Slender3D: Curve-Guided Multi-View Reconstruction of Slender Structures

Suqin Wang[1], Zeyi Wang[1], Min Shi[1*], Zhaoxin Li[2,3], Qi Wang[1], Xiujuan Chai[2,3], Dengming Zhu[4,5]
[1]North China Electric Power University, China
[2]Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[3]Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing 100081, China
[4]Institute of Computing Technology, Chinese Academy of Sciences, China
[5]Taicang–ZK Institute of Information and Technology, China
shi_min@ncepu.edu.cn

Abstract

Although geometric reconstruction of general objects from images has made remarkable progress in recent years, slender structures remain largely underexplored, despite their critical importance in engineering, biomedical, and agricultural applications. To bridge this gap, we propose a dedicated 2DGS-based geometric reconstruction framework tailored for slender structures, achieving accurate and faithful geometry recovery. Our method first addresses the challenge that most slender objects are texture-less, which hinders reliable feature matching and pose estimation in traditional SfM pipelines. By leveraging the curve-like nature of slender structures, we perform a curve-guided SfM process that provides robust camera poses and accurate 3D curve initialization for Gaussian primitives. To ensure SfM reliability, we introduce a high-precision mask extraction strategy that integrates geometric priors with a segmentation network, effectively handling self-occlusion and thin geometry. Furthermore, to enhance fine geometric recovery, we incorporate a differentiable Poisson reconstruction module to extract an initial mesh during training, which is then refined via image-space iterative optimization using differentiable mesh rasterization. In contrast to conventional approaches that rely on differentiable Gaussian rasterization followed by TSDF-based mesh extraction, our method avoids the additional geometric errors and artifacts introduced during the intermediate TSDF conversion, thereby improving the overall reconstruction quality. Comprehensive experiments on both synthetic and real-world datasets validate that our method achieves superior reconstruction quality compared to state-of-the-art approaches.

1 Introduction

Reconstructing 3D geometry from multi-view RGB images is a fundamental task in computer vision and graphics, with broad applications in virtual reality (Kamran-Pishhesari, Moniri-Morad, and Sattarvand 2024), video games (Kargas, Loumos, and Varoutas 2019), autonomous driving (Liu et al. 2023), and robotics (Lee et al. 2022). While existing methods have achieved remarkable progress in general scene reconstruction, accurately recovering slender structures remains a persistent challenge. Such objects are often defined by thin geometry, weak textures, ambiguous image boundaries between the object and the background, making conventional multi-view stereo (MVS) (Schönberger and Frahm 2016) approaches prone to incomplete structure, fragmented geometry, and reduced accuracy.

Recently, neural radiance fields (NeRFs) (Mildenhall et al. 2021) and neural implicit surfaces (neural SDFs) (Wang et al. 2021) have advanced reconstruction quality by leveraging continuous implicit representations and geometric constraints. However, these methods typically require dense ray sampling, resulting in high computational overhead and long optimization times. In contrast, 3D Gaussian Splatting (3DGS) (Kerbl et al. 2023) introduces a differentiable rasterization mechanism that enables efficient rendering and fast optimization. Nevertheless, due to the limited representational capacity of 3D Gaussian primitives, its geometric fidelity remains suboptimal. Further, 2DGS (Huang et al. 2024) extended 3DGS by projecting 3D Gaussian primitives into the image plane as 2D primitives, enhancing surface compactness and clarity, especially in fine-detailed regions. Despite these advances, existing Gaussian Splatting-based methods still struggle with slender reconstruction, often leading to floating artifacts, discontinuities, or missing geometry. These problems primarily arise from the absence of structural modeling and geometric constraints specifically designed for the unique characteristics of slender shapes.

To address these challenges, we propose *Slender3D*, a structure-aware optimization framework based on 2DGS, specialized in the high-fidelity reconstruction of slender objects. First, we leverage the inherently curve-like nature of slender structures and integrate a curve-based SfM method into our pipeline. This module provides reliable camera poses and 3D curve initialization, serving as geometric anchors for the Gaussian primitives. To enhance the robustness of both the SfM stage and subsequent optimization, we introduce a structure-aware foreground extraction module that combines a 2D segmentation network with 3D geometric priors. This module produces high-precision masks capable of handling thin geometry and self-occlusion, thereby reducing background interference and improving multi-view correspondence. Finally, to overcome the representational limitations of Gaussian primitives in modeling fine structures, we introduce a differentiable surface optimization strategy. During training, we perform Poisson surface reconstruction to generate an initial mesh from the current set of Gaussians, which is then refined via image-space optimization using dif-

ferentiable mesh rasterization (Laine et al. 2020). Rather than relying on intermediate volumetric fusion (e.g., TSDF extraction) to finally reconstruct meshes from Gaussian primitives, we directly optimize the surface geometry via differentiable mesh rendering, significantly reducing artifacts and yielding cleaner, more accurate reconstructions.

Extensive experiments on both synthetic and real data demonstrate that our method significantly improves geometric accuracy, structural completeness, and surface smoothness in slender structure reconstruction, achieving the state-of-the-art (SOTA) performance.

Our main contributions are summarized as follows:

- We propose Slender3D, a novel optimization framework based on 2D Gaussian Splatting, designed to address the unique challenges of slender object reconstruction, including thin geometry, weak textures, and frequent background confusion.
- We identify that slender objects inherently exhibit curve-like structures, and leverage this insight to guide structural initialization, which improves the robustness and effectiveness of subsequent optimization.
- We propose a surface-based refinement strategy that achieves higher geometric fidelity and fewer artifacts compared to conventional TSDF-based approaches.

