ANCHORGEO: CONFIDENCE-GATED DEPTH ALIGNMENT FOR WEAKLY OBSERVED REGIONS IN GAUSSIAN SURFACE RECONSTRUCTION

Anonymous authors
Paper under double-blind review

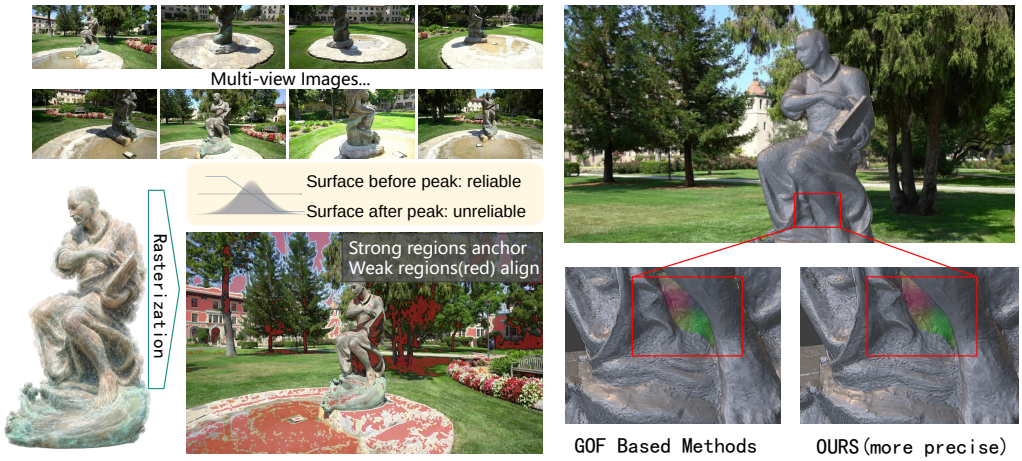

Figure 1: Weakly-observed regions drift under purely photometric supervision. From multi-view images, our transmittance-based gate identifies whether the $T{=}0.5$ surface lies before the Gaussian peak (reliable) or after it (unreliable, red); reliable regions then anchor the weak ones through a per-segment shape model. In GOF-based methods, insufficient observation leaves low-frequency, blurry Gaussians, so the mesh in these weak regions loses precision or even vanishes, whereas our reconstruction recovers the surface.

ABSTRACT

Reconstructing accurate surfaces from multi-view images with radiance fields has achieved remarkable progress in recent years. However, prevailing approaches, primarily based on Gaussian Splatting, remain unreliable in weakly observed regions, where the photometric loss is nearly flat and geometry drifts freely: self-consistency regularizers cannot resolve such regions, while external priors applied indiscriminately fight the photometric evidence exactly where it is decisive. In this paper, we introduce AnchorGeo, a confidence-gated depth-alignment framework that commits geometry precisely where observations run out. Our key finding is that the reliability of the rendered depth is detectable per pixel: a crossing in front of the transmittance peak extracts a view-independent surface, whereas one behind the peak is view-dependent by construction. To locate trustworthy geometry, we first propose a Transmittance-based Confidence Gate that separates reliable anchors from weak pixels directly at rendering time. Subsequently, Confidence-Anchored Depth Alignment is designed to align a monocular depth prior to the anchored geometry in closed form on each object segment and to pull only the weak pixels toward the aligned target, routing external cues into exactly the regions the images cannot constrain. Extensive experiments on DTU and Tanks-and-Temples demonstrate state-of-the-art surface quality with the largest gains concentrated in weakly observed regions, and a per-point recovery/damage decomposition verifies that these gains come from recovering weak regions rather than damaging well-observed ones.

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.

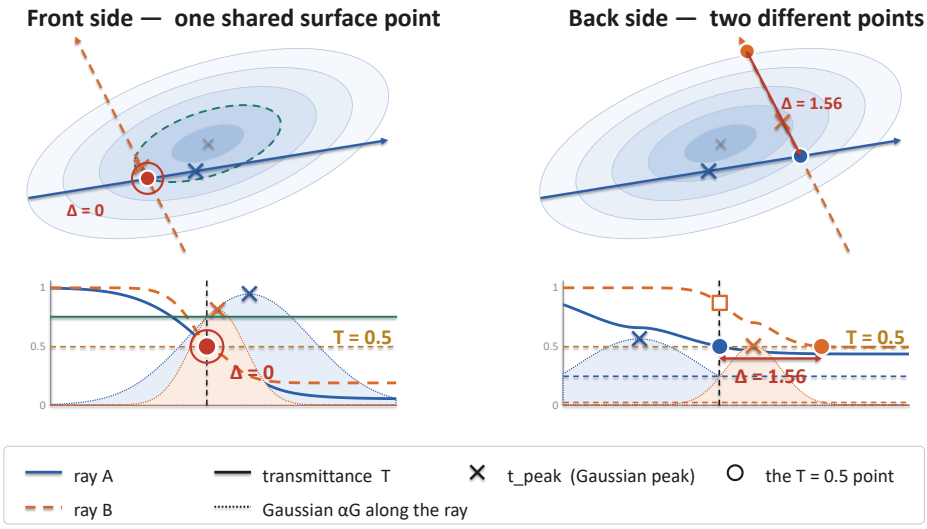

Figure 2: Co-point experiment on a single Gaussian: we take the $T{=}0.5$ point of ray A and send a second ray B through the same 3D point. *Left (front)*: when the crossing lies in front of the transmittance peak ($\times$), $T{=}0.5$ is equivalent to $\alpha\tilde{G}(\mathbf{x}) = \frac{3}{4}$—a condition on the *point* alone (dashed iso-opacity shell). Ray B then also reads $T{=}0.5$ exactly there ($\Delta{=}0$): both views share one surface point. *Right (back)*: behind the peak the crossing condition becomes $\alpha\tilde{G}(\mathbf{x}) = 1 - 4(1 - O_p)^2$, which carries each ray's own peak opacity O_p. Ray B passes through ray A's surface point without crossing 0.5 there, and reaches $T{=}0.5$ only a distance Δ farther: each view writes its own surface. Bottom: transmittance T (solid/dashed) and the Gaussian contribution $\alpha\tilde{G}$ (dotted) along each ray.

