



MILO: Mesh-In-the-Loop Gaussian Splatting for Detailed and Efficient Surface Reconstruction

Antoine Guédon, Diego Gomez, Nissim Maruani, Bingchen Gong, George Drettakis, Maks Ovsjanikov

► To cite this version:

Antoine Guédon, Diego Gomez, Nissim Maruani, Bingchen Gong, George Drettakis, et al.. MILO: Mesh-In-the-Loop Gaussian Splatting for Detailed and Efficient Surface Reconstruction. ACM Transactions on Graphics, 2025. hal-05327792

HAL Id: hal-05327792

<https://inria.hal.science/hal-05327792v1>

Submitted on 23 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

MILo: Mesh-In-the-Loop Gaussian Splatting for Detailed and Efficient Surface Reconstruction

ANTOINE GUÉDON* and DIEGO GOMEZ*, École Polytechnique, France

NISSIM MARUANI, Inria, Université Côte d’Azur, France

BINGCHEN GONG, École Polytechnique, France

GEORGE DRETTAKIS, Inria, Université Côte d’Azur, France

MAKS OVSJANIKOV, École Polytechnique, France



Fig. 1. **Mesh-in-the-Loop Gaussian Splatting.** Our method introduces a novel differentiable mesh extraction framework that operates during the optimization of 3D Gaussian Splatting representations. At every training iteration, we differentially extract a mesh—including both vertex locations and connectivity—directly from Gaussian parameters. This enables gradient flow from the mesh to Gaussians, allowing us to promote bidirectional consistency between volumetric (Gaussians) and surface (extracted mesh) representations. This approach guides Gaussians toward configurations better suited for surface reconstruction, resulting in higher quality meshes with significantly fewer vertices. In this example, our method reconstructs the entire bicycle scene—including background—with almost 10 times fewer vertices than previous methods while preserving fine geometric details. Our framework can be plugged into any Gaussian splatting representation, increasing performance while generating an order of magnitude fewer mesh vertices. MILo makes the reconstructions more practical for downstream applications like physics simulations and animation.

While recent advances in Gaussian Splatting have enabled fast reconstruction of high-quality 3D scenes from images, extracting accurate surface meshes remains a challenge. Current approaches extract the surface through costly post-processing steps, resulting in the loss of fine geometric details or requiring significant time and leading to very dense meshes with millions of vertices. More fundamentally, the *a posteriori* conversion from a volumetric to a surface representation limits the ability of the final mesh to preserve all geometric structures captured during training. We present MILo, a novel Gaussian Splatting framework that bridges the gap between volumetric and surface representations by differentially extracting a mesh from the 3D Gaussians. We design a fully differentiable procedure that constructs the

mesh—including both vertex locations and connectivity—at every iteration directly from the parameters of the Gaussians, *which are the only quantities optimized during training*.

Our method introduces three key technical contributions: (1) a bidirectional consistency framework ensuring both representations—Gaussians and the extracted mesh—capture the same underlying geometry during training; (2) an adaptive mesh extraction process performed at each training iteration, which uses Gaussians as differentiable pivots for Delaunay triangulation; (3) a novel method for computing signed distance values from the 3D Gaussians that enables precise surface extraction while avoiding geometric erosion.

Our approach can reconstruct complete scenes, including backgrounds, with state-of-the-art quality while requiring an order of magnitude fewer mesh vertices than previous methods.

Due to their light weight and empty interior, our meshes are well suited for downstream applications such as physics simulations and animation.

The code for our approach and an online gallery are available at <https://anttwo.github.io/milo/>.

*Both authors contributed equally to the paper.

Authors’ Contact Information: Antoine Guédon, antoine.guedon@enpc.fr; Diego Gomez, diego.gomez@polytechnique.edu; Nissim Maruani, Inria, Université Côte d’Azur, France, nissim.maruani@inria.fr; Bingchen Gong, École Polytechnique, France, bingchen.gong@polytechnique.edu; George Drettakis, Inria, Université Côte d’Azur, France, George.Drettakis@inria.fr; Maks Ovsjanikov, École Polytechnique, France, maks@lix.polytechnique.fr.

CCS Concepts: • **Computing methodologies** → **Reconstruction; Mesh models**; *Point-based models*; *Shape representations*; Rendering; Image-based rendering.

Additional Key Words and Phrases: Mesh, Gaussian Splatting

1 Introduction

Recent methods [Chen et al. 2023; Guédon and Lepetit 2024b; Huang et al. 2024; Li et al. 2023; Reiser et al. 2024a; Yu et al. 2024; Zhang et al. 2024] for reconstructing surface meshes from images commonly adopt a two-stage pipeline: Optimize a volumetric representation via differentiable rendering—typically using Neural Radiance Fields (NeRFs) or 3D Gaussian Splatting—then extract a surface mesh during *postprocessing* by defining a surface, usually as an isosurface. None of these methods considers the reconstructed mesh *during optimization*.

This strategy offers no guarantee that the final mesh will be consistent with the volumetric representation—fine details *that are present when rendering the representation* can be lost.

This postprocessing step collapses complex volumetric information into a surface, often leading to the loss of fine or semitransparent structures, and can introduce artifacts in the extracted mesh. Moreover, Gaussian Splatting and NeRF-based methods are known to adjust their opacity and view-dependent colors independently of the geometry. This allows them to fit the training images more precisely, but often at the expense of geometric consistency. “*Cheating*” leads to hallucinated structures such as floaters or cavities, which are *particularly hard to resolve during mesh extraction*. Fixing them is non-trivial, as the underlying volumetric representation has already absorbed these inconsistencies in its optimization. Several works have highlighted this issue in challenging settings—such as low-information [Gomez et al. 2025; Guédon et al. 2025; Warburg et al. 2023; Yang et al. 2023] or high specularities [Verbin et al. 2022]—yet a general and robust solution remains elusive.

Crucially, this process provides *no feedback during optimization*, leaving mesh quality undefined until training is complete. This means that there is no guarantee that the optimized Gaussian representation will give rise to a high quality mesh capturing all relevant structures. For instance, fine details such as bicycle spokes can disappear or be inaccurately represented in the final mesh *even if visually they exist in the volumetric representation*.

Moreover, many existing methods [Huang et al. 2024; Wang et al. 2021; Yariv et al. 2021; Zhang et al. 2024a] focus on object-centric reconstructions, evaluating only the quality of the foreground geometry. Similarly, benchmarks like DTU [Jensen et al. 2014] or Tanks and Temples [Knapitsch et al. 2017] offer ground truth solely for foreground objects, discouraging full-scene reconstructions that have backgrounds.

In this work, we introduce a novel pipeline for reconstructing compact, high-fidelity meshes of complete 3D scenes using Gaussian Splatting [Kerbl et al. 2023], addressing all of the aforementioned limitations. Our core idea is to generate a mesh at every training iteration from a set of points entangled with the Gaussians, which we call **Gaussian Pivots**.

This setup enables direct gradient backpropagation from the mesh to the Gaussian parameters, allowing us to impose consistency between the surface mesh, volumetric field, and input data. Thanks to the Pivots, Gaussians serve as an implicit parameterization of an explicit mesh, progressively optimized to yield high-quality geometry, vertex positions, and connectivity.

As a result, our approach **sidesteps fixed-topology constraints**, given that the connectivity is dynamically updated as Gaussians move around the scene. On the other hand, the mesh provides a geometric prior that regularizes Gaussians to address “*cheating*”, mitigating undesirable geometric artifacts. This tight, bidirectional coupling enables effective mesh-based regularization, enhancing both fidelity and compactness. In other words, **both representations help each other**.

To further control the complexity of outputs, we propose novel densification and simplification strategies inspired by [Fang and Wang 2024b].

We achieve state-of-the-art geometric quality while using an order of magnitude fewer vertices than competing methods, making our surfaces far more practical for downstream applications. Inspired in previous works which focus on surface-based view synthesis methods [Reiser et al. 2024b] we use a proxy metric based on mesh rendering to evaluate mesh quality in background regions where no ground-truth geometry exists, training a neural color field over the mesh and comparing rendered images across methods. This decoupling rendering fidelity from mesh resolution and helps us to evaluate reconstruction quality across the entire scene. Our work makes the following contributions.

- We introduce the *first radiance field pipeline in which extracting a surface mesh is an integral part of the optimization*, leveraging expressive 3D Gaussian Splatting to parameterize and jointly refine both representations.
- We propose mesh-based regularization strategies that improve geometric quality, especially for thin structures.
- Our method achieves state-of-the-art results in mesh quality and compactness across multiple complex 3D scenes, improving upon previous approaches in both scalability and visual fidelity.
- We adapt a metric introduced in the context of surface-based view synthesis and use it as an evaluation protocol to assess full-scene geometry, even in the absence of ground-truth 3D models. We use this protocol to demonstrate that our method achieves state-of-the-art results in mesh quality and compactness across multiple complex 3D scenes,

We provide detailed definitions of 3D Gaussian Splatting and Delaunay Triangulation in the supplementary material.

