DGS: DUAL-EVIDENCE INTEGRATION FOR ACCURATE SPARSE-VIEW SURFACE RECONSTRUCTION WITH GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Accurate surface reconstruction with 3D Gaussian Splatting (3DGS) remains challenging under sparse views, where limited multi-view observations provide insufficient geometric constraints. Recent methods alleviate this issue by supervising reconstruction with aligned depth predictions from foundation models. However, these predictions typically lack explicit dense multi-view matching and can be suboptimal where reliable photometric correspondences are available. More importantly, their depth reliability is inherently range-dependent: aligned VGGT depth exhibits smoothly varying bias across surfaces, while local depth variations remain substantially more accurate and degrade gradually with spatial distance. Motivated by this, we propose DGS, a dual-evidence integration framework for sparse-view surface reconstruction. Rather than directly using predicted depth as supervision, DGS combines two complementary forms of geometric evidence: point evidence from multi-view photometric matching provides accurate absolute anchor depths at reliable correspondences, while edge evidence from local log-depth differences of VGGT captures dense local depth variations. A global integration couples the two, extending anchor depths through local edge relations while constraining accumulated drift. We further introduce prior-based aggregation and anchor depth refinement to improve the coverage and accuracy of point evidence under sparse views. Experiments on DTU and Tanks and Temples demonstrate state-of-the-art surface reconstruction quality.

1 INTRODUCTION

Recovering faithful surface geometry from multi-view images remains a fundamental challenge in 3D computer vision (Hartley et al., 2003). The emergence of differentiable rendering (Mildenhall et al., 2021) has driven substantial progress in multi-view surface reconstruction (Wang et al., 2021; Yariv et al., 2021; Fu et al., 2022). More recently, 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023), with its explicit primitive-based representation and efficient differentiable rasterization, has provided a powerful representation for high-fidelity surface reconstruction (Huang et al., 2024a; Chen et al., 2024; Li et al., 2026; Yu et al., 2024; Guédon et al., 2025a). Despite these advances, accurate geometry still relies heavily on sufficiently dense multi-view observations to establish reliable cross-view constraints. Under sparse views, reduced overlap and limited correspondences weaken these constraints, leaving the underlying surface geometry severely under-constrained.

A natural way to compensate for insufficient multi-view constraints is to leverage geometric priors from pretrained foundation models (Wang et al., 2025; 2024b; Eftekhar et al., 2021; Kong et al., 2026). Recent methods (Jin et al., 2026; Xiang et al., 2026; Turkulainen et al., 2025; Yu et al., 2022; Han et al., 2025) typically align predicted depth to the rendered geometry and use it as dense supervision, implicitly treating the aligned depth as a reliable geometric target. However, these models typically lack explicit dense multi-view matching, resulting in suboptimal depth supervision in regions with reliable photometric evidence. Their predictions also exhibit inherently range-dependent reliability. Taking VGGT (Wang et al., 2025) as a representative example, its aligned depth exhibits a smoothly varying bias whose magnitude and even sign vary across surfaces, resulting in inaccurate absolute surface locations. In contrast, depth differences between nearby pixels are considerably more accurate, with errors increasing gradually with spatial distance (Fig. 1). This obser-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

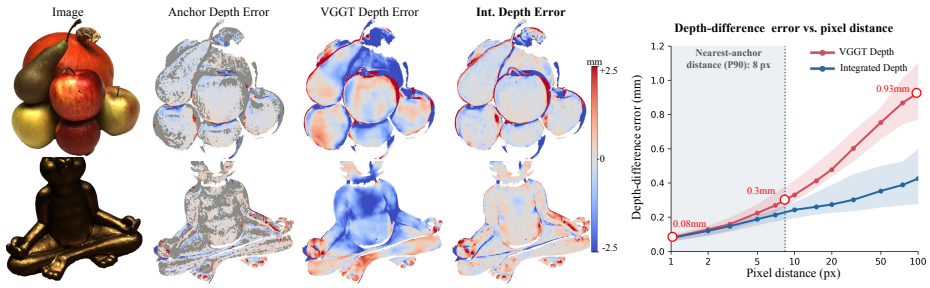

Figure 1: **Range-dependent reliability of the VGGT depth prior.** *Left:* Signed depth errors on two representative scenes. Photometric anchors provide accurate depth estimates at reliable correspondences, whereas the aligned VGGT depth exhibits a smoothly varying bias that varies in sign and magnitude across surfaces. The integrated depth substantially mitigates this bias. *Right:* For each pixel, we pair it with pixels at a given distance along the four axis directions, compute their depth differences in both the prediction and ground truth, and average the absolute discrepancies over the four pairs. VGGT provides accurate local depth differences, but the error increases progressively with distance. With 90% of pixels lying within 8 px of an anchor (shaded region), the integration primarily exploits VGGT within its more reliable local range, mitigating long-range drift.

vation suggests that direct depth supervision overlooks more accurate depths available from reliable multi-view correspondences and imposes the less reliable absolute depth of the learned prior on the reconstruction.

Based on these observations, we propose DGS, a dual-evidence integration framework for sparse-view surface reconstruction. Rather than using the predicted depth itself as supervision, DGS combines two complementary forms of geometric evidence: point evidence from multi-view photometric matching provides accurate absolute anchor depths at pixels with reliable correspondences, while edge evidence from first-order differences of the VGGT log-depth captures dense local depth variations. Their combination naturally leads to a global integration problem (Quéau et al., 2018a), in which edge evidence extends the incomplete anchors into a dense depth field, while the anchors constrain the accumulated drift across surfaces. The accuracy of the integrated depth depends on the distribution and quality of these anchors. We further introduce prior-based aggregation and anchor depth refinement to improve their coverage and accuracy. In our experiments, 90% of pixels lie within 8 px of an anchor. Over this spatial range, the average local depth-difference error of VGGT remains below approximately 0.30 mm (Fig. 1), allowing the integration to exploit the learned prior predominantly within its more reliable local range.

In summary, our main contributions are as follows:

- We analyze the VGGT depth prior for sparse-view reconstruction, characterizing its limited accuracy in regions with reliable photometric evidence and its range-dependent prediction reliability.
- We propose DGS, which integrates point evidence from multi-view photometric matching and edge evidence from VGGT into dense depth supervision that preserves accurate surface locations and faithful local variations. Prior-based aggregation (PBA) and anchor depth refinement (ADR) further improve anchor coverage and accuracy under sparse views.
- DGS achieves state-of-the-art results on DTU and Tanks and Temples, with a mean Chamfer Distance of 0.71 mm and a mean F-score of 0.310.

2 Related Work

2.1 Multi-view Reconstruction

Classical multi-view stereo methods recover scene geometry from calibrated images by establishing photometric correspondences to estimate per-view depth maps, followed by geometric filtering and multi-view fusion for surface reconstruction (Galliani et al., 2015; Schönberger et al., 2016;

Xu & Tao, 2019; Yao et al., 2018; Zhang et al., 2023; Wu et al., 2024). Neural Radiance Fields (NeRFs) (Mildenhall et al., 2021) introduce differentiable rendering into neural scene representations, with subsequent methods incorporating geometry-aware representations or additional constraints for accurate surface reconstruction (Wang et al., 2021; Yariv et al., 2021; Fu et al., 2022). More recently, 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) combines explicit Gaussian primitives with differentiable splatting, motivating extensive efforts toward accurate and efficient geometric reconstruction (Huang et al., 2024a; Chen et al., 2024; Yu et al., 2024; Guédon et al., 2025a; Li et al., 2026; Gomez et al., 2026). While these methods achieve high-quality reconstruction with sufficient observations, their geometric constraints become substantially weaker as the number of input views decreases, motivating recent efforts toward sparse-view reconstruction.