INTRODUCTION

NeRF (Mildenhall et al., 2020) and 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) achieve high-quality novel view synthesis, yet accurate surface extraction remains challenging: volume rendering favors soft, semi-transparent representations, while surfaces require well-defined boundaries.

Implicit methods optimize signed-distance or occupancy fields (Wang et al., 2021; Li et al., 2023), often requiring hours per scene. Explicit methods extract surfaces through flattened splats (Huang et al., 2024), unbiased depth rasterization (Zhang et al., 2024; Chen et al., 2025), or transmittance level sets (Yu et al., 2024c; Zhang et al., 2026), matching implicit quality in well-captured regions at a fraction of the cost.

The remaining challenge is *where* the read-out can be trusted. With few views or little texture, geometry can change with little photometric penalty—a regime we call the *photometric null space*. Depth–normal, distortion, and multi-view consistency regularizers constrain quantities rendered by the model, but cannot uniquely resolve such regions. Monocular priors (Li et al., 2025a) supply external geometry while introducing their own estimation biases.

We introduce **AnchorGeo**: a transmittance-based confidence gate identifies unreliable read-outs, and a monocular depth prior is aligned in closed form to trusted geometry within each object segment (Kirillov et al., 2023). Supervision pulls weak pixels toward the aligned target, adding about 10 ms per iteration. AnchorGeo achieves mean F1 0.62 on Tanks-and-Temples and mean Chamfer 0.45 mm on DTU, with the largest gains in weakly observed scenes. Our contributions are as follows.

- Geometric error in weakly observed regions lives in the photometric null space: the field's representational advances remove *global* biases in how a surface is read out, but local correctness in weak regions is information-limited, not representation-limited.
- The view-independence of the transmittance level-set read-out admits a closed-form, per-pixel criterion—the crossing must precede the transmittance peak, equivalently the composited peak opacity must exceed three quarters—which certifies whose surface can be

trusted across the opacity-field family, and directly implies that multi-view evidence must be conditioned on it.

- We introduce an object-aware reconstruction framework that couples local geometric reliability with texture cues to guide targeted refinement. Segment-wise prior calibration and confidence-guided supervision focus geometric optimization on poorly constrained regions while preserving well-observed structures.

Related Work

Neural and Gaussian Surface Reconstruction

Neural implicit methods reconstruct surfaces through unbiased SDF rendering (NeuS (Wang et al., 2021)), SDF-to-density mappings (VolSDF (Yariv et al., 2021)), and scalable grid representations (Li et al., 2023; Müller et al., 2022; Fridovich-Keil et al., 2022). BakedSDF extracts meshes (Yariv et al., 2023), while MonoSDF adds monocular cues (Yu et al., 2022). These methods provide strong surface priors but require costly field optimization.

3DGS enables real-time rendering with explicit Gaussians (Kerbl et al., 2023). Surface-oriented variants align or flatten primitives (Guédon & Lepetit, 2024a;b; Huang et al., 2024; Dai et al., 2024), improve depth read-outs (Chen et al., 2025; Zhang et al., 2024), or extract opacity level sets (Yu et al., 2024c). MILo integrates a differentiable mesh (Guédon et al., 2025); hybrid methods couple Gaussians to distance fields (Chen et al., 2023; Yu et al., 2024a; Li et al., 2025b). Related work studies volumetric formulations and filtering (Miller et al., 2023; Jiang et al., 2025; Govindarajan et al., 2025; Yu et al., 2024b). Our focus is the remaining lack of geometric constraints in weakly observed regions.

Geometric Regularization for Gaussian Splatting

Gaussian-surface regularizers enforce depth–normal consistency (Huang et al., 2024; Yu et al., 2024c; Dai et al., 2024; Ren et al., 2025), reduce depth distortion (Barron et al., 2022), impose multi-view consistency (Chen et al., 2025), or sharpen surfaces through flatness and opacity penalties (Guédon & Lepetit, 2024a; Huang et al., 2024). Such self-consistency permits ambiguous geometry under weak observations. Observability-aware methods acquire or weight data (Jiang et al., 2024; Pan et al., 2022; Ran et al., 2023; Lee et al., 2022; Goli et al., 2024; Hu et al., 2024a); we use representation confidence to guide geometric supervision.

Monocular Geometric Priors

MonoSDF (Yu et al., 2022), DN-Splatter, and GausSurf (Turkulainen et al., 2025; Wang et al., 2024) use monocular depth or normals to constrain reconstruction. Gaussian methods draw on Omnidata and DSINE (Eftekhar et al., 2021; Dai et al., 2024; Chen et al., 2024; Bae & Davison, 2024), stereo distillation (Wolf et al., 2024), and DepthAnythingV2 (Li et al., 2024; 2025a; Yang et al., 2024b). Monocular foundation models supply these cues (Yin et al., 2023; Hu et al., 2024b; Yang et al., 2024a), but scale ambiguity and prediction biases can propagate into reconstruction. AnchorGeo calibrates the prior and routes supervision toward weak regions while preserving image-supported geometry.

