DIGS: DIFFUSION PRIORS GUIDED BY INFORMATION GAIN FOR EXTRAPOLATIVE GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

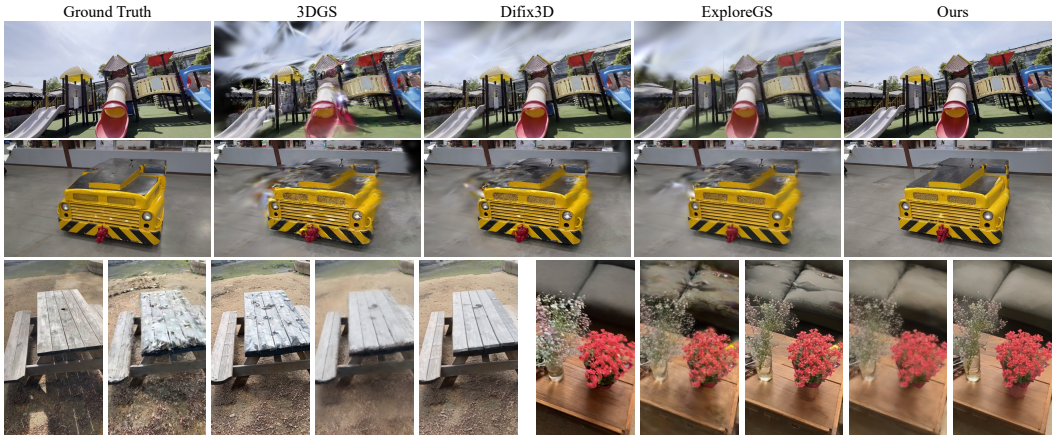

Figure 1: **Visual comparison of reconstruction on extrapolated views.** DIGS (Ours) preserves sharp textures with fewer geometric distortions and blurring artifacts compared to 3DGS (Kerbl et al., 2023), Difix3D (Wu et al., 2025), and ExploreGS (Kim et al., 2025). Guided by reliable information gain, DIGS effectively improves reconstruction quality in under-constrained regions where baseline methods suffer from severe degradation.

ABSTRACT

Recent advances in neural radiance fields and 3D Gaussian Splatting (3DGS) have enabled photorealistic, real-time novel view synthesis. However, reconstruction quality often degrades severely at extrapolated views that deviate from the training views. Existing methods improve extrapolative reconstruction by generating pseudo ground truths at pseudo views with diffusion models. Yet they overlook either the information gain at pseudo views or the reliability of the generated images, both of which are pivotal to effectively integrating diffusion priors. In this work, we introduce DIGS, a framework that integrates **D**iffusion priors guided by **I**nformation gain for extrapolative **G**aussian **S**platting. We first propose a Group-Block-Diagonal (GBD) approximation of the Fisher information matrix to accurately and efficiently estimate the information gain of pseudo views. We further propose Reliable Information Gain (RIG), which accounts for the reliability of the diffusion-generated pseudo ground truths. Building on RIG, DIGS selects informative and reliable pseudo views and optimizes the 3DGS model in an uncertainty-aware manner. Experiments on three challenging benchmarks show that DIGS consistently outperforms existing methods, enabling a more principled and effective integration of diffusion priors into 3DGS.

1 INTRODUCTION

3D scene reconstruction and novel view synthesis (NVS) are fundamental problems in computer vision with wide downstream applications (Chen et al., 2024; Li et al., 2024b; Wang et al., 2024a). Rapid progress has been made in this field since the introduction of Neural Radiance Fields (NeRF) (Mildenhall et al., 2021). 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) further enables high-quality photorealistic rendering in real time. Recent feed-forward 3DGS approaches (Charatan et al., 2024; Ye et al., 2024; Jiang et al., 2025; Ye et al., 2025) reconstruct unseen 3D scenes with a single forward pass by pretraining neural networks on large-scale datasets. Nonetheless, existing evaluations of both standard 3DGS and feed-forward methods often assume interpolated test views uniformly drawn from

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the training view distribution. As shown in Fig. 2(a), the reconstruction quality degrades severely when the test view deviates from the training views (Wu et al., 2025; Kim et al., 2025), which is common in in-the-wild datasets and real-world applications (Warburg et al., 2023; Ling et al., 2024; Kim et al., 2025). Extrapolative 3D reconstruction (Chen et al., 2025; Zhou et al., 2026), i.e., reconstruction under large viewpoint changes, and out-of-distribution settings as shown in Fig. 2(b), remains challenging.

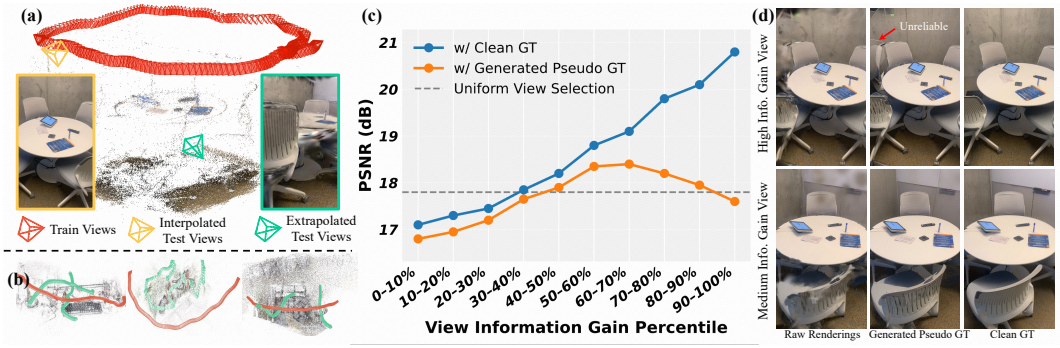

Figure 2: (a) Comparison of interpolated and extrapolated views. (b) Examples of extrapolative reconstruction (Kim et al., 2025). (c) Reconstruction quality when selecting 10% of the candidate views at different information gain percentiles, supervised by clean or generated pseudo ground truths. (d) Views with high information gain can lead to unreliable generated pseudo ground truths.

Recent works (Liu et al., 2024b; 2026; 2024a; Wu et al., 2025; Zhong et al., 2025; Kim et al., 2025; Zhou et al., 2026) attribute the poor performance on extrapolated views to insufficient coverage of the training views, leaving large regions of the 3D model under-constrained. To this end, they propose to incorporate generative priors from diffusion models to enhance these under-constrained regions. More specifically, diffusion priors are injected into the 3D scenes by sampling pseudo views in the 3D space and generating pseudo ground truth images at these views (Wu et al., 2025; Kim et al., 2025; Ni et al., 2026). Then, the 3D model is fine-tuned under the supervision of the generated images, which offers more coverage over under-constrained regions in the initial training views (Wu et al., 2025; Kim et al., 2025). Notably, while feed-forward 3DGS methods can generalize to scenes unseen in the training data, it is still challenging for them to directly generalize to test views that deviate from the training views (Charatan et al., 2024; Jiang et al., 2025; Ye et al., 2025). Diffusion priors are shown to be effective for the feed-forward methods as well (Chen et al., 2026).

While generated images can expand coverage over under-constrained regions (Liu et al., 2024b; Wu et al., 2025), effectively incorporating diffusion priors remains challenging. As shown in Fig. 2(c), pseudo views uniformly sampled by interpolation along predefined trajectories (Liu et al., 2024b; Wu et al., 2025; Liu et al., 2026) may provide only modest improvements. In contrast, selecting views with higher information gain yields larger improvements in reconstruction quality when real captured images are used for supervision (Jiang et al., 2023; Wilson et al., 2025). In practice, however, real captures are unavailable at pseudo views, and supervision often relies on diffusion-generated pseudo ground truths. While some methods (Kim et al., 2025; Ni et al., 2026) propose to integrate diffusion priors at the most informative pseudo views, the reliability of the generated pseudo ground truths is not sufficiently addressed. Highly informative views tend to reveal poorly constrained regions where degraded renderings can lead to unreliable pseudo ground truths, as shown in Fig. 2(d). Consequently, pseudo supervision at the most informative views can yield even smaller improvements than at moderately informative views. Effectively incorporating diffusion priors therefore requires jointly considering the information gain of pseudo views and the reliability of the generated supervision.

In this work, we propose DIGS, a framework that enhances 3DGS with diffusion priors guided by reliable information gain, which considers both the informativeness of pseudo views and the reliability of their generated pseudo ground truths. Our contributions are summarized as follows:

- To quantify the information gain for 3DGS models more accurately, we propose a Group-Block-Diagonal (GBD) approximation of the Fisher information matrix that captures parameter correlations across spatially adjacent Gaussians.
- We propose Reliable Information Gain (RIG), the first to incorporate the reliability of generated pseudo ground truths into information gain estimation. The reliability is assessed by the geometric consistency between the generated content and the training views.

- We propose the DIGS training framework, which leverages RIG to select informative and reliable pseudo views and to optimize the 3DGS model in an uncertainty-aware manner. DIGS is a general framework that can be integrated into existing diffusion-enhanced pipelines.
- Experiments on three extrapolative benchmarks show that DIGS outperforms existing diffusion-enhanced methods, and that GBD leads to better view selection than existing approximations of the Fisher information matrix.

2 Related Work

3D Reconstruction and Novel View Synthesis. Neural Radiance Fields (NeRF) (Mildenhall et al., 2021) and their variants have advanced novel view synthesis by representing 3D scenes as continuous radiance fields parameterized by neural networks (Barron et al., 2021; 2022). However, NeRF-based methods typically suffer from heavy training costs and slow rendering. Recently, 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) has significantly improved efficiency with an explicit representation, enabling real-time rendering while maintaining reconstruction quality. A line of work has been proposed to improve the geometric accuracy and training efficiency of 3DGS by introducing additional regularization terms and pruning model parameters (Guédon & Lepetit, 2024; Huang et al., 2024; Dai et al., 2024; Li et al., 2024a; Hanson et al., 2025). Recent 3D foundation models (Wang et al., 2024b; 2025) have further enabled feed-forward 3DGS (Charatan et al., 2024; Jiang et al., 2025; Chen et al., 2026). Yet both per-scene and feed-forward methods are mostly evaluated on interpolated views, whereas we focus on per-scene 3DGS in the extrapolative setting.

Table 1: Comparison with existing works. "Uncer. Quan." stands for Uncertainty Quantification.

Method	Method Type	Diffusion Prior	Uncer. Quan.	View Selection	Render. Uncer.	Diff. Reliab.
3DGS (Kerbl et al., 2023) SIGGRAPH'23	Per-Scene	✗	✗	✗	✗	✗
3DGS-Enhancer (Liu et al., 2024b) NeurIPS'24	Per-Scene	Video	✗	✗	Scale	✗
ExploreGS (Kim et al., 2025) ICCV'25	Per-Scene	Video	✗	Visibility	LPIPS	✗
G4Splat (Ni et al., 2026) ICLR'26	Per-Scene	Video	✗	Geometry	✗	✗
Difix3D (Wu et al., 2025) CVPR'25	Per-Scene	Image	✗	✗	✗	✗
FreeFix (Zhou et al., 2026) 3DV'26	Per-Scene	Image	✓	✗	Fisher Info.	✗
AnySplat (Jiang et al., 2025) TOG'25	Feed-Forward	✗	✗	✗	✗	✗
YoNoSplat (Ye et al., 2025) ICLR'26	Feed-Forward	✗	✗	✗	✗	✗
UAR-Scenes (Bose et al., 2025) ICCV'25	Feed-Forward	Video	✗	✗	✗	✓
GIFSplat (Chen et al., 2026) CVPR'26	Feed-Forward	Image	✗	✗	✗	✗
Ours	Per-Scene	Video/Image	✓	Fisher Info.	Fisher Info.	✓

Diffusion-Enhanced 3D Reconstruction. To improve reconstruction quality in extrapolative settings, recent works have explored injecting generative priors into 3D reconstruction (Roessle et al., 2023; Liu et al., 2023; 2024b;a; Wu et al., 2025; Kim et al., 2025; Liu et al., 2026; Zhou et al., 2026). Early approaches train a per-scene generative adversarial network, which significantly increases the training cost (Roessle et al., 2023). Recent methods inject diffusion priors in a more general and explicit manner by generating pseudo ground-truth images at pseudo views. For example, 3DGS-Enhancer (Liu et al., 2024b) samples pseudo views by interpolating between training views and generates pseudo images with a video diffusion model (Xing et al., 2024). Difix3D (Wu et al., 2025) further improves efficiency by replacing the video model with a single-step image diffusion model (Sauer et al., 2024). Recent feed-forward 3DGS methods also explore the integration of diffusion priors (Bose et al., 2025; Chen et al., 2026). We leave a more detailed discussion to App. A.