2 Related Work

Novel View Synthesis. Early novel view synthesis (NVS) techniques modeled scenes as continuous volumetric fields, notably with NeRF [Mildenhall et al. 2020] introducing a multilayer perceptron to learn a radiance field. Subsequent works explored alternative representations, including hash-based encodings [Müller et al. 2022], discrete voxel grids [Fridovich-Keil et al. 2022], and low-rank tensor

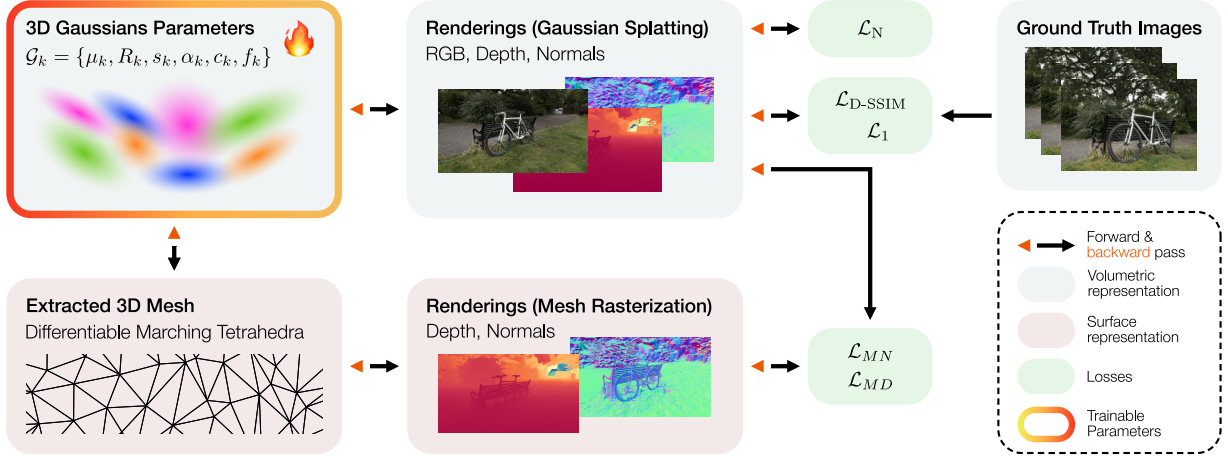


Fig. 2. **MILo Pipeline.** We use Gaussians as pivots for the parametrization of a mesh we can differentially extract at every iteration. Supervision on this mesh allows us to impose a soft prior on the Gaussians which results in better reconstructed surfaces

decompositions [Chen et al. 2022], as well as solutions focused on antialiasing [Barron et al. 2021; Barron et al. 2022, 2023].

Recently, 3D Gaussian Splatting [Kerbl et al. 2023] has redefined the NVS landscape by enabling faster, high-quality synthesis from point-based primitives. This sparked a wave of follow-up work aiming to improve efficiency and scalability, including Mip-Splatting [Yu et al. 2024a], Mini-Splatting v1 and v2 [Fang and Wang 2024a,b], MCMC-Gaussians [Kheradmand et al. 2024], BetaSplatting [Liu et al. 2025], and Taming 3DGS [Mallick et al. 2024]. Additional work like 3DGUT [Wu et al. 2024] and 3DGRT [Moenne-Loccoz et al. 2024] extend these models to arbitrary camera systems and differentiable ray tracing, respectively.

Despite their impressive rendering performance, none of these methods aim at explicit surface mesh reconstruction, limiting their applicability in tasks requiring geometric reasoning or downstream editing.

Surface reconstruction from images. Recovering explicit surface meshes exclusively from images is a longstanding challenge. Well-established volumetric approaches relying on implicit functions [Li et al. 2023; Wang et al. 2021; Yariv et al. 2021] proved efficient for reconstructing accurate surfaces. Training an implicit function is a strong regularization that allows to obtain smooth surfaces, thus these methods excel at extracting the foreground of object-centric scenes.

However, these methods generally require very long training times (often exceeding 24 hours) before mesh extraction can be performed. The success of Gaussian Splatting has naturally inspired an emerging line of work that trains splats jointly with implicit functions, notably Gaussian-UDF [Li et al. 2025], GS-Pull [Zhang et al. 2024b], and GSDF [Yu et al. 2024b]. By combining these two representations, such methods mitigate the time bottleneck by querying the implicit function through strategies that exploit the positions of the Gaussian splats. While this improves training efficiency, it also **exacerbates scalability issues**, since two representations must be optimized simultaneously. Moreover, although neural SDFs excel at

reconstructing isolated objects, their expressivity remains bounded by the MLP architecture [Li et al. 2025; Yu et al. 2024b; Zhang et al. 2024b]. Consequently, while these approaches can in principle be applied to more complex scenes, they tend to underperform beyond single-object or foreground-dominant settings.

Moreover, these volumetric representations typically cannot achieve real-time rendering, limiting their practical applications. Some approaches address this by first optimizing the volumetric representation, extracting a mesh, and then baking the rendering capabilities into the mesh to enable real-time rendering [Reiser et al. 2024a; Yariv et al. 2023]. Thus, these methods are split into two stages, which **prevents any refinement of the mesh**, while still exhibiting lengthy optimization times.

Other methods based on discrete structures [Munkberg et al. 2022; Shen et al. 2023] also proved efficient to some extent, but focus on individual objects and **are not scalable to large complex scenes**.

Recent methods relying on Gaussian Splatting offer promising efficient alternatives [Dai et al. 2024; Wang et al. 2023] namely: 2DGS [Huang et al. 2024], RaDe-GS [Zhang et al. 2024], NeuSG [Chen et al. 2023], VCR-GauS [Chen et al. 2024], and Quadratic Gaussian Splatting [Zhang et al. 2024a] explore different rendering or regularization strategies that rely on Truncated Signed Distance Fields (TSDF), while SuGaR [Guédon and Lepetit 2024b] and Gaussian Opacity Fields [Yu et al. 2024] propose their own mesh extraction procedures.

All previously highlighted methods perform mesh extraction *after* optimization, treating it as a post-processing step. Some of these go slightly further, adopting a two-stage approach where they first optimize a volumetric representation, extract a mesh, and then refine it with differentiable rendering [Guédon and Lepetit 2024a,b; Yariv et al. 2023]. This line of work, however, neither avoids the initial extraction issues **nor adjusts the mesh topology**, which stays fixed during the refinement process. This separation, in both cases, introduces potential inconsistencies: **there is no guarantee that the extracted surface accurately reflects the underlying volumetric representation**. Since Gaussians are continuous by

nature, naive isosurfacing often results in geometric artifacts such as over-inflation or erosion. These artifacts are particularly noticeable around thin structures (see Fig. 1). Moreover, all of these methods typically produce overly dense meshes (up to tens of millions of vertices) that are difficult to scale to large scenes.

In contrast, our method efficiently leverages Gaussians to scale to full-scene reconstruction, and is the **first to integrate mesh extraction directly into the optimization loop**. This design naturally enables the refinement of both vertex positions and mesh topology, ensuring consistency with the underlying volumetric representation (Gaussian splats).

Voronoi & Delaunay-based methods. Our work also exploits the Voronoi diagram and specifically its dual the Delaunay triangulation, which are classical geometric constructions with deep theoretical and practical relevance [Aurenhammer 1991]. These structures have long been applied to surface reconstruction problems [Amenta et al. 1998, 2001; Dey and Goswami 2003], with explicit guarantees under specific sampling conditions. More recently, they have also gained traction in machine learning contexts: in 2D vision [Williams et al. 2020], 3D geometry [Maruani et al. 2023, 2024], or even novel view synthesis [Elsner et al. 2023; Govindarajan et al. 2025]. While GOF [Yu et al. 2024] applies Delaunay triangulation as a post-processing step for mesh extraction, Radiant Foam [Govindarajan et al. 2025] incorporates it into its training pipeline, but focuses solely on view synthesis and not surface reconstruction. In contrast, *our work is the first to leverage Delaunay triangulation in-the-loop with the explicit goal of producing high-quality meshes, enabling direct supervision of surface geometry throughout training.*

3 Overview

Our method aims to address the fundamental challenge of extracting high-quality surface meshes from 3D Gaussian Splatting representations. Rather than treating mesh extraction as a post-processing step, we integrate it directly into the optimization process and ensure that both the volumetric and surface representations are consistent. By extracting a mesh at every training iteration and back-propagating gradients to the Gaussians, we guide them toward configurations optimized for accurate surface extraction. See Fig. 2 for a visual overview of our method.

At each training iteration, our pipeline consists of five main steps:

- (1) Fetch the *trainable* Delaunay vertices derived from the Gaussian Pivots (Sec. 4.1).
- (2) Update the Delaunay triangulation (Sec. 4.1).
- (3) Fetch the *trainable* signed distance values for each Delaunay vertex (Sec. 4.2).
- (4) Apply GPU-based differentiable Marching Tetrahedra to extract the mesh (Sec. 4.3).
- (5) Back-propagate image-based and consistency losses to the Gaussian parameters, by simultaneously rendering the extracted mesh and the 3D Gaussians. (Sec. 5).

During optimization, we enforce geometric consistency between the mesh and the Gaussians by rendering and comparing depth and normal maps with both representations, encouraging the extracted surface to match the geometry encoded in the 3D Gaussians. This

Table 1. Resource requirements for different methods. We compare the computational resources required for training on high-resolution images, as well as the resulting model sizes across different approaches. All measurements are averaged across scenes from the Tanks & Temples dataset. Our method achieves high-quality reconstruction with significantly lower resource requirements, particularly regarding mesh complexity. The base version of our model uses up to 10x fewer Gaussians than competing methods depending on the scene, producing meshes with fewer vertices and triangles while preserving details and achieving better accuracy, making it more suitable for downstream applications.

	Training Resources			Output Mesh		
	#Gaussians (M)	GPU Mem (GB)	Time	#Vertices	#Triangles	Size (MB)
2DGS	0.98	4.7 GiB	29m	16.39 M	21.68 M	557.1
GOF	1.55	10.6 GiB	93m	16.49 M	33.17 M	600.0
RaDe-GS	1.56	12.4 GiB	42m	14.75 M	29.59 M	592.0
Ours (base)	0.28	10.0 GiB	50m	4.36 M	8.97 M	179.6
Ours (dense)	2.11	16.5 GiB	110m	6.89 M	13.79 M	276.1

bidirectional consistency framework not only improves the quality of the extracted mesh but also encourages Gaussians to converge toward better, solid surfaces with multi-view consistent geometry.

4 Differentiable Mesh Extraction

In this section, we provide more details on the different steps of our differentiable mesh extraction, allowing gradients to flow from the vertices of the mesh back to the parameters of the Gaussians during optimization.

