2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

ROI-GSurFisher: Next Best View Selection for Active Gaussian Splatting via Fisher Information of ROI-Selected Gaussian Surfels

Wei Wang[1], Wei Ma[1,*], Hongliang Zhang[1], Hongbin Zha[2,3]

***Abstract*— Next Best View (NBV) selection is critical for achieving high-quality 3D reconstruction in unknown environments. This paper presents an active NBV selection approach tailored for Gaussian Splatting (GS), a widely adopted 3D reconstruction technique that has recently gained significant attention and been extended to Simultaneous Localization and Mapping (SLAM) systems. Existing state-of-the-art NBV methods for GS focus on minimizing uncertainties of GS parameters but often fail to prioritize views that improve geometric reconstruction quality. To address this limitation, we propose an active view selection method for GS-based reconstruction, with its core being ROI-GSurFisher. This method calculates Fisher Information on Gaussian surfels selected via a Region of Interest (ROI) mechanism. Both the use of surfels for computation and the ROI constraint enhance ROI-GSurFisher's ability to evaluate geometric information gain. We further introduce a close-front view scoring module that prioritizes viewpoints conducive to high-quality reconstruction. The final NBV is selected by maximizing the combined geometric information gain and close-front score. Experimental results on 3D reconstruction of various objects and scenes demonstrate consistent qualitative and quantitative improvements. Beyond standalone 3D reconstruction, the proposed NBV method can be integrated into SLAM systems to select fewer but more valuable keyframes. Code is available at `https://github.com/WW11111/ROI-GSurFisher`.**

I. INTRODUCTION

3D reconstruction is fundamental to computer vision and robotics, while active Next Best View (NBV) selection underpins high-quality 3D reconstruction in unknown environments. In this paper, we focus on active NBV for Gaussian Splatting (GS)-based 3D reconstruction [1], a method that has gained significant attention in recent years and been extended to various Simultaneous Localization and Mapping (SLAM) systems [2].

For GS-based 3D reconstruction, existing NBV selection methods often aim to directly quantify model uncertainty [3]. One common approach [4], [5] is to introduce an additional uncertainty branch, enabling the model to output both rendered images and uncertainty maps; however, this method is typically sensitive to loss function design and relies on ground-truth images of candidate views. Another strategy

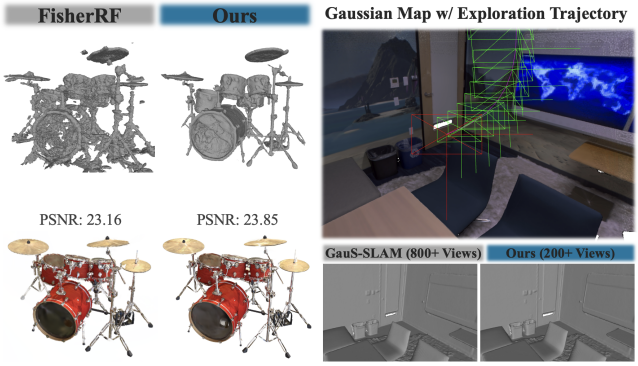

Fig. 1: Our NBV selection method outperforms baseline approaches in both active GS-based 3D reconstruction (left panel) and SLAM-based mapping (right panel). For 3D reconstruction: Both the FisherRF-based method [7] and our NBV-based method stop at the same number of views. The two methods yield comparable rendering results (23.16 vs 23.85 in PSNR), while our method achieves significantly better geometric reconstruction quality. For SLAM experiments: The SLAM system with our NBV method selects only 200+ views as keyframes, yet obtains reconstruction results comparable to those of the baseline GauS-SLAM [2] with 800+ views.

[6] retains the original GS architecture and approximates epistemic uncertainty by training multiple GS models with different random seeds and measuring cross-model variance. While straightforward, these strategies fail to convert uncertainty into optimizable information gain, thereby limiting their applicability in active mapping tasks. To address this gap, FisherRF [7] leverages Fisher Information to quantify the parameter uncertainty of GS models, and thus maximize the information gain over candidate views, with no requirement for ground-truth view data. However, FisherRF still struggles to select new views to improve geometric reconstruction performance, as GS models unify both the geometric and non-geometric properties of scenes.

To measure geometric information gain during NBV selection for active Gaussian Splatting, some approaches [8], [9] explicitly construct a voxel map and then accumulate the "to-be-observed" information gain along rays traversing the volume. Reed et al. [10] distinguish between known and unknown regions in such voxel maps and prioritize sampling views at the boundaries to reduce motion costs. While these voxel map-based strategies can effectively explore the environment by acquiring additional views, they

This work is supported by Beijing Natural Science Foundation (Nos. 4252029), National Natural Science Foundation of China (No. 62176010) and the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (No. VRLAB 2024B02).

*Corresponding author. Emails: mawei@bjut.edu.cn.

[1]Wei Wang, Wei Ma, and Hongliang Zhang are with College of Computer Science, Beijing University of Technology, Beijing, 100124, China.

[2]School of Computer and Information, Anqing Normal University, Anqing, 246000, China.

[3]Key Laboratory of Machine Perception, School of IST, Peking University, Beijing, 100871, China.

overlook mapping quality, especially in regions with complex geometry.

In this paper, we propose a new NBV selection method for Gaussian Splatting, with its core being the proposed ROI-GSurFisher. Unlike FisherRF, ROI-GSurFisher specifically emphasizes geometric uncertainty modeling. We introduce an uncertainty-based view scoring module to compute ROI-GSurFisher, which calculates Fisher Information gain on 2D Gaussian surfels [11] over candidate views. By focusing on surfels adhering to the scene surface, this design inherently emphasizes geometric information gain more effectively. Additionally, the view scoring module performs calculation on ROI surfels that are close to the viewpoint, within the frustum, and front-facing to each candidate view, thus avoiding interference from view-irrelevant primitives. We further present a close-front view scoring module that prioritizes viewpoints close to and front-facing the surface, ensuring better coverage of critical surface regions. Our NBV selection is performed by maximizing the summed Fisher score and close-front score. As shown in Fig. 1, compared to FisherRF, GS reconstruction with our NBV selection method achieves significantly higher geometric quality with the same number of views, despite comparable rendered view quality.

