GURecon: Learning Detailed 3D Geometric Uncertainties for Neural Surface Reconstruction

Zesong Yang[1*], Jiale Shi[1*], Ru Zhang[1], Zixiang Ai[1], Boming Zhao[1],
Hujun Bao[1], Luwei Yang[2], Zhaopeng Cui[1†]

[1]State Key Lab of CAD & CG, Zhejiang University, [2]Simon Fraser University
zesongyang0@zju.edu.cn, jlshi@zju.edu.cn, ruzhang@zju.edu.cn, zxai@zju.edu.cn, bmzhao@zju.edu.cn,
bao@cad.zju.edu.cn, luweiy@sfu.ca, zhpcui@gmail.com

Abstract

Neural surface representation has demonstrated remarkable success in the areas of novel view synthesis and 3D reconstruction. However, assessing the geometric quality of 3D reconstructions in the absence of ground truth mesh remains a significant challenge, due to its rendering-based optimization process and entangled learning of appearance and geometry with photometric losses. In this paper, we present a novel framework, *i.e*, GURecon, which establishes a geometric uncertainty field for the neural surface based on geometric consistency. Different from existing methods that rely on rendering-based measurement, GURecon models a continuous 3D uncertainty field for the reconstructed surface, and is learned by an online distillation approach without introducing real geometric information for supervision. Moreover, in order to mitigate the interference of illumination on geometric consistency, a decoupled field is learned and exploited to fine-tune the uncertainty field. Experiments on various datasets demonstrate the superiority of GURecon in modeling 3D geometric uncertainty, as well as its plug-and-play extension to various neural surface representations and improvement on downstream tasks such as incremental reconstruction. The code and supplementary material are available on the project webpage: https://zju3dv.github.io/GURecon/.

1 Introduction

Image-based 3D reconstruction is a long-standing problem in computer vision with a wide range of applications like AR/VR, autonomous driving, digital heritage preservation, etc. Recently, learning-based methods have attracted much attention with the development of neural radiance representations like Neural Radiance Fields (NeRF) (Mildenhall et al. 2021). Unlike traditional methods, NeRF and its variants (*e.g*, NSVF (Liu et al. 2020), NeuS (Wang et al. 2021)) encode scene geometry and appearance with neural networks, which can be optimized by leveraging the differentiable rendering given a set of calibrated images.

Although the neural representations demonstrate remarkable performance in novel view synthesis and surface reconstruction with high levels of detail and photorealism, assessing the reconstruction quality remains challenging. Some
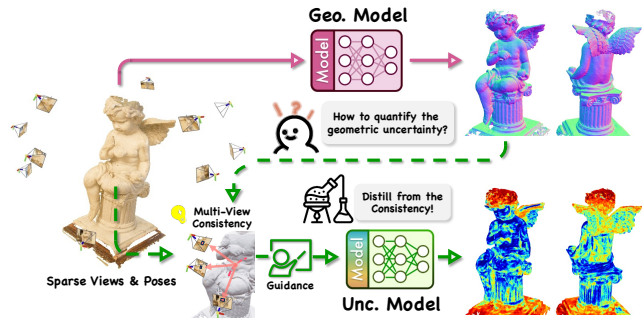

Figure 1: **A brief overview.** By leveraging multi-view consistency as guidance, GURecon learns detailed 3D geometric uncertainties for neural surface reconstruction.

existing work incorporates uncertainty estimation into NeRF models to identify areas with poor rendering quality. NeRF-W (Martin-Brualla et al. 2021) and its following works (Pan et al. 2022; Ran et al. 2023) take the radiance field as Gaussian distributions to model the uncertainty of rendered RGB. Some other works model uncertainty as the entropy of the weight distribution along rays in NeRF models (Zhan et al. 2022; Lee et al. 2022). Besides, the deep learning techniques are also applied to NeRF to quantify the uncertainty via ensemble learning (Sünderhauf, Abou-Chakra, and Miller 2023) or variational inference (Shen et al. 2021, 2022). However, all these methods evaluate the uncertainty of the neural fields in a single-view pixel-wise manner via volumetric rendering, which does not support direct evaluation of 3D geometry accurately, and the uncertainty for the same surface point may vary across different views due to the multi-view inconsistencies in images caused by varying lighting and observation angles, disobeying the view-independent nature of 3D geometric uncertainties.

In this paper, we present a novel framework, *i.e*, GURecon, which is able to learn detailed 3D geometric uncertainty for neural surface reconstruction as shown in Fig. 1. Different from existing methods relying on rendering-based pixel-wise uncertainty measurement, GURecon directly models the 3D uncertainty for surface points and ensures consistency over viewpoints. However, designing such a system is nontrivial. Without ground truth geometric supervision (*e.g*,

---

*Zesong Yang and Jiale Shi contributed equally to this work.
†Corresponding author.

input depth), it is difficult to model the geometric uncertainty just based on photometric error between the rendered and input images. This is because, as with previous methods, the neural radiance field tends to overfit the input images in sparse viewpoint settings, resulting in minor photometric errors but significant geometric errors, just as the ambiguity problem highlighted in NeRF++ (Zhang et al. 2020).

Motivated by the traditional multi-view stereo (Hu and Mordohai 2012; Schönberger et al. 2016) where photometric consistency is widely used to assess the confidence of reconstructed geometry, we employ the multi-view consistency as a cue to quantify the quality of reconstruction. We compute the patch-based warping consistency of surface points projected onto the input images, and utilize it as a pseudo label of geometric accuracy to supervise a continuous geometric uncertainty field based on a novel online distillation approach. We consider the estimated uncertainty derived from such pseudo label as epistemic uncertainty, which reflects the geometric confidence of the model per-scene trained with given images (*i.e*, reconstruction error), and serves as a reference identifying areas where reconstruction is inadequate and unreliable.

Besides, inevitable illumination in real-world scenes poses a challenge to modeling geometric uncertainty based on inconsistent color observations. To handle this problem, we propose to learn additional decoupled fields and further fine-tune the uncertainty field by removing view-dependent factors from each image. Our method can be extended to various neural surface representations. With accurate 3D geometric uncertainty estimation, GURecon can be integrated into tasks like incremental reconstruction to boost the quality of surface reconstruction.

Our main contributions are summarized as follows:

- We present a novel framework, *i.e*, GURecon, to quantify geometric uncertainty for neural surface reconstruction.
- We proposed a new strategy to distill geometric uncertainty based on multi-view consistency, thus decoupling geometric uncertainty with rendering-related uncertainty.
- Additional decoupled fields are learned and exploited to eliminate view-dependent factors for robust estimation.
- Extensive experiments on diverse datasets demonstrate the superior performance of our framework in modeling geometric uncertainty and the potential for application in downstream tasks such as incremental reconstruction.