2 Related Works

Traditional reconstruction of slender structures Traditional methods for reconstructing slender structures can be broadly categorized into three approaches: scan-based geometric reconstruction (Berger et al. 2014), image-based multi-view reconstruction (Tabb 2013), and depth sensor fusion techniques (Liu et al. 2021). Early scan-based methods employed structured light, LiDAR, or RGB-D cameras to capture dense point clouds, which were then processed using surface reconstruction algorithms such as Poisson reconstruction. However, these approaches are fundamentally constrained by sensor resolution and the scale of the target object, often leading to incomplete or inaccurate reconstructions (Fan et al. 2016; Wu et al. 2014). Image-based multi-view methods (Hsiao, Huang, and Chu 2018; Li et al. 2018) rely on geometric cues such as contours, edges, or curves across views to perform structure recovery. Despite their potential, these methods are sensitive to errors from camera pose inaccuracies, self-occlusion, and insufficient texture, which significantly limits their robustness in complex environments. Depth-based methods like CurveFusion (Liu et al. 2021) avoid correspondence matching by directly extracting curve skeletons from depth data for integration. Nonetheless, their accuracy is heavily dependent on the depth sensor's resolution and noise characteristics, restricting their applicability in practice. In summary, while traditional methods provide effective solutions under controlled or simple conditions, they remain inadequate for accurately reconstructing slender structures in real-world scenarios with complex topology, weak textures, and severe occlusion.

Gaussian Splatting for Slender Structures In recent years, Gaussian Splatting (Kerbl et al. 2023) has been widely adopted for novel view synthesis and 3D reconstruction tasks. 3D Gaussian Splatting employs learnable Gaussian primitives to efficiently model scenes, but its reliance solely on image supervision and lack of explicit geometric constraints often leads to disorganized Gaussian distributions, particularly in geometrically complex regions such as slender structures. To improve geometric accuracy, methods like Dn-splatter (Turkulainen et al. 2025) and GSDF (Yu et al. 2024) incorporate SDFs or explicit geometric priors as supervision. However, these strategies mainly focus on global geometric consistency and struggle to capture local structural details.

2DGS (Huang et al. 2024) simplifies 3D Gaussians into planar 2D Gaussians, enhancing surface modeling performance, especially in textureless areas. Nonetheless, its lack of structural constraints on Gaussian positions and normals leads to reconstruction errors in thin or blurred-edge regions. Therefore, a key challenge in this field is how to introduce explicit structural guidance into the optimization of Gaussian primitives to better accommodate the geometric characteristics of slender shapes.

3 Preliminary

The key innovation of 2D Gaussian Splatting (2DGS) (Huang et al. 2024) is to represent 3D surfaces using planar 2D Gaussians (elliptical disks), simplifying scene modeling while enabling accurate geometry reconstruction without mesh refinement. Each disk lies on a local tangent plane and is parameterized by a center $\mathbf{p}_k$, two orthogonal tangent vectors $\mathbf{t}_u$, $\mathbf{t}_v$, and scaling factors (s_u, s_v) that control the spread. The normal is given by $\mathbf{t}_w = \mathbf{t}_u \times \mathbf{t}_v$, forming a rotation matrix $\mathbf{R} = [\mathbf{t}_u, \mathbf{t}_v, \mathbf{t}_w]$, and the scale matrix is $\mathbf{S} = \mathrm{diag}(s_u, s_v, 0)$. A 2D Gaussian can thus be expressed as:

$$\mathrm{P}(u, v) = \mathbf{p}_k + s_u \mathbf{t}_u u + s_v \mathbf{t}_v v = \mathbf{H}(u, v, 1, 1)^\top \quad (1)$$

where $\mathbf{H} \in \mathbb{R}^{4\times4}$ denotes the homogeneous transformation matrix representing the 2D Gaussian geometry:

$$\mathbf{H} = \begin{bmatrix} s_u \mathbf{t}_u & s_v \mathbf{t}_v & \mathbf{0} & \mathbf{p}_k \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{RS} & \mathbf{p}_k \\ \mathbf{0} & 1 \end{bmatrix} \quad (2)$$

In the Gaussian's tangent frame (u, v), the Gaussian value at position $\mathbf{u} = (u, v)$ is:

$$\mathcal{G}(\mathbf{u}) = \exp\left(-\frac{u^2 + v^2}{2}\right). \quad (3)$$

Each Gaussian is projected to the screen via the world-to-camera matrix $\mathbf{W}$:

$$\mathbf{x} = (x, y, z, z)^\top = \mathbf{WH}(u, v, 1, 1)^\top \quad (4)$$

To compute the Gaussian-ray intersection in (u, v) space, two planes are defined per pixel: $\mathbf{h}_x = (-1, 0, 0, x)$ and $\mathbf{h}_y = (0, -1, 0, y)$. These are transformed as :

$$\mathbf{h}_u = (\mathbf{WH})^\top \mathrm{h}_x, \quad \mathbf{h}_v = (\mathbf{WH})^\top \mathrm{h}_y. \quad (5)$$

The intersection point $(u(\mathbf{x}), v(\mathbf{x}))$ on the 2D Gaussian splat can be computed by applying a homography transformation, as follows:

$$u(\mathbf{x}) = \frac{\mathbf{h}_u^2 \mathbf{h}_v^4 - \mathbf{h}_u^4 \mathbf{h}_v^2}{\mathbf{h}_u^1 \mathbf{h}_v^2 - \mathbf{h}_u^2 \mathbf{h}_v^1}, \quad v(\mathbf{x}) = \frac{\mathbf{h}_u^4 \mathbf{h}_v^1 - \mathbf{h}_u^1 \mathbf{h}_v^4}{\mathbf{h}_u^1 \mathbf{h}_v^2 - \mathbf{h}_u^2 \mathbf{h}_v^1} \quad (6)$$

where $\mathbf{h}_u^i$ and $\mathbf{h}_v^i$ are the components of the transformed planes in the Gaussian's tangent frame.