Preliminary and Motivation

We focus on opacity-field Gaussian splatting, where the scene surface is defined as the level set at which transmittance falls to 0.5 (Yu et al., 2024c; Zhang et al., 2026). Our goal is to characterize *when* this read-out defines a multi-view-consistent surface, and to regularize the field exactly where it does not.

Transmittance depth in Gaussian opacity fields. Along a ray, the transmittance contributed by a single Gaussian is piecewise: $T(t) = v(t)$ before the peak and $T(t) = v(t_{peak})^2/v(t)$ after it, with $v = \sqrt{1 - \alpha\tilde{G}(\mathbf{x})}$, where α is the opacity and $\tilde{G}$ the normalized Gaussian evaluated at the 3D point

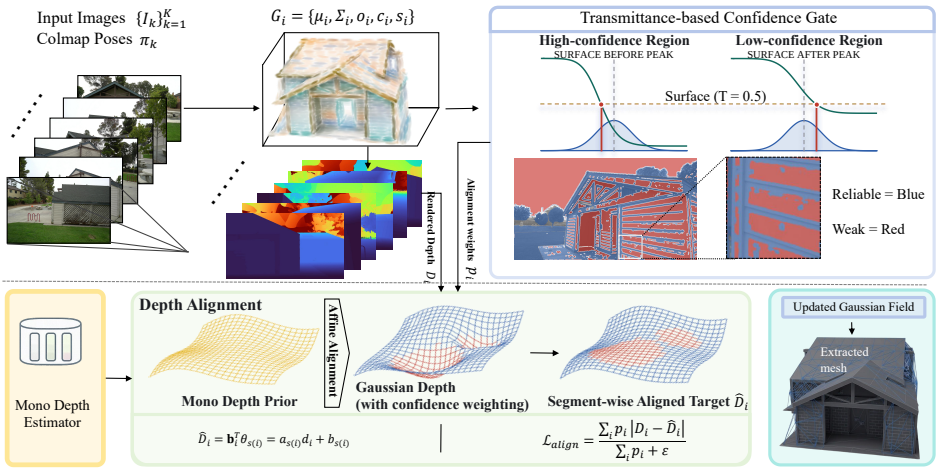

Figure 3: Overview of our pipeline. From posed input images we optimize Gaussians and render per-view transmittance depth D_i. A transmittance-based confidence gate distinguishes front-side (reliable, blue) and back-side (unreliable, red) read-outs using the rendered front-bias, contributing to alignment weights p_i. The illustrated $T{=}0.5$ crossings precede and follow the Gaussian peak, respectively. A monocular depth prior is affine-aligned per segment to the confidence-weighted Gaussian depth, producing the segment-wise aligned target $\hat{D}_i$; the alignment loss $\mathcal{L}_{\text{align}}$ pulls the weak pixels toward it, and the mesh is extracted from the updated Gaussian field.

(Yu et al., 2024c). The rendered depth D is read at the first $T{=}0.5$ crossing (the median depth), whose iso-surface yields the mesh.

The co-point experiment. Figure 2 isolates, on a single Gaussian, why this read-out is trustworthy on one side of the peak and not on the other. In front of the peak, the surface condition $T{=}0.5$ reduces to $\alpha\tilde{G}(\mathbf{x}) = \frac{3}{4}$: it constrains the 3D point alone, so every view that reaches the point before its own peak extracts the *same* iso-opacity shell—the surface is view-independent. Behind the peak, the condition becomes $\alpha\tilde{G}(\mathbf{x}) = 1 - 4(1 - O_p)^2$ and inherits the per-ray peak opacity $O_p = \alpha\tilde{G}(\mathbf{x}_{peak})$, which varies with viewing direction: every view gets its own surface. A Gaussian that keeps its crossing in front ($O_p \geq \frac{3}{4}$) therefore writes one consistent surface, whereas one that pushes it behind ($\frac{1}{2} \leq O_p < \frac{3}{4}$) writes a different surface per view. The regime is decided by a quantity we can render—whether the crossing precedes the peak—so untrustworthy geometry is detectable per pixel. We call this test the *front-transmittance criterion*.

Core result: a shared dominant Gaussian with residual contributions. Using the same incoming normalization, $T_r(-\infty) = 1$, fix a candidate dominant Gaussian g shared across the views under comparison, with spatial field $q_g(\mathbf{x}) = \alpha_g \tilde{G}_g(\mathbf{x})$ and ray-dependent peak opacity $O_{p,g}(r)$. Let $T_{j,r}$ be the complete piecewise transmittance of Gaussian j along ray r, including its back branch when applicable, and define

$$P_r(\mathbf{x}) = \prod_{j \neq g} T_{j,r}(\mathbf{x}), \qquad \epsilon_r(\mathbf{x}) = 1 - P_r(\mathbf{x}). \tag{1}$$

Thus ϵ_r is computed from all remaining contributions, not a free constant. Factoring out g gives, with arguments suppressed,

$$T_r = \begin{cases} \sqrt{1 - q_g}\,(1 - \epsilon_r), & \text{before or at } g\text{'s ray peak}, \\ \dfrac{1 - O_{p,g}(r)}{\sqrt{1 - q_g}}\,(1 - \epsilon_r), & \text{after } g\text{'s ray peak}. \end{cases} \tag{2}$$

