EUROGRAPHICS 2026 / B. Masia and J. Thies
(Guest Editors)

COMPUTER GRAPHICS *forum*
*Volume 45 (2026), Number 2*

# OUGS: Active View Selection via Object-aware Uncertainty Estimation in 3DGS

Haiyi Li[1], Qi Chen[1], Denis Kalkofen[2], and Hsiang-Ting Chen[1]

[1]University of Adelaide, Australia
[2]Graz University of Technology

**Abstract**

*Recent advances in 3D Gaussian Splatting (3DGS) have achieved state-of-the-art results for novel view synthesis. However, efficiently capturing high-fidelity reconstructions of specific objects within complex scenes remains a significant challenge. A key limitation of existing active reconstruction methods is their reliance on scene-level uncertainty metrics, which are often biased by irrelevant background clutter and lead to inefficient view selection for object-centric tasks. We present OUGS, a novel framework that addresses this challenge with a more principled, physically-grounded uncertainty formulation for 3DGS. Our core innovation is to derive uncertainty directly from the **explicit physical parameters** of the 3D Gaussian primitives (e.g., position, scale, rotation). By propagating the covariance of these parameters through the rendering Jacobian, we establish a highly interpretable uncertainty model. This foundation allows us to then seamlessly integrate semantic segmentation masks to produce a targeted, **object-aware** uncertainty score that effectively disentangles the object from its environment. This allows for a more effective active view selection strategy that prioritizes views critical to improving object fidelity. Experimental evaluations on public datasets demonstrate that our approach significantly improves the efficiency of the 3DGS reconstruction process and achieves higher quality for targeted objects compared to existing state-of-the-art methods, while also serving as a robust uncertainty estimator for the global scene.*

**CCS Concepts**

• ***Computing methodologies*** → ***Reconstruction; Active vision;*** *Image-based rendering;*

**Keywords:** 3D Reconstruction, Computing methodologies Shape analysis, Active Vision, 3D Gaussian Splatting

## 1. Introduction

Efficient 3D scene reconstruction is a foundational goal in computer vision and robotics. The advent of Neural Radiance Fields (NeRF) [MST*20] marked a breakthrough, enabling photorealistic novel view synthesis by learning implicit volumetric representations. More recently, 3D Gaussian Splatting (3DGS) [KKLD23] has emerged as a compelling alternative. By modeling scenes with explicit 3D Gaussian primitives and leveraging a fast, differentiable rasterization pipeline, 3DGS achieves real-time rendering speeds without compromising visual fidelity, addressing the high computational cost that limits NeRF's practicality.

Despite these advances, both NeRF and 3DGS remain highly data-intensive, typically requiring dense image captures to produce high-quality reconstructions [NBM*22, CKD*25]. This motivates the need for active reconstruction [CLK11], where a human [STC*25,MZC*25] or robotic agent actively selects a minimal subset of views that maximally reduces uncertainty. Achieving this requires models to estimate their own information needs in real-time—a capability that hinges on accurate uncertainty estimation.

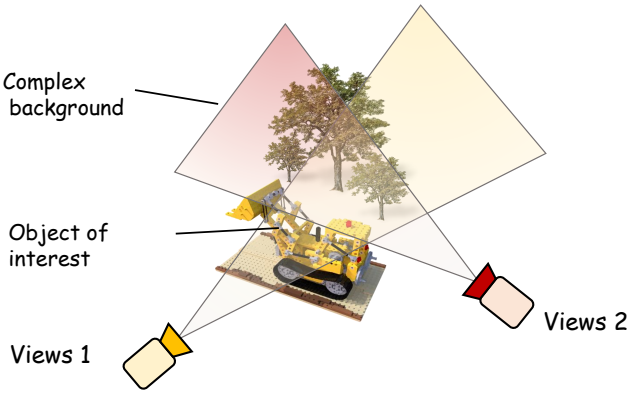


Figure 1: **A complex background can inflate image-level uncertainty.** *This can mislead active view selection away from the object.*

In this context, recent work has explored incorporating uncertainty into both NeRFs [GRS*24, KMKS25] and 3DGS. For the latter, emerging research has methods based on ensemble variance [HD25] or Fisher Information approximations [JLD25]. How-

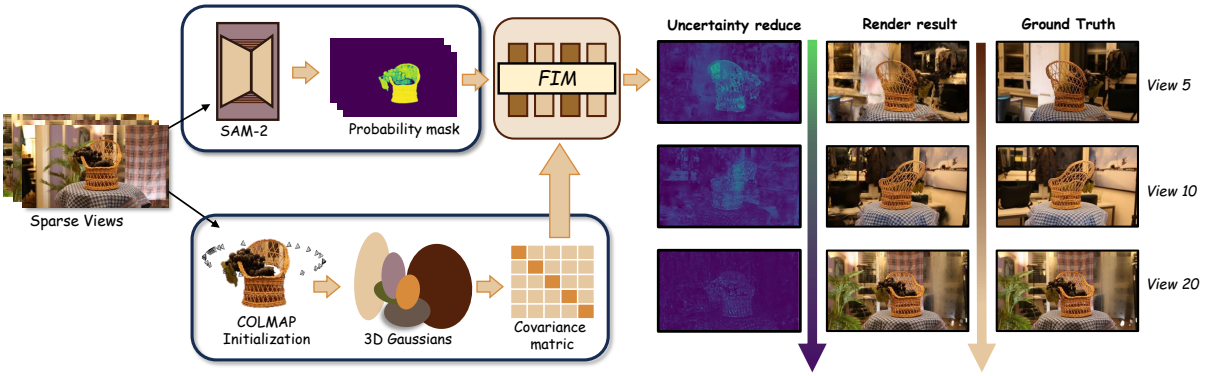


Figure 2: **Object-aware uncertainty guides 3DGS view planning for precise object reconstruction.** *Our physically-grounded uncertainty model, derived from the explicit parameters of the 3D Gaussians, is combined with a semantic mask to generate an object-level uncertainty score. This score effectively guides the active view selection to improve object fidelity, as shown for 5, 10, and 20 selected views.*

ever, these pioneering methods share a fundamental limitation: they typically estimate uncertainty at the scene level. As illustrated in figure 1, a global uncertainty score is often dominated by complex but irrelevant background clutter, misleading the view selection process. This is particularly problematic for applications where the primary goal is to capture a **specific object of interest** with the highest possible fidelity, such as in cultural heritage preservation, industrial inspection, and AR [CLDC23, WC23]. These scenarios demand object-centric planning, where background noise can distort the focus.

To address this critical gap, we introduce OUGS, a framework designed specifically for **object-centric active reconstruction** that prioritizes the most informative camera viewpoints during training to improve reconstruction quality and efficiency (Figure 2) Our work is built upon a key insight: to effectively isolate an object's uncertainty, one must first model uncertainty from a more fundamental, physically-grounded source, as is common practice, our method is the first to establish a rigorous framework that quantifies uncertainty directly from the **explicit physical parameters** of the 3D Gaussian primitives—their position, scale, rotation, and appearance. We begin by treating these parameters as random variables and propagate their covariance through the differentiable rendering pipeline via the Jacobian. This yields a pixel-wise visual uncertainty map that is not only robust but also highly interpretable. This covariance captures epistemic uncertainty over existing Gaussian parameters—precisely what next-best-view selection aims to reduce during refinement. Since unobserved regions may lack instantiated Gaussians after sparse initialization, our method focuses on improving already-represented areas rather than discovery of entirely new structure.

This explicit projection to pixel space represents a crucial advantage over implicit methods like FisherRF [JLD25]. While FisherRF optimizes an information-theoretic objective in the abstract parameter space of a neural field, our approach generates a native spatial uncertainty field. This spatial modularity allows us to seamlessly integrate semantic masks as direct filters, effectively disentangling the target object's uncertainty from its environment.

To ensure scalability, we approximate the parameter covariance using a diagonal Fisher Information Matrix (FIM), updated efficiently through an exponential moving average. This complete formulation enables an active view selection strategy that is powerfully and precisely focused on the object of interest. Our contributions are threefold:

- We introduce a novel active reconstruction framework specifically designed for **object-centric tasks** in 3DGS, addressing a key limitation of existing scene-level methods.
- We propose a new, **physically-grounded uncertainty model** based on the explicit parameters of 3D Gaussians, offering greater accuracy and interpretability compared to implicit, weight-based approaches.
- Through extensive experiments, we demonstrate that our method significantly outperforms state-of-the-art approaches in object-focused reconstruction while maintaining strong performance on global scene metrics.

## 2. Related Work

### 2.1. Uncertainty in 3D Splatting

Quantifying uncertainty in 3DGS is an emerging research area crucial for real-world applications. Current approaches can be categorized into four main directions: **(1) Variational/Bayesian.** A principled approach is to treat Gaussian parameters as distributions. Li & Cheung [LC24] use hierarchical Bayesian priors, while Savant et al. [AVM25] employ variational inference. While mathematically rigorous, these methods incur significant computational overhead, limiting their real-time applicability. **(2) Sensitivity-based pruning.** Alternatively, some methods measure the model's sensitivity to its parameters. PUP 3D-GS, for instance, uses a Hessian-based metric to prune Gaussians with high uncertainty. This approach is efficient but offers a less direct measure of predictive uncertainty. **(3) Learned uncertainty fields.** Several works train auxiliary predictors to output uncertainty maps or uncertainty-related signals. UNG-GS [TCZ*25] introduces an uncertainty-aware field to better handle sparse inputs, while Han & Dumery [HD25] learn a view-dependent uncertainty field for 3DGS. These methods are flexible, but the resulting uncertainty is typically model-dependent and may be less physically interpretable than parameter-centric formulations. We note that some concurrent 3DGS uncertainty systems (including UNG-GS at the time of our experiments) did not provide

a publicly available implementation, which prevented us from reproducing their exact pipelines and performing controlled, apples-to-apples comparisons under our object-centric evaluation protocol. **(4) Information-theoretic.** Finally, information theory can be used to quantify the information value of additional views. GauSS-MI [XCZ*25], for instance, selects views that maximize mutual information. While powerful for view selection, this paradigm focuses on the information value of candidate views rather than directly modeling the inherent uncertainty of the current reconstruction.