2.2 Sparse-View Surface Reconstruction

Existing methods have explored various strategies to improve surface reconstruction from sparse views. One line of work learns generalizable reconstruction models across scenes (Long et al., 2022; Ren et al., 2023; Liang et al., 2023; Xu et al., 2023; Peng et al., 2023; Na et al., 2024), but typically requires extensive pre-training and may generalize poorly to out-of-distribution scenes. Optimization-based methods instead incorporate additional geometric priors (Yin et al., 2023; Eftekhar et al., 2021) to compensate for insufficient geometric constraints under sparse observations. MonoSDF (Yu et al., 2022) incorporates monocular depth and normal cues into neural implicit reconstruction. FatesGS (Huang et al., 2025) distills local depth rankings from monocular depth predictions. MAtCha (Guédon et al., 2025b) regularizes surface normals and mean curvature to preserve fine-scale geometry. DP-GS (Han et al., 2026) uses monocular normals to guide depth propagation from high-confidence regions and employs the resulting depth as geometric supervision. Beyond monocular priors, some methods incorporate MVS (Wu et al., 2025; Younes et al., 2024; Huang et al., 2024b) and stereo matching (Gu et al., 2026) to enhance geometric estimation. More recently, multi-view geometry foundation models (Wang et al., 2024b; 2025; 2026; Lin et al., 2025) have shown strong capabilities in recovering 3D geometry directly from unposed images. Building on these advances, PlanarGS (Jin et al., 2026) aligns DUSt3R depth with sparse SfM points to regularize surface geometry. VGGS (Xiang et al., 2026) uses multi-view consistent anchors to calibrate VGGT depth for geometric supervision. Rather than relying on aligned absolute depth, our method integrates local depth differences from VGGT with anchor depths obtained through multi-view photometric matching, providing more reliable geometric supervision.

2.3 Gradient-Domain Integration

Recovering a scalar field from its differences is a classical problem, extensively studied in normal integration (Horn & Brooks, 1986; Frankot & Chellappa, 1988; Quéau et al., 2018a), with subsequent methods improving robustness to noise and depth discontinuities (Quéau et al., 2018b; Cao et al., 2022). Since differential constraints alone cannot determine absolute depth, absolute measurements have been introduced to resolve this ambiguity, including control points for correcting photometric stereo bias (Horovitz & Kiryati, 2004) and triangulated positions jointly optimized with normals (Nehab et al., 2005). Recent methods extend this idea to learned geometric cues, integrating predicted local geometry with observed or reconstructed depth (Zhang & Funkhouser, 2018; Pataki et al., 2025). In our setting, we instead recover accurate and well-distributed absolute constraints from sparse RGB observations, which establish reliable surface locations while bounding the spatial range of learned priors, enabling dense geometric supervision for 3DGS.

3 Methods

Given N posed RGB images $\{\boldsymbol{I}_i\}_{i=1}^{N}$ captured from sparse views, where $\boldsymbol{I}_i \in \mathbb{R}^{H\times W\times 3}$, our goal is to reconstruct accurate scene geometry with Gaussian Splatting. The overall framework is illustrated in Fig. 2. Instead of directly using the predicted geometry as supervision, DGS constructs two complementary forms of evidence and integrates them into a dense depth map for 3DGS optimization. In this section, we first review the preliminaries (Sec. 3.1) and formulate the dual-evidence integration problem (Sec. 3.2). We then describe the construction of point and edge evidence (Sec. 3.3), followed by the optimization objectives (Sec. 3.4).

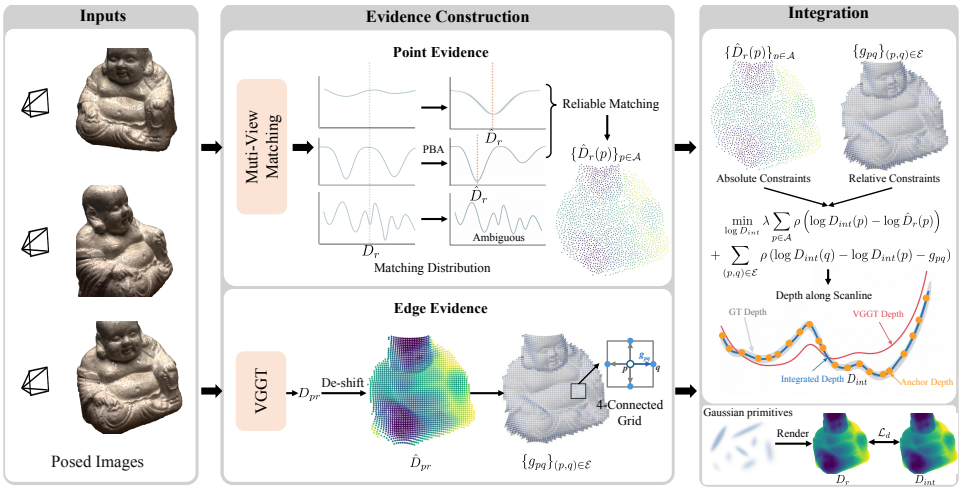

Figure 2: Given sparse input views, DGS constructs complementary geometric evidence from multi-view photometric matching and the VGGT depth prior. Point evidence provides accurate absolute depth anchors, further improved by prior-based aggregation and anchor depth refinement, while edge evidence captures reliable local depth variations from the prior. The two forms of evidence are globally integrated into dense depth, providing geometric supervision for 3DGS reconstruction.

3.1 Preliminaries

3DGS (Kerbl et al., 2023) represents a scene using a set of anisotropic Gaussians $\{\mathcal{G}_i\}$, where each Gaussian $\mathcal{G}_i$ is parameterized by a center $\boldsymbol{\mu}_i$, a covariance matrix $\boldsymbol{\Sigma}_i$, an opacity α_i, and spherical harmonic coefficients $\boldsymbol{f}_i$:

$$\mathcal{G}_i(\boldsymbol{x}) = \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right). \tag{1}$$

Following PGSR (Chen et al., 2024), we flatten each Gaussian into a planar disk with normal $\boldsymbol{n}_i$. The rendered normal $\boldsymbol{N}$ and plane distance $\boldsymbol{\mathcal{D}}$ are obtained via α-blending, and the unbiased rendered depth $\boldsymbol{D}_r$ is then recovered through ray-plane intersection:

$$\begin{bmatrix}\boldsymbol{N}\\ \boldsymbol{\mathcal{D}}\end{bmatrix} = \sum_{i\in\mathcal{S}} \begin{bmatrix}\boldsymbol{R}_c^\top \boldsymbol{n}_i\\ d_i\end{bmatrix} \alpha_i \prod_{j=1}^{i-1}(1-\alpha_j), \qquad \boldsymbol{D}_r(\boldsymbol{p}) = \frac{\boldsymbol{\mathcal{D}}(\boldsymbol{p})}{\boldsymbol{N}(\boldsymbol{p})^\top \boldsymbol{K}^{-1}\tilde{\boldsymbol{p}}}, \tag{2}$$

where $\mathcal{S}$ denotes the set of Gaussians contributing to the pixel, $\boldsymbol{R}_c$ is the camera rotation, $\boldsymbol{K}$ is the camera intrinsic matrix, $\tilde{\boldsymbol{p}}$ is the homogeneous coordinate of pixel $\boldsymbol{p}$, and d_i denotes the distance from the plane of $\mathcal{G}_i$ to the camera center.

3.2 Dual-Evidence Depth Integration

We formulate depth recovery as an integration problem over two complementary forms of evidence. Point evidence $\{\hat{\boldsymbol{D}}_r(\boldsymbol{p})\}_{\boldsymbol{p}\in\mathcal{A}}$ provides absolute depth constraints at a set of anchor pixels $\mathcal{A}$, while edge evidence $\{g_{\boldsymbol{pq}}\}_{(\boldsymbol{p},\boldsymbol{q})\in\mathcal{E}}$ encodes log-depth differences over a set of retained neighboring pixel pairs $\mathcal{E}$:

$$g_{\boldsymbol{pq}} = \log \hat{\boldsymbol{D}}_{pr}(\boldsymbol{q}) - \log \hat{\boldsymbol{D}}_{pr}(\boldsymbol{p}), \qquad (\boldsymbol{p},\boldsymbol{q}) \in \mathcal{E}, \tag{3}$$

where $\hat{\boldsymbol{D}}_{pr}$ is derived from VGGT depth $\boldsymbol{D}_{pr}$. We integrate the two forms of evidence by solving for a dense depth map $\boldsymbol{D}_{int}$:

$$\min_{\log \boldsymbol{D}_{int}} \lambda \sum_{\boldsymbol{p}\in\mathcal{A}} \rho\Big(\log \boldsymbol{D}_{int}(\boldsymbol{p}) - \log \hat{\boldsymbol{D}}_r(\boldsymbol{p})\Big) + \sum_{(\boldsymbol{p},\boldsymbol{q})\in\mathcal{E}} \rho(\log \boldsymbol{D}_{int}(\boldsymbol{q}) - \log \boldsymbol{D}_{int}(\boldsymbol{p}) - g_{\boldsymbol{pq}}), \tag{4}$$

where λ balances the point and edge evidence, and $\rho(\cdot)$ denotes the Huber penalty. We solve the resulting optimization problem using iteratively reweighted least squares (IRLS) with a preconditioned

conjugate gradient (PCG) solver (Appendix A.1). By coupling local edge relations with distributed absolute anchors, the integration limits error accumulation over spatial distance and produces dense depth supervision for 3DGS.

