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

In this supplementary material, we provide more details of our GURecon framework, including 1) detailed network architecture of GURecon (Sec. A); 2) modifications to other compared baselines (Sec. B); 3) more experiment results (Sec. C); 4) details of incremental reconstruction (Sec. D). Additionally, we provide a video supplementary material to summarize our method and demonstrate the results.

A Model Architecture

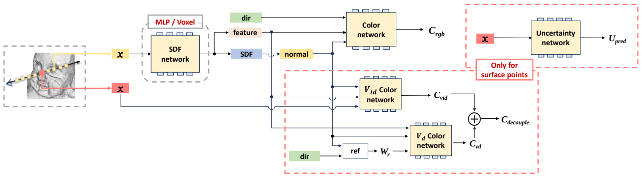

Figure A. The network architecture of GURecon. In addition to the SDF network and the Color network proposed in NeuS (Wang et al. 2021), we introduce decoupled fields to mitigate the interference of view-dependent factors, and an uncertainty field to predict the geometric uncertainty.

The detailed architecture of GURecon is depicted in Fig.A. Taking NeuS (Wang et al. 2021) as a representative in neural surface representations, we establish an SDF network and a Color network to learn the neural geometric representation and radiance field of the scene based on the volume rendering formulation. Note that the representation of the SDF network can be substituted with various architectures such as MLP (Wang et al. 2021; Yariv et al. 2021), voxel grid (Zhao et al. 2022; Li et al. 2023; Wu et al. 2022) and tetrahedral lattice (Rosu and Behnke 2023), as our method serves as a plug-and-play module applicable to diverse neural surface reconstructions. In our paper, we use a multi-resolution hash feature grid (Müller et al. 2022) for memory and time efficiency. Specifically, the feature grid adopts a coarse-to-fine structure consisting of 16 levels, with each level growing exponentially from the lowest resolution of 32^3 to the highest resolution of 2048^3. At each level, the grid is organized and indexed through a hash table, where each hash entry has a feature with a channel size of 2. Towards the end of the SDF network, a two-layer MLP with 128 hidden units in each layer is incorporated. Following NeuS (Wang et al. 2021), we initialize the SDF as a sphere and train the network to produce both a 256-dimensional feature and an SDF value for the input spatial position. The Color Network encodes the appearance in radiance fields and consists of two-layer MLPs, each with 64 hidden units.

Furthermore, to disentangle appearance components for better modeling view-dependent factors, two networks are employed to represent view-independent and view-dependent factors as (Fan et al. 2023). The view-independent network relates solely to spatial position and normal vector, while the view-dependent network additionally considers the direction of reflections. Each module is implemented with an MLP comprising two layers, with 64 and 32 hidden units respectively.

To speed up distilling the uncertainty field, we utilize another feature grid structure similar to the SDF network. The configuration of the grid parameter is L=8, N_{min}=16, N_{max}=512, Feature$_{dim}$=2, followed by a two-level MLP with a dimension of 64 and a ReLU activation function to prevent negative outputs. The network takes the spatial position of surface points as input and outputs the corresponding view-independent geometric uncertainties.

B Modifications to Baselines

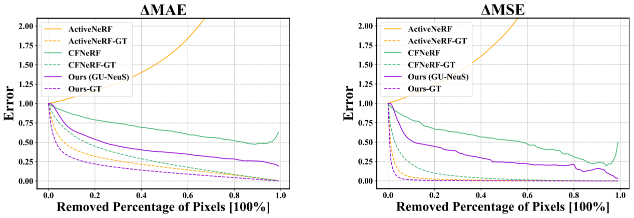

Figure B: **Sparsification curves.** The figure illustrates the AUSE curves based on ΔMAE and ΔMSE for ActiveNeRF (Pan et al. 2022), CFNeRF (Shen et al. 2022) and our method. The dashed lines depict varying actual geometric errors corresponding to different methods. The discrepancies introduce interference in quantifying the capability for uncertainty estimation. Therefore, as shown in Fig. 5 in our main paper, we unify the representation using an SDF-based representation to mitigate the significant interference of different actual geometric errors on uncertainty quantification.