Our work carves a distinct path by adopting an efficient, FIM-based approximation of parameter uncertainty. We apply this formulation directly to *object-centric active view selection*—a critical application gap not fully addressed by prior works.

### 2.2. Uncertainty for Active View Selection

Active view selection, or Next-Best-View (NBV) planning, is a long-standing problem in computer vision and robotics [CLK11], aiming to intelligently choose views to maximize reconstruction quality while minimizing cost. Methodologies have evolved significantly over time. **(1) Traditional and geometric methods.** Early approaches in robotics often relied on geometric heuristics. For instance, receding-horizon planners like the one by Bircher et al. [BKA*16] aim to maximize exploration of unknown free space using occupancy maps. Other classical NBV methods use voxel-grid representations and select views based on metrics like Shannon entropy or frontier exploration [KSH22]. While effective for coverage, these discretized methods can struggle to capture fine geometric details and are less suited for the continuous representations used in modern neural rendering. **(2) Uncertainty- and information-driven neural rendering.** With neural rendering, view selection can be guided by predictive uncertainty or information gain. ActiveNeRF [PLSH22] uses variance-driven criteria, while FisherRF [JLD25] proposes Fisher-information-based objectives for principled NBV selection. Concurrent works extend these ideas to 3DGS: POp-GS [WAM*25] uses Fisher-matrix-based planning, while GauSS-MI [XCZ*25] ranks views by mutual information. However, these methods optimize *global* scene-level objectives, where background clutter can dominate the planning signal—suboptimal when reconstructing a specific object. As POp-GS lacked public code during our study, we compare against baselines with available implementations under controlled settings, using identical masks where applicable. **(3) Other view-selection paradigms.** Beyond uncertainty and information-theoretic approaches, other paradigms have been explored. Learning-based methods, such as NeurAR [RZH*23], employ reinforcement learning to train an agent that learns an optimal view selection policy directly from simulation. Concurrently, other works focus on explicitly modeling visibility. For instance, Neural Visibility Fields [XDM*24] learn to predict which parts of a scene are visible from a given viewpoint, guiding selection towards views that maximize observable new area. While powerful, these methods either require extensive training or shift the focus from reconstruction fidelity to geometric coverage.

Our work addresses the critical limitation of scene-level uncertainty methods. By introducing an object-aware mechanism built on a physically grounded, parameter-centric uncertainty formulation, we enable the view selection process to focus on semantically important regions of the scene, a challenge not explicitly addressed by any of these prior paradigms.

## 3. Method

### 3.1. Preliminary: 3D Gaussian Splatting

Our method builds upon the 3D Gaussian Splatting (3DGS) framework [KKLD23], which represents a scene as a collection of anisotropic 3D Gaussian primitives $\mathcal{G} = \{\mathcal{G}_i\}_{i=1}^{N_g}$. To ground our uncertainty analysis, we first provide a detailed list of the parameters used in the differentiable rendering pipeline. Each Gaussian $\mathcal{G}_i$ is fully described by a parameter vector $\theta_i$:

$$\theta_i = \Big[\underbrace{\mu_i}_{\text{Center}}, \underbrace{\mathbf{s}_i}_{\text{Scale}}, \underbrace{\mathbf{q}_i}_{\text{Rotation}}, \underbrace{\alpha_i}_{\text{Opacity}}, \underbrace{\mathbf{f}_i^{\text{dc}}, \mathbf{f}_i^{\text{sh}}}_{\text{Color (SH)}}\Big]^\top \tag{1}$$

where the components are:

- Geometry: The 3D center $\mu_i \in \mathbb{R}^3$, an anisotropic scaling vector $\mathbf{s}_i \in \mathbb{R}^3_+$, and an orientation quaternion $\mathbf{q}_i \in \mathbb{S}^3$. Together, these define the Gaussian's position, size, and orientation.
- Appearance: A scalar opacity value $\alpha_i \in \mathbb{R}$ and view-dependent color modeled by Spherical Harmonics (SH). The color is parameterized by the degree-0 (DC) term $\mathbf{f}_i^{\text{dc}} \in \mathbb{R}^3$ and a set of higher-order coefficients $\mathbf{f}_i^{\text{sh}} \in \mathbb{R}^{3\times 15}$.

A visual breakdown of these parameters is provided in figure 3.

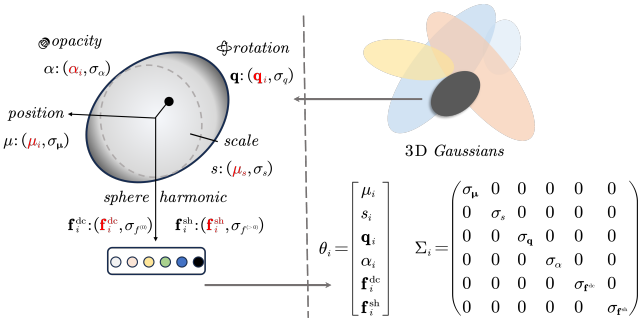


Figure 3: **The parameterization of a 3D Gaussian primitive.** *Each Gaussian, the fundamental building block of our scene representation, is defined by a set of explicit physical parameters.*

Rendering in 3DGS uses a differentiable splatting approach based on standard alpha compositing. First, the 3D Gaussians are projected onto the 2D image plane and sorted in front-to-back order based on their depth. The color $C(u)$ for a pixel $u$ is then:

$$C(u) = \sum_{i\in\mathcal{I}(u)} c_i(u)\alpha_i'(u)\prod_{j=1}^{i-1}\big(1-\alpha_j'(u)\big) \tag{2}$$

where $\mathcal{I}(u)$ is the list of Gaussians overlapping the pixel $u$, $c_i(u)$ is the view-dependent color evaluated from SH. The effective opacity $\alpha_i'(u)$ is determined by modulating the Gaussian's learned opacity parameter $\alpha_i$ by its projected 2D profile at the pixel location.

### 3.2. Mapping 3D Gaussian Parameter Uncertainty to Pixel-wise Object-aware Uncertainty

To quantify uncertainty, we treat the parameter vector of each Gaussian as a random variable and initialize it with a Gaussian prior, $\theta_i \sim \mathcal{N}(\theta_i^0, \Sigma_i^0)$. For notational clarity, we stack the per-Gaussian covariances into a block-diagonal matrix

$$\Sigma = \mathrm{diag}\big(\Sigma_1, \ldots, \Sigma_{N_g}\big) \in \mathbb{R}^{d\times d} \tag{3}$$

where $d = N_g \cdot d_g$. Throughout the rest of this section, $\Sigma$ refers to this global covariance, while $\Sigma_i$ denotes its $i$-th block. We then project the uncertainty of the 3DGS parameters into pixel space and describe how it interacts with a soft object mask. Here, we present simplified expressions; full derivations and a second-order error bound are provided in the section A of the Supplementary Material.

**Decomposition of Uncertainty Sources.** A key advantage of our explicit, parameter-centric formulation is the ability to decompose the total uncertainty into its underlying physical sources. We can partition the Gaussian parameter vector $\theta_i$ into a **geometry** component, $\theta_i^g = [\mu_i, \mathbf{s}_i, \mathbf{q}_i]^\top$, and an **appearance** component, $\theta_i^a = [\alpha_i, \mathbf{f}_i^{\mathrm{dc}}, \mathbf{f}_i^{\mathrm{sh}}]^\top$. Consequently, the total parameter vector $\theta$ and its Jacobian $J_u$ can be similarly partitioned:

$$\theta = \begin{bmatrix}\theta^g \\ \theta^a\end{bmatrix}, \quad J_u = \begin{bmatrix}J_u^g & J_u^a\end{bmatrix} \tag{4}$$

where $J_u^g = \partial C(u)/\partial\theta^g$ and $J_u^a = \partial C(u)/\partial\theta^a$. Assuming independence between geometric and appearance parameters (a reasonable simplification enforced by our diagonal FIM approximation), the pixel-wise color covariance from Eq. 6 can be expressed as a sum of two distinct sources:

$$\Sigma_C(u) \approx \underbrace{J_u^g \Sigma^g (J_u^g)^\top}_{\text{Geometric Uncertainty}} + \underbrace{J_u^a \Sigma^a (J_u^a)^\top}_{\text{Appearance Uncertainty}} \tag{5}$$

This decomposition is theoretically significant. It allows us to differentiate between uncertainty arising from poorly constrained object geometry (e.g., ambiguous boundaries, fine structures) and uncertainty from poorly observed appearance (e.g., complex materials, view-dependent effects). Implicit methods like FisherRF, which operate on abstract network weights, lack this inherent interpretability.