4.1 Sampling Delaunay Vertices from Gaussians

The first step for extracting a surface mesh is to identify appropriate points in 3D space that will serve as the vertices of our Delaunay triangulation. While a naive approach would consist in using the **centers** of **all** Gaussians as Delaunay vertices, this simple strategy presents two significant limitations:

- (1) The **centers** of Gaussians are typically located on or near the surface of the scene, whereas for effective application of marching tetrahedra, one needs Delaunay vertices straddling the target surface.
- (2) Using **all** Gaussians as Delaunay vertices can be computationally expensive for large scenes.

To address the first issue, we follow the strategy proposed by Gaussian Opacity Fields (GOF) [Yu et al. 2024]: we sample 9 points per Gaussian, including its center and 8 corner points aligned with its principal axes. These are obtained by scaling and rotating the unit bounding box vertices and center $\{b_0, b_1 \dots b_8\}$:

$$p_{k,i} = \mu_k + R_k \times (s_k \odot b_i) \text{ for } i = 0 \dots 8, \quad (1)$$

where $\odot$ is the Hadamard product.

This sampling strategy ensures that the triangulation adapts to the anisotropic nature of the primitives, creating a more accurate representation of the surface.

To address the scalability challenge, we propose to sample Delaunay vertices only from Gaussians that are most likely to be located near the surface. To do so, we leverage the *importance-weighted sampling* introduced in Mini-Splatting2 [Fang and Wang 2024b], which ranks Gaussians based on their contribution to rendering across all training views, as reflected by the average magnitude of their blending coefficients along camera rays. We use these importance

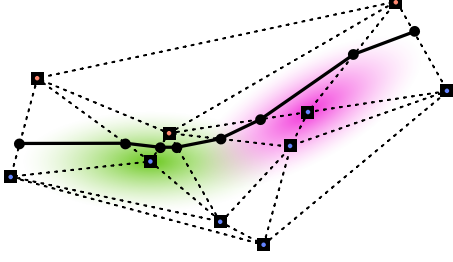


Fig. 3. **Overview of Gaussian Pivots in 2D.** Two Gaussians each generate Delaunay vertices (black squares) with continuous SDF values (red and blue dots). Their Delaunay triangulation (dashed lines) is iso-surfaced with Marching Tetrahedra, which places final mesh vertices (black dots) on edges where sign changes occur, resulting in the extracted mesh (bold lines).

scores as sampling probabilities, allowing us to select a subset of Gaussians that effectively preserve the geometric structure and fine details of the scene.

While only selected Gaussians receive direct gradient updates from the mesh regularization, since we impose consistency between the 3DGS and mesh rendering (see section 5), all Gaussians are constrained in practice. The result is a computationally efficient and geometrically accurate triangulation.

We propose two variants of our method:

- **Base model:** We sample a set of Gaussians using Importance-weighted sampling [Fang and Wang 2024b] and remove all other Gaussians from the scene. We extract a mesh at each iteration from the remaining Gaussians. This results in a lightweight set of Gaussians—between 0.1M and 0.5M depending on the complexity of the scene—and a correspondingly lightweight mesh that still captures fine details.
- **Dense model:** We also sample a set of Gaussians using Importance-weighted sampling [Fang and Wang 2024b], but we do not remove the other Gaussians from the scene: We maintain the large set of Gaussians (typically between 2M and 5M) for Gaussian Splatting but use only the sampled ones as pivots for generating the Delaunay vertices. The number of Delaunay vertices is approximately the same as the base model, still resulting in a lightweight mesh; however, having more Gaussians helps in learning better SDF values at these vertices. This model results in longer optimization times, but better performance.

Once we obtain the Delaunay vertices, we compute their **Delaunay triangulation** which provides a tetrahedralization required for the following step.

4.2 Computing Signed Distance Values

To compute a surface mesh from the previously obtained Delaunay triangulation, we rely on the Marching Tetrahedra algorithm [Doi and Koide 1991]. This algorithm, which we will describe in Sec. 4.3, requires a tetrahedral grid augmented with scalar values, typically derived from a Signed Distance Field (SDF). To apply it, we must assign a (signed) scalar value to each Delaunay vertex. Although these values do not originate from a true SDF, we refer to them as SDF values for simplicity.

We propose a simple yet effective strategy: **augmenting each Gaussian G_k with 9 optimizable SDF values $f_k \in \mathbb{R}^9$** —one for each Delaunay vertex. Importantly, these SDF values are decoupled from the Gaussians’ other parameters—such as opacity, scale, and rotation—allowing for localized control over the extracted isosurface level. This decoupling proves highly beneficial for accurately capturing fine surface details and ensuring strong consistency between the extracted mesh and the underlying volumetric representation. For an even faster convergence, we devise a custom initialization algorithm: please refer to the supplementary material for more details. An illustration of our approach is provided in Fig 3.

4.3 Differentiable Marching Tetrahedra

Once the Delaunay vertices and the SDF values are computed, we apply Marching Tetrahedra [Doi and Koide 1991] to extract a triangle mesh. For each tetrahedron with vertices having SDF values of opposite signs (indicating that the surface intersects the tetrahedron), the algorithm computes the intersection points of the surface with the edges of the tetrahedron. These 3 (or 4) intersection points form the vertices $\{v_n\}$ of the final mesh, and they are trivially connected with 1 (or 2) triangle faces. Specifically, given two Delaunay vertices $p_{k,i}$ and $p_{k',j}$ of a tetrahedron with SDF values $f_{k,i}$ and $f_{k',j}$ of opposite signs, the position of v_n is given by:

$$v_n = \frac{f_{k,i}p_{k',j} - f_{k',j}p_{k,i}}{f_{k,i} - f_{k',j}}, \quad (2)$$

Note that this mesh extraction process allows gradients to flow from the vertices of the extracted mesh back to the Gaussians through (a) the learnable SDF values and (b) the coordinates of the Delaunay vertices, computed from the means and covariances of the Gaussians.

5 Mesh-in-the-Loop Optimization

We build on top of previous works [Huang et al. 2024; Kerbl et al. 2023; Yu et al. 2024; Zhang et al. 2024], whose optimization is based on successive iterations of rendering the scene with 3D Gaussians and comparing the resulting image to the target image from the captured dataset. Our approach can be plugged into any Gaussian-based method able to render depth maps and normal maps through differentiable rasterization [Huang et al. 2024; Yu et al. 2024; Zhang et al. 2024]. Please note that depending on the method, the definition of the rendered depth and normal maps may differ.

A key advantage of our approach lies in the ability to extract a triangle mesh during optimization, perform any differentiable operation on the surface, and backpropagate the resulting gradients to the Gaussians. To couple the two representations, we rely on differentiable mesh rendering: we directly compare depth and normal maps rendered from both representations to enforce geometric consistency between the Gaussians and the extracted surface. Below, we describe the different loss functions involved during optimization.

5.1 Volumetric rendering

To optimize the Gaussians, we rely on the same volumetric rendering loss $\mathcal{L}_{\text{vol}}$ as [Huang et al. 2024; Yu et al. 2024; Zhang et al. 2024],

Table 2. **Quantitative comparison on the Tanks & Temples Dataset [Knapitsch et al. 2017]**. We report the F1-score and average optimization time. All results are evaluated with the official evaluation scripts. Ours_X uses the differentiable rasterizer from X. VCR-GauS [Chen et al. 2024] relies on a pre-trained normal estimation model. Best results for implicit methods are highlighted in blue; best results for explicit methods are highlighted in red. When a method fails (OOM), we report the mean of the successful scenes, denoted by a superscript asterisk (*). Our method achieves the best F1 score among explicit representations, even though it outputs lighter meshes than other approaches.

	Implicit						Explicit									
	Gaussian UDF	GS-Pull	GSDF	NeuS	Geo-Neus	Neuralangelo	SuGaR	3DGS	2DGS	GOF	RaDe-GS	QGS	VCR-GauS	Ours _{RaDe-GS} (base)	Ours _{GOF} (base)	Ours _{RaDe-GS} (dense)
Barn	0.27	0.60	0.16	0.29	0.33	0.70	0.14	0.13	0.36	0.51	0.43	0.43	0.62	0.55	0.59	0.64
Caterpillar	0.17	0.37	0.11	0.29	0.26	0.36	0.16	0.08	0.23	0.41	0.32	0.31	0.26	0.38	0.39	0.38
Courthouse	0.03	0.16	OOM	0.17	0.12	0.28	0.08	0.09	0.13	0.28	0.21	0.26	0.19	0.28	0.29	0.31
Ignatius	0.38	0.71	0.34	0.83	0.72	0.89	0.33	0.04	0.44	0.68	0.69	0.79	0.61	0.73	0.78	0.76
Meetingroom	0.17	0.22	0.01	0.24	0.20	0.32	0.15	0.01	0.16	0.28	0.25	0.25	0.19	0.25	0.26	0.28
Truck	0.30	0.52	0.25	0.45	0.45	0.48	0.26	0.19	0.26	0.59	0.51	0.60	0.52	0.60	0.62	0.59
Mean	0.22	0.43	0.18*	0.38	0.35	0.50	0.19	0.09	0.30	0.46	0.40	0.44	0.40	0.47	0.49	0.49
Time	90m	37.6m	70m	>24h	>24h	>24h	73 m	7.9 m	12.3 m	69m	11.5 m	120 m	53 m	50 m	150 m	110 m



Fig. 4. **Examples of surface reconstruction results on the Tanks and Temples and MipNeRF 360 datasets.** Our method produces meshes with clean surfaces and very fine details, while being lighter than previous methods (Stump: 320MB, Garden: 301MB, Barn: 313MB)

which consists of two photometric terms and a regularization term:

$$\mathcal{L}_{\text{vol}} = (1 - \lambda_{\text{RGB}})\mathcal{L}_1 + \lambda_{\text{RGB}}\mathcal{L}_{\text{D-SSIM}} + \lambda_{\text{N}}\mathcal{L}_{\text{N}}, \quad (3)$$

where $\mathcal{L}_1$ is the L1 loss, $\mathcal{L}_{\text{D-SSIM}}$ is a D-SSIM term and $\mathcal{L}_{\text{N}}$ is a normal consistency loss defined as the following sum over pixels i :

$$\mathcal{L}_{\text{N}} = \sum_i \left(1 - \mathbf{N}(i) \cdot \tilde{\mathbf{N}}(i)\right), \quad (4)$$

where $\mathbf{N}_i$ is the expected normal at pixel i computed from volumetric rendering, and $\tilde{\mathbf{N}}_i$ is the normal direction at pixel i obtained by applying finite difference on the rendered depth map, following [Huang et al. 2024; Yu et al. 2024; Zhang et al. 2024]. This term encourages Gaussians to align with their neighbors and drastically reduces the noise in the rendered depth and normal maps.

