SA-ResGS: Self-Augmented Residual 3D Gaussian Splatting for Next Best View Selection

Kim Jun-Seong[1*] Tae-Hyun Oh[2]
Eduardo Pérez-Pellitero[3] Youngkyoon Jang[4*†]

[1]POSTECH [2]KAIST [3]Huawei Noah's Ark Lab [4]Rivian
junseong.kim@postech.ac.kr
https://saresgs.github.io/

Abstract. We propose Self-Augmented Residual 3D Gaussian Splatting, a novel framework for stabilizing uncertainty quantification and enhancing uncertainty-aware supervision in Next-Best-View selection for active scene reconstruction. To efficiently estimate scene coverage, SA-ResGS generates geometry-consistent Self-Augmented point clouds (SA-Points) via triangulation between observed training views and rasterized extrapolated views. To address the lack of learning signals in underrepresented regions within sparse, wide-baseline settings, we introduce the first skip-connection-inspired residual learning strategy tailored for 3DGS. This mechanism amplifies gradient flow to weakly contributing, high-uncertainty Gaussians. Our contributions are threefold: (1) a physically grounded, diversified view selection strategy; (2) an uncertainty-aware residual supervision scheme that improves gradient flow and learning stability; and (3) implicitly debiased uncertainty quantification resulting from constrained view selection and residual supervision. Experiments on NeRF Synthetic, Mip-NeRF 360, and challenging extended benchmark from Deep Blending and Tanks and Temples demonstrate that SA-ResGS consistently outperforms state-of-the-art competing methods in both reconstruction quality and view selection robustness.

Keywords: Active Learning · 3D Gaussian Splatting · Next-best-view

1 Introduction

Recent advances in neural rendering, particularly Neural Radiance Fields (NeRFs) [38] and 3D Gaussian Splatting (3DGS) [25], have significantly advanced photorealistic scene reconstruction [9, 31, 40, 63], enabling high-fidelity, real-time applications across diverse environments [23, 26, 27]. Beyond static scene capture, these methods have also spurred broader interest in tackling complex challenges, such as active view selection [5, 60] and uncertainty quantification for next-best-view (NBV) selection [22]. Although pre-captured, dense-view training

*This work was conducted while authors were with Huawei Noah's Ark Lab in London
†denotes the corresponding author

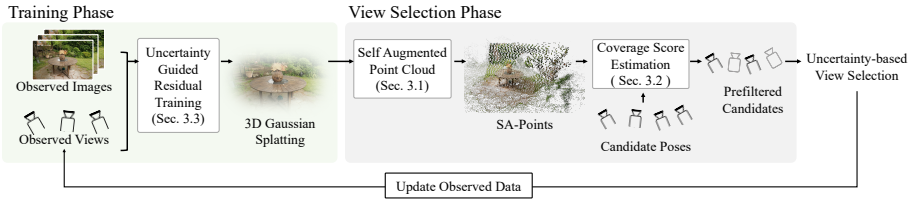

Fig. 1: Overview of SA-ResGS. The framework alternates between view selection and training. At each NBV step, Self-Augmented Points (SA-Points) are generated via triangulation between a training view and its extrapolated render, enabling surface-aware coverage estimation (Sec. 3.1). Candidate views are first physically filtered using hash-encoded feature dissimilarity, then ranked by uncertainty quantification scores for final selection, encouraging exploration of under-observed regions (Sec. 3.2). During training, residual supervision (Sec. 3.3) combines full and uncertainty-intensified renders to reinforce gradients toward weakly contributing Gaussians, improving training stability and reconstruction quality under sparse views.

methods can achieve impressive reconstruction quality, in-situ (active) reconstruction—where views are selected and added progressively—remains challenging due to artifacts caused by shape-radiance ambiguity, further exacerbated by limited training views and the dynamics of the view-addition strategy. Despite the difficulty of reliable uncertainty estimation in this setting, post-hoc methods—such as Laplacian-approximation-based, model-agnostic approaches [12, 22]—remain promising, as they provide uncertainty signals without changing the rendering pipeline. However, we observe that the following challenges have been overlooked:

- **Disregarded physical constraints:** Computational uncertainty is often misaligned with the physical plausibility of reconstructed geometry.
- **Underutilized supervision:** Existing methods rarely convert uncertainty into learning signals, leaving weakly contributing Gaussians undersupervised.
- **Performance dependency:** Reliability of uncertainty remains coupled with training dynamics, especially in early stages with incomplete scene coverage.

In response to these challenges, we propose SA-ResGS, a Self-Augmented Residual 3D Gaussian Splatting framework that stabilizes uncertainty quantification and enhances uncertainty-aware supervision for next-best-view selection in progressive scene reconstruction, as shown in Fig. 1. SA-ResGS strategically decouples view selection from heavy reliance on uncertainty estimates that are sensitive to internal learning dynamics, thereby promoting more robust, geometry-aware surface coverage. Concretely, we first prefilter candidate views using geometric dissimilarity and then apply uncertainty-based scoring within this subset, effectively implementing a physically grounded, uncertainty-informed selection strategy. To support this process, we construct SA-Points by reconstructing 3D point maps from a training view and its rasterized extrapolated views at each view-selection step, after training on a fixed number of initial views. We encode these SA-Points with a hash-based scene representation to efficiently measure

similarity between candidate and previously selected views. We then select the most dissimilar candidate views to improve coverage of unseen regions.

While the physically grounded and uncertainty-informed selection strategy enhances overall scene coverage, it inadvertently reduces multi-view overlap, since more dissimilar views are less likely to observe shared regions, thereby weakening multi-view geometric constraints. To counterbalance this effect without sacrificing the benefits of diverse view selection, we introduce a residual supervision mechanism that provides additional learning signals targeted at under-optimized Gaussians. These Gaussians, which typically correspond to underrepresented regions in sparse views [20], are otherwise often overlooked due to their minimal contribution to the 3DGS rendering process. Specifically, SA-ResGS additionally rasterizes color images using a targeted subset that combines a small fraction of the most uncertain Gaussians with a majority of the originally visible ones. Inspired by Dropout [42, 50] and Hard Negative Mining [19, 62], this strategy acts similarly to ResNet skip connections [17]; it amplifies supervision for under-optimized Gaussians that otherwise receive weak gradients, while remaining fully compatible with conventional 3DGS pipelines.

The main contributions of SA-ResGS are threefold:

- **Physically grounded view selection:** We propose a geometry-aware strategy using SA-Points to guide next-best-view selection, enforcing physical plausibility and promoting more balanced, coverage-oriented exploration.
- **Residual learning for 3DGS:** We introduce the first residual supervision framework specifically designed for 3DGS, addressing the vanishing gradient problem by reinforcing weakly supervised Gaussians and improving both optimization stability and reconstruction quality.
- **Unbiased uncertainty quantification:** By jointly improving the view distribution and supervising under-optimized Gaussians, SA-ResGS mitigates geometric sparsity and density bias, leading to fairer and more reliable uncertainty estimates throughout training.