2 Related Work

Neural Surface Reconstruction. Neural representations have achieved great success in various tasks such as multi-view 3D reconstructions and novel view synthesis. Among them, NeRF (Mildenhall et al. 2021) encodes scenes within an MLP through differentiable volume rendering, enabling high-quality novel view synthesis. SDF-based variants (Wang et al. 2021; Yariv et al. 2021) constrain the scene as an SDF field and achieve smooth surface reconstruction. Subsequent works utilize monocular geometric priors (Yu et al. 2022; Xiao et al. 2024) and geometric consistency to enhance the quality of reconstruction. (Fu et al. 2022; Darmon et al. 2022) utilize the homography warp as a constraint,

while (Ge et al. 2023; Wang et al. 2022) use multi-view consistency to filter the interferences in input data. In this paper, we use a hash-based NeuS (Zhao et al. 2022) as the scene representation and first utilize multi-view consistency as guidance for uncertainty quantification.

Uncertainty Modeling in NeRF. Considering the various interferences such as dynamic objects and limited observations present in input data, integrating uncertainty modeling becomes crucial for achieving robust reconstructions. Uncertainty estimation in NeRF can be divided into epistemic uncertainty and aleatoric uncertainty. The former typically arises from data limitations, while the latter is generally associated with the inherent randomness of data. NeRF-Wild (Martin-Brualla et al. 2021) mitigate the interference of transient objects by modeling rendered colors as Gaussian distributions. Subsequent works build upon it to address the Next Best View (NBV) problem (Pan et al. 2022; Chen et al. 2023). Other approaches tackle uncertainty through sampling techniques to establish a probability model, such as ensemble learning (Sünderhauf, Abou-Chakra, and Miller 2023) or variational inference (Shen et al. 2021, 2022), the former is time and memory-consuming, while the latter involves major network architecture modifications. In contrast to predicting probability, (Lee et al. 2022) computes uncertainty as the entropy of weight distribution along the rays. All these methods utilize probabilistic models to model uncertainty, focusing on network convergence rather than constructing uncertainty from a geometric perspective. Bayes' Rays (Goli et al. 2023) simulates spatially parameterized perturbation of the radiance field and uses a Laplace approximation to produce a volumetric uncertainty field. Another work, FisherRF (Jiang, Lei, and Daniilidis 2023), introduces fisher information for uncertainty modeling. However, they still model uncertainty in a pixel-wise manner based on rendering RGB values and need to measure uncertainty by rendering at a pixel level, not approaching the problem from a 3D geometric perspective. All existing methods are designed for uncertainty estimation in NeRF, considering only the rendering perspective, with no work addressing geometric uncertainty estimation for neural surface representation. In contrast, we introduce GURecon, the first framework that models geometric uncertainty for the neural surface from the perspective of multi-view geometric consistency.

3 Method

In this paper, we introduce a novel framework, *i.e*, GURecon, which enables accurate geometric uncertainty estimation for various neural surface representations without GT geometric information for supervision. As shown in Fig. 2, with given posed images, we learn a render field and an SDF field through differentiable rendering. As the training progresses, with the currently learned geometry field, we first utilize a root-finding method to identify the zero-crossing points intersected with the implicit surface and calculate the multi-view consistency of these points as pseudo supervision to guide the learning of geometric uncertainty (Sec. 3.2). Then we present a novel online distillation method that simultaneously learns a spatially continuous uncertainty field with other fields in a self-supervised manner
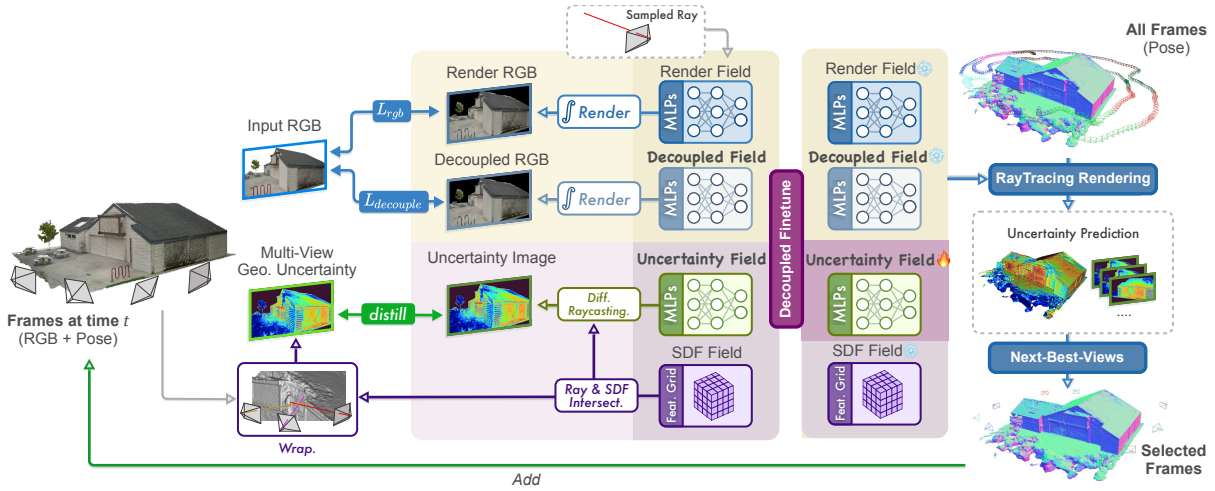
Figure 2: **System Overview.** The proposed GURecon models a geometric uncertainty field supervised by the pseudo labels computed based on the multi-view geometry consistency. To deal with the view-dependent factors, additional decoupled fields are also learned and exploited to fine-tune the uncertainty field. With the predicted uncertainty fields, GURecon can boost the downstream tasks such as incremental reconstruction.

by utilizing the multi-view consistency as pseudo ground truth labels (Sec. 3.3). In order to overcome the interference caused by view-dependent factors in the calculation of multi-view consistency, we propose to simultaneously learn additional decoupled fields and exploit them to fine-tune the geometric uncertainty field (Sec. 3.4).

3.1 Neural Surface Representation

Taking NeuS (Wang et al. 2021) as a representative, we denote each pixel by casting a ray as $\mathbf{p}(t) = \mathbf{o} + t\mathbf{v}$, where $\mathbf{o}$ is the camera origin and $\mathbf{v}$ is the view direction. We define the surface as the zero-level set $\mathcal{S} = \{x \in \mathbb{R}^3 | f_s(x) = 0\}$ using a geometry encoder $f_s(x; \Phi_s)$, which predicts the signed distance field (SDF) value s and a hidden geometric feature h at point x. Additionally, we employ a radiance encoder $f_c(x, \mathbf{v}, \mathbf{n}, h; \Phi_c)$ to predict the color c based on the view direction $\mathbf{v}$, where $\mathbf{n}$ is the normal at point x computed from the gradient of the SDF. The color of each pixel is computed by accumulating colors of sampled points along the ray:

$$\hat{C}(\mathbf{p}) = \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_i, \ T_i = \prod_{j=1}^{i-1}(1 - \alpha_j),$$
$$\alpha_j = \max\left(\frac{\psi_s(s_i) - \psi_s(s_{i+1})}{\psi_s(s_i)}, 0\right), \quad (1)$$

where T_i is accumulated transmittance at $\mathbf{p}(t_i)$, ψ_s is the sigmoid function, and α_i is the opacity of the i-th ray segment. Similar neural surface representations can also be adopted as long as the geometric surface can be computed on the fly.

