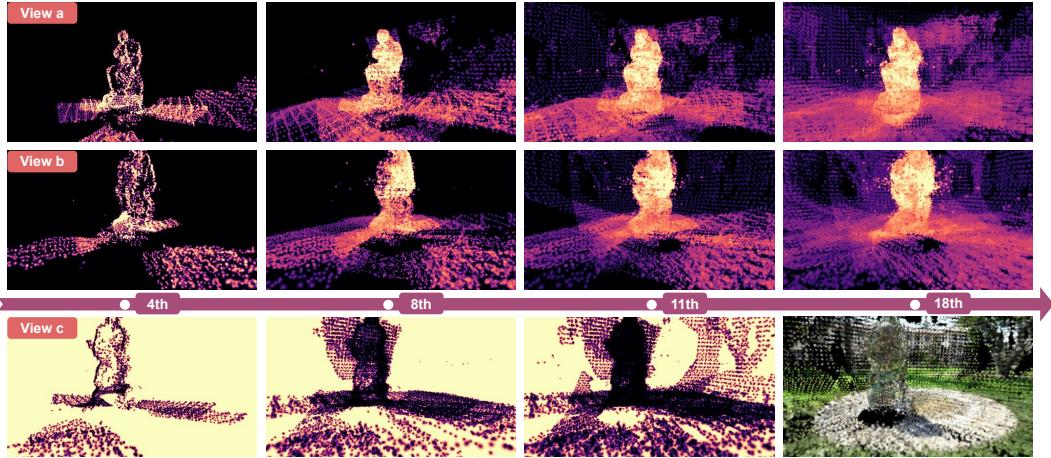


# 000 001 002 003 004 005 CREAT3R: CONFIDENCE REAGGREGATION FOR 006 EXPLORATION-AWARE ACTIVE 3D RECONSTRUCTION 007 008 009

010 **Anonymous authors**  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024

Paper under double-blind review



025  
026 Figure 1: The proposed *Creat3r* progressively refines scene geometry and confidence field across  
027 successive selection rounds. The top two rows respectively illustrate the progression of the *confidence*  
028 maps of views *a* and *b*. The bottom row shows the *exploration map* of view *c* alongside the  
029 reconstructed geometry, which becomes increasingly detailed as more images are incorporated. At  
030 each iteration, the confidence and exploration maps jointly quantify the exploration measure of every  
031 candidate view. The exploration map highlights unobserved or weakly constrained regions (bright)  
032 to drive exploration, whereas the confidence map quantifies the reliability of reconstructed points,  
033 enabling refinement in uncertain yet already observed areas.

## 034 ABSTRACT

035  
036 We introduce *Creat3r*, an active view selection framework designed for efficient  
037 and high-quality 3D reconstruction using a limited subset of image-pose pairs.  
038 Given an initial set of selected views, our method iteratively identifies the most  
039 informative candidate views to maximize reconstruction accuracy while adhering to  
040 computational constraints. Our approach begins by generating an intermediate 3D  
041 point cloud through dense pixel correspondences and stereo triangulation, refining  
042 point estimates via the Direct Linear Transform (DLT). To assess reconstruction  
043 reliability, we introduce a *3D confidence field* that integrates camera support  
044 and view consistency, enabling a quantitative evaluation of point quality. This  
045 confidence information is then propagated to all candidate views using an efficient  
046 Gaussian projection technique, generating *2D confidence and exploration maps*  
047 for each potential viewpoint. We define an exploration measure based on these  
048 maps to evaluate and optimally select the next best view. By balancing exploration,  
049 reconstruction accuracy, and computational efficiency, *Creat3r* is well-suited for  
050 applications in autonomous 3D scanning, robotic vision, and multi-view scene  
051 reconstruction. To demonstrate its effectiveness, our method is evaluated against  
052 baselines using the standard 3DGS representation for 3D reconstruction from the  
053 selected views. The experimental results show that our method excels in novel  
view synthesis and surface reconstruction, achieving significant improvements in  
SSIM and F1-score.

054  
055  

## 1 INTRODUCTION

056  
057  
058  
059  
060  
061  
062  
Recent advances in 3D Gaussian Splatting (3DGS) have enabled high-quality, real-time novel view  
synthesis from multi-view images. By representing a scene with a set of differentiable 3D Gaussians,  
3DGS has become a leading method for immersive and realistic scene rendering. However, its state-  
of-the-art performance remains critically dependent on a dense set of input views, often requiring  
hundreds of images for a single scene. This reliance creates a significant practical bottleneck, leading  
to heavy computational costs, protracted optimization times, and intensive manual data acquisition,  
particularly for large-scale environments.063  
064  
065  
066  
067  
068  
To alleviate this data acquisition burden, *active view selection* has emerged as a promising solution,  
aiming to intelligently choose a minimal yet informative subset of views/images. Existing methods  
for neural rendering typically operate in an iterative loop: they optimize the scene representation with  
a small set of known views, estimate uncertainty for unobserved regions, and select the “next best  
view” to reduce this uncertainty. For example, methods like FisherRF (Jiang et al., 2024) and Lyu  
et al. (2024) leverage Fisher information or variational inference to estimate model uncertainty.069  
070  
071  
072  
073  
074  
075  
076  
While effective, these approaches suffer from two major limitations. First, they are fundamentally  
tied to the iterative optimization of the underlying 3D representation. Each selection step requires  
re-initializing and re-optimizing the Gaussian model, introducing computational redundancy and  
inefficiency. Second, most existing pipelines initialize using Structure-from-Motion (SfM) methods  
such as COLMAP, which are typically run over the entire candidate image pool. This practice leaks  
information from views that are supposed to be “unseen” and biases the evaluation of selection  
strategies. *As a result, existing approaches do not fully respect the principle of active selection, since  
they implicitly assume access to geometric priors derived from all candidate views.*077  
078  
079  
080  
081  
To address these limitations, we introduce Creat3r, a novel active view selection framework designed  
to be efficient, robust, and fully decoupled from costly 3DGS optimization. Instead of relying on a  
resource-intensive, full Gaussian model, we propose a lightweight, proxy 3D model composed of  
spherical Gaussians. This proxy model is not optimized iteratively but is instead built on a more  
fundamental and robust geometric representation of the scene.082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
Creat3r operates through two key mechanisms. It incrementally estimates a robust geometry of  
the scene by establishing pairwise correspondences between known and candidate views, then  
triangulating them into 3D points via Direct Linear Transformation (DLT). This procedure produces  
a dynamic scaffold that grows with each selection step, thereby avoiding both the leakage bias of  
global SfM and the instability of random initialization in sparse-view settings. Building upon this  
scaffold, Creat3r defines a novel exploration-exploitation criterion using two geometry-derived maps.  
The *confidence map* encodes the reliability of reconstructed regions, guiding refinement in uncertain  
but already observed areas, while the *exploration map* highlights regions that remain unobserved or  
poorly constrained, directing the system toward novel content. (See Figure 1.) Together, these signals  
balance local detail refinement with global scene expansion. The overall pipeline is illustrated in  
Figure 2, where geometry and confidence are re-estimated each round to guide view selection.093  
094  
095  
096  
097  
Through successive selections, our method produces a compact set of images sufficient for high-  
quality 3DGS reconstruction, while also generating a robust scene scaffold that can serve as an  
effective initialization for downstream tasks. We demonstrate that Creat3r consistently outperforms  
prior view-selection methods on both novel view synthesis and surface reconstruction tasks, achieving  
superior performance while significantly reducing computational and data requirements.098  
Our key contributions can be summarized as follows:100  
101  
102  
103  
104  
105  
106  
107  
1. We introduce Creat3r, a novel view selection approach for 3DGS that is fully decoupled  
from the iterative optimization process, yielding substantial computational savings.  
2. We present a robust, geometry-based initialization that incrementally builds a sparse 3D  
scaffold using only the selected views, thereby avoiding leakage bias caused by global SfM  
pipelines that are misapplied in existing approaches.  
3. We propose a new exploration-exploitation selection measure that leverages a confidence  
map and an exploration map to intelligently guide view selection.  
4. We show that Creat3r achieves state-of-the-art results for both novel view synthesis and  
surface reconstruction, demonstrating superior performance and data efficiency.