It is widely acknowledged that SDF-based methods outperform NeRF-based methods in the quality of geometric reconstruction. To validate the impact of geometric errors on uncertainty quantification, we first make a comparison

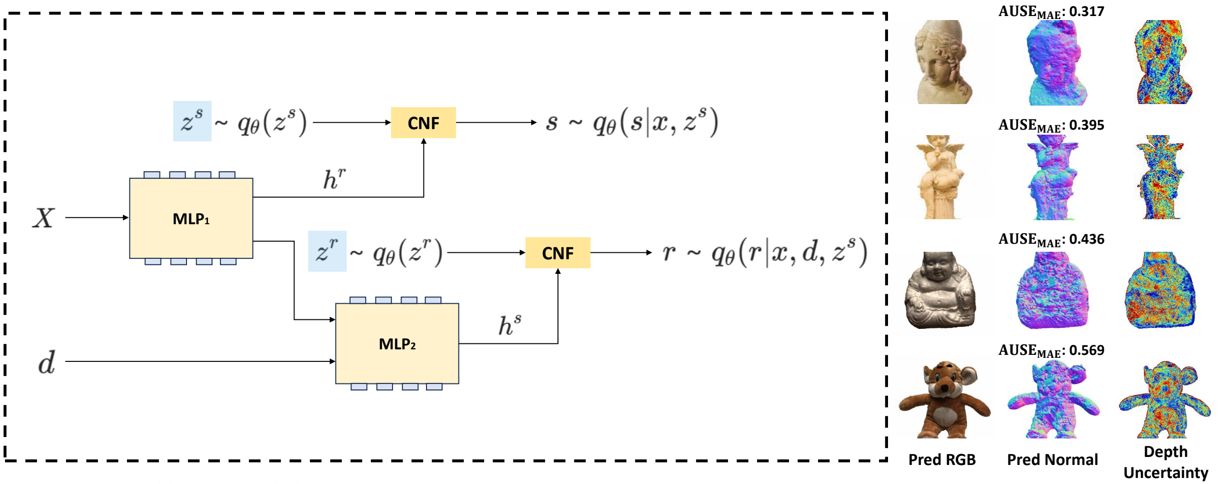
$z^s \sim q_\theta(z^s)$
CNF
$s \sim q_\theta(s|x, z^s)$
h^r
X
MLP₁
$z^r \sim q_\theta(z^r)$
CNF
$r \sim q_\theta(r|x, d, z^s)$
h^s
d
MLP₂
AUSE$_{\text{MAE}}$: 0.317
AUSE$_{\text{MAE}}$: 0.395
AUSE$_{\text{MAE}}$: 0.436
AUSE$_{\text{MAE}}$: 0.569
Pred RGB
Pred Normal
Depth Uncertainty

Figure C: **The modification of CFNeRF**. We make structural modifications to the CFNeRF architecture to adapt it for the SDF-based representation. We show the rendered results and reconstruction quality of the modified CFNeRF.

with the NeRF-based uncertainty estimation methods: CFNeRF (Shen et al. 2022) and ActiveNeRF (Pan et al. 2022). As demonstrated by the AUSE curves in Fig. B, the quantification of uncertainty is disrupted by the discrepancies of the actual geometric errors associated with different methods.