3.3 Evidence Construction

3.3.1 Point Evidence

We construct point evidence through multi-view photometric matching between the input images. Specifically, for each pixel in the reference view, we evaluate photometric consistency over a set of candidate depths to obtain a matching-cost distribution. The resulting distribution characterizes both correspondence reliability and the associated depth, providing the basis for anchor selection and depth refinement.

Matching cost construction. For each pixel $\boldsymbol{p}$ in the reference view, we uniformly sample K depth hypotheses in the disparity domain around the rendered depth $\boldsymbol{D}_r(\boldsymbol{p})$. For each hypothesis, the reference patch centered at $\boldsymbol{p}$ is warped to each source view using the corresponding plane-induced homography. The matching cost is defined by the best photometric agreement across source views:

$$c_k(\boldsymbol{p}) = \min_{j \in \mathcal{V}} \left(1 - \mathrm{NCC}_{j,k}(\boldsymbol{p})\right), \tag{5}$$

where $\mathcal{V}$ denotes the set of source views and $\mathrm{NCC}_{j,k}(\boldsymbol{p})$ measures the normalized cross-correlation between the reference patch and its warped counterpart in source view j. The costs over all hypotheses constitute the matching-cost distribution at $\boldsymbol{p}$.

Prior-based aggregation. Per-pixel matching distributions can be ambiguous under sparse views, exhibiting flat responses or multiple competing minima that make individual depth hypotheses difficult to assess reliably. Inspired by cost aggregation in learning-based MVS, we introduce a parameter-free prior-based aggregation scheme that uses local relative depth from the learned prior to align neighboring matching distributions before aggregation. Specifically, for a reference pixel $\boldsymbol{p}$ and its neighbor $\boldsymbol{q}$, their relative depth ratio is computed as:

$$\gamma = \frac{\boldsymbol{D}_{pr}(\boldsymbol{q})}{\boldsymbol{D}_{pr}(\boldsymbol{p})}. \tag{6}$$

Given the k-th depth hypothesis at $\boldsymbol{p}$, the corresponding hypothesis index at $\boldsymbol{q}$ is

$$t(k) = k_0 + \frac{k - k_0}{\gamma} + \delta, \qquad \delta = \frac{1}{\gamma}\frac{fb}{\boldsymbol{D}_r(\boldsymbol{p})} - \frac{fb}{\boldsymbol{D}_r(\boldsymbol{q})}, \tag{7}$$

where k_0 denotes the center index of the hypothesis space, f is the focal length, and b is the average baseline from the reference view to the source views. The offset δ accounts for the difference between the disparity centers at $\boldsymbol{p}$ and $\boldsymbol{q}$. For non-integer $t(k)$, the corresponding matching cost is obtained by linear interpolation. We then aggregate the aligned matching distributions within the local 5×5 window $\mathcal{W}$:

$$\tilde{c}_k(\boldsymbol{p}) = \frac{1}{|\mathcal{W}|} \sum_{\boldsymbol{q} \in \mathcal{W}} c_{t(k)}(\boldsymbol{q}). \tag{8}$$

The aggregated distribution $\tilde{c}(\boldsymbol{p})$ suppresses ambiguous responses and yields a more distinct minimum around the matched depth (Fig. 5).

Reliability assessment and anchor depth refinement. Following conventional MVS (Wang et al., 2023), we assess matching reliability from the shape of the aggregated cost distribution. A pixel $\boldsymbol{p}$ is retained as an anchor if its global minimum is sufficiently low and well separated from the other local minima, indicating an unambiguous photometric match. All retained pixels constitute the anchor set $\mathcal{A}$. The reliability assessment determines where point constraints can be established, while their depths are determined from the aggregated matching distribution. Recall that the rendered depth $\boldsymbol{D}_r(\boldsymbol{p})$ only defines the center of the hypothesis search and may deviate from the true surface. For each $\boldsymbol{p} \in \mathcal{A}$, we therefore locate the global minimum of the aggregated distribution and use its corresponding depth to correct the initial rendered depth. The final anchor depth $\hat{\boldsymbol{D}}_r(\boldsymbol{p})$ is obtained with three-point parabolic interpolation around the minimum.

3.3.2 EDGE EVIDENCE

We construct edge evidence on a 4-connected pixel grid using log-depth differences from the VGGT depth prior. Edges crossing significant depth discontinuities are removed, and the remaining neighboring pixel pairs constitute the edge set $\mathcal{E}$. VGGT predicts depth $\boldsymbol{D}_{pr}$ up to a global scale, which cancels in the log-depth domain. In practice, however, we observe a residual depth shift and model the prior as $\boldsymbol{D}_{pr} \approx s(\boldsymbol{D}^* + t)$, where s and t are shared across the image. While s cancels in log-depth differences, the residual shift t introduces an error which can be approximated as:

$$\Delta g_{pq} \approx -t \frac{\boldsymbol{D}^*(\boldsymbol{q}) - \boldsymbol{D}^*(\boldsymbol{p})}{\boldsymbol{D}^*(\boldsymbol{p})\boldsymbol{D}^*(\boldsymbol{q})}, \tag{9}$$

where Δg_{pq} denotes the error between the predicted and true log-depth differences. Such errors can accumulate over spatial distance during integration. To mitigate this effect, we perform *de-shift* using the reliable anchor depths. Specifically, we fit $\boldsymbol{D}_{pr}(\boldsymbol{p}) \approx a\hat{\boldsymbol{D}}_r(\boldsymbol{p}) + b$ over $\boldsymbol{p} \in \mathcal{A}$, where a captures the unknown scale and b the residual shift. Since the scale cancels in log-depth differences, we remove only the estimated shift, yielding $\hat{\boldsymbol{D}}_{pr} = \boldsymbol{D}_{pr} - \hat{b}$ for constructing the edge evidence.

3.4 OPTIMIZATION OBJECTIVES

Our optimization objective combines the standard reconstruction and geometric regularization terms with our proposed geometric supervision. Specifically, we adopt the image reconstruction loss $\mathcal{L}_{rgb}$ from 3DGS and the flattening loss $\mathcal{L}_s$ and depth-normal consistency loss $\mathcal{L}_{dn}$ from PGSR (Chen et al., 2024), together with normal loss $\mathcal{L}_n$ and our depth loss $\mathcal{L}_d$:

$$\mathcal{L} = \mathcal{L}_{rgb} + \lambda_s\mathcal{L}_s + \lambda_{dn}\mathcal{L}_{dn} + \lambda_n\mathcal{L}_n + \lambda_d\mathcal{L}_d, \qquad \mathcal{L}_d = \|\boldsymbol{D}_r - \boldsymbol{D}_{int}\|_1 . \tag{10}$$

Here, $\mathcal{L}_s$ flattens each Gaussian into a plane by minimizing its smallest scale factor, and $\mathcal{L}_n$ supervises the rendered normal with the normal derived from the first-order differences of the VGGT depth. We set the loss weights to $\lambda_s = 100$, $\lambda_{dn} = 0.015$, $\lambda_n = 3$, and $\lambda_d = 5$.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETTINGS

Implementation details. We adopt Planar-based Gaussian Splatting from PGSR (Chen et al., 2024) as the scene representation and initialize the Gaussian primitives from the MASt3R-SfM (Duisterhof et al., 2025) point cloud following (Guédon et al., 2025b). Each scene is optimized for 6,000 iterations. The depth loss is introduced after 2,000 iterations, with the integrated depth $\boldsymbol{D}_{int}$ updated every 1,000 iterations thereafter. After optimization, the final mesh is extracted by applying TSDF fusion (Newcombe et al., 2011) to the rendered depth maps. All experiments are conducted on a single NVIDIA RTX 3090 GPU.

Datasets. We evaluate our method on DTU (Aanæs et al., 2016) and Tanks and Temples (TnT) (Knapitsch et al., 2017). For DTU, we follow the three-view low-overlap protocol of PixelNerf (Yu et al., 2021) and evaluate on 15 scenes using Chamfer Distance (CD). For TnT, we follow the 10-view setting of MAtCha (Guédon et al., 2025b) and report the standard F-score for evaluation.