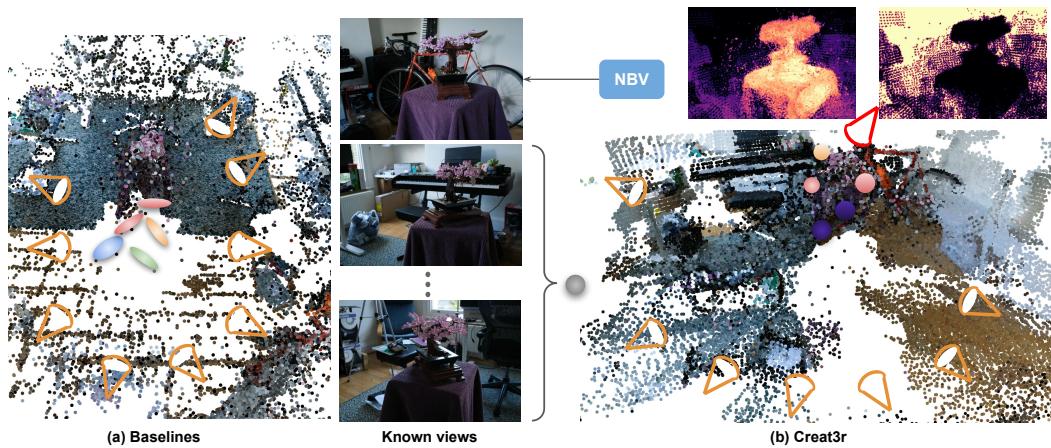


Figure 2: (a) Baseline methods re-initialize and re-optimize Gaussian ellipsoids at every iteration. Their initialization depends on the point cloud reconstructed with all the candidate views (orange cones), leading to information leakage. (b) Creat3r estimates Gaussian spheres directly from pairwise pixel correspondences. The resulting confidence field and geometry render confidence and exploration maps for each candidate view. The view with the highest exploration measure (red cone) is then selected as the *next best view* (NBV) and the newly acquired image is added to the known set.

## 2 RELATED WORK

3D reconstruction with classical structure-from-motion (SfM) (Schonberger & Frahm, 2016; Pan et al., 2024) or multi-view stereo (MVS) (Schönberger et al., 2016; Yao et al., 2018) is experiencing a renaissance with the advent of emerging radiance field (Mildenhall et al., 2021; Sun et al., 2022; Fridovich-Keil et al., 2022; Müller et al., 2022), signed distance fields (Wang et al., 2021; 2023; Li et al., 2023; Liu et al., 2023), and 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023; Huang et al., 2024; Dai et al., 2024).

Active 3D reconstruction determines the next best view that will most significantly enhance the quality of the reconstruction. ActiveNeRF (Pan et al., 2022) assumes parameters to be independent and estimates the uncertainty. ActiveNeuS (Ichimaru et al., 2024) extends active selection to surface reconstruction, but only for small objects. ActiveGAMER (Chen et al., 2025) uses 3DGS and RGBD inputs for next best view selection, but only in a synthetic world. NARUTO (Feng et al., 2024) and ActiveGS (Jin et al., 2024) also take RGBD inputs and extend to the real environment.

The method of Kopanas & Drettakis (2023) samples points in space and models the point camera relationship to select new views. They have adopted InstantNGP (Müller et al., 2022) to reconstruct the scene and place cameras in the empty space. FisherRF (Jiang et al., 2024) quantifies the uncertainty of each candidate using Fisher information. It uses Laplace’s approximation and computes Jacobians instead of the Hessian matrix. It modifies the rasterize function in 3DGS to speed up computation. Similar to FisherRF, the method of Goli et al. (2024) also uses Laplace’s approximation for uncertainty computation in NeRF. Like FisherRF, GauSS-MI (Xie et al., 2025) also uses information gains for view selection. Their computation does not involve known views, resulting in constant search time.

The manifold sampling technique proposed by Lyu et al. (2024) takes a different approach and uses variational inference to model the parameter distribution of Gaussian primitives. They find an effective low-dimensional manifold that can speed up computation and a differentiable scheme to optimize uncertainty. Also relying on variance, the method of Ewen et al. (2025) computes pixel-wise higher moments for each candidate. For selection, they compute the variance for each candidate.

Unlike the methods mentioned above, we do not rely on optimization of 3DGS or NeRF during view selection. Instead, we use 2D correspondence predictions (Lindenberger et al., 2023; Sarlin et al., 2020; Sun et al., 2021; Leroy et al., 2024) to estimate robust geometry for active view selection.

162 **3 METHOD**  
 163

164 Consider a collection of image–pose pairs  $\mathcal{S} = \{(I, W)\}$  representing a given scene or object, where  
 165 each image  $I$  is associated with its corresponding camera pose  $W$ . Selecting a view therefore entails  
 166 including both its image and pose. Our goal is to develop an *active view selection* framework that  
 167 enables high-quality 3D reconstruction while relying on only a limited subset of these pairs.

168 We initialize with a small set of *known* views,  $\mathcal{S}^K = \{(I^K, W^K)\}$ , which serve as the starting point  
 169 for reconstruction. The remaining *candidate* views form the set  $\mathcal{S}^C$ , such that the entire dataset  
 170 is partitioned as  $\mathcal{S} = \mathcal{S}^K \cup \mathcal{S}^C$ . The objective is to iteratively select the most “informative” views  
 171 from  $\mathcal{S}^C$  to expand  $\mathcal{S}^K$ , while respecting constraints on the number of views and optimization steps  
 172 allowed for active 3D reconstruction.  
 173

174 **3.1 ROBUST POINT ESTIMATION**  
 175

176 During active view selection, performing 3D reconstruction using the known views  $\mathcal{S}^K$  yields an  
 177 intermediate estimate of 3D point cloud  $\tilde{\mathcal{P}}$ , providing insight into the expected quality of the final  
 178 reconstruction. However, reconstruction quality is not just the goal of our method—it also plays a  
 179 critical role in shaping the effectiveness of the view selection criteria, as elaborated later.

180 Given the known set  $\mathcal{S}^K = \{(I^K, W^K)\}$ , we leverage correspondence networks such as Light-  
 181 Glue (Lindenberger et al., 2023) to extract pixel correspondences between all pairwise views in  $\mathcal{S}^K$ .  
 182 These correspondences are triangulated to recover 3D points, explicitly incorporating stereo geometry  
 183 for accurate predictions. More formally, consider two images  $I_a$  and  $I_b$  in  $\mathcal{S}^K$ , captured by cameras  
 184  $a$  and  $b$ , respectively. The network identifies a set of corresponding pixels, denoted as  
 185

$$\{(u_a, v_a) \leftrightarrow (u_b, v_b)\}. \quad (1)$$

187 For camera  $a$  with projection matrix  $\mathbf{P}^a$ , the 3D-to-2D projection equation can be expressed as  
 188

$$\kappa_a \cdot \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{11}^a & \mathbf{P}_{12}^a & \mathbf{P}_{13}^a & \mathbf{P}_{14}^a \\ \mathbf{P}_{21}^a & \mathbf{P}_{22}^a & \mathbf{P}_{23}^a & \mathbf{P}_{24}^a \\ \mathbf{P}_{31}^a & \mathbf{P}_{32}^a & \mathbf{P}_{33}^a & \mathbf{P}_{34}^a \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \quad (2)$$

193 where  $\kappa_a$  is a scalar depth factor. The same formulation applies to camera  $b$  by substituting its  
 194 corresponding projection matrix  $\mathbf{P}^b$ .