Beyond 3D reconstruction, the proposed NBV method can be integrated into SLAM systems (e.g., GauS-SLAM [2]) by replacing their original keyframe selection modules, enabling robots to actively select information-rich views for mapping in unknown environments. As demonstrated in Fig. 1, with only 200+ views, the NBV-enhanced GauS-SLAM system yields reconstruction results comparable to those of the original system using 800+ views.

Our main contributions are as follows:

- We propose an uncertainty-based view scoring module grounded in 2D Gaussian Splatting. This module selects ROI Gaussian surfels for Fisher Information evaluation, specifically targeting improved geometric accuracy on complex surfaces.
- To further enhance surface reconstruction quality, we design a close-front view scoring module that prioritizes viewpoints close to and front-facing the surface.
- We propose a NBV selection method that jointly considers Gaussian surfels' uncertainty and close-front deviation. This method is generalizable and can be seamlessly integrated into any 2D GS-based SLAM system.

II. RELATED WORK

A. GS-based Next Best View Selection

NBV selection for Gaussian Splatting generally falls into two categories: approaches [3], [7], [13] that rely on estimating model uncertainty, as well as approaches [8], [14] that construct proxy maps for view planning. Recent work [3] compared various uncertainty estimation strategies, including learned uncertainty outputs and model ensembles. In particular, Li et al. [15] proposed a multiscale variational representation that learned an offset table for parameter space sampling, producing posterior uncertainty for each Gaussian and inferring predictive uncertainty. Although effective, this incurs additional computational overhead and depends on prior knowledge. On the other hand, RIGI [6] trains multiple GS models with different random seeds and approximates epistemic uncertainty through cross-model variance, but such methods are computationally expensive. In contrast, FisherRF [7] and HGS-Planner [16] quantify uncertainty by constructing parameter information matrices based on the sensitivity of 3DGS rendering to model parameters. However, Fisher Information derived solely from rendering optimization tends to overemphasize color cues, leading to view selections that neglect geometric information gain.

GS-Planner [17] and HGS-Planner [16] adopt hybrid scene representations with GS maps and voxel maps, incorporating unobserved voxels into the splatting pipeline to compute volumetric gains weighted by opacity. Similarly, ActiveSplat [9] combines GS with Voronoi-based map representations, projecting low-visibility pixels from panoramic GS images into 3D space and approximating unseen geometry using convex hull volume [18]. These geometry-driven methods align information gain closely with structural metrics and avoid the need for learning, but their limitation is that they remain insensitive to texture and appearance, making it difficult to distinguish views that are visible yet not informative in complex visual conditions.

B. Active View Selection for Reconstruction and Mapping

To enable efficient 3D reconstruction and embodied mapping, recent research has focused on active view selection, where candidate viewpoints are evaluated according to geometric or uncertainty-based criteria. A typical strategy [19]–[21] is the exploration of volume entropy to discretize space, score the occupied voxels by entropy, and select views accordingly. Another strategy [22] is frontier-based coverage, which defines the boundary between known free space and unknown space as frontiers (frontier cells on a 2D grid or frontier voxels in a 3D OctoMap) and prioritizes those locations.

In parallel, uncertainty-driven methods have been extensively explored in the context of Neural Radiance Fields (NeRF) [23]. Some works [24]–[26] attempt to quantify the uncertainty of the scene from density distributions along cast rays, while others [4], [5], [27]–[29] reparameterize NeRF to design a Bayesian model, but address only predictive uncertainty rather than uncertainty related to parameter observability. Regarding concurrent work, Goli et al. [30] introduced a hypothetical perturbation field and interpolated the uncertainty over NeRF-style query points for uncertainty quantification and a NeRF clean-up task. However, when it comes to active learning, they are incompatible with the latest Gaussian Splatting models.

In recent years, RGB-D SLAM with NeRF or 3D Gaussians as map representation has been widely studied, creating a need for active mapping algorithms that maximize mapping efficiency in embodied robot systems. ActiveNeRF [4] performs trajectory planning for object-centric autonomous reconstruction by predicting variance as an extra network