Motivation. As shown in Tab. 1, existing diffusion-enhanced methods do not jointly consider the information gain of pseudo views and the reliability of their pseudo ground truths. First, many methods do not consider the information gain of pseudo views, and simply sample interpolated views along predefined trajectories (Liu et al., 2024b; Wu et al., 2025; Liu et al., 2026). Fisher information has been used to quantify information gain in 3DGS active mapping (Jiang et al., 2023; Wilson et al., 2025), but these methods ignore the inter-Gaussian parameter correlations. Second, while ExploreGS (Kim et al., 2025) and G4Splat (Ni et al., 2026) guide view selection with visibility and geometric cues, which reflect the information gain to some extent, they overlook the reliability of the pseudo ground truths generated at the selected views. As more informative views tend to cover poorly constrained regions where pseudo ground truths are less reliable, the two factors should be considered jointly. DIGS addresses the above challenges with GBD for principled information gain estimation and RIG for incorporating the reliability of pseudo ground truths.

3 METHODOLOGY

Fig. 3 illustrates an overview of DIGS. In Sec. 3.1, we introduce a Group-Block-Diagonal (GBD) approximation of the Fisher information matrix to quantify the information gain of pseudo views for the 3DGS model. Next, in Sec. 3.2, we propose Reliable Information Gain (RIG), which further accounts for the reliability of the pseudo ground truths generated by the diffusion model. Finally, in Sec. 3.3, we present the DIGS training framework, which leverages RIG to select informative and reliable pseudo views and optimizes the 3DGS model in an uncertainty-aware manner.

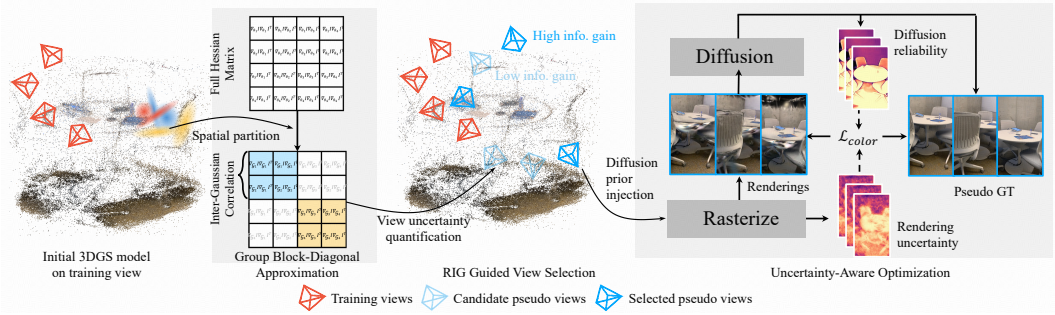

Figure 3: Overview of DIGS, which integrates diffusion priors into 3DGS guided by reliable information gain.

Preliminaries on 3DGS. 3DGS represents a 3D scene with a set of Gaussian primitive $\mathcal{G} = \{\mathcal{G}_i\}_{i=1}^{N}$, where N is the number of Gaussians. Each Gaussian primitive is parameterized by a position $\boldsymbol{\mu} \in \mathbb{R}^3$, scaling $\boldsymbol{s} \in \mathbb{R}^3$, rotation $\boldsymbol{r} \in \mathbb{R}^4$, opacity $\eta \in \mathbb{R}$ and color $\boldsymbol{c}$. We use $f(\mathcal{V}; \mathcal{G})$ to denote rendering the 3DGS model $\mathcal{G}$ at view $\mathcal{V}$. The rendering is done by computing the color of each pixel $\boldsymbol{p}$ via a process known as α-blending: $C(x_p) = \sum_{i \in N} \boldsymbol{c}_i \alpha_i \prod_{j=1}^{i-1}(1-\alpha_j)$, where α_i is computed by: $\alpha_i = \eta_i \exp(-\frac{1}{2}(\boldsymbol{p}-\boldsymbol{\mu}_i')^{\mathrm{T}} \boldsymbol{\Sigma}_i'^{-1}(\boldsymbol{p}-\boldsymbol{\mu}_i'))$. $\boldsymbol{\Sigma}_i'$ and $\boldsymbol{\mu}'$ are the projected Gaussian covariance and center via affine approximation (Kerbl et al., 2023). We denote the training view as $\mathcal{V}_{train}$, the ground truth images as $\mathcal{I}$. When using diffusion priors, we use $\bar{\mathcal{V}}_{candidate}$ and $\bar{\mathcal{V}}_{pseudo}$ to denote the candidate and selected pseudo views, and $\bar{\mathcal{I}}$ to denote the generated pseudo ground truth images.

3.1 INFORMATION GAIN VIA GROUP-BLOCK-DIAGONAL APPROXIMATION

Quantifying the information gain of pseudo views is important for guiding the integration of diffusion priors. However, 3DGS rasterization is a deterministic process that does not directly produce predictive uncertainty (Li & Cheung, 2024), and thus provides no direct measure of the information gain. In this section, we first formulate the information gain based on Fisher information, and then propose a novel Group-Block-Diagonal (GBD) approximation to make it tractable for 3DGS.

3DGS is optimized with an $\mathcal{L}_1$ loss, which is essentially a Maximum Likelihood Estimation (MLE). Maximizing the likelihood function is equivalent to minimizing the negative log-likelihood function given the monotonicity of the log function (Jiang et al., 2023): $-\log p(\mathcal{I}|\mathcal{V}, \mathcal{G}) = (f(\mathcal{V};\mathcal{G}) - \mathcal{I})^{\mathrm{T}}(f(\mathcal{V};\mathcal{G}) - \mathcal{I})$. The Fisher information matrix of the model is essentially the Hessian of the above equation with respect to the model parameters $\mathcal{G}$ under regularity conditions (Schervish, 2012; Jiang et al., 2023), as:$I(\mathcal{G}) = -\mathbb{E}[\nabla_{\mathcal{G}}^2 \log p(\mathcal{I}|\mathcal{V}, \mathcal{G})]$. Since the computation of the expectation is intractable, the Fisher information is computed based on observed data points in practice. More formally, given Fisher information is additive, the Fisher information matrix of a model parameterized by $\mathcal{G}$ trained on $\mathcal{V}_{train}$ can be computed by:

$$I_{train}(\mathcal{G}) = -\sum_{\mathcal{V} \in \mathcal{V}_{train}} \nabla_{\mathcal{G}}^2 \log p(\mathcal{I}|\mathcal{V}, \mathcal{G}) = \sum_{\mathcal{V} \in \mathcal{V}_{train}} H[\mathcal{I}|\mathcal{V}, \mathcal{G}] \tag{1}$$

where the $H[\cdot]$ is the Hessian matrix; that is $H[\mathcal{I}|\mathcal{V}, \mathcal{G}] = \nabla_{\mathcal{G}} f(\mathcal{V};\mathcal{G}) \nabla_{\mathcal{G}} f(\mathcal{V};\mathcal{G})^{\mathrm{T}} + (f(\mathcal{V};\mathcal{G}) - \mathcal{I}) \nabla_{\mathcal{G}}^2 f(\mathcal{V};\mathcal{G})$. The residual term $||f(\mathcal{V};\mathcal{G}) - \mathcal{I}||_1$ typically approaches zero (Wilson et al., 2025), which leads to a simplified formulation of the Hessian at view $\mathcal{V}$ as:

$$H_{\mathcal{V}} = H[\mathcal{I}|\mathcal{V}, \mathcal{G}] = \nabla_{\mathcal{G}} f(\mathcal{V};\mathcal{G}) \nabla_{\mathcal{G}} f(\mathcal{V};\mathcal{G})^{\mathrm{T}} \tag{2}$$

where $\nabla_{\mathcal{G}} f(\mathcal{V};\mathcal{G})$ is the Jacobian. With the Hessian formulation in Eq. (2), we can quantify the information gain of a view without the need of actual ground truth images. Yet, the dimension of the Hessian matrix $H \in \mathbb{R}^{\mathcal{G} \times \mathcal{G}}$ is quadratic w.r.t. the number of model parameters, which is intractable for 3DGS models consisting of millions of parameters.

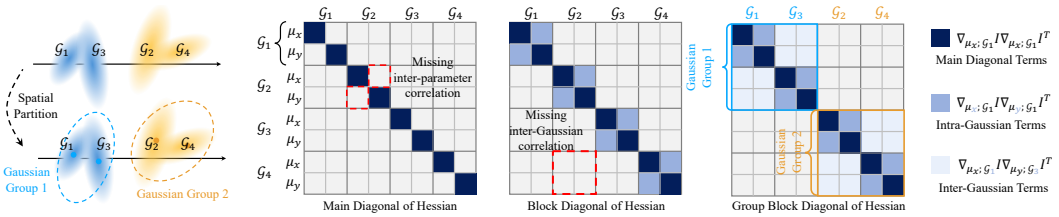

Figure 4: Illustration of the Group-Block-Diagonal (GBD) approximation of the Fisher information matrix. (Left) Spatially adjacent Gaussians are highly correlated during rasterization. (Right) Existing approximations (Jiang et al., 2023; Wilson et al., 2025; Hanson et al., 2025) discard inter-Gaussian correlations, which are preserved by GBD.

To this end, we introduce the Group-Block-Diagonal (GBD) approximation for tractable and accurate estimation of the information gain. Previous approaches approximate the Hessian matrix by its main diagonal (Jiang et al., 2023) or main block-diagonal (Wilson et al., 2025; Hanson et al., 2025). As illustrated in Fig. 4 (left), our key insight is that spatially adjacent Gaussians are highly correlated during the alpha-blending process in 3DGS rasterization. Accordingly, both intra-Gaussian and inter-Gaussian parameter correlations are non-negligible when computing the Fisher information. As shown in Fig. 4 (right), the diagonal approximation considers only the sensitivity of individual parameters, such as $\nabla_{\mu_x;\mathcal{G}_1} f \nabla_{\mu_x;\mathcal{G}_1} f^{\mathrm{T}}$. While the block-diagonal approximation considers intra-Gaussian correlations, such as $\nabla_{\mu_x;\mathcal{G}_1} f \nabla_{\mu_y;\mathcal{G}_1} f^{\mathrm{T}}$, the inter-Gaussian correlations are discarded.

In contrast, our GBD approximation additionally preserves the inter-Gaussian correlations, such as $\nabla_{\mu_x;\mathcal{G}_1} f \nabla_{\mu_y;\mathcal{G}_2} f^{\mathrm{T}}$, among spatially adjacent Gaussians, leading to a more accurate estimation of the information gain. Specifically, we partition the Gaussian primitives $\{\mathcal{G}_i\}_{i=1}^N$ into M non-overlapping groups $\{\mathcal{S}_m\}_{m=1}^M$ of size K based on the spatial proximity of Gaussian means $\boldsymbol{\mu}_i$, such that $\bigcup_{m=1}^M \mathcal{S}_m = \{\mathcal{G}_i\}_{i=1}^N$ and $|\mathcal{S}_m| = K$. Based on the group partition, we approximate the Hessian matrix as $H \approx \mathrm{diag}(H_{\mathcal{S}_1}, \ldots, H_{\mathcal{S}_M})$. We compute the log determinant of the Hessian matrix to quantify the sensitivity of parameters as:

$$U(H) = \log|H| = \textstyle\sum_{m=1}^{M} \log|H_{\mathcal{S}_m}| \tag{3}$$

where $H_{\mathcal{S}_m}$ denotes a block of Gaussians grouped by spatial location. Our GBD approximation in Eq. (3) effectively captures necessary intra-Gaussian and inter-Gaussian parameter correlations. We leave a more detailed comparison with other approximation methods (Jiang et al., 2023; Hanson et al., 2025; Wilson et al., 2025) to App. A.

3.2 Diffusion-Aware Reliable Information Gain

While the Fisher information matrix measures the information gain a 3DGS model can obtain by adding a certain view, it assumes that a clean ground truth image can be captured at that view. This assumption does not hold for diffusion-enhanced methods, where the pseudo ground truths are generated by the diffusion model rather than captured. Therefore, the actual information gain of a pseudo view should also account for the reliability of its pseudo ground truth. In this section, we propose a Reliable Information Gain (RIG) formulation based on geometric consistency.