195 To eliminate the depth scale factor  $\kappa$ , we construct a linear system by multiplying the third row by  $u$   
 196 and  $v$ , and subtracting it from the first and second rows, respectively, for both cameras. This results  
 197 in the following system:  $\mathbf{A}[x, y, z]^T = \mathbf{b}$ , where

$$\mathbf{A} = \begin{bmatrix} u_a \mathbf{P}_{31}^a - \mathbf{P}_{11}^a & u_a \mathbf{P}_{32}^a - \mathbf{P}_{12}^a & u_a \mathbf{P}_{33}^a - \mathbf{P}_{13}^a \\ v_a \mathbf{P}_{31}^a - \mathbf{P}_{21}^a & v_a \mathbf{P}_{32}^a - \mathbf{P}_{22}^a & v_a \mathbf{P}_{33}^a - \mathbf{P}_{23}^a \\ u_b \mathbf{P}_{31}^b - \mathbf{P}_{11}^b & u_b \mathbf{P}_{32}^b - \mathbf{P}_{12}^b & u_b \mathbf{P}_{33}^b - \mathbf{P}_{13}^b \\ v_b \mathbf{P}_{31}^b - \mathbf{P}_{21}^b & v_b \mathbf{P}_{32}^b - \mathbf{P}_{22}^b & v_b \mathbf{P}_{33}^b - \mathbf{P}_{23}^b \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} \mathbf{P}_{14}^a - u_a \mathbf{P}_{34}^a \\ \mathbf{P}_{24}^a - v_a \mathbf{P}_{34}^a \\ \mathbf{P}_{14}^b - u_b \mathbf{P}_{34}^b \\ \mathbf{P}_{24}^b - v_b \mathbf{P}_{34}^b \end{bmatrix}. \quad (3)$$

203 We solve for  $[x, y, z]^T$  using the Direct Linear Transform (DLT):  $[x, y, z]^T = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$ .  
 204 Applying this to all pixel correspondences across view pairs in  $\mathcal{S}^K$  yields *co-visible* points whose 3D  
 205 coordinates are refined under stereo geometry, ensuring robust and accurate predictions for active  
 206 view selection.  
 207

208 **3.2 CONFIDENCE FIELD VIA REAGGREGATION**  
 209

210 Given the intermediate 3D reconstruction  $\tilde{\mathcal{P}}$ , we assign each point a confidence value to guide the  
 211 selection of the next best view from  $\mathcal{S}^C$ . Specifically, for each predicted point  $p \in \tilde{\mathcal{P}}$  with color  $\mathbf{c}$   
 212 and position  $\mathbf{p}$ , we evaluate its visibility and consistency across the known cameras in  $\mathcal{S}^K$ .

213 For every camera  $a \in \mathcal{S}^K$ , point  $p$  can be projected onto the image plane if it lies within the camera  
 214 frustum. We define the binary *support* of camera  $a$  for point  $p$  as  
 215

$$g(a, p) \in \{0, 1\}, \quad g(a, p) = 1 \text{ iff } p \text{ is visible in camera } a. \quad (4)$$

Visibility alone does not guarantee reliability, as points may suffer from occlusions, clutter, or false correspondences. To capture photometric consistency, we project  $p$  onto all supporting cameras and measure its color agreement with the corresponding pixels. We define the *viewing consistency* by

$$H(p) = \exp \left( -\frac{1}{\sum_{a \in \mathcal{S}^K} g(a, p)} \times \sum_{n=1}^{|\mathcal{S}^K|} g(a_n, p) \|\mathbf{c}_n - \mathbf{c}\|_2 \right), \quad (5)$$

where  $\mathbf{c}$  is the color of  $p$ , and  $\mathbf{c}_n$  is the observed pixel color in camera  $a_n$ . If  $p$  is not supported by any camera, we set  $H(p) = 0$ .

Finally, we define the **3D confidence field**  $\mathcal{M}_{\text{Conf}}$  by weighting  $H(p)$  with  $\bar{g}(p)$ , the fraction of known cameras in which  $p$  is visible:

$$\mathcal{M}_{\text{Conf}}(p) = H(p) \times \bar{g}(p) = H(p) \times \frac{1}{|\mathcal{S}^K|} \sum_{a \in \mathcal{S}^K} g(a, p). \quad (6)$$

After each round of active view selection, we update the confidence field  $\mathcal{M}_{\text{Conf}}$  for the newly constructed  $\tilde{\mathcal{P}}$  by pointwise reaggregating the support and reevaluating the color consistency from all relevant cameras in  $\mathcal{S}^K$ .

### 3.3 VIEW-SPECIFIC CONFIDENCE AND EXPLORATION MAPS

Thus far, we have leveraged stereo geometry to improve the quality of the intermediate 3D reconstruction  $\tilde{\mathcal{P}}$  and to compute its confidence field  $\mathcal{M}_{\text{Conf}}$ . The next step is to propagate this information to each candidate camera view in  $\mathcal{S}^C$ , generating corresponding 2D confidence and exploration maps. These maps provide a quantitative basis for assessing which candidate views will contribute most effectively to the next iteration of active view selection.

Inspired by 3D Gaussian Splatting (3DGS), we adopt a simplified projection scheme to transfer 3D point information to candidate views. Each point  $p \in \tilde{\mathcal{P}}$  is modeled as a sphere centered at  $\mathbf{p}$  with radius  $r$  denoting its *influence region*. For an arbitrary position  $\mathbf{x} \in \mathbb{R}^3$ , the influence of  $p$  is approximated using an *isotropic* Gaussian function:

$$G(\mathbf{x}) = o \cdot \exp \left( -\frac{1}{2r^2} \|\mathbf{x} - \mathbf{p}\|^2 \right), \quad (7)$$

where  $o$  is a constant opacity. Compared to the full 3DGS formulation, the isotropic case simplifies projection: the projected radius in the image plane is  $r^{2D} = r \cdot f/\lambda$ , where  $f$  is the focal length and  $\lambda$  denotes the depth of the Gaussian. Projected Gaussians are depth-sorted and alpha-composited to obtain the final pixel value. Leveraging this projection formulation, we determine the influence region  $r$  by constraining the projected radius  $r^{2D}$  in the source view to correspond to exactly one pixel.

Using  $\tilde{\mathcal{P}}$  and  $\mathcal{M}_{\text{Conf}}$ , we generate the **2D exploration map**  $\mathbf{M}_{\text{Exp}}$  and **2D confidence map**  $\mathbf{M}_{\text{Conf}}$  for each candidate view. For  $\mathbf{M}_{\text{Exp}}$ , we assign each point a grayscale value of one, project it, and then invert the rendered image to emphasize unexplored or weakly constrained regions. For  $\mathbf{M}_{\text{Conf}}$ , we set the intensity of each point to its confidence value in  $[0, 1]$ . In both cases, we fix the opacity in Equation (7) to  $o = 0.8$ .

More concretely, these maps are generated for each candidate view by rendering from a simplified 3DGS model. For the exploration map, reconstructed points are modeled as small spheres, projected with maximum grayscale intensity (white), and blurred by the Gaussian kernel in Equation (7). The rendered image is then inverted so that unobserved or weakly constrained regions appear bright. The confidence map is produced analogously, except each point is colored by its confidence score rather than a uniform white value, thereby encoding the reliability of observed regions.