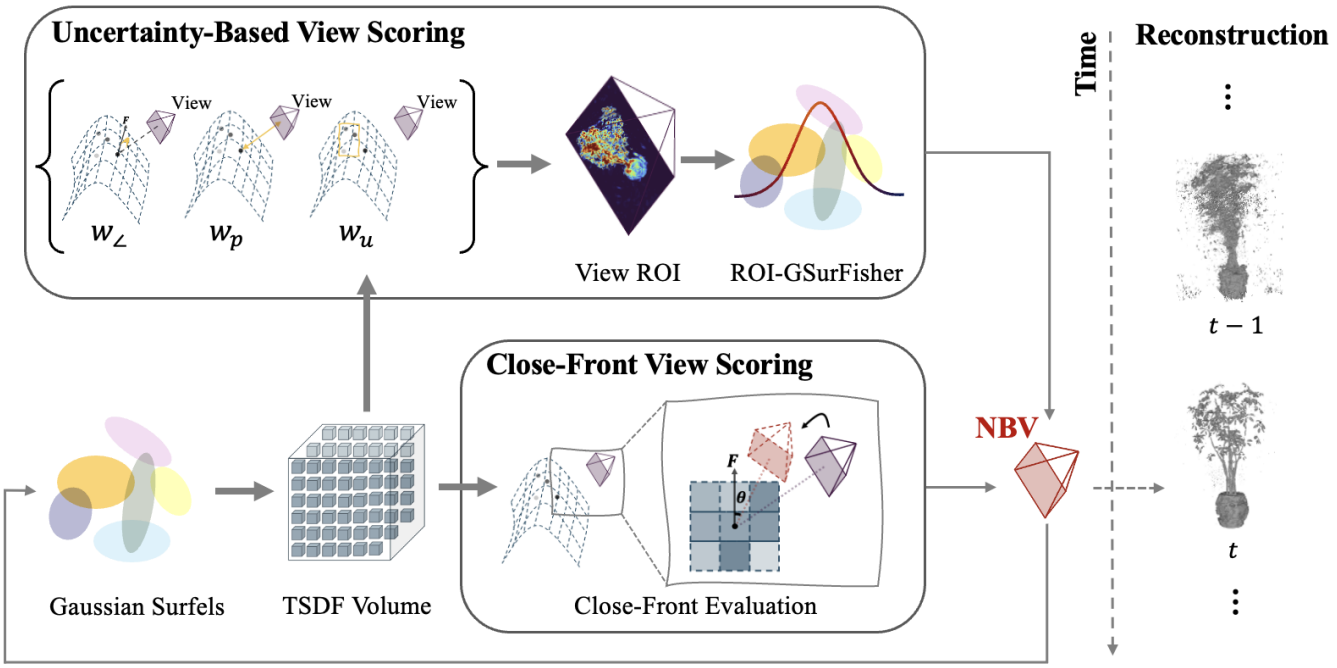

Fig. 2: **Overview of the proposed NBV selection method for Gaussian Splatting-based 3D reconstruction.** Our NBV method, designed for 2D Gaussian Splatting (2DGS), takes few-shot RGB images and corresponding camera poses as input to a 2DGS model. Rendered depth/normal maps are incrementally fused to build a TSDF volume [12]. We generate a pixel-level geometric ROI map for each candidate view based on the TSDF, to guide Fisher Information gain calculation on Gaussian surfels. We also leverage TSDF weights to compute a close-front view score. The optimal NBV is selected by maximizing the linearly fused two scores.

output. Yan et al. [31] discuss the intuition of flat minima and quantifies the predictive uncertainty of neural mapping models from the perspective of the loss landscape, but they only approximate the variance uncertainty of the output of neighboring points. ActiveGAMER [32] selects NBVs via opacity-based rendering of 3DGS for real-time mapping, but its rendering-based information gain requires extensive candidate evaluation, leading to higher motion cost. ActiveSplat [9] uses a hybrid 3DGS–Voronoi map to approximate unseen volumetric gain, but its precision is limited by voxel resolution. For the larger scope of Active Simultaneous Localization and Mapping (ASLAM), please refer to literature reviews [33], [34].

III. Preliminaries

2D Gaussian Splatting: 2D Gaussian Splatting explicitly models a 3D scene as a set of oriented planar Gaussian disks. The 2D Gaussian primitive is parameterized as $\Theta = \{\mathbf{p}, t_u, t_v, s_u, s_v, \alpha, c\}$, where $\mathbf{p}$ denotes the center point, t_u and t_v are the principal tangential vectors in the uv space that form an orthogonal tangential basis $t_\top = t_u \times t_v$, a scaling vector $\mathbf{S} = diag(s_u, s_v, 0)$ that controls the variances of the 2D Gaussian, $\mathbf{R} = [t_u, t_v, t_\top]$ is a rotation matrix. Together, $\mathbf{R}$, $\mathbf{S}$ and $\mathbf{p}$ define the homogeneous transform $\mathbf{H} = \begin{bmatrix} \mathbf{RS} & \mathbf{p} \\ 0 & 1 \end{bmatrix}$. Each primitive further carries opacity α and view-dependent appearance c (low-order spherical harmonics).

For a point $\mathbf{u} = (u, v)$ in the uv space, its 2D Gaussian value is given by the standard Gaussian as follows:

$$\mathcal{G}(\mathbf{u}) = exp(-\frac{u^2 + v^2}{2}). \tag{1}$$

Let the i-th Gaussian's opacity be α_i, the view-dependent color be c_i, and the depth be d_i. The effective contribution along the view is $m_i = \alpha_i \mathcal{G}(\mathbf{u})$. The color and depth of a pixel are accumulated with front-to-back compositing:

$$C = \sum_{i=1}^{n} c_i m_i \prod_{j=1}^{i-1}(1 - m_j), \tag{2}$$

$$D = \sum_{i=1}^{n} d_i m_i \prod_{j=1}^{i-1}(1 - m_j). \tag{3}$$

Fisher Radiance Fields: Fine-grained NBV selection is achieved by maximizing the Expected Information Gain (EIG) of the GS model using Fisher Information. Given a GS model $\mathcal{T}(y|x, w)$ parametrized by w, where x denotes a camera pose and the corresponding image observation y captured at that pose. The objective of GS optimization is to minimize the Negative Log-Likelihood (NLL) between rendered and ground-truth images, which is defined as follows:

$$-\log \mathcal{T}(y|x, w) = \|y - f(x; w)\|_2^2 . \tag{4}$$

Here, $f(x; w)$ is the rendering model. Fisher Information is defined as the Hessian matrix of the NLL with respect to the model parameters, denoted as $\mathbf{H}''[y|x, w]$.

