables 1 and 2 demonstrated that GeoProp-GS achieved state-of-the-art results across both datasets, substantially outperforming the second-best methods on all metrics. These improvements spanned diverse environments from urban buildings and stadiums to natural mountains and port scenes, demonstrating robustness across varying geometric complexity, texture patterns, and viewpoint sparsity levels. Fig. 3 and Fig. 4 presented visualization comparisons. NeRF-based methods produced severely blurred NVS, reflecting their struggle with implicit representation under extreme sparsity. Sparse-view 3DGS methods exhibited varying degradation: vanilla 3DGS suffered from floating artifacts and incomplete geometry due to failed SfM initialization; FSGS and DNGaus-

Table 2: Quantitative comparison with baseline methods on 3D-AS dataset (7 training views). The best, second-best, and third-best entries are marked in , , and , respectively.

Method	City									Country									Port								
	Scene 0			Scene 1			Scene 2			Scene 0			Scene 1			Scene 2			Scene 0			Scene 1			Scene 2		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
NeRF [19]	13.26	0.15	0.65	14.85	0.20	0.64	15.11	0.23	0.72	15.46	0.33	0.71	17.52	0.47	0.64	16.42	0.42	0.70	15.63	0.47	0.50	14.84	0.22	0.65	15.49	0.32	0.61
FreeNeRF [44]	13.52	0.12	0.58	14.28	0.15	0.56	14.31	0.17	0.61	15.52	0.27	0.58	18.30	0.42	0.52	15.78	0.32	0.55	18.98	0.59	0.31	15.04	0.23	0.54	15.89	0.35	0.48
RegNeRF [20]	13.41	0.08	0.62	15.46	0.16	0.57	15.54	0.18	0.59	16.31	0.25	0.58	18.43	0.37	0.55	16.83	0.33	0.58	17.74	0.54	0.37	15.45	0.20	0.60	16.27	0.31	0.54
3DGS [12]	20.66	0.67	0.24	20.62	0.52	0.37	19.43	0.41	0.45	19.89	0.42	0.45	24.35	0.65	0.32	19.43	0.42	0.45	23.67	0.88	0.21	24.03	0.84	0.13	24.41	0.87	0.12
FSGS [55]	19.67	0.64	0.26	18.37	0.47	0.38	14.93	0.23	0.59	16.88	0.36	0.55	22.87	0.66	0.35	22.86	0.69	0.35	23.81	0.87	0.15	24.02	0.84	0.14	24.07	0.86	0.14
DNGaussian [15]	17.59	0.49	0.38	14.62	0.19	0.59	13.33	0.22	0.68	15.31	0.34	0.65	16.95	0.48	0.58	19.98	0.60	0.57	21.62	0.80	0.20	19.31	0.61	0.30	20.86	0.75	0.22
DropoutGS [43]	19.18	0.58	0.33	18.37	0.47	0.38	14.79	0.24	0.62	15.86	0.34	0.60	17.93	0.47	0.52	16.55	0.43	0.61	21.92	0.81	0.19	21.55	0.72	0.21	22.28	0.78	0.19
Guidedvd-3DGS [54]	14.60	0.25	0.59	16.49	0.28	0.53	16.63	0.29	0.57	17.59	0.37	0.51	19.94	0.55	0.47	17.61	0.45	0.54	21.88	0.81	0.16	17.28	0.43	0.50	21.89	0.73	0.22
Ours	22.20	0.75	0.18	21.98	0.66	0.25	20.81	0.56	0.32	21.71	0.54	0.37	25.55	0.69	0.25	26.16	0.74	0.25	25.36	0.89	0.10	25.09	0.85	0.13	26.49	0.89	0.10

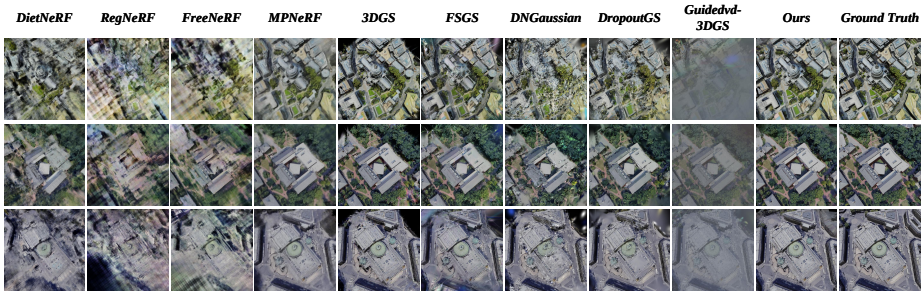

Fig. 3: Qualitative comparisons on LEVIR-NVS dataset (3 training views).

sian showed distorted structures despite depth regularization; DropoutGS and Guidedvd-3DGS achieved moderate but inconsistent quality. Even aerial-specific MPNeRF exhibited noticeable artifacts in complex regions.