Reliability of Generated Pseudo Ground Truth. Diffusion-enhanced methods rely on the diffusion model to generate content for under-constrained regions. However, the diffusion model can generate unreliable content with high confidence. Our key insight is that the reliability of a generated pseudo ground truth can be assessed by whether its content is geometrically consistent with the training views. Accordingly, we leverage the 3D geometric foundation model (GFM), Depth Anything 3 (DA3) (Lin et al., 2025), to check the geometric consistency. Specifically, as shown in Fig. 5, we feed DA3 with both the initial training images $\mathcal{I}$ and the pseudo ground truths $\bar{\mathcal{I}}$ generated at the candidate views, together with their camera poses, and obtain a pixel-wise confidence map for each view:

$$\{\mathbf{C}_i\}_{\mathcal{V}_i \in \mathcal{V}_{train}} \cup \{\bar{\mathbf{C}}_j\}_{\bar{\mathcal{V}}_j \in \bar{\mathcal{V}}_{candidate}} = \mathrm{DA3}\Big(\{(\mathcal{V}_i, \mathcal{I}_i)\}_{\mathcal{V}_i \in \mathcal{V}_{train}} \cup \{(\bar{\mathcal{V}}_j, \bar{\mathcal{I}}_j)\}_{\bar{\mathcal{V}}_j \in \bar{\mathcal{V}}_{candidate}}\Big) \tag{4}$$

where $\mathbf{C}_i, \bar{\mathbf{C}}_j \in \mathbb{R}^{h \times w}$ denote the pixel-wise confidence maps. The posed training images provide consistent geometry, which aligns with the 3DGS model. Thus, low confidence in pseudo ground truth images indicates that the generated content is inconsistent with the existing geometry, which

typically corresponds to diffusion artifacts, as shown in Fig. 5. We normalize $\bar{\mathbf{C}}_j$ to $[0, 1]$ and use it as the pixel-wise reliability map $\mathbf{R}_j$ of $\bar{\mathcal{I}}_j$. More details are provided in App. I.

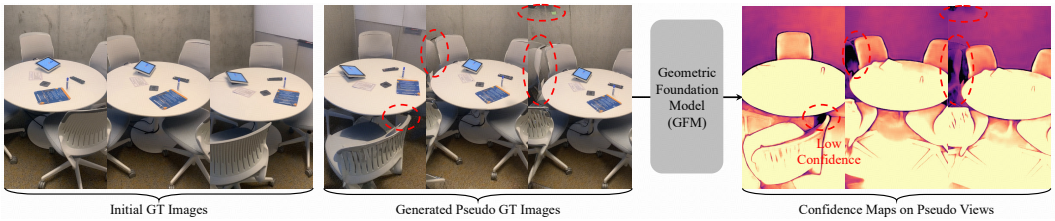

Figure 5: Evaluating the reliability of pseudo ground truths via geometric consistency. $\mathcal{I}$ and $\bar{\mathcal{I}}$ are jointly fed into DA3 (Lin et al., 2025). Artifacts generated by the diffusion model (red circles) are inconsistent with the geometry of the training images and receive low confidence.

Reliable Information Gain. The Hessian in Eq. (2) implicitly assumes that every pixel at view $\mathcal{V}$ is observed as reliably as in a real capture. This assumption holds for training views, but not for pseudo views supervised by generated images. We therefore model the pseudo ground truth $\bar{\mathcal{I}}_j$ as a noisy observation whose pixel-wise precision is given by its reliability map, $W_j = \mathrm{diag}(\mathbf{R}_j)$. Accordingly, the squared residual in the negative log-likelihood in Sec. 3.1 is weighted by W_j, and following the same derivation as Eq. (2), the Hessian at the pseudo view becomes $\tilde{H}_{\bar{\mathcal{V}}_j} = \nabla_{\mathcal{G}} f(\bar{\mathcal{V}}_j; \mathcal{G})\, W_j\, \nabla_{\mathcal{G}} f(\bar{\mathcal{V}}_j; \mathcal{G})^{\mathrm{T}}$. Since W_j is diagonal, the information contributed by each pixel is scaled by the reliability of its pseudo ground truth. The above equation reduces to Eq. (2) when all pixels are fully reliable, i.e., W_j is the identity matrix. Given the Hessian of the training views $H_{\mathcal{V}_{train}} = \sum_{\mathcal{V} \in \mathcal{V}_{train}} H_{\mathcal{V}}$, we define the Reliable Information Gain (RIG) of a pseudo view as the increase in the log determinant in Eq. (3):

Algorithm 1: RIG-guided 3DGS Training with Diffusion Priors

Input : Initial 3DGS model $\mathcal{G}$, training views $\mathcal{V}_{train}$ with images $\mathcal{I}$, number of pseudo views B.

$\bar{\mathcal{V}}_{pseudo} \leftarrow \emptyset$, $\bar{\mathcal{V}}_{candidate} \leftarrow \mathrm{Sample}(\mathcal{V}_{train})$, $\mathcal{L} \leftarrow 0$.
$\bar{\mathcal{I}} \leftarrow \mathrm{Diffusion}(f(\bar{\mathcal{V}}; \mathcal{G}))$, $\forall \bar{\mathcal{V}} \in \bar{\mathcal{V}}_{candidate}$ // Pseudo GT
$\mathbf{R} \leftarrow \mathrm{DA3}(\mathcal{I}, \bar{\mathcal{I}})$ // Reliability, Eq. (4)
$H \leftarrow \sum_{\mathcal{V} \in \mathcal{V}_{train}} \nabla_{\mathcal{G}} f(\mathcal{V}; \mathcal{G}) \nabla_{\mathcal{G}} f(\mathcal{V}; \mathcal{G})^{\mathrm{T}}$
while $|\bar{\mathcal{V}}_{pseudo}| < B$ **do**
  $\bar{\mathcal{V}}^* \leftarrow \arg\max_{\bar{\mathcal{V}} \in \bar{\mathcal{V}}_{candidate}} U(H + \tilde{H}_{\bar{\mathcal{V}}})$ // Eq. (6)
  $\bar{\mathcal{V}}_{pseudo} \leftarrow \bar{\mathcal{V}}_{pseudo} \cup \{\bar{\mathcal{V}}^*\}$
  $H \leftarrow H + \tilde{H}_{\bar{\mathcal{V}}^*}$ // Update Fisher info.
  $\bar{\mathcal{V}}_{candidate} \leftarrow \bar{\mathcal{V}}_{candidate} \setminus \{\bar{\mathcal{V}}^*\}$
for $\mathcal{V} \in \mathcal{V}_{train} \cup \bar{\mathcal{V}}_{pseudo}$ **do**
  $\hat{\mathcal{I}} \leftarrow f(\mathcal{V}; \mathcal{G})$ // Render at view
  if $\mathcal{V} \in \mathcal{V}_{train}$ **then**
    $\mathcal{L} \leftarrow \mathcal{L} + \mathcal{L}_{train}(\hat{\mathcal{I}}, \mathcal{I})$ // Clean GT
  else
    $\mathcal{L} \leftarrow \mathcal{L} + \mathcal{L}_{pseudo}(\hat{\mathcal{I}}, \bar{\mathcal{I}})$ // Pseudo GT
$\mathcal{G}' \leftarrow \mathcal{G} - \nabla_{\mathcal{G}} \mathcal{L}$ // Update the 3DGS model
return Updated model $\mathcal{G}'$

$$\mathrm{RIG}(\bar{\mathcal{V}}_j) = U\big(H_{\mathcal{V}_{train}} + \tilde{H}_{\bar{\mathcal{V}}_j}\big) - U\big(H_{\mathcal{V}_{train}}\big) \tag{5}$$

where $U(\cdot)$ is the log determinant of the Hessian matrix based on GBD approximation in Eq. (3). A pseudo view thus obtains a high RIG only if it is informative to the current 3DGS model and geometrically reliable in the regions that bring the information gain.

3.3 DIGS Training

We introduce the DIGS training framework guided by RIG. Specifically, we start from a base 3DGS model trained on $\mathcal{V}_{train}$, progressively select pseudo views to maximize RIG, and use the generated pseudo ground truths to further fine-tune the 3DGS model.

View Selection with RIG. As shown in Algorithm 1, the pseudo views are selected in three steps. (1) *Candidate view sampling.* Following (Kim et al., 2025; Ni et al., 2026), we sample a set of candidate pseudo views $\bar{\mathcal{V}}_{candidate}$ around the training views within the scene bounds, e.g., by translating, rotating, and orbiting the training cameras. (2) *RIG evaluation.* For each candidate view, we render the current 3DGS model, generate its pseudo ground truth with the diffusion model, and estimate the reliability map via Eq. (4). (3) *Greedy selection.* Starting from $H = H_{\mathcal{V}_{train}}$, we iteratively select the candidate with the highest RIG, computed by Eq. (5) with $H_{\mathcal{V}_{train}}$ replaced by the current H, and add its reliability-weighted Hessian to H:

$$\bar{\mathcal{V}}^* = \underset{\bar{\mathcal{V}} \in \bar{\mathcal{V}}_{candidate}}{\arg\max}\ \mathrm{RIG}(\bar{\mathcal{V}}; H) = \underset{\bar{\mathcal{V}} \in \bar{\mathcal{V}}_{candidate}}{\arg\max}\ U\big(H + \tilde{H}_{\bar{\mathcal{V}}}\big) \tag{6}$$

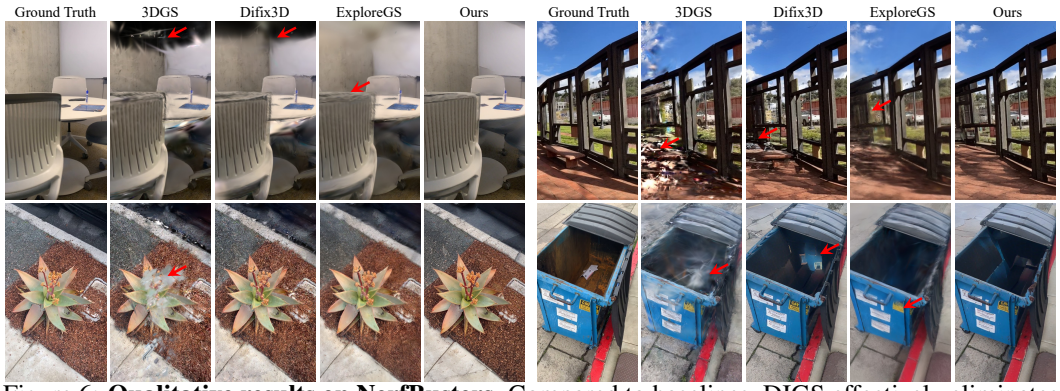

Figure 6: **Qualitative results on NerfBusters.** Compared to baselines, DIGS effectively eliminates floaters and geometric distortions in extrapolated views, with more complete scene boundaries.

Table 2: Quantitative comparison on NerfBusters, WildExplore and DL3DV datasets. The best, second-best, and third-best entries are highlighted respectively.

Method	NerfBusters			WildExplore			DL3DV		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
3DGS (Kerbl et al., 2023)	14.51	0.417	0.454	15.81	0.487	0.434	16.19	0.537	0.363
3DGS* (Kerbl et al., 2023)	15.00	0.435	0.436	15.78	0.483	0.433	16.49	0.566	0.331
DNGaussian (Li et al., 2024a)	14.01	0.442	0.525	15.65	0.504	0.450	15.48	0.496	0.422
ViewExtrapolator (Liu et al., 2024a)	14.98	0.450	0.455	16.02	0.510	0.422	16.88	0.581	0.347
Difix3D (Wu et al., 2025)	15.78	0.468	0.433	16.56	0.514	0.423	17.33	0.588	0.324
ExploreGS (Kim et al., 2025)	16.29	0.479	0.434	16.96	0.534	0.413	17.65	0.592	0.319
Ours	17.15	0.523	0.381	17.78	0.586	0.349	18.14	0.623	0.279

Since each $\tilde{H}_{\bar{\mathcal{V}}}$ is positive semi-definite and does not depend on the other selected views, the log-determinant objective is monotone submodular, and the greedy strategy yields a $(1 - 1/e)$-approximation guarantee (Nemhauser et al., 1978; Jiang et al., 2023). We provide the analysis of the greedy strategy in App. J, and visualizations and details on candidate sampling in App. E.