For active view selection, starting from the training set D^{train}, the initial parameter estimate w^* is obtained. The objective is to select the NBV that maximizes the information gain between the candidate view $x \in D^{pool}$ and D^{train},where D^{pool} denotes the candidate view set. Taking advantage of the additivity of Fisher Information across views, the EIG associated with x admits a tractable upper bound that reduces to a trace-based objective, defined as:

$$\mathbf{I}(x, w^*) = \arg\max_x \ \mathrm{tr}\left(\mathbf{H}''[y \mid x, w^*]\, \mathbf{H}''[w^* \mid D^{\mathrm{train}}]^{-1}\right). \tag{5}$$

Under the Gauss–Newton approximation [35], the Hessian simplifies to $\mathbf{H}''[y|x, w^*] = J^T J$, where $J = \bigtriangledown_w f(x; w^*)$. Consequently, Hessian $\mathbf{H}''[y|x, w^*]$ can also be approximated as diagonal values as follows:

$$\begin{aligned} \mathbf{H}''[y|x, w^*] &= \nabla_w f(x; w^*)^T \nabla_w f(x; w^*) \\ &\approx \mathrm{diag}\Big(\nabla_w f(x; w^*)^T \\ &\qquad \times \nabla_w f(x; w^*)\Big) + \lambda E, \end{aligned} \tag{6}$$

where λE is a log-prior regularizer.

IV. METHOD

Our method consists of three components: 1. **Uncertainty-Based View Scoring**, 2. **Close-Front View Scoring** and 3. **Pixel-Wise Uncertainty with Volumetric Rendering**. The two scoring modules are complementary: the uncertainty-based view scoring emphasizes challenging regions and evaluates the uncertainty of Gaussian surfels. The close-front view scoring penalizes viewpoints that are distant or oriented away from the surface. These are fused into a final score for greedy NBV selection. We also render pixel-wise uncertainty to validate the correlation between the current uncertainty estimate and the depth error. Using the real-time rendering capability of 2DGS, our method enables efficient NBV selection during training.

A. Uncertainty-Based View Scoring

We emphasize the information gain from regions with poor geometric quality caused by incidence angle, scale, and occlusion. To align the selection criterion with the goal of reducing geometric uncertainty, we design an uncertainty-based view scoring scheme.

1) Construction of the View ROI: In the few-shot 2DGS setting, we incrementally construct a TSDF volume from Gaussian depth and normal maps. Based on the TSDF visibility weights, we select voxels that are near the surface, front-facing with respect to the candidate view, and inside the candidate frustum. These voxels are then projected onto the image plane to form the view region of interest (ROI). This encodes geometric informativeness per-pixel from three uncertainty-related factors: incidence angle, scale adequacy, and coverage uncertainty.

Incidence angle weight $\mathbf{W}_{\angle}$: Depth and tangential perturbation accuracy depend on parallax: small angles between the view ray and the surface normal poorly constrain geometry, while large angles raise occlusion risk. We therefore favor a texture-adaptive mid-range incidence angle: about 30° in texture-poor regions to maintain stable geometric constraints, and approximately 40° in texture-rich areas, where stronger photometric cues support increased parallax.

Scale adequacy weight $\mathbf{W}_{\mathbf{scale}}$: Voxels that are very close to the camera tend to be projected onto many pixels, leading to redundancy and texture bias, while distant voxels collapse to subpixels and contribute little photometric variation. Therefore, we assign higher weights to voxels within suitable ranges, specifically those whose projection covers approximately 2.5 pixels.

Coverage uncertainty weight $\mathbf{W}_{\mathbf{cover}}$: We boost weakly supported regions using the inverse of the TSDF fusion weight. Lower confidence yields higher per-pixel weights after projection, steering the ROI toward under-reconstructed geometry.

Given $\beta, \gamma > 0$, $\mu \geq 0$, we apply a linear gain to the geometric weight of the voxel ℓ in the candidate camera pose x to avoid over-amplification, defined as:

$$W(\ell, x) = W_{\angle}(\ell, x)^{\beta} W_{scale}(\ell, x)^{\gamma} \left(1 + \mu W_{cover}(\ell, x)\right). \tag{7}$$

We splat the voxel weights onto the candidate camera plane and normalize the resulting ROI weight map. The weight of the ROI pixel p for the candidate camera x is

$$W_{\mathrm{roi}}(p; x) = \mathrm{Norm}\left(\sum_{\ell \in \mathcal{S}_x} K(p, \pi_x(\ell))\, W(\ell, x)\right). \tag{8}$$

Here, S_x is the set of front-facing surface voxels under pose x, π_x is the camera projection operator, and K is a bilinear kernel.

2) ROI-GSurFisher Evaluation: We perform Fisher Information to evaluate candidate views so that their ranking reflects the expected reduction in geometric uncertainty. Given a candidate camera pose x, we calculate its utility only on ROI pixels $\mathcal{P}_{roi}(x)$ and on the geometry-related parameters of the Gaussian surfel $\{\mathbf{p}, t_u, t_v, s_u, s_v\}$. Let w^* denote the current parameter estimate obtained from the training set D^{train}. The rendered output at pixel p under pose x is denoted as $f_p(x; w^*)$. Taking σ_p as the noise scale per-pixel, we construct a diagonal Fisher approximation per-view within the ROI from the rendering sensitivity to geometric parameters w^{geo} as follows:

$$\mathbf{F}^{\mathrm{diag}}(x, w^*) = \sum_{p \in \mathcal{P}_{\mathrm{roi}}(x)} \frac{W_{\mathrm{roi}}(p; x)}{\sigma_p^2} \left\| \nabla_{w^{geo}} f_p(x; w^*) \right\|_2^2. \tag{9}$$

To avoid repeatedly selecting views that reinforce already well-constrained geometry, we normalize the per-view diagonal Fisher by the accumulated training Fisher. Then, the final Fisher Information score of Gaussian surfels within the

ROI is

$$\mathbf{F}(x, w^*) = \Big\langle \mathbf{F}^{\text{diag}}(x, w^*), \big(\mathbf{H}''[w^{geo} \mid D^{\text{train}}]^{-1} + \phi\big)^{-1} \Big\rangle. \tag{10}$$

Specifically, $\mathbf{H}''[w^{geo} \mid D^{\text{train}}]$ denotes the diagonal Fisher accumulated over the training set, $\phi > 0$ stabilizes the inverse, and $\langle ., . \rangle$ denotes the element-wise inner product.