At an existing $T_r = \frac{1}{2}$ crossing, the two read-out conditions are

$$q_g(\mathbf{x}) = \begin{cases} 1 - \dfrac{1}{4(1-\epsilon_r)^2}, & \text{front,} \\ 1 - 4(1-\epsilon_r)^2(1-O_{p,g}(r))^2, & \text{back.} \end{cases} \tag{3}$$

If the residual opacity at the compared front-side crossings satisfies $0 \le \epsilon_r \le \epsilon < \frac{1}{2}$, then

$$\left| q_g(\mathbf{x}) - \frac{3}{4} \right| = \frac{2\epsilon_r - \epsilon_r^2}{4(1-\epsilon_r)^2} \le B(\epsilon) \equiv \frac{2\epsilon - \epsilon^2}{4(1-\epsilon)^2}. \tag{4}$$

Proof. Substitute $T_r = \frac{1}{2}$ into Eq. (2) to obtain Eq. (3); the front-side deviation is $B(\epsilon_r)$, and $B'(u) = 1/[2(1-u)^3] > 0$ for $u < \frac{1}{2}$. Consequently, as the residual contributions vanish, front-side crossings approach the same view-independent level $q_g = \frac{3}{4}$. Back-side read-outs retain $q_g = 1 - 4(1 - O_{p,g}(r))^2$, and hence a potential dependence on the dominant Gaussian's own ray peak, even when the residual vanishes.

The distinction is the source of view dependence: on the front branch it enters through the residual factors; on the back branch it also enters through $O_{p,g}(r)$. Selecting the largest-weight Gaussian does not ensure that the remaining *total* contribution is small or that the same Gaussian dominates across views. This is a conditional field-value stability bound, not a pointwise accuracy guarantee or a millimeter-scale error bound. The full proof, finite-camera-origin correction, and conditional depth bounds are in Supplementary Section A.

Validation on DTU. Across all fifteen DTU scenes, geometric error increases monotonically when reconstructed vertices are grouped into quartiles by their cross-view back-read fraction. The highest back-read quartile has 2.1–2.8$\times$ the error of the lowest. This trend supports the geometric relevance of the front/back stability analysis in real multi-Gaussian reconstructions.

The photometric null space. Which pixels end up in the untrustworthy regime? Precisely the weakly observed ones. In regions seen from few views or lacking texture, the photometric loss is nearly flat: the geometry can move within a range without changing the rendering error, so nothing forces the transmittance to commit in front of the peak. Gradients there do not point toward the true surface, the underlying Gaussians drift, and recall collapses exactly where observations are scarce.

What recovery requires. Detection alone does not say what the drifting surface should be. Recovering a weak pixel needs three ingredients: *membership* (which surface it belongs to), *confidence* (which pixels can be trusted), and a *propagation law* (how geometry extends from one to the other). Membership is available from Segment Anything (SAM) masks (Kirillov et al., 2023) and confidence from local image texture; the missing ingredient is the propagation law, and supplying it is the subject of the next section. Among prior work, GGGS (Zhang et al., 2026) refines the read-out but leaves weak regions unconstrained, while GeoSVR (Li et al., 2025a) supervises all pixels indiscriminately. We instead *route*: the gate decides where the scene's own geometry is trustworthy, and the prior is applied only where it is not.

METHOD

Method overview. Our method operates as follows (Figure 3). At every training view we render the transmittance depth D_i; a local texture confidence grades which pixels of each SAM segment can serve as anchors, and the transmittance-based confidence gate of the previous section decides where rendered geometry can be used as evidence. A monocular depth prior is then aligned to the anchored geometry in closed form on each segment, and a loss active only on the weak pixels pulls them toward the aligned target; fitting and pulling alternate every iteration, propagating trusted geometry into the photometric null space.

TEXTURE CONFIDENCE AND ANCHORS

The trustworthy pixels are the richly textured ones, where the photometric loss genuinely constrains geometry. We measure texture by the local image-gradient magnitude g_i, average-pooled over a 7$\times$7

window and normalized as $c_i = \mathrm{clip}(g_i/g_{\mathrm{ref}}, 0, 1)$, and call pixels with $c_i \geq 0.5$ *anchors*. Its role is *information availability*—who can anchor, who needs help—not whether the rendered geometry is correct.

Per-Segment Affine Alignment of the Prior

A monocular estimator provides a relative depth d_i whose scale and shift are unknown and vary across the image. Within one SAM segment s, we align it to the scene's own trusted geometry by an affine map $\hat{D}_i = a_s\, d_i + b_s$ fitted to the anchors in closed form,

$$(a_s, b_s) \;=\; \arg\min_{a,b} \sum_{i \in s} w_i \big(a\, d_i + b - \mathrm{sg}[D_i]\big)^2, \tag{5}$$

a 2×2 system per segment, where the weights w_i favor high-confidence anchors and $\mathrm{sg}[\cdot]$ denotes stop-gradient. The affine freedom absorbs the prior's per-segment scale/shift ambiguity, while the anchors pin the aligned surface $\hat{D}_i$ to geometry the images themselves certify; because the target is detached, the loss below cannot be reduced by deforming the fit toward the weak pixels.