**Pixel-wise uncertainty** Assume complete set of scene parameters $\theta = \{\theta_i\}$ and $\theta^\star$ is the MAP estimate after optimization. For small parameter perturbations $\delta\theta = \theta - \theta^\star$, the change in pixel color $\delta C(u)$ can be linearly approximated using a first-order Taylor expansion:

$$\delta C(u) \approx J_u \delta\theta.$$

where the Jacobian $J_u = \partial C(u;\theta)/\partial\theta \in \mathbb{R}^{3\times d}$. Given $\mathbb{E}[\delta\theta] = 0$ under a prior normal distribution, the induced pixel-colour covariance can be written as

$$\Sigma_C(u) = \mathrm{Var}[C(u;\theta)] \;\approx\; J_u \Sigma J_u^\top \tag{6}$$

where $\Sigma$ is the full parameter covariance. Eq 6 shows that the Jacobian acts as a lever arm that magnifies (or attenuates) each parameter's uncertainty in proportion to that parameter's influence on the pixel.

**Pixel-wise object-aware uncertainty** To estimate the uncertainty of a specific object $k$, we introduce a soft mask $M_k(u) \in [0,1]$ based on semantic probabilities. This allows us to define an object-specific pixel covariance $\Sigma_{C,k}(u)$ by masking the standard error propagation formula:

$$\Sigma_{C,k}(u) \;=\; \big(M_k(u)\big)^2 \left(J_u \Sigma J_u^\top\right) \tag{7}$$

The mask term $M_k(u)$ is squared because covariance propagates quadratically via the Jacobian and its transpose.

### 3.3. Updating Uncertainty With FIM

While our formulation provides a physically interpretable model of uncertainty, direct computation of the full covariance matrix $\Sigma$ is intractable. We therefore approximate it with the inverse of the Fisher Information Matrix (FIM), $\Sigma \approx \sigma^2 \mathcal{I}^{-1}$ [LMV*17]. Crucially, our FIM is defined over the space of the 3D Gaussians' physical parameters, capturing how perturbations in geometry and appearance affect the rendered output. This stands in contrast to implicit-representation methods where the FIM is computed over abstract neural network weights.

**Online Diagonal Approximation.** To ensure computational tractability, we make a key simplifying assumption: we approximate the full FIM with its diagonal entries only, effectively assuming that the different physical parameters of a Gaussian are locally independent. This diagonal approximation, $\mathcal{I} \approx \mathrm{diag}(\mathcal{I})$, aligns perfectly with the uncertainty decomposition presented in Eq. 5. It implies that a Gaussian's geometric uncertainty is decoupled from its appearance uncertainty. While this is a strong simplification, it is a common and effective strategy that allows us to efficiently estimate the parameter-wise variances. We update the diagonal entries $\mathcal{I}_{t,i}$ online during training using an exponential moving average (EMA) of squared gradients [KB17]:

$$\mathcal{I}_{t,i} = \alpha_t \mathcal{I}_{t-1,i} + (1-\alpha_t)\left[\nabla_{\theta_i}\ell_t\right]^2, \tag{8}$$

where $\ell_t$ denotes the photometric reconstruction loss at iteration $t$, $\ell_t = \frac{1}{2\sigma^2}\sum_{u\in\Omega_t}\|C_{\mathrm{pred}}(u) - C_{\mathrm{gt}}(u)\|_2^2$ computed on the sampled pixels (or patches) at step $t$, and $[\cdot]^2$ is element-wise. We employ a time-dependent momentum schedule $\alpha_t = 0.95^{1.5t/T}$. This schedule gradually decreases the momentum over iterations (from 1 at $t$=0 to $\approx 0.926$ at $t$=$T$), increasing the contribution of newly observed gradients as optimization stabilizes.

**Object-Aware Uncertainty Score.** Substituting the diagonal approximation into Eq. 7 yields:

$$\Sigma_{C,k}(u) = \big(M_k(u)\big)^2 J_u \big(\mathrm{diag}\,\mathcal{I}_t + \lambda I\big)^{-1} J_u^\top, \tag{9}$$

where $\lambda$ is a small damping factor to ensure numerical stability.

Note that unlike expected information gain objectives that quantify parameter-space entropy reduction, our score sums pixel-space predictive variances, enabling direct spatial modulation via object masks. Key differences from FisherRF [JLD25]: (1) we project parameter covariance into pixel space, allowing native 2D mask integration for object-centric planning; (2) we maintain an online EMA estimate during training rather than snapshot-based gradient evaluation, stabilizing uncertainty under stochastic mini-batches.

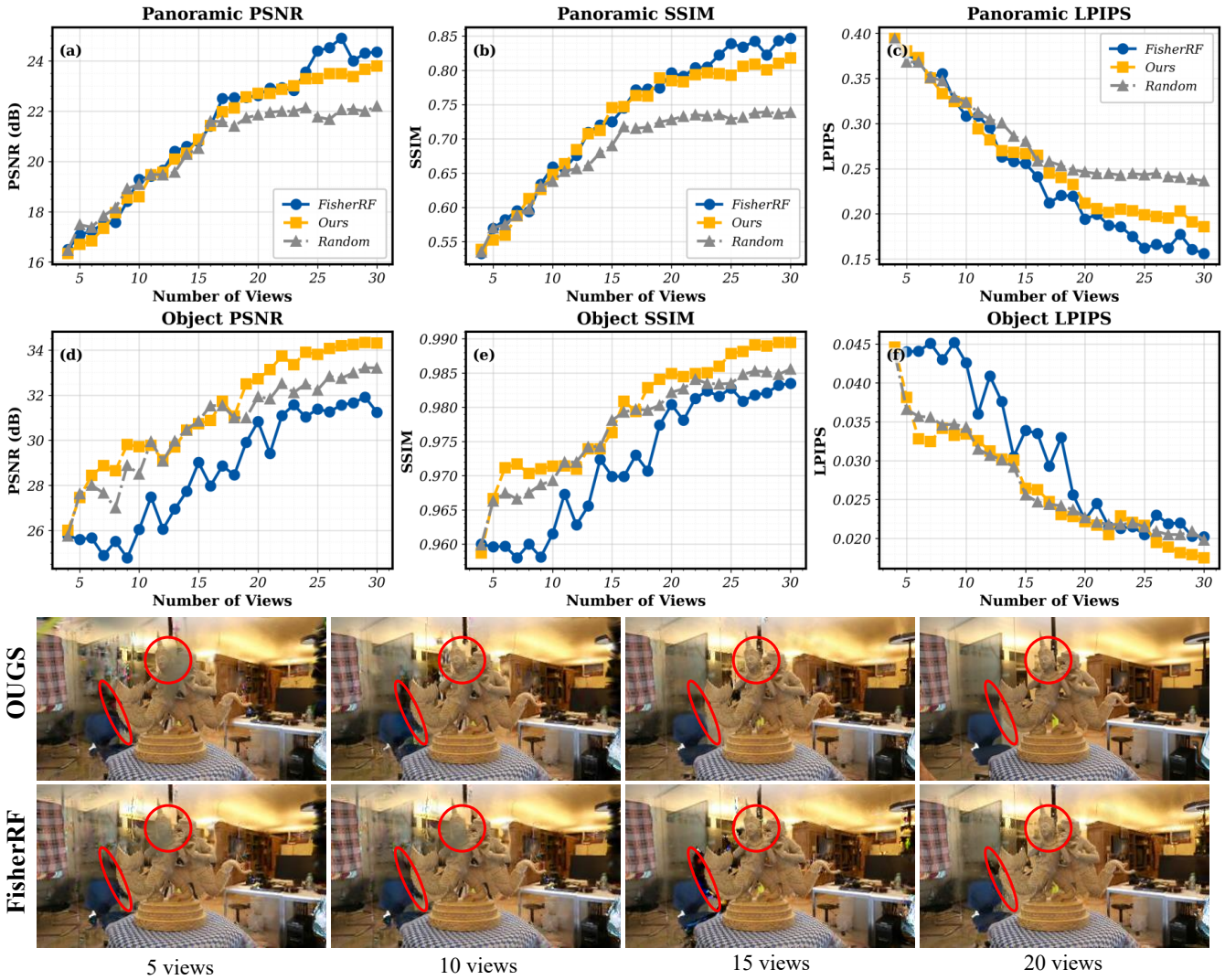


Figure 4: **Object-aware approach speeds up convergence.** *Curves are recorded on the statue scene of the LF dataset as new views are added. Top: panoramic PSNR/SSIM/LPIPS; middle: the same metrics evaluated only inside the object. Bottom: visual progression at 5, 10, 15 and 20 views; red circles mark regions where the competing method keeps struggling while our reconstruction sharpens steadily.*

## 4. Experiments

### 4.1. Experimental Setup

**Dataset.** We conduct our evaluation across three public datasets to ensure coverage of diverse capture scenarios. **Mip-NeRF 360** [BMV*22] comprises four bounded indoor scenes with strong specular clutter and five unbounded outdoor environments featuring foliage occlusion and high dynamic range lighting. Following the standard 3DGS protocol [KKLD23], every 8$^{th}$ view is held out for testing. We also evaluate on the **Light-Field** dataset [YSHWSH16], which features four tabletop objects (*torch*, *statue*, *basket*, and *africa*) captured via a motorised gantry. A salient characteristic of this dataset is that the target object remains centred across all views. This yields stable masks from SAM-2, making it an ideal testbed for evaluating per-object uncertainty calibration. Additionally, we select the *train* and *truck* scenes from **Tanks & Temples** [KPZK17]. These large-scale, drone-style outdoor captures are characterised by long-baseline parallax and strong depth discontinuities. Inspired by the sparse-view protocol of Shen [SAMNR22], we adopt an intentionally biased and imbalanced camera orbit configuration. Such a setup exacerbates reconstruction artefacts, serving as a rigorous stress test for our Fisher-based uncertainty modelling. To preserve fine geometric cues, all RGB frames are processed at their full original resolution without cropping.