Uncertainty-Aware Optimization. For initial training views with clean ground truth images, we adopt the standard 3DGS loss (Kerbl et al., 2023): $\mathcal{L}_{train} = \|f(\mathcal{V}; \mathcal{G}) - \mathcal{I}\|_1 + \mathcal{L}_{SSIM}(f(\mathcal{V}; \mathcal{G}), \mathcal{I})$. For pseudo views, we use the generated images to supervise regions where the 3DGS model is less constrained and less certain, while avoiding conflicts with well-constrained regions. We first derive the rendering uncertainty from the Fisher information in Sec. 3.1. For a Gaussian $\mathcal{G}_i$ in group $\mathcal{S}_m$, we define its uncertainty as $u_i = -\beta \log |H_{\mathcal{G}_i}| - (1 - \beta) \log |H_{\mathcal{S}_m}|/|\mathcal{S}_m|$, where $H_{\mathcal{G}_i}$ and $H_{\mathcal{S}_m}$ are the blocks of H corresponding to $\mathcal{G}_i$ and $\mathcal{S}_m$. The group term accounts for the correlations among adjacent Gaussians, since a Gaussian with high individual uncertainty can still be well constrained by its group. We then rasterize u_i into a pixel-wise uncertainty map U_{render} by replacing the color $\boldsymbol{c}_i$ with u_i in α-blending, and normalize it to $[0, 1]$. However, uncertain regions are also where the pseudo ground truth is more likely to contain artifacts. We therefore further weight the supervision by the reliability map $\mathbf{R}_j$ from Sec. 3.2, so that the model learns from generated content only where it is both needed and reliable. The loss for pseudo view $\bar{\mathcal{V}}_j$ is:

$$\mathcal{L}_{pseudo} = \left(U_{render} \odot \mathbf{R}_j\right) \odot \|f(\bar{\mathcal{V}}_j; \mathcal{G}) - \bar{\mathcal{I}}_j\|_1 + \mathcal{L}_{SSIM}(f(\bar{\mathcal{V}}_j; \mathcal{G}), \bar{\mathcal{I}}_j). \tag{7}$$

4 Experiments

Datasets and Metrics. Our evaluation focuses on extrapolative 3D reconstruction, where the test views deviate from the training views. Specifically, we use NerfBusters (Warburg et al., 2023), WildExplore (Kim et al., 2025), and DL3DV (Ling et al., 2024). For DL3DV, we use its standalone evaluation benchmark to avoid information leakage. Following (Liu et al., 2024b; Wu et al., 2025; Kim et al., 2025), we report PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018). We leave a detailed discussion and visualization of the datasets to App. B.

Baselines. Following (Wu et al., 2025; Kim et al., 2025), we compare DIGS with 3DGS and its regularized variants. For depth-regularized baselines, we consider 3DGS* with depth prior

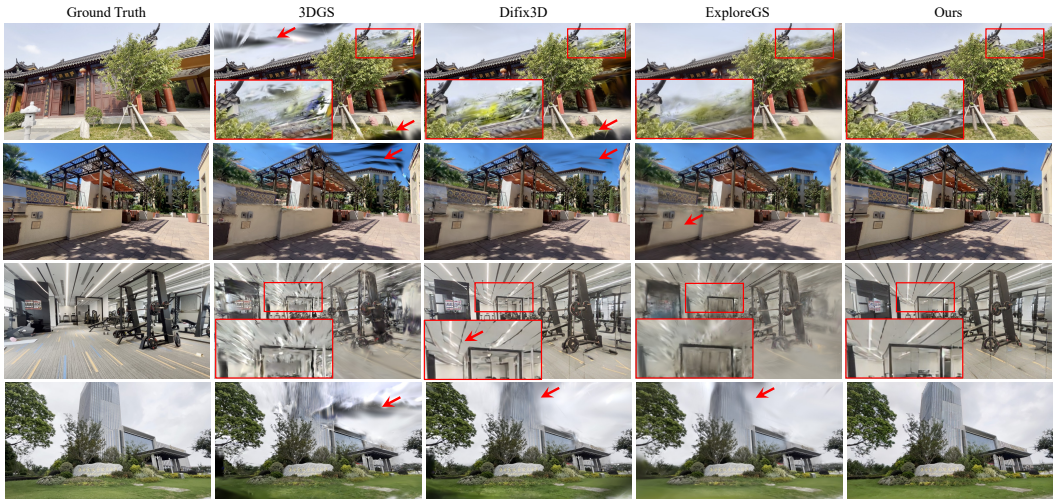

Figure 7: **Qualitative results on the DL3DV dataset.** DIGS effectively mitigates artifacts in extrapolated views, with more complete scene boundaries compared to baselines.

from Depth Anything V2 (Yang et al., 2024), and DNGaussian with depth prior from DPT (Ranftl et al., 2021). For diffusion-enhanced baselines, we compare with ViewExtrapolator (Liu et al., 2024a), Difix3D (Wu et al., 2025), and ExploreGS (Kim et al., 2025). We do not compare with G4Splat (Ni et al., 2026), which focuses on geometry reconstruction. We also compare with the feed-forward 3DGS method AnySplat (Jiang et al., 2025). We provide a more detailed discussion of the baselines in App. A.

Implementation Details. In our GBD approximation, we set the Gaussian group size to $K = 5$. For pseudo view selection, we select $B = 80$ pseudo views during the fine-tuning stage. For the uncertainty-aware optimization, we set $\beta = 0.5$ when computing the Gaussian uncertainty. More details about the implementation, parameter settings, and experimental environment, as well as the efficiency analysis, are provided in App. B and App. F.

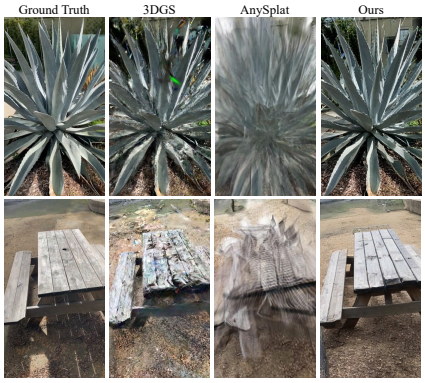

Figure 8: Qualitative comparison with feed-forward methods (Jiang et al., 2025).

Table 3: Comparison with feed-forward methods with varying training views.

# View	Method	PSNR ↑	SSIM ↑	LPIPS ↓
16 Views	3DGS	13.51	0.339	0.516
	AnySplat	11.57	0.276	0.591
	Ours	**14.37**	**0.376**	**0.469**
32 Views	3DGS	13.58	0.344	0.501
	AnySplat	11.93	0.291	0.59
	Ours	**16.29**	**0.467**	**0.443**
64 Views	3DGS	13.98	0.374	0.469
	AnySplat	12.09	0.306	0.593
	Ours	**16.56**	**0.484**	**0.417**

4.1 Experimental Results

Qualitative Results. As shown in Fig. 7, we compare our method with baselines on the DL3DV dataset. The 3DGS baseline suffers from significant artifacts in both indoor and outdoor scenes, which is expected since the evaluation focuses on regions that are under-constrained by the training views. While both Difix3D and ExploreGS reconstruct clearer details and object boundaries compared to 3DGS, they still suffer from minor artifacts and blurred regions. In contrast, our method consistently produces clearer rendered images. We provide additional qualitative comparisons on the NerfBusters and WildExplore datasets in Fig. 6 and Fig. 9.

Quantitative Results. As shown in Table 2, we provide quantitative comparisons with the baselines. Consistent with our qualitative results, 3DGS performs poorly across all datasets, indicating weak reconstruction quality at extrapolated views. Depth priors do not effectively improve the performance. 3DGS* improves the performance on NerfBusters and DL3DV but slightly degrades PSNR on WildExplore, while DNGaussian generally underperforms the 3DGS baseline. By selecting informative and reliable pseudo views with RIG, our method achieves the best performance across all datasets.

Comparison with Feed-Forward Methods. We compare with the feed-forward baseline AnySplat (Jiang et al., 2025) on the NerfBusters dataset with different numbers of training views, as shown in Tab. 3 and Fig. 8. Although AnySplat generalizes to unseen scenes, it underperforms the 3DGS

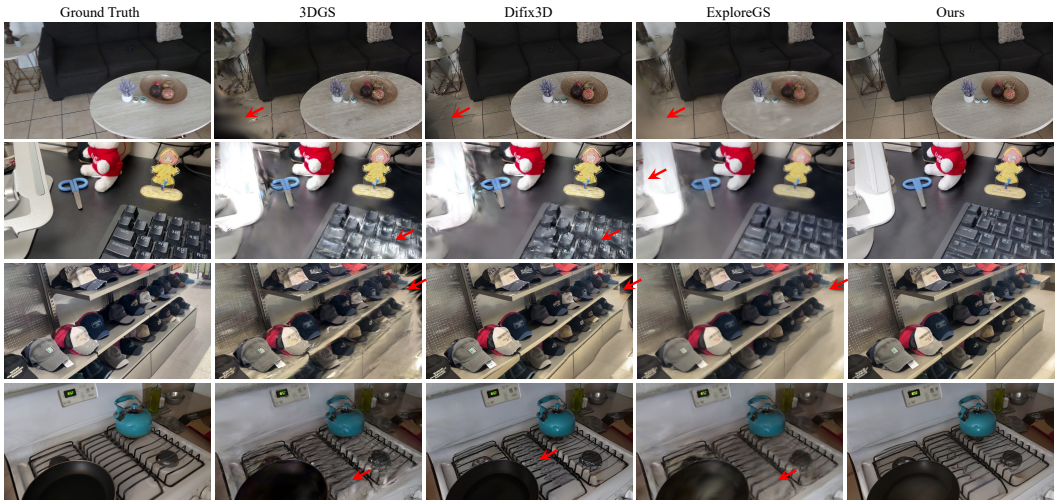

Figure 9: **Qualitative results on WildExplore.** Compared to baselines, DIGS renders extrapolated views with finer details, cleaner geometry, and more complete scene boundaries.

baseline on all metrics in the extrapolative setting, and its renderings suffer from heavy artifacts. In contrast, our method outperforms both 3DGS and AnySplat by a significant margin.

4.2 Ablation Study

Information Gain Quantification. To quantitatively evaluate the effectiveness of GBD, we compare it with prior Fisher information approximations (Jiang et al., 2023; Wilson et al., 2025) on the Mip-NeRF 360 dataset (Barron et al., 2022) in the active view selection setting. We measure the quality of the estimation by the AUSE following (Jiang et al., 2023). As shown in Tab. 4, GBD achieves lower AUSE and better reconstruction quality than prior works. We further ablate the view selection strategy in the extrapolative setting in Tab. 5, while keeping the diffusion model fixed. Selecting views purely by information gain brings limited gains over the uniform baseline. Although GBD provides a more accurate estimation and further improves the performance, it still overlooks the reliability of pseudo ground truths. By incorporating the reliability, RIG yields a clear improvement over GBD alone, validating that the information gain and the reliability should be considered jointly.

Table 4: Ablation on information gain quantification methods.

Method	AUSE ↓	PSNR ↑	SSIM ↑	LPIPS ↓
FisherRF	0.28	16.81	0.493	0.445
POp-GS	0.24	18.15	0.548	0.401
Ours (GBD)	**0.19**	**19.63**	**0.588**	**0.382**

Table 5: Ablation on view selection methods on WildExplore.

View Sel.	Method	PSNR ↑	SSIM ↑	LPIPS ↓
Coverage	Uniform	16.43	0.508	0.427
	Visibility	16.65	0.517	0.421
Fisher Info.	Diag.	16.41	0.497	0.427
	Block Diag.	16.63	0.522	0.418
	Ours(GBD)	16.82	0.535	0.402
	Ours(GBD+RIG)	17.27	0.547	0.391

Effectiveness of Diffusion Prior Integration. Tab. 6 ablates each component of DIGS on WildExplore. Diffusion priors improve the performance of 3DGS even with uniform view sampling, and RIG-based view selection further improves reconstruction quality. For uncertainty-aware optimization, incorporating the rendering uncertainty map improves the performance by preventing the pseudo supervision from altering well-constrained regions. Further incorporating the reliability map leads to a clear improvement, as it prevents the model from learning unreliable generated content in under-constrained regions. We leave a detailed discussion of limitations to App. L.

Table 6: Ablation on WildExplore.

Diff. Prior	RIG Sel.	Render Uncer.	Diff. Reliab.	PSNR ↑	SSIM ↑	LPIPS ↓
✗	✗	✗	✗	15.81	0.487	0.434
✓	✗	✗	✗	16.43	0.508	0.427
✓	✓	✗	✗	17.27	0.547	0.391
✓	✓	✓	✗	17.43	0.554	0.374
✓	✓	✓	✓	17.78	0.586	0.349

5 Conclusion

In this work, we propose DIGS, which integrates diffusion priors guided by information gain for extrapolative 3D reconstruction. DIGS estimates the information gain of pseudo views with GBD, accounts for the reliability of generated pseudo ground truths with RIG, and leverages RIG for both view selection and uncertainty-aware optimization. Extensive experiments show that DIGS consistently improves reconstruction quality in extrapolative settings.

AI Use Statement. In this work, we did not use generative AI tools for tasks requiring mandatory disclosure. We used generative AI tools to assist with debugging code and improving the grammar, clarity, and readability of the manuscript. We have reviewed all AI-assisted work. In particular, AI-assisted code modifications were manually reviewed and tested by the authors, and AI-assisted manuscript edits were checked for technical accuracy and consistency with the authors' intended meaning. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5855–5864, 2021.

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5470–5479, 2022.

Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

Sarosij Bose, Arindam Dutta, Sayak Nag, Junge Zhang, Jiachen Li, Konstantinos Karydis, and Amit K Roy-Chowdhury. Uncertainty-aware diffusion-guided refinement of 3d scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 28271–28281, 2025.

Henry Carrillo, Ian Reid, and José A Castellanos. On the comparison of uncertainty criteria for active slam. In *2012 IEEE International Conference on Robotics and Automation*, pp. 2080–2087. IEEE, 2012.

David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, and Vincent Sitzmann. pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 19457–19467, 2024.

Tianyu Chen, Wei Xiang, Kang Han, Yu Lu, Di Wu, Gaowen Liu, and Ramana Rao Kompella. Gifsplat: Generative prior-guided iterative feed-forward 3d gaussian splatting from sparse views. *arXiv preprint arXiv:2602.22571*, 2026.

Yutong Chen, Marko Mihajlovic, Xiyi Chen, Yiming Wang, Sergey Prokudin, and Siyu Tang. Splatformer: Point transformer for robust 3d gaussian splatting. In *The Thirteenth International Conference on Learning Representations*, 2025.

Zilong Chen, Feng Wang, Yikai Wang, and Huaping Liu. Text-to-3d using gaussian splatting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 21401–21412, 2024.

Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, and Weiwei Xu. High-quality surface reconstruction using gaussian surfels. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–11, 2024.

Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5354–5363, 2024.

Alex Hanson, Allen Tu, Vasu Singla, Mayuka Jayawardhana, Matthias Zwicker, and Tom Goldstein. Pup 3d-gs: Principled uncertainty pruning for 3d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5949–5958, 2025.

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*, pp. 1–11, 2024.

Lihan Jiang, Yucheng Mao, Linning Xu, Tao Lu, Kerui Ren, Yichen Jin, Xudong Xu, Mulin Yu, Jiangmiao Pang, Feng Zhao, et al. Anysplat: Feed-forward 3d gaussian splatting from unconstrained views. *ACM Transactions on Graphics (TOG)*, 44(6):1–16, 2025.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherrf: Active view selection and uncertainty quantification for radiance fields using fisher information. *arXiv preprint arXiv:2311.17874*, 2023.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139:1–139:14, 2023.

Minsu Kim, Subin Jeon, In Cho, Mijin Yoo, and Seon Joo Kim. Exploregs: Explorable 3d scene reconstruction with virtual camera samplings and diffusion priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 27042–27051, 2025.

Durk P Kingma, Tim Salimans, and Max Welling. Variational dropout and the local reparameterization trick. *Advances in neural information processing systems*, 28, 2015.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20775–20785, 2024a.

Pengzhi Li, Chengshuai Tang, Qinxuan Huang, and Zhiheng Li. Art3d: 3d gaussian splatting for text-guided artistic scenes generation. *arXiv preprint arXiv:2405.10508*, 2024b.

Ruiqi Li and Yiu-ming Cheung. Variational multi-scale representation for estimating uncertainty in 3d gaussian splatting. *Advances in Neural Information Processing Systems*, 37:87934–87958, 2024.

Haotong Lin, Sili Chen, Junhao Liew, Donny Y Chen, Zhenyu Li, Guang Shi, Jiashi Feng, and Bingyi Kang. Depth anything 3: Recovering the visual space from any views. *arXiv preprint arXiv:2511.10647*, 2025.

Lu Ling, Yichen Sheng, Zhi Tu, Wentian Zhao, Cheng Xin, Kun Wan, Lantao Yu, Qianyu Guo, Zixun Yu, Yawen Lu, et al. Dl3dv-10k: A large-scale scene dataset for deep learning-based 3d vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22160–22169, 2024.

Fangfu Liu, Wenqiang Sun, Hanyang Wang, Yikai Wang, Haowen Sun, Junliang Ye, Jun Zhang, and Yueqi Duan. Reconx: Reconstruct any scene from sparse views with video diffusion model. *IEEE Transactions on Image Processing*, 2026.

Kunhao Liu, Ling Shao, and Shijian Lu. Novel view extrapolation with video diffusion priors. *arXiv preprint arXiv:2411.14208*, 2024a.

Xi Liu, Chaoyi Zhou, and Siyu Huang. 3dgs-enhancer: Enhancing unbounded 3d gaussian splatting with view-consistent 2d diffusion priors. *Advances in Neural Information Processing Systems*, 37: 133305–133327, 2024b.

Xinhang Liu, Jiaben Chen, Shiu-hong Kao, Yu-Wing Tai, and Chi-Keung Tang. Deceptive-nerf: Enhancing nerf reconstruction using pseudo-observations from diffusion models. 2023.

Pierre Marza, Laetitia Matignon, Olivier Simonin, Dhruv Batra, Christian Wolf, and Devendra S Chaplot. Autonerf: Training implicit scene representations with autonomous agents. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 13442–13449. IEEE, 2024.

Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (ToG)*, 38(4):1–14, 2019.

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(1):99–106, 2021.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 14(1):265–294, 1978.

Junfeng Ni, Yixin Chen, Zhifei Yang, Yu Liu, Ruijie Lu, Song-Chun Zhu, and Siyuan Huang. G4splat: Geometry-guided gaussian splatting with generative prior. In *The Fourteenth International Conference on Learning Representations*, 2026.

Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. Activenerf: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision*, pp. 230–246. Springer, 2022.

René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 12179–12188, 2021.

Barbara Roessle, Norman Müller, Lorenzo Porzi, Samuel Rota Bulo, Peter Kontschieder, and Matthias Nießner. Ganerf: Leveraging discriminators to optimize neural radiance fields. *ACM Transactions on Graphics (TOG)*, 42(6):1–14, 2023.

Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *European Conference on Computer Vision*, pp. 87–103. Springer, 2024.

Mark J Schervish. *Theory of statistics*. Springer Science & Business Media, 2012.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5294–5306, 2025.

Linhan Wang, Kai Cheng, Shuo Lei, Shengkun Wang, Wei Yin, Chenyang Lei, Xiaoxiao Long, and Chang-Tien Lu. Dc-gaussian: Improving 3d gaussian splatting for reflective dash cam videos. *arXiv preprint arXiv:2405.17705*, 2024a.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20697–20709, 2024b.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Frederik Warburg, Ethan Weber, Matthew Tancik, Aleksander Holynski, and Angjoo Kanazawa. Nerfbusters: Removing ghostly artifacts from casually captured nerfs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 18120–18130, 2023.

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*, pp. 3646–3655, 2025.

Jay Zhangjie Wu, Yuxuan Zhang, Haithem Turki, Xuanchi Ren, Jun Gao, Mike Zheng Shou, Sanja Fidler, Zan Gojcic, and Huan Ling. Difix3d+: Improving 3d reconstructions with single-step diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26024–26035, 2025.

Jinbo Xing, Menghan Xia, Yong Zhang, Haoxin Chen, Wangbo Yu, Hanyuan Liu, Gongye Liu, Xintao Wang, Ying Shan, and Tien-Tsin Wong. Dynamicrafter: Animating open-domain images with video diffusion priors. In *European Conference on Computer Vision*, pp. 399–417. Springer, 2024.

Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Botao Ye, Sifei Liu, Haofei Xu, Xueting Li, Marc Pollefeys, Ming-Hsuan Yang, and Songyou Peng. No pose, no problem: Surprisingly simple 3d gaussian splats from sparse unposed images. *arXiv preprint arXiv:2410.24207*, 2024.

Botao Ye, Boqi Chen, Haofei Xu, Daniel Barath, and Marc Pollefeys. Yonosplat: You only need one model for feedforward 3d gaussian splatting. *arXiv preprint arXiv:2511.07321*, 2025.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018.

Saining Zhang, Baijun Ye, Xiaoxue Chen, Yuantao Chen, Zongzheng Zhang, Cheng Peng, Yongliang Shi, and Hao Zhao. Drone-assisted road gaussian splatting with cross-view uncertainty. *arXiv preprint arXiv:2408.15242*, 2024.

Yingji Zhong, Zhihao Li, Dave Zhenyu Chen, Lanqing Hong, and Dan Xu. Taming video diffusion prior with scene-grounding guidance for 3d gaussian splatting from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6133–6143, 2025.

Hongyu Zhou, Zisen Shao, Sheng Miao, Pan Wang, Dongfeng Bai, Bingbing Liu, and Yiyi Liao. Freefix: Boosting 3d gaussian splatting via fine-tuning-free diffusion models. In *Thirteenth International Conference on 3D Vision*, 2026.

APPENDIX

In this Appendix, we provide additional details about the design and parameter settings of DIGS, theoretical analysis of our view selection method using submodular formulation, and extended quantitative and qualitative results. Specifically, the appendix is organized as follows:

- Section A **Related Works**. We introduce related works in more detail, including diffusion-enhanced 3DGS, feed-forward 3DGS, and 3DGS uncertainty quantification.
- Section B **Implementation Details**. We describe the implementation details, key parameter settings, and the evaluation protocol for reproducibility, including the active view selection setting we used in Tab. 4.
- Section C **Additional Visualizations**. We present more qualitative comparisons, including additional rendered images and videos (in `video.zip`).
- Section D **Additional Evaluation**. We report additional quantitative results on the LLFF dataset (Mildenhall et al., 2019), which focuses on the forward-facing setting.
- Section E **Additional Discussion on View Selection**. We provide more details on candidate view sampling and RIG-based view selection.
- Section F **Efficiency Analysis**. We analyze the efficiency of our method in terms of model size, as well as training and inference efficiency.
- Section G **Ablation on Group Size**. We analyze the effect of the group size K in GBD on the computational cost and reconstruction quality.
- Section H **Ablation on Different Diffusion Models**. We provide experimental results when using our method on image and video diffusion models.
- Section I **Discussion on Diffusion Reliability**. We provide more analysis as well as visualizations to show how diffusion model can potentially generate unreliable pseudo ground truths, and how our geometric consistency-based reliability map can effectively mitigate this issue.
- Section J **Submodular Formulation of View Selection**. We provide a theoretical analysis of the view selection strategy used in Algorithm 1 based on submodular optimization.
- Section K **Per-Scene Results**. We report the per-scene results on the WildExplore dataset.
- Section L **Broader Impact and Limitation**. We discuss the broader impact and limitations of this work and outline possible future directions.

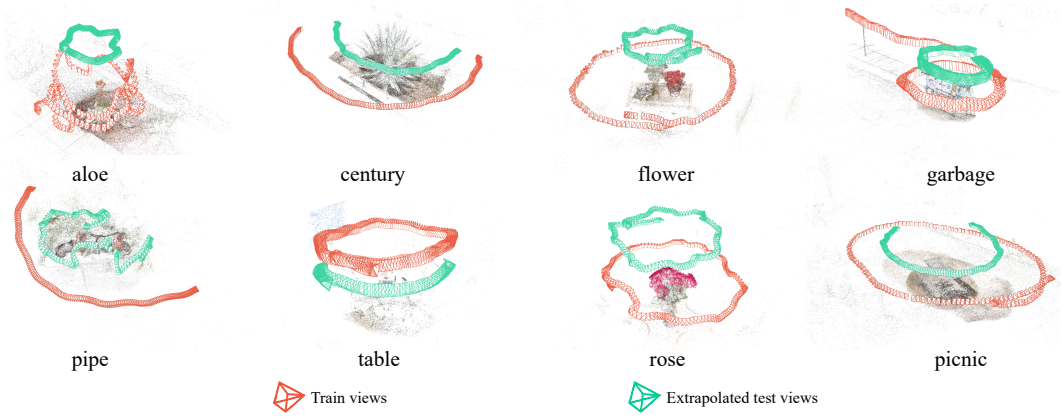

Figure 10: Visualization of the training and test views in the NerfBusters (Warburg et al., 2023) dataset.

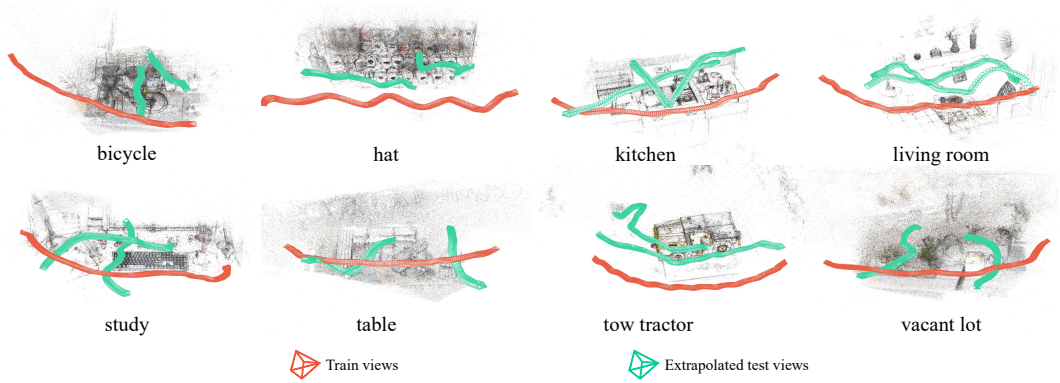

Figure 11: Visualization of the training and test views in the WildExplore (Kim et al., 2025) dataset.