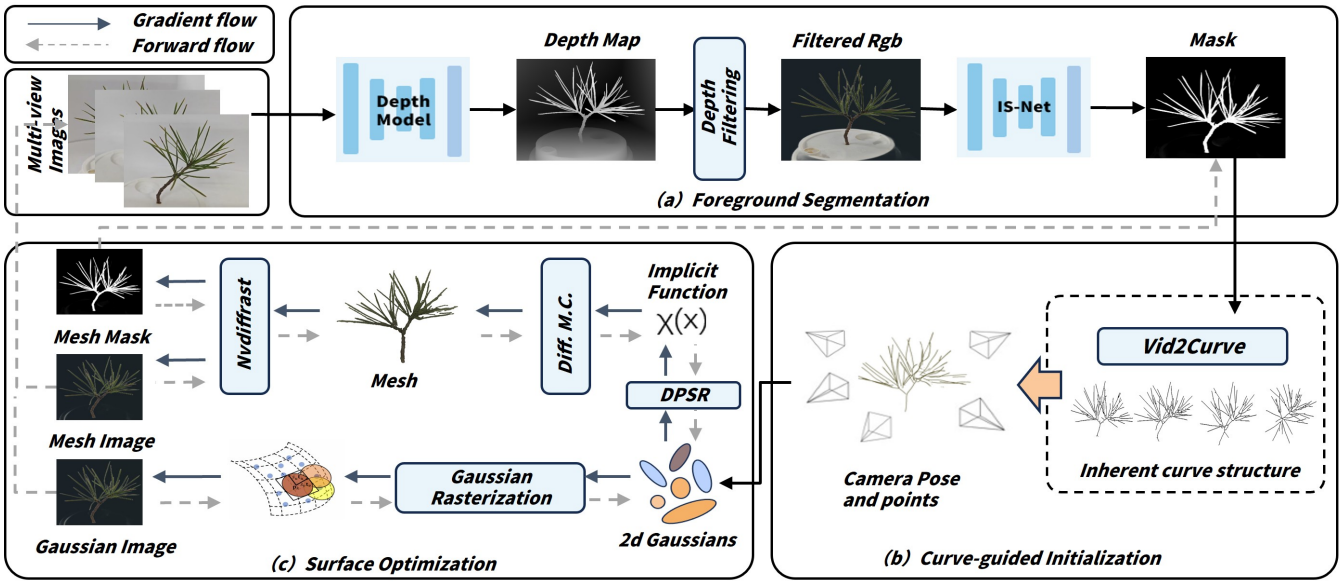


Figure 1: Overview of Slender3D. Our pipeline consists of three stages: (1) Multi-view images are first processed by a depth estimation network and IS-Net to obtain foreground-filtered depth maps and fine-grained segmentation masks. (2) We then perform curve-guided joint pose optimization to extract geometric priors, yielding accurate initial point clouds and camera poses. (3) Finally, we reconstruct an initial mesh via differentiable Poisson sampling and Marching Cubes, and refine it through differentiable rasterization along with Gaussian primitive optimization. This compact pipeline leverages structured priors and geometry-aware optimization to improve 3D reconstruction for textureless, slender objects.

4 Method

Given multi-view images $\mathbf{I} = \{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_n\}$, we aim to accurately reconstruct slender objects. To this end, we first design a two-stage foreground extraction module that integrates geometric priors with a segmentation network to produce high-quality masks $\boldsymbol{M} = \{\boldsymbol{M}_1, \boldsymbol{M}_2, \ldots, \boldsymbol{M}_n\}$, which improve both the SfM stage and the subsequent optimization process (Sec. 4.1). Next, we incorporate a curve-guided SfM approach that jointly estimates camera poses and sparse 3D curves through cross-frame curve matching, providing reliable geometric anchors for initializing Gaussian primitives (Sec. 4.2). Finally, we propose a mesh-based optimization strategy that generates an initial surface via differentiable Poisson reconstruction (Peng et al. 2021) and refine it using differentiable mesh rasterization (Laine et al. 2020), which in turn updates the positions and normals of the Gaussian primitives during training (Sec. 4.3). The overall pipeline of our method is illustrated in Fig. 1.

4.1 Foreground Segmentation

Accurate foreground segmentation is essential for 3D reconstruction of slender structures. Without explicit masks, background clutter with a similar appearance can lead to reconstruction errors such as shape distortion, fragmentation, or complete failure.

To obtain these masks, a straightforward approach is to utilize an existing SOTA 2D image segmentation network for prediction. Therefore, we first experiment with Dichotomous Image Segmentation (DIS) (Qin et al. 2022), a high-precision binary segmentation method based on the IS-Net (Qin et al. 2022) architecture. DIS has demonstrated strong performance in extracting fine structures such as cables and fences, particularly on well-annotated datasets like DIS5K (Qin et al. 2022).