5.2 Volume-to-Surface Consistency

To enforce consistency between the geometry encoded in (a) the Gaussians as a volumetric representation and (b) the extracted surface mesh, we introduce the following loss:

$$\mathcal{L}_{\text{mesh}} = \lambda_{\text{MD}}\mathcal{L}_{\text{MD}} + \lambda_{\text{MN}}\mathcal{L}_{\text{MN}}. \quad (5)$$

Here $\mathcal{L}_{\text{MD}}$ is a depth consistency loss defined as follows:

$$\mathcal{L}_{\text{MD}} = \sum_i \log(1 + |D(i) - D_{\text{M}}(i)|), \quad (6)$$

with D the depth map rendered from the Gaussians and D_{M} the depth map rendered from the mesh. The normal consistency loss $\mathcal{L}_{\text{MN}}$ is defined as

$$\mathcal{L}_{\text{MN}} = \sum_i \left(1 - \tilde{\mathbf{N}}(i) \cdot \mathbf{N}_{\text{M}}(i)\right), \quad (7)$$

with $\mathbf{N}_{\text{M}}(i)$ the normal of the face of the mesh rasterized at pixel i .

5.3 Regularization

Despite their effectiveness, the previous losses are not sufficient for reconstructing optimal meshes, and two key challenges remain.

Erosion. During optimization, if all SDF values inside a tetrahedron become positive, fine details of the geometry can be eroded or even lost. Due to the sharp nature of mesh rasterization, it becomes extremely difficult for the representation to recover this geometry, even with antialiasing applied to mesh renderings. Once a region is eroded, the gradient signal becomes weak or non-existent, making it challenging for optimization to restore the missing surface.

To address this issue, we propose a regularization term $\mathcal{L}_{\text{erosion}}$ that encourages geometry to be preserved:

$$\mathcal{L}_{\text{erosion}} = \sum_{g \in G_{\text{Del}}} \max(0, f_{\mu_g}), \quad (8)$$

where G_{Del} is the set of Gaussians sampled for the Delaunay triangulation and f_{μ_g} is the SDF value of the center of the g -th Gaussian. As mentioned in Section 4, the centers of all Gaussians sampled for the triangulation are used as Delaunay vertices: this regularization term aims to include the centers of these Gaussians inside the surface by encouraging their SDF values to become negative, preventing erosion in all areas where at least one Gaussian has been sampled. $\mathcal{L}_{\text{erosion}}$ is applied only to the centers of selected Gaussian pivots, not to all tetrahedral vertices, preventing collapse. This ensures regularization is effective without harming mesh integrity.

Interior artifacts. The previous losses do not provide adequate constraints on occluded parts of the scene. As a consequence, the interior of the surface mesh, which does not receive proper supervision through depth and normal renderings, tends to produce chaotic structures and internal cavities, rather than being empty as it should

be. These interior artifacts can impact downstream applications such as mesh editing or physics simulations.

To prevent such artifacts inside the mesh, we introduce a novel loss based on a feedback loop between the SDF values and the extracted mesh. This loss aims to enforce occluded points located inside the mesh to have negative SDF values. Specifically, after extracting the surface using the Gaussians and their SDF values as described in section 4.3, we use the mesh to build an occupancy label $o_p \in \{0, 1\}$ for each Delaunay site p , indicating whether these points are inside or outside the visible portion of the surface mesh as described below. We finally enforce the Delaunay vertices labeled as inside to have negative SDF values, without enforcing any additional constraint for Delaunay vertices labeled as outside—as these points already receive supervision from the renderings.

This regularization term is computed as follows:

$$\mathcal{L}_{\text{interior}} = \sum_p H(\sigma(-f_p), o_p) \cdot o_p, \quad (9)$$

where H is the cross-entropy loss, σ is the sigmoid function, f_p is the SDF value of the Delaunay site p , and o_p is the occupancy label of p .

To compute the occupancy labels o_p , we first render depth maps of the mesh from all training viewpoints. Then, for any Delaunay site p , we classify it as inside the mesh if it located “behind” all depth maps containing p in their field of view. Updating the occupancy labels requires a few seconds and can be performed every 200 iterations, resulting in a computationally efficient and effective regularization term for producing meshes with clean interiors.

Our full optimization loss is finally defined as:

$$\mathcal{L} = \mathcal{L}_{\text{vol}} + \mathcal{L}_{\text{mesh}} + \mathcal{L}_{\text{reg}}, \quad (10)$$

with

$$\mathcal{L}_{\text{reg}} = \lambda_{\text{erosion}} \mathcal{L}_{\text{erosion}} + \lambda_{\text{interior}} \mathcal{L}_{\text{interior}}, \quad (11)$$

where λ_{erosion} and $\lambda_{\text{interior}}$ are hyperparameters controlling the strength of the anti-erosion and interior regularization, respectively.

6 Experiments

In this section, we evaluate our method on various datasets and compare it with state-of-the-art approaches for surface reconstruction from multi-view images. We first describe our experimental setup, then present quantitative and qualitative results for surface reconstruction quality, followed by ablation studies and applications in mesh editing and animation. We show that the obtained meshes are clean, accurate and light as showcased in Fig. 4.

Implementation Details. Our method is implemented in PyTorch, building upon the RaDe-GS codebase for Gaussian Splatting with depth and normal maps rendering. We train the Gaussian representation using standard 3D Gaussian Splatting techniques with aggressive densification and importance-weighted sampling inspired by Mini-Splatting2 [Fang and Wang 2024b]. We provide more details on the optimization procedure, SDF normalization, mesh rendering and Delaunay triangulation in the supplementary material.

Datasets. We evaluate our method on several standard datasets: (1) the Tanks and Temples (T&T) dataset [Knapitsch et al. 2017], which contains complex real-world scenes with varying scales and challenging geometry; (2) the DTU dataset [Jensen et al. 2014], which provides scans of real objects with ground truth geometry; (3) the Mip-NeRF 360 dataset [Barron et al. 2022], which contains unbounded and complex scenes; (4) and DeepBlending [Hedman et al. 2018] which contains real-world scenes with challenging geometry. For T&T, we follow previous works [Huang et al. 2024; Yu et al. 2024; Zhang et al. 2024] and use the scenes Barn, Caterpillar, Ignatius, Courthouse, Meetingroom, and Truck. For DTU, we use the same 15 scenes as in previous works [Huang et al. 2024; Yu et al. 2024; Zhang et al. 2024].

Metrics. For quantitative evaluation of surface reconstruction quality, we follow standard practice [Yu et al. 2024; Zhang et al. 2024] and report F1-score for T&T and Chamfer Distance for DTU. For novel view synthesis on Mip-NeRF 360 and DeepBlending, we report PSNR, SSIM, and LPIPS metrics.

No dataset provides ground truth geometry for background objects. To mitigate this issue we use Mesh-Based Novel View Synthesis (see Section. 6.3) to measure the visual consistency between the extracted mesh and reference views, which contain background information. This evaluation measures the ability of reconstructed meshes to accurately represent full scenes, including background objects, while requiring only a set of ground truth images.

6.1 Resources Requirements.

The addition of mesh-based constraints inevitably leads to increased computational load. However, we observed that only a subset of Gaussians is critical for producing an effective tetrahedral structure.

Building on this, inspired by the sampling strategy introduced in Mini-Splatting2 [Fang and Wang 2024b], we employ it for a distinct purpose: selecting the Gaussians which will spawn the Gaussian pivots that serve as Delaunay vertices. Unlike Mini-Splatting2, our dense model preserves all Gaussians throughout optimization, and the sampling criterion is used exclusively to identify those most suitable for mesh construction rather than to prune the representation.

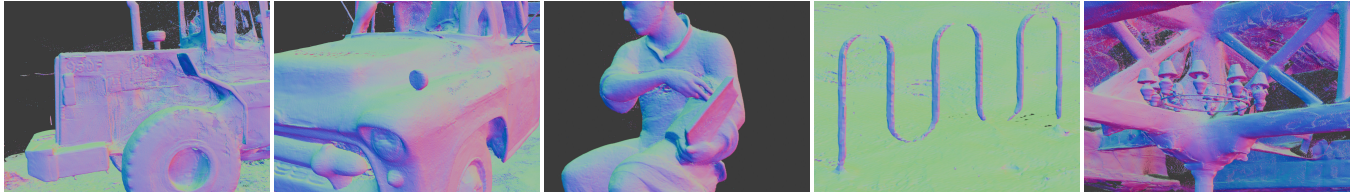
This targeted selection yields lighter and more accurate meshes, significantly reducing the computational cost of triangulation, Marching Tetrahedra, and mesh rasterization. Table 1 provides a comprehensive comparison of the computational resources required by concurrent methods.

Memory Requirements. All experiments were conducted on a single NVIDIA RTX 4090 GPU with 24GB of VRAM. The base model contains between 0.1 and 0.5 million Gaussians, requiring 10GB of VRAM during training. The dense model generally contains between 2 and 4 million Gaussians requiring up to 17GB.

Time Requirements. A complete training run for the base model on a single GPU takes 25 minutes for bounded scenes from DTU and between 40 and 50 minutes for unbounded scenes from Tanks and Temples and Mip-NeRF 360, while the dense model requires up to 2 hours for unbounded scenes.