**Exploration measure** The exploration and confidence maps,  $\mathbf{M}_{\text{Exp}}$  and  $\mathbf{M}_{\text{Conf}}$ , provide a quantitative way to evaluate the contribution of each candidate view in  $\mathcal{S}^C$ , given the set of already selected views  $\mathcal{S}^K$ . Intuitively, selecting a new view reduces the unexplored content of nearby candidates due to overlapping coverage, while views observing disjoint regions remain more valuable for selection. Building on this intuition, we define the *exploration measure* for a candidate view  $(I, W) \in \mathcal{S}^C$  as

$$\text{Exploration}(W) = \sum \mathbf{M}_{\text{Exp}}(W) - \tau \cdot \bar{\mathbf{M}}_{\text{Conf}}(W), \quad (8)$$

270 Table 1: Novel view synthesis evaluation on Mip-NeRF 360 dataset. We present the evaluations of  
 271 3DGS optimized with 10 and 20 selected views. (\*) denotes methods initialized with COLMAP-  
 272 induced subsampled points. (‡) indicates methods initialized with Creat3r-LightGlue points, and (†)  
 273 indicates methods initialized with Creat3r-MASt3R. **Best** and **second-best** results are highlighted.  
 274

Method	10 cameras			20 cameras		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
FPS*	12.529	0.261	0.613	14.918	0.389	0.528
FPS†	13.560	0.362	0.555	14.940	0.439	0.514
FisherRF (Jiang et al., 2024)*	12.625	0.264	0.608	15.434	0.390	0.515
FisherRF (Jiang et al., 2024)†	14.196	0.392	0.546	16.028	0.474	0.491
Kopanas & Drettakis (2023)*	13.022	0.284	0.596	15.658	0.407	0.506
Kopanas & Drettakis (2023)†	13.677	0.383	0.546	15.727	0.470	0.499
Lyu et al. (2024)*	12.561	0.264	0.612	15.446	0.401	0.518
Lyu et al. (2024)†	14.282	0.377	0.547	16.264	0.503	0.480
Creat3r ‡	<b>16.040</b>	<b>0.449</b>	<b>0.536</b>	<b>19.637</b>	<b>0.567</b>	<b>0.443</b>
Creat3r †	<b>17.809</b>	<b>0.511</b>	<b>0.523</b>	<b>20.678</b>	<b>0.601</b>	<b>0.397</b>

288  
 289 where the first term quantifies the total unexplored regions visible from  $W$ , the second term penalizes  
 290 views with high average confidence (i.e., already well-covered), and  $\tau$  is a scaling factor that balances  
 291 the two terms. The next best view is then selected by maximizing this exploration measure:  
 292

$$(I^*, W^*) = \arg \max_{(I, W) \in S^C} \text{Exploration}(W). \quad (9)$$

## 4 EXPERIMENT

293 To evaluate our method, we perform comprehensive comparisons on 3D reconstruction. The first  
 294 task is novel view synthesis. The experimental setting follows previous methods, and the results are  
 295 detailed in Section 4.2. The second task is surface reconstruction. This is a more severe task and has  
 296 not been discussed by previous active view-selection methods. It is shown that our method is capable  
 297 of reconstructing the surface under limited views. Further discussion is presented in Section 4.3.  
 298

### 4.1 IMPLEMENTATION DETAILS

307 Creat3r uses pixel correspondences to estimate robust geometry. Any correspondence estimation  
 308 method can be used in our framework. In the experiment, we report the evaluation with two different  
 309 correspondence estimation methods, LightGlue and MASt3R, for novel view synthesis. Note that we  
 310 simply treat MASt3R as a correspondence network for pixel matching. We do not use their point  
 311 estimate in the entire process of view selection. Our framework is 3D model agnostic, meaning  
 312 it can accept any 3D reconstruction technique. For a fair comparison, we use 3DGS as our 3D  
 313 representation method, following the baselines. Since Creat3r provides a reliable 3D scaffold, 3DGS  
 314 converges in a very short time: We finish 3DGS optimization in 5,000 iterations. Computational  
 315 efficiency is discussed in the appendix, and Table 5 shows the average selection time per iteration.  
 316

317 We compare Creat3r with the previous state of the arts. Specifically, we consider FisherRF (Jiang  
 318 et al., 2024), Lyu et al. (2024) and Kopanas & Drettakis (2023) as competitive counterparts. Since  
 319 the method of Kopanas & Drettakis (2023) is originally designed for NeRF, we adapt their method to  
 320 3DGS. We also use the 3DGS version of FisherRF in a single selection manner. All of the baseline  
 321 methods have unified searching and optimization iterations. We follow the setting of Lyu et al. (2024),  
 322 which takes 20,000 iterations for searching. They take another 10,000 iterations for final optimization.  
 323 In addition to the uncertainty estimation approaches, we construct a simple baseline through farthest  
 324 point sampling (FPS). This method only considers the position of each camera and collects the views  
 325 with the largest inner distances. Despite the absence of camera orientation, this strategy can be useful,  
 326 especially in an inward-captured dataset.

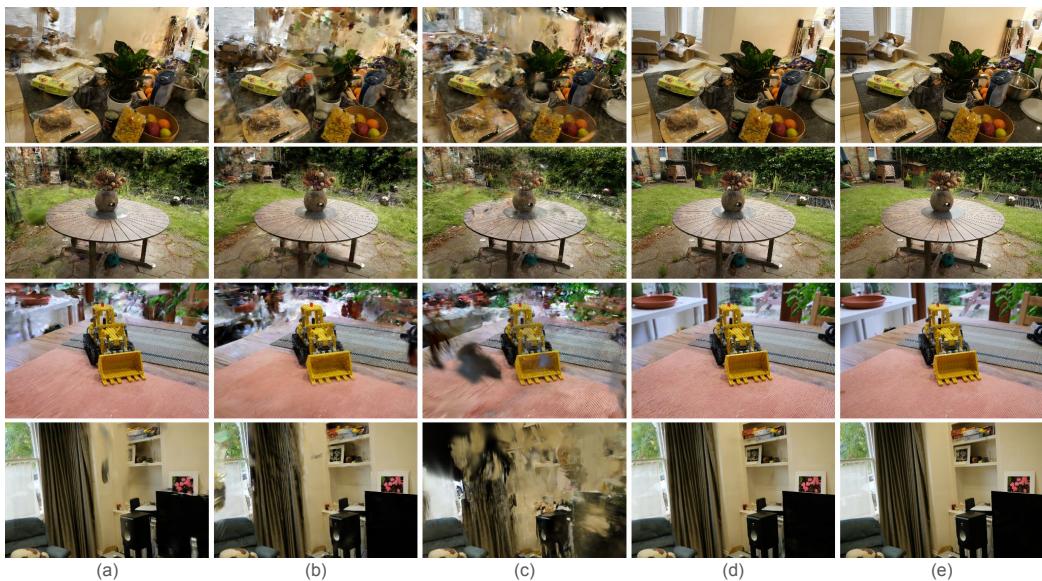


Figure 3: Qualitative comparison of active view selection on the Mip-NeRF 360 dataset for 20 selected views. The results demonstrate novel view renderings produced by competing methods: (a) FisherRF (Jiang et al., 2024), (b) Kopanas & Drettakis (2023), (c) Lyu et al. (2024), and (d) our proposed method, Creat3r. Column (e) serves as the ground truth novel view reference. The comparison highlights the superior ability of Creat3r to synthesize high-fidelity views with finer details and fewer artifacts.

#### 4.2 NOVEL VIEW SYNTHESIS

The evaluation process for novel view synthesis includes several steps. First, each method selects a certain number of views for 3DGS optimization. The optimized model renders images in novel poses. The evaluation compares the image quality between rendered and ground-truth images. Standard metrics include peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS). The three metrics reflect different perspectives of image quality. When an image has better quality, it should have higher PSNR and SSIM, and also lower LPIPS. In this task, we use the popular MipNeRF-360 dataset for evaluation. The dataset contains nine real-world scenes, including indoor and outdoor captures. To fully expose the data efficiency of each method, the number of selections is set to 10 and 20, respectively. In the original data split, each scene has hundreds of views for optimization. Here, the training set is used as the candidate pool to find the optimal selections.