However, when applied directly to real-world images containing slender objects, DIS often fails to produce reliable masks. As illustrated in Fig. 2(a), the segmentation result may contain missing parts or include erroneous background regions when the background shares similar textures with the foreground. These defects significantly limit the quality of reconstruction results.

To address this limitation, we introduce a depth-guided prefiltering step using Depth-Anything V2 (Yang et al. 2024), a monocular depth estimation network. Although originally designed for depth prediction, **its predicted depth maps implicitly separate foreground and background regions through relative depth cues**. We repurpose these cues to guide the first stage coarse segmentation.

Given a predicted depth map $\mathbf{D}$, we define a global threshold τ to filter out distant background pixels:

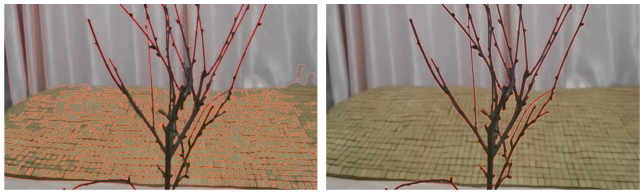
$$\mathbf{M}_{\text{depth}}(x, y) = \begin{cases} 1 & \text{if } \mathbf{D}(x, y) < \tau, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

τ is a depth value satisfying:

$$\frac{|\{\mathbf{D}(\cdot) \mid \mathbf{D}(\cdot) \leq \tau\}|}{N} = 0.85 \tag{8}$$

where $\mathbf{D}(\cdot)$ are all valid depths, N is the number of $\mathbf{D}(\cdot)$.

We apply the binary mask to the original image to obtain a filtered version as a coarse segmentation result, formulated



(a) (b)

Figure 2: Comparison of foreground masks. (a) Result of applying DIS directly on raw images. (b) Output of our two-stage pipeline with depth-guided pre-filtering, showing clearer boundaries and less background leakage.

as $\mathbf{I}'(x, y) = \mathbf{I}(x, y) \cdot \mathbf{M}_{\text{depth}}(x, y)$, which is then fed into the DIS network to perform final segmentation with reduced background interference, ultimately producing the desired fine-grained mask $\boldsymbol{M}$.

This two-stage strategy improves both the completeness and boundary accuracy of the masks, thereby providing a more reliable geometric prior for the reconstruction of slender objects (see Fig. 2(b)).

4.2 Curve-guided Initialization

Traditional structure-from-motion (SfM) pipelines such as COLMAP (Schönberger and Frahm 2016) are known to perform well on textured, rigid scenes. However, when applied to slender and textureless objects, these methods often encounter difficulties. The absence of distinctive appearance features leads to unreliable keypoint detection and sparse or unstable correspondences, resulting in inaccurate camera pose estimation and incomplete 3D reconstructions.

To address this, we observe that **slender objects inherently exhibit a curve-like geometric structure**, which offers a more stable representation than sparse keypoints in low-texture scenarios. Motivated by this insight, we incorporate the Vid2Curve algorithm (Wang et al. 2020) into our reconstruction pipeline as an initialization module. Unlike point-based SfM methods that rely on dense textures, Vid2Curve leverages the geometric continuity of curve structures for alignment and reconstruction. It exploits stable 2D skeletal curves instead of appearance features, making it particularly suitable for slender, low-texture scenarios where traditional methods fail. In our case, the accurate multi-view foreground masks $\boldsymbol{M}$ obtained in Section 4.1 enable the precise extraction of single-pixel-width 2D skeletal curves via skeletonization. We denote the resulting curve set as $\mathbf{C} = \{C_1, C_2, \ldots, C_n\}$, where each C_k represents the skeleton extracted from the k-th view. These curves provide clean and structurally meaningful constraints that support the initialization process.

The process begins by selecting a pair of keyframes with a large parallax. Pixel-wise curve correspondences are estimated using a combination of curve matching and optical flow, followed by a two-view bundle adjustment to jointly estimate the relative poses and triangulate a sparse set of 3D points denoted as $\mathbf{P}$.

To recover a structured 3D curve, a variant of Kruskal's algorithm (Kruskal 1956) is used to form a minimum spanning tree over the 3D points $\mathbf{P}$, resulting in an initial curve graph $\mathcal{C} = (\mathbf{P}, \mathbf{E})$, where $\mathbf{P} = \{\mathbf{p}_i \in \mathbb{R}^3\}$ and $\mathbf{E}$ denote the curve points and their connectivity, respectively.

The estimated $\mathcal{C}$ and camera poses $(\mathbf{R}_k, \mathbf{T}_k)$ are then jointly refined by minimizing a projection-based alignment loss, which encourages consistency between the 3D curve and the 2D skeletons across views:

$$F(\mathbf{R}_k, \mathbf{T}_k; \mathcal{C}) = \sum_k \text{dist}^2(c'_k, c_k), \tag{9}$$

where $c'_k = \pi(\mathbf{R}_k, \mathbf{T}_k; \mathcal{C})$ denotes the projection of the 3D curve onto view $\mathbf{I}_k$, and the distance is defined as:

$$\text{dist}^2(c'_k, c_k) = \int_{p \in c'_k} \min_{q \in c_k} |p - q|_2^2. \tag{10}$$

In this expression, p and q denote points on the projected 3D curve c'_k and the 2D skeleton curve c_k in image $\mathbf{I}_k$, respectively.

The remaining frames are then incrementally integrated through an alternating optimization process, jointly refining both the global camera poses and the 3D curve structure.

This initialization yields a globally consistent and structurally faithful 3D curve representation $\mathcal{C}$ along with reliable camera pose estimates, which naturally provide geometric and volumetric guidance for subsequent modeling stages.