In contrast, GeoProp-GS tackled this bottleneck through three complementary components. HDD established complete geometric coverage of observed regions. GVA then propagated structure into unobserved areas beyond the photometric supervision boundary. AGO further stabilized optimization where direct supervision was absent.

4.3 Ablation Studies

To validate the effectiveness of each component in our GeoProp-GS framework, we conducted ablation studies on the LEVIR-NVS dataset by progressively adding our proposed modules. Table 3 presented quantitative results across three configurations, organized by strategies for handling observed regions (with photometric supervision) and unobserved regions (lacking direct supervision). Since GVA and AGO were co-designed as a unified solution, they were ablated

Table 3: Progressive Component Ablation study on LEVIR-NVS dataset (3-views).

Region Handling Strategy		Metrics		
Observed *(w/ supervision)*	Unobserved *(w/o supervision)*	PSNR↑	SSIM↑	LPIPS↓
COLMAP	-	21.10	0.75	0.23
HDD	-	21.57	0.82	0.15
HDD	GVA+AGO	**24.35**	**0.83**	**0.15**

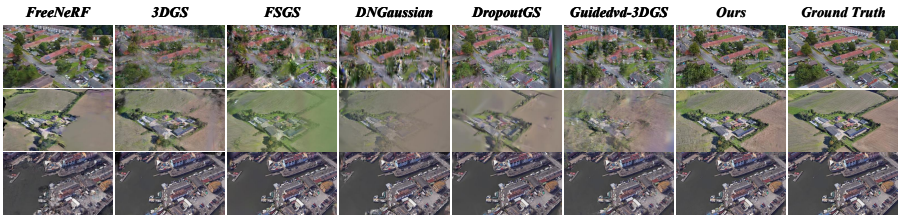

Fig. 4: Qualitative comparisons on 3D-AS dataset (7 training views).

together: GVA generated proposal points for unobserved regions, while AGO stabilized them via anchor-residual parameterization to provide optimization signals where direct photometric supervision was absent.

Progressive Component Analysis. Using standard COLMAP initialization within FSGS (without depth supervision) established the baseline, confirming that COLMAP struggled with repetitive aerial textures and produced incomplete initialization. Adding HDD improved all metrics, demonstrating that densifying SfM points with metric-aligned monocular depth provided more complete geometric coverage. The substantial LPIPS improvement indicated better perceptual quality, though the modest PSNR gain revealed that handling observed regions alone could not address unobserved areas. The full model incorporating GVA and AGO achieved dramatic improvements across all metrics. This validated the critical importance of addressing unobserved regions: GVA generated geometrically plausible proposals through depth-conditioned inpainting, while AGO stabilized them via anchor-residual parameterization, enabling stable NVS beyond direct photometric supervision. Fig. 5 (left) provided visualization results across three representative scenes. The COLMAP baseline produced artifacts and incomplete structures. Adding HDD dramatically improved rendering quality. Full GeoProp-GS achieved photorealistic synthesis matching ground truth.

Summary. The ablation results validated our two-stage design philosophy: HDD established robust initialization for observed regions with photometric supervision, while GVA+AGO enabled stable NVS optimization in unobserved regions lacking direct supervision. Together, these components addressed both failure modes identified in Section 1.