**Baselines.** We compare our approach against several state-of-the-art methods that explicitly model uncertainty for *next-best-view* (NBV) selection. Our baseline pool includes: *ActiveNeRF* [PLSH22], *FisherRF* [JLD25], and *Bayes' Rays* [GRS*24]. Furthermore, to ensure a thorough comparison, we include ***GauSS-MI*** [XCZ*25], a recent information-theoretic method that represents the state-of-the-art in scene-level active reconstruction. We also note the concurrent work ***POp-GS*** [WAM*25], which also fo-

Table 1: **Active View Selection on Mip-NeRF360, Tanks&Temples, and LF datasets**. *"Panoramic" evaluates the full image; "Object-aware" evaluates only inside the object mask. Rows denote selection policies.*

| Method \ Metrics | Mip-NeRF360 | | | LF | | | Tanks & Temples | | |
|---|---|---|---|---|---|---|---|---|---|
| | PSNR ↑ | SSIM ↑ | LPIPS ↓ | PSNR ↑ | SSIM ↑ | LPIPS ↓ | PSNR ↑ | SSIM ↑ | LPIPS ↓ |
| Random | 17.9140 | 0.5640 | 0.4300 | 19.0857 | 0.6646 | 0.2669 | 15.8784 | 0.5365 | 0.3720 |
| ActiveNeRF | 17.8890 | 0.5330 | 0.4140 | 21.2263 | 0.7691 | 0.1742 | 16.2918 | 0.5892 | 0.2514 |
| BayesRays | 18.8120 | 0.5730 | 0.4210 | 21.9232 | 0.7628 | 0.1752 | 16.9322 | 0.6091 | 0.3252 |
| FisherRF | 20.3510 | 0.6010 | 0.3610 | 23.6450 | 0.8323 | 0.1651 | 17.3684 | 0.6296 | 0.3091 |
| GauSS-MI | 20.8150 | 0.6433 | 0.2710 | 23.9820 | 0.8354 | 0.1628 | 17.4210 | 0.6315 | 0.2425 |
| OUGS | 20.6099 | 0.6453 | 0.2727 | 23.7014 | 0.8058 | 0.1726 | 17.2666 | 0.6125 | 0.2468 |
| **Object-Aware** | | | | | | | | | |
| Random | 26.3382 | 0.9732 | 0.0266 | 31.0709 | 0.9866 | 0.0227 | 24.6778 | 0.9306 | 0.1349 |
| FisherRF | 26.4312 | 0.9731 | 0.0276 | 30.82740 | 0.9813 | 0.0231 | 23.4830 | 0.9208 | 0.1396 |
| GauSS-MI | 27.3012 | 0.9764 | 0.0258 | 30.9742 | 0.9830 | 0.0226 | 24.9832 | 0.9318 | 0.1336 |
| OUGS | 29.6099 | 0.9813 | 0.0241 | 32.1856 | 0.9888 | 0.0221 | 26.2533 | 0.9333 | 0.1169 |

cuses on NBV selection. However, as its implementation was not publicly available during our experimental phase, a direct quantitative comparison was not feasible. All baselines are trained using their official repositories with default hyper-parameters. To control for segmentation bias, we feed identical SAM-2 object masks to every method where applicable. NBV selection is executed with identical setting to ensure a fair comparison.

**Metrics.** We evaluate reconstruction quality at two levels: **Scene-Level (Panoramic)** and **Object-Level**. Panoramic metrics, including Peak Signal-to-Noise Ratio (PSNR) [GW08] for reconstruction accuracy, Structural Similarity Index (SSIM) [WBSS04] for structural fidelity, and Learned Perceptual Image Patch Similarity (LPIPS) [ZIE*18] for perceptual quality, are computed on the full, unmasked image to assess global consistency. Object-Level metrics are restricted strictly to the masked region of interest. Following the experimental protocol of [GRS*24], we calibrate our predictive uncertainty using the Area Under the Sparsification Error curve (AUSE). For every test image, we compute the per-pixel absolute error and the corresponding predicted uncertainty, then iteratively mask out the top $t\%$ of pixels ($t = 1,\ldots,100$) in descending order of predicted uncertainty. Integrating the resulting error curve yields $\text{AUSE}_{\text{MAE}}$ and $\text{AUSE}_{\text{MSE}}$, where lower values indicate better alignment between predicted uncertainty and actual reconstruction error.

**Implementation Details.** Following SoTA NBV protocols [WAM*25], we evaluate on the benchmark datasets described in Sec. 4.1. The Gaussians are initialised with COLMAP [SF16], and object masks are obtained from SAM-2 [RGH*24]. NBV planning follows FisherRF [JLD25]: four initial views are selected using the farthest-point strategy, followed by 100 epochs of training and the addition of a new view chosen by the highest predicted uncertainty within the object mask. The selection process repeats until 20 views are reached, after which the model is optimised for 30k iterations with the default 3DGS schedule. To ensure practical scalability, OUGS incorporates strategic optimizations to minimize the computational overhead of view selection. The main overhead arises during NBV planning, where we evaluate the explicit Jacobian projection $J_u = \partial C(u)/\partial\theta$ for candidate views (Eq. 9); computing this densely for every pixel at full resolution would be prohibitive. We therefore combine a diagonal FIM approximation, which reduces covariance storage and propagation from $O(d^2)$ to $O(d)$, with a strided patch-based sampling scheme during planning: we score candidate views strictly within the masked region of interest using uniformly sampled $16\times16$ patches at a stride of 16 pixels. This strategy ignores vast background areas and reduces the evaluated pixel count by $\sim 256\times$ (relative to the mask) while retaining sufficient spatial coverage for reliable ranking. These choices keep uncertainty tracking lightweight during training and make Jacobian-based projection practical at selection time, enabling scalability to large scenes (validated on TNT sequences with $>$ 1.2M Gaussians).

### 4.2. Quantitative Results

Our primary quantitative evaluation, presented in table 1, analyzes reconstruction quality at two levels of granularity: the full scene and the object of interest.

**Scene-Level Panoramic Performance.** In the panoramic evaluation, which assesses the entire rendered view, GauSS-MI demonstrates strong performance due to its entropy-based global objective. OUGS remains highly competitive, consistently outperforming FisherRF and ranking second only to GauSS-MI in our benchmarks. This reflects an intentional design choice: we prioritize robust global consistency to prevent catastrophic background degradation, while strategically reserving the active selection budget for the target object. Notably, panoramic metrics are computed on the full image without masking; moreover, all methods receive identical masks during view selection, so improvements over FisherRF cannot be attributed to privileged semantic access.

**Object-Aware Performance.** The decisive advantage of our framework is revealed in the object-centric evaluation. In this critical setting, **OUGS consistently and substantially outperforms all baselines, including GauSS-MI, across every dataset, table 2**. Crucially, employing **identical SAM-2 masks for all baselines** eliminates segmentation quality as a confounding variable. The results validate our core hypothesis: while scene-level methods dilute their view budget on high-gradient background textures, our