4.3 Surface Optimization

Most 3DGS-based methods rely on TSDF fusion for surface extraction by aggregating depth maps rendered from Gaussians. While effective for general scenes, this approach struggles with slender objects due to its dependence on accurate depth. Slender structures typically occupy small image areas and contain sparse, unevenly distributed Gaussian primitives, especially around fine regions. As a result, the rendered depth maps are often noisy or incomplete, leading to noticeable discontinuities and floating artifacts in the TSDF-fused surfaces.

To fundamentally address the limitations of TSDF fusion, we propose a differentiable reconstruction pipeline that **bypasses TSDF entirely and directly optimizes mesh geometry**. We first apply the differentiable Poisson solver to estimate the indicator function $\chi(x)$ from the current Gaussian-based normal field. Then we extract the initial mesh utilizing differentiable Marching Cubes (Wei et al. 2025). This mesh is then projected back to the image plane via Nvdiffrec (Laine et al. 2020), and optimized by RGB and mask supervision. The above pipeline forms a positive feedback loop: image-space supervision improves the mesh, which enables more accurate gradients to flow back to the Gaussian primitives. The refined Gaussians lead to better mesh extraction via Poisson and Marching Cubes, which in turn simplifies and enhances downstream optimization through Nvdiffrec. Through this formulation, our approach achieves significantly improved surface continuity and fine-grained geometric detail, without relying on precise depth rendering or TSDF fusion.

Method	Tree					VMR				
	Acc↓	Comp↓	Prec↑	Recall↑	F1↑	Acc↓	Comp↓	Prec↑	Recall↑	F1↑
GOF	0.0507	0.0207	0.3607	0.8363	0.4989	0.0181	0.0636	0.2585	0.4721	0.3211
SuGaR	0.0930	0.0488	0.5195	0.8611	0.6362	0.0452	0.0110	0.4794	0.7361	0.5661
2DGS	0.0386	**0.0188**	0.4935	0.7760	0.5969	0.0156	0.0781	0.5219	0.6871	0.5932
PGSR	0.0374	0.0214	0.5447	0.8521	0.6549	0.0146	0.0191	0.4558	0.6026	0.5123
Ours	**0.0271**	0.0211	**0.6729**	**0.8889**	**0.7573**	**0.0113**	**0.0095**	**0.7011**	**0.7914**	**0.7432**

Table 1: Performance comparison across different datasets and scenes. For Acc and Comp, lower is better (↓). For Prec, Recall and F1, higher is better (↑). The best results in each column are highlighted in **bold**.

Let $\boldsymbol{I}$ denote the input ground-truth RGB image, and $\hat{\boldsymbol{I}}$ the predicted RGB image rendered from the mesh via differentiable rasterization (Laine et al. 2020). Similarly, Let $\boldsymbol{M}$ denote the set of refined binary masks obtained in Section 4.1, and $\hat{\boldsymbol{M}}$ the predicted mask obtained by rasterizing the mesh. Then, the mesh-related loss can be formulated as:

$$\mathcal{L}_{\text{mesh}} = \mathcal{L}_{\text{rgb}} + \mathcal{L}_{\text{mask}} = \|\boldsymbol{I} - \hat{\boldsymbol{I}}\|_1 + \text{BCE}(\boldsymbol{M}, \hat{\boldsymbol{M}}). \quad (11)$$

where $\text{BCE}(\cdot, \cdot)$ denotes the binary cross-entropy loss, which measures the pixel-wise difference between the predicted mask $\hat{\boldsymbol{M}}$ and the ground-truth mask $\boldsymbol{M}$.

To better preserve the fine-grained structure of the mesh surface, we incorporate a mesh Laplacian regularization term to penalize local curvature variations. Specifically, we compute a uniformly weighted differential value δ_i for each vertex v_i, representing the difference between v_i and the mean position of its neighboring vertices. The Laplacian regularization term is defined as the mean squared value of δ_i over all mesh vertices, where n denotes the total number of vertices:

$$\mathcal{L}_{\text{lap}} = \frac{1}{n}\sum_{i=0}^{n} \|\delta_i\|^2, \delta_i = v_i - \frac{1}{|N_i|}\sum_{k \in N_i} v_k \quad (12)$$

4.4 Loss Function

Let $\mathcal{L}_{\text{2dgs}}$ denote the photometric supervision term originally introduced in 2DGS, which aims to minimize the discrepancy between the rendered and input images. Our final training loss function $\mathcal{L}$ is thus defined as:

$$\mathcal{L} = \mathcal{L}_{2dgs} + \lambda_{rgb}\mathcal{L}_{rgb} + \lambda_{mask}\mathcal{L}_{mask} + \lambda_{lap}\mathcal{L}_{lap} \quad (13)$$

where $\mathcal{L}_{rgb}$ and $\mathcal{L}_{mask}$ denote the mesh-related RGB and mask losses.

5 Experiments

5.1 Experimental Setup

Dataset and metrics. We evaluate our method on both synthetic and real-world datasets. The synthetic set includes six virtual models: three tree-like structures and three vascular geometries from the Vascular Model Repository (VMR) (Wilson, Ortiz, and Johnson 2013; Walker et al. 2025), all rendered using a custom Blender pipeline. Real-world data is captured with consumer devices and covers four common slender object types.However,Ground-truth geometry is not available for the real-world datasets.We report standard metrics—Accuracy (Acc), Completeness (Comp), Precision (Prec), Recall, and F-score (F1) to assess deviation, coverage, and geometric fidelity.overage, and geometric fidelity.