4.4 Analysis Experiments

To further validate the design choices and generalization capability of GeoProp-GS, we conducted a series of analysis experiments. We compared HDD against alternative initialization strategies to demonstrate its superiority in establishing dense geometric priors. We then integrated HDD with existing 3DGS-based methods to evaluate its effectiveness as a plug-and-play module. Additional experiments examined the impact of depth estimator choice on overall performance,

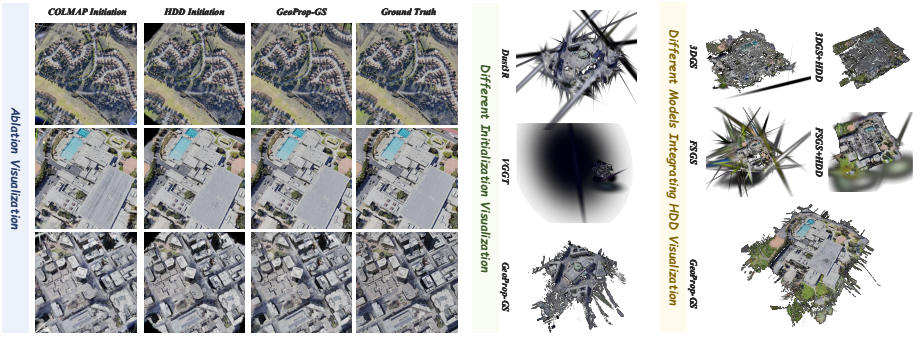

Fig. 5: Ablation and analysis visualizations for sparse-view aerial 3DGS. ***Left***: Progressive ablation showing rendering quality improvements. ***Middle***: Comparison of different initialization methods (VGGT, DUSt3R, and our HDD) and their impact on 3DGS quality. ***Right***: HDD as a plug-and-play module consistently improves 3DGS quality for both vanilla 3DGS and FSGS.

Table 4: Comparison of different initialization methods.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
VGGT [35]	16.43	0.30	0.46
DUSt3R [36]	18.65	0.56	0.34
HDD (Ours)	**21.57**	**0.82**	**0.15**

Table 5: Performance improvement by integrating our HDD.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
3DGS [12]	20.20	0.72	0.23
+ HDD	21.80 (+1.60)	0.83 (+0.11)	0.14 (-0.09)
FSGS [55]	20.93	0.71	0.27
+ HDD	21.75 (+0.82)	0.80 (+0.09)	0.19 (-0.08)

and we assessed the time and memory efficiency of the full pipeline to confirm its practical applicability.

Comparison of Initialization Strategies. Tab. 4 compared our HDD against two alternative initialization methods: VGGT [35] and DUSt3R [36]. Our method consistently outperformed both baselines across all metrics. VGGT[3] suffered from point cloud-pose misalignment. DUSt3R employed learning-based feature matching but remained fundamentally limited by correspondence establishment. In repetitive aerial textures with sparse discriminative features, both methods produced incomplete point clouds with geometric inconsistencies. Our HDD bypassed these limitations by leveraging monocular depth priors with hierarchical refinement, ensuring both metric consistency and fine-grained geometric accuracy. Fig. 5 (middle) visualized the learned 3DGS point clouds. VGGT produced degenerate Gaussians with massive scales, while DUSt3R exhibited oversized Gaussians in incomplete regions. Our HDD generated well-behaved 3DGS points with uniformly-scaled Gaussians throughout the scene.

Generalization to Existing Methods. To demonstrate HDD's generalizability, we integrated it with existing 3DGS methods by replacing their SfM initialization with our $\mathcal{C}_{dense}$ generated by HDD. Table 5 showed that both

[3] We ran VGGT without bundle adjustment (BA) due to memory limits.

Table 6: Performance comparison across different monocular depth estimators on LEVIR-NVS (3-views).

		MiDas	Marigold	Depth Anything V2
HDD Only	PSNR ↑	21.18	21.26	**21.57**
	SSIM ↑	0.80	0.81	**0.82**
	LPIPS ↓	0.17	0.16	**0.15**
Full Model	PSNR ↑	23.79	23.93	**24.35**
	SSIM ↑	0.81	0.82	**0.83**
	LPIPS ↓	0.17	0.16	**0.15**

Table 7: Training time comparison and GPU cost on LEVIR-NVS dataset. "m" denotes minutes, "h" denotes hours with an RTX 3090.

Method	Iterations	Training Time	GPU Memory (GB)
DNGaussian	6K	3.5 m	0.88
DropoutGS	6K	4.5 m	0.93
FSGS	10K	6.1 m	2.64
Guidedvd-3DGS	10K	2.5 h	21.30
GeoProp-GS (Ours)	10K	6.1 m	2.73

vanilla 3DGS [12] and FSGS [55] achieved consistent improvements across all metrics when equipped with HDD. These results validated that HDD addressed a fundamental bottleneck shared by existing methods: dependence on sparse and incomplete SfM point clouds for observed regions. By providing dense and geometrically consistent initialization in observed areas, HDD effectively compensated for insufficient geometric coverage that limited existing methods in sparse aerial NVS scenarios. Fig. 5 (right) demonstrated substantial 3DGS quality improvements for both methods, validating HDD as an effective plug-and-play module that enhanced sparse-view aerial NVS.