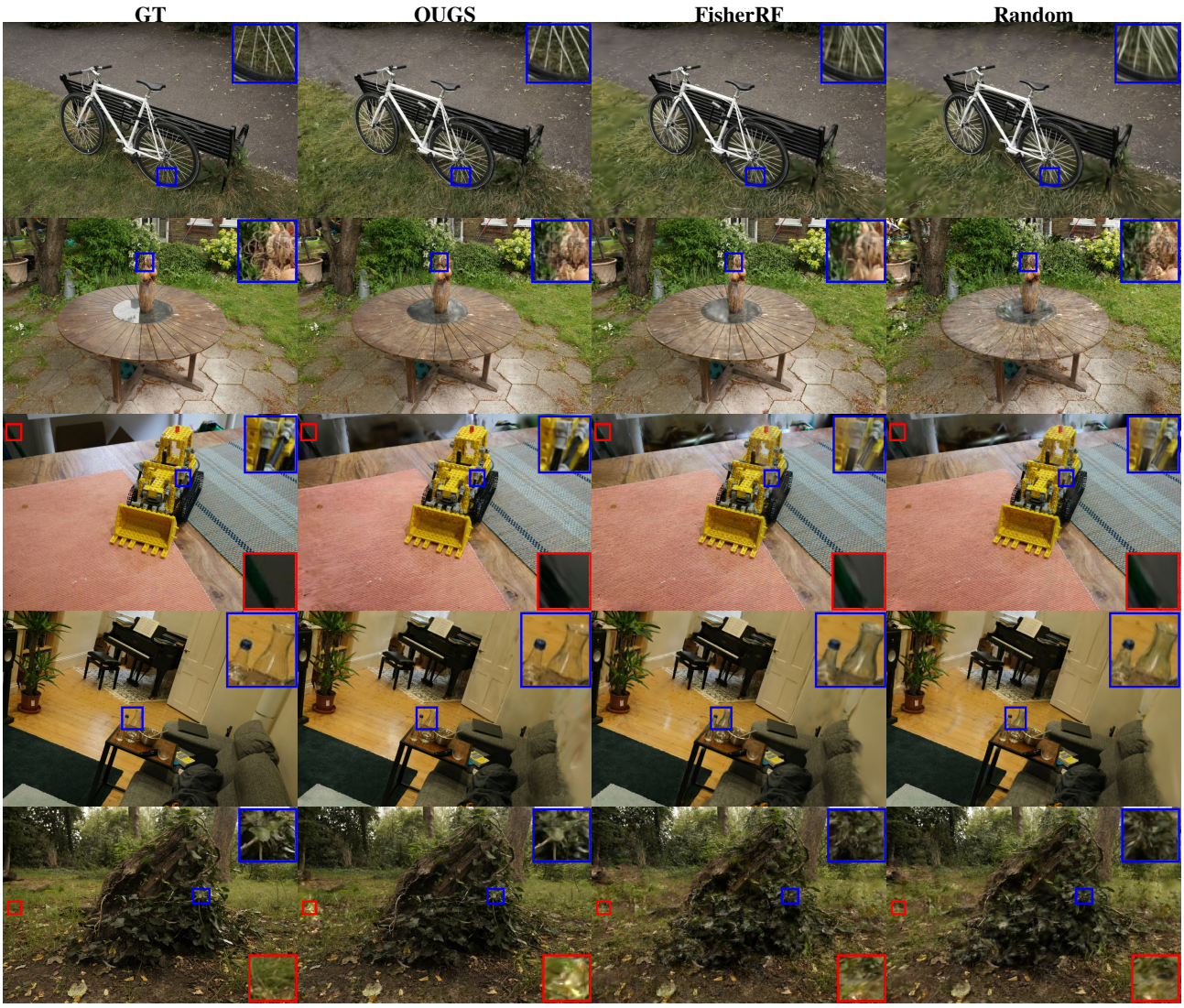


Figure 5: **Qualitative result on Mip-NeRF360.** *From left to right are ground truth, Ours (OUGS), FisheRF, and Random. Each row corresponds to a different scene.* ***20*** *views were selected to train a model and render the result on the test set. The blue box circles the object of our interest, while the red box circles some of the background.*

method successfully **reallocates the finite view budget** to the target, achieving significantly higher fidelity on the regions of interest.

### 4.3. Qualitative Results

Figure 4 illustrates the convergence behavior, comparing our approach against the FisherRF. While both methods track closely in the panoramic evaluation, OUGS achieves sharp gains in PSNR and SSIM early in the process by prioritizing object centric views, whereas FisherRF's progress on the object is slower due to distractions from background gradients. In addition, Figure 5 provides a qualitative comparison on the Mip-NeRF 360 dataset. To specifically analyze the behavior of uncertainty estimation based on the Fisher Information Matrix, we focus the visual comparison on our method, FisherRF, and a random baseline.

The results clearly illustrate the impact of our object-aware strategy. Across all scenes, our reconstructions remain visually closest to the ground truth in the regions of interest (blue boxes). On the *stump* scene, for example, the slender bark fibres are sharply delineated in our result, whereas they are rendered as blurred or are entirely missing by FisherRF. This reveals the fundamental difference in FIM-based approaches: FisherRF, however, often produces a cleaner background (red boxes). This stems from its scene-level Fisher information score; high-gradient background textures can dominate the score and steer the next-best-view search away from the object. Our method, by design, resists this distraction. Random sampling, while uninformed, spreads views uniformly and therefore sometimes captures object-centric angles, leading to occasional details that surpass FisherRF, but this comes at the cost of high variance and no guarantees. The visual evidence strongly indicates that our method preserves object detail most faithfully.

Table 2: **Validation of our parameter-centric FIM as a standalone uncertainty estimator.** *To isolate the quality of our core uncertainty model, we evaluate it on the full scene without any object-aware masks. The table reports the Area Under the Sparsification Error (AUSE), a metric for uncertainty quality (lower is better).*

| | africa | | basket | | statue | | torch | | TNT-Train | | TNT-Truck | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ΔMSE↓ | ΔMAE↓ | ΔMSE↓ | ΔMAE↓ | ΔMSE↓ | ΔMAE↓ | ΔMSE↓ | ΔMAE↓ | MSE↓ | MAE↓ | MSE↓ | MAE↓ |
| ActiveNeRF | 1.123 | 0.958 | 0.642 | 0.546 | 0.818 | 0.732 | 1.513 | 1.246 | 1.279 | 1.076 | 0.994 | 0.438 |
| Bayes' Rays | 0.445 | 0.271 | 0.326 | 0.284 | 0.192 | 0.182 | 0.342 | 0.224 | 0.822 | 0.689 | 0.865 | 0.529 |
| FisherRF | 0.181 | 0.186 | 0.212 | 0.225 | 0.191 | 0.178 | 0.247 | 0.254 | 0.892 | 0.632 | 0.843 | 0.589 |
| OUGS | 0.192 | 0.187 | 0.122 | 0.131 | 0.181 | 0.181 | 0.248 | 0.217 | 0.787 | 0.589 | 0.651 | 0.487 |

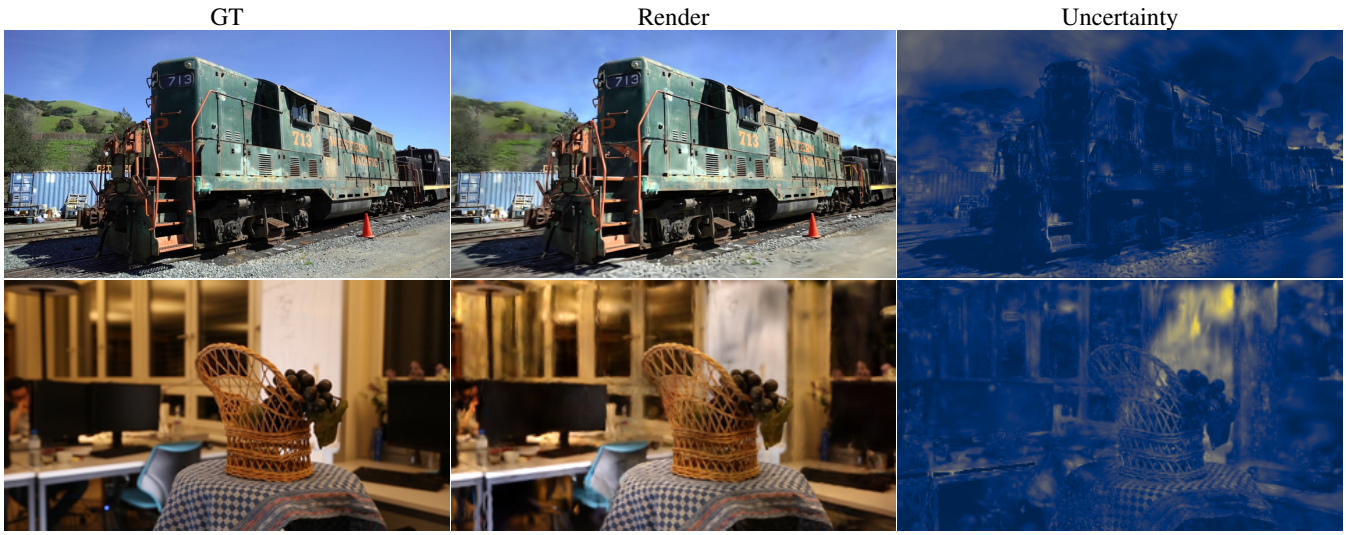


Figure 6: **Qualitative validation of our parameter-centric uncertainty.** *These results on the TNT-Train and LF-Basket scenes showcase the remarkable accuracy of our uncertainty estimation. The uncertainty heatmap (right, yellow indicates high uncertainty) precisely localizes the regions where the final rendering (middle) deviates from the ground truth (left), such as blurry structures and ghosting artifacts. This strong visual correlation demonstrates the effectiveness of our physically-grounded model in predicting and explaining rendering errors.*

The modest artifacts that may appear at the image periphery do not outweigh the substantial and consistent gains in the target region, confirming the effectiveness of our object-aware formulation.

### 4.4. Ablation Study

In this section, we conduct a rigorous analysis to isolate the sources of our performance gains, verify robustness against segmentation failures, and quantify EMA update strategy.

**Quantitative Validation. As reported in table 2, OUGS demonstrates robust uncertainty calibration. While performing on par with FisherRF [JLD25] on simpler, foreground-centric scenes (e.g., *Africa*, *Torch*), OUGS achieves superior calibration on geometrically complex environments. Notably, on the challenging *TNT-Train* and *TNT-Truck* scenes, our method reduces AUSE significantly compared to the baselines. This confirms that our parameter-to-pixel Jacobian formulation provides a more accurate proxy for reconstruction error, particularly in complex scenarios. The semantic mask therefore acts as a spatial filter, but the high-quality uncertainty signal is intrinsic to our FIM formulation.**

**Qualitative Visualization.** Figure 6 complements these findings. The uncertainty highlighted by our model precisely concentrates on background regions that later exhibit blur or ghosting artifacts, while high-confidence areas remain artifact-free. This visual alignment confirms that our parameter-level uncertainty estimation successfully localizes residual errors in image space.

**Source of Gains: Intrinsic Uncertainty Quality** A core question raised in our analysis is whether our gains stem solely from object masking or from a fundamental improvement in uncertainty estimation. To isolate the quality of our core uncertainty model from semantic guidance, we refer to the **Area Under the Sparsification Error (AUSE)** evaluated on the **full scene** without applying any object masks in table 2.

**Robustness to Imperfect Segmentation** Since our method leverages semantic masks for object targeting, a critical question is how robust OUGS is to segmentation errors. We analyze this in two dimensions: sensitivity to thresholding parameters and resilience to severe segmentation noise.