Transmittance-Based Confidence Gate

The co-point analysis says the read-out is view-independent only when the $T{=}0.5$ crossing precedes the transmittance peak. With many Gaussians along a ray, $T = \prod_i v_i$ remains a purely local function of the crossing point as long as the crossing precedes *every* contributing peak. We test this with a rendered front-bias

$$b_i \;=\; \frac{D_i - \bar{z}_i}{D_i}, \tag{6}$$

where $\bar{z}_i$ is the alpha-weighted mean depth of the contributing Gaussian centers (splatted with the same compositing weights and normalized by the accumulated alpha): $b_i \leq 0$ means the surface is read in front of the contributing mass, i.e., in the trustworthy regime. For a single Gaussian this reduces exactly to the front-of-peak condition; for composited rays it aggregates every contributing peak into one test, and the exact composited criterion ($O_p \geq \frac{3}{4}$) behaves consistently. The gate's role is *representation trust*: it decides where rendered geometry may be used as *evidence*, not where the pull acts. Multi-view reprojection errors measured at back-side pixels are artifacts of the representation rather than information about geometry, so their evidence weight is reset to neutral where $b_i > 0$—the pull itself is never switched off by the gate.

Pulling Only the Weak Pixels

The alignment loss acts only where the images cannot:

$$\mathcal{L}_{\mathrm{align}} \;=\; \frac{\sum_i p_i \left|D_i - \hat{D}_i\right|}{\sum_i p_i + \varepsilon}, \tag{7}$$

with the purely multiplicative selector $p_i = (1 - c_i)\, \mathbf{1}\big[|s| > 800\big]\, w_i^v\, G_s$, whose factors play four distinct roles: *need* $(1{-}c_i)$, *error evidence* (w_i^v), *representation trust* (the gate: whether that evidence counts), and *target reliability* (G_s). Here $w_i^v = \mathrm{clip}\big((e_i - e_0)/(e_1 - e_0), 0, 1\big)$ turns the multi-view reprojection error e_i into evidence ($e_0{=}0.5$, $e_1{=}2$ px), reset to 1 for pixels invisible to the neighboring view or gated as back-side by Eq. (6); and $G_s = \mathbf{1}[n_a \geq 50] \cdot \mathrm{clip}\big(2 - r_s/(\rho \bar{D}_s), 0, 1\big)$ softly discounts segments whose affine fit is poor (n_a anchors, residual r_s, $\rho{=}0.03$). The full training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{rgb}} + \lambda_{dn}\mathcal{L}_{dn} + \lambda_{ncc}\mathcal{L}_{ncc} + \lambda_{geo}\mathcal{L}_{geo} + \lambda_n\mathcal{L}_n + \lambda_a\mathcal{L}_{\mathrm{align}}, \tag{8}$$

where the first four terms follow the base pipeline (Zhang et al., 2026), $\mathcal{L}_n$ supervises rendered normals with a monocular normal prior ($\lambda_n{=}0.05$), and $\lambda_a{=}0.10$.

Design Choices and Cost

The regularizer runs entirely in image space—SAM masks, rendering, and median depth are already aligned there—and the per-segment 2×2 solves add about 10 ms per iteration.

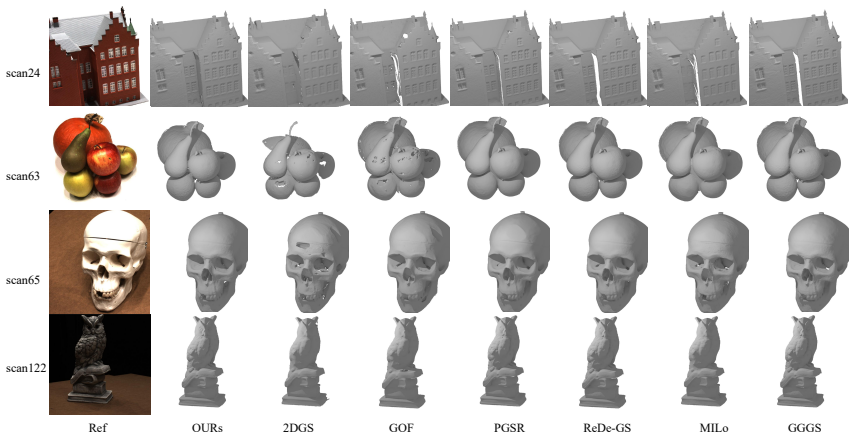

Figure 4: Qualitative comparison on the DTU dataset (Jensen et al., 2014). Rows are scans 24, 63, 65, and 122; columns show the reference image, our reconstruction, and the baselines, including the mesh-in-the-loop method MILo (Guédon et al., 2025). Competing methods leave holes, cracks, and dented patches precisely in weakly observed or low-texture regions (fruit contacts in scan 63, the cranium in scan 65, the facade seam in scan 24), whereas our method produces complete, clean surfaces while preserving detail.