Impact of Depth Estimator Choice. To assess the sensitivity of HDD to the choice of monocular depth estimator, we replaced Depth Anything V2 with MiDas [28] and Marigold [11] while keeping all other components unchanged. As shown in Tab. 6, all three variants consistently outperformed all baselines in Table 1, including sparse-view 3DGS methods. This finding confirmed that the performance gains of GeoProp-GS were not contingent on a specific depth model. Performance scaled with depth estimation quality: higher-quality depth estimates produced more accurate dense point clouds within HDD, which in turn provided more reliable geometric scaffolds for optimization. Depth Anything V2 achieved the best results across all metrics and was therefore adopted as the default estimator in all other experiments.

Time and Memory Efficiency. GeoProp-GS required a one-time initialization before training: HDD generated $\mathcal{C}_{\text{dense}}$ in 3 seconds, and GVA generated $\mathcal{C}_{\text{proposal}}$ in approximately 15 minutes (a local RTX 4090) or 8 minutes (cloud API). As shown in Tab. 7, subsequent training took 6.1 minutes for 10K iterations, matching FSGS and substantially faster than Guidedvd-3DGS, as anchor-constrained optimization replaced expensive per-iteration regularization with implicit architectural constraints. In terms of GPU memory, GeoProp-GS (2.73 GB) remained comparable to FSGS while achieving substantially higher quality. The one-time initialization overhead was therefore a worthwhile investment, making GeoProp-GS practical for real-world aerial reconstruction.

5 Conclusion

We present GeoProp-GS to address the fundamental failure of sparse-view 3DGS in aerial domains, where repetitive textures preclude reliable SfM initialization. Our method introduces HDD to establish dense geometric coverage, GVA to propagate structure into unobserved regions, and AGO to stabilize optimization through anchor-residual decomposition. Extensive experiments demonstrate state-of-the-art performance on two aerial benchmarks, with consistent and substantial improvements over existing methods across diverse scene types. Ablation studies validate the necessity of each component, and generalization experiments confirm HDD as an effective plug-and-play module compatible with existing 3DGS pipelines. We believe this work establishes a strong foundation for bridging the gap between ground-level and aerial-view 3D scene representation, enabling robust novel view synthesis in challenging sparse-view scenarios.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 62571387 and by the Program of China Scholarship Council.

References

1. Bao, Y., Ding, T., Huo, J., Liu, Y., Li, Y., Li, W., Gao, Y., Luo, J.: 3d gaussian splatting: Survey, technologies, challenges, and opportunities. IEEE Transactions on Circuits and Systems for Video Technology (2025)
2. Birkl, R., Wofk, D., Müller, M.: Midas v3. 1–a model zoo for robust monocular relative depth estimation. arXiv preprint arXiv:2307.14460 (2023)
3. Chen, Y., Xu, H., Zheng, C., Zhuang, B., Pollefeys, M., Geiger, A., Cham, T.J., Cai, J.: Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images. In: European Conference on Computer Vision. pp. 370–386. Springer (2024)
4. Chung, J., Oh, J., Lee, K.M.: Depth-regularized optimization for 3d gaussian splatting in few-shot images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 811–820 (2024)
5. Fei, B., Xu, J., Zhang, R., Zhou, Q., Yang, W., He, Y.: 3d gaussian splatting as new era: A survey. IEEE Transactions on Visualization and Computer Graphics (2024)
6. Gao, R., Holynski, A., Henzler, P., Brussee, A., Martin-Brualla, R., Srinivasan, P., Barron, J.T., Poole, B.: Cat3d: Create anything in 3d with multi-view diffusion models. arXiv preprint arXiv:2405.10314 (2024)
7. Gao, Z., Jiao, L., Li, L., Liu, X., Liu, F., Chen, P., Guo, Y.: Multiplane prior guided few-shot aerial scene rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5009–5019 (2024)
8. Han, L., Zhou, J., Liu, Y.S., Han, Z.: Binocular-guided 3d gaussian splatting with view consistency for sparse view synthesis. Advances in Neural Information Processing Systems **37**, 68595–68621 (2024)
9. Jain, A., Tancik, M., Abbeel, P.: Putting nerf on a diet: Semantically consistent few-shot view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5885–5894 (2021)