Ground Truth Images



(a) 2DGS



(b) GOF



(c) RaDe-GS



(d) ours

Fig. 5. **Qualitative comparison of surface reconstruction results on the Tanks and Temples dataset.** MLo produces meshes of full scenes that better fit the target surfaces, with reduced artifacts such as cavities (three left images). Our approach effectively addresses the erosion problem that plagues previous methods (two right images), while producing lighter meshes. In addition, it is particularly effective at recovering peripheral objects (e.g., the bike rack or chandelier) that are often missed by post-hoc methods.

Table 3. **Surface reconstruction metrics on the DTU dataset.** We report the Chamfer Distance across 15 scenes. Our method achieves competitive performance with significantly fewer Gaussians and mesh vertices compared to previous approaches.

		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
	3D GS [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	5.2m
explicit	SuGaR [Guédon and Lepetit 2024b]	1.47	1.33	1.13	0.61	2.25	1.71	1.15	1.63	1.62	1.07	0.79	2.45	0.98	0.88	0.79	1.33	52m
	2D GS [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	8.9m
	GOF [Yu et al. 2024]	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	55m
	RaDe-GS [Zhang et al. 2024]	0.46	0.73	0.33	0.38	0.79	0.75	0.76	1.19	1.22	0.62	0.70	0.78	0.36	0.68	0.47	0.68	8.3m
	Ours (base)	0.43	0.74	0.34	0.37	0.80	0.74	0.70	1.21	1.22	0.66	0.62	0.80	0.37	0.76	0.48	0.68	25m

Storage Requirements. Our base model uses significantly fewer Gaussians (0.1-0.5M) and produces meshes with fewer vertices compared to previous approaches. Despite using fewer resources, our method achieves superior performance in terms of surface reconstruction quality. The dense variant of our model maintains a larger set of Gaussians but still produces meshes with fewer vertices and triangles than competing methods, making it more suitable for downstream applications that require efficient mesh representations.

6.2 Surface Reconstruction

Table 2 presents the quantitative results on the Tanks and Temples dataset. Our method consistently outperforms previous approaches in terms of F-score, demonstrating the effectiveness of our mesh-in-the-loop optimization strategy. The dense variant of our method achieves the best results among all Gaussian-based approaches.

We provide results on the DTU dataset, consisting of small, object-centric scenes. Table 3 shows the results on this dataset. Note that, DTU consists of isolated objects in highly controlled scenes, where standard post-hoc mesh extraction is already effective. MILo is particularly effective for complex full-scene reconstruction. Still, our mesh-in-the-loop optimization maintains competitive performance in terms of Chamfer Distance (CD) on DTU. In line with established practice, we evaluate performance using F-Score on the Tanks and Temples dataset and Chamfer Distance (CD) on DTU.

Figure 5 shows qualitative comparisons of our method with previous approaches on selected scenes from the Tanks and Temples and Mip-NeRF 360 datasets. Our method produces meshes with significantly cleaner surfaces and better preservation of fine details. In particular, our approach effectively addresses the erosion problem present in previous methods, resulting in more complete reconstructions of thin structures and complex geometry: see Fig. 6 for a detailed analysis.

6.3 Mesh-Based Novel View Synthesis

Despite extensive research on mesh reconstruction from images, current evaluation benchmarks remain limited. Common datasets—DTU, MipNeRF 360, and Tanks & Temples—all have notable drawbacks: MipNeRF 360 lacks ground-truth geometry; DTU features overly controlled, object-centric scenes; and Tanks & Temples provides sparse ground-truth limited to foreground regions. Moreover, evaluation protocols to measure the alignment of extracted surfaces with

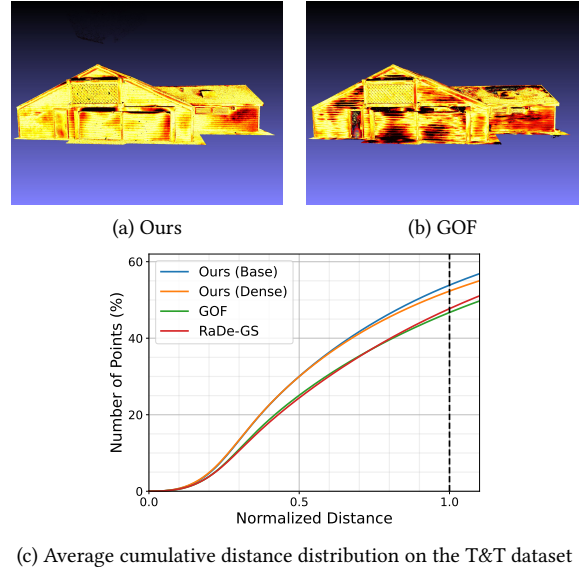


Fig. 6. **Reconstruction → Ground truth Distance Evaluation.** Qualitatively, our reconstructions using the differentiable rasterizer from RaDe-GS (a) are closer to the ground truth point clouds than both GOF (b) and RaDe-GS. The T&T [Knapitsch et al. 2017] evaluation toolkit reports cumulative histograms (c) showing the proportion of reconstructed points within a given distance from the ground truth surface. The dashed line marks the threshold used to compute the F1-score. Our method produces significantly more points close to the true surface (within a normalized distance of 1.0), demonstrating higher reconstruction accuracy compared to prior approaches.

ground truth images remain uncommon in surface reconstruction research.

We highlight this as an important issue in surface reconstruction. Inspired by previous works which evaluate surface-based view synthesis methods [Reiser et al. 2024b], we use Mesh-Based Novel View Synthesis as an evaluation method for meshes that relies solely on ground-truth RGB images. The core intuition is that better geometry should yield better mesh-based renderings: for each test view, we render the scene using the mesh and compare the resulting image to the corresponding ground-truth image.

By measuring the visual consistency between mesh renderings and reference views, this metric allows us to evaluate:

Table 4. **Mesh-Based Novel View synthesis metrics on real-world datasets.** We report PSNR, SSIM, and LPIPS metrics, as well as the total number of Gaussians and vertices (in millions). Our method demonstrates superior performance on these challenging unbounded scenes. Ours (base) uses the differentiable rasterizer from RaDe-GS.

	MipNeRF 360					Tanks & Temples					DeepBlending				
	PSNR ↑	SSIM ↑	LPIPS ↓	#Gaussians ↓	#Verts ↓	PSNR ↑	SSIM ↑	LPIPS ↓	#Gaussians ↓	#Verts ↓	PSNR ↑	SSIM ↑	LPIPS ↓	#Gaussians ↓	#Verts ↓
2DGS	15.36	0.4987	0.4749	1.88	4.31	14.23	0.5697	0.4854	0.98	16.39	17.46	0.7133	0.5556	1.51	16.21
GOF	20.78	0.5730	0.4654	2.99	32.80	21.69	0.6904	0.3261	1.25	11.63	27.44	0.8078	0.2549	1.64	12.64
RaDe-GS	23.56	0.6685	0.3612	2.91	30.85	20.51	0.6595	0.3447	1.18	10.06	26.74	0.7964	0.2621	1.65	11.53
Ours (base)	24.09	0.6885	0.3235	0.46	6.73	21.46	0.7067	0.3489	0.28	4.36	28.04	0.8336	0.2285	0.38	5.40

- Geometric artifacts and misalignment between the reconstructed surface and the ground truth images (e.g., surface erosion or inflation);
- Mesh completeness, since missing geometry usually results in degraded rendering performance;
- Background reconstruction, even when 3D ground truth is unavailable.

To render the mesh from a given viewpoint, we first rasterize the triangles using Nvdiffrast [Laine et al. 2020]. Then, we associate a color with each pixel based on the rasterized triangle.

A naive approach would rely on vertex colors, assigning RGB values to each mesh vertex and interpolating those values to determine a pixel’s color. However, this method biases the metric towards dense meshes, as image quality would suffer for sparse, yet accurate, meshes due to limited color resolution.

To overcome this, we decouple color from mesh resolution by employing a neural color field $F_{\text{color}} : \mathbb{R}^3 \rightarrow [0, 1]^3$ for texturing the mesh. Specifically, to obtain the rendered color for any pixel p , we first calculate the 3D location $P \in \mathbb{R}^3$ of the triangle point rasterized onto the pixel p . This is achieved by backprojecting the depth value using camera parameters. We then query the neural color field at this backprojected surface point to retrieve its RGB value $F_{\text{color}}(P)$. Unlike vertex colors, this mesh texturing process stores color values in a neural field rather than per-vertex, ensuring color assignment is independent of mesh resolution.

In practice, for each mesh under evaluation, we first train the neural field using only the training views. We adopt TensorRF [Chen et al. 2022] as the backbone representation and optimize it for 5k iterations. We then render test views for evaluation. This metric effectively quantifies the alignment between mesh geometry and the image data. Note that, this evaluation protocol assumes dense testing views, as achieving good mesh-based rendering performance on a sparse test set does not necessarily translate into high quality of the underlying geometry.

Table 4 shows how Mesh-Based Novel View Synthesis allows us to compare against other state-of-the-art methods. We use the differentiable rasterizer from RaDe-GS for training Ours (base). Our approach outperforms 2DGS [Huang et al. 2024] and RaDe-GS [Zhang et al. 2024] across all metrics. GOF [Yu et al. 2024] produces highly detailed meshes, yielding strong PSNR results, but our method achieves comparable PSNR while outperforming in SSIM and LPIPS. This suggests that our meshes are not only accurate but also exhibit fewer visual artifacts and less noise compared to those from GOF. We include a qualitative comparison in the supplementary material.

Moreover, the performance on Tanks & Temples under the Mesh-Based Novel View Synthesis metric (Table 4) shows strong correlation with the F1 score computed against ground-truth geometry

Table 5. **Ablation study on the Tanks and Temples dataset.** We report only the average F_1 -score across all scenes to emphasize the impact of each component.