2 Related Work

Next-best-view selection. NBV selection originated from the robotics community as a strategy to efficiently guide in-situ scene capture, where the goal is to incrementally select viewpoints that maximally reduce reconstruction ambiguity [7, 8, 47]. Classical NBV methods primarily relied on geometric heuristics, selecting views based on geometric coverage [10, 14, 15], viewpoint entropy [54], or visibility [2, 4, 11, 52]. While subsequent learning-based approaches attempted to model scene-specific view policies via reinforcement learning [6, 55], they often struggled with cross-scene generalization. More recent efforts explore information theoretic formulations, such as FisherRF [22], offer a principled formulation for uncertainty-based NBV in neural fields. In our work, we build on this line [12, 16, 22, 59] by integrating physically grounded geometry priors to stabilize early-stage view planning, particularly when minimal visual input is available.

Uncertainty estimation for 3DGS. Quantification of uncertainty plays a pivotal role in active 3D reconstruction, particularly for guiding view selection. In the context of 3DGS, earlier attempts to estimate uncertainty primarily relied on ensemble-based estimates [53] or variational inference [32, 37, 48, 49]. However, these approaches require redundant multiple inferences or modifications to the 3DGS parameters, making them incompatible with the standard training and rendering pipelines of 3DGS. To avoid these limitations, recent research has shifted toward post-hoc uncertainty estimation [12, 13, 16, 22, 56, 59], which extracts uncertainty signals, after training is done, without altering the underlying model architecture. Due to its architectural agnostic nature, post-hoc methods are increasingly explored across diverse applications [21, 33, 51].

Representative post-hoc methods, such as FisherRF [22] and BayesRays [12], adapted classical Laplacian approximations on 3DGS setting, pioneering Bayesian approaches to 3DGS uncertainty estimation. In subsequent work, ActiveViewSelector [56] has introduced image-level rendering quality metrics, and PRIMU [13] suggests 3D unprojection of 2D spatial errors to localize regions of high uncertainty. Despite their promise, these existing post-hoc approaches often overlook the unique training dynamics inherent to 3DGS. Specifically, these methods remain strongly coupled to the density of underlying Gaussians, leading to biased uncertainty estimates in early training stages when geometry is sparse or unevenly distributed and often misinterpreting under-observed regions as confident. To mitigate this limitation, we introduce residual learning, assisted by physically grounded view selection, enabling more loosely coupled uncertainty estimation during early-stage view selection while emphasizing supervision focused on high-uncertainty Gaussians.

Residual supervision in 3DGS. While accurate uncertainty estimation helps localize regions having deficient supervision, it alone is not sufficient to counterbalance under-optimized Gaussians in the 3DGS pipeline. Existing 3DGS methods mainly rely on direct photometric losses [25] or external depth priors [35, 61], which often fail to sufficiently supervise Gaussians with low opacity or minimal rendering contributions. Recent studies, such as pixelSplat [3], PAPR [64], and PAPR-in-Motion [44], explicitly discuss the vanishing gradient issue and propose solutions including differentiable parameterization of Gaussians, proximity attention-based differentiable rendering, adaptive updates, and activation tuning.

Despite the various strategies mitigating the vanishing gradient problem, prior approaches lack an explicit mechanism to correct weakly supervised Gaussians, leaving the problem largely unresolved due to insufficient gradient signals. Although dropout-based approaches [42] help increase gradient diversity, they operate stochastically and do not target supervision toward the most uncertain or least-updated Gaussians. Our method addresses these limitations by introducing the residual supervision strategy for 3DGS. Residual learning, as popularized by ResNet [17], has proven effective in mitigating vanishing gradients and improving training stability through skip connections and additive refinement, yet it remains underexplored in the context of 3D Gaussian Splatting. We replicate skip-connections among 3D Gaussians along the ray, applying uncertainty-guided ren-

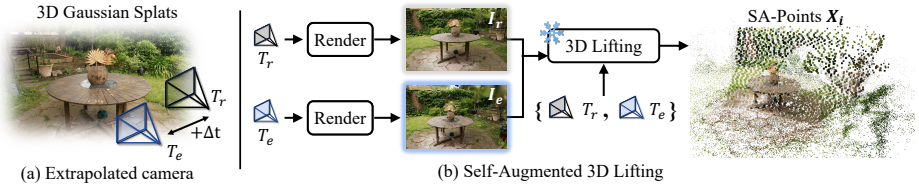

Fig. 2: SA-Points Generation process. (a) The virtual camera pose T_e is extrapolated from the reference camera pose T_r by adding translation Δt. (b) The reference image I_r and extrapolated render I_e are then jointly processed by the 3D lifting module to reconstruct SA-Points X_i. More details in Suppl. Sec. A.

dering to intentionally amplify gradients for under-supervised Gaussians—without altering the underlying rasterization process.

3 Method

The proposed SA-ResGS framework is illustrated in Fig. 1. Built on the next-best-view selection method, FisherRF [22] as a baseline, SA-ResGS extends it with SA-Points around two core ideas: (1) physically grounded view selection with reduced reliance on uncertainty estimates, and (2) residual supervision that explicitly strengthens weakly contributing Gaussians, which minimally affect rasterized pixels and therefore receive insufficient gradients. Implementation details and hyperparameter settings are provided in Suppl. Sec. A.

3.1 Generation of Self-Augmented Points

The physically grounded view selection is enabled by a physically aligned surface representation, constructed using SA-Points derived from a single training view. The overall SA-Points generation pipeline is visualized in Fig. 2. Given a reference image I_r with camera pose $\mathbf{T}_r = [\mathbf{R}_r \mid \mathbf{t}_r]$, we render an extrapolated image I_e from a perturbed pose $\mathbf{T}_e = [\mathbf{R}_r \mid \mathbf{t}_r + \Delta\mathbf{t}]$ using 3DGS. Dense correspondences $\{(\mathbf{p}_r^i, \mathbf{p}_e^i)\}$ between I_r and I_e are predicted using the pretrained MASt3R model [34], which is robust to moderate viewpoint changes and produces contextually meaningful matches even in the presence of minor geometric distortions. SA-Point $\mathbf{X}^i$ is triangulated from a 2D correspondences using the projection matrices $\mathbf{P}_r$, $\mathbf{P}_e$, derived by $\mathbf{T}_r$, $\mathbf{T}_e$, with intrinsic from COLMAP [45, 46].