In previous literature, the evaluation process includes the sparse reconstruction from SfM. As mentioned earlier, the point cloud is reconstructed from hundreds of views that are actually treated as candidates during the selection. Using this point cloud for 3DGS initialization would reveal geometric information and lead to unfair/biased comparisons. To develop a fair comparison, we provide three different kinds of initialization. The first one is subsampling initialization. To prevent unlimited space sampling, we use the extreme values of the SfM point cloud coordinates (induced by COLMAP) as boundaries and sample within them. We are also interested in the case where other competing methods have the same initial points as ours. The second and third initializations share the same point sets estimated with Creat3r and initial views. Each of them relates to LightGlue and MASt3R-matching, respectively.

The evaluation results are listed in Table 1. A more complete comparison can be found in the Table 6. We use  $(^*)$  to indicate that the methods use COLMAP-induced sampling initialization.  $(^{\ddagger})$  and  $(^{\dagger})$  indicate the methods start with robust points generated with Creat3r and initial views. Each baseline methods report two results with COLMAP-induced subsampling and Creat3r-MASt3R-matching initialization. Due to space limit, we report the baseline results with Creat3r-LightGlue in

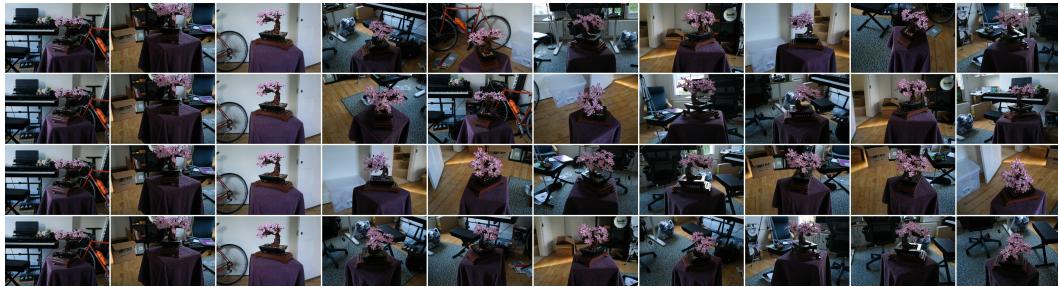


Figure 4: **Visualization of active view selection sequences on the ‘bonsai’ scene.** The rows correspond to (from top to bottom): Creat3r, Lyu et al. (2024), FisherRF (Jiang et al., 2024), and Kopanas & Drettakis (2023). The first three columns display the fixed initial set, adopted from the ReconFusion benchmark (Wu et al., 2024).

the appendix. Our Creat3r-MASt3R-matching initialization has consistent improvements in all the baselines, compared to the COLMAP-induced subsampling counterpart.

For different active selection strategies, FPS shows basic performance as it only considers the positions of the camera and ignores the orientations. FisherRF and Lyu et al. (2024) have similar performance. FisherRF performs better when the view is sparse. The method finds a candidate with the most information gain and achieves a significant improvement in the initial selections. The manifold sampling method of Lyu et al. (2024) performs better when there are more views. They use posterior to the scene, which is more accurate when there are more observations. The method of Kopanas & Drettakis (2023) considers the visibility and viewing directions of sampled points. Their method has better performance in indoor or area-constrained scenes. Interestingly, the method of Kopanas & Drettakis (2023) outperforms FisherRF and Lyu et al. (2024) when the initialization is COLMAP-induced subsampling, while FisherRF and Lyu et al. (2024) surpass Kopanas & Drettakis (2023) when using Creat3r estimated points as initialization. This suggests that uncertainty estimation, either with information gain or variational inference, is more beneficial from robust geometry, while Kopanas & Drettakis (2023) is less dependent on geometry.

Creat3r outperforms all baselines in all metrics, regardless of the number of selections. Due to its robust geometry, our 3D representation requires only half of the iterations for optimization. Note that optimization is difficult due to sparse views and the absence of ground-truth points. Creat3r estimates robust geometry, projects confidence and exploration maps to each candidate, and carefully selects the next best view by exploration measure. All efforts significantly improve the novel view quality. The comparison validates that our design is effective in various real-world scenes.

The qualitative results are illustrated in Figure 3. The figure demonstrates novel view renderings of four independent scenes in Mip-NeRF 360 dataset. The comparison highlights the superior ability of Creat3r to synthesize high-fidelity views with finer details and fewer artifacts. Other competing methods render with some artifacts due to a suboptimal selection set. More qualitative comparisons are shown in Figure 5 in the appendix.

The view selection sequences generated by different approaches are visualized in Figure 4. The initial set, comprising the first three views, is adopted from the ReconFusion benchmark (Wu et al., 2024). The top row of Figure 4 illustrates the selection process of Creat3r, which exhibits a spatially diverse distribution and progressively achieves comprehensive scene exploration. In contrast, the second and third rows—representing selections driven by the uncertainty estimates of Lyu et al. and FisherRF—reveal that while these methods explore the scene, they suffer from intermittent redundancy. Finally, the bottom row indicates that the approach of Kopanas & Drettakis results in a highly repetitive selection pattern.

### 4.3 SURFACE RECONSTRUCTION

The task aims to reconstruct the actual surface of the scene. While 3DGS does not produce an actual surface, we adopt the mesh extraction pipeline from 2DGS (Huang et al., 2024). After optimization,

432 Table 2: Surface reconstruction evaluation on Tansk&Temples dataset. **Best** results are highlighted.  
433

	Precision(%)	Recall(%)	F1-score(%)
FisherRF (Jiang et al., 2024)	7.91	9.81	8.61
Kopanas & Drettakis (2023)	<b>17.6</b>	0.58	0.84
Lyu et al. (2024)	5.85	5.86	5.61
Creat3r	14.09	<b>25.93</b>	<b>18.05</b>

443 we render depth maps for selected views. The depths are then fused into a voxel grid using truncated  
444 signed distance fusion Curless & Levoy (1996) and extracted via marching cubes Lorensen & Cline  
445 (1998). The evaluation process densely samples the reconstructed surface and compares it against  
446 the ground truth. The predicted point is considered valid if it is within a 5-millimeter distance from  
447 the ground-truth points. The reported metrics are precision, recall, and F1-score. In this evaluation,  
448 we use the popular Tanks&Temples dataset as the benchmark. Our setting is similar to GOF (Yu  
449 et al., 2024), which samples three scenes for evaluation, including “Caterpillar”, “Ignatius”, and  
450 “Truck”. The scenes are more difficult than Mip-NeRF 360 scenes and exhibit a wide variety of  
451 lighting conditions, such as sunshine and reflective surfaces. Each scene provides the surface ground  
452 truth of the foreground object. Only the foreground surface is evaluated. To reconstruct the surface,  
453 all the competitors must find optimal view collections of 20 views that cover most of the appearance  
454 and regional detail.

455 The results are shown in Table 2. Note that the numerical values are shown in percentages. Since  
456 the three scenes are outdoors and have different exposures across views, it is challenging for 3DGS  
457 to model the scene, as the optimization solely relies on appearance differences. While the methods  
458 of FisherRF, Kopanas & Drettakis (2023), and Lyu et al. (2024) depend on optimized 3DGS for  
459 view selection, they face challenges when the optimization fails. On the other hand, Creat3r is not  
460 affected by 3DGS performance. **Although our method yields lower precision compared to Kopanas**  
461 **& Drettakis (2023), their approach suffers from extremely limited surface coverage (low recall),**  
462 **resulting in a compromised F1-score. In contrast, Creat3r maintains a superior balance between**  
463 **precision and recall.**

464 The robust geometry enables view selection even in challenging scenes, especially for luminance  
465 variation across views. To further validate the point, we compare Creat3r with the original 3DGS.  
466 To prevent the influence of sparse views, the 3DGS optimization uses the entire training set, which  
467 includes 200 to 400 images. The comparison is listed in Table 3. The result aligns with our point.  
468 The scene “Caterpillar” is captured in the rural field. The images have severe exposure differences.  
469 Our method is not affected by this adversary. On the other hand, “Ignatius” has a sculpture in the  
470 foreground. The material of the sculpture reflects specular light, which leads to inconsistencies across  
471 views. Creat3r has the same performance with 3DGS on the “Truck” scene, while 3DGS uses 11  
472 times more images for training. It suggests that our selection criterion is effective and provides data  
473 efficiency for 3DGS optimization.