B. Close-Front View Scoring

Selecting views solely based on uncertainty neglects surface orientation and geometric visibility. To address this, we propose a close-front scoring method that favors candidate views with close-range and front-facing observations, while avoiding costly geometric reasoning.

Based on the TSDF volume introduced in Section A, we assign to each voxel ℓ a per-view observation weight $W_r(\ell)$ under training view $r \in D^{train}$, where D^{train} is the training view set. We set the fused weight $W_f(\ell)$ as the maximum of its per-view weights in all training views.

$$W_r(\ell) = \cos\theta_r(\ell)/z_r(\ell), \tag{11}$$

$$W_f(\ell) = \max_{r \in D^{train}} W_r(\ell), \tag{12}$$

where $\theta_r(\ell)$ is the angle between the view direction of the training view r and the surface normal $\mathbf{n}_\ell$ obtained from the 2D GS normal map. $z_r(\ell)$ denotes the depth of voxel ℓ in the coordinate frame of view r.

For each candidate camera pose x, let $\theta_x(\ell)$ denote the angle between the viewing direction of pose x and the surface normal $\mathbf{n}_\ell$. The new weight of the voxel ℓ in candidate pose x is

$$W_x(\ell) = \cos\theta_x(\ell)/z_x(\ell). \tag{13}$$

Consequently, the visibility information gain contributed by candidate camera pose x to the current reconstruction is calculated as

$$\mathbf{I}_{\text{vis}}(x) = \sum_{\ell \in \mathcal{X}} \big[W_x(\ell) - W_f(\ell)\big]_+, \tag{14}$$

where $[a]_+ = \max(a, 0)$, $\mathcal{X}$ denotes the set of near-surface, front-facing voxels that projected into the candidate view. Accordingly, we compute a surface-aware global geometric score for each candidate view and rank the candidates in descending order. A higher score indicates that the candidate view is more front-facing and closer to the surface, thus providing stronger geometric observability for under-explored regions.

Finally, we apply z-score normalization to uncertainty-based and close-front view scores, and linearly fuse them to obtain the final information gain $\tilde{\mathbf{I}}$, defined as:

$$\tilde{\mathbf{I}}(x, w^*) = \varepsilon z(\mathbf{F}(x, w^*)) + \delta z(\mathbf{I}_{\text{vis}}(x)). \tag{15}$$

Here, $\varepsilon > 0$ and $\delta > 0$ are fusion weights that balance the contributions of the two components. Let D^{pool} denote the candidate view set. We select NBV x_{NBV} by maximizing the final information gain as follows:

$$x_{NBV} = \arg\max_{x \in D^{pool}} \tilde{\mathbf{I}}(x, w^*). \tag{16}$$

TABLE I: **Novel View Synthesis Results on the Blender Dataset with 20 Training Views.** All the metrics are averaged over the 5 objects/scenes in the dataset.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
3DGS [SIGGRAPH 2023]	28.73	0.939	0.053
FisherRF [ECCV 2024]	29.52	0.944	0.043
2DGS [SGGRAPH 2024]	29.64	0.948	0.048
Ours	**30.83**	**0.958**	**0.040**

C. Pixel-Wise Uncertainty with Volumetric Rendering

Previously, we quantified Fisher Information within the ROI for any candidate views. Our formulation can also be extended to obtain pixel-wise uncertainties. As discussed in Section A, uncertainty of the parameters can be approximated using the diagonal elements of the Hessian matrix. For each ROI pixel p under candidate camera pose x, we cast a ray and compute its uncertainty through front-to-back volume compositing. The contribution of the i-th Gaussian surfel along the ray is weighted by its opacity α_i and pre-transmittance $T_i = \prod_{k<i}(1 - \alpha_k)$. Following FisherRF, the uncertainty of pixel p is

$$U_x(p) = \sum_{i=1}^{N} T_i \alpha_i \,\text{tr}(\mathbf{G}_i). \tag{17}$$

Here, $\mathbf{G}_i$ denotes the Hessian block of the training NLL evaluated at the optimum w^{geo}, corresponding to the i-th Gaussian surfel. The trace summarizes the information content of the i-th Gaussian. The computation is restricted to the ROI pixel set and reweighted per pixel by $W_{roi}(p; x)$. The uncertainty of ROI pixel p is defined as:

$$U_{\text{roi}}(p; x) = \frac{W_{\text{roi}}(p; x)}{\sigma_p^2} U_x(p). \tag{18}$$

V. RESULTS

A. Experimental Setup

Datasets: We evaluate the performance of our method on multiple datasets, including Blender [23], DTU [36], LF [37], and Replica [38].

Metrics: We evaluate geometric completeness using distance-based metrics, including Acc and Comp [39]. Higher scores indicate better geometric accuracy and completeness. For novel view synthesis, we report standard photometric metrics, including the peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM), and the learned perceptual image patch similarity (LPIPS).

Baselines: We compare our approach against state-of-the-art methods, including FisherRF [7] and 3DGS [1]. For 2DGS [11], we use the random view selection strategy. For active mapping experiments, we further benchmark against GauS-SLAM [2].