A Related Works

Diffusion-Enhanced 3DGS. Existing works have explored injecting diffusion priors into 3DGS to enhance the reconstruction quality, especially in under-constrained regions. However, most works rely on simple view interpolation to inject the diffusion priors. For example, 3DGS-Enhancer (Liu et al., 2024b) targets the sparse-view setting and interpolates between training views to generate pseudo views. Similarly, Difix3D (Wu et al., 2025) assumes that query poses are known, and interpolates between training views and test views. We note that in our evaluation, we assume the test views are not known during reconstruction, which is common in real-world applications. We reproduce the results for Difix3D using uniform pseudo-view sampling based on grid view coverage (Marza et al., 2024). We note that ExploreGS (Kim et al., 2025) leverages visibility cues for view selection, which reflect the information gain to some extent, but overlooks the reliability of the generated pseudo ground truths. Similarly, G4Splat (Ni et al., 2026) leverages geometry cues for view selection and focuses on geometry reconstruction. Since we do not consider geometry reconstruction in this work, we did not compare with G4Splat. FreeFix (Zhou et al., 2026) uses Fisher information to guide diffusion denoising at given views rather than selecting pseudo views, and thus targets a different problem from ours. There are also some feed-forward 3DGS methods that use diffusion priors. However, their setting is different from ours. For example, UAR-Scenes (Bose et al., 2025) models diffusion uncertainty via semantic entropy, but focuses on 3D scene generation from a single image, and the test views are assumed to be continuous and in-distribution. GIFSplat (Chen et al., 2026) incorporates Difix3D in the feed-forward setting.

Uncertainty Quantification in 3DGS. Uncertainty is useful for evaluating the information of different camera views and the model's confidence in rendered images for 3D reconstruction. However, 3DGS rasterization is a deterministic process that is optimized in a non-Bayesian manner (Li & Cheung, 2024). As a result, standard 3DGS lacks the ability to directly quantify predictive uncertainty. Prior works have explored several ways to estimate the uncertainty of 3DGS via Bayesian learning (Bishop & Nasrabadi, 2006; Lakshminarayanan et al., 2017; Kingma et al., 2015). Ensemble-based methods train multiple 3DGS models and derive predictive uncertainty from the variance of the model outputs (Lakshminarayanan et al., 2017; Zhang et al., 2024). Variational inference-based methods quantify uncertainty by approximating the model posterior and jointly optimizing the likelihood of pixel colors and the discrepancy between the posterior and a variational distribution (Kingma et al., 2015; Li & Cheung, 2024). Recent Fisher information-based approaches leverage the formulation of the reconstruction objective, which corresponds to Maximum Likelihood Estimation (MLE) with error following a Laplace distribution (Jiang et al., 2023). Accordingly, uncertainty can be quantified through the approximation of the Fisher information matrix (Jiang et al., 2023; Hanson et al., 2025; Wilson et al., 2025; Zhou et al., 2026).

In the field of active perception, maximizing information gain in 3D reconstruction through view selection has been widely studied (Carrillo et al., 2012; Pan et al., 2022; Jiang et al., 2023; Marza et al., 2024; Wilson et al., 2025). Existing methods estimate the information gain through model uncertainty quantification. While model ensemble (Lakshminarayanan et al., 2017; Zhang et al., 2024) and variational inference (Kingma et al., 2015; Li & Cheung, 2024) based methods are

straightforward, they often incur large computational overhead. Recent methods leverage the Fisher information matrix to quantify model uncertainty, which provides a better balance between utility and efficiency (Jiang et al., 2023; Wilson et al., 2025; Hanson et al., 2025). In practice, uncertainty estimation with the Fisher information relies on approximating the Hessian matrix. While existing works introduce efficient approximations based on diagonal (Jiang et al., 2023) or block-diagonal structures (Hanson et al., 2025; Wilson et al., 2025), they ignore inter-Gaussian parameter correlations. Such correlations are important for 3DGS models, since spatially adjacent Gaussians are strongly correlated. In addition, these view selection methods do not consider diffusion priors or the reliability of the generated pseudo ground truths, which is crucial in extrapolative reconstruction. We note that FisherRF (Jiang et al., 2023) and POp-GS (Wilson et al., 2025) target active view selection, while PUP 3D-GS (Hanson et al., 2025) focuses on Gaussian pruning. Therefore, we compare with FisherRF and POp-GS in the active view selection setting in Tab. 4, and do not compare with PUP 3D-GS.

Feed-Forward 3DGS. Recent feed-forward 3DGS methods achieve more generalizable reconstruction by pretraining on large datasets (Charatan et al., 2024; Ye et al., 2024; Jiang et al., 2025; Ye et al., 2025). They can reconstruct 3DGS from sparse view inputs in a single forward pass. However, they typically assume that the test views are in the distribution of training views. In contrast, this work focuses on the extrapolative setting where the test views significantly deviate from training views. Moreover, these feed-forward methods typically only support low-resolution images, such as 224 or 448 resolution. Feed-forward methods generally struggle to generalize to extrapolated views, as also shown in Sec. 4. We have also attached the results on NerfBusters obtained by AnySplat (Jiang et al., 2025) in the supplementary material.

B Implementation Details

Extrapolative Datasets. NerfBusters consists of 12 scenes where the test views lie on a separate trajectory from the training views. We follow ExploreGS (Kim et al., 2025) to partition the training and test views to simulate more significant viewpoint changes. WildExplore is an evaluation benchmark consisting of 8 scenes. Each scene contains two distinct test view trajectories that both deviate from the training views.

For DL3DV, we select 12 scenes from the standalone evaluation benchmark, as listed in Tab. 7. This differs from prior works such as Difix3D, which use a subset of the DL3DV-10K training data. Since some diffusion models in the baselines are fine-tuned on DL3DV-10K (Liu et al., 2024b; Wu et al., 2025; Kim et al., 2025), evaluating on the standalone evaluation benchmark ensures fair comparison and avoids potential information leakage. Following (Liu et al., 2024b; Wu et al., 2025; Kim et al., 2025), we report PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018).

Table 7: Scenes selected from the DL3DV evaluation benchmark (Ling et al., 2024).

Category	Name	Hash
Outdoor	Apartment	c9fa8e4f985a871be37f8bcb5baab74cfc535b12c3898e7469536599a63b9184
Outdoor	Building	924228bfa8983b5aa7921a609d5046036c96022f8458c037460386e74f3ffec6
Outdoor	Museum	1e5ce991775e9266dcd553306ae9cc153ffd19101b789ff66526bc3877c54fdf
Outdoor	Park	ae90ea23130506b51f0aa19e518d2cc920ebb8c9a2b89a2178c00d3f5959440b
Outdoor	Playground	494835cf61b77ab1cb7dd9fb71995be8117ccebd3b9be69717060be759d1336b
Outdoor	Booth	14452271416631599c6e7a2362b2c5330e08fbf0ec2adfea7930d84cba4e3d69
Indoor	Grocery	39c7fd84f7908d0a3feb90090cd6ad77d45aee8f11a0fcc9d0db08989fb544e1
Indoor	Gym	f2ee054873ffbfae0269060f956ef713579852b6a94ded2faceee282baf3448a
Indoor	Hotel Room	b93f5691ba1cc8890d0b0fb5792668d8c8e084e3b3ae1a476fc5e2c1e38b79ae
Indoor	Lobby	d2ce0396dc6df5a579dbd8189885ce978c19490709875043e3566efba054236f
Indoor	Restaurant	9f361436f272b69706cdba2747ab7b86d7227bbbd95458cc927255d11f6e9bbb
Indoor	Library	9c567331ea71ab56d8a96c10ab4d17b8ae3350fcaf689908cdcf608e1d8ae6db

To better illustrate the extrapolative setting, we visualize the camera distributions of datasets where the test views significantly deviate from the training views. As shown in Fig. 10, the NerfBusters dataset contains training and test views that orbit the objects. Following (Kim et al., 2025), we select the inner trajectory as the test views to simulate a forward-moving setting. As shown in Fig. 11, the WildExplore dataset also contains test views that are outside the distribution of the training views. We note that the WildExplore dataset provides two independent test trajectories, and the evaluation results

are averaged over these two trajectories. For DL3DV, following (Zhou et al., 2026), we partition the views of each scene into two groups according to their camera positions, and use one group for training and the other for evaluation. For LLFF (Mildenhall et al., 2019), we follow (Zhou et al., 2026) to set up a sparse-view extrapolative setting, as detailed in App. D.

Training Schedule and Parameter Setting. Following (Kim et al., 2025), we train all baselines for a total of 45,000 iterations. Following (Liu et al., 2024b; Kim et al., 2025), we first train the standard 3DGS model for 30,000 iterations and then fine-tune the models for an additional 15,000 iterations using generated pseudo ground-truth images. We follow the original 3DGS for densification and pruning (Kerbl et al., 2023). For the 3DGS* baseline, which uses depth maps from Depth Anything V2 (Yang et al., 2024) as additional regularization, we set the initial weight of the depth loss to 1.0 and decay it exponentially to a final weight of 0.01. For the Difix3D (Wu et al., 2025) baseline, we re-run all experiments without assuming access to the test views during training. In their original setting, Difix3D uses test views to perform view interpolation.

During the information gain estimation in Sec. 3.1, we use the GBD approximation. We partition Gaussians into groups of size $K = 5$ using balanced K-Means. Note that the total number of Gaussians may not be a multiple of K; in such cases, the remaining Gaussians are computed in a smaller group. In balanced K-Means, we consider only the Gaussian positions, i.e., $\boldsymbol{\mu} \in \mathbb{R}^3$, to perform spatial partitioning. Similar to prior work (Hanson et al., 2025; Wilson et al., 2025), when computing the log determinant, we consider the spatial parameters of each Gaussian, $\{\boldsymbol{\mu}, \boldsymbol{s}\}$. This is because the uncertainty of Gaussian primitives is mainly determined by the spatial parameters. Ignoring appearance parameters such as color and opacity does not reduce performance and can even improve it (Wilson et al., 2025).

During RIG-guided view selection in Algorithm 1, pseudo views are selected from a set of candidate views. Following (Kim et al., 2025), we generate candidate views **without relying on test views**, which is more practical for real-world applications. Specifically, we sample 320 candidate pseudo views following the sampling strategy of (Kim et al., 2025; Ni et al., 2026), from which $B = 80$ pseudo views are selected. More details on candidate sampling are provided in App. E. Unless otherwise stated, we use the video diffusion model from ExploreGS (Kim et al., 2025) by default. All experiments are conducted on A5000 GPUs.

Active View Selection Setting. In Tab. 4, we follow (Wilson et al., 2025) to set up the active view selection setting (i.e., active mapping) on the Mip-NeRF 360 dataset (Barron et al., 2022). The 3DGS model is initialized with 2 training views. When v training views are available, we train the model for $100v$ iterations, and then select one new view from the remaining candidate views according to the criterion of each method, i.e., the information gain estimated by the diagonal (Jiang et al., 2023), block-diagonal (Wilson et al., 2025), or our GBD approximation. The selected view is added to the training set, and this process is repeated until $v = 10$. After that, we continue training the model until the total number of training iterations reaches 10,000. Unlike the extrapolative setting, the selected views are supervised by real captured images in this setting, which allows us to evaluate the information gain estimation independently of diffusion priors.

To evaluate the quality of the estimation, we compute the Area Under the Sparsification Error (AUSE) following (Jiang et al., 2023). Specifically, for each test view, we render the pixel-wise uncertainty map with each approximation, and sort the pixels by their estimated uncertainty in descending order. We then progressively remove the pixels with the highest uncertainty and compute the average rendering error of the remaining pixels, which yields a sparsification curve. An oracle curve is obtained in the same way by sorting the pixels according to their actual rendering errors. AUSE is defined as the area between the two curves, and a lower AUSE indicates that the estimated uncertainty better ranks the actual rendering errors.

C Additional Visualizations

Rendered Videos. To better demonstrate the improvement in reconstruction quality under extrapolative settings, we provide rendered videos in the `video.zip` file included in the supplementary materials. The RIG-guided diffusion prior generally improves the reconstruction quality in under-constrained regions across different datasets.