4.2 BENCHMARK COMPARISONS

We first evaluate our method on DTU against both implicit and explicit reconstruction methods. For a fair comparison, all explicit methods are initialized with the same MASt3R-SfM point cloud. As shown in Tab. 1, our method achieves the best overall performance with an average Chamfer Distance of 0.71, improving upon DP-GS (0.85) and VGGS (0.91) by 16% and 22%, respectively. Across the 15 scenes, our method ranks first on 13 scenes and second on the remaining two. Fig. 3 further presents qualitative comparisons of the reconstructed meshes. Our method recovers more complete surfaces and preserves finer geometric details under sparse-view inputs.

We further conduct comparisons on the Tanks and Temples (TnT) dataset to assess performance on more complex real-world scenes. As shown in Tab. 2, our method achieves the best average F-score

Table 1: **Quantitative comparison of Chamfer Distance on the DTU dataset**. Red, orange and yellow denote the best, second-best and third-best results.

	Scan ID	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	Mean ↓
Implicit	SparseNeuS$_{ft}$	4.81	5.56	5.81	2.68	3.30	3.88	2.39	2.91	3.08	2.33	2.64	3.12	1.74	3.55	2.31	3.34
	VolRecon	3.05	4.45	3.36	3.09	2.78	3.68	3.01	2.87	3.07	2.55	3.07	2.77	1.59	3.44	2.51	3.02
	GenS$_{ft}$	7.67	8.58	5.41	5.88	7.18	5.26	4.40	6.29	6.19	4.08	6.04	5.29	4.69	4.35	4.01	5.69
	ReTR	3.30	3.60	3.37	2.85	2.91	3.07	2.26	2.29	2.25	2.05	2.80	2.80	1.52	2.29	2.11	2.63
	UFORecon	1.51	2.58	1.82	1.44	1.60	1.81	1.04	1.56	0.96	1.40	1.20	0.93	0.66	1.26	1.26	1.40
	MonoSDF	3.47	3.61	2.10	1.05	2.37	1.38	1.41	1.85	1.74	1.10	1.46	2.28	1.25	1.44	1.45	1.86
	NeuSurf	1.35	3.25	2.50	0.80	1.21	2.35	0.77	1.19	1.20	1.05	1.05	1.21	0.41	0.80	1.08	1.35
	SparseCraft	2.42	2.79	2.78	0.74	1.44	2.51	1.26	1.42	1.65	1.10	1.34	5.24	0.65	0.88	1.16	1.83
	SparseRecon	1.26	1.46	1.39	0.71	1.20	2.38	0.70	1.23	0.92	0.80	0.94	0.77	0.44	0.83	0.91	1.06
Explicit	PGSR	2.29	3.85	3.26	1.29	2.62	2.53	1.36	2.02	2.31	1.75	1.68	2.19	0.97	2.10	1.89	2.14
	FatesGS	0.65	2.80	1.48	0.71	0.90	1.58	0.66	1.11	0.86	0.76	0.96	0.79	0.43	0.71	0.71	1.01
	MAtCha	0.88	2.79	1.40	0.83	0.93	1.35	0.87	1.25	1.51	0.95	1.15	1.02	0.56	1.01	0.92	1.16
	DP-GS	0.57	1.71	1.37	0.71	0.84	1.11	0.64	1.04	0.70	0.74	0.76	0.65	0.45	0.66	0.76	0.85
	VGGS	0.60	1.84	1.05	0.73	1.08	0.94	0.71	1.06	0.90	0.78	0.79	1.20	0.50	0.74	0.78	0.91
	Ours	0.43	1.38	0.69	0.62	0.89	0.73	0.55	1.00	0.72	0.71	0.65	0.48	0.39	0.65	0.69	0.71

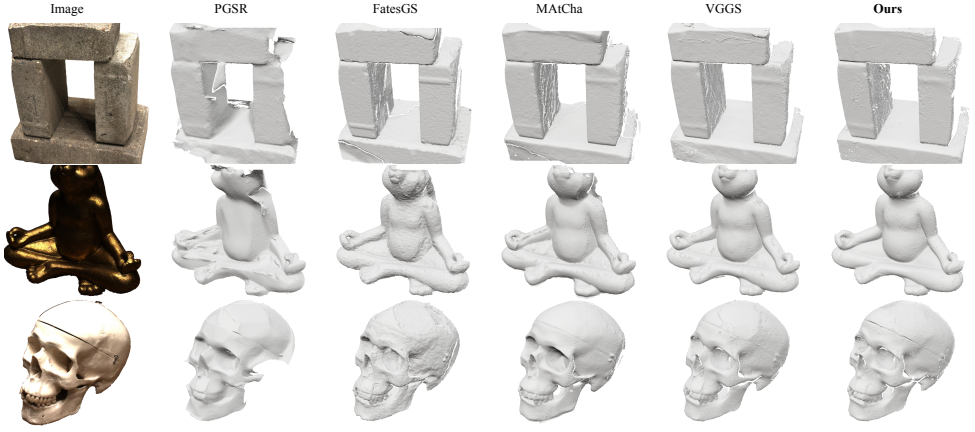

Figure 3: **Visual comparison on the DTU dataset.** Our method achieves improved reconstruction quality across different scenes.

of 0.3100, surpassing the previous best result of 0.2659 from MAtCha by 16.6%. The qualitative comparisons in Fig. 4 reveal distinct geometric degradation under sparse-view inputs, including large-scale distortions in PGSR, noisy surfaces in VGGS, and spurious structures in MAtCha. Our method substantially alleviates these artifacts, yielding cleaner and more coherent surfaces with better-preserved geometric structures. Additional qualitative comparisons on the BlendedMVS Yao et al. (2020) dataset are provided in Appendix A.3.3.

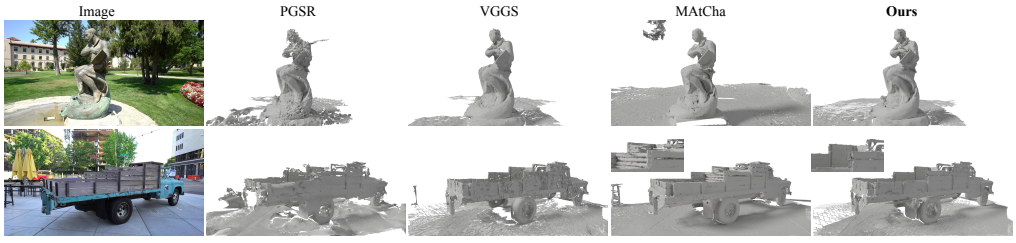

Figure 4: **Mesh visualization on the Tanks and Temples dataset.** Our method produces cleaner and more coherent surfaces with fewer geometric artifacts under sparse-view inputs.

Table 2: **Quantitative comparison on the Tanks and Temples dataset.** Our method achieves the best F-score (↑) across all scenes.

Scene	PGSR	VGGS	MAtCha	Ours
Barn	0.0434	0.1322	0.2010	**0.2371**
Caterpillar	0.0431	0.0967	0.1300	**0.1331**
Ignatius	0.2168	0.2215	0.3087	**0.3530**
Truck	0.2841	0.3384	0.4240	**0.5168**
Mean	0.1469	0.1972	0.2659	**0.3100**

Table 3: **Ablation study on the DTU dataset.** PBA and ADR denote the Prior-based Aggregation and Anchor Depth Refinement, respectively.

Settings	$\mathcal{L}_n$	$\mathcal{L}_d$				CD ↓
		Point evidence		Edge evidence		
		PBA	ADR	De-shift		
(a)						1.55
(b)	✓					1.49
(c)	✓	✓	✓			0.76
(d)	✓		✓	✓		0.83
(e)	✓	✓		✓		0.78
(f)	✓	✓	✓	✓		**0.71**

4.3 Ablations and Analysis

To assess the contribution of each component, we conduct ablation studies on the DTU dataset, with the results summarized in Tab. 3. Introducing normal supervision reduces the CD from 1.55 to 1.49. Adding our integrated depth supervision further reduces the CD from 1.49 to 0.71 (rows (b) and (f)), corresponding to a 52.3% reduction and demonstrating its effectiveness in improving reconstruction accuracy. We further ablate the components used to construct point evidence. Introducing PBA (rows (d) and (f)) reduces the CD from 0.83 to 0.71. As illustrated in Fig. 5, PBA reduces ambiguity in the raw matching cost distributions, yielding cleaner anchor depths with broader spatial coverage. ADR further improves anchor reliability, with its removal (rows (e) and (f)) increasing the CD by 0.07. Together, PBA and ADR provide more reliable point constraints for depth integration. For edge evidence, introducing de-shift further improves the reconstruction performance (rows (c) and (f)). The combination of all these components achieves the best performance.