Method	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	Mean	Time
Implicit																	
NeRF (Mildenhall et al., 2020)	1.90	1.60	1.85	0.58	2.28	1.27	1.47	1.67	2.05	1.07	0.88	2.53	1.06	1.15	0.96	1.49	>12h
VolSDF (Yariv et al., 2021)	1.14	1.26	0.81	0.49	1.25	0.70	0.72	1.29	1.18	0.70	0.66	1.08	0.42	0.61	0.55	0.86	>12h
NeuS (Wang et al., 2021)	1.00	1.37	0.93	0.43	1.10	0.65	0.57	1.48	1.09	0.83	0.52	1.20	0.35	0.49	0.54	0.84	>12h
Neuralangelo (Li et al., 2023)	0.37	0.72	0.35	0.35	0.87	0.54	0.53	1.29	0.97	0.73	0.47	0.74	0.32	0.41	0.43	0.61	>12h
Explicit																	
3DGS (Kerbl et al., 2023)	2.14	1.53	2.08	1.68	3.49	2.21	1.43	2.07	2.22	1.75	1.79	2.55	1.53	1.52	1.50	1.96	7.8m
2DGS (Huang et al., 2024)	0.48	0.91	0.39	0.39	1.01	0.83	0.81	1.36	1.27	0.76	0.70	1.40	0.40	0.76	0.52	0.80	11.3m
GOF (Yu et al., 2024c)	0.50	0.82	0.37	0.37	1.12	0.74	0.73	1.18	1.29	0.68	0.77	0.90	0.42	0.66	0.49	0.74	52m
3DGSR (Lyu et al., 2024)	0.44	0.96	0.40	0.36	1.02	0.80	0.64	1.20	1.08	0.97	0.54	0.72	0.37	0.52	0.42	0.70	–
RaDe-GS (Zhang et al., 2024)	0.43	0.75	0.35	0.37	0.81	0.74	0.74	1.19	1.20	0.65	0.61	0.84	0.35	0.66	0.46	0.68	8.2m
PGSR (Chen et al., 2025)	0.34	0.54	0.44	0.37	0.78	0.57	0.49	1.06	0.63	0.59	0.47	0.50	0.30	0.37	0.34	0.52	30.5m
GGGS (Zhang et al., 2026)	0.37	**0.50**	**0.27**	**0.31**	0.81	0.43	0.42	1.05	0.64	0.52	**0.32**	0.58	0.30	**0.31**	0.33	0.48	25.3m
GeoSVR (Li et al., 2025a)	**0.32**	0.51	0.30	0.33	0.71	0.48	0.42	1.03	**0.62**	0.56	0.33	**0.46**	0.30	0.34	**0.32**	0.47	53.3m
Ours	0.35	0.53	0.30	0.33	**0.66**	**0.42**	**0.40**	**0.93**	0.63	**0.51**	0.33	**0.46**	**0.29**	0.34	0.34	**0.45**	29.1m

Table 1: Chamfer distance (mm, ↓) on the DTU dataset (Jensen et al., 2014); columns are DTU scan IDs. Best per column in bold. Implicit methods run for more than twelve hours per scene.

EXPERIMENTS

We evaluate and ablate AnchorGeo on standard benchmarks, and decompose the metrics per ground-truth point into weak-region *recovery* and well-observed *damage*.

EXPERIMENTAL SETUP

Datasets & evaluations. We evaluate on the standard fifteen DTU scenes (Jensen et al., 2014) and six Tanks-and-Temples (T&T) scenes (Knapitsch et al., 2017) listed in Tables 1 and 2. Meshes extracted from the 0.5 occupancy iso-surface are scored against reference scans using the official protocols.

Baselines. Tables 1 and 2 compare neural implicit, Gaussian, and sparse-voxel methods, including our immediate baseline GGGS (Zhang et al., 2026). All baselines use their public implementations and reported settings.

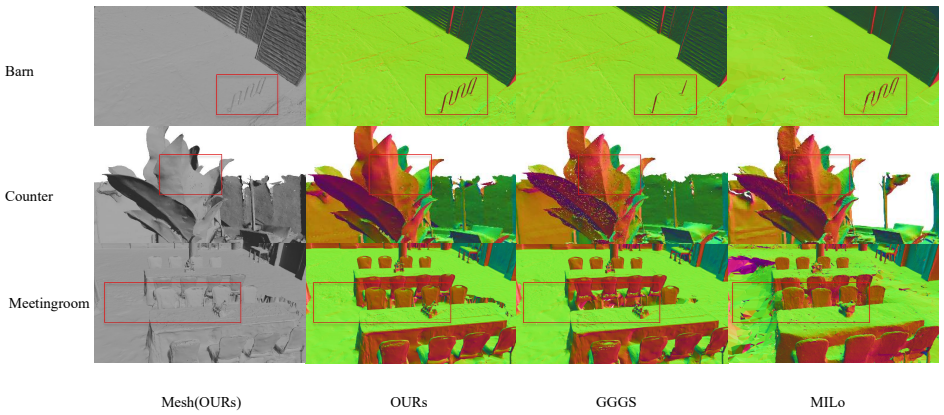

Figure 5: Qualitative comparison on large scenes: *Barn* and *Meetingroom* from Tanks-and-Temples and *Counter* from Mip-NeRF 360 (Barron et al., 2022). From left: our extracted mesh (shaded), then normal renderings of ours, GGGS, and MILo. The boxed close-ups mark weakly observed structures: the thin ladder in *Barn* (truncated by GGGS), the leaf in *Counter* (speckled by GGGS, with MILo corrupting the surrounding geometry), and the table region in *Meetingroom* (left bumpy by MILo). Our method recovers these structures cleanly.

Method	Barn	Cat.	Cour.	Igna.	Meet.	Truc.	Mean	Time
Implicit								
NeuS (Wang et al., 2021)	0.29	0.29	0.17	0.83	0.24	0.45	0.38	>24h
Geo-NeuS (Fu et al., 2022)	0.33	0.26	0.12	0.72	0.20	0.45	0.35	>24h
Neuralangelo (Li et al., 2023)	0.70	0.36	0.28	**0.89**	0.32	0.48	0.50	>24h
Explicit								
2DGS (Huang et al., 2024)	0.36	0.23	0.13	0.44	0.16	0.26	0.30	15.5m
GOF (Yu et al., 2024c)	0.51	0.41	0.28	0.68	0.28	0.59	0.46	71.6m
RaDe-GS (Zhang et al., 2024)	0.49	0.36	0.27	0.72	0.27	0.61	0.45	**12.1m**
PGSR (Chen et al., 2025)	0.66	0.44	0.20	0.81	0.33	0.66	0.52	42.9m
GeoSVR (Li et al., 2025a)	0.68	0.49	0.34	0.83	0.37	0.66	0.56	66.4m
GGGS (Zhang et al., 2026)	0.70	0.56	0.38	0.81	0.42	**0.70**	0.60	32.1m
Ours	**0.72**	**0.57**	**0.40**	0.82	**0.50**	**0.70**	**0.62**	35.3m

Table 2: Quantitative comparison on the Tanks-and-Temples dataset (Knapitsch et al., 2017). We report the per-scene F1-score, the mean F1-score, and the average optimization time. Best F1 per column in bold.