**Sensitivity to Probability Thresholds.** To investigate how the quality of the soft mask influences our object-aware uncertainty estimation, we simulate mask degradation by varying the threshold value from 0.1 to 0.9 and plot the corresponding object level AUSE scores. Pixels with probabilities above the threshold are considered

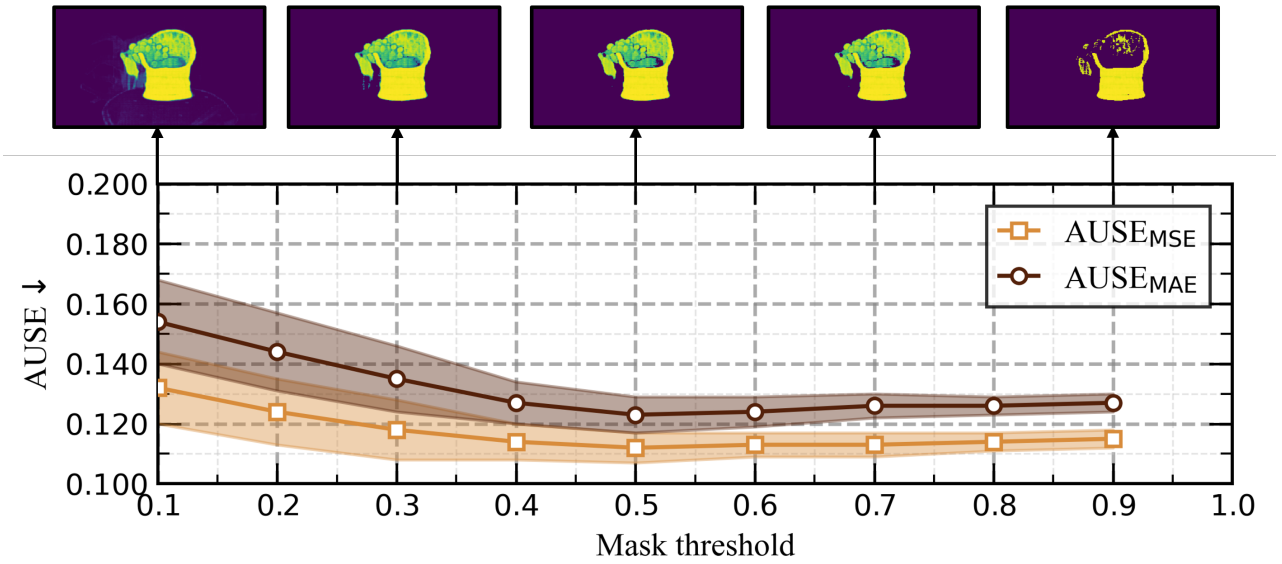


Figure 7: **Sensitivity to mask quality.** *We analyze object-level AUSE (lower is better) by varying the mask binarization threshold. The clear basin of stability indicates that OUGS is robust to a wide range of segmentation hyperparameters and effectively utilizes the mask to filter background clutter without requiring pixel-perfect thresholds.*

part of the object. At low threshold values (left side of the plot), the mask includes more unwanted background regions, leading to suboptimal view selection and higher (worse) AUSE scores. As the threshold increases, the mask becomes more focused on the object, improving performance and reaching an optimal AUSE around 0.5—where the mask best isolates the object while excluding background clutter. Beyond this point, further increases in the threshold aggressively remove less salient object regions, slightly degrading performance as informative areas are excluded.

**Robustness to Segmentation Noise.** To evaluate robustness under realistic segmentation failures, we stress-test both the *view-selection policy* and the *final object reconstruction* on *TNT–Truck* by injecting three controlled perturbations into the planning masks: **(i) Under-segmentation**, which removes portions of the true foreground; **(ii) Over-segmentation**, which introduces spurious foreground blobs in the background; and **(iii) Boundary Blur**, which dilates the mask to include surrounding clutter.

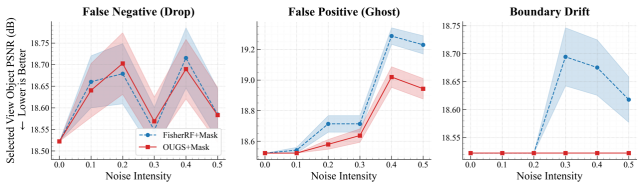


Figure 8: **NBV robustness under segmentation noise (TNT–Truck).** *Selected-view object PSNR (*lower is better*) versus noise intensity for Drop (left), Ghost (center), and Boundary Drift (right).*

Figure 8 analyzes one-step selection robustness via *Selected View Object PSNR* (*lower is better*), as an effective NBV policy should select views where the object is currently worst reconstructed. Under **Under-segmentation** (left), both methods degrade similarly—missing foreground pixels reduce evidence for both scoring rules. Under **Over-segmentation** (center), the baseline increasingly selects already well-reconstructed views (higher PSNR), wasting view budget on background artifacts. Under **Boundary Blur** (right), this gap widens: as masks expand into clutter, the baseline shifts toward "easy" views while OUGS remains stable, demonstrating robustness to coarse boundary errors.

Table 3: **Robustness of object-aware reconstruction on TNT under degraded object masks during NBV planning.**

| Method | PSNR↑ | PSNR Drop↓ | SSIM↑ | LPIPS↓ |
|---|---|---|---|---|
| *I. Clean masks (ideal)* | | | | |
| FisherRF | 23.48 | — | 0.921 | 0.140 |
| **OUGS (ours)** | **26.25** | — | **0.933** | **0.117** |
| *II. "Ghost" noise (false positives)* | | | | |
| FisherRF | 20.15 | 3.33 | 0.845 | 0.210 |
| **OUGS (ours)** | **24.05** | **2.20** | **0.895** | **0.152** |
| *III. "Boundary drift" (coarse edges)* | | | | |
| FisherRF | 21.60 | 1.88 | 0.872 | 0.185 |
| **OUGS (ours)** | **24.45** | **1.80** | **0.902** | **0.141** |

Table 3 confirms this selection stability translates to end-to-end robustness. Under **Over-segmentation**, FisherRF drops 3.33,dB by overfitting to background artifacts, while OUGS degrades only 2.20,dB. This resilience stems from Eq. 9: our uncertainty score combines semantic mask weights with physical uncertainty from the Jacobian. False positive regions typically have low geometric uncertainty, suppressing their contribution and acting as a physics-based safeguard against mask noise.

**Reliability in Fully Autonomous Settings.** A key challenge in object-centric reconstruction is obtaining object masks for unseen candidate views. Beyond offline SAM2-based propagation, we

Table 4: **Quantitative comparison across four scenes of the LF dataset.** *We compare our Warp-based mask generation against ground-truth SAM2 masks (GT). Metrics are evaluated on object regions after training with 20 views for 30k iterations.*

| | africa | | | basket | | | statue | | | torch | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ |
| **Online FisherRF** | 24.80 | 0.975 | 0.018 | 24.95 | 0.958 | 0.048 | 27.60 | 0.972 | 0.032 | 22.60 | 0.958 | 0.029 |
| **Offline FisherRF** | 32.45 | 0.991 | 0.011 | 29.35 | 0.973 | 0.038 | 31.85 | 0.983 | 0.024 | 29.66 | 0.978 | 0.020 |
| **Online OUGS** | 27.94 | 0.988 | 0.014 | 27.35 | 0.968 | 0.041 | 30.02 | 0.981 | 0.023 | 25.63 | 0.969 | 0.025 |
| **Offline OUGS** | 33.82 | 0.996 | 0.010 | 30.65 | 0.982 | 0.036 | 33.25 | 0.989 | 0.022 | 31.02 | 0.988 | 0.019 |

evaluate an online depth-guided warping strategy that uses the current 3DGS geometry to transfer masks to unvisited viewpoints, but its quality is limited by view coverage and depth fidelity. Table 4 shows that this online variant underperforms the SAM2 baseline; the gap is most pronounced at the beginning of scanning, where limited view coverage yields noisier depth-warped masks and consequently less effective NBV decisions. Although mask quality improves over NBV iterations as coverage increases, the view budget consumed early cannot be recovered, leading to lower final reconstruction quality. This trend aligns with Table 3, where degraded masks (e.g., projection-induced boundary drift due to geometric errors) measurably harm planning; we therefore recommend high-quality mask acquisition (e.g., SAM2) whenever available.

### Analysis of the EMA Update Schedule

Our FIM approximation relies on an online update of the diagonal entries using an EMA of squared gradients. A key design choice is our decaying momentum schedule for the EMA parameter, $\alpha_t = 0.95^{1.5t/T}$, which prioritizes stability early in training and faster adaptation later on. To validate this choice, we conduct an ablation study comparing our proposed schedule against simpler alternatives with a constant momentum. We evaluate the final object-aware PSNR on the LF scene after 20 selected views.

Table 5: **Ablation on the EMA update schedule, evaluated on the LF-Statue scene.** *Our decaying momentum strategy outperforms all constant momentum alternatives.*

| **EMA Schedule Strategy** | **Object-Aware PSNR (dB) ↑** |
|---|---|
| No EMA (Instantaneous) | 29.52 |
| Low Momentum ($\alpha_t = 0.90$) | 31.63 |
| Medium Momentum ($\alpha_t = 0.95$) | 31.91 |
| High Momentum ($\alpha_t = 0.99$) | 32.04 |
| **Ours (Decaying Momentum)** | **32.18** |

The results, presented in table 5, confirm the effectiveness of our proposed strategy. A constant high momentum ($\alpha_t = 0.99$) is overly cautious, smoothing too much and preventing the model from adapting quickly enough, resulting in lower PSNR. Conversely, a constant low momentum ($\alpha_t = 0.9$) is too reactive to noisy gradients, leading to an unstable FIM estimate and the worst performance. Our decaying schedule strikes the optimal balance, achieving the highest PSNR. This validates that our carefully designed "slow-start, fast-finish" update strategy is crucial for robustly estimating the FIM online and contributes significantly to the final reconstruction quality.