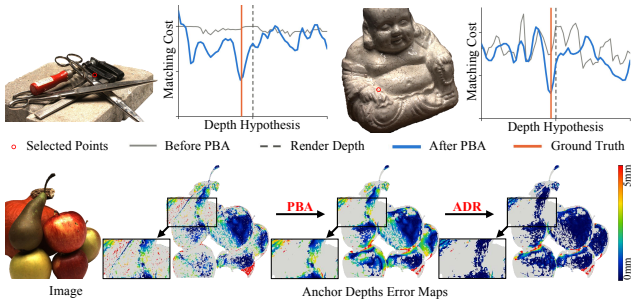

Figure 5: **Visualization of point evidence construction.**

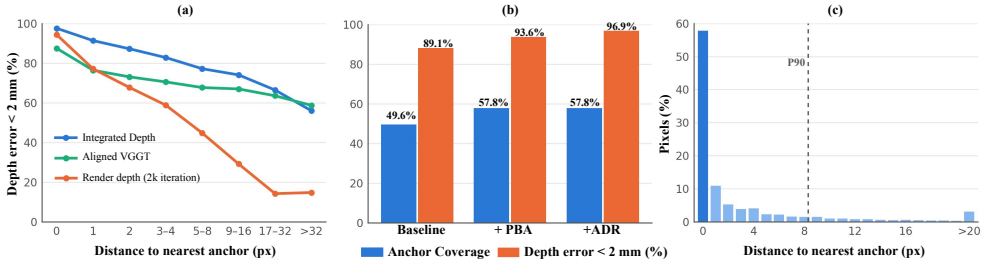

Figure 6: **Analysis of point evidence.** (a) Pixels with depth error below 2 mm versus distance to the nearest anchor. (b) Anchor coverage and accuracy with PBA and ADR. (c) Distribution of the distance to the nearest anchor (dashed line: 90th percentile).

Analysis of point evidence. Fig. 6 analyzes the anchors and their effect on depth integration at the first depth update (iteration 2k) on the DTU dataset. Fig. 6a reports depth accuracy as a function of the distance to the nearest anchor. The rendered depth is highly accurate near the anchors but degrades rapidly with distance, whereas VGGT remains relatively stable as photometric evidence weakens. Notably, VGGT is less accurate than the rendered depth near the anchors, suggesting that directly using its predicted depth as supervision can introduce errors in regions already supported by reliable photometric evidence. Our integrated depth combines the complementary strengths of both sources, outperforming both within 32 px of the nearest anchor and remaining comparable

Table 4: **Ablation on different pseudo-GT depth construction strategies.**

Method	d2s↓	s2d↓	Overall↓
Baseline (with normal sup.)	1.382	1.588	1.485
(a) Aligned VGGT	0.670	1.144	0.907
(b) PatchMatch	1.300	1.568	1.434
(c) Anchor Depth	0.896	1.222	1.059
(d) Binary Fusion	0.602	1.010	0.806
(e) Integrated Depth	**0.504**	**0.910**	**0.707**

Table 5: **Comparison of efficiency and performance on the DTU dataset.**

Method	CD ↓	Time ↓	Memory (GB) ↓
NeuSurf	1.35	14 h	8.0
SparseCraft	1.83	10 min	10.0
FatesGS	1.01	23 min	4.5
MAtCha	1.16	8.6 min	4.3
VGGS	0.91	**1.8 min**	**2.4**
Ours	**0.71**	4.3 min	6.0

to VGGT at larger distances. Fig. 6b examines the quality of the constructed anchors. PBA improves both coverage and accuracy over per-pixel photometric matching, and ADR further refines their depths, increasing the anchor accuracy to 96.9%. Together, they provide accurate and well-distributed absolute references (Fig. 6c) for the depth integration.

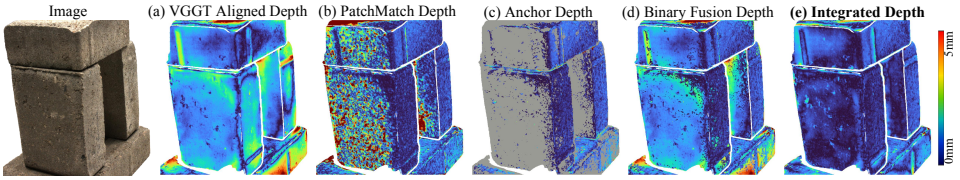

Figure 7: **Depth error visualization of different pseudo-GT depth construction strategies.**

Comparison of different depth supervision. We compare different depth supervision strategies under a unified backbone: (a) anchor-aligned VGGT depth (Xiang et al., 2026); (b) PatchMatch-refined depth (Wang et al., 2024a); and three variants based on our point evidence: (c) sparse anchor depths only; (d) binary fusion, which uses anchor depths where available and aligned VGGT depth elsewhere; and (e) our integrated depth. All methods share the same backbone and optimization schedule. Each scene is optimized for 6k iterations, with depth supervision introduced at iteration 2k and updated every 1k iterations thereafter based on the current reconstruction. Tab. 4 reports the quantitative results. Aligned VGGT depth improves reconstruction by providing dense geometric supervision, although errors in its absolute depth limit the resulting surface accuracy. PatchMatch refinement brings only marginal improvement over the baseline, as its depth estimates become noisy under sparse-view inputs (Fig. 7(b)). Anchors provide more accurate absolute depth constraints, but their limited coverage prevents them from effectively constraining the entire surface. Binary fusion reduces the overall error to 0.806, confirming the complementarity of the two sources. However, simply replacing the prior at anchor locations leaves errors elsewhere unchanged (Fig. 7(d)). Our integration instead incorporates the anchor constraints into dense depth recovery, achieving the lowest errors across all metrics.

Efficiency and performance. We compare different methods in terms of reconstruction quality, optimization time, and peak memory usage. As shown in Tab. 5, our method achieves the best reconstruction quality with moderate computational overhead.

5 Conclusion

We presented DGS, a dual-evidence integration framework for sparse-view surface reconstruction. We show that aligned VGGT depth remains inaccurate even with reliable photometric evidence, and its local depth variations exhibit range-dependent reliability. Based on these observations, DGS integrates accurate photometric point evidence with locally reliable edge evidence, further enhanced by PBA and ADR. DGS achieves state-of-the-art results on DTU and Tanks and Temples datasets. The effectiveness of DGS relies on the availability and reliability of the two types of geometric evidence. Regions with sparse photometric anchors remain under-constrained, with the benefit of integration diminishing as the anchor distance increases and becoming limited beyond 32 pixels (Fig. 6a). Moreover, in regions where both point and edge evidence contain substantial errors, their integration may fail to recover the correct geometry. Future work will explore complementary geometric constraints to improve reconstruction in these challenging regions.

AI USE STATEMENT

We used large language models (LLMs) for three purposes: (i) editing the manuscript to improve the clarity and readability of the writing; (ii) assisting with the layout and visual presentation of figures. All AI-assisted outputs were reviewed and verified by the authors. AI-edited text was checked to ensure that the intended technical meaning was preserved and that no unsupported claims were introduced.

REPRODUCIBILITY STATEMENT

We have taken care to provide the information necessary to reproduce the results reported in this paper. Appendix A provides a detailed description of our method, including its formulation, implementation details, and experimental settings. The datasets, preprocessing procedures, evaluation protocols, hyperparameter settings, and computational resources used in our experiments are specified in the main paper and appendix.

REFERENCES

Henrik Aanæs, Rasmus Ramsbøl Jensen, George Vogiatzis, Engin Tola, and Anders Bjorholm Dahl. Large-scale data for multiple-view stereopsis. *International Journal of Computer Vision*, 120: 153–168, 2016.

Xu Cao, Hiroaki Santo, Boxin Shi, Fumio Okura, and Yasuyuki Matsushita. Bilateral normal integration. In *European Conference on Computer Vision*, pp. 552–567. Springer, 2022.

Danpeng Chen, Hai Li, Weicai Ye, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 31 (9):6100–6111, 2024.

Bardienus Pieter Duisterhof, Lojze Zust, Philippe Weinzaepfel, Vincent Leroy, Yohann Cabon, and Jerome Revaud. Mast3r-sfm: a fully-integrated solution for unconstrained structure-from-motion. In *2025 International Conference on 3D Vision (3DV)*, pp. 1–10. IEEE, 2025.

Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10766–10776. IEEE, 2021.

Robert T. Frankot and Rama Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 10(4):439–451, 1988.

Qiancheng Fu, Qingshan Xu, Yew Soon Ong, and Wenbing Tao. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 35:3403–3416, 2022.