Metrics. We report DTU Chamfer distance (mm, ↓), T&T F1-score (↑), and mean optimization time. F1 combines precision and recall at the official distance thresholds. For recovery/damage analysis, aligned per-reference-point distances identify points *recovered* (missed by the baseline, covered by ours), *damaged* (the reverse), or unchanged; net gain is recovery minus damage (Tatarchenko et al., 2019).

Implementation details. We build on GGGS (Zhang et al., 2026) with COLMAP poses (Schönberger & Frahm, 2016), Metric3D v2 normals (Hu et al., 2024b) (λ_n=0.05), and Depth Anything 3 depth (Lin et al., 2025) (λ_a=0.10). Both datasets share one configuration, detailed in Supplementary Section B.

Main Results

Geometry on Tanks-and-Temples. AnchorGeo achieves the best mean F1 (0.62), ahead of GGGS (0.60) and Neuralangelo (0.50), and ranks first or tied-first on five of six scenes (Table 2). It matches or improves GGGS on every scene, with gains from +0.01 on *Ignatius* to +0.08 on

Meetingroom. Neuralangelo remains stronger on *Ignatius* (0.89 vs. 0.82). Optimization takes 35.3m versus GGGS's 32.1m, below GeoSVR (66.4m) and GOF (71.6m). Supplementary Section C analyzes *Courthouse*.

Geometry on DTU. AnchorGeo leads DTU with 0.45 mm mean Chamfer, ahead of GeoSVR (0.47), GGGS (0.48), and PGSR (0.52), ranking first or tied-first on seven of fifteen scans (Table 1), with gains on weakly observed scans 63 and 83. The weak-to-well-observed error ratio drops from 2.1–2.8× to 1.52× on average. Comparing anchor and weak regions shows most gains in the latter.

Qualitative comparisons. Figures 4 and 5 show improved completeness and detail in weak regions, including the ladder in *Barn* and the leaf in *Counter*. Supplementary Section C provides results on all fifteen DTU scans and all six T&T scenes.

Ablation: The Gate and the Prior Multiply

Table 3 compares ungated and gated normal and DA3 depth priors across multiple T&T scenes.

Gating makes the priors effective. The baseline and gate-only configuration both achieve 0.60 mean F1. Adding the normal prior with ungated supervision also yields 0.60, while gating the normal prior raises the score to 0.61. The depth-prior comparison shows an even larger benefit: directly applying the ungated DA3 depth prior reduces mean F1 from 0.60 to 0.58, whereas its gated counterpart reaches 0.61, improving by 0.03 over ungated supervision and by 0.01 over the baseline. These results demonstrate that indiscriminate depth-prior supervision can degrade reconstruction quality, while confidence gating turns the same prior into a source of geometric improvement.

Normal and depth cues complement each other. The full model combines the normal prior with gated depth alignment and achieves the best mean F1 of 0.62: +0.02 over the baseline and +0.01 over either gated prior alone. The multi-scene results support the central design of AnchorGeo: confidence-guided integration makes external geometric cues effective, and combining normal and aligned depth supervision further improves surface reconstruction.

Configuration	Mean F1	Δ vs. baseline
GGGS baseline (no prior, no gate)	0.60	–
Gate only (no external prior)	0.60	0.00
Normal prior only (ungated)	0.60	0.00
Normal prior only (gated)	0.61	+0.01
DA3 depth prior (ungated)	0.58	−0.02
DA3 depth prior (gated)	0.61	+0.01
Full (normal + gated depth align.)	**0.62**	**+0.02**

Table 3: Multi-scene ablation on Tanks-and-Temples: mean F1 (↑) and gain over GGGS. Gating improves both priors; their combination achieves the best result.

Conclusion

We presented AnchorGeo, a confidence-gated depth-alignment framework for Gaussian surface reconstruction. Our starting point is a diagnosis: the transmittance read-out of opacity-field methods is trustworthy only where the crossing happens in front of the transmittance peak; behind it, every view writes its own surface, and the distinction is detectable per pixel. AnchorGeo turns this diagnosis into a mechanism: a confidence gate separates reliable anchors from weak pixels and decides where rendered geometry may serve as evidence, and a closed-form, segment-wise alignment calibrates a monocular depth prior against the anchors and pulls only the weak pixels toward it. The prior is thus routed into exactly the photometric null space while the images keep their authority elsewhere. This routing is what makes the prior valuable: applied indiscriminately, the same prior hurts the baseline; gated, it sets a new state of the art on both benchmarks, with the largest gains where observations are scarce. Limitations include the inherited segmentation granularity and the co-evolution of gate and geometry.

REFERENCES

Gwangbin Bae and Andrew J. Davison. Rethinking inductive biases for surface normal estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5470–5479, 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, 2025.

Hanlin Chen, Chen Li, and Gim Hee Lee. Neusg: Neural implicit surface reconstruction with 3d gaussian splatting guidance. *arXiv preprint arXiv:2312.00846*, 2023.