## 5. Limitations and Future Work

While our framework demonstrates a significant advancement in object-centric active reconstruction, we acknowledge several limitations that open up exciting avenues for future research. We employ SAM-2 as an offline oracle to simulate reliable masks, decoupling planning efficacy from segmentation failures. Consequently, our method's effectiveness is contingent upon the availability and quality of a semantic mask for the object of interest. While our ablation study (Sec. 4.4) demonstrates robustness to moderate mask degradation, significant segmentation errors inevitably degrade reconstruction quality; we observe that severe noise such as spurious Ghost regions can cause performance drops comparable to baselines. Additionally, while we extend this method to handle online mask estimation (e.g., via depth image warping), achieving high-fidelity reconstruction without oracle masks still remains a quality challenge. While we present a theoretical extension to multiple objects in appendix, which is further detailed in the Supplementary Material, the experimental evaluation in this work focuses on single-object scenarios; empirical validation of multi-object active reconstruction and optimal weighting strategies remain to be explored. Furthermore, our method approximates the full Fisher Information Matrix with its diagonal entries, assuming independence between different parameters of a Gaussian—a strong simplification that could be relaxed with more structured FIM approximations.

## 6. Conclusion

We introduced OUGS, an object-aware uncertainty estimation framework for active view selection in 3DGS. Our work presents a fundamental shift in how uncertainty is modeled for explicit representations. By deriving uncertainty directly from the physical parameters of the 3D Gaussian primitives, we establish a more principled and interpretable link between the 3D scene representation and the rendered 2D image. Our method projects this parameter-level covariance into pixel space and, by coupling it with semantic masks, produces an uncertainty score that effectively disentangles the object of interest from its environment. This is enabled by an efficient diagonal FIM update scheme that makes the approach computationally tractable. When integrated into a next-best-view loop, our method consistently and substantially outperforms state-of-the-art baselines in the critical task of object-centric reconstruction, achieving sharper results under the same view budget. Notably, our underlying uncertainty model also proves to be highly competitive for global scene reconstruction. Ultimately, these results underscore the importance of disentangling object-level uncertainty from background clutter for efficient, high-fidelity active reconstruction.

## References


[AVM25] AIRA L. S., VALSESIA D., MAGLI E.: Modeling uncertainty for gaussian splatting, 2025. doi:10.1109/TNNLS.2025.3553582. 2

[BKA*16] BIRCHER A., KAMEL M., ALEXIS K., OLEYNIKOVA H., SIEGWART R.: Receding horizon "next-best-view" planner for 3d exploration. In *2016 IEEE International Conference on Robotics and Automation (ICRA)* (2016), pp. 1462–1468. doi:10.1109/ICRA.2016.7487281. 3

[BMV*22] BARRON J. T., MILDENHALL B., VERBIN D., SRINIVASAN P. P., HEDMAN P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 5460–5469. doi:10.1109/CVPR52688.2022.00539. 5

[CKD*25] CELAREK A., KOPANAS G., DRETTAKIS G., WIMMER M., KERBL B.: Does 3d gaussian splatting need accurate volumetric rendering?, 2025. URL: https://arxiv.org/abs/2502.19318, arXiv:2502.19318. 1

[CLDC23] CORTES C. A. T., LIN C.-T., DO T.-T. N., CHEN H.-T.: An eeg-based experiment on vr sickness and postural instability while walking in virtual environments. In *IEEE Conference Virtual Reality and 3D User Interfaces (VR)* (2023), p. 8. 2

[CLK11] CHEN S., LI Y., KWOK N. M.: Active vision in robotic systems: A survey of recent developments. *The International Journal of Robotics Research 30*, 11 (2011), 1343–1377. 1, 3

[GRS*24] GOLI L., READING C., SELLÁN S., JACOBSON A., TAGLIASACCHI A.: Bayes' Rays: Uncertainty quantification in neural radiance fields. *CVPR* (2024). 1, 5, 6

[GW08] GONZALEZ R. C., WOODS R. E.: *Digital Image Processing*, 3rd ed. Pearson, 2008. 6

[HD25] HAN C., DUMERY C.: View-dependent uncertainty estimation of 3d gaussian splatting, 2025. URL: https://arxiv.org/abs/2504.07370, arXiv:2504.07370. 1, 2

[JLD25] JIANG W., LEI B., DANIILIDIS K.: Fisherrf: Active view selection and mapping with radiance fields using fisher information. In *Computer Vision – ECCV 2024* (Cham, 2025), Leonardis A., Ricci E., Roth S., Russakovsky O., Sattler T., Varol G., (Eds.), Springer Nature Switzerland, pp. 422–440. 1, 2, 3, 4, 5, 6, 8

[KB17] KINGMA D. P., BA J.: Adam: A method for stochastic optimization, 2017. URL: https://arxiv.org/abs/1412.6980, arXiv:1412.6980. 4

[KKLD23] KERBL B., KOPANAS G., LEIMKÜHLER T., DRETTAKIS G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics 42*, 4 (July 2023). URL: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/. 1, 3, 5

[KMKS25] KLASSON M., MEREU R., KANNALA J., SOLIN A.: Sources of uncertainty in 3d scene reconstruction, 2025. 1

[KPZK17] KNAPITSCH A., PARK J., ZHOU Q.-Y., KOLTUN V.: Tanks and temples: benchmarking large-scale scene reconstruction. *ACM Trans. Graph. 36*, 4 (July 2017). URL: https://doi.org/10.1145/3072959.3073599. 5

[KSH22] KIM M., SEO S., HAN B.: Infonerf: Ray entropy minimization for few-shot neural volume rendering. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 12902–12911. doi:10.1109/CVPR52688.2022.01257. 3

[LC24] LI R., CHEUNG Y.-M.: Variational multi-scale representation for estimating uncertainty in 3d gaussian splatting. In *Advances in Neural Information Processing Systems* (2024), Globerson A., Mackey L., Belgrave D., Fan A., Paquet U., Tomczak J., Zhang C., (Eds.), vol. 37, Curran Associates, Inc., pp. 87934–87958. doi:10.52202/079017-2791. 2

[LMV*17] LY A., MARSMAN M., VERHAGEN J., GRASMAN R. P., WAGENMAKERS E.-J.: A tutorial on fisher information. *Journal of Mathematical Psychology 80* (2017), 40–55. doi:https://doi.org/10.1016/j.jmp.2017.05.006. 4

[MST*20] MILDENHALL B., SRINIVASAN P. P., TANCIK M., BARRON J. T., RAMAMOORTHI R., NG R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In *CVPR* (2020), pp. 612–621. 1

[MZC*25] MINTON A. G., ZHU H. Y., CHEN H.-T., WANG Y.-K., ZHUANG Z., NOTARO G., GALVAN R., ALLEN J., ZIEGLER M. D., LIN C.-T.: A longitudinal study on the effects of circadian fatigue on sound source identification and localization using a heads-up display. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (2025), pp. 1–12. 1

[NBM*22] NIEMEYER M., BARRON J. T., MILDENHALL B., SAJJADI M. S. M., GEIGER A., RADWAN N.: Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 5470–5480. doi:10.1109/CVPR52688.2022.00540. 1

[PLSH22] PAN X., LAI Z., SONG S., HUANG G.: Activenerf: Learning where to see with uncertainty estimation. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII* (2022), Springer, pp. 230–246. 3, 5

[RGH*24] RAVI N., GABEUR V., HU Y.-T., HU R., RYALI C., MA T., KHEDR H., RÄDLE R., ROLLAND C., GUSTAFSON L., MINTUN E., PAN J., ALWALA K. V., CARION N., WU C.-Y., GIRSHICK R., DOLLÁR P., FEICHTENHOFER C.: Sam 2: Segment anything in images and videos, 2024. URL: https://arxiv.org/abs/2408.00714. 6

[RZH*23] RAN Y., ZENG J., HE S., CHEN J., LI L., CHEN Y., LEE G., YE Q.: Neurar: Neural uncertainty for autonomous 3d reconstruction with implicit neural representations. *IEEE Robotics and Automation Letters 8*, 2 (Feb. 2023), 1125–1132. URL: http://dx.doi.org/10.1109/LRA.2023.3235686, doi:10.1109/lra.2023.3235686. 3

[SAMNR22] SHEN J., AGUDO A., MORENO-NOGUER F., RUIZ A.: Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In *Computer Vision – ECCV 2022* (Cham, 2022), Avidan S., Brostow G., Cissé M., Farinella G. M., Hassner T., (Eds.), Springer Nature Switzerland, pp. 540–557. 5

[SF16] SCHÖNBERGER J. L., FRAHM J.-M.: Structure-from-motion revisited. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 4104–4113. doi:10.1109/CVPR.2016.445. 6

[STC*25] SHEN S., TAN C. T., CHEN H.-T., RAFFE W. L., LEONG T. W.: Educator perceptions of xrauthor: An accessible tool for authoring learning content with different immersion levels. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (2025), pp. 1–11. 1

[TCZ*25] TAN Z., CHEN X., ZHANG J., FENG L., HU D.: Uncertainty-aware normal-guided gaussian splatting for surface reconstruction from sparse image sequences, 2025. URL: https://arxiv.org/abs/2503.11172, arXiv:2503.11172. 2

[WAM*25] WILSON J., ALMEIDA M., MAHAJAN S., LABRIE M., GHAFFARI M., GHASEMALIZADEH O., SUN M., KUO C.-H., SEN A.: Pop-gs: Next best view in 3d-gaussian splatting with p-optimality. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2025), pp. 3646–3655. doi:10.1109/CVPR52734.2025.00345. 3, 5, 6