Implementation Details: We evaluate NBV selection with a 2D Gaussian Splatting backend on the DTU and Blender datasets. All models are initialized with the same random seed and trained from four uniformly sampled initial views

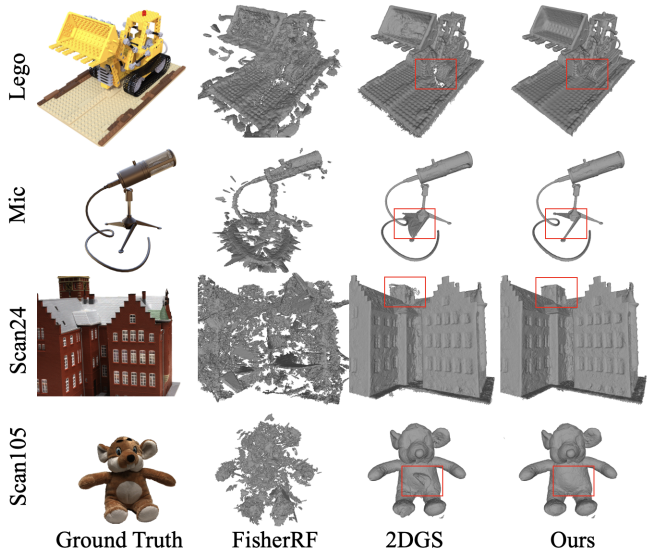

Fig. 3: **Qualitative Results obtained with 20 Training Views.** The examples are from the Blender Dataset and the DTU Dataset.

for 30,000 iterations following the default 2DGS configuration. We reset the opacity whenever new views are added to prevent training instability. For view scoring fusion, we set the weighting parameters to $\varepsilon = 0.4$ and $\delta = 0.6$. All other settings remain identical across experiments.

B. Active View Selection

We evaluate both the baseline and our method with sequential active view selection, setting the same initial views and view selection schedules. And we select one new view every 100 epochs until the model reaches 20 training views. The quantitative results are reported in Table I, showing that our method consistently outperforms the baseline in different metrics.

Qualitative comparisons on the Blender and DTU datasets are presented in Fig. 3. The results show that, under reconstruction with 20 views, FisherRF, which selects views by reducing color uncertainty, renders better than random selection in 3DGS but fails to improve geometric reconstruction, because it suffers from geometry distortion caused by perspective projection. In contrast, 2DGS models the scene using surface elements instead of volumetric ones and performs perspective-correct splatting via explicit ray-disk intersection, yielding a geometric reconstruction quality higher than FisherRF. However, random view selection with 2DGS leads to missing geometry or local redundancy in sparse-view regions. Our method shows clear superiority in both rendering quality and geometric completeness.

In addition, following previous work [7] on uncertainty estimation, we evaluated our approach on the LF dataset using the Area Under Sparsification Error (AUSE) metric. Specifically, AUSE is the area between two sparsification curves obtained by progressively removing pixels ranked by (i) absolute depth error to ground truth and (ii) predicted uncertainty; a smaller AUSE means uncertainty aligns better with erro (see Table II and Fig. 4). Our method achieves a stronger correlation between uncertainty and depth error than previous state-of-the-art approaches.

TABLE II: **Uncertainty Estimation Results on the LF Dataset.** The reported Numbers are AUSE (a lower value indicates better performance).

Method	Statue	Africa	Torch	Basket	Average
CF-NeRF [40]	0.54	0.34	0.50	**0.14**	0.38
ActiveNeRF [4]	0.40	0.39	**0.21**	0.34	0.33
BayesRays [30]	0.17	0.27	0.22	0.28	0.23
FisherRF [7]	0.21	**0.26**	0.24	0.18	0.22
Ours	**0.11**	0.31	0.29	0.16	**0.21**

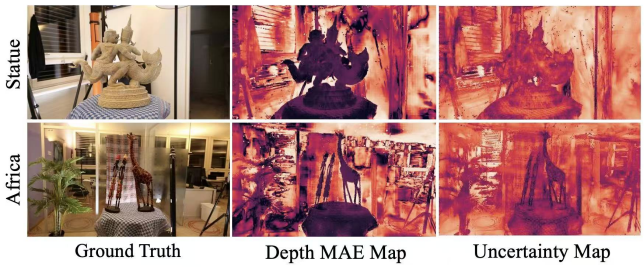

Fig. 4: **Visualizations of Uncertainty maps and Depth MAE maps.** Our per-pixel uncertainty maps demonstrate a strong correlation with the depth error maps. The examples are from the LF Dataset.

C. Active view Selection for Mapping

SLAM systems typically rely on posed RGB-D inputs and ground-truth images, selecting a large number of keyframes for mapping. To simulate active mapping in unknown environments, we integrate our NBV strategy into the SLAM pipeline. Unlike standard SLAM, our approach does not require image ground truth for keyframe selection. Instead, given only the candidate camera poses, we evaluate them by their contribution to the reconstruction quality. This enables performance comparable to multi-view training while using significantly fewer views, thereby reducing the motion cost of acquiring large image sets in real robotic scenarios.

For evaluation, we adopt the state-of-the-art GauS-SLAM [2] as the baseline. GauS-SLAM represents scenes with 2D Gaussian surfels and performs localization and mapping through a front-end/back-end framework. In our system, we replace the keyframe selection module with our NBV method while keeping all other components identical. Active mapping experiments on room-scale scene datasets demonstrate that, with only 200+ selected views, our approach achieves mesh quality comparable to GauS-SLAM trained with 800+ views (see Table III and Fig. 5).

D. Ablation Study

We perform an ablation study on the Blender dataset. We examine the effects of uncertainty-based view scoring (VS),

TABLE III: **Comparison of Reconstruction Performance on the Replica Dataset.** With only 200+ selected keyframes, our method achieves reconstruction results comparable to those of GauS-SLAM (which uses 800+ views).