Silvano Galliani, Katrin Lasinger, and Konrad Schindler. Massively parallel multiview stereopsis by surface normal diffusion. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 873–881, 2015.

Diego Gomez, Antoine Guédon, Nissim Maruani, Bingchen Gong, and Maks Ovsjanikov. From blobs to spokes: High-fidelity surface reconstruction via oriented gaussians. In *European Conference on Computer Vision*, pp. 54–72. Springer, 2026.

Meiying Gu, Jiawei Zhang, Jiahe Li, Xiaohan Yu, Haonan Luo, Jin Zheng, and Xiao Bai. Sparsesurf: Sparse-view 3d gaussian splatting for surface reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 4311–4319, 2026.

Antoine Guédon, Diego Gomez, Nissim Maruani, Bingchen Gong, George Drettakis, and Maks Ovsjanikov. Milo: Mesh-in-the-loop gaussian splatting for detailed and efficient surface reconstruction. *ACM Transactions on Graphics (TOG)*, 44(6):1–15, 2025a.

Antoine Guédon, Tomoki Ichikawa, Kohei Yamashita, and Ko Nishino. Matcha gaussians: Atlas of charts for high-quality geometry and photorealism from sparse views. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6001–6011. IEEE, 2025b.

Liang Han, Xu Zhang, Haichuan Song, Kanle Shi, Yu-Shen Liu, and Zhizhong Han. Sparserecon: Neural implicit surface reconstruction from sparse views with feature and depth consistencies. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 28514–28524. IEEE, 2025.

Liang Han, Bangcai Wei, Junsheng Zhou, Yu-Shen Liu, and Zhizhong Han. Sparse-view surface reconstruction using gaussian splatting through high-confidence depth propagation with normal priors. In *European Conference on Computer Vision*, pp. 491–511. Springer, 2026.

Richard Hartley, Andrew Zisserman, et al. *Multiple view geometry in computer vision*, volume 2. Cambridge university press Cambridge, 2003.

Berthold KP Horn and Michael J Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, 1986.

Itsik Horovitz and Nahum Kiryati. Depth from gradient fields and control points: Bias correction in photometric stereo. *Image and Vision Computing*, 22(9):681–694, 2004.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, pp. 1–11, 2024a.

Han Huang, Yulun Wu, Junsheng Zhou, Ge Gao, Ming Gu, and Yu-Shen Liu. Neusurf: On-surface priors for neural surface reconstruction from sparse input views. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 2312–2320, 2024b.

Han Huang, Yulun Wu, Chao Deng, Ge Gao, Ming Gu, and Yu-Shen Liu. Fatesgs: Fast and accurate sparse-view surface reconstruction using gaussian splatting with depth-feature consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3644–3652, 2025.

Xirui Jin, Renbiao Jin, Boying Li, Danping Zou, and Wenxian Yu. Planargs: High-fidelity indoor 3d gaussian splatting guided by vision-language planar priors. *Advances in Neural Information Processing Systems*, 38:36342–36370, 2026.

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

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.

Lingyu Kong, Ruicheng Li, Ruicheng Wang, Sicheng Xu, Chengtang Yao, Jianfeng Xiang, and Jiaolong Yang. Moge-3: Fine-detail monocular geometry estimation with self-guided sparse volumetric refinement. *arXiv e-prints*, pp. arXiv–2607, 2026.

Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xiaohan Yu, Lin Gu, and Gim Hee Lee. Revisiting photometric ambiguity for accurate gaussian-splatting surface reconstruction. *arXiv preprint arXiv:2605.12494*, 2026.

Yixun Liang, Hao He, and Yingcong Chen. Retr: Modeling rendering via transformer for generalizable neural surface reconstruction. *Advances in neural information processing systems*, 36: 62332–62351, 2023.

Haotong Lin, Sili Chen, Junhao Liew, Donny Y Chen, Zhenyu Li, Guang Shi, Jiashi Feng, and Bingyi Kang. Depth anything 3: Recovering the visual space from any views. *arXiv preprint arXiv:2511.10647*, 2025.

Xiaoxiao Long, Cheng Lin, Peng Wang, Taku Komura, and Wenping Wang. Sparseneus: Fast generalizable neural surface reconstruction from sparse views. In *European Conference on Computer Vision*, pp. 210–227. Springer, 2022.

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

Youngju Na, Woo Jae Kim, Kyu Beom Han, Suhyeon Ha, and Sung-Eui Yoon. Uforecon: Generalizable sparse-view surface reconstruction from arbitrary and unfavorable sets. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5094–5104. IEEE, 2024.

Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM transactions on graphics (TOG)*, 24(3): 536–543, 2005.

Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE international symposium on mixed and augmented reality*, pp. 127–136. Ieee, 2011.

Zador Pataki, Paul-Edouard Sarlin, Johannes L Schönberger, and Marc Pollefeys. Mp-sfm: Monocular surface priors for robust structure-from-motion. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21891–21901. IEEE, 2025.

Rui Peng, Xiaodong Gu, Luyang Tang, Shihe Shen, Fanqi Yu, and Ronggang Wang. Gens: Generalizable neural surface reconstruction from multi-view images. *Advances in Neural Information Processing Systems*, 36:56932–56945, 2023.

Yvain Quéau, Jean-Denis Durou, and Jean-François Aujol. Normal integration: a survey. *Journal of Mathematical Imaging and Vision*, 60(4):576–593, 2018a.

Yvain Quéau, Jean-Denis Durou, and Jean-François Aujol. Variational methods for normal integration. *Journal of Mathematical Imaging and Vision*, 60(4):609–632, 2018b.

Yufan Ren, Fangjinhua Wang, Tong Zhang, Marc Pollefeys, and Sabine Süsstrunk. Volrecon: Volume rendering of signed ray distance functions for generalizable multi-view reconstruction. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16685–16695. IEEE, 2023.

Johannes L Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise view selection for unstructured multi-view stereo. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pp. 501–518. Springer, 2016.

Matias Turkulainen, Xuqian Ren, Iaroslav Melekhov, Otto Seiskari, Esa Rahtu, and Juho Kannala. Dn-splatter: Depth and normal priors for gaussian splatting and meshing. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2421–2431. IEEE, 2025.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5294–5306. IEEE, 2025.

Jiepeng Wang, Yuan Liu, Peng Wang, Cheng Lin, Junhui Hou, Xin Li, Taku Komura, and Wenping Wang. Gaussurf: Geometry-guided 3d gaussian splatting for surface reconstruction. *arXiv preprint arXiv:2411.19454*, 2024a.

Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20697–20709. IEEE, 2024b.

Yifan Wang, Jianjun Zhou, Haoyi Zhu, Wenzheng Chang, Yang Zhou, Zizun Li, Junyi Chen, Jiangmiao Pang, Chunhua Shen, and Tong He. π^3: Permutation-equivariant visual geometry learning. In *International Conference on Learning Representations*, volume 2026, pp. 10481–10497, 2026.

Yuesong Wang, Zhaojie Zeng, Tao Guan, Wei Yang, Zhuo Chen, Wenkai Liu, Luoyuan Xu, and Yawei Luo. Adaptive patch deformation for textureless-resilient multi-view stereo. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1621–1630. IEEE, 2023.

Jiang Wu, Rui Li, Haofei Xu, Wenxun Zhao, Yu Zhu, Jinqiu Sun, and Yanning Zhang. Gomvs: Geometrically consistent cost aggregation for multi-view stereo. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20207–20216. IEEE, 2024.

Jiang Wu, Rui Li, Yu Zhu, Rong Guo, Jinqiu Sun, and Yanning Zhang. Sparse2dgs: Geometry-prioritized gaussian splatting for surface reconstruction from sparse views. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11307–11316. IEEE, 2025.

Peng Xiang, Liang Han, Hui Zhang, Yu-Shen Liu, and Zhizhong Han. Vggs: Vggt-guided gaussian splatting for efficient and faithful sparse-view surface reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 10969–10977, 2026.

Luoyuan Xu, Tao Guan, Yuesong Wang, Wenkai Liu, Zhaojie Zeng, Junle Wang, and Wei Yang. C2f2neus: Cascade cost frustum fusion for high fidelity and generalizable neural surface reconstruction. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 18245–18255. IEEE, 2023.

Qingshan Xu and Wenbing Tao. Multi-scale geometric consistency guided multi-view stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5483–5492, 2019.

Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. Mvsnet: Depth inference for unstructured multi-view stereo. In *European conference on computer vision*, pp. 785–801. Springer, 2018.

Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1787–1796. IEEE, 2020.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in neural information processing systems*, 34:4805–4815, 2021.