Baselines and implementation. To evaluate our model's capability in reconstructing slender structures, we conduct comprehensive comparisons against several state-of-the-art methods, including 2DGS (Huang et al. 2024), GOF (Yu, Sattler, and Geiger 2024), SuGar (Guédon and Lepetit 2024), and PGSR (Chen et al. 2024). Our implementation is primarily built on the PyTorch framework, and all experiments are carried out on an Nvidia 3090 GPU. Most hyperparameters follow the default settings of 2DGS.We set $\lambda_{\text{rgb}} = 2.0$, $\lambda_{\text{mask}} = 1000$, and $\lambda_{\text{lap}} = 1000$ in our experiments. Our pipeline is trained for 30k iterations. To ensure the positions and normals of 2D Gaussian primitives better conform to object surfaces, we introduce a differentiable Poisson Solver (DPSR) module and differentiable Marching Cubes at the 15k iteration mark to extract mesh geometry from the Gaussian primitives. The DPSR module requires an average of 19 GB GPU memory and approximately 64.5 minutes of runtime, which is comparable to existing methods while significantly improving geometric detail. For mesh rasterization, we adopt the Nvdiffrast library and utilize an MLP network to store vertex color information. The final supervision combines dual rendering losses from both Gaussian and mesh-based renderings.

5.2 Quantitative Comparisons

To evaluate the quality of the reconstructed mesh, we first conducted quantitative comparisons on synthetic datasets. Given the geometric properties of slender structures, we adopted different distance thresholds when computing evaluation metrics: 3 cm for tree-like datasets and 1.2 cm for the medical VMR dataset. Since the COLMAP (Schönberger and Frahm 2016) results are suboptimal on some datasets and could negatively affect evaluation, all methods listed in Table 1 employ curve-based SfM for camera pose estimation.

As shown in Table 1, the experimental results clearly demonstrate that our proposed method achieves consistently superior performance across various evaluation metrics and diverse scenarios. It is noteworthy that some baseline methods attain relatively high Recall but suffer from significantly

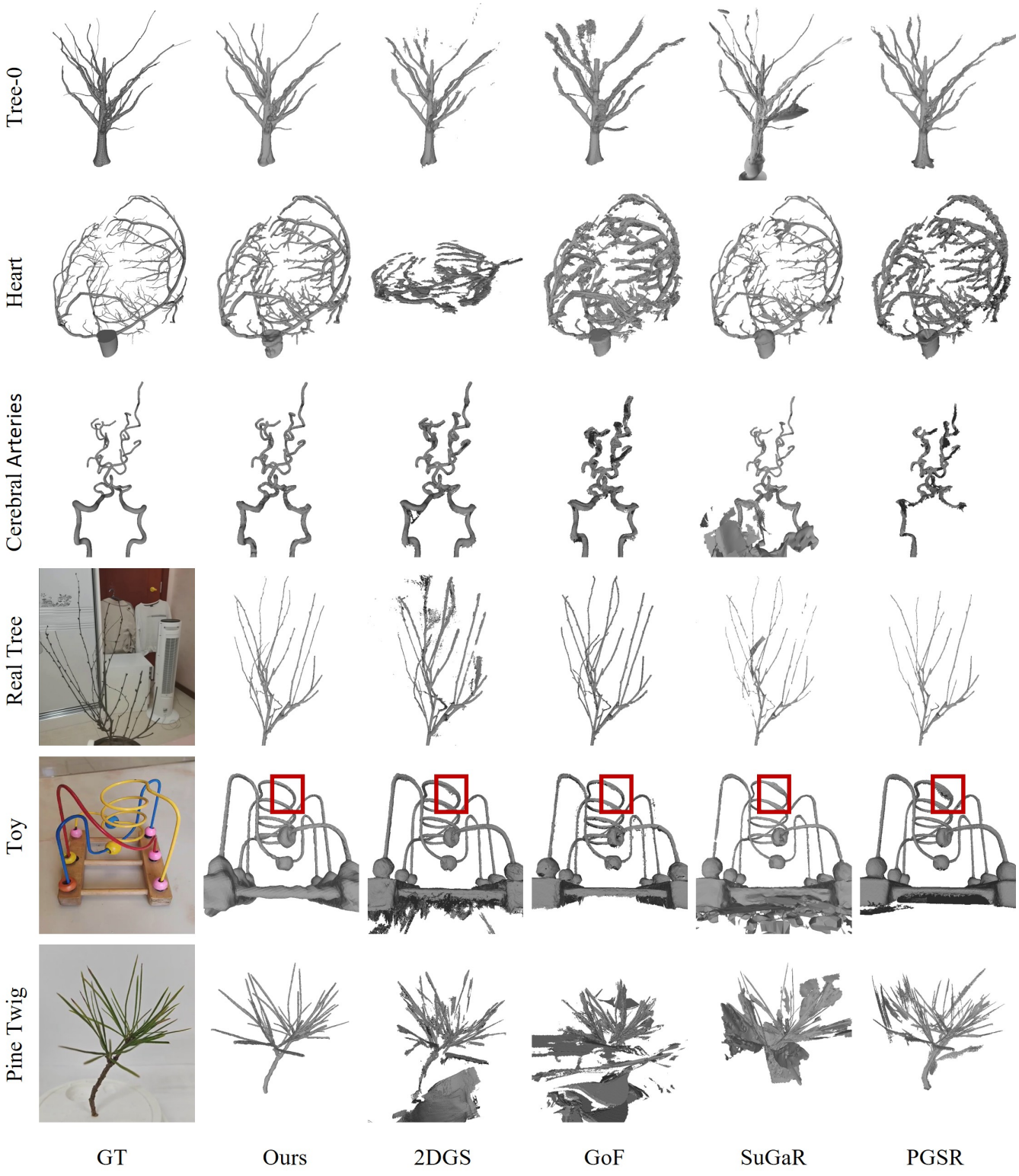


Figure 3: Qualitative comparison of reconstruction results. Our method achieves cleaner and more complete reconstructions on both synthetic and real-world datasets, including Tree-0, Heart, Cerebral Arteries, Real Tree, Toy, and Ping Twig.