3.2 Patch-based Multi-view Consistency

Motivated by the traditional MVS works (Stereopsis 2010; Schönberger et al. 2016) that leverage photometric consistency among different views as a geometric constraint, we

exploit it as a cue to guide the learning of geometric uncertainty in the neural surface.

Surface Interaction Retrieval. The primary step is to identify the surface points of the neural representation. Following the existing works (Fu et al. 2022; Oechsle, Peng, and Geiger 2021), root finding is a widely used method to locate the intersection with the neural surface. As our approach is based on SDF representation, and the SDF values of sampling points along the ray are precomputed for volume rendering, we employ linear interpolation to locate the zero-crossing points $\mathcal{T}$ as follows:

$$\mathcal{T} = \left\{ \mathbf{p}(t_i^*) \mid t_i^* = \frac{f_s(\mathbf{p}(\hat{t}_i))\hat{t}_{i+1} - f_s(\mathbf{p}(\hat{t}_{i+1}))\hat{t}_i}{f_s(\mathbf{p}(\hat{t}_i)) - f_s(\mathbf{p}(\hat{t}_{i+1}))} \right\},$$
$$\hat{t}_i = \arg\min_i \{t_i \mid f_s(\mathbf{p}(t_i)) \cdot f_s(\mathbf{p}(t_{i+1})) < 0\}, \quad (2)$$

where $\hat{t}_i$ is the ray segment of the zero-crossing point.

Patch-based Multi-view Photometric Consistency. With the intersected points $\mathcal{T}$ of the neural surface, we acquire the multi-view photometric information by projecting these points onto visible views following (Fu et al. 2022; Darmon et al. 2022). For robustness, we consider the consistency of the pixel patches around the projection of surface points rather than a single pixel. We approximate the small region around the point as a local plane and use the homography warp to compute the patch-based multi-view photometric consistency for computational efficiency. The tangent plane (Stereopsis 2010; Schönberger et al. 2016) at the surface point $\mathbf{p}'$ can be modeled as follows:

$$\mathbf{\Pi} = \{\mathbf{n}', \mathbf{p}' \mid \mathbf{n}'^T \mathbf{p}' + d = 0\}, \text{ where } \mathbf{p}' \in \mathcal{T}, \quad (3)$$

where $\mathbf{n}'$ is the normal computed from the gradient of the SDF values, $i.e$, $\nabla f_s(\mathbf{p}')$, and d is the distance to the origin of the coordinate system. Then, the homography warping matrix $\mathbf{H}$ can be constructed based on the local plane
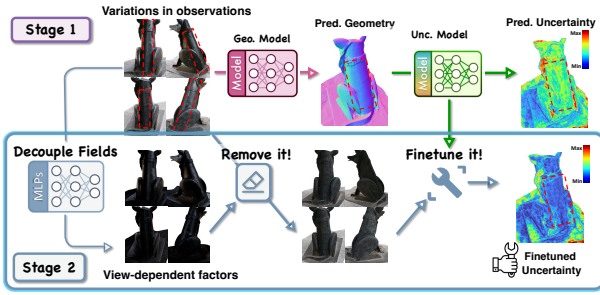
Figure 3: **Uncertainty finetuning with the decoupled fields.** Even if the geometry has been well reconstructed, the uncertainty field still erroneously estimates it with high uncertainty caused by light interference across different views. We employ the decoupled fields to remove the view-dependent factor from the training images.
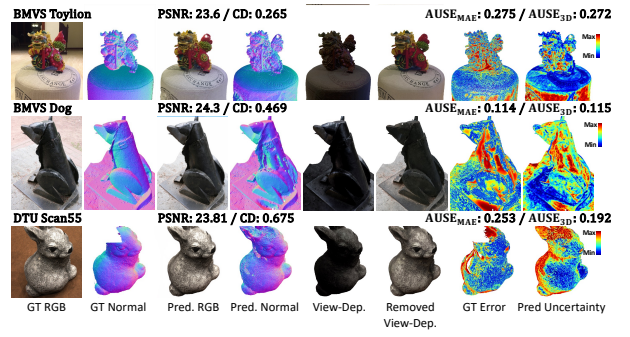

Figure 4: **Visualization of the learned fields.** Our method presents accurate decoupled results for view-dependent factors, and the learned geometric uncertainties are well aligned with the GT geometric error.

and enables the mutual projection of image patches between viewpoints as the following:

$$\mathbf{H}_{rel} = \mathbf{K}_{src} \left(\mathbf{R}_{rel} - \mathbf{t}_{rel} \frac{\mathbf{n}'^T}{d} \right) \mathbf{K}_{ref}^{-1}, \ \mathbf{P}_j = \mathbf{H}_{ij}\mathbf{P}_i, (4)$$

where $\mathbf{K}$ corresponds to the camera's intrinsic matrix, $[\mathbf{R}_{rel}, \mathbf{t}_{rel}]$ corresponds to the relative transformation matrix from the reference view i to the source view j, $\mathbf{P}_i$ and $\mathbf{P}_j$ represent the corresponding patch coordinates of the local plane projected on reference and source view respectively.

Finally, we convert the color images $\mathbf{I}_i$ into gray images $\mathbf{I}'_i$, and utilize the Structural Similarity Index Measure (SSIM) (Campbell et al. 2008) to measure the correlation coefficient $\mathbb{C}$ between pairs of projected patches as:

$$\mathbb{C}_{ij}^k = 1 - SSIM(\mathbf{I}'_i(\mathbf{P}_i^k), \mathbf{I}'_j(\mathbf{P}_j^k)). \quad (5)$$

As the similarity between $\mathbf{I}'_i(\mathbf{P}_i^k)$ and $\mathbf{I}'_j(\mathbf{P}_j^k)$ increases, the score of $\mathbb{C}_{ij}^k$ decreases and the corresponding geometric quality reconstructs better. Considering the potential occlusion and large deviations in projection viewing angles, for robustness we ultimately select the four patch pairs with the lowest computed scores $\mathbb{C}_n^{k*}$ and compute the average score to represent the final geometric consistency $\mathbb{G}_k$ of $\mathbf{p}_k$:

$$\mathbb{G}_k = (\sum_{n=1}^4 \mathbb{C}_n^{k*})/4, \ \text{where} \ \mathbb{C}_n^{k*} \in \underset{i,j}{\arg\min} \left\{\mathbb{C}_{ij}^k\right\}. \quad (6)$$

We utilize the computed consistency as pseudo-ground-truth labels to guide the learning of geometric uncertainty.

3.3 Distillation of Geometric Uncertainty Field

Considering it is impractical and inefficient to perform such consistency calculations for each pixel during inference due to its high computational cost, we propose to learn a geometric uncertainty field distilled from the above geometric consistency, which is conducted simultaneously with the learning process of geometric and radiance fields.