10. Kathariya, B., Lee, D.Y., Huang, T.W., Shao, T., Yin, P., Su, G.M.: Gaussian splatting: State-of-the-arts and future trends. In: 2025 International Conference on Computing, Networking and Communications (ICNC). pp. 876–881. IEEE (2025)
11. Ke, B., Obukhov, A., Huang, S., Metzger, N., Daudt, R.C., Schindler, K.: Repurposing diffusion-based image generators for monocular depth estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9492–9502 (2024)
12. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
13. Kim, H., Lee, I.K.: Is 3DGS useful?: Comparing the effectiveness of recent reconstruction methods in vr. In: 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 71–80. IEEE (2024)
14. Labs, B.F., Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., Esser, P., Kulal, S., Lacey, K., Levi, Y., Li, C., Lorenz, D., Müller, J., Podell, D., Rombach, R., Saini, H., Sauer, A., Smith, L.: Flux.1 kontext: Flow matching for in-context image generation and editing in latent space (2025)
15. Li, J., Zhang, J., Bai, X., Zheng, J., Ning, X., Zhou, J., Gu, L.: Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20775–20785 (2024)
16. Li, X., Zhang, Q., Kang, D., Cheng, W., Gao, Y., Zhang, J., Liang, Z., Liao, J., Cao, Y.P., Shan, Y.: Advances in 3d generation: A survey. arXiv preprint arXiv:2401.17807 (2024)
17. Liu, F., Sun, W., Wang, H., Wang, Y., Sun, H., Ye, J., Zhang, J., Duan, Y.: Reconx: Reconstruct any scene from sparse views with video diffusion model. arXiv preprint arXiv:2408.16767 (2024)
18. Liu, X., Zhou, C., Huang, S.: 3DGS-enhancer: Enhancing unbounded 3d gaussian splatting with view-consistent 2d diffusion priors. Advances in Neural Information Processing Systems **37**, 133305–133327 (2024)
19. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
20. Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S., Geiger, A., Radwan, N.: Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5480–5490 (2022)
21. Pal, O.K., Shovon, M.S.H., Mridha, M.F., Shin, J.: In-depth review of ai-enabled unmanned aerial vehicles: trends, vision, and challenges. Discover Artificial Intelligence **4**(1), 97 (2024)
22. Paliwal, A., Ye, W., Xiong, J., Kotovenko, D., Ranjan, R., Chandra, V., Kalantari, N.K.: Coherentgs: Sparse novel view synthesis with coherent 3d gaussians. In: European Conference on Computer Vision. pp. 19–37. Springer (2024)
23. Paliwal, A., Zhou, X., Ye, W., Xiong, J., Ranjan, R., Kalantari, N.K.: RI3D: Few-shot gaussian splatting with repair and inpainting diffusion priors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 25094–25103 (2025)
24. Pan, L., Baráth, D., Pollefeys, M., Schönberger, J.L.: Global structure-from-motion revisited. In: European Conference on Computer Vision. pp. 58–77. Springer (2024)