However, as triangulation is performed repeatedly during training—while the model is still fitting to a sparse and incomplete geometry—rasterized extrapolated images may occasionally contain rendering noise due to inaccurately placed Gaussians. To ensure reliable geometry while fully leveraging the generalization capability of MASt3R, we apply reprojection error-based filtering. The reprojection error is defined as :

$$\varepsilon^i = \tfrac{1}{2}\left(\left\|\mathbf{p}_r^i - \pi(\mathbf{P}_r\mathbf{X}^i)\right\|_2 + \left\|\mathbf{p}_e^i - \pi(\mathbf{P}_e\mathbf{X}^i)\right\|_2\right), \tag{1}$$

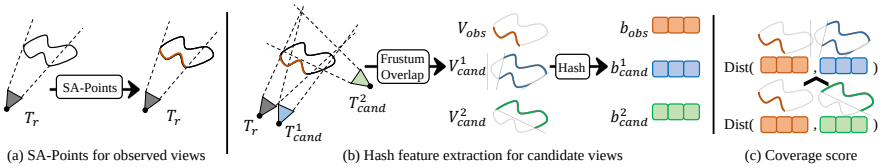

Fig. 3: Physically grounded candidate view selection via surface coverage. (a) SA-Points from training views define observed voxels $\mathcal{V}_{\text{obs}}$. (b) Each candidate view generates a binary hash-encoded feature $\mathbf{b}$, via frustum-based visibility estimation. (c) Normalized Hamming distance between hash-encoded features quantifies coverage dissimilarity, enabling efficient selection of geometrically complementary views without rendered images or uncertainty scores.

where $\|\cdot\|_2$ is the ℓ_2 norm in pixel space, and SA-Points with $\varepsilon^i < \tau$ are retained. This filtering step discards geometrically inconsistent points while preserving accurate SA-Points from dense, context-aware matches, even when the extrapolated image is noisier than the original training view. Compared to prior methods such as CoMapGS [20] or MP-SfM [43], our triangulation pipeline produces scale-consistent, surface-aware geometry from a single image by leveraging extrapolated viewpoints rather than requiring multi-view input or monocular depth estimates. Also, note that, while we use MASt3R as our default 3D lifting module, our framework is modular and compatible to other 3D foundational reconstruction methods.[1]

3.2 Physically Grounded View Selection Algorithm

We present our physically grounded view selection algorithm for next-best-view (NBV) selection, illustrated in Fig. 3. As discussed in Sec. 2, NBV selection in 3D Gaussian Splatting (3DGS) is particularly challenging due to the tight coupling between quality of uncertainty estimation and the quality of reconstructed geometry—both are highly sensitive to the sparsity and distribution of Gaussian splats. Under sparse-view settings, where reconstruction starts from as few as four images and new views are incrementally added every 100 training iterations, uncertainty-based NBV strategies often become unreliable. This occurs because uncertainty signals are inherently biased or unstable when the geometry is incomplete or under-constrained. To address this, we introduce a surface-aware guidance mechanism based on SA-Points, which allows view selection to operate independently of the computed uncertainty quantification. By decoupling view selection from 3DGS internal training dynamics, our method provides more stable and physically meaningful candidate views during the early reconstruction, even before the model accumulates sufficient confidence to produce reliable uncertainty.

We begin by discretizing the 3D scene into a voxel grid $\mathcal{V} = \{v_k\}_{k=1}^{K}$, where each voxel represents a unit volume. The bounding volume of $\mathcal{V}$ is defined by the

[1] Refer to Secs. 4.1 and 4.3 for evaluation with an alternative, Depth-Anything-v3 [36].

sparse point cloud obtained via structure-from-motion (SfM). A voxel $v_k \in \mathcal{V}$ is marked as observed if it intersects any SA-Point $\mathbf{X}^i$ (Sec. 3.1), forming the subset $\mathcal{V}_{\text{obs}} \subset \mathcal{V}$. To account for potential localization errors and promote coverage continuity, we dilate each occupied voxel using a 3D kernel $\mathcal{K}_r$ of radius r:

$$\tilde{\mathcal{V}}_{\text{obs}} = \bigcup\nolimits_{v_k \in \mathcal{V}_{\text{obs}}} \mathcal{K}_r(v_k), \tag{2}$$

where $\tilde{\mathcal{V}}_{\text{obs}}$ denotes the dilated observed region for the current training views.

For each candidate view j, from the index set of all candidate views $\mathcal{C} = \{1, ..., M\}$, we compute a frustum $\mathcal{F}_j \subset \mathcal{V}$, defined by the camera intrinsics (field of view) and near/far planes estimated from the SfM point distribution. A voxel is considered potentially visible from view j if its center lies within the frustum:

$$\mathcal{V}_{\text{cand}}^{(j)} = \{v_k \in \mathcal{V} \mid v_k \in \mathcal{F}_j\}. \tag{3}$$

To estimate geometric dissimilarity between current coverage and a candidate view (Fig. 3), we compute the normalized Hamming distance:

$$d_j = \frac{1}{K} \left\| \mathbf{b}_{\text{obs}} \oplus \mathbf{b}_{\text{cand}}^{(j)} \right\|_1, \tag{4}$$

where $\oplus$ denotes the element-wise XOR operation between binary vectors, and $\|\cdot\|_1$ is the ℓ_1 norm (i.e., the number of differing entries). Here, $\mathbf{b}$ is a binary occupancy vector obtained by mapping voxel coordinates through a fixed random hashing function, following the spatial hashing strategy of Instant-NGP [28, 39]. The resulting value $d_j \in [0, 1]$ measures the proportion of voxels with inconsistent occupancy status between the currently observed volume and the candidate view; it quantifies the volume of non-overlapping occupancy between the observed voxel set and a candidate view's frustum voxels, *i.e.*, how many newly covered voxels a candidate adds. Candidate views are then ranked in descending order of their normalized Hamming distances d_j, and the top $N\%$ (e.g., $N = 20$) are retained as the physically filtered candidate set $\mathcal{C}' = \text{Top}_{\text{N\%}}(\mathcal{C}, d)$.

We apply uncertainty quantification[2] only within $\mathcal{C}'$, and finalize view selection via finer-level scoring. This two-stage pipeline follows a coarse-to-fine strategy: SA-Points provide an explicit estimate of which regions are already observed, and we first select views that maximally expose the remaining unobserved regions to form a stable candidate subset and refines the choice through uncertainty-aware reasoning. Restricting uncertainty evaluation to $\mathcal{C}'$ reduces computation by avoiding uncertainty scoring for every candidate view. At the same time, as $\mathcal{C}'$ is pre-filtered to favor views that expose unobserved regions, the final selection avoids redundant viewpoints and achieves more balanced scene coverage.

[2]We adopt uncertainty estimation formulation from FisherRF [22], the uncertainty and expected-infromation-gain (EIG) formulation is summarized in Suppl. Sec. A.

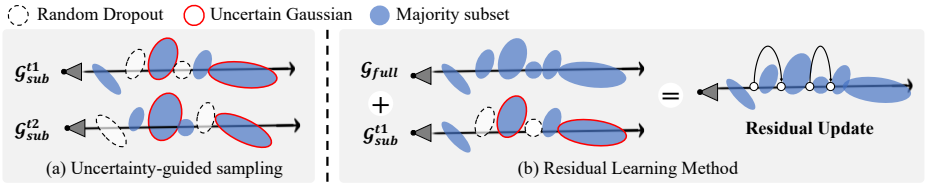

Fig. 4: Residual supervision in 3DGS. Overview of the proposed residual supervision strategy. (a) At each iteration (t_1, t_2), $\mathcal{G}_{\text{sub}}$ combines random and top-uncertain Gaussians; (b) residual supervision in 3DGS mimics ResNet-style skip connections.

3.3 Uncertainty-Guided Residual Learning in 3DGS

We propose the first residual learning framework for 3DGS that emulates skip connections to address vanishing gradients in weakly contributing Gaussians, as shown in Fig. 4. These Gaussians often receive insufficient supervision due to their limited impact on rasterized pixels, particularly in sparse or ambiguous regions. While ResNet [17] mitigates similar issues through skip connections, such mechanisms are infeasible in 3DGS given the dynamic, view-dependent nature of Gaussian properties. Instead, we introduce a rasterizer-agnostic strategy that improves gradient flow by generating auxiliary renders emphasizing high-uncertainty Gaussians. These renders are supervised with input RGB images, forming the basis of the residual supervision scheme described below.