#### 474 4.4 ABLATION STUDY

475 In the ablation study, we focus on three components of Creat3r: namely, robust point estimate,  
476 confidence reaggregation, and exploration. The study excludes one component at a time, evaluating  
477

478 Table 3: F1-score comparison between Creat3r and original 3DGS. Creat3r only optimizes with 20  
479 selected views. 3DGS uses all of the training set for optimization (more than 200.)  
480

	Caterpillar	Ignatius	Truck
3DGS (Kerbl et al., 2023)	0.08	0.04	<b>0.19</b>
Creat3r	<b>0.10</b>	<b>0.25</b>	<b>0.19</b>

486 Table 4: Ablation study of Creat3r. Influence comparison of robust point estimate, exploration, and  
 487 confidence reaggregation. We use nine different scenes in Mip-NeRF 360 for evaluation.

489	Point	Exploration	Confidence	PSNR↑	SSIM↑	LPIPS↓
490		✓	✓	16.479	0.458	0.574
491	✓		✓	17.022	0.507	0.525
492	✓	✓		17.457	0.502	0.543
493	✓	✓	✓	<b>17.809</b>	<b>0.511</b>	<b>0.523</b>

498 the performance drop for the exclusion. The evaluation uses nine scenes in the Mip-NeRF 360, and  
 499 the number of selections is set to 10. The comparison is shown in Table 4. The first row excludes  
 500 the point estimate. Instead, we use MASt3R predicted points as an alternative. Compared to our  
 501 full model in the fourth row, the robust point estimate is the most effective technique, yielding an  
 502 improvement of 1.26 in PSNR. Exploration and confidence reaggregation provide different aspects  
 503 of improvement. While confidence reaggregation performs better in SSIM and LPIPS metrics,  
 504 the exploration has a higher performance in PSNR. In the experiment, we find that exploration  
 505 performs better in constrained scenes, such as indoor environments. On the other hand, confidence  
 506 performs better on outdoor scenes. With this in mind, we achieve the best of both worlds through  
 507 the exploration measure and obtain an overall better performance. [The qualitative evaluation of  
 508 the ablated components is presented in Figure 6 \(Appendix\), clearly demonstrating their respective  
 509 functionalities.](#)

## 511 5 LIMITATIONS AND DISCUSSION

512  
 513 A primary limitation of Creat3r stems from its reliance on geometric co-visibility between the  
 514 candidate views and the current reconstruction scaffold. While our approach proves highly effective  
 515 for inward-facing (object-centric) and forward-facing trajectories, outward-facing scenarios (e.g.,  
 516 the ‘room’ scene) present a distinct challenge. In such cases, candidate views often observe disjoint  
 517 regions of the scene and may share minimal overlap with the initial estimated geometry. Consequently,  
 518 these views yield null confidence maps and uniformly high-intensity exploration maps. This ambiguity  
 519 can inadvertently bias the exploration measure towards redundant sampling of unconstrained regions,  
 520 leading to suboptimal convergence. We emphasize that this vulnerability is inherent to the active  
 521 selection paradigm; all baseline methods similarly struggle to identify informative views in the  
 522 absence of initial geometric overlap. To mitigate this, we implement a regularization strategy  
 523 that temporarily masks such candidates from the selection pool. These views are subsequently  
 524 reintroduced as the confidence field expands and sufficient geometric connectivity is established to  
 525 meaningfully constrain them.

## 526 527 6 CONCLUSION

528 We introduced Creat3r, a novel active view selection framework for 3D reconstruction that is  
 529 computationally efficient and robust. Unlike prior methods that rely on iterative optimization of the  
 530 3D representation, Creat3r is fully decoupled from this process, leading to substantial computational  
 531 savings. Our method incrementally builds a robust 3D scaffold from only the selected views,  
 532 effectively avoiding the information leakage bias inherent in global SfM pipelines. Through the  
 533 use of confidence map and exploration map, Creat3r balances the need for local detail refinement  
 534 with global scene expansion. Our approach consistently achieves state-of-the-art results for both  
 535 novel view synthesis and surface reconstruction. By providing a reliable geometry, Creat3r also  
 536 significantly reduces the optimization time for downstream tasks, such as 3DGS, highlighting its  
 537 superior performance and data efficiency. This work represents a significant step forward in making  
 538 high-quality 3D reconstruction feasible with a minimal and informative set of images.

540 ETHICS STATEMENT  
541542 Our study uses only publicly available image datasets, accessed and used under the research licenses  
543 of the datasets. We conducted no interaction or intervention with individuals and accessed no private  
544 or non-public data. We have carefully considered potential impacts and do not anticipate ethical risks  
545 beyond those commonly encountered in computer vision and machine learning research.546  
547 REPRODUCIBILITY STATEMENT  
548549 We detail the training pipeline and evaluation protocols in both the main paper and the appendix,  
550 providing explicit hyperparameters and dataset partitions to facilitate exact replication. Additionally,  
551 we outline the implementation settings and reporting conventions to ensure our results can be  
552 accurately reproduced. To further support reproducibility, we will release the complete training and  
553 evaluation code, along with runnable scripts, in the camera-ready version.554  
555 THE USE OF LARGE LANGUAGE MODELS (LLMs)  
556557 We used LLMs to assist with (i) improving prose (grammar, flow, and clarity), (ii) reorganizing and  
558 refining section structure, captions, and titles, and (iii) generating keywords and query strings to  
559 explore related work.560 For literature discovery, every citation in the paper was located through standard search engines  
561 or digital libraries and then read and verified by the authors; we did not accept model-generated  
562 references without inspection. Numerical results, comparisons, and quotes were cross-checked  
563 against the original sources.564 We reviewed all text suggested by the model to ensure that the writing aligns with our intentions,  
565 thereby confirming accuracy and originality. No confidential or sensitive information was shared  
566 with the model. The authors conducted a thorough review of the final manuscript, including all tables  
567 and figures, to verify factual accuracy and completeness.568 REFERENCES  
569570 Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, and Yi Xu.  
571 Activegamer: Active gaussian mapping through efficient rendering. In *Proceedings of the Computer*  
572 *Vision and Pattern Recognition Conference*, pp. 16486–16497, 2025.573 Brian Curless and Marc Levoy. A volumetric method for building complex models from range images.  
574 In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*,  
575 pp. 303–312, 1996.576 Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, and Weiwei Xu. High-quality  
577 surface reconstruction using gaussian surfels. In *ACM SIGGRAPH 2024 Conference Papers*.  
578 Association for Computing Machinery, 2024.579 Parker Ewen, Hao Chen, Seth Isaacson, Joey Wilson, Katherine A Skinner, and Ram Vasudevan.  
580 These magic moments: Differentiable uncertainty quantification of radiance field models. *arXiv*  
581 *preprint arXiv:2503.14665*, 2025.582 Ziyue Feng, Huangying Zhan, Zheng Chen, Qingan Yan, Xiangyu Xu, Changjiang Cai, Bing Li,  
583 Qilun Zhu, and Yi Xu. Naruto: Neural active reconstruction from uncertain target observations.  
584 In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.  
585 21572–21583, 2024.586 Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinzhong Chen, Benjamin Recht, and Angjoo  
587 Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF*  
588 *conference on computer vision and pattern recognition*, pp. 5501–5510, 2022.589 Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes’ rays:  
590 Uncertainty quantification for neural radiance fields. In *Proceedings of the IEEE/CVF Conference*  
591 *on Computer Vision and Pattern Recognition*, pp. 20061–20070, 2024.