Specifically, considering that geometric uncertainty is a view-independent factor which only related to the position

x of points, we use an uncertainty field $f_u(x; \Phi_u)$ with position input and learn the geometric uncertainty solely for surface points $\mathbf{p}' \in \mathcal{T}$ corresponding to the current SDF field during training process.

As described in Sec 3.2, we firstly use the root-finding method to locate the surface point corresponding to the current iteration at each training step, and then take the multi-view patch-based consistency of the point as a pseudo label to supervise a continuous and accurate uncertainty field using online distillation with the following loss:

$$\mathcal{L}_{distill} = \frac{1}{\mathcal{R}'} \sum_{r \in \mathcal{R}'} |f_u(\mathbf{p}_r') - \mathbb{G}_r|, \quad (7)$$

where $\mathcal{R}'$ corresponds to the set of rays intersected with the surface, and $\mathbf{p}_r'$ is the intersection sampled on ray r.

3.4 Finetuning with Decoupled Fields

Variations in light across different views can lead to inconsistent observations and subsequently impact the computation of multi-view consistency as Fig. 3 shows. For more accurate modeling of the geometric uncertainty, inspired by prior works for radiance decomposition (Verbin et al. 2022; Fan et al. 2023; Tang et al. 2023), we further introduce an additional branch to decouple view-dependent factors as:

$$\bar{C} = C_{vi}(\mathbf{n}', \mathbf{p}', h') + C_{vd}(\mathbf{n}', \mathbf{p}', w_r', h'), \quad (8)$$

where C_{vd} and C_{vi} correspond to view-dependent and view-independent components respectively. As the same with uncertainty field, we only decouple the points on the surface, and w_r' is the reflection of the view direction around the normal $\mathbf{n}'$. We model view-dependent components using the reflection direction rather than the view direction as it allows for better interpolation of factors like specular following (Ge et al. 2023; Fan et al. 2023; Verbin et al. 2022). The decouple fields are trained with surface rendering as follows:

$$\mathcal{L}_{decouple} = \frac{1}{|\mathcal{R}'|} \sum_{r \in \mathcal{R}'} |\bar{C}^r - C_{gt}^r|. \quad (9)$$

Once the fields are decoupled, we eliminate lighting, reflections, and other interferences by subtracting the rendered
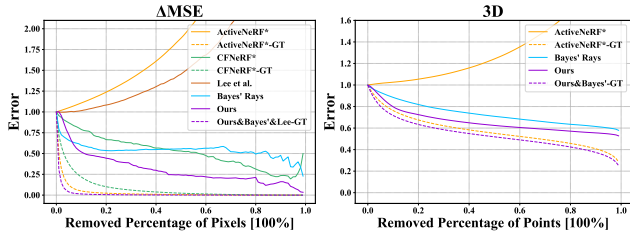
Figure 5: **Sparsification curves of different methods.** The dashed and solid lines correspond to the average error of the remaining pixels filtered using GT-error-based and uncertainty-based criteria, the area between them is AUSE. Bayes' Rays and Lee et al. (Lee et al. 2022) share the same GT curve with ours as post-hoc frameworks.

view-dependent factor $\mathbf{I}_{vd}$ from the true RGB image as shown in Fig. 4, and then use the processed image to recompute the multi-view consistency as the same in Sec. 3.2, thus fine-tuning the uncertainty field. We divided our pipeline into two stages. In the first stage we sample on the ground-truth image to generate pseudo labels and supervise the geometric uncertainty while simultaneously learning the decoupled fields, and in the second stage, our method initially renders the view-dependent component for each training viewpoint and processes the ground-truth image, then freezes other fields and uses the processed image for N_{ft} iterations to finetune the uncertainty field as shown in Fig. 3.

3.5 Loss Function and Implementation Details

Loss Function. Our total loss is defined as the following:

$$\mathcal{L} = \mathcal{L}_{color} + \alpha_1 \mathcal{L}_{reg} + \alpha_2 \mathcal{L}_{mask} + \alpha_3 \mathcal{L}_{decouple} + \alpha_4 \mathcal{L}_{distill}. \quad (10)$$

Following the definition in NeuS (Wang et al. 2021), $\mathcal{L}_{color}$ is the rgb loss between the ground truth pixel colors and the rendered colors, and $\mathcal{L}_{reg}$ is the eikonal loss to regularize the gradients of SDF. Since we only focus on the geometric quality of the target to be reconstructed, we use the mask to filter irrelevant regions and $\mathcal{L}_{mask}$ corresponds to the constraint. We set $\alpha_1 = 0.1$, $\alpha_2 = 1.0$, $\alpha_3 = 0.1$ and $\alpha_4 = 0.1$.

Implementation Details. The GURecon serves as a plug-and-play module applicable to various neural surface representations. Our preference for the fundamental 3D representation leans towards the hash-based variant (Zhao et al. 2022) due to its time efficiency. For each scene, we sample 1024 rays per batch and train for 50k iterations, which takes nearly 30 minutes on an NVIDIA RTX 3090. After completing the training stage, we run an additional 10k iterations to finetune the uncertainty field while keeping other fields frozen. Since the geometry is fixed, we utilize sphere tracing instead of inefficient sampling to locate the intersection points with the neural surface, and the fine-tuning stage takes approximately 5 minutes. Please refer to the supplementary materials for more details.

Methods	ActiveNeRF*		Bayes' Rays		Ours	
Scenes	$AUSE_{3D}(\downarrow)$	$CD(\downarrow)$	$AUSE_{3D}(\downarrow)$	$CD(\downarrow)$	$AUSE_{3D}(\downarrow)$	$CD(\downarrow)$
TNT-Barn (100 images)	1.076	1.079	0.438	0.994	**0.327**	0.994
TNT-Truck (65 images)	0.989	3.082	0.289	2.965	**0.243**	2.965
TNT-Caterpillar (100 images)	1.247	0.808	0.346	0.747	**0.198**	0.747

Table 1: **Uncertainty Quantification for TNT dataset**.

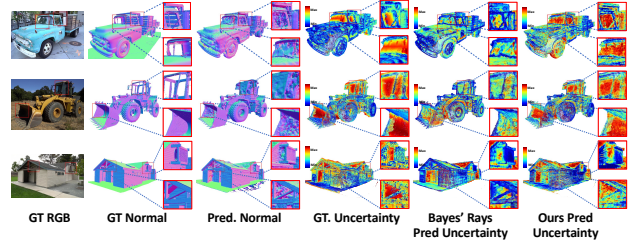GT RGB   GT Normal   Pred. Normal   GT. Uncertainty   Bayes' Rays Pred Uncertainty   Ours Pred Uncertainty

Figure 6: **Geometric uncertainty in TNT dataset.** We present the reconstruction results and the corresponding GT 3D error. GURecon predicts more accurate uncertainty than Bayes' Rays (Goli et al. 2023), especially in areas with texture repetition and reflection.

4 Experiment

In this section, we first assess the efficacy of GURecon in uncertainty quantification. Then we perform ablation studies to validate each component within our framework, demonstrating its versatility across different numbers of training images and various neural surface models. Lastly, we demonstrate our plug-and-play capability of GURecon by applying it to the task of incremental reconstruction and comparing it against other NeRF-based NBV selection methods.