[WBSS04] WANG Z., BOVIK A., SHEIKH H., SIMONCELLI E.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (2004), 600–612. doi:10.1109/TIP.2003.819861. 6

[WC23] WU R., CHEN H.-T.: The effect of visual and auditory modality mismatching between distraction and warning on pedestrian street crossing behavior. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2023), p. 8. 2

[XCZ*25] XIE Y., CAI Y., ZHANG Y., YANG L., PAN J.: Gauss-mi: Gaussian splatting shannon mutual information for active 3d reconstruction, 2025. URL: https://arxiv.org/abs/2504.21067, arXiv:2504.21067. 3, 5

[XDM*24] XUE S., DILL J., MATHUR P., DELLAERT F., TSIOTRA P., XU D.: Neural visibility field for uncertainty-driven active mapping. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2024), pp. 18122–18132. doi:10.1109/CVPR52733.2024.01716. 3

[YSHWSH16] YÜCER K., SORKINE-HORNUNG A., WANG O., SORKINE-HORNUNG O.: Efficient 3d object segmentation from densely sampled light fields with applications to 3d reconstruction. *ACM Trans. Graph. 35*, 3 (Mar. 2016). URL: https://doi.org/10.1145/2876504, doi:10.1145/2876504. 5

[ZIE*18] ZHANG R., ISOLA P., EFROS A. A., SHECHTMAN E., WANG O.: The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 586–595. doi:10.1109/CVPR.2018.00068. 6

**Appendix A:** Jacobian–Covariance Propagation

Let $C(u;\theta) \in \mathbb{R}^3$ denote the rendered RGB colour at pixel $u$ under parameters $\theta \in \mathbb{R}^d$, and let $\theta^\star$ be the MAP estimate after optimisation. We model local parameter uncertainty via $\delta\theta := \theta - \theta^\star$ with mean $\mathbb{E}[\delta\theta] = 0$ and covariance $\Sigma = \mathbb{E}[\delta\theta\,\delta\theta^\top]$.

**Second-order expansion.** A Taylor expansion of $C(u;\theta)$ around $\theta^\star$ yields

$$C(u;\theta) = C(u;\theta^\star) + J_u\,\delta\theta + \frac{1}{2}\,\delta\theta^\top H_u\,\delta\theta + \mathcal{O}(\|\delta\theta\|^3), \tag{10}$$

where

$$J_u := \left.\frac{\partial C(u;\theta)}{\partial\theta}\right|_{\theta^\star} \in \mathbb{R}^{3\times d} \quad \text{and} \quad H_u := \left.\frac{\partial^2 C(u;\theta)}{\partial\theta^2}\right|_{\theta^\star} \in \mathbb{R}^{d\times d}.$$

**First-order (Jacobian) covariance.** For uncertainty propagation we use the standard first-order approximation, which is accurate when $\delta\theta$ is small (equivalently, when $\Sigma$ is small in operator/Frobenius norm). Retaining only the linear term in (10) gives

$$C(u;\theta) \;\approx\; C(u;\theta^\star) + J_u\,\delta\theta. \tag{11}$$

Taking the second central moment yields the classic Jacobian–covariance law:

$$\mathrm{Var}\big[C(u;\theta)\big] \;\approx\; J_u\,\Sigma\,J_u^\top. \tag{12}$$

Intuitively, $J_u$ acts as a sensitivity map: its element $[J_u]_{kl} = \partial C_k(u;\theta^\star)/\partial\theta_l$ quantifies how the $l$-th parameter perturbs the $k$-th colour channel at pixel $u$, and (12) weights these sensitivities by parameter variance.

**Remainder control.** Including the quadratic term in (10) contributes corrections of higher order in $\Sigma$. Under mild smoothness, one obtains a bound of the form

$$\big\|\mathrm{Var}[C(u;\theta)] - J_u\Sigma J_u^\top\big\| \;\le\; c_u\,\|\Sigma\|^2 \;+\; \mathcal{O}(\|\Sigma\|^3), \tag{13}$$

Hence the approximation error vanishes as the Gaussian parameters become well-constrained.

**Appendix B:** FIM Approximation and Online Update

**Fisher Matrix under Gaussian Image Noise.** Assume independent Gaussian image noise $\varepsilon(u) \sim \mathcal{N}(0,\sigma^2\mathbf{I}_3)$ so that $C_{\mathrm{gt}}(u) = C(u;\theta) + \varepsilon(u)$. The (per-iteration) negative log-likelihood is the photometric error

$$\ell_t(\theta) = \frac{1}{2\sigma^2}\sum_{u\in\Omega_t}\big\|C_{\mathrm{pred},t}(u;\theta) - C_{\mathrm{gt},t}(u)\big\|_2^2, \tag{14}$$

where $\Omega_t$ denotes the set of sampled pixels (or patches) at step $t$. Let $r_t(u) := C_{\mathrm{pred},t}(u;\theta) - C_{\mathrm{gt},t}(u)$ be the residual. Then

$$\nabla_\theta\ell_t = \frac{1}{\sigma^2}\sum_{u\in\Omega_t} J_u^\top\, r_t(u). \tag{15}$$

The Fisher information matrix (FIM) is defined as the noise expectation of the outer product of the log-likelihood gradients:

$$\mathcal{I} = \mathbb{E}_\varepsilon\big[\nabla_\theta\ell_t\;\nabla_\theta\ell_t^\top\big]. \tag{16}$$

Under the i.i.d. Gaussian noise assumption, cross-terms vanish and $\mathbb{E}[r_t(u)r_t(u)^\top] = \sigma^2\mathbf{I}_3$, yielding

$$\mathcal{I} = \frac{1}{\sigma^2}\sum_{u\in\Omega_t} J_u^\top J_u. \tag{17}$$

Equivalently, stacking per-pixel Jacobians row-wise into $J \in \mathbb{R}^{(3|\Omega_t|)\times d}$ gives $\mathcal{I} = \sigma^{-2}J^\top J$. This dense $d\times d$ matrix captures parameter correlations but is prohibitively expensive to store or invert at scale.

**Diagonal Approximation and Damping.** A local Laplace / Cramér–Rao approximation gives the posterior covariance

$$\Sigma \;\approx\; \sigma^2\big(\mathcal{I} + \lambda\mathbf{I}_d\big)^{-1}, \tag{18}$$

where $\lambda > 0$ is a small Tikhonov damping term for numerical stability. To obtain a tractable estimate, we retain only the diagonal Fisher entries:

$$\mathcal{I}^{\mathrm{diag}} = \mathrm{diag}(\mathcal{I}) = \frac{1}{\sigma^2}\,\mathrm{diag}\Big(\sum_{u\in\Omega_t} J_u^\top J_u\Big), \qquad \Sigma \approx \sigma^2\big(\mathcal{I}^{\mathrm{diag}} + \lambda\mathbf{I}_d\big)^{-1}. \tag{19}$$

This approximation discards off-diagonal correlations while preserving parameter-wise uncertainty magnitudes, which is sufficient for our pixel-space propagation through $J_u$.

**EMA Update and Object-Aware Pixel Covariance.** During optimisation we maintain an online estimate of the diagonal Fisher via an exponential moving average (EMA) of squared gradients:

$$\mathcal{I}_{t,i}^{(j)} = \alpha_t\,\mathcal{I}_{t-1,i}^{(j)} + (1-\alpha_t)\big[\nabla_{\theta_i^{(j)}}\ell_t\big]^2, \qquad \alpha_t = 0.95^{\,1.5\,t/T}. \tag{20}$$

Here $\mathcal{I}_{t,i}^{(j)}$ denotes the $(j,j)$ entry (diagonal element) for the $j$-th parameter of Gaussian $i$, and $T$ is the total number of optimisation steps. The schedule starts at $\alpha_0 = 1$ and decays smoothly, gradually increasing the contribution of new gradient information as optimisation stabilises. With the diagonal Fisher approximation, the pixel-space covariance follows from (12). Introducing a soft object mask $M_k(u) \in [0,1]$ yields the object-aware covariance used in the main paper:

$$\Sigma_{C,k}(u) = (M_k(u))^2\,J_u\Big(\sigma^2\big(\mathcal{I}_t^{\mathrm{diag}} + \lambda\mathbf{I}_d\big)^{-1}\Big)J_u^\top. \tag{21}$$

Although $\mathcal{I}_t^{\mathrm{diag}}$ is diagonal, $J_u$ generally has nonzero entries across many parameters, so the projection couples parameter effects when forming pixel-space uncertainty. Summing $\mathrm{tr}(\Sigma_{C,k}(u))$ over pixels with $M_k(u) > 0$ produces an object-centric uncertainty score that drives next-best-view selection.

**Appendix C:** Multi-Object Extension

While this paper focuses on single-object active reconstruction, the framework naturally extends to multiple targets. For $K$ objects with masks $M_k(u)$ and covariances $\Sigma_{C,k}(u)$, the aggregated score is:

$$\mathrm{Score}(v) = \sum_{k=1}^{K} w_k \cdot \mathrm{Trace}\left(\int_\Omega \Sigma_{C,k}(u)\,du\right), \tag{22}$$

where weights $w_k$ satisfy $\sum_k w_k = 1$. The computational complexity remains $\mathcal{O}(K\cdot d)$ under diagonal FIM approximation. Determining optimal weighting strategies for specific applications (e.g., task-specific priorities, adaptive updates) represents an interesting direction for future work.