594 Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for  
 595 geometrically accurate radiance fields. In *SIGGRAPH 2024 Conference Papers*. Association for  
 596 Computing Machinery, 2024. doi: 10.1145/3641519.3657428.

597

598 Kazuto Ichimaru, Takaki Ikeda, Diego Thomas, Takafumi Iwaguchi, and Hiroshi Kawasaki. Ac-  
 599 tiveneus: Neural signed distance fields for active stereo. In *2024 International Conference on 3D*  
 600 *Vision (3DV)*, pp. 539–548. IEEE, 2024.

601

602 Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherrf: Active view selection and uncertainty  
 603 quantification for radiance fields using fisher information. In *European Conference on Computer*  
 604 *Vision*, pp. 422–440. Springer, 2024.

605

606 Liren Jin, Xinguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. Activevgs:  
 607 Active scene reconstruction using gaussian splatting. *arXiv preprint arXiv:2412.17769*, 2024.

608

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

611

612 Georgios Kopanas and George Drettakis. Improving NeRF Quality by Progressive Camera Placement  
 613 for Free-Viewpoint Navigation. In Michael Guthe and Thorsten Gorsch (eds.), *Vision, Modeling,*  
 614 *and Visualization*. The Eurographics Association, 2023. ISBN 978-3-03868-232-5. doi: 10.2312/  
 615 vnv.20231222.

616

617 Vincent Leroy, Yohann Cabon, and Jérôme Revaud. Grounding image matching in 3d with mast3r.  
 618 In *European Conference on Computer Vision*, pp. 71–91. Springer, 2024.

619

620 Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and  
 621 Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *IEEE Conference*  
 622 *on Computer Vision and Pattern Recognition (CVPR)*, 2023.

623

624 Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching  
 625 at light speed. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp.  
 626 17627–17638, 2023.

627

628 Yu-Tao Liu, Li Wang, Jie Yang, Weikai Chen, Xiaoxu Meng, Bo Yang, and Lin Gao. Neudf: Leaning  
 629 neural unsigned distance fields with volume rendering. In *Computer Vision and Pattern Recognition*  
 630 (*CVPR*), 2023.

631

632 William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction  
 633 algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pp. 347–353. 1998.

634

635 Linjie Lyu, Ayush Tewari, Marc Habermann, Shunsuke Saito, Michael Zollhöfer, Thomas Leimkühler,  
 636 and Christian Theobalt. Manifold sampling for differentiable uncertainty in radiance fields. In  
 637 *SIGGRAPH Asia 2024 Conference Papers*, pp. 1–11, 2024.

638

639 Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and  
 640 Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications*  
 641 *of the ACM*, 65(1):99–106, 2021.

642

643 Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics  
 644 primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):  
 645 1–15, 2022.

646

647 Linfei Pan, Dániel Baráth, Marc Pollefeys, and Johannes L Schönberger. Global structure-from-  
 648 motion revisited. In *European Conference on Computer Vision*, pp. 58–77. Springer, 2024.

649

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

652

653 Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Super glue:  
 654 Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference*  
 655 *on computer vision and pattern recognition*, pp. 4938–4947, 2020.

648 Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of*  
 649 *the IEEE conference on computer vision and pattern recognition*, pp. 4104–4113, 2016.  
 650

651 Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view  
 652 selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*,  
 653 2016.

654 Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast conver-  
 655 gence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF conference on computer*  
 656 *vision and pattern recognition*, pp. 5459–5469, 2022.  
 657

658 Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local  
 659 feature matching with transformers. In *Proceedings of the IEEE/CVF conference on computer*  
 660 *vision and pattern recognition*, pp. 8922–8931, 2021.  
 661

662 Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus:  
 663 Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *NeurIPS*,  
 664 2021.

665 Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu.  
 666 Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of*  
 667 *the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.  
 668

669 Rundi Wu, Ben Mildenhall, Philipp Henzler, Keunhong Park, Ruiqi Gao, Daniel Watson, Pratul P  
 670 Srinivasan, Dor Verbin, Jonathan T Barron, Ben Poole, et al. Reconfusion: 3d reconstruction with  
 671 diffusion priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern*  
 672 *recognition*, pp. 21551–21561, 2024.

673 Yuhang Xie, Yixi Cai, Yinqiang Zhang, Lei Yang, and Jia Pan. Gauss-mi: Gaussian splatting shannon  
 674 mutual information for active 3d reconstruction. *arXiv preprint arXiv:2504.21067*, 2025.  
 675

676 Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. Mvsnet: Depth inference for unstructured  
 677 multi-view stereo. In *Proceedings of the European conference on computer vision (ECCV)*, pp.  
 678 767–783, 2018.  
 679

680 Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient adaptive surface  
 681 reconstruction in unbounded scenes. *ACM Transactions on Graphics*, 2024.  
 682

683

## A APPENDIX

### A.1 EXPERIMENTAL DETAIL

688 For novel view synthesis, we use Mip-NeRF 360 dataset as the benchmark. For every scene, 1/8  
 689 of the views are split as novel views. The remaining views form the candidate set. Each scene has  
 690 three initial views as the known set. We collect the known set as described in ReconFusion (Wu  
 691 et al., 2024). Like Lyu et al. (2024), we set the active selection to gather 10 and 20 views for 3DGS  
 692 optimization.

693 For surface reconstruction, we use the Tanks&Temples dataset as the benchmark. We follow the  
 694 practice of GOF (Yu et al., 2024) and sample three scenes from the dataset’s training set for evaluation.  
 695 For each scene, 1/8 of the views are split as novel views. We assign three views in each scene as the  
 696 initial known set. The number of view collections is set to 20.

697 Creat3r is data and computational efficient. The efficiency stems from two aspects. The first one is  
 698 active selection efficiency. Since the selection does not include 3DGS optimization, it takes less time  
 699 to find the next best view. Average selection iteration costs of various methods are listed in Table 5.  
 700 Creat3r takes less than half time needed in other counterparts. The duration is measured with a single  
 701 NVIDIA V100 GPU. The second attribute is the optimization efficiency. Due to the robust geometry,  
 Creat3r only uses half of the optimization time to converge.

702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
Table 5: Average selection iteration duration.

Method	Time(sec)
FisherRF (Jiang et al., 2024)	24.124
Kopanas & Drettakis (2023)	24.965
Lyu et al. (2024)	39.470
Creat3r	10.075

713  
714  
715  
716  
717  
718  
719  
720  
A.2 MORE QUANTITATIVE RESULTS721  
722  
723  
724  
725  
726  
The complete evaluation of novel view synthesis is listed in Table 6. For each baseline method, we  
evaluated with three different initialization strategies. For all the methods, initialization with Creat3r-  
MASt3R-matching consistently performs best among all the initializations, and Creat3r-LightGlue  
initialization performs better than COLMAP-induced subsampling initialization. We observe that  
MASt3R produces larger amounts of 2D correspondences than LightGlue, as it estimates more 3D  
points and is beneficial for active view selection. Creat3r  $\ddagger$  and Creat3r  $\dagger$  progressively estimate more  
and more points during the selection, leading to improvement by a large margin compared to other  
competitors.727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
Table 6: Complete novel view synthesis evaluation on Mip-NeRF 360 dataset. The left and right  
columns show evaluations of 3DGS optimized with 10 and 20 selected views. (\*) denotes methods  
initialized with COLMAP-induced subsampling points. ( $\ddagger$ ) indicates methods initialized with Creat3r-  
LightGlue and ( $\dagger$ ) indicates methods initialized with Creat3r-MASt3R. **Best** and **second-best**  
results are highlighted.