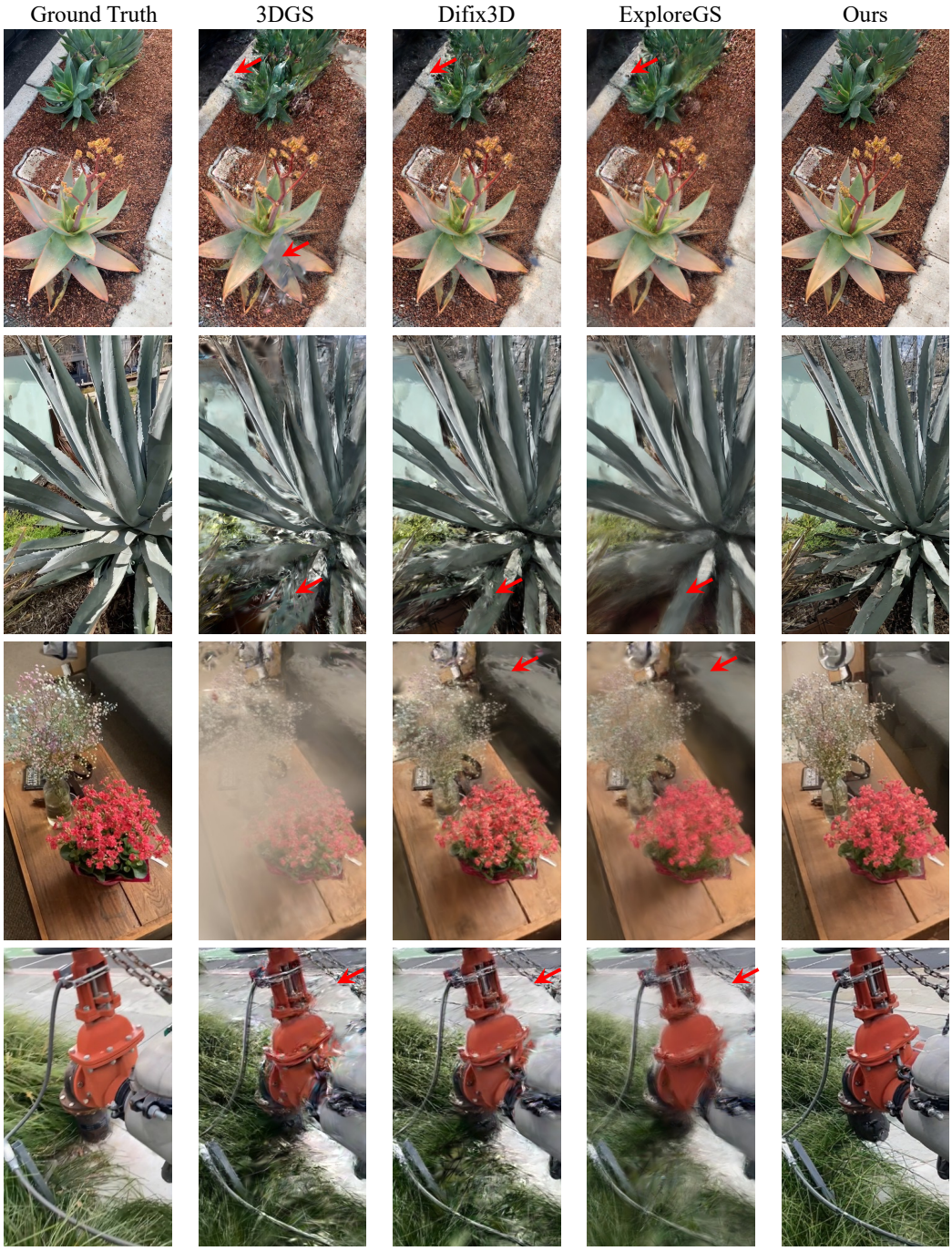

Figure 12: **Additional Qualitative Results on NerfBusters.** Red arrows highlight regions where our RIG-guided diffusion priors preserve sharper textures and more complete structures in under-constrained regions.

Qualitative Comparison. We provide additional qualitative results on the NerfBusters, WildExplore, and DL3DV-Eval datasets. As shown in Fig. 12, 3DGS often produces floaters and artifacts in under-constrained regions due to the lack of supervision in the NerfBusters dataset. In contrast, diffusion-enhanced methods improve the reconstruction quality in these regions. Additional results on the WildExplore dataset are shown in Fig. 13, and results on DL3DV are presented in Fig. 14. Our method generally outperforms Difix3D and ExploreGS across these datasets, in both indoor and

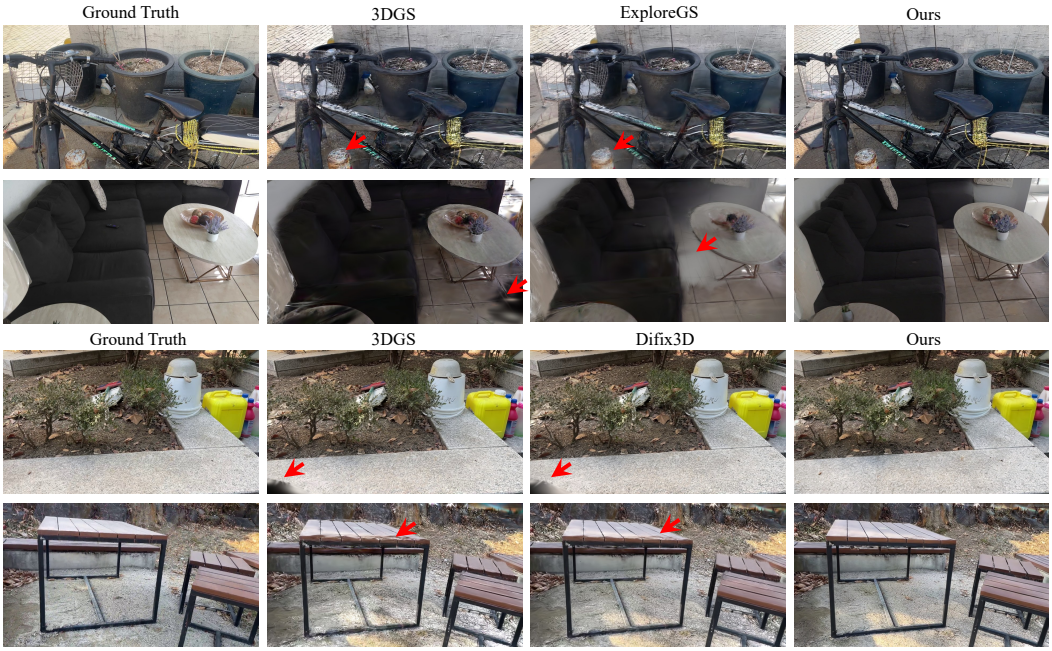

Figure 13: **Additional Qualitative Results on WildExplore.** Red arrows highlight regions where our RIG-guided diffusion priors preserve sharper textures and more complete structures in under-constrained regions.



Figure 14: **Additional Qualitative Results on DL3DV.** Red arrows highlight regions where our RIG-guided diffusion priors preserve sharper textures and more complete structures in under-constrained regions.

outdoor scenes. RIG allows diffusion priors to be injected into informative and reliable pseudo views, providing more effective supervision during training.

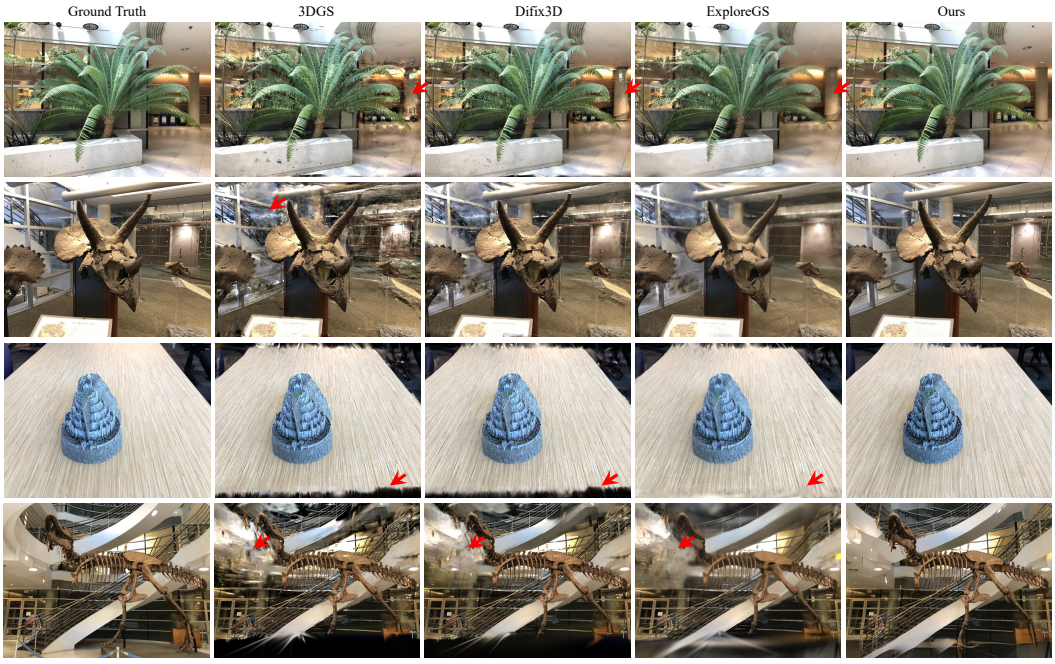

Figure 15: **Qualitative Comparisons on LLFF.** Red arrows highlight regions where our RIG-guided diffusion priors preserve sharper textures and more complete structures in under-constrained regions.

Table 8: Comparison on the LLFF dataset. 3DGS* denotes 3DGS with depth regularization.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
3DGS	18.10	0.633	0.265
3DGS*	18.28	0.641	0.267
Difix3D	18.86	0.658	0.239
ExploreGS	19.73	0.691	0.235
Ours	20.74	0.728	0.209

D Additional Evaluation on the LLFF Dataset

Our evaluation on the NerfBusters, WildExplore, and DL3DV datasets has already covered indoor scenes, outdoor scenes, and object-centric scenes. To further validate the performance of our method, we conduct additional experiments on the LLFF dataset (Mildenhall et al., 2019), which focuses on the forward-facing setting. We note that LLFF has fewer training and test views and is relatively simpler than the NerfBusters, WildExplore, and DL3DV datasets. Following prior work (Zhou et al., 2026), we adopt a sparse-view extrapolative setting, where 10 views are used for training and the remaining views are used for evaluation.

As shown in Table 8, we provide quantitative comparisons with the baseline methods on the LLFF dataset. In the forward-facing setting, the depth regularization in 3DGS* provides little improvement. In contrast, diffusion-enhanced methods improve the quality of rendered images at extrapolative test views by a significant margin. We also present qualitative comparisons on the LLFF dataset in Fig. 15. Our method produces more complete structures and fewer rendering artifacts.

E Additional Discussion on View Selection

As described in Sec. 3.3, we first sample a set of candidate pseudo views and then select pseudo views from them with RIG. Here we detail the candidate sampling. Following (Kim et al., 2025; Ni et al., 2026), we first extract the scene boundary from the initial 3DGS model and discretize

Table 9: Comparison of efficiency. FT Time denotes time for the fine-tuning iterations. Time is reported in minutes. We use FPS to measure inference efficiency.

Method	#Gaussians ↓	Train Time ↓	FT Time ↓	FPS ↑
3DGS	1.039M	21.3	10.9	130.9
3DGS*	1.038M	21.3	10.9	131.2
Difix3D	1.039M	21.3	19.5	131.0
ExploreGS	1.039M	21.3	29.7	130.9
Ours	1.039M	21.3	35.4	130.9

the scene into a voxel grid. We then generate candidate views by perturbing the training cameras, including translation, rotation, and orbiting, where the translation step is set according to the voxel size. Additional perturbation types can be included to further enlarge the candidate set. Note that we do not restrict the candidates to stay close to the training views in order to avoid unreliable pseudo ground truths. Instead, this trade-off is handled by RIG, which accounts for both the information gain of a view and the reliability of its pseudo ground truth.

F Efficiency Analysis

In this section, we analyze the efficiency of the methods. As described in Sec. B, all methods are trained for 45,000 iterations. The first 30,000 iterations follow the standard 3DGS training procedure with standard density control. The last 15,000 iterations mainly refine the 3DGS model using supervision from pseudo ground-truth images generated by diffusion models. For the 3DGS and 3DGS* baselines, we simply extend the training iterations without introducing pseudo images. As a result, different models have similar model sizes, measured by the number of Gaussians, and the training time is also similar.