Residual supervision. To reinforce under-supervised Gaussians, we introduce a residual supervision scheme that uses two rendered images from the same training view: one rendered with the full Gaussian set $\mathcal{G}_{\text{full}}$, and another rendered with a guided subset $\mathcal{G}_{\text{sub}}$, as shown in Fig. 4(a). We define this subset as:

$$\mathcal{G}_{\text{sub}} = \mathcal{G}_{\text{rand}} \cup \mathcal{G}_{\text{uncertain}}, \tag{5}$$

where $\mathcal{G}_{\text{rand}}$ is a random sample comprising $\alpha\%$ of $\mathcal{G}$ (e.g., α=90), and $\mathcal{G}_{\text{uncertain}}$ contains the top-β most uncertain Gaussians (e.g., β=10). To estimate uncertainty, we analyze two per-Gaussian attributes: opacity and scale. Gaussians with low opacity contribute minimally to alpha blending during rasterization, while those with large scales blur across pixels and tend to dominate ambiguous or low-texture regions. This ranking identifies Gaussians that are both visually suppressed and spatially diffuse, making them key targets for correction.

We compute two rendered images: I_{full} from the full Gaussian set $\mathcal{G}_{\text{full}}$, and I_{sub} from the uncertainty-intensified subset $\mathcal{G}_{\text{sub}}$. Each is supervised independently against the input image I_{gt} using ℓ_1 and SSIM losses:

$$\mathcal{L} = \sum_{i \in \{\text{full},\text{sub}\}} \lambda_i \left[\mathcal{L}_{\text{rgb}}(I_i, I_{\text{gt}}) + \mathcal{L}_{\text{ssim}}(I_i, I_{\text{gt}}) \right], \tag{6}$$

where $\lambda_{\text{full}} + \lambda_{\text{sub}} = 1$, and we simply set both to 0.5. We denote the losses for I_{full} and I_{sub} as the full loss ($\mathcal{L}_{\text{full}}$) and subset loss ($\mathcal{L}_{\text{sub}}$), respectively.

This uncertainty-intensified rasterization strategy is conceptually inspired by Dropout [42, 50] and Hard Negative Mining [19, 62]. Random sampling of $\mathcal{G}_{\text{rand}}$ provides stochastic diversity, allowing weakly contributing Gaussians to receive supervision when dominant ones are excluded. Meanwhile, deterministic inclusion of $\mathcal{G}_{\text{uncertain}}$ ensures consistent gradient flow to persistently under-optimized Gaussians. This dual mechanism reinforces learning in uncertain or ambiguous regions without modifying the rasterization process, while complementing full-image supervision to maintain global photometric fidelity.

To see why the subset branch amplifies supervision for these Gaussians, we examine the gradient it induces. For a pixel ray, the full branch renders $C_{\text{full}} = \sum_i T_i \alpha_i c_i$ with $T_i = \prod_{j<i}(1-\alpha_j)$. In the residual branch, we exclude the Gaussians not selected in $\mathcal{G}_{\text{sub}}$; for mathematical clarity, this is equivalent to masking them with zero opacity in that branch. Considering the effect of a single masked Gaussian k in isolation, the modified transmittance becomes $\tilde{T}_i = T_i/(1-\alpha_k)$ for every later Gaussian $i > k$. which can be generalized by accumulating one such factor per masked Gaussian ahead of i. As $\tilde{T}_i$ multiplies the subset gradient with respect to *any* parameter θ_i, the amplification carries over to the whole parameter set; we illustrate with the color c_i, as the compositing map is linear ($\partial C/\partial c_i = T_i\alpha_i$). For $i > k$:

$$\frac{\partial \mathcal{L}}{\partial c_i} = \lambda_{\text{full}} \frac{\partial \mathcal{L}_{\text{full}}}{\partial C_{\text{full}}} T_i \alpha_i + \lambda_{\text{sub}} \frac{\partial \mathcal{L}_{\text{sub}}}{\partial C_{\text{sub}}} \frac{T_i \alpha_i}{1-\alpha_k}. \tag{7}$$

The amplification factor $1/(1-\alpha_k)$ grows as the Gaussian k becomes more opaque. Intuitively, dropping a high-opacity occluder redirects gradient toward previously occluded, weakly contributing Gaussians, the source of our residual supervision. By supervising both full and uncertainty-intensified images, we promote stronger gradient flow toward uncertain or low-opacity Gaussians without compromising photometric quality. This strategy mirrors the effect of residual skip connections in ResNet [17] (Fig. 4(b)), enabling more stable convergence and reducing overfitting in sparse or wide-baseline training settings. It is particularly effective during early next-best-view selection, when reconstruction is sensitive to both sparsely initialized regions and supervision bias from limited views.

4 Experiments

Dataset. We evaluate our approach on two benchmark datasets: NeRF-Synthetic [38] and Mip-NeRF 360 [1]. Although these datasets span scenes from synthetic object-scale setups to real-world outdoor environments with full 360-degree coverage, their uniform, curated camera trajectories pose only a limited challenge for active view selection, as even simple heuristics (e.g., furthest-distance selection) perform reliably under balanced coverage [60]. To address this limitation, we curate an extended benchmark with five diverse scenes from Deep Blending [18] and Tanks and Temples [29], which introduce unbalanced view distributions and varied scene scales that better reflect practical conditions. All experiments use

Table 1: Quantitative results for the Active View Selection. We compare our model with (1) Rule-based models (Random, ACP [30]), (2) 2D-based models [5] (MUSIQ, CrossScore), and 3D-based model(FisherRF [22]). Results are averaged over 9 scenes from the Mip-NeRF 360, 7 scenes from NeRF-synthetic dataset and 5 additional scenes from the Deep Blending and Tanks and Temples. We conduct four trials for each scene and report the average. For further statistics please refer to the Suppl. Sec. I.

Category	Method	NeRF Synthetic			Mip-NeRF 360			Extended Benchmark		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Rule-based	Random	24.847	0.893	0.117	19.969	0.584	0.456	19.262	0.699	0.375
	ACP	22.718	0.855	0.138	20.325	0.596	<u>0.449</u>	19.950	0.718	<u>0.361</u>
2D-based	MUSIQ	25.237	0.889	0.119	19.850	0.575	0.466	18.699	0.688	0.391
	CrossScore	23.746	0.868	0.130	21.076	<u>0.612</u>	**0.448**	19.942	0.727	**0.356**
3D-based	FisherRF	25.190	0.892	0.116	20.642	0.595	0.450	19.654	0.711	0.370
	Ours (MASt3R)	**26.580**	**0.907**	**0.110**	**21.410**	**0.613**	0.451	**20.401**	<u>0.732</u>	<u>0.361</u>
	Ours (DA-v3)	<u>26.437</u>	<u>0.905</u>	<u>0.111</u>	<u>21.361</u>	**0.613**	<u>0.449</u>	<u>20.348</u>	**0.733**	0.365

images at their original resolutions; for further dataset curation details, please refer to Suppl. Sec. A.