Wei Yin, Chi Zhang, Hao Chen, Zhipeng Cai, Gang Yu, Kaixuan Wang, Xiaozhi Chen, and Chunhua Shen. Metric3d: Towards zero-shot metric 3d prediction from a single image. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9009–9019. IEEE, 2023.

Mae Younes, Amine Ouasfi, and Adnane Boukhayma. Sparsecraft: Few-shot neural reconstruction through stereopsis guided geometric linearization. In *European Conference on Computer Vision*, pp. 37–56. Springer, 2024.

Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4576–4585. IEEE, 2021.

Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in neural information processing systems*, 35:25018–25032, 2022.

Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics (ToG)*, 43(6):1–13, 2024.

Yinda Zhang and Thomas Funkhouser. Deep depth completion of a single rgb-d image. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 175–185. IEEE, 2018.

Zhe Zhang, Rui Peng, Yuxi Hu, and Ronggang Wang. Geomvsnet: Learning multi-view stereo with geometry perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21508–21518, 2023.

A Appendix

A.1 Details of Dual-Evidence Depth Integration

IRLS. We solve the integration objective in Eq. 4 using IRLS. For an anchor $\boldsymbol{p} \in \mathcal{A}$ and an edge $(\boldsymbol{p}, \boldsymbol{q}) \in \mathcal{E}$, the corresponding residuals are

$$\begin{aligned} r_{\boldsymbol{p}} &= \log \boldsymbol{D}_{int}(\boldsymbol{p}) - \log \hat{\boldsymbol{D}}_r(\boldsymbol{p}), \\ r_{\boldsymbol{pq}} &= \log \boldsymbol{D}_{int}(\boldsymbol{q}) - \log \boldsymbol{D}_{int}(\boldsymbol{p}) - g_{\boldsymbol{pq}}. \end{aligned} \tag{11}$$

For the Huber penalty with threshold ϵ, the IRLS weight of a residual r is

$$w(r;\epsilon) = \begin{cases} 1, & |r| \le \epsilon, \\ \dfrac{\epsilon}{|r|}, & |r| > \epsilon. \end{cases} \tag{12}$$

We set separate thresholds for point and edge residuals as $\epsilon_{point} = 0.25R/256$ and $\epsilon_{edge} = 0.05R/256$, where $R = (\max \log \boldsymbol{D}_r - \min \log \boldsymbol{D}_r)$ is the log-depth range of the current view. Accordingly, $\nu_{\boldsymbol{p}} = w(r_{\boldsymbol{p}}; \epsilon_{point})$ and $w_{\boldsymbol{pq}} = w(r_{\boldsymbol{pq}}; \epsilon_{edge})$. The Huber weighting reduces the influence of occasional outlier residuals.

Let $\boldsymbol{u} = \log \boldsymbol{D}_{int}$ denote the unknown log-depth field. With the weights fixed, each IRLS iteration reduces to a weighted least-squares problem:

$$(\boldsymbol{L}_w + \boldsymbol{\Lambda})\, \boldsymbol{u} = \boldsymbol{b}, \tag{13}$$

where $\boldsymbol{L}_w$ is the weighted graph Laplacian induced by the edge weights and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda \nu_{\boldsymbol{p}} [\boldsymbol{p} \in \mathcal{A}])$ contains the weighted anchor constraints. The right-hand side $\boldsymbol{b}$ is assembled from the edge targets $w_{\boldsymbol{pq}} g_{\boldsymbol{pq}}$ and the anchor targets $\lambda \nu_{\boldsymbol{p}} \log \hat{\boldsymbol{D}}_r(\boldsymbol{p})$.

Linear solve. We solve Eq. 13 using Jacobi-preconditioned conjugate gradient (PCG), with the stopping criterion $\|\boldsymbol{r}_{cg}\| \le 10^{-6} \|\boldsymbol{b}\|$ and a maximum of 20,000 iterations. We perform up to 4 IRLS iterations.

Effect of the robust penalty. Tab. 6 ablates the penalty $\rho(\cdot)$ on the two terms of Eq. 4. With a quadratic penalty on both terms, the objective reduces to a single sparse linear solve and already provides accurate depth, indicating that the integration itself is effective. Introducing the Huber penalty on either term improves the accuracy by suppressing outlier anchors and edges, and applying it to both yields a slightly better result.

Table 6: **Ablation of the penalty on the anchor and edge terms.** Depth accuracy of the integrated depth on DTU.

Anchor term	Edge term	<2 mm (%) ↑	<4 mm (%) ↑	MAE (mm) ↓
ℓ_2	ℓ_2	89.83	95.48	1.67
ℓ_2	Huber	90.19	95.65	1.65
Huber	ℓ_2	91.04	96.45	1.53
Huber	Huber (ours)	**91.16**	**96.49**	**1.51**

Effect of the balancing factor. Tab. 7 varies the balancing factor λ over three orders of magnitude. The accuracy is stable for small λ, whereas a large λ forces the solution to follow every anchor closely, so that anchor errors are passed on to the integrated depth, especially in regions far from other anchors. We use $\lambda = 0.03$ in all experiments.

A.2 Details of Point Evidence Construction

Hypothesis sampling. Recall that f is the focal length and b the average baseline from the reference view to the source views. The K depth hypotheses at pixel $\boldsymbol{p}$ are uniformly spaced in the disparity domain and centered on the rendered depth:

$$\boldsymbol{D}_r^{(k)}(\boldsymbol{p}) = \frac{fb}{\dfrac{fb}{\boldsymbol{D}_r(\boldsymbol{p})} + (k - k_0)}, \qquad k = 0, \dots, K-1, \tag{14}$$

Table 7: **Ablation of the balancing factor λ.** Depth accuracy of the integrated depth on DTU.

λ	<2 mm (%) ↑	<4 mm (%) ↑	<8 mm (%) ↑	MAE (mm) ↓
0.003	90.37	96.25	98.27	**1.51**
0.03	**91.16**	**96.49**	**98.32**	**1.51**
0.3	90.69	95.14	97.71	1.67
3	89.59	93.95	97.00	1.85

where $k_0 = (K-1)/2$ is the center index, so that $\boldsymbol{D}_r^{(k_0)}(\boldsymbol{p}) = \boldsymbol{D}_r(\boldsymbol{p})$. We use $K = 61$ with a spacing of one pixel of disparity, which keeps the sampling independent of the scene scale.

Source views. For each reference view, source views are ranked by the angle between their viewing direction and that of the reference view, and those with the smallest angles are selected. We use 2 and 4 views on DTU and Tanks and Temples datasets, respectively.

Derivation of the hypothesis alignment (Eq. 7). Since hypothesis indices are defined relative to each pixel's rendered depth, the same index generally corresponds to different depths across neighboring pixels. We therefore align neighboring hypotheses according to the prior depth ratio γ in Eq. 6. Given a hypothesis $\boldsymbol{D}_r^{(k)}(\boldsymbol{p})$ at pixel $\boldsymbol{p}$, its corresponding hypothesis at a neighboring pixel $\boldsymbol{q}$ is defined by

$$\boldsymbol{D}_r^{(t)}(\boldsymbol{q}) = \gamma\, \boldsymbol{D}_r^{(k)}(\boldsymbol{p}), \tag{15}$$

where t denotes a continuous hypothesis index at $\boldsymbol{q}$. Taking reciprocals and substituting Eq. 14 yields

$$\begin{aligned} \frac{fb}{\boldsymbol{D}_r(\boldsymbol{q})} + (t - k_0) &= \frac{1}{\gamma}\left(\frac{fb}{\boldsymbol{D}_r(\boldsymbol{p})} + (k - k_0)\right) \\ \Longrightarrow \quad t(k) &= k_0 + \frac{k - k_0}{\gamma} + \underbrace{\frac{1}{\gamma}\frac{fb}{\boldsymbol{D}_r(\boldsymbol{p})} - \frac{fb}{\boldsymbol{D}_r(\boldsymbol{q})}}_{\delta}, \end{aligned} \tag{16}$$

which gives Eq. 7. Since $t(k)$ is affine in k, the cost curve at $\boldsymbol{q}$ can be efficiently resampled at the aligned indices using linear interpolation. Hypotheses mapped outside the valid range, i.e., $t(k) \notin [0, K-1]$, are excluded, and the aggregation in Eq. 8 is normalized over the remaining valid samples. This alignment introduces no additional hyperparameter beyond the aggregation window size and naturally reduces to standard index-wise aggregation when both the prior and rendered depths are locally constant.