Method	Data	PSNR ↑	Acc ↑	Comp ↑	#Views
GauS-SLAM [2]	Office 0	44.00	0.937	0.768	800+
	Office 1	44.57	0.950	0.718	800+
Ours	Office 0	43.36	0.938	0.762	200+
	Office 1	44.62	0.959	0.706	200+

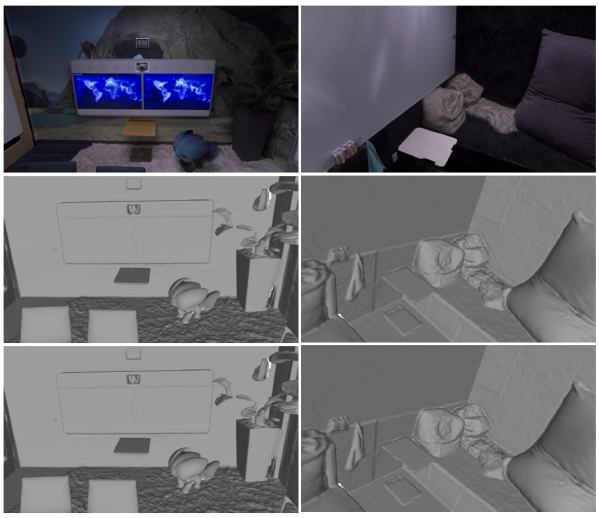

Fig. 5: **Comparison of Mesh Results on Replica.** The first row shows the ground truth, the second row shows GauS-SLAM (800+ views), and the third row shows our method (200+ views). Compared with many-view mapping, our method achieves equally smooth meshes using far fewer views.

close-front VS, and their fusion on the reconstruction quality. Using 2DGS with random view selection as the baseline, we remove each module in turn and evaluate reconstruction accuracy. In all experiments, a new view is added every 100 epochs until the model reaches 20 training views.

Quantitative and qualitative results are reported in Table IV and Fig. 6, respectively. The close-front vs module yields the lowest PSNR and worst LPIPS but a higher SSIM, indicating coarse structural alignment, because it tends to select views that are close to and front-facing the surface. The uncertainty-based VS module attains the lowest SSIM and slightly worse PSNR/LPIPS than the full model, suggesting that uncertainty guidance is critical for enhancing appearance details but overlooks geometric structure. The full model achieves the best PSNR and LPIPS, with SSIM slightly below the surface-aware variant, because adding the uncertainty-based VS module favors views that reduce photometric uncertainty and recover richer textures, while also introducing mild color variations and subpixel misalignments under larger parallax, effects that SSIM penalizes as structural differences. From

TABLE IV: **The Quantitative Results of the Ablation Study.**

Methods	PSNR ↑	SSIM ↑	LPIPS ↓
Rand	26.82	0.852	0.146
w/o Uncertainty-Based VS	24.70	**0.857**	0.155
w/o Close-Front VS	27.24	0.841	0.135
Full Model	**28.25**	0.852	**0.134**

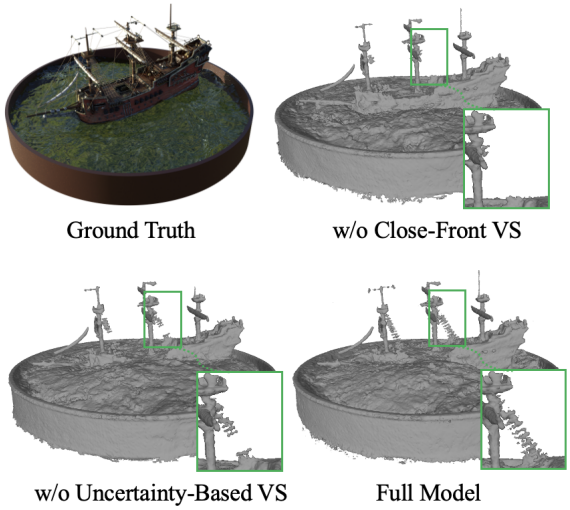

Fig. 6: **The Qualitative Results of the Ablation Study.**

mesh accuracy, the close-front VS module produces more complete geometry than the uncertainty-based VS module, and the full model is the most complete with smoother surfaces. Overall, the full model achieves a balanced trade-off between geometry and appearance.

VI. CONCLUSIONS

We presented a novel NBV selection method tailored for 2D GS-based active 3D reconstruction and SLAM. Specifically, we designed a view selection strategy that leverages two key cues, including geometry-guided uncertainty and view-to-surface reliability, to identify the most informative viewpoints for 2DGS training. This strategy enables an accurate evaluation of geometric information gain, which directly supports high-fidelity 3D reconstruction. We further integrated this NBV strategy into a SLAM system, allowing simulated robots to actively select viewpoints to reconstruct objects and scene surfaces in unknown environments.

For future work, we plan to extend our model by incorporating semantic uncertainty. This extension is motivated by its alignment with human visual perception: in human observation, semantic ambiguities often stem from poorly reconstructed surfels, and addressing this gap will enhance the robustness of our view selection against such ambiguities.

REFERENCES

[1] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (TOG)*, 42(4), 2023.

[2] Yongxin Su, Lin Chen, Kaiting Zhang, Zhongliang Zhao, Chenfeng Hou, and Ziping Yu. Gaus-slam: Dense rgb-d slam with gaussian surfels. *IEEE Robotics and Automation Letters*, pages 1–8, 2026.

[3] Marcus Klasson, Riccardo Mereu, Juho Kannala, and Arno Solin. Sources of uncertainty in 3d scene reconstruction. In *European Conference on Computer Vision (ECCV)*, pages 271–289, 2024.

[4] Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. Activenerf: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision (ECCV)*, pages 230–246, 2022.

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

[6] Jiacheng Wang, Zhedong Zheng, Wei Xu, and Ping Liu. Rigi: Rectifying image-to-3d generation inconsistency via uncertainty-aware learning. *arXiv preprint arXiv:2411.18866*, 2024.

[7] Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherrf: Active view selection and mapping with radiance fields using fisher information. In *European Conference on Computer Vision (ECCV)*, pages 422–440, 2024.