For diffusion-enhanced methods, pseudo-view sampling and diffusion model inference introduce additional computational overhead in the fine-tuning stage. Difix3D uses a single-step image generation model (Sauer et al., 2024), which is slightly more efficient. In our method, the additional overhead mainly comes from computing the information gain and generating pseudo ground truths at candidate views with the diffusion model, while the DA3 inference for the reliability maps is fast and adds negligible cost. Compared to ExploreGS, our method increases the fine-tuning time from 29.7 to 35.4 minutes, which increases the overall training time by about 11%, while outperforming both Difix3D and ExploreGS by a clear margin. We note that the additional overhead occurs only during training. During inference, the rendering speed mainly depends on the model size and the rasterization method. Since all methods have similar model sizes and use the same rasterizer at inference time, the FPS is nearly the same. Results are averaged from experiments on the DL3DV dataset.

G Ablation on Group Size

Table 10: Ablation on the group size K in GBD on WildExplore.

K	Runtime (s) ↓	Memory (GB) ↓	PSNR ↑
1	131.3	0.91	16.63
3	179.2	1.65	16.77
5	237.5	2.34	16.82
10	366.1	4.92	16.88
15	685.7	8.79	16.89

The group size K in GBD controls the trade-off between the accuracy of information gain estimation and the computational cost. Given the total number of Gaussians N, the number of groups M is determined by K, i.e., $M \approx N/K$. Note that GBD reduces to the block-diagonal approximation when $K = 1$. We follow the setting of the GBD-based view selection in Tab. 5, where only the diffusion prior and view selection are applied, and conduct the ablation on the WildExplore dataset.

Table 11: Ablation on different diffusion models on NerfBusters.

Diffusion	PSNR ↑	SSIM ↑	LPIPS ↓
Difix3D (Wu et al., 2025)	15.78	0.468	0.433
Ours (Image)	16.73	0.488	0.401
ExploreGS (Kim et al., 2025)	16.29	0.479	0.434
Ours (Video)	17.15	0.523	0.381

As shown in Tab. 10, the computational cost increases with K, since the size of each block grows quadratically with K. In contrast, the improvement in reconstruction quality becomes marginal when K exceeds 10. We empirically find that $K = 5$ provides a good balance between computational efficiency and reconstruction quality.

H Ablation on Different Diffusion Models

In Table 11, we study the impact of different diffusion models. Our method primarily focuses on leveraging RIG to guide the integration of diffusion priors, and is therefore applicable to both image and video diffusion models in a plug-and-play manner. We observe that video diffusion models generally outperform image diffusion models, which can be attributed to their cross-frame consistency. Our RIG-guided framework improves the performance of both image- and video-diffusion-based baselines.

I Discussion on Diffusion Reliability

As shown in Fig. 16, the raw renderings from 3DGS at pseudo views are typically noisy due to insufficient constraints from the training views. When generating pseudo ground truths from such renderings, the diffusion model effectively removes most artifacts, but also produces unreliable content that deviates from the clean ground truths, as highlighted by the red arrows. However, the rendering uncertainty map of 3DGS fails to identify these regions, since it reflects the uncertainty of the 3DGS model rather than the reliability of the generated content, and its values are spread over the entire image. In contrast, our reliability map assesses the geometric consistency between the generated content and the training views, and assigns low reliability to the regions where the diffusion model produces unreliable artifacts. This enables DIGS to down-weight unreliable pseudo supervision in both view selection and optimization. We note that weakly textured regions that are rarely observed in the training views (e.g., the upper walls) also receive relatively low reliability, which conservatively reduces the supervision in these regions.

J Submodular Formulation of View Selection

In this section, we provide a theoretical analysis of our method for selecting the most informative pseudo views for diffusion prior injection. Our view selection method in Algorithm 1 is supported by submodular optimization. This analysis provides theoretical guarantees for the performance of our method compared to coverage-based view selection approaches (Marza et al., 2024; Kim et al., 2025).

Let $\bar{\mathcal{V}}_{candidate}$ denote the candidate pseudo-view pool, and let each candidate view $v \in \bar{\mathcal{V}}_{candidate}$ contribute a positive semi-definite reliability-weighted information matrix $\tilde{H}_v \succeq 0$. For any selected subset $S \subseteq \mathcal{V}_{candidate}$, define the accumulated information matrix as

$$H(S) = H_{\mathcal{V}_{train}} + \sum_{v \in S} \tilde{H}_v,$$

where $H_{\mathcal{V}_{train}}$ is the information from the observed training views, which is assumed to be positive definite. Using the log determinant measure defined in Sec. 3.1, $U(H) = \log |H|$, the pseudo-view selection objective can be written as

$$F(S) = U(H(S)) = \log \det(H(S)).$$

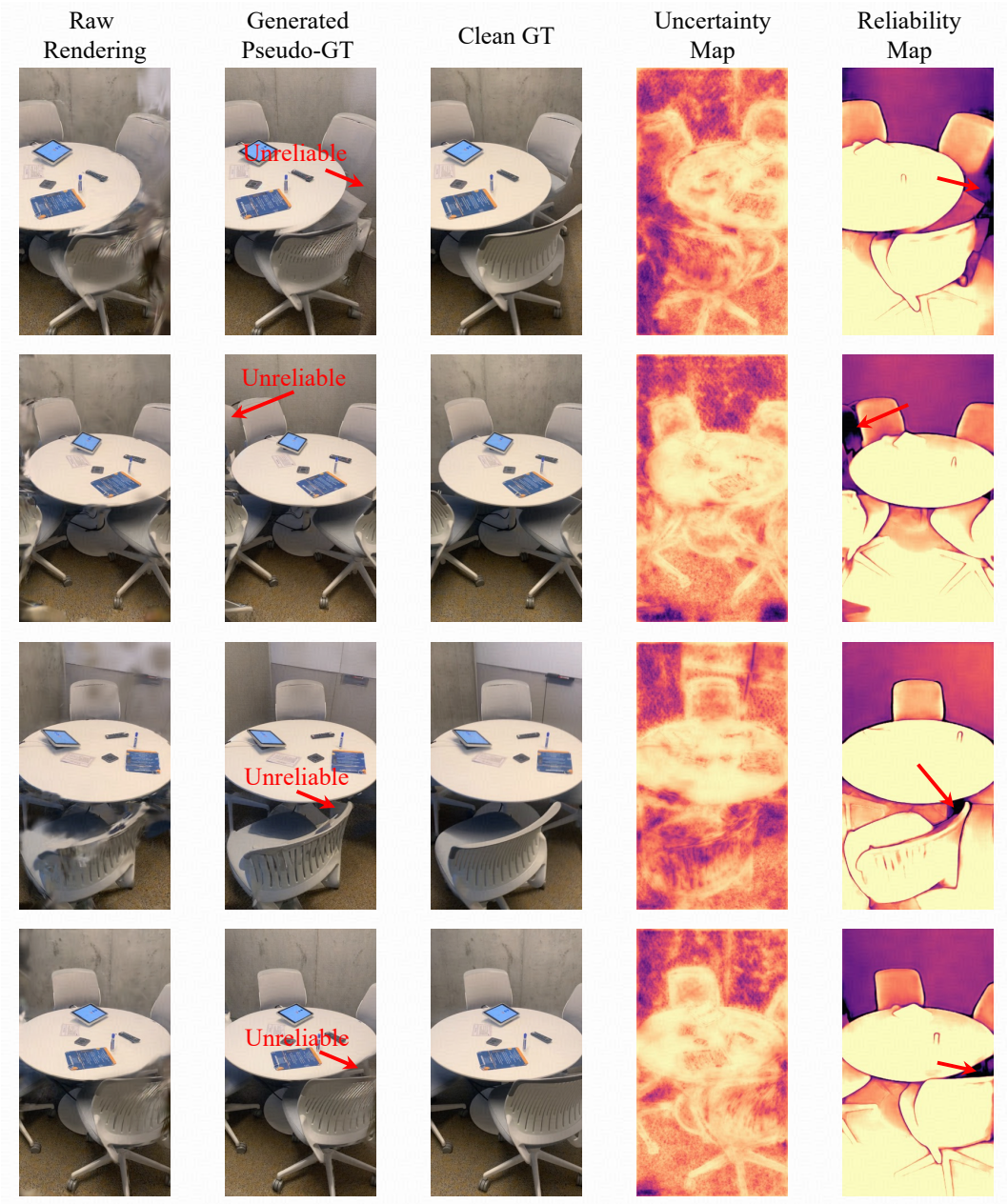

Figure 16: Visualization of unreliable content generated by the diffusion model. From left to right: raw renderings at pseudo views, generated pseudo ground truths, clean ground truths, rendering uncertainty maps of 3DGS, and our geometric consistency-based reliability maps. Red arrows highlight unreliable content, which is captured by our reliability map but not by the rendering uncertainty map.

This objective is a standard log-determinant information criterion (i.e., D-optimal design (Wilson et al., 2025)), since each $\tilde{H}_v \succeq 0$, the set function $F(S)$ is monotone submodular. In particular, for any $A \subseteq B \subseteq \bar{\mathcal{V}}_{candidate}$ and $s \in \bar{\mathcal{V}}_{candidate} \setminus B$, we have:

$$F(A \cup \{s\}) - F(A) \geq F(B \cup \{s\}) - F(B).$$

That is, adding a new pseudo-view yields a smaller marginal gain when the currently selected set is already more informative, which is precisely the diminishing-returns property.

Therefore, the view selection process in the while-loop in Algorithm 1 can be interpreted as a greedy algorithm for maximizing a monotone submodular function under the cardinality constraint $|S| \leq B$.

By the classical result for greedy submodular maximization (Nemhauser et al., 1978), if S_{greedy} is the subset returned by the algorithm and $S^\star$ is the optimal subset of size at most B, then the normalized objective $f(S) = F(S) - F(\emptyset)$, which equals the accumulated RIG of the selected views and satisfies $f(\emptyset) = 0$, is lower-bounded as

$$f(S_{\text{greedy}}) \geq \left(1 - \frac{1}{e}\right) f(S^\star).$$

Such a theoretical guarantee supports the effectiveness of our Fisher information–based view selection approach in Algorithm 1, which is generally more effective than coverage-based methods (Kim et al., 2025).

K Per-Scene Results

Table 12: Per-scene PSNR on the WildExplore dataset.

Method	Bicycle	Hat	Kitchen	Livingroom	Study	Table2	Towtractor	Vacantlot	**Avg.**
3DGS (Kerbl et al., 2023)	15.27	15.35	19.11	17.51	14.13	13.40	17.78	13.93	15.81
ExploreGS (Kim et al., 2025)	15.94	16.51	20.31	18.57	15.89	14.18	18.87	15.41	16.96
Ours	**16.63**	**16.99**	**21.01**	**19.64**	**16.72**	**15.38**	**19.52**	**16.35**	**17.78**

We report the per-scene PSNR on the WildExplore dataset in Tab. 12. Our method consistently outperforms 3DGS and ExploreGS across all scenes, covering both indoor (e.g., Kitchen and Study) and outdoor (e.g., Bicycle and Vacantlot) environments. This indicates that GBD effectively estimates the information gain of pseudo views, and RIG reliably identifies unreliable generated content across diverse scenes, rather than benefiting only a few specific scenes.

L Broader Impact and Limitation

Broader Impact. This work advances extrapolative 3D scene reconstruction by improving the reliability of 3D Gaussian Splatting in under-constrained regions through diffusion priors guided by reliable information gain. The proposed method can benefit applications that require robust 3D reconstruction from sparse or limited observations, such as robotics, AR/VR, digital twins, cultural heritage preservation, and autonomous navigation. By explicitly accounting for the rendering uncertainty and the reliability of generated pseudo ground truths, the framework also encourages more cautious use of generative priors and may reduce hallucinated or geometrically inconsistent reconstructions compared with existing diffusion-enhanced approaches. However, improved 3D reconstruction can also raise potential risks when applied to privacy-sensitive environments, surveillance, or unauthorized reconstruction of private spaces. In addition, diffusion-generated pseudo views may still introduce hallucinated content, which could be problematic in safety-critical applications if reconstructions are treated as factual measurements. We therefore recommend that downstream users clearly distinguish reconstructed or generated content from directly observed data, avoid deployment in privacy-sensitive settings without consent and safeguards, and perform task-specific validation before using the method in high-stakes applications.

Limitation. In this work, we introduce DIGS, a novel framework that incorporates diffusion priors into 3DGS guided by reliable information gain. Experimental results show that DIGS effectively improves the reconstruction quality under the extrapolative setting. However, we mainly focus on leveraging generative priors from diffusion models (Sauer et al., 2024; Liu et al., 2024a). Extending this work to Flow-Matching based models might further improve the performance. Also, we mainly considered standard 3DGS in the per-scene optimization setting. Combining model uncertainty in feed-forward 3DGS (Charatan et al., 2024; Wang et al., 2025; Chen et al., 2026) is a promising direction that we leave to future work.