Competing Methods. We compare our method quantitatively and qualitatively against several active 3DGS methods that operate solely on RGB images: FisherRF (baseline) [22], ACP [30], and random view selection. We also include 2D-based view selection methods from the Active View Selector framework [56], which incorporates two image quality assessment (IQA) models, MUSIQ [24] and CrossScore [57], to evaluate perceptual quality. Both models are re-implemented following the authors' official instructions and publicly available code. We additionally compare against earlier methods such as BayesRays [12] and ActiveNeRF [41] adapted to 3DGS, results are reported in Suppl. Sec. H.

4.1 Active View Selection

Experiment setup. Following the protocol in [22], we adopt the prescribed initial view configurations and view selection schedule. Specifically, each experiment starts with four uniformly distributed views, then adds one view every 100 epochs until reaching 20 training views (for fewer training views, see Suppl. Sec. G). We apply the same active selection strategy across all datasets. For consistency, each model is initialized with the same random seed, trained for 20,000 iterations, and shares the same COLMAP SfM initialization[3]. All other settings remain unchanged across experiments, except for the view selection algorithm. Because each process includes stochasticity, we repeat all experiments four times and report average scores (for details, see Suppl. Sec. I). **Novel View Synthesis Results.** Quantitative and qualitative results on NeRF-Synthetic, Mip-NeRF

[3]SA-Points are used only for view selection and are not included in the COLMAP SfM or 3DGS initialization.

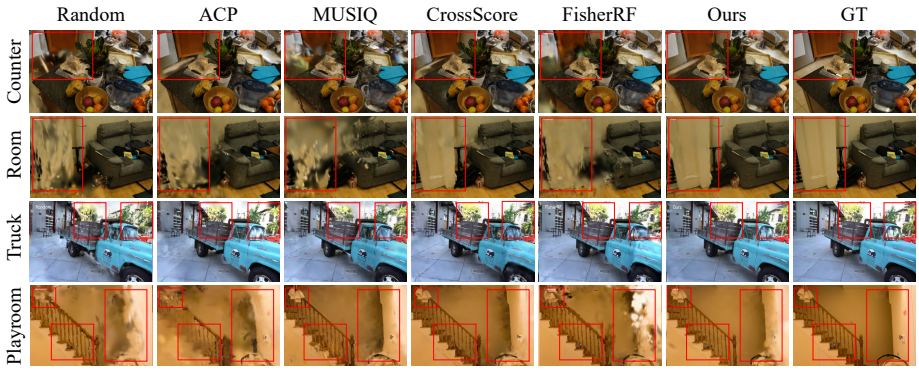

Fig. 5: Qualitative Comparison of Active View Selection. Reconstruction from 20 selected views per scene. Our method shows improved completeness and fewer artifacts compared to counterparts models. Multi-view visualization and 360° rendering is provided in Suppl. Sec. C and Suppl. video.

360, and the extended benchmark (Deep Blending and Tanks and Temples) are summarized in Table 1 and Fig. 5. Competing methods show limited 3D reconstruction performance, especially in sparsely observed regions, primarily due to biased view selection and overfitting induced by vanishing gradients. This leads to incomplete reconstructions, characterized by floating artifacts and missing geometry (*i.e.* holes and missing objects). In contrast, our method achieves the highest PSNR and SSIM on NeRF-Synthetic and Mip-NeRF 360 while maintaining comparable LPIPS performance. The smaller LPIPS gains likely stem from smoother color reconstruction in uncertain regions, consistent with trends reported in [20]. As also reflected in Table 1, variants using DA-v3 as the 3D lifting module show consistent gains over all competing methods, indicating our improvements are not tied to specific backbone.

Consistent with the trends above, results in Table 1 on the extended datasets further validate the generalizability of our method. In particular, Deep Blending results show robustness under diverse, real-world-like camera distributions. Likewise, numerical improvements on Tanks and Temples indicate stronger scene coverage in large outdoor settings, as exemplified by the Truck scene in Fig. 5. Even under challenging, realistic view configurations, our model remains superior in reconstruction quality and scene coverage while preserving comparable high-frequency details. For multi-view evaluation, we provide visualizations and 360° rendering results in Suppl. Sec. C and the Suppl. video.

Camera Distribution Results. To further support these quantitative and qualitative trends, we analyze the camera view distributions selected by different methods. As illustrated in Fig. 6, competing methods often select clustered or redundant viewpoints, leading to uneven scene coverage. This behavior is especially evident in FisherRF, whose selections concentrate around high-response regions due to its tight coupling with 3DGS training dynamics. In contrast, our

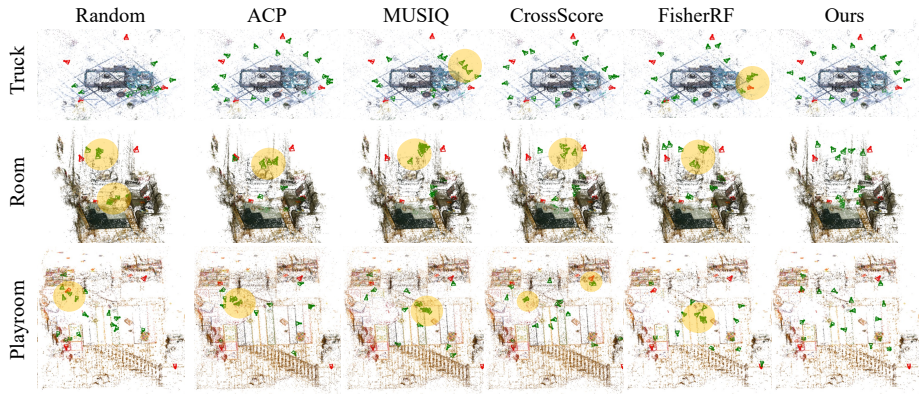

Fig. 6: Selected Camera View Distribution. Visualization of camera poses selected by each method on *Truck, Playroom* scenes from Extended dataset, and *Room* scene from Mip-NeRF 360 dataset. Red frustums indicate the initial views, while green frustums denote views added during active selection. Our method produces a more uniformly distributed set of viewpoints, while counterpart models exhibit cluttering (highlighted by yellow circles), due to its reliance on uncertainty signals entangled with 3DGS training dynamics. For more visualization please refer to Suppl. Sec. B.

method selects more spatially dispersed and geometrically diverse viewpoints. Together, these observations show that our physically grounded view selection and uncertainty-aware learning strategy improves reconstruction fidelity and coverage consistency. For more camera-distribution results please refer to Suppl. Sec. B.

4.2 Comparison on Uncertainty Estimation

Experiment setup. We evaluate how effectively our method improves uncertainty estimation quality. Specifically, we examine whether residual loss and self-augmented pre-filtering improve alignment between depth errors and predicted uncertainties under controlled conditions. To measure, we adopt the Area Under the Sparsification Error (AUSE) metric, to evaluate uncertainty calibration adopted in [12,22,48]. AUSE evaluates how accurately the high-magnitude regions of the uncertainty heatmap correspond to the largest actual depth errors.