4.1 Uncertainty Quantification

Datasets. We evaluate our method over three widely used benchmark datasets: the DTU dataset (Jensen et al. 2014), the BlendedMVS dataset (Yao et al. 2020), and the Tank and Template (TNT) dataset (Knapitsch et al. 2017). These datasets offer calibrated multi-view images, along with object masks and high-fidelity 3D models serving as the ground truth. The DTU dataset comprises object scans, with each scene containing 49 or 64 views from concentrated perspectives. We selected eight scenes with diverse materials, all exposed to challenging ambient lighting conditions. The BlendedMVS dataset contains a large collection of indoor and outdoor scenes, each featuring 360-degree surround view captures with varying scales and numbers of images. Additionally, we conduct experiments on 3 large-scale outdoor scenes from the TNT dataset with more randomized viewpoints and biased captures. As discussed in (Shen et al. 2022), training with sparse images can ensure variations in the reconstruction quality, providing an ideal setup to evaluate the uncertainty modeling ability. Therefore, based on the spatial distribution within each scene, we uniformly sample a sparse number of views for the training ($\sim$6) and test ($\sim$3) sets in the DTU dataset, for the BlendedMVS and TNT datasets, we uniformly sample 25% images for the training set and choose 4 adjacent images as the test set.

Table 2 — Block 1

Scenes	BMVS-Angel				BMVS-Dog				BMVS-Egg				BMVS-Jade			
Metrics	AUSE(↓)			CD(↓)	AUSE(↓)			CD(↓)	AUSE(↓)			CD(↓)	AUSE(↓)			CD(↓)
Methods	ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D	
CFNeRF*	0.384	0.318	—	1.324	0.354	0.287	—	1.904	0.337	0.306	—	1.841	0.335	0.291	—	1.433
ActiveNeRF*	1.425	1.158	1.471	0.385	0.796	0.686	0.786	0.632	0.608	0.631	0.606	1.548	1.513	0.946	1.067	0.912
Lee et al.*	1.945	1.380	—		1.929	1.599	—		2.644	1.500	—		1.782	1.615	—	
Bayes' Rays	0.521	0.469	0.201	0.365†	0.258	0.225	0.157	0.469†	0.147	0.172	0.078	0.875†	0.233	0.238	0.143	0.904†
Ours	0.295	0.271	0.111		0.112	0.114	0.115		0.224	0.206	0.069		0.130	0.207	0.129	

Table 2 — Block 2

Scenes	BMVS-Sculpture				BMVS-Soilder				BMVS-Stonelion				BMVS-Toylion			
Metrics	AUSE(↓)			CD(↓)	AUSE(↓)			(CD↓)	AUSE(↓)			CD(↓)	AUSE(↓)			CD(↓)
Methods	ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D	
CFNeRF*	0.348	0.399	—	1.346	0.395	0.406	—	2.053	0.463	0.495	—	1.712	0.372	0.315	—	1.987
ActiveNeRF*	1.270	1.056	1.279	0.572	0.704	1.103	1.196	0.569	1.237	1.085	1.158	0.493	0.926	1.172	0.852	0.462
Lee et al.*	2.129	1.988	—		1.696	1.070	—		2.282	1.640	—		2.329	1.720	—	
Bayes' Rays	0.720	0.564	0.259	0.560†	0.147	0.195	0.081	0.541†	0.296	0.357	0.223	0.477†	0.345	0.227	0.175	0.265†
Ours	0.167	0.205	0.212		0.093	0.146	0.079		0.226	0.232	0.192		0.299	0.275	0.272	

Table 2 — Block 3

Scenes	DTU-scan55				DTU-scan63				DTU-scan83				DTU-scan105			
Metrics	AUSE(↓)			CD(↓)	AUSE(↓)			CD(↓)	AUSE(↓)			CD(↓)	AUSE(↓)			CD(↓)
Methods	ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D	
CFNeRF*	0.367	0.463	—	4.205	0.385	0.426	—	4.357	0.370	0.465	—	4.409	0.585	0.624	—	3.978
ActiveNeRF*	0.634	0.671	0.781	0.630	0.512	0.502	0.923	1.458	1.352	1.407	1.006	0.961	1.491	1.132	0.754	0.849
Lee et al.*	1.274	1.297	—		1.375	1.697	—		1.941	1.449	—		2.096	1.219	—	
Bayes' Rays	0.348	0.290	0.313	0.675†	0.252	0.402	0.499	1.154†	0.934	0.878	0.609	1.035†	2.136	1.440	0.547	0.839†
Ours	0.231	0.253	0.192		0.207	0.256	0.284		0.232	0.249	0.303		0.208	0.328	0.363	

Table 2 — Block 4

Scenes	DTU-scan106				DTU-scan114				DTU-scan118				DTU-scan122			
Metrics	AUSE(↓)			CD(↓)	AUSE(↓)			(CD↓)	AUSE(↓)			CD(↓)	AUSE(↓)			CD(↓)
Methods	ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D		ΔMSE	ΔMAE	3D	
CFNeRF*	0.462	0.437	—	3.530	0.366	0.419	—	3.879	0.390	0.387	—	3.914	0.398	0.381	—	3.750
ActiveNeRF*	1.222	1.253	1.049	0.778	1.489	1.167	1.160	0.706	0.470	0.583	1.312	0.910	0.438	0.551	0.923	0.907
Lee et al.*	2.177	1.515	—		1.898	1.633	—		1.341	1.504	—		1.846	1.816	—	
Bayes' Rays	0.181	0.246	0.356	0.656†	0.204	0.341	0.312	0.519†	0.267	0.355	0.393	0.824†	0.330	0.291	0.290	0.864†
Ours	0.150	0.204	0.213		0.171	0.319	0.188		0.217	0.188	0.214		0.233	0.252	0.205	