Method	Average F_1 -score ↑
Baseline	0.41
+ $\mathcal{L}_{\text{MD}}$	0.46
+ $\mathcal{L}_{\text{mesh}}$	0.44
+ $\mathcal{L}_{\text{mesh}} + \mathcal{L}_{\text{erosion}}$	0.44
Ours (Baseline + $\mathcal{L}_{\text{mesh}} + \mathcal{L}_{\text{reg}}$)	0.47
Ours (GOF)	0.49

(Table 2). For example, among the common baselines, 2DGS and RaDe-GS rank fourth and third, respectively, under both metrics. Ours (base) and GOF also perform consistently, with the Ours (base) achieving a slightly higher F1 score and GOF performing marginally better on Mesh-Based Novel View Synthesis. These results indicate that our proposed metric is well aligned with ground-truth geometric accuracy.

6.4 Novel View Synthesis

While our primary focus is on surface reconstruction, we also provide in supplementary material an evaluation of the novel view synthesis quality of our optimized Gaussians. Our method maintains competitive rendering quality compared to previous approaches, demonstrating that our mesh-in-the-loop optimization does not compromise the visual fidelity of the Gaussian representation.

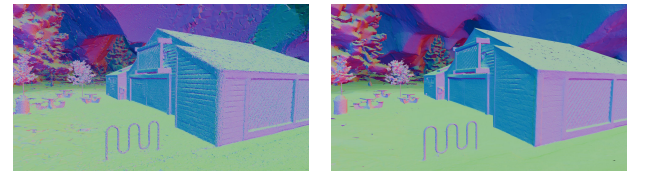


Fig. 7. **Effect of adding Normal supervision.** Despite showcasing an apparent decrease in performance in our ablation study (see the supplementary material), this figure supports our statement that the combination of depth and normal supervision is essential. The resulting mesh when using only $\mathcal{L}_{\text{MD}}$ is significantly noisier than when supervising with $\mathcal{L}_{\text{mesh}}$.

6.5 Ablation Study

To evaluate the contribution of each component, we conduct ablation studies on the Tanks and Temples dataset (see Tab. 5). Adding depth supervision via the mesh loss (+ $\mathcal{L}_{\text{MD}}$) substantially boosts F_1 -score, while further including the normal loss (+ $\mathcal{L}_{\text{mesh}}$) slightly

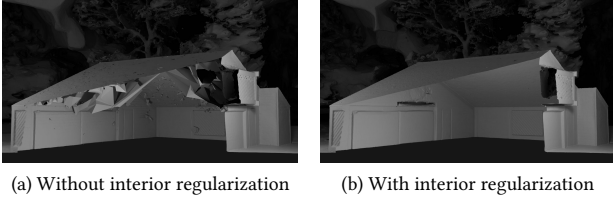


Fig. 8. **Comparison of mesh interiors.** We show a cross-section of the reconstructed mesh of the *Barn* scene to reveal its internal structure. Our interior regularization $\mathcal{L}_{\text{interior}}$ effectively eliminates internal artifacts, producing clean, watertight meshes with empty interiors, which is crucial for downstream applications such as physics simulations and animation.

reduces it. Still, Fig. 7 shows that combining both is critical for removing noise from meshes.

Introducing the anti-erosion loss ($+\mathcal{L}_{\text{mesh}} + \mathcal{L}_{\text{erosion}}$) improves reconstruction by preserving thin structures and fine details—especially in challenging regions like fences or vegetation—by penalizing surface erosion. Our loss also acts as a mechanism to recover geometry lost due to SDF sign flips yielding a modest F_1 -score gain (Fig. 5).

Fig. 8 highlights the benefit of our interior regularization, which eliminates internal cavities and artifacts, producing watertight surfaces suitable for downstream tasks like physics simulation.

Combining all components leads to the highest F_1 -score, validating the effectiveness of our full method. We also demonstrate its plug-and-play nature by integrating it into GOF’s pipeline.

6.6 Analysis of Mesh Extraction

As discussed in Section 2, many existing approaches to mesh extraction from images rely on the TSDF algorithm. These methods typically follow a simple yet seemingly effective pipeline: they first estimate accurate depth maps using a variety of techniques, and then fuse these maps into a mesh via TSDF. However, we argue that TSDF fusion does not scale well to large-scale reconstructions. This limitation arises from its reliance on a fixed 3D grid, which drives up memory requirements as resolution increases, and from its tendency to oversmooth geometric detail.

In Figure 9, we present a quantitative comparison between our extraction method and TSDF fusion. For TSDF fusion, we optimized Gaussians with the same densification strategy as MILo, but removed the mesh-in-the-loop regularization term and employed the same loss function as in [Huang et al. 2024; Yu et al. 2024; Zhang et al. 2024] to avoid bias. TSDF fusion requires substantial memory and is prone to out-of-memory failures; consequently, we exclude the Courthouse scene from this benchmark, as TSDF fusion could not be run across all mesh resolutions due to memory constraints. As shown in Figure 9, our method consistently achieves higher reconstruction quality, demonstrating the robustness of our learnable SDF strategy. We highlight that, TSDF fusion requires **iterating over all training viewpoints for each extraction**, leading to a long and computationally expensive process that **prevents integration into the training loop of Gaussian Splatting representations**.

By contrast, MILo leverages learnable SDF values and a scalable pivot set to enable lightweight and efficient mesh extraction, making it practical to incorporate mesh extraction directly into the training loop.

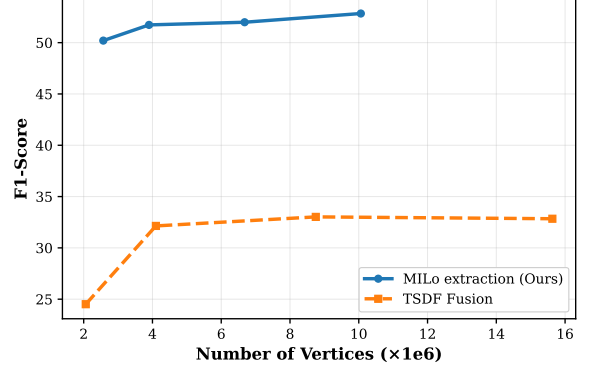


Fig. 9. Quantitative evaluation of mesh extraction methods on the Tanks & Temples dataset across different mesh resolutions. We compare (a) traditional TSDF fusion, which relies on a regular 3D grid, with (b) our mesh extraction method, which learns SDF values at Gaussian pivots.

6.7 Limitations and Conclusions

While our method significantly improves surface reconstruction quality, it still has some limitations. First, the computational cost of extracting and processing the mesh at every iteration increases the training time compared to standard Gaussian Splatting, although it still remains manageable compared to baselines. Second, the quality of the reconstruction still depends on the initial distribution of Gaussians, which may not be optimal for all scenes.

Conceptually, our framework unlocks many surface-based processing tools for Gaussians that can be applied during training. This integration of mesh-based operations with Gaussian representations opens new possibilities for geometry manipulation, regularization, and enhancement throughout the optimization process, not just as a post-processing step.

Future work could explore novel adaptive sampling strategies for Delaunay sites, and improved initialization techniques. Additionally, extending our approach to *dynamic* scenes and exploring its applications in real-time rendering and interactive applications would be promising directions for future research.

Further exploration of surface-based processing tools that can be integrated into the training pipeline could also yield significant improvements in reconstruction quality and versatility.

Finally, through a mesh-based novel view synthesis evaluation we highlight the lack of standardized quantitative protocols to evaluate the alignment of the reconstructed surface with the training images. This is true on widely used datasets such as DTU, MipNeRF 360, and Tanks & Temples. While this approach offers a step toward addressing this gap, it remains an initial effort; future work could

focus on refining the evaluation methodology to better capture reconstruction quality in challenging real-world scenes.

7 Acknowledgements

Parts of this work were supported by the ERC Consolidator Grant “VEGA” (No. 101087347), the ERC Advanced Grant “explorer” (No. 101097259), the ANR AI Chair AIGRETTE. Co-funded by the European Union (EU) (ERC Advanced grant FUNGRAPH No 788065 and ERC Advanced Grant NERPHYS No 101141721). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the EU or the European Research Council. Neither the EU nor the granting authority can be held responsible for them.