Following CF-NeRF [48], we use depth maps from NerfingMVS [58], optimized at test time with stereo depth from COLMAP. Experiments are conducted on all nine Mip-NeRF 360 scenes under an identical view-selection schedule and evaluated on all test views.

Results. Starting from the baseline (FisherRF†), adding residual learning on a fixed baseline view sequence (‡+ResGS) reduces AUSE from 0.327 to 0.323. Adding prefiltering with SA-Points (‡+SA-ResGS) further reduces AUSE to 0.297. These reductions indicate that both residual learning and our prefiltering improve uncertainty calibration. Fig. 7 highlights representative miscalibration cases, including overconfident predictions in high-error regions and conservative

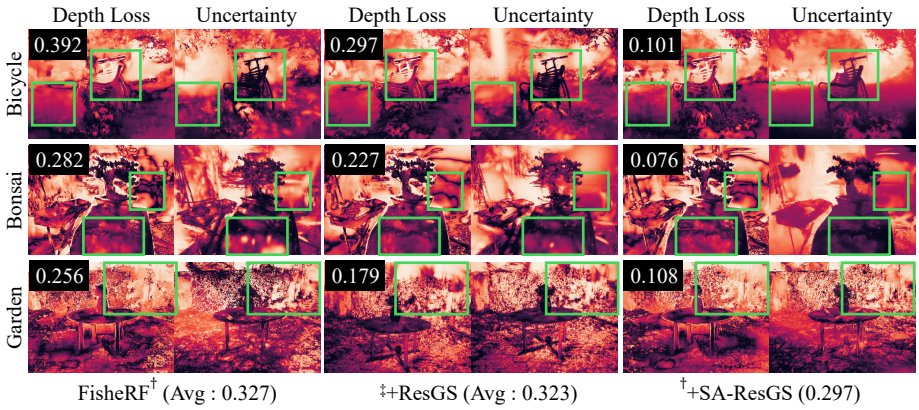

Fig. 7: Uncertainty Comparison. We visualize depth loss with their corresponding uncertainty with AUSE metric (lower is better). Highlighted green boxes show that †+SA-ResGS yields better alignment between large depth errors and high uncertainty. Please note that ‡ denotes fixed-order view selection, replicating the original selection sequence from FisherRF†, whereas † indicates dynamically updated view selection.

uncertainty in low-error areas. Compared with the baseline, our method reduces these ranking mismatches and aligns predicted uncertainty more closely with ground-truth depth error.

We attribute these gains to the complementary roles of residual learning and our physically grounded prefiltering strategy. Residual learning stabilizes confidence in low-error regions while preserving adaptive refinement in uncertain areas via skip connections. Meanwhile, prefiltering promotes more balanced spatial coverage during training, reducing localized miscalibration and improving global uncertainty–error consistency. Consequently, our model achieves better structural and quantitative uncertainty calibration, improving accuracy in uncertainty-driven tasks such as active mapping.

4.3 Ablation Studies

In the following section, we present ablation studies that analyze (1) the effect of individual components and (2) the effect of the full loss. Additional ablations are provided in Supplementary material. Specifically, we analyze our physically grounded view selection algorithm in Suppl. Sec. E and the uncertainty-guided residual learning in Suppl. Sec. F. In Suppl. Sec. G, we further study robustness to correspondence noise, the hash-encoding size, and selected view-count.

Effect of individual components. The ablation study in Table 2 highlights the individual and synergistic contributions of our proposed SA-ResGS framework. While residual learning (ResGS) alone improves reconstruction quality under a fixed baseline view sequence (‡+ResGS), its gains are attenuated when combined

Table 2: Ablation study on Mip-NeRF 360 and Extended dataset. ‡ denotes fixed-order view selection, whereas † indicates dynamically updated view selection. We conduct four trials for each scene and report the average. Visual comparison is included in Suppl. Sec. D, and further statistics in Suppl. Sec. I.

Methods	3D Lifting	Our proposed methods		Mip-NeRF 360			Extended dataset		
		Sec. 3.2	Sec. 3.3	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
FisherRF†	-	-	-	20.642	0.595	0.450	19.490	0.708	0.373
‡+ResGS	-	-	✓	21.045	0.604	0.453	19.602	0.709	0.378
†+ResGS	-	-	✓	20.732	0.594	0.461	19.791	0.715	0.374
†+SA-HashGS	MASt3R	✓	-	21.093	0.609	**0.441**	19.702	0.718	0.363
†+SA-ResGS	MASt3R	✓	✓	**21.410**	**0.613**	0.451	**20.401**	0.732	0.361
†+SA-HashGS	DA-v3	✓	-	21.076	0.609	**0.441**	20.025	0.726	**0.357**
†+SA-ResGS	DA-v3	✓	✓	21.361	**0.613**	0.449	20.348	**0.733**	0.365

with purely dynamic selection (†+ResGS). This indicates that training improvements alone are insufficient, especially under large uncertainty quantification errors. As shown in Sec. 4.2, our training module effectively aligns predicted uncertainty with actual errors, yet purely uncertainty-driven view selection remains vulnerable to biases from internal learning dynamics.

Our physically grounded prefiltering module (†+SA-HashGS) stabilizes view selection by restricting the candidate set for Fisher uncertainty. The full configuration (†+SA-ResGS) achieves the best performance, demonstrating a synergy in which SA-HashGS provides a reliable view sequence that enables residual supervision to further improve reconstruction in sparse or ambiguous regions. The same trend appears when replacing the 3D lifting module with DA-v3: prefiltering consistently stabilizes view selection, and adding ResGS yields further gains.

Effect of the full loss. To assess the contribution of the Full Loss, we compare ResGS with a w/o Full Loss variant (Fig. 8), where Gaussians are updated only through the guided subset term ($\mathcal{L}_{sub}$) and the default w/ Full Loss setting ($\mathcal{L}_{full}$+$\mathcal{L}_{sub}$). The w/o Full Loss model tends to oversmooth low-confidence regions, causing a loss of high-frequency detail. Without $\mathcal{L}_{full}$ reinforcing these regions, the subset-only variant cannot preserve or reactivate Gaussians that require continued refinement. In contrast, w/ Full Loss prevents under-updated Gaussians from collapsing and preserves both global structure and fine details. For an ablation on uncertainty-guided sampling, see Suppl. Sec. D, Fig. S10.

4.4 Computation Efficiency Analysis

A key challenge in active view selection is computational cost, as FisherRF computes per-Gaussian Fisher information via backpropagation across all candidate views, creating a bottleneck in large-scale datasets. To evaluate this overhead, we conduct a runtime analysis on the *Bonsai* scene using a mid-range GPU (38 TFLOPS fp32), as summarized in Table 3. SA-ResGS replaces exhaustive Fisher

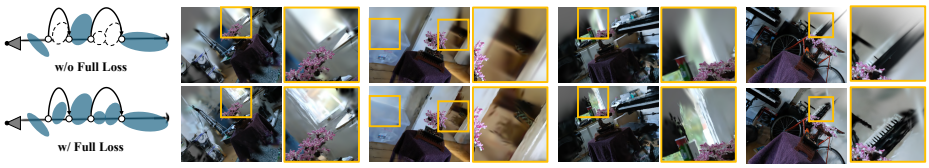

Fig. 8: Comparison of novel-view renderings with/without the Full Loss. We use a pre-selected view order from FisherRF† to ignore view-selection effects. The w/o Full Loss ($\mathcal{L}_{sub}$ only) variant exhibits smoothing artifacts, whereas w/ Full Loss ($\mathcal{L}_{full} + \mathcal{L}_{sub}$) preserves scene structure and high-frequency details.