25. Park, H., Ryu, G., Kim, W.: Dropgaussian: Structural regularization for sparse-view gaussian splatting. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 21600–21609 (2025)
26. Qian, J., Yan, Y., Gao, F., Ge, B., Wei, M., Shangguan, B., He, G.: C3DGS: Compressing 3D gaussian model for surface reconstruction of large-scale scenes based on multi-view uav images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2025)
27. Radl, L., Steiner, M., Parger, M., Weinrauch, A., Kerbl, B., Steinberger, M.: Stopthepop: Sorted gaussian splatting for view-consistent real-time rendering. ACM Transactions on Graphics (TOG) **43**(4), 1–17 (2024)
28. Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., Koltun, V.: 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)
29. Schönberger, J.L., Zheng, E., Frahm, J.M., Pollefeys, M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (ECCV) (2016)
30. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
31. Schönberger, J.L., Price, T., Sattler, T., Frahm, J.M., Pollefeys, M.: A vote-and-verify strategy for fast spatial verification in image retrieval. In: Asian Conference on Computer Vision (ACCV) (2016)
32. Sun, W., Chen, S., Liu, F., Chen, Z., Duan, Y., Zhang, J., Wang, Y.: Dimensionx: Create any 3d and 4d scenes from a single image with controllable video diffusion. arXiv preprint arXiv:2411.04928 (2024)
33. Tang, X., Jia, J., Wang, Y., Ma, J., Zhang, X.: Semantic-aware dropsplat: Adaptive pruning of redundant gaussians for 3d aerial-view segmentation. arXiv preprint arXiv:2508.09626 (2025)
34. Wang, G., Chen, Z., Loy, C.C., Liu, Z.: SparseNeRF: Distilling depth ranking for few-shot novel view synthesis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9065–9076 (2023)
35. Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., Novotny, D.: VGGT: Visual geometry grounded transformer. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5294–5306 (2025)
36. Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., Revaud, J.: Dust3r: Geometric 3d vision made easy. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20697–20709 (2024)
37. Wang, Y., Tang, X., Ma, J., Zhang, X., Zhu, C., Liu, F., Jiao, L.: Pseudo-viewpoint regularized 3d gaussian splatting for remote sensing few-shot novel view synthesis. In: IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. pp. 8660–8663. IEEE (2024)
38. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
39. Wu, J.Z., Zhang, Y., Turki, H., Ren, X., Gao, J., Shou, M.Z., Fidler, S., Gojcic, Z., Ling, H.: Difix3d+: Improving 3d reconstructions with single-step diffusion models. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 26024–26035 (2025)
40. Wu, Y., Zou, Z., Shi, Z.: Remote sensing novel view synthesis with implicit multi-plane representations. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–13 (2022)

41. Xiong, H., Muttukuru, S., Upadhyay, R., Chari, P., Kadambi, A.: Sparsegs: Real-time 360 sparse view synthesis using gaussian splatting. arXiv preprint arXiv:2312.00206 (2023)
42. Xu, H., Peng, S., Wang, F., Blum, H., Barath, D., Geiger, A., Pollefeys, M.: Depthsplat: Connecting gaussian splatting and depth. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 16453–16463 (2025)
43. Xu, Y., Wang, L., Chen, M., Ao, S., Li, L., Guo, Y.: Dropoutgs: Dropping out gaussians for better sparse-view rendering. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 701–710 (2025)
44. Yang, J., Pavone, M., Wang, Y.: FreeNeRF: Improving few-shot neural rendering with free frequency regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8254–8263 (2023)
45. Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H.: Depth anything: Unleashing the power of large-scale unlabeled data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10371–10381 (2024)
46. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything v2. Advances in Neural Information Processing Systems **37**, 21875–21911 (2024)
47. Ye, B., Liu, S., Xu, H., Li, X., Pollefeys, M., Yang, M.H., Peng, S.: No pose, no problem: Surprisingly simple 3d gaussian splats from sparse unposed images. arXiv preprint arXiv:2410.24207 (2024)
48. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelNeRF: Neural radiance fields from one or few images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4578–4587 (2021)
49. Yu, W., Xing, J., Yuan, L., Hu, W., Li, X., Huang, Z., Gao, X., Wong, T.T., Shan, Y., Tian, Y.: Viewcrafter: Taming video diffusion models for high-fidelity novel view synthesis. arXiv preprint arXiv:2409.02048 (2024)
50. Zhang, J., Li, Y., Chen, A., Xu, M., Liu, K., Wang, J., Long, X.X., Liang, H., Xu, Z., Su, H., et al.: Advances in feed-forward 3d reconstruction and view synthesis: A survey. arXiv preprint arXiv:2507.14501 (2025)
51. Zhang, J., Li, J., Yu, X., Huang, L., Gu, L., Zheng, J., Bai, X.: Cor-gs: sparse-view 3d gaussian splatting via co-regularization. In: European Conference on Computer Vision. pp. 335–352. Springer (2024)
52. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
53. Zheng, Y., Jiang, Z., He, S., Sun, Y., Dong, J., Zhang, H., Du, Y.: Nexusgs: Sparse view synthesis with epipolar depth priors in 3d gaussian splatting. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 26800–26809 (2025)
54. Zhong, Y., Li, Z., Chen, D.Z., Hong, L., Xu, D.: Taming video diffusion prior with scene-grounding guidance for 3d gaussian splatting from sparse inputs. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 6133–6143 (2025)
55. Zhu, Z., Fan, Z., Jiang, Y., Wang, Z.: Fsgs: Real-time few-shot view synthesis using gaussian splatting. In: European conference on computer vision. pp. 145–163. Springer (2024)