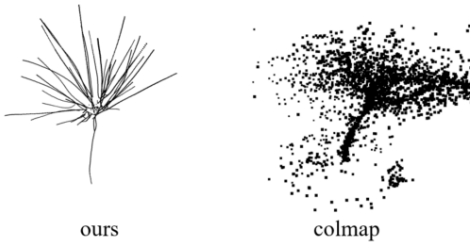


Figure 4: Comparison between our point cloud and the one generated by COLMAP. Our method produces cleaner and more structured results.

lower Precision. This indicates that although these methods tend to broadly cover the target regions, they struggle with accurate localization and boundary delineation, resulting in a higher false positive rate. Such a tendency towards "Recall over Precision" is often observed in models lacking effective structural awareness of slender objects, leading to suboptimal performance in fine-grained reconstruction.

In contrast, our method achieves strong performance in both Precision and Recall, demonstrating its ability not only to reconstruct the full extent of slender structures but also to delineate them accurately. This balanced performance can be attributed to our task-specific design and the incorporation of geometric constraints tailored to thin, elongated objects, enabling more accurate and reliable 3D reconstructions.

5.3 Qualitative Comparisons

We qualitatively compare our method with several state-of-the-art approaches on both synthetic and real-world datasets, including the synthetic tree (Tree-0), vascular structures (Heart and Cerebral Arteries), and three real examples: Real tree, Toy, and Ping Twig (Fig. 3). Our method produces cleaner reconstructions with minimal noise and achieves superior completeness. Notably, it demonstrates strong structural awareness by faithfully preserving the thin and elongated features of slender objects.

5.4 Ablation Studies

Efficacy of the Curve-Guided SfM and DPSR We perform ablation studies to assess the contributions of curve-guided SfM and the differentiable Poisson solver (DPSR) for slender structure reconstruction. As shown in Fig. 4, curve-guided SfM reduces noise and floating points, producing point clouds better aligned with slender contours, while DPSR enforces surface continuity and improves accuracy. Table 2 demonstrates that both modules enhance performance, and their combination achieves the best results, highlighting their complementary roles.

Ablation Study on Mask Extraction Strategy. We compare two settings to assess our mask extraction strategy: (1) using masks from a standard segmentation network without geometric priors, and (2) using our full two-stage pipeline with geometric cues. As shown in Figure 5, the baseline suffers from background noise and missing thin structures,

Method	Acc↓	Comp↓	Prec↑	Recall↑	F1↑
w/o curve sfm	0.0775	0.0650	0.2793	0.3487	0.3018
w/o dpsr	0.0372	0.0265	0.5418	0.8528	0.6530
Full Model	**0.0271**	**0.0211**	**0.6729**	**0.8889**	**0.7573**

Table 2: Quantitative comparison on the synthetic tree dataset. We conduct ablation to assess the effect of curve-guided SfM and the DPSR module. Curve-guided SfM improves pose estimation for slender, low-texture branches, while DPSR enhances geometric consistency and completeness. For Acc and Comp, lower is better (↓); for Prec, Recall, and F1, higher is better (↑).

Figure 5: Comparison of final reconstructed meshes. (a) Using single-stage masks without geometric priors results in background noise and missing structures. (b) Using our two-stage mask extraction with geometric priors improves foreground completeness and detail, confirming the benefit of geometric guidance.

degrading reconstruction quality. In contrast, our method effectively suppresses clutter and preserves slender details, highlighting the importance of geometric priors for reliable SfM and reconstruction.

Conclusion

We present **Slender-3D**, an efficient framework for reconstructing slender structures. By exploiting their inherent curve-like geometry, we introduce a curve-guided Structure-from-Motion (SfM) pipeline that enhances feature matching and camera pose estimation under textureless conditions, enabling accurate initialization of Gaussian primitives. To further improve segmentation in cluttered scenes, we incorporate a geometry-aware mask extraction strategy. During training, Slender-3D integrates differentiable Poisson reconstruction with image-space mesh rasterization to achieve fine-grained, globally consistent optimization. Extensive experiments demonstrate that our method delivers state-of-the-art performance in terms of geometric completeness and detail preservation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants 62172392, 61972379), the Innovation Program of the Chinese Academy of Agricultural Sciences (Grants CAAS-CSSAE-202401, CAAS-ASTIP-2025-AII), the Major Industrial Tackling Project of Shandong Province (Film and Television Production Industrialization Technology Platform), and the Suzhou Science and Technology Program (SYG202327).

References

Berger, M.; Tagliasacchi, A.; Seversky, L.; Alliez, P.; Levine, J.; Sharf, A.; and Silva, C. 2014. State of the art in surface reconstruction from point clouds. *Eurographics 2014-State of the Art Reports*, 1(1): 161–185.