Table 3: Runtime comparison. Breakdown of temporal training costs in the active view selection on *Bonsai* scene. SA-ResGS introduces additional prefiltering steps but reduces total runtime by 40%, with modest increases in GPU memory usage.

Method	**Active View Selection** [s]					**Raster.** [s/iter]	**End-to-end**	**GPU** [K]
	MASt3R	Triang.	Prefilter	Fisher	**Total**			
FisherRF	—	—	—	28.00	28.00	0.005	32 m 59 s	8.0
Ours	0.40	1.49	5.00	5.60	12.50	0.027	19 m 45 s	10.5

evaluation with a four-step process: dense correspondence prediction (MASt3R), SA-Points triangulation (Triang.), physically grounded prefiltering (Prefilter), and Fisher computation on filtered views (Fisher). Despite these additional steps, view selection is 55% faster, with only a modest increase in per-iteration cost. End-to-end runtime is reduced by 40%. GPU memory usage rises slightly but remains within standard limits, supporting scalability and practicality.

5 Conclusion

This paper presents SA-ResGS, a framework that stabilizes uncertainty quantification and enhances uncertainty-aware supervision for next-best-view selection in active scene reconstruction. We introduce Self-Augmented Points, reconstructed from a training view and a rasterized extrapolated view. These points enable physically grounded view selection and help mitigate bias in uncertainty estimates. Furthermore, we propose the first residual learning strategy tailored to 3D Gaussian Splatting, with an emulated skip connection, enabling effective supervision for both uncertain image regions and weakly contributing Gaussian splats. This leads to improved photometric reconstruction in novel view synthesis. Extensive experiments for NBV selection and novel view synthesis demonstrate the effectiveness of SA-ResGS across a range of realistic scenes.

Acknowledgements. This work was partially supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant (No. RS-2026-25518317, Development of AI memory mechanism that reflects human cognitive principles), the National Research Foundation of Korea (NRF) grant (No. RS-2024-00451947; No. RS-2024-00453301), and the InnoCORE program of the Ministry of Science and ICT(26-InnoCORE-01) funded by the Korea government (MSIT).

References

1. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In: IEEE Conf. Comput. Vis. Pattern Recog. (2022)
2. Bircher, A., Kamel, M., Alexis, K., Oleynikova, H., Siegwart, R.: Receding horizon "next-best-view" planner for 3d exploration. In: IEEE Int. Conf. on Robotics and Automation (2016)
3. Charatan, D., Li, S.L., Tagliasacchi, A., Sitzmann, V.: pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In: IEEE Conf. Comput. Vis. Pattern Recog. (2024)
4. Chen, L., Zhan, H., Chen, K., Xu, X., Yan, Q., Cai, C., Xu, Y.: Activegamer: Active gaussian mapping through efficient rendering. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
5. Chen, X., Li, Q., Wang, T., Xue, T., Pang, J.: Gennbv: Generalizable next-best-view policy for active 3d reconstruction. In: IEEE Conf. Comput. Vis. Pattern Recog. (2024)
6. Christopher, C., William, J., B., Manfred, H.: Learning the next best view for 3d point clouds via topological features. In: IEEE Int. Conf. on Robotics and Automation (2021)
7. Connolly, C.I.: The determination of next best views. In: IEEE Int. Conf. on Robotics and Automation (1985)
8. Delmerico, J., Isler, S., Sabzevari, R., Scaramuzza, D.: A comparison of volumetric information gain metrics for active 3d object reconstruction. Autonomous Robots **42**(2) (2018)
9. Dong-Yeon, S., Jun-Seong, K., Byung-Ki, K., Oh, T.H.: Hdr-nsff: High dynamic range neural scene flow fields. In: ICLR (2026)
10. Dunn, E., Frahm, J.M.: Next best view planning for active model improvement. In: Brit. Mach. Vis. Conf. (2009)
11. Feng, Z., Zhan, H., Chen, Z., Yan, Q., Xu, X., Cai, C., Li, B., Zhu, Q., Xu, Y.: Naruto: Neural active reconstruction from uncertain target observations. In: IEEE Conf. Comput. Vis. Pattern Recog. (2024)
12. Goli, L., Reading, C., Sellán, S., Jacobson, A., Tagliasacchi, A.: Bayes' Rays: Uncertainty quantification for neural radiance fields. In: IEEE Conf. Comput. Vis. Pattern Recog. (2024)
13. Gottwald, T., Heinert, E., Stehr, P., Galappaththige, C.J., Rottmann, M.: Primu: Uncertainty estimation for novel views in gaussian splatting from primitive-based representations of error and coverage. In: IEEE Conf. Comput. Vis. Pattern Recog. (2026)
14. Guédon, A., Monasse, P., Lepetit, V.: Scone: surface coverage optimization in unknown environments by volumetric integration. In: Adv. Neural Inform. Process. Syst. (2022)
15. Guédon, A., Monnier, T., Monasse, P., Lepetit, V.: Macarons: Mapping and coverage anticipation with rgb online self-supervision. In: IEEE Conf. Comput. Vis. Pattern Recog. (2023)
16. Hanson, A., Tu, A., Singla, V., Jayawardhana, M., Zwicker, M., Goldstein, T.: Pup 3d-gs: Principled uncertainty pruning for 3d gaussian splatting. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conf. Comput. Vis. Pattern Recog. (2016)