Method	10 cameras			20 cameras		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
FPS*	12.529	0.261	0.613	14.918	0.389	0.528
FPS $\ddagger$	12.687	0.318	0.591	14.409	0.413	0.539
FPS $\dagger$	13.560	0.362	0.555	14.940	0.439	0.514
FisherRF (Jiang et al., 2024)*	12.625	0.264	0.608	15.434	0.390	0.515
FisherRF (Jiang et al., 2024) $\ddagger$	13.238	0.342	0.582	15.254	0.432	0.523
FisherRF (Jiang et al., 2024) $\dagger$	14.196	0.392	0.546	16.028	0.474	0.491
Kopanas & Drettakis (2023)*	13.022	0.284	0.596	15.658	0.407	0.506
Kopanas & Drettakis (2023) $\ddagger$	13.039	0.332	0.585	15.771	0.458	0.509
Kopanas & Drettakis (2023) $\dagger$	13.677	0.383	0.546	15.727	0.470	0.499
Lyu et al. (2024)*	12.561	0.264	0.612	15.446	0.401	0.518
Lyu et al. (2024) $\ddagger$	13.308	0.343	0.578	15.708	0.454	0.511
Lyu et al. (2024) $\dagger$	14.282	0.377	0.547	16.264	0.503	0.480
Creat3r $\ddagger$	<b>16.040</b>	<b>0.449</b>	<b>0.536</b>	<b>19.637</b>	<b>0.567</b>	<b>0.443</b>
Creat3r $\dagger$	<b>17.809</b>	<b>0.511</b>	<b>0.523</b>	<b>20.678</b>	<b>0.601</b>	<b>0.397</b>

744  
745  
746  
A.3 MORE QUALITATIVE RESULTS747  
748  
749  
750  
751  
Figure 5 presents more instances of novel view rendering with Creat3r and competing counterparts.  
When the selection is suboptimal, the 3D model cannot correctly render the novel views due to less  
exploration or a lack of finer detail. The renderings present artifacts or holes. Creat3r demonstrates  
renderings closest to the ground truth. Other methods have different kinds of artifacts.752  
753  
A.4 VISUALIZATION OF ABLATION STUDY754  
755  
We visualize the qualitative impact of these design choices in Figure 6. Direct reliance on raw MASt3R  
predictions introduces geometric scale inconsistencies, yielding noisy and blurred renderings. As  
observed in Figure 6(b), relying solely on the confidence map biases the selection towards local detail

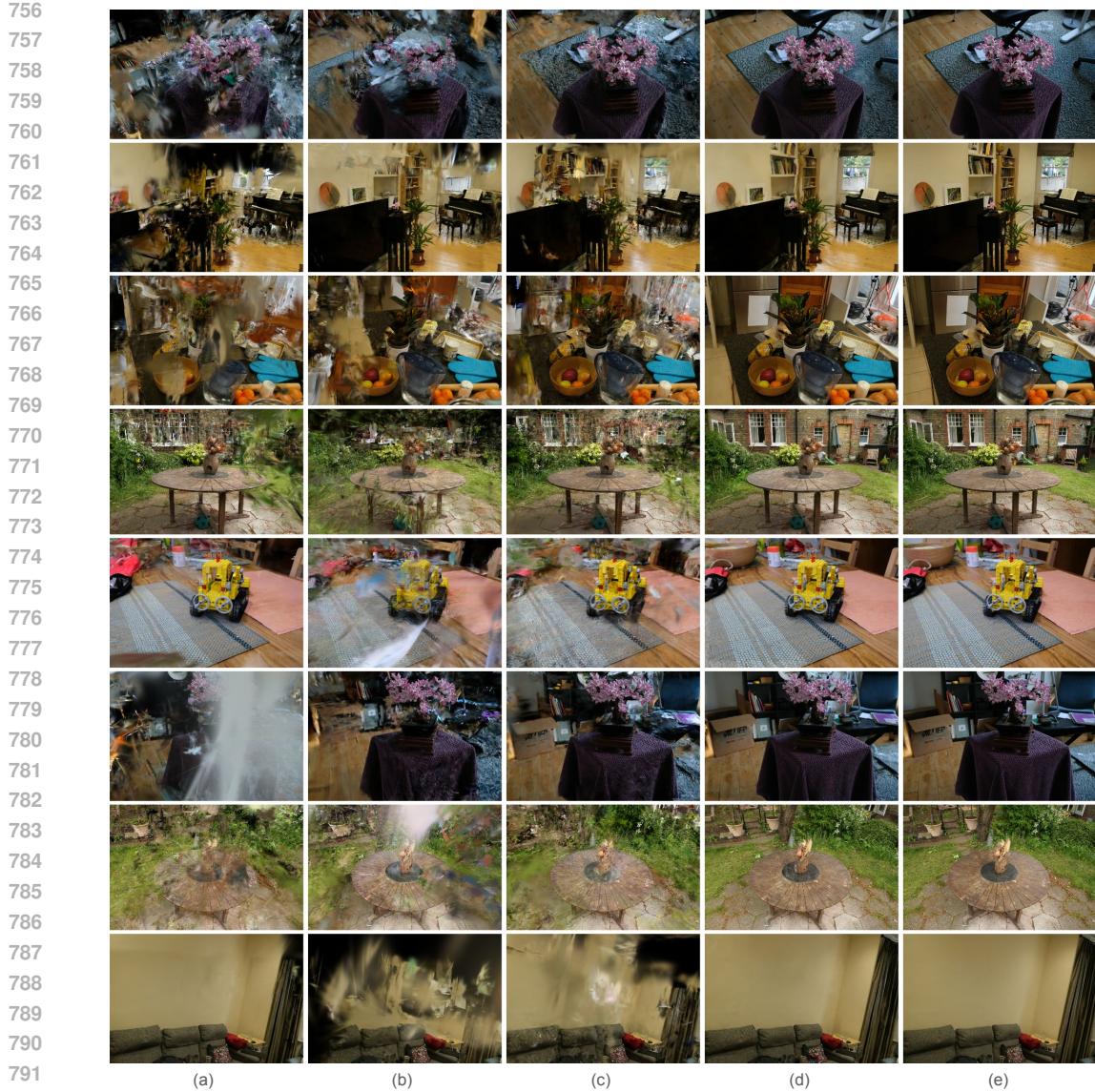


Figure 5: More qualitative comparisons of active view selection on the Mip-NeRF 360 dataset for 20 selected views. The results demonstrate novel view renderings produced by competing methods: (a) FisherRF (Jiang et al., 2024), (b) Kopanas & Drettakis (2023), (c) Lyu et al. (2024), and (d) our proposed method, Creat3r. Column (e) serves as the ground truth novel view reference. The comparison highlights the superior ability of Creat3r to synthesize high-fidelity views with finer details and fewer artifacts.

refinement; this produces high-fidelity reconstruction in observed regions (e.g., the grass) but leaves the background largely unexplored and degraded. Conversely, utilizing only the exploration map ensures broader coverage of both foreground and background but fails to resolve high-frequency details, resulting in noticeable blurring. Finally, Creat3r synergizes both exploration and confidence objectives, achieving globally consistent and highly detailed reconstructions.

810  
 811  
 812  
 813  
 814  
 815  
 816  
 817  
 818  
 819  
 820  
 821  
 822  
 823  
 824  
 825  
 826  
 827  
 828  
 829  
 830  
 831  
 832  
 833  
 834  
 835  
 836  
 837  
 838  
 839  
 840  
 841



842 **Figure 6: Qualitative evaluation of ablation components on the ‘garden’ scene. The rendered novel**  
 843 **views illustrate the impact of distinct design choices: (a) substituting robust point estimation with raw**  
 844 **MASt3R predictions, (b) relying exclusively on confidence for selection, (c) relying exclusively on**  
 845 **exploration for selection, and (d) the full Creat3r framework. The ground truth is provided in (e).**  
 846 **This comparison highlights the specific contribution of each component and the superior reconstruction**  
 847 **fidelity achieved by our holistic approach.**

848  
 849  
 850  
 851  
 852  
 853  
 854  
 855  
 856  
 857  
 858  
 859  
 860  
 861  
 862  
 863