Reliability assessment. We assess matching reliability following the cost-profile criterion of ADP-MVS (Wang et al., 2023), applied to the aggregated cost profile $\tilde{\boldsymbol{c}}(\boldsymbol{p})$ (Algorithm 1). Omitting the pixel argument for brevity, let $k^* = \arg\min_k \tilde{c}_k$ denote the global minimum, and define the set of local minima as

$$\mathcal{M} = \{1 \le k \le K-2 : \tilde{c}_k \le \tilde{c}_{k-1},\, \tilde{c}_k \le \tilde{c}_{k+1}\}. \tag{17}$$

The distinctiveness of the global minimum relative to the other local minima is measured by

$$\sigma = \sqrt{\frac{1}{|\mathcal{M}| - 1} \sum_{m \in \mathcal{M}} \left(\tilde{c}_m - \tilde{c}_{k^*}\right)^2}. \tag{18}$$

In addition to a low and distinctive minimum, we require the photometric optimum to remain close to the current rendered depth, $|k^* - k_0| \le \eta$, since a large displacement from k_0 is more likely to indicate an unreliable correspondence.

Algorithm 1 Reliability assessment of pixel $\boldsymbol{p}$

Require: aggregated profile $\tilde{\boldsymbol{c}}(\boldsymbol{p})$
$k^* \leftarrow \arg\min_k \tilde{c}_k$
$\mathcal{M} \leftarrow$ local minima of $\tilde{\boldsymbol{c}}$
if $|k^* - k_0| > \eta$ **or** $\tilde{c}_{k^*} > \tau_1$ **then**
  return unreliable
end if
if $|\mathcal{M}| \leq 1$ **then**
  return reliable **if** $\tilde{c}_{k^*} < \tau_2$ **else** unreliable
end if
return reliable **if** $\sigma > \tau_3$ **else** unreliable

Table 8: **Parameters of point evidence construction**.

Parameter	Value
Hypotheses K	61
Patch size	5×5
Source views (DTU/TnT)	2 / 4
Aggregation window $\mathcal{W}$	5×5
τ_1 / τ_2 / τ_3	0.50/0.15/0.20
Max. index offset η	4

A.3 More Experiments

A.3.1 Comparison with Spatially Varying Alignment

We further examine whether the anchors can improve the prior without explicit integration by using them to construct a spatially varying correction. Specifically, the log-depth residuals at the anchors, $e(\boldsymbol{p}) = \log \hat{\boldsymbol{D}}_r(\boldsymbol{p}) - \log \hat{\boldsymbol{D}}_{pr}(\boldsymbol{p})$, are propagated into a dense correction field via Nadaraya–Watson kernel regression:

$$c(\boldsymbol{x}) = \frac{\sum_{\boldsymbol{p}\in\mathcal{A}} G_\sigma(\boldsymbol{x}-\boldsymbol{p})\, e(\boldsymbol{p})}{\sum_{\boldsymbol{p}\in\mathcal{A}} G_\sigma(\boldsymbol{x}-\boldsymbol{p})}. \tag{19}$$

We compute this field as a normalized convolution and use the locally corrected depth $\hat{\boldsymbol{D}}_{pr}\exp(c)$ in place of $\boldsymbol{D}_{int}$ for supervision.

As shown in Tab. 9, spatially varying correction consistently improves over global alignment across different kernel widths, but remains inferior to the proposed integration. The distance-wise analysis in Tab. 10 further reveals how this difference arises from the propagation of anchor information. While both approaches achieve comparable accuracy near the anchors and gradually approach the prior as the distance to the nearest anchor increases, the benefit of kernel-based correction diminishes more rapidly at intermediate distances.

Table 9: **Local correction with the same anchors on DTU.**

Supervision	CD ↓
Global alignment	0.907
Kernel interp.($\sigma = 2$)	0.775
Kernel interp.($\sigma = 4$)	0.781
Kernel interp.($\sigma = 8$)	0.786
Integrated depth (ours)	**0.707**

Table 10: **Depth accuracy versus distance to the nearest anchor.** Percentage of pixels with depth error below 2 mm.

Dis.(px)	KI.($\sigma = 2$)	Integrated
0	96.54%	97.15%
1–2	85.60%	91.38%
3–5	68.53%	83.88%
6–10	66.32%	77.74%
>10	69.63%	69.26%

A.3.2 Generalization to Other Geometric Priors

To examine whether DGS depends on a specific prior, we replace VGGT with three other feed-forward geometry models, DA3 (Lin et al., 2025), MoGe3 (Kong et al., 2026), and π^3 (Wang et al., 2026), keeping all other settings unchanged. Tab. 11 compares DGS with anchor-based global alignment of each prior. DGS consistently improves over global alignment regardless of the prior.

Table 11: **DGS with different geometric priors on DTU.**

Prior	VGGT	DA3	MoGe3	π^3
Global alignment	0.907	0.965	1.150	0.947
DGS (ours)	**0.707**	**0.766**	**0.820**	**0.757**

A.3.3 Qualitative Results on BlendedMVS Dataset

We further evaluate DGS on BlendedMVS (Yao et al., 2020) to assess its generalization to more diverse objects and scenes. For each scene, we select three input views with large viewpoint changes, resulting in a challenging sparse-view setting. Since there is no established protocol for quantitative mesh evaluation on BlendedMVS under sparse views, we provide qualitative comparisons. As shown in Fig. 8, DGS produces more complete surfaces with fewer geometric artifacts. The improvements are particularly evident in the highlighted regions, including the top of the bread, the holes on the front panel of the clock, and the thin structures of the robot, where fine geometric structures are better preserved.

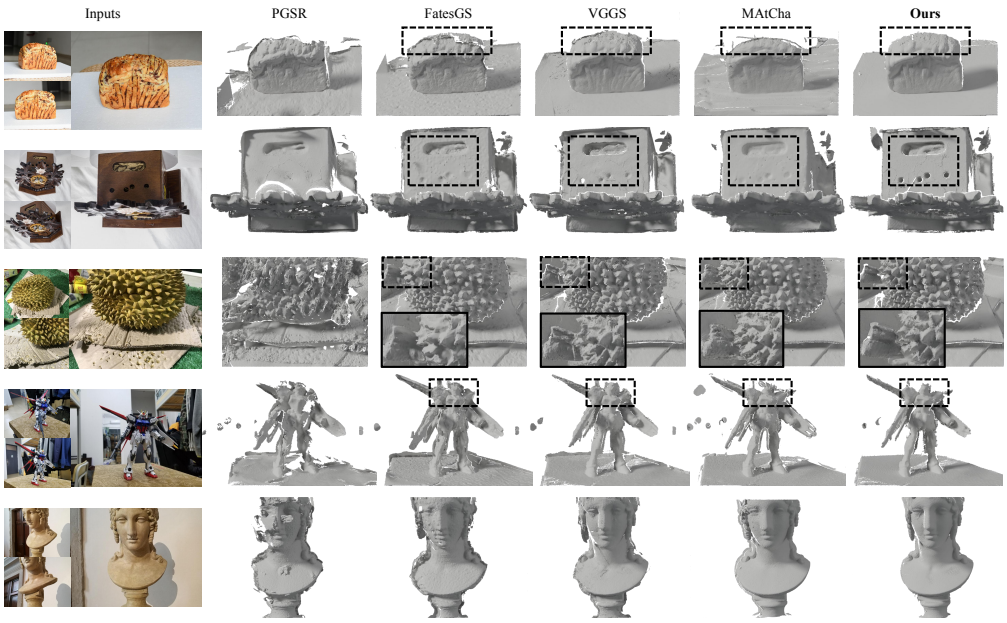

Figure 8: **Qualitative comparison on BlendedMVS with three input views.**

A.3.4 Evolution over Depth Updates

Anchors are re-estimated at each depth update by searching around the current rendered depth. As the reconstruction improves, the rendered depth provides a better initialization for photometric matching, allowing reliable anchors to be established in regions that were not supported in earlier updates. Fig. 9 shows this progressive process. In the dashed region, for example, the roof initially contains few reliable anchors; as the geometry is refined, additional anchors are established and provide new constraints for subsequent depth integration.

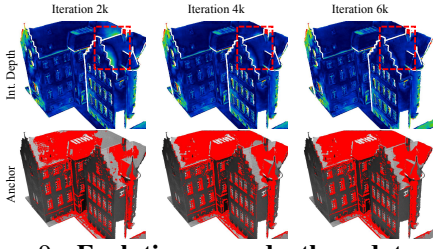

Figure 9: **Evolution over depth updates.** Error of the integrated depth (top) and anchor pixels (bottom) at different iterations.