Hanlin Chen, Fangyin Wei, Chen Li, Tianxin Huang, Yunsong Wang, and Gim Hee Lee. Vcr-gaus: View consistent depth-normal regularizer for gaussian surface reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, and Weiwei Xu. High-quality surface reconstruction using gaussian surfels. In *ACM SIGGRAPH 2024 Conference Papers*, 2024. doi: 10.1145/3641519.3657441.

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 *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

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

Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes' rays: Uncertainty quantification for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Shrisudhan Govindarajan, Daniel Rebain, Kwang Moo Yi, and Andrea Tagliasacchi. Radiant foam: Real-time differentiable ray tracing. *arXiv preprint arXiv:2502.01157*, 2025.

Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024a.

Antoine Guédon and Vincent Lepetit. Gaussian frosting: Editable complex radiance fields with real-time rendering. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024b.

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 (Proc. SIGGRAPH Asia)*, 44(6), 2025. doi: 10.1145/3763339.

Jiarui Hu, Xianhao Chen, Boyin Feng, Guanglin Li, Liangjing Yang, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui. Cg-slam: Efficient dense rgb-d slam in a consistent uncertainty-aware 3d gaussian field. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024a.

Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3d v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 46(12):10579–10596, 2024b.

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*, 2024. doi: 10.1145/3641519.3657428.

Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 406–413, 2014.

Kaiwen Jiang, Venkataram Sivaram, Cheng Peng, and Ravi Ramamoorthi. Geometry field splatting with gaussian surfels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherrf: Active view selection and uncertainty quantification for radiance fields using fisher information. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. doi: 10.1145/3592433. URL https://arxiv.org/abs/2308.04079.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4):1–13, 2017. doi: 10.1145/3072959.3073599.

Soomin Lee, Le Chen, Jiahao Wang, Alexander Liniger, Suryansh Kumar, and Fisher Yu. Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields. *IEEE Robotics and Automation Letters*, 7(4):12070–12077, 2022.

Jiahe Li, Jiawei Zhang, Youmin Zhang, Xiao Bai, Jin Zheng, Xiaohan Yu, and Lin Gu. Geosvr: Taming sparse voxels for geometrically accurate surface reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025a.

Kunyi Li, Michael Niemeyer, Zeyu Chen, Nassir Navab, and Federico Tombari. Monogsdf: Exploring monocular geometric cues for gaussian splatting-guided implicit surface reconstruction. *arXiv preprint arXiv:2411.16898*, 2024.

Shujuan Li, Yu-Shen Liu, and Zhizhong Han. Gaussianudf: Inferring unsigned distance functions through 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025b.

Zhaoshuo Li, Thomas Mueller, Alex Evans, Russell H. Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. URL https://arxiv.org/abs/2306.03092.

Haotong Lin, Sili Chen, Jun Hao 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.

Xiaoyang Lyu, Yang-Tian Sun, Yi-Hua Huang, Xiuzhe Wu, Ziyi Yang, Yilun Chen, Jiangmiao Pang, and Xiaojuan Qi. 3dgsr: Implicit surface reconstruction with 3d gaussian splatting. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 43(6), 2024. doi: 10.1145/3687952.

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. In *European Conference on Computer Vision*, 2020. URL https://arxiv.org/abs/2003.08934.

Bailey Miller, Hanyu Chen, Alice Lai, and Ioannis Gkioulekas. Objects as volumes: A stochastic geometry view of opaque solids, 2023.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4), 2022.

Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. Activenerf: Learning where to see with uncertainty estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022.

Yunlong Ran, Jing Zeng, Shibo He, Lincheng Li, Yingfeng Chen, Gimhee Lee, Jiming Chen, and Qi Ye. Neurar: Neural uncertainty for autonomous 3d reconstruction with implicit neural representations. *IEEE Robotics and Automation Letters*, 8(2):1125–1132, 2023.

Haofan Ren, Qingsong Yan, Ming Lu, Rongfeng Lu, and Zunjie Zhu. 2dgs-r: Revisiting the normal consistency regularization in 2d gaussian splatting. *arXiv preprint*, 2025.

Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Maxim Tatarchenko, Stephan R. Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3405–3414, 2019.

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 *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2421–2431, 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*, 2024.

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. In *Advances in Neural Information Processing Systems*, 2021. URL https://arxiv.org/abs/2106.10689.

Yaniv Wolf, Amit Bracha, and Ron Kimmel. Gs2mesh: Surface reconstruction from gaussian splatting via novel stereo views. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024.

Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10371–10381, 2024a.

Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024b.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *Advances in Neural Information Processing Systems*, 2021. URL https://lioryariv.github.io/volsdf/.

Lior Yariv, Peter Hedman, Christian Reiser, Dor Verbin, Pratul P. Srinivasan, Richard Szeliski, Jonathan T. Barron, and Ben Mildenhall. Bakedsdf: Meshing neural sdfs for real-time view synthesis. *ACM SIGGRAPH 2023 Conference Proceedings*, 2023.

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 *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Mulin Yu, Tao Lu, Linning Xu, Lihan Jiang, Yuanbo Xiangli, and Bo Dai. Gsdf: 3dgs meets sdf for improved rendering and reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024a.

Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024b.

Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 43(6), 2024c. doi: 10.1145/3687937.

Baowen Zhang, Chuan Fang, Rakesh Shrestha, Yixun Liang, Xiaoxiao Long, and Ping Tan. Rade-gs: Rasterizing depth in gaussian splatting. *arXiv preprint arXiv:2406.01467*, 2024.

Baowen Zhang, Chenxing Jiang, Heng Li, Shaojie Shen, and Ping Tan. Geometry-grounded gaussian splatting. *arXiv preprint arXiv:2601.17835*, 2026.