18. Hedman, P., Philip, J., Price, T., Frahm, J.M., Drettakis, G., Brostow, G.: Deep blending for free-viewpoint image-based rendering. ACM Trans. Graph. (2018)
19. Jang, Y., Gunes, H., Patras, I.: Registration-free face-ssd: Single shot analysis of smiles, facial attributes, and affect in the wild. Computer Vision and Image Understanding **182**, 17–29 (2019)
20. Jang, Y., Pérez-Pellitero, E.: Comapgs: Covisibility map-based gaussian splatting for sparse novel view synthesis. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 26779–26788 (June 2025)
21. Jiang, W., Lei, B., Ashton, K., Daniilidis, K.: Multimodal llm guided exploration and active mapping using fisher information. In: Int. Conf. Comput. Vis. (2025)
22. Jiang, W., Lei, B., Daniilidis, K.: Fisherrf: Active view selection and uncertainty quantification for radiance fields using fisher informations. In: Eur. Conf. Comput. Vis. (2024)
23. Jun-Seong, K., GeonU, K., Yu-Ji, K., Wang, Y.C.F., Choe, J., Oh, T.H.: Dr. splat: Directly referring 3d gaussian splatting via direct language embedding registration. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
24. Ke, J., Wang, Q., Wang, Y., Milanfar, P., Yang, F.: Musiq: Multi-scale image quality transformer. In: Int. Conf. Comput. Vis. (2021)
25. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. (2023)
26. Kim, G., Youwang, K., Hyoseok, L., Oh, T.H.: Fpgs: Feed-forward semantic-aware photorealistic style transfer of large-scale gaussian splatting (2026)
27. Kim, G., Youwang, K., Oh, T.H.: FPRF: Feed-forward photorealistic style transfer of large-scale 3D neural radiance fields. In: AAAI (2024)
28. Kim, J.S., Kim, M., Kim, G., Oh, T.H., Kim, J.H.: Factorized multi-resolution hashgrid for efficient neural radiance fields: Execution on edge-devices. In: IEEE Robotics and Automation Letters. IEEE (2024)
29. Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V.: Tanks and temples: Benchmarking large-scale scene reconstruction. ACM Trans. Graph. **36**(4) (2017)
30. Kopanas, G., Drettakis, G.: Improving NeRF Quality by Progressive Camera Placement for Free-Viewpoint Navigation. In: Proceedings of the Vision Modeling and Visualization (2023)
31. Kulhanek, J., Peng, S., Kukelova, Z., Pollefeys, M., Sattler, T.: WildGaussians: 3D gaussian splatting in the wild. In: Adv. Neural Inform. Process. Syst. (2024)
32. Lee, S., Kang, K., Ha, S., Yu, H.: Bayesian nerf: Quantifying uncertainty with volume density for neural implicit fields. IEEE Robotics Autom. Lett. (2025)
33. Lei, B., Jiang, W., Daniilidis, K.: Activegrasp: Information-guided active grasping with calibrated energy-based model (2025), arXiv preprint arXiv:2511.12795
34. Leroy, V., Cabon, Y., Revaud, J.: Grounding image matching in 3d with mast3r. In: Eur. Conf. Comput. Vis. (2024)
35. Li, J., Zhang, J., Bai, X., Zheng, J., Ning, X., Zhou, J., Gu, L.: Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In: IEEE Conf. Comput. Vis. Pattern Recog. (2024)
36. Lin, H., Chen, S., Liew, J.H., Chen, D.Y., Li, Z., Shi, G., Feng, J., Kang, B.: Depth anything 3: Recovering the visual space from any views (2025), arXiv preprint arXiv:2511.10647
37. Lyu, L., Tewari, A., Habermann, M., Saito, S., Zollhöfer, M., Leimküehler, T., Theobalt, C.: Manifold sampling for differentiable uncertainty in radiance fields. In: ACM Trans. Graph. (2024)

38. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: Eur. Conf. Comput. Vis. (2020)
39. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. (2022)
40. Niedermayr, S., Stumpfegger, J., Westermann, R.: Compressed 3d gaussian splatting for accelerated novel view synthesis. In: IEEE Conf. Comput. Vis. Pattern Recog. (2024)
41. Pan, X., Lai, Z., Song, S., Huang, G.: Activenerf: Learning where to see with uncertainty estimation. In: Eur. Conf. Comput. Vis. (2022)
42. Park, H., Ryu, G., Kim, W.: Dropgaussian: Structural regularization for sparse-view gaussian splatting. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
43. Pataki, Z., Sarlin, P.E., Schönberger, J.L., Pollefeys, M.: MP-SfM: Monocular Surface Priors for Robust Structure-from-Motion. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
44. Peng, S., Zhang, Y., Li, K.: Papr in motion: Seamless point-level 3d scene interpolation. In: IEEE Conf. Comput. Vis. Pattern Recog. (2024)
45. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: IEEE Conf. Comput. Vis. Pattern Recog. (2016)
46. Schönberger, J.L., Zheng, E., Frahm, J.M., Pollefeys, M.: Pixelwise view selection for unstructured multi-view stereo. In: Eur. Conf. Comput. Vis. (2016)
47. Scott, W.R., Roth, G., Rivest, J.F.: View planning for automated three-dimensional object reconstruction and inspection. ACM Computing Surveys (CSUR) **35**(1) (2003)
48. Shen, J., Agudo, A., Moreno-Noguer, F., Ruiz, A.: Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In: Eur. Conf. Comput. Vis. (2022)
49. Shen, J., Ruiz, A., Agudo, A., Moreno-Noguer, F.: Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In: Int. Conf. on 3D Vis. (2021)
50. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research (2014)
51. Strong, M., Lei, B., Swann, A., Jiang, W., Daniilidis, K., III, M.K.: Next best sense: Guiding vision and touch with fisherrf for 3d gaussian splatting. In: IEEE Int. Conf. on Robotics and Automation (2025)
52. Sun, Y., Huang, Q., Hsiao, D.Y., Guan, L., Hua, G.: Learning view selection for 3d scenes. In: IEEE Conf. Comput. Vis. Pattern Recog. (June 2021)
53. Sünderhauf, N., Abou-Chakra, J., Miller, D.: Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. In: IEEE Int. Conf. on Robotics and Automation (2023)
54. Vázquez, P.P., Feixas, M., Sbert, M., Heidrich, W.: Viewpoint selection using viewpoint entropy. In: Proceedings of the Vision Modeling and Visualization (2001)
55. Wang, T., Xi, W., Cheng, Y., Han, H., Yang, Y.: Rl-nbv: A deep reinforcement learning based next-best-view method for unknown object reconstruction. Pattern Recognition Letters (2024)
56. Wang, Z., Bhalgat, Y., Li, R., Prisacariu, V.A.: Active view selector: Fast and accurate active view selection with cross reference image quality assessment (2025), arXiv preprint arXiv:2506.19844
57. Wang, Z., Bian, W., Prisacariu, V.A.: Crossscore: Towards multi-view image evaluation and scoring. In: Eur. Conf. Comput. Vis. (2024)

58. Wei, Y., Liu, S., Rao, Y., Zhao, W., Lu, J., Zhou, J.: Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo. In: Int. Conf. Comput. Vis. (2021)
59. Wilson, J., Almeida, M., Mahajan, S., Labrie, M., Ghaffari, M., Ghasemalizadeh, O., Sun, M., Kuo, C.H., Sen, A.: Pop-gs: Next best view in 3d-gaussian splatting with p-optimality. In: IEEE Conf. Comput. Vis. Pattern Recog. (2025)
60. Xiao, W., Santa Cruz, R., Ahmedt-Aristizabal, D., Salvado, O., Fookes, C., Lebrat, L.: Nerf director: Revisiting view selection in neural volume rendering. In: IEEE Conf. Comput. Vis. Pattern Recog. (June 2024)
61. Xu, W., Gao, H., Shen, S., Rui Peng, J.J., Wang, R.: Mvpgs: Excavating multi-view priors for gaussian splatting from sparse input views. In: Eur. Conf. Comput. Vis. (2024)
62. Xuan, H., Stylianou, A., Liu, X., Pless, R.: Hard negative examples are hard, but useful. In: Eur. Conf. Comput. Vis. (2020)
63. Yu, Z., Chen, A., Huang, B., Sattler, T., Geiger, A.: Mip-splatting: Alias-free 3d gaussian splatting. In: IEEE Conf. Comput. Vis. Pattern Recog. (June 2024)
64. Zhang, Y., Peng, S., Moazeni, A., Li, K.: Papr: Proximity attention point rendering. In: Adv. Neural Inform. Process. Syst. (2023)