[8] Liren Jin, Xingguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. Activegs: Active scene reconstruction using gaussian splatting. *IEEE Robotics and Automation Letters*, 10(4):4866–4873, 2025.

[9] Yuetao Li, Zijia Kuang, Ting Li, Qun Hao, Zike Yan, Guyue Zhou, and Shaohui Zhang. Activesplat: High-fidelity scene reconstruction through active gaussian splatting. *IEEE Robotics and Automation Letters*, 10(8):8099–8106, 2025.

[10] Alec Reed, Lorin Achey, Brendan Crowe, Bradley Hayes, and Christoffer Heckman. Online diffusion-based 3d occupancy prediction at the frontier with probabilistic map reconciliation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2846–2852, 2025.

[11] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 Conference Papers (SIGGRAPH)*, pages 1–11, 2024.

[12] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 303–312, 1996.

[13] Joey Wilson, Marcelino Almeida, Sachit Mahajan, Martin Labrie, Maani Ghaffari, Omid Ghasemalizadeh, Min Sun, Cheng-Hao Kuo, and Arnab Sen. Pop-gs: Next best view in 3d-gaussian splatting with p-optimality. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 3646–3655, 2025.

[14] Jiaxin Wei and Stefan Leutenegger. Gsfusion: Online rgb-d mapping where gaussian splatting meets tsdf fusion. *IEEE Robotics and Automation Letters*, 9(12):11865–11872, 2024.

[15] Ruiqi Li and Yiu-ming Cheung. Variational multi-scale representation for estimating uncertainty in 3d gaussian splatting. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 87934–87958, 2024.

[16] Zijun Xu, Rui Jin, Ke Wu, Yi Zhao, Zhiwei Zhang, Jieru Zhao, Fei Gao, Zhongxue Gan, and Wenchao Ding. Hgs-planner: Hierarchical planning framework for active scene reconstruction using 3d gaussian splatting. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14161–14167, 2025.

[17] Rui Jin, Yuman Gao, Yingjian Wang, Yuze Wu, Haojian Lu, Chao Xu, and Fei Gao. Gs-planner: A gaussian-splatting-based planning framework for active high-fidelity reconstruction. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11202–11209, 2024.

[18] David P. Dobkin Barber, C. Bradford and Hannu Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*, 22(4):469–483, 1996.

[19] Stefan Isler, Reza Sabzevari, Jeffrey Delmerico, and Davide Scaramuzza. An information gain formulation for active volumetric 3d reconstruction. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3477–3484, 2016.

[20] Pere-Pau Vázquez, Miquel Feixas, Mateu Sbert, and Wolfgang Heidrich. Viewpoint selection using viewpoint entropy. In *Vision, Modeling and Visualization (VMV)*, pages 273–280, 2001.

[21] Kelsey Saulnier, Nikolay Atanasov, George J Pappas, and Vijay Kumar. Information theoretic active exploration in signed distance fields. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4080–4085, 2020.

[22] Boyang Sun, Hanzhi Chen, Stefan Leutenegger, Cesar Cadena, Marc Pollefeys, and Hermann Blum. Frontiernet: Learning visual cues to explore. *IEEE Robotics and Automation Letters*, 10(7):6576–6583, 2025.

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

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

[25] Huangying Zhan, Jiyang Zheng, Yi Xu, Ian Reid, and Hamid Rezatofighi. Activermap: Radiance field for active mapping and planning. *arXiv preprint arXiv:2211.12656*, 2022.

[26] Dongyu Yan, Jianheng Liu, Fengyu Quan, Haoyao Chen, and Mengmeng Fu. Active implicit object reconstruction using uncertainty-guided next-best-view optimization. *IEEE Robotics and Automation Letters*, 8(10):6395–6402, 2023.

[27] Jianxiong Shen, Antonio Agudo, Francesc Moreno-Noguer, and Adria Ruiz. Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In *European Conference on Computer Vision (ECCV)*, pages 540–557, 2022.

[28] Jianxiong Shen, Adria Ruiz, Antonio Agudo, and Francesc Moreno-Noguer. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In *2021 International Conference on 3D Vision (3DV)*, pages 972–981, 2021.

[29] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7210–7219, 2021.

[30] Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes' rays: Uncertainty quantification for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20061–20070, 2024.

[31] Zike Yan, Haoxiang Yang, and Hongbin Zha. Active neural mapping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10981–10992, 2023.

[32] Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, and Yi Xu. Activegamer: Active gaussian mapping through efficient rendering. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 16486–16497, 2025.

[33] Iker Lluvia, Elena Lazkano, and Ander Ansuategi. Active mapping and robot exploration: A survey. *Sensors*, 21:2445, 2021.

[34] Julio A Placed, Jared Strader, Henry Carrillo, Nikolay Atanasov, Vadim Indelman, Luca Carlone, and José A Castellanos. A survey on active simultaneous localization and mapping: State of the art and new frontiers. *IEEE Transactions on Robotics*, 39(3):1686–1705, 2023.

[35] Weijun Zhou and Xiaojun Chen. Global convergence of a new hybrid gauss–newton structured bfgs method for nonlinear least squares problems. *SIAM Journal on Optimization*, 20(5):2422–2441, 2010.

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

[37] Kaan Yücer, Alexander Sorkine-Hornung, Oliver Wang, and Olga Sorkine-Hornung. Efficient 3d object segmentation from densely sampled light fields with applications to 3d reconstruction. *ACM Transactions on Graphics (TOG)*, 35(3):1–15, 2016.

[38] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, et al. The replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019.

[39] Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. Mvsnet: Depth inference for unstructured multi-view stereo. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 767–783, 2018.

[40] Jianxiong Shen, Antonio Agudo, Francesc Moreno-Noguer, and Adria Ruiz. Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In *European Conference on Computer Vision (ECCV)*, pages 540–557, 2022.