Table 2: **Uncertainty Quantification for the BlendedMVS and DTU datasets**. Best results are highlighted as **first**, **second**.
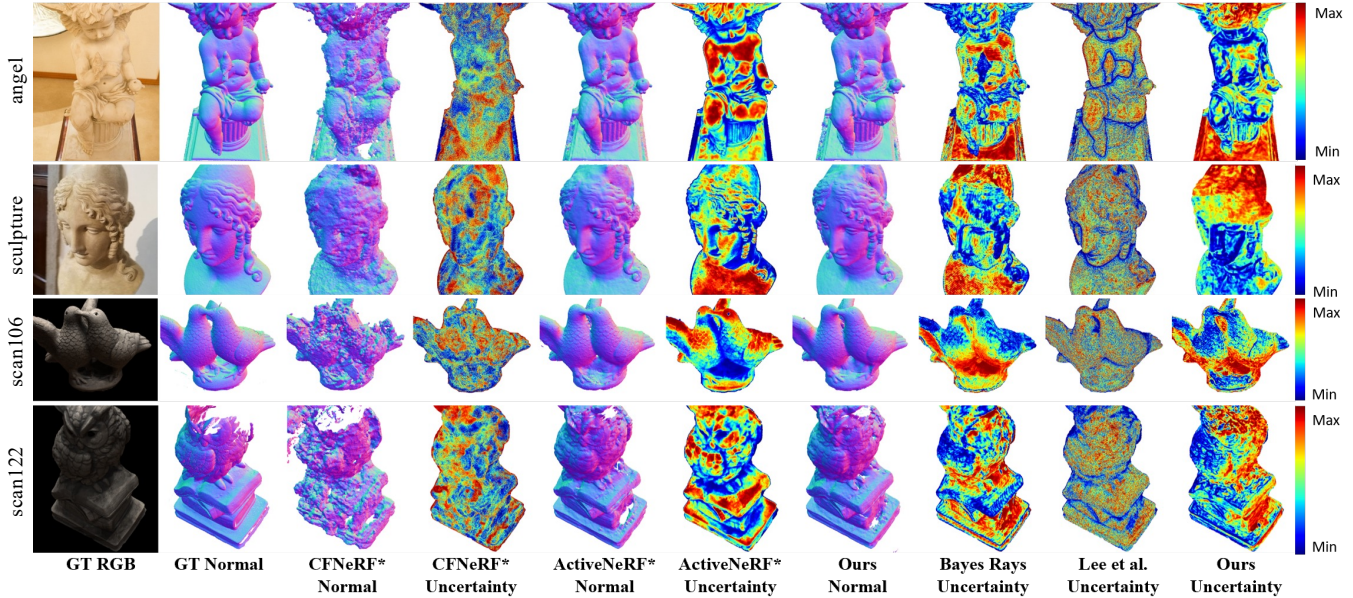† Bayes' Rays and Unc-NeRF evaluate our trained model as post-hoc frameworks and share the same CD metric.

Figure 7: **Qualitative comparison of the uncertainty estimation with 2D depth.** Our estimated uncertainties are more closely aligned with the error between the GT and predicted geometry than other baselines (Pan et al. 2022; Shen et al. 2022; Goli et al. 2023; Lee et al. 2022). Following (Goli et al. 2023), uncertainties are colored based on the ranking order of uncertainty scores.

Scheme	AUSE_{MSE}	AUSE_{MAE}	AUSE_{3D}
w/o decouple finetune	0.104	0.137	0.125
with pixel-based consistency	0.675	0.713	0.792
with smaller patch size 7	0.135	0.254	0.179
with larger patch size 15	0.207	0.195	0.214
Full Model	**0.058**	**0.094**	**0.091**

Table 3: **Ablation studies on BlendedMVS dataset**.

Figure 8: **Extension to 2DGS.** Our uncertainty distillation can be migrated to various neural surface representations. We show the extension to 2DGS (Huang et al. 2024).

Metrics. In the experiment for uncertainty quantification, following previous methods (Shen et al. 2022; Goli et al. 2023), we calculate the Area Under Sparsification Error (AUSE), a widely used metric to assess the quality of model uncertainty. Given the predicted depth error and predicted uncertainty of each pixel in the test image, we gradually remove the top $t\%(t = 1 \sim 100)$ pixels according to two criteria: once based on GT depth error, once based on predicted uncertainty, and compute the average depth error for the remaining pixels. The area between the curves obtained from the two criteria is the AUSE, which reflects the correlation between predicted uncertainty and actual depth error. In addition to calculating AUSE based on the Mean Absolute depth Error (ΔMAE) and Mean Squared Error (ΔMSE), we also compute the AUSE based on the closest distance of each point to the ground truth geometry as the 3D geometric error, *i.e*, AUSE_{3D}, which reflects geometric uncertainty from a 3D perspective. Finally, we evaluate the accuracy of surfaces reconstructed by different methods using the Chamfer Distance (CD). Since the scales among scenes are inconsistent in both BlendedMVS and TNT datasets, we uniformly normalize the scenes to fit within the unit sphere to compute geometric metrics.

Baselines. We compare ours with previous works designed for uncertainty estimation in NeRF: CFNeRF (Shen et al. 2022), ActiveNeRF (Pan et al. 2022), Bayes' Rays (Goli et al. 2023), and Uncertainty-Guided NeRF (Lee et al. 2022). Although the ability to model uncertainty should be independent of reconstruction quality, considering they are all designed based on NeRF (Mildenhall et al. 2021), while ours is designed for the SDF backend, to avoid differences in geometric errors caused by representation affecting the assessment of uncertainty modeling capabilities, we make structural modifications to each method (labeled with [*]), making them applicable to neural surface representation. Please refer to the Supp. Mat. for details. To be noted, for Bayes' Rays (Goli et al. 2023) and Uncertainty-Guided NeRF (Lee et al. 2022), we directly migrate them as post-hoc frameworks to evaluate the model $(f_s(\Phi_s), f_c(\Phi_c))$ trained by our method. From the geometric perspective, although ActiveNeRF[*] and Bayes' Rays cannot directly model the uncertainty of 3D points since they rely on the pixel-level volumetric rendering process, we feed the surface points into their uncertainty model as a rough estimation and compare them in terms of AUSE_{3D} as a reference for better understanding.

Results. As illustrated in Fig. 7 and Table 2, under sparse-view setting, CFNeRF[*] exhibits severe degradation in reconstruction quality and inadequate predictive capabilities for uncertainty, particularly evident in the DTU datasets where training is conducted with only six views. Considering the
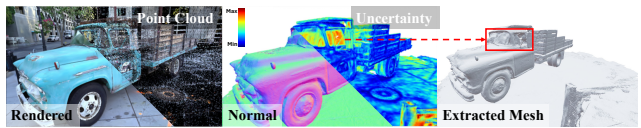
uncertainty prediction of ActiveNeRF[*] is based on modeling the predicted RGB, which is vulnerable to significant disruption when substantial color variations exist across different viewpoints, it does not adequately reflect geometric uncertainty. As NeuS (Wang et al. 2021) is designed under the assumption of an ideal impulse function distribution, the estimation of uncertainty based on the entropy of the weight distribution on sampled rays in Uncertainty-Guided NeRF[*] becomes ineffective. Bayes' Rays has a relatively reasonable performance of modeling uncertainty among existing approaches. However, it still models the uncertainty from the perspective of RGB rendering by perturbing sampling points and tends to predict high uncertainty for regions with abundant repetitive textures regardless of the reconstruction, such as the case of Angel and Sculpture shown in Fig. 7. Compared to other methods, as our approach leverages the advantages of multi-view geometric consistency and decoupling view-dependent factors, GURecon achieves a significant improvement. Even facing scenes with texture repetitions, lighting interferences, and sparse training views as shown in Fig. 7, our method accurately distinguishes unreliable regions and quantitatively evaluates the uncertainty. As the AUSE curves shown in Fig. 5, the depth error identified by ours closely approximates the actual variation curve. From Fig. 4 and Fig. 6, we can also see that our learned 3D geometric uncertainties align well with the real 3D geometric error in both indoor and outdoor large-scale scenes.