Chen, D.; Li, H.; Ye, W.; Wang, Y.; Xie, W.; Zhai, S.; Wang, N.; Liu, H.; Bao, H.; and Zhang, G. 2024. PGSR: Planar-based Gaussian Splatting for Efficient and High-Fidelity Surface Reconstruction. *arXiv preprint arXiv:2406.06521*.

Fan, X.; Zhang, L.; Brown, B.; and Rusinkiewicz, S. 2016. Automated view and path planning for scalable multi-object 3D scanning. *ACM Transactions on Graphics (TOG)*, 35(6): 1–13.

Guédon, A.; and Lepetit, V. 2024. SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering. *CVPR*.

Hsiao, K.-W.; Huang, J.-B.; and Chu, H.-K. 2018. Multi-view wire art. *ACM Trans. Graph.*, 37(6): 242.

Huang, B.; Yu, Z.; Chen, A.; Geiger, A.; and Gao, S. 2024. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. In *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery.

Kamran-Pishhesari, A.; Moniri-Morad, A.; and Sattarvand, J. 2024. Applications of 3d reconstruction in virtual reality-based teleoperation: A review in the mining industry. *Technologies*, 12(3): 40.

Kargas, A.; Loumos, G.; and Varoutas, D. 2019. Using different ways of 3D reconstruction of historical cities for gaming purposes: The case study of Nafplio. *Heritage*, 2(3).

Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4).

Kruskal, J. B. 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1): 48–50.

Laine, S.; Hellsten, J.; Karras, T.; Seol, Y.; Lehtinen, J.; and Aila, T. 2020. Modular primitives for high-performance differentiable rendering. *ACM Transactions on Graphics (ToG)*, 39(6): 1–14.

Lee, S.; Chen, L.; Wang, J.; Liniger, A.; Kumar, S.; and Yu, F. 2022. Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields. *IEEE Robotics and Automation Letters*, 7(4): 12070–12077.

Li, S.; Yao, Y.; Fang, T.; and Quan, L. 2018. Reconstructing thin structures of manifold surfaces by integrating spatial curves. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2887–2896.

Liu, L.; Chen, N.; Ceylan, D.; Theobalt, C.; Wang, W.; and Mitra, N. J. 2021. CurveFusion: Reconstructing thin structures from RGBD sequences. *arXiv preprint arXiv:2107.05284*.

Liu, Y.; Zhu, K.; Wu, G.; Ren, Y.; Liu, B.; Liu, Y.; and Shan, J. 2023. MV-DeepSDF: Implicit modeling with multi-sweep point clouds for 3D vehicle reconstruction in autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8306–8316.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Peng, S.; Jiang, C.; Liao, Y.; Niemeyer, M.; Pollefeys, M.; and Geiger, A. 2021. Shape as points: A differentiable poisson solver. *Advances in Neural Information Processing Systems*, 34: 13032–13044.

Qin, X.; Dai, H.; Hu, X.; Fan, D.-P.; Shao, L.; and Gool, L. V. 2022. Highly Accurate Dichotomous Image Segmentation. In *ECCV*.

Schönberger, J. L.; and Frahm, J.-M. 2016. Structure-from-Motion Revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.

Tabb, A. 2013. Shape from silhouette probability maps: reconstruction of thin objects in the presence of silhouette extraction and calibration error. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 161–168.

Turkulainen, M.; Ren, X.; Melekhov, I.; Seiskari, O.; Rahtu, E.; and Kannala, J. 2025. Dn-splatter: Depth and normal priors for gaussian splatting and meshing. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2421–2431. IEEE.

Walker, T.; Esposito, S.; Rebain, D.; Vaxman, A.; Onken, A.; Li, C.; and Mac Aodha, O. 2025. CrossSDF: 3D Reconstruction of Thin Structures from Cross-Sections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Wang, P.; Liu, L.; Chen, N.; Chu, H.-K.; Theobalt, C.; and Wang, W. 2020. Vid2Curve: simultaneous camera motion estimation and thin structure reconstruction from an RGB video. *ACM Transactions on Graphics (TOG)*, 39(4): 132–1.

Wang, P.; Liu, L.; Liu, Y.; Theobalt, C.; Komura, T.; and Wang, W. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*.

Wei, X.; Xiang, F.; Bi, S.; Chen, A.; Sunkavalli, K.; Xu, Z.; and Su, H. 2025. Neumanifold: Neural watertight manifold reconstruction with efficient and high-quality rendering support. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 731–741. IEEE.

Wilson, N. M.; Ortiz, A. K.; and Johnson, A. B. 2013. The vascular model repository: a public resource of medical imaging data and blood flow simulation results. *Journal of medical devices*, 7(4): 040923.

Wu, S.; Sun, W.; Long, P.; Huang, H.; Cohen-Or, D.; Gong, M.; Deussen, O.; and Chen, B. 2014. Quality-driven poisson-guided autoscanning. *ACM Trans. Graph.*, 33(6): 203–1.

Yang, L.; Kang, B.; Huang, Z.; Zhao, Z.; Xu, X.; Feng, J.; and Zhao, H. 2024. Depth Anything V2. *arXiv:2406.09414*.

Yu, M.; Lu, T.; Xu, L.; Jiang, L.; Xiangli, Y.; and Dai, B. 2024. Gsdf: 3dgs meets sdf for improved neural rendering and reconstruction. *Advances in Neural Information Processing Systems*, 37: 129507–129530.

Yu, Z.; Sattler, T.; and Geiger, A. 2024. Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes. *ACM Transactions on Graphics*.