References

- Nina Amenta, Marshall Bern, and Manolis Kamvyselis. 1998. A new Voronoi-based surface reconstruction algorithm. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques - SIGGRAPH '98*. ACM Press, 415–421. doi:10.1145/280814.280947
- Nina Amenta, Sunghee Choi, and Ravi Krishna Kolluri. 2001. The power crust. In *Proceedings of the sixth ACM symposium on Solid modeling and applications* (New York, NY, USA) (SMA '01). Association for Computing Machinery, 249–266. doi:10.1145/376957.376986
- Franz Aurenhammer. 1991. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM computing surveys (CSUR)* 23, 3 (1991), 345–405.
- Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin Brualla, and Pratul P Srinivasan. 2021. Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. In *ICCV*. 5855–5864.
- Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *CVPR*. IEEE, 5460–5469. doi:10.1109/CVPR52688.2022.00539
- Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2023. Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. In *ICCV*. 19697–19705.
- Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. 2022. TensoRF: Tensorial Radiance Fields. In *ECCV*. 333–350.
- Hanlin Chen, Chen Li, and Gim Hee Lee. 2023. 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. 2024. VCR-GauS: View Consistent Depth-Normal Regularizer for Gaussian Surface Reconstruction. In *NeurIPS*.
- Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, and Weiwei Xu. 2024. High-quality surface reconstruction using gaussian surfels. In *ACM SIGGRAPH 2024 conference papers*. 1–11.
- Tamal K. Dey and Samrat Goswami. 2003. Tight cocone: a water-tight surface reconstructor. In *Proceedings of the eighth ACM symposium on Solid modeling and applications* (Seattle Washington USA). ACM, 127–134. doi:10.1145/781606.781627
- Akio Doi and Akio Koide. 1991. An efficient method of triangulating equi-valued surfaces by using tetrahedral cells. *IEICE TRANSACTIONS on Information and Systems* 74, 1 (1991), 214–224.
- Tim Elsner, Victor Czech, Julia Berger, Zain Selman, Isaak Lim, and Leif Kobbelt. 2023. Adaptive Voronoi NeRFs. *arXiv:2303.16001 [cs]* <http://arxiv.org/abs/2303.16001>
- Guangchi Fang and Bing Wang. 2024a. Mini-splatting: Representing scenes with a constrained number of gaussians. In *European Conference on Computer Vision*. Springer, 165–181.
- Guangchi Fang and Bing Wang. 2024b. Mini-Splatting2: Building 360 Scenes within Minutes via Aggressive Gaussian Densification. *arXiv preprint arXiv:2411.12788* (2024).
- Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance Fields without Neural Networks. In *CVPR*. 5501–5510.
- Diego Gomez, Bingchen Gong, and Maks Ovsjanikov. 2025. FourierRF: Few-Shot NeRFs via Progressive Fourier Frequency Control. In *2025 International Conference on 3D Vision (3DV)*. 607–615. doi:10.1109/3DV66043.2025.00061
- Shrividhan Govindarajan, Daniel Rebain, Kwang Moo Yi, and Andrea Tagliasacchi. 2025. Radiant Foam: Real-Time Differentiable Ray Tracing. *arXiv:2502.01157 [cs]* doi:10.48550/arXiv.2502.01157
- Antoine Guédon, Tomoki Ichikawa, Kohei Yamashita, and Ko Nishino. 2025. MATCha Gaussians: Atlas of Charts for High-Quality Geometry and Photorealism From Sparse Views. In *CVPR*.
- Antoine Guédon and Vincent Lepetit. 2024a. Gaussian Frosting: Editable Complex Radiance Fields with Real-Time Rendering. In *ECCV*.
- Antoine Guédon and Vincent Lepetit. 2024b. SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering. In *CVPR*. 5354–5363.
- Peter Hedman, Julien Philip, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel J. Brostow. 2018. Deep blending for free-viewpoint image-based rendering. *ACM TOG* 37, 6 (2018), 257. doi:10.1145/3272127.3275084
- Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2024. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. In *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery. doi:10.1145/3641519.3657428
- Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2024. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. In *ACM SIGGRAPH Conference Papers*. 1–11.
- Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. 2014. Large scale multi-view stereopsis evaluation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 406–413.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM TOG* 42, 4 (July 2023). <https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/>
- Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hosam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 2024. 3d gaussian splatting as markov chain monte carlo. *Advances in Neural Information Processing Systems* 37 (2024), 80965–80986.
- Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1–13.
- Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. 2017. Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction. *ACM TOG* 36, 4 (2017).
- Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. 2020. Modular Primitives for High-Performance Differentiable Rendering. *ACM TOG* 39, 6 (2020).
- Shujuan Li, Yu-Shen Liu, and Zhizhong Han. 2025. Gaussianudf: Inferring unsigned distance functions through 3d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 27113–27123.
- Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. 2023. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8456–8465.
- Rong Liu, Dylan Sun, Meida Chen, Yue Wang, and Andrew Feng. 2025. Deformable Beta Splatting. *arXiv preprint arXiv:2501.18630* (2025).
- Saswat Subhajiyo Mallick, Rahul Goel, Bernhard Kerbl, Markus Steinberger, Francisco Vicente Carrasco, and Fernando De La Torre. 2024. Taming 3dgs: High-quality radiance fields with limited resources. In *SIGGRAPH Asia 2024 Conference Papers*. 1–11.
- Nissim Maruani, Roman Klokov, Maks Ovsjanikov, Pierre Alliez, and Mathieu Desbrun. 2023. VoroMesh: Learning Watertight Surface Meshes with Voronoi Diagrams. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. 14565–14574. https://openaccess.thecvf.com/content/ICCV2023/html/Maruani_VoroMesh_Learning_Watertight_Surface_Meshes_with_Voronoi_Diagrams_ICCV_2023_paper.html
- Nissim Maruani, Maks Ovsjanikov, Pierre Alliez, and Mathieu Desbrun. 2024. PoNQ: A Neural QEM-Based Mesh Representation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Seattle, WA, USA). IEEE, 3647–3657. doi:10.1109/CVPR52733.2024.00350
- Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*. 99–106.
- Nicolas Moenne-Loccoz, Ashkan Mirzaei, Or Perel, Riccardo de Lutio, Janick Martinez Esturo, Gavriel State, Sanja Fidler, Nicholas Sharp, and Zan Gojcic. 2024. 3D Gaussian Ray Tracing: Fast Tracing of Particle Scenes. *ACM Transactions on Graphics (TOG)* 43, 6 (2024), 1–19.
- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM TOG* 41, 4, Article 102 (July 2022), 15 pages. doi:10.1145/3528223.3530127
- J. Munkberg, J. Hasselgren, T. Shen, J. Gao, W. Chen, A. Evans, T. Müller, and S. Fidler. 2022. Extracting Triangular 3D Models, Materials, and Lighting From Images. In *CVPR*.
- Christian Reiser, Stephan Garbin, Pratul P Srinivasan, Dor Verbin, Richard Szeliski, Ben Mildenhall, Jonathan T Barron, Peter Hedman, and Andreas Geiger. 2024a. Binary Opacity Grids: Capturing Fine Geometric Detail for Mesh-Based View Synthesis. *arXiv preprint arXiv:2402.12377* (2024).
- Christian Reiser, Stephan Garbin, Pratul P. Srinivasan, Dor Verbin, Richard Szeliski, Ben Mildenhall, Jonathan T. Barron, Peter Hedman, and Andreas Geiger. 2024b. Binary Opacity Grids: Capturing Fine Geometric Detail for Mesh-Based View Synthesis.

- SIGGRAPH (2024).
- T. Shen, J. Munkberg, G. Hasselgren, K. Yin, Z. Wang, W. Chen, Z. Gojcic, S. Fidler, N. Sharp, and J. Gao. 2023. Flexible Isosurface Extraction for Gradient-Based Mesh Optimization. *ACM TOG* 42, 4 (2023).
- Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. 2022. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. *CVPR* (2022).
- Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Proceedings of the 35th International Conference on Neural Information Processing Systems (NIPS)*. Article 2081, 13 pages.
- Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. 2023. NeuS2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3295–3306.
- Frederik Warburg, Ethan Weber, Matthew Tancik, Aleksander Holynski, and Angjoo Kanazawa. 2023. Nerfbusters: Removing ghostly artifacts from casually captured nerfs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 18120–18130.
- Francis Williams, Jerome Parent-Levesque, Derek Nowrouzezahrai, Daniele Panozzo, Kwang Moo Yi, and Andrea Tagliasacchi. 2020. VoronoiNet : General Functional Approximators with Local Support. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (Seattle, WA, USA). IEEE, 1069–1073. doi:10.1109/CVPRW50498.2020.00140
- Qi Wu, Janick Martinez Esturo, Ashkan Mirzaei, Nicolas Moenne-Loccoz, and Zan Gojcic. 2024. 3DGUT: Enabling Distorted Cameras and Secondary Rays in Gaussian Splatting. *arXiv preprint arXiv:2412.12507* (2024).
- Jiawei Yang, Marco Pavone, and Yue Wang. 2023. FreeNeRF: Improving Few-shot Neural Rendering with Free Frequency Regularization. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
- Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. 2021. Volume rendering of neural implicit surfaces. (2021).
- L. Yariv, P. Hedman, C. Reiser, D. Verbin, P. P. Srinivasan, R. Szeliski, J. T. Barron, and B. Mildenhall. 2023. BakedSDF: Meshing Neural SDFs for Real-Time View Synthesis. *ACM SIGGRAPH Conference Papers* (2023).
- Mulin Yu, Tao Lu, Linning Xu, Lihan Jiang, Yuanbo Xiangli, and Bo Dai. 2024b. GSDF: 3DGS Meets SDF for Improved Rendering and Reconstruction. In *NeurIPS*.
- Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. 2024a. Mip-Splatting: Alias-free 3D Gaussian Splatting. In *CVPR*. 19447–19456.
- Zehao Yu, Torsten Sattler, and Andreas Geiger. 2024. Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes. *ACM Transactions on Graphics* (2024).
- Zehao Yu, Torsten Sattler, and Andreas Geiger. 2024. Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes. *ACM TOG* (2024).
- Baowen Zhang, Chuan Fang, Rakesh Shrestha, Yixun Liang, Xiaoxiao Long, and Ping Tan. 2024. RaDe-GS: Rasterizing Depth in Gaussian Splatting. *arXiv preprint arXiv:2406.01467* (2024).
- Wenyuan Zhang, Yu-Shen Liu, and Zhizhong Han. 2024b. Neural Signed Distance Function Inference through Splatting 3D Gaussians Pulled on Zero-Level Set. In *Advances in Neural Information Processing Systems*.
- Ziyu Zhang, Binbin Huang, Hanqing Jiang, Liyang Zhou, Xiaojun Xiang, and Shunhan Shen. 2024a. Quadratic Gaussian Splatting for Efficient and Detailed Surface Reconstruction. *arXiv preprint arXiv:2411.16392* (2024).

Supplementary Material

A Preliminaries

In this supplementary material, we start by briefly reviewing the key concepts that form the foundation of our approach: 3D Gaussian Splatting and Delaunay tetrahedralization.

A.1 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [Kerbl et al. 2023] has emerged as a powerful technique for novel view synthesis, offering high-quality rendering with real-time performance. In this representation, a 3D scene is modeled as a collection of 3D Gaussians, each defined by its position $\mu \in \mathbb{R}^3$, covariance matrix $\Sigma \in \mathbb{R}^{3 \times 3}$ (typically parameterized by a scaling vector $s \in \mathbb{R}^3$ and a rotation matrix $R \in \mathbb{R}^{3 \times 3}$ encoded using a quaternion $q \in \mathbb{R}^4$), opacity $\alpha \in (0, 1)$, and appearance attributes (such as spherical harmonics coefficients $c \in \mathbb{R}^5$ for

view-dependent color). During rendering, these 3D Gaussians are projected onto the image plane as 2D Gaussians and composited in a front-to-back order to produce the final image.