4.2 Ablation Study

We conduct an ablation study to demonstrate the effectiveness of each module in the proposed method.

Fine-tuning with Decoupling Modules. As shown in the Fig. 3 and Table 3, fine-tuning with decoupling modules effectively addresses the misclassification issue wherein regions with good reconstruction quality are erroneously classified as reconstruction failures due to lighting interference in the calculation of geometric consistency.

Different Sizes of Patches in Consistency. As shown in Table 3, the utilization of pixel-based and small patch-based consistency fails to accurately reflect geometric uncertainty due to their sensitivity to view-dependent factors such as lighting. Large patches fail to capture geometric consistency in detailed areas such as edges and corners. The patch size used is 11x11. Please refer to more ablation in Supp. Mat.

Plug-and-play extension to 2DGS. Our proposed uncertainty distillation can be migrated to various neural surface representations. We extend it to the latest surface reconstruction work 2DGS (Huang et al. 2024). We utilize the GS corresponding to the median depth of each pixel as the intersec-
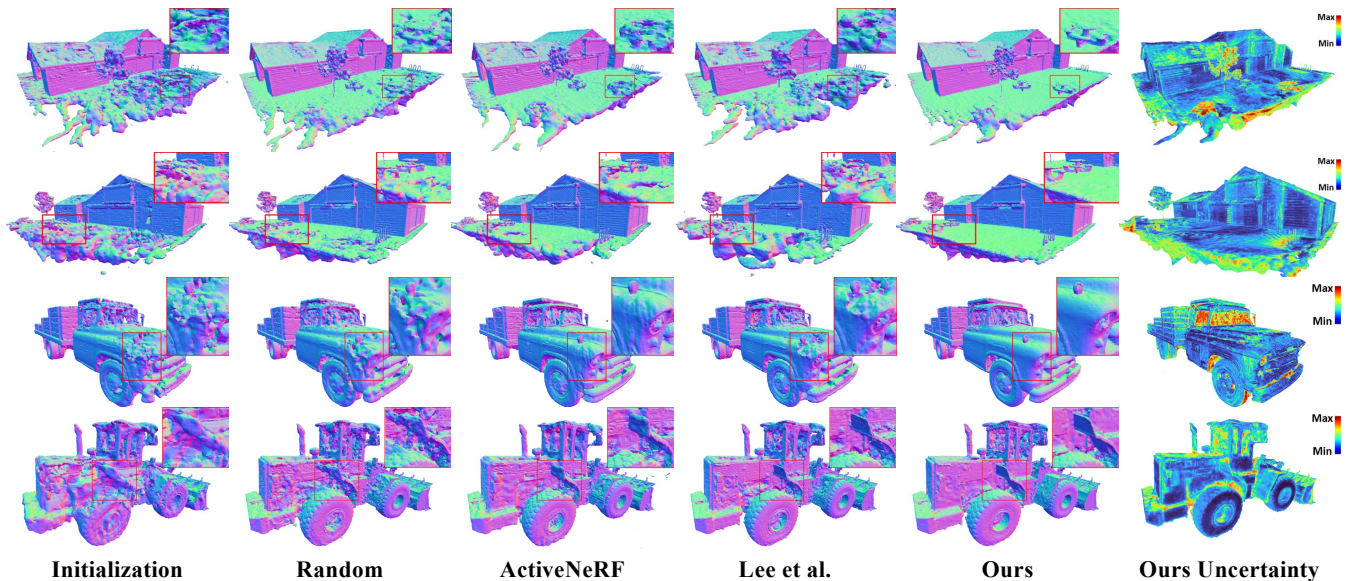
Figure 9: **Incremental results on TNT dataset**. Our scheme reconstructs more details while ensuring a smoother surface.

Scenes	Barn		Caterpillar		Truck	
Methods	CD(↓)	PSNR(↑)	CD(↓)	PSNR(↑)	CD(↓)	PSNR(↑)
Random	1.033	22.69	0.809	19.71	2.248	20.89
ActiveNeRF* (Pan et al. 2022)	1.002	22.72	0.733	19.24	2.123	20.73
Lee et al.* (Lee et al. 2022)	0.983	22.02	0.778	19.24	2.212	20.40
Ours	**0.947**	**23.21**	**0.705**	**20.06**	**2.059**	**21.49**

Table 4: **Quantitative comparison of NBV strategies on TNT dataset.** The best results are highlighted in **bold.**

Figure 10: Limitation in extreme scenarios.

tion with the surface and employ the direction of its shortest axis as the normal for homography warping. With the proposed distillation method, we can supervise an additional attribute of uncertainty for the GS located on the surface, as shown in Fig. 8. Please refer to more details in Supp. Mat.

4.3 Evaluation on Incremental Reconstruction

Datasets and Metrics. We select the same large-scale scenarios from the TNT dataset as used in Sec. 4.1 to measure the effectiveness of the incremental reconstruction strategy. We report Chamfer Distance for surface evaluation and peak signal-to-noise ratio (PSNR) for image synthesis qualities.

Baselines. We compare ours with two representative NeRF-based NBV methods: Uncertainty-Guided NeRF (Lee et al. 2022) and ActiveNeRF (Pan et al. 2022) discussed in Sec. 4.1, alongside a completely randomized NBV strategy. Considering the substantial differences in geometric quality between NeRF and NeuS representations, we adopt the same strategy described in Sec. 4.1.

Implementation Details. We follow the initialization strategy in (Lee et al. 2022) to divide the space into several regions uniformly and select one viewpoint from each region for both the initial training set and the test set. During each selection for Next Best View (NBV), we assess the uncertainty of the remaining viewpoints in each region and select the one with the highest score to augment the training set.
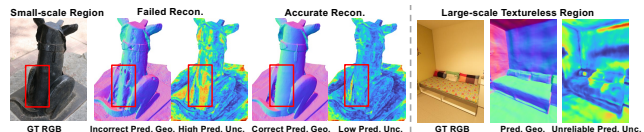
As our method directly models the uncertainty of surface points, we utilize sphere tracing for root-finding and achieve rapid surface rendering of uncertainty for new viewpoints.

Results. As shown in Fig. 9 and Table 4, the NBV strategy with our geometric uncertainty achieves the best reconstruction results under the same limited number of views (30% of the total image). Compared to other methods, our approach reconstructs more details while ensuring a smoother surface. Please refer to Supp. Mat. for a more qualitative comparison.

5 Conclusion

In this paper, we introduce GURecon, a novel approach for learning a 3D geometric uncertainty field for neural surface models. Unlike existing methods that model rendering-based pixel-wise uncertainty, the proposed GURecon exploits the multi-view consistency to accurately model the geometric uncertainty. Moreover, additional decoupled fields are learned for robust uncertainty estimation. Comprehensive experiments have demonstrated our superior performance compared to existing methods. As Fig. 10 shows, while our approach works well in small-scale textureless regions, its performance is limited in extreme scenarios with large textureless areas (*e.g*, white walls), where high-level semantic information can be incorporated in future work.

References

Campbell, N. D.; Vogiatzis, G.; Hernández, C.; and Cipolla, R. 2008. Using multiple hypotheses to improve depth-maps for multi-view stereo. In *Eur. Conf. Comput. Vis.*, 766–779. Springer.

Chen, L.; Chen, W.; Wang, R.; and Pollefeys, M. 2023. Leveraging Neural Radiance Fields for Uncertainty-Aware Visual Localization. *arXiv preprint arXiv:2310.06984*.

Darmon, F.; Bascle, B.; Devaux, J.-C.; Monasse, P.; and Aubry, M. 2022. Improving neural implicit surfaces geometry with patch warping. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 6260–6269.

Fan, Y.; Skorokhodov, I.; Voynov, O.; Ignatyev, S.; Burnaev, E.; Wonka, P.; and Wang, Y. 2023. Factored-NeuS: Reconstructing Surfaces, Illumination, and Materials of Possibly Glossy Objects. *arXiv preprint arXiv:2305.17929*.

Fu, Q.; Xu, Q.; Ong, Y. S.; and Tao, W. 2022. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Adv. Neural Inform. Process. Syst.*, 35: 3403–3416.

Ge, W.; Hu, T.; Zhao, H.; Liu, S.; and Chen, Y.-C. 2023. Ref-NeuS: Ambiguity-Reduced Neural Implicit Surface Learning for Multi-View Reconstruction with Reflection. *arXiv preprint arXiv:2303.10840*.

Goli, L.; Reading, C.; Selllán, S.; Jacobson, A.; and Tagliasacchi, A. 2023. Bayes' Rays: Uncertainty Quantification for Neural Radiance Fields. *arXiv preprint arXiv:2309.03185*.

Hu, X.; and Mordohai, P. 2012. A Quantitative Evaluation of Confidence Measures for Stereo Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11): 2121–2133.

Huang, B.; Yu, Z.; Chen, A.; Geiger, A.; and Gao, S. 2024. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 Conference Papers*, 1–11.

Jensen, R.; Dahl, A.; Vogiatzis, G.; Tola, E.; and Aanæs, H. 2014. Large scale multi-view stereopsis evaluation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 406–413.

Jiang, W.; Lei, B.; and Daniilidis, K. 2023. FisherRF: Active View Selection and Uncertainty Quantification for Radiance Fields using Fisher Information. *arXiv preprint arXiv:2311.17874*.

Knapitsch, A.; Park, J.; Zhou, Q.-Y.; and Koltun, V. 2017. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Trans. Graph.*, 36(4): 1–13.

Lee, S.; Chen, L.; Wang, J.; Liniger, A.; Kumar, S.; and Yu, F. 2022. Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields. *IEEE Robotics and Automation Letters*, 7(4): 12070–12077.

Liu, L.; Gu, J.; Zaw Lin, K.; Chua, T.-S.; and Theobalt, C. 2020. Neural sparse voxel fields. *Adv. Neural Inform. Process. Syst.*, 33: 15651–15663.

Martin-Brualla, R.; Radwan, N.; Sajjadi, M. S.; Barron, J. T.; Dosovitskiy, A.; and Duckworth, D. 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 7210–7219.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Oechsle, M.; Peng, S.; and Geiger, A. 2021. Unisurf: Unifying neural implicit surfaces and radiance fields for multiview reconstruction. In *Int. Conf. Comput. Vis.*, 5589–5599.

Pan, X.; Lai, Z.; Song, S.; and Huang, G. 2022. Activenerf: Learning where to see with uncertainty estimation. In *Eur. Conf. Comput. Vis.*, 230–246.

Ran, Y.; Zeng, J.; He, S.; Chen, J.; Li, L.; Chen, Y.; Lee, G.; and Ye, Q. 2023. NeurAR: Neural Uncertainty for Autonomous 3D Reconstruction With Implicit Neural Representations. *IEEE Robotics and Automation Letters*, 8(2): 1125–1132.

Schönberger, J. L.; Zheng, E.; Frahm, J.-M.; and Pollefeys, M. 2016. Pixelwise view selection for unstructured multiview stereo. In *Eur. Conf. Comput. Vis.*, 501–518. Springer.

Shen, J.; Agudo, A.; Moreno-Noguer, F.; and Ruiz, A. 2022. Conditional-flow NeRF: Accurate 3D modelling with reliable uncertainty quantification. In *Eur. Conf. Comput. Vis.*, 540–557. Springer.

Shen, J.; Ruiz, A.; Agudo, A.; and Moreno-Noguer, F. 2021. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In *International Conference on 3D Vision (3DV)*, 972–981. IEEE.

Stereopsis, R. M. 2010. Accurate, Dense, and Robust Multiview Stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(8).

Sünderhauf, N.; Abou-Chakra, J.; and Miller, D. 2023. Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. 9370–9376. IEEE.

Tang, J.; Zhou, H.; Chen, X.; Hu, T.; Ding, E.; Wang, J.; and Zeng, G. 2023. Delicate textured mesh recovery from nerf via adaptive surface refinement. *arXiv preprint arXiv:2303.02091*.

Verbin, D.; Hedman, P.; Mildenhall, B.; Zickler, T.; Barron, J. T.; and Srinivasan, P. P. 2022. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Wang, J.; Wang, P.; Long, X.; Theobalt, C.; Komura, T.; Liu, L.; and Wang, W. 2022. Neuris: Neural reconstruction of indoor scenes using normal priors. In *European Conference on Computer Vision*, 139–155. Springer.

Wang, P.; Liu, L.; Liu, Y.; Theobalt, C.; Komura, T.; and Wang, W. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Adv. Neural Inform. Process. Syst.*

Xiao, Y.; Xu, J.; Yu, Z.; and Gao, S. 2024. Debsdf: Delving into the details and bias of neural indoor scene reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Yao, Y.; Luo, Z.; Li, S.; Zhang, J.; Ren, Y.; Zhou, L.; Fang, T.; and Quan, L. 2020. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 1790–1799.

Yariv, L.; Gu, J.; Kasten, Y.; and Lipman, Y. 2021. Volume rendering of neural implicit surfaces. *Adv. Neural Inform. Process. Syst.*, 34: 4805–4815.

Yu, Z.; Peng, S.; Niemeyer, M.; Sattler, T.; and Geiger, A. 2022. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Adv. Neural Inform. Process. Syst.*, 35: 25018–25032.

Zhan, H.; Zheng, J.; Xu, Y.; Reid, I.; and Rezatofighi, H. 2022. ActiveRMAP: Radiance Field for Active Mapping And Planning. *arXiv preprint arXiv:2211.12656*.

Zhang, K.; Riegler, G.; Snavely, N.; and Koltun, V. 2020. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*.

Zhao, F.; Jiang, Y.; Yao, K.; Zhang, J.; Wang, L.; Dai, H.; Zhong, Y.; Zhang, Y.; Wu, M.; Xu, L.; et al. 2022. Human performance modeling and rendering via neural animated mesh. *ACM Trans. Graph.*, 41(6): 1–17.