A.2 Delaunay Triangulation

Given a set of 3D points, the Delaunay triangulation divides their convex hull into tetrahedra such that no point lies inside the circumsphere of any tetrahedron. This structure has been extensively studied due to its desirable theoretical and practical properties; see [Aurenhammer 1991] for more details.

In MILo, we rely on the Delaunay triangulation as it naturally adapts to the local density of the input point cloud. Specifically, it creates smaller tetrahedra in regions with higher point density and larger ones in sparser regions. This adaptive nature makes it more efficient than uniform grid-based approaches for extracting surfaces from non-uniform point distributions, such as those formed by optimized 3D Gaussians. Furthermore, the regularity of the Delaunay structure facilitates GPU-accelerated mesh extraction via the marching tetrahedra algorithm [Doi and Koide 1991], which is the cornerstone of our differentiable mesh extraction process.

B Experiments

B.1 Implementation Details

Optimization. For optimization, we use the Adam optimizer with the same learning rates for the Gaussian parameters as [Zhang et al. 2024]. We use a learning rate of 0.025 for the SDF values. We densify Gaussians for 3,000 iterations using aggressive densification [Fang and Wang 2024b], during which we only rely on our photometric loss. We introduce the volumetric rendering regularization loss $\mathcal{L}_N$ after densification, at iteration 3,000. We let Gaussians populate the scene and refine their parameters for 5,000 additional iterations, during which we rely on our volumetric rendering loss $\mathcal{L}_{vol}$. After this point we stop the densification and pruning procedures. Thus, our method does not introduce additional sensitivity to initialization beyond what is already present in standard 3DGS pipelines.

We introduce our mesh extraction pipeline at iteration 8,000. We extract a mesh at every iteration and apply our full loss $\mathcal{L}$ to our representation. The mesh-in-the-loop optimization continues for an additional 10,000 iterations, for a total of 18,000 iterations. For our base model, we prune Gaussians at iteration 8,000 with importance-weighted sampling [Fang and Wang 2024b] to maintain only the most important Gaussians. For the dense model, we maintain a larger set of Gaussians but compute the importance scores at iteration 8,000 and sample Delaunay sites only from the most important Gaussians. Our weight parameters are $\lambda_{RGB} = 0.2$, $\lambda_N = \lambda_{MD} = \lambda_{MN} = 0.05$, and $\lambda_{erosion} = \lambda_{interior} = 0.005$.

Delaunay triangulation. For the Delaunay triangulation, we use the CGAL library’s 3D Delaunay triangulation implementation, which provides robust and efficient computation of the tetrahedralization. Although Delaunay triangulation is inherently non-differentiable, this does not pose a problem in our setting, as gradients propagate from the mesh back to the Gaussians via the Gaussian pivots. We found that updating the Delaunay triangulation at every iteration is not necessary for stable optimization; We therefore only update the Delaunay triangulation every 500 iterations.

Table 6. **Quantitative results for novel view synthesis on MipNeRF 360 [Barron et al. 2022]**. We report PSNR, SSIM, and LPIPS. Our method maintains competitive rendering quality while significantly improving surface reconstruction.

	Indoor Scenes			Outdoor Scenes		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
3DGS	30.41	0.920	0.189	24.64	0.731	0.234
Mip-Splatting	30.90	0.921	0.194	24.65	0.729	0.245
BakedSDF	27.06	0.836	0.258	22.47	0.585	0.349
SuGaR	29.43	0.906	0.225	22.93	0.629	0.356
2DGS	30.40	0.916	0.195	24.34	0.717	0.246
GOF	30.79	0.924	0.184	24.82	0.750	0.202
RaDe-GS	30.74	0.928	0.165	25.17	0.764	0.199
Ours (base)	29.96	0.920	0.191	24.47	0.718	0.290
Ours (dense)	30.76	0.934	0.155	24.81	0.744	0.229

SDF normalization. In practice, for more stable optimization, we optimize Truncated SDF values normalized to be in the range $[-1, 1]$ using a tanh function. When introducing our Mesh-in-the-Loop optimization, the initial SDF values are computed using a custom, scalable depth-fusion algorithm that operates directly on our Delaunay sites rather than on a regular grid: For each Delaunay site, we compute the signed difference between the depth map rendered from the Gaussians and the depth of the site, and fuse these distances by averaging across all views. This initialization provides

a good starting point for the SDF optimization; While being over-smoothed and missing many fine details, the initial mesh is already a reasonable approximation of the surface encoded by the Gaussians.

Mesh rendering. For differentiable mesh rendering, we use nvd-iffrastr [Laine et al. 2020], which provides efficient GPU-accelerated rasterization with gradient support. We render the mesh at the same resolution as the Gaussian rendering, using the same camera parameters. To smooth out discontinuities and ensure the propagation of gradients between neighbor pixels, we apply antialiasing smoothing to the depth maps rendered from the mesh.

B.2 Mesh-Based Novel View Synthesis

We showcase a qualitative comparison of our mesh-based novel view synthesis evaluation in Fig. 10.

B.3 Novel View Synthesis

While our primary focus is on surface reconstruction, we also evaluate the novel view synthesis quality of our optimized Gaussians. Table 6 presents the PSNR, SSIM, and LPIPS metrics on the MipNeRF 360 [Barron et al. 2022] dataset. Our method maintains competitive rendering quality compared to previous approaches, demonstrating that our mesh-in-the-loop optimization does not compromise the visual fidelity of the Gaussian representation.

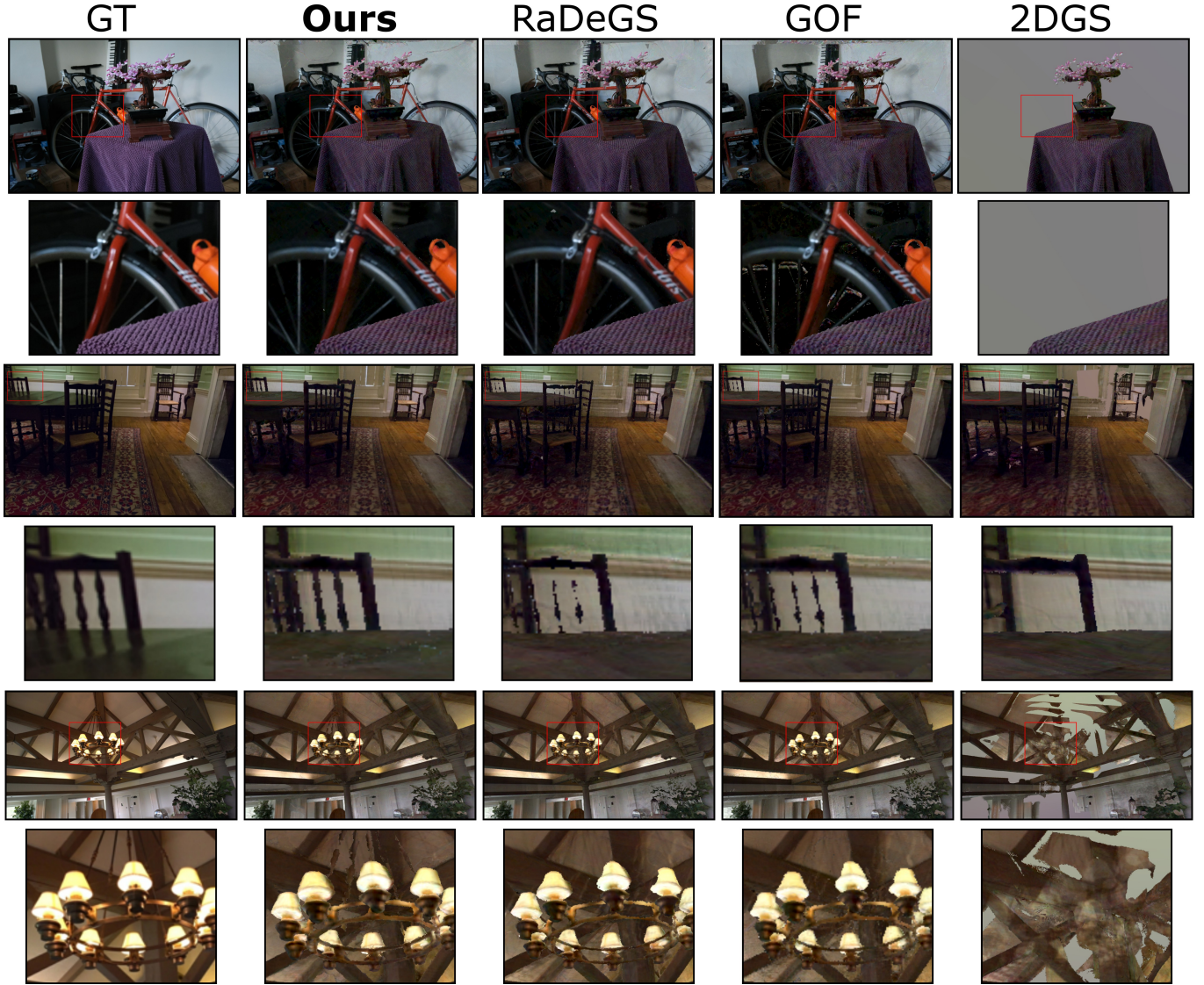


Fig. 10. Our mesh-based evaluation serves as a proxy for quantitatively assessing geometric artifacts, mesh completeness, and background reconstruction. MLo demonstrates superior performance in capturing fine details—such as the intricate chandelier stem and chair spindles—and consistently recovers high-quality background geometry.