Therefore, we make structural modifications to each method for fairness considerations, converting the NeRF-based architecture into the SDF-based architecture to avoid significant differences in geometric errors that could affect the assessment of uncertainty modeling capabilities. The specific modifications made to each method are outlined below:

CFNeRF (Shen et al. 2022): As Fig. C shows, we follow the setting in (Shen et al. 2022) and use two Conditional Normalizing Flow (CNF) models to sample radiance and SDF values instead of density values from distributions $q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, z)$ and $q_{\boldsymbol{\theta}}(\mathbf{s}|\mathbf{x}, z)$. Each CNF computes a transformation of a sample from the latent distribution $q_\psi(z)$ conditioned on an embedding h, which is computed by an MLP with the location-view pair $(\mathbf{x}, \mathbf{d})$ as input. We follow the optimization process of CFNeRF (Shen et al. 2022) and adopt a Variational Bayesian approach to learn the posterior distribution defined in (Shen et al. 2022) based on the volume rendering formulation proposed in (Wang et al. 2021). When computing the uncertainty, we sample a set of random variables $\mathbf{z}_{1:K}$ from latent distribution $q_\vartheta(z)$ and obtain a set of estimated depth-values for each pixel. The mean and variance over the K samples are treated as the estimated depth and its associated uncertainty. The results of the modified methods are presented in Fig. C. Despite ensuring satisfactory rendering quality under sparse viewpoints, there is a decline in the reconstruction quality, as sampling from probability distributions in CNF introduces noise in geometry.

ActiveNeRF (Pan et al. 2022): Follow (Pan et al. 2022), we define the radiance color of a location $\mathrm{r}(t)$ as a Gaussian distribution $c(\mathrm{r}(t)) \sim \mathcal{N}(\bar{c}(\mathrm{r}(t)), \bar{\beta}^2(\mathrm{r}(t)))$. We replace the output of the density MLP with an SDF value s and incorporate an additional branch to the MLP network to model the variance β^2 as follows:

$$\begin{aligned}[\mathrm{s}, f, \beta^2(\mathrm{r}(t))] =& \mathrm{MLP}_{\theta_1,\theta_3}(\gamma_{\mathrm{x}}(\mathrm{r}(t))), \quad (1)\\ \bar{c}(\mathrm{r}(t)) =& \mathrm{MLP}_{\theta_2}(f, \gamma_{\mathrm{d}}(\mathrm{d})). \quad (2)\end{aligned}$$

We optimize the model by minimizing the negative log-likelihood of rays $\{r_{i=1}^N\}$:

$$\min_\theta \frac{1}{N}\sum_{i=1}^{N} \frac{\|C(\mathrm{r}_i) - \bar{C}(\mathrm{r}_i)\|_2^2}{2\bar{\beta}^2(\mathrm{r}_i)} + \frac{\log \bar{\beta}^2(\mathrm{r}_i)}{2}. \quad (3)$$

The predicted color $\bar{C}(\mathrm{r}_i)$ is rendered with the volume rendering formulation as (Wang et al. 2021), and the variance of the rendered RGB $\beta^2(\mathrm{r}_i)$ is treated as the uncertainty.

Uncertainty-Guided NeRF (Lee et al. 2022): Follow (Lee et al. 2022), we reconstruct the scene utilizing the SDF-based representation (Wang et al. 2021) and compute the entropy of the weight distribution along the rays, considering the regions with non-concentrated weights indicate areas where the 3D geometry can be improved. Since the weight $w(t)$ can be viewed as a Probability Density Function (PDF), the entropy of a ray is defined as:

$$H(r) = -\sum_{i=1}^{N} w(r(t_i))log(w(r(t_i))). \quad (4)$$

We calculate the entropy of each pixel's corresponding ray and utilize it as the uncertainty for that pixel.

C Additional Experiments Results

Additional decoupled results and estimated geometric uncertainty. We present additional results on the BlendedMVS dataset and DTU dataset with the decoupled results and the estimated geometric uncertainties in Fig. D.

Details of ablation study. We present more results on different sizes of image patches, the versatility across different numbers of training images, and other factors influencing consistency computation on uncertainty quantification.

As shown in Fig. E, the utilization of small patch-based consistency fails to accurately reflect geometric uncertainty due to their sensitivity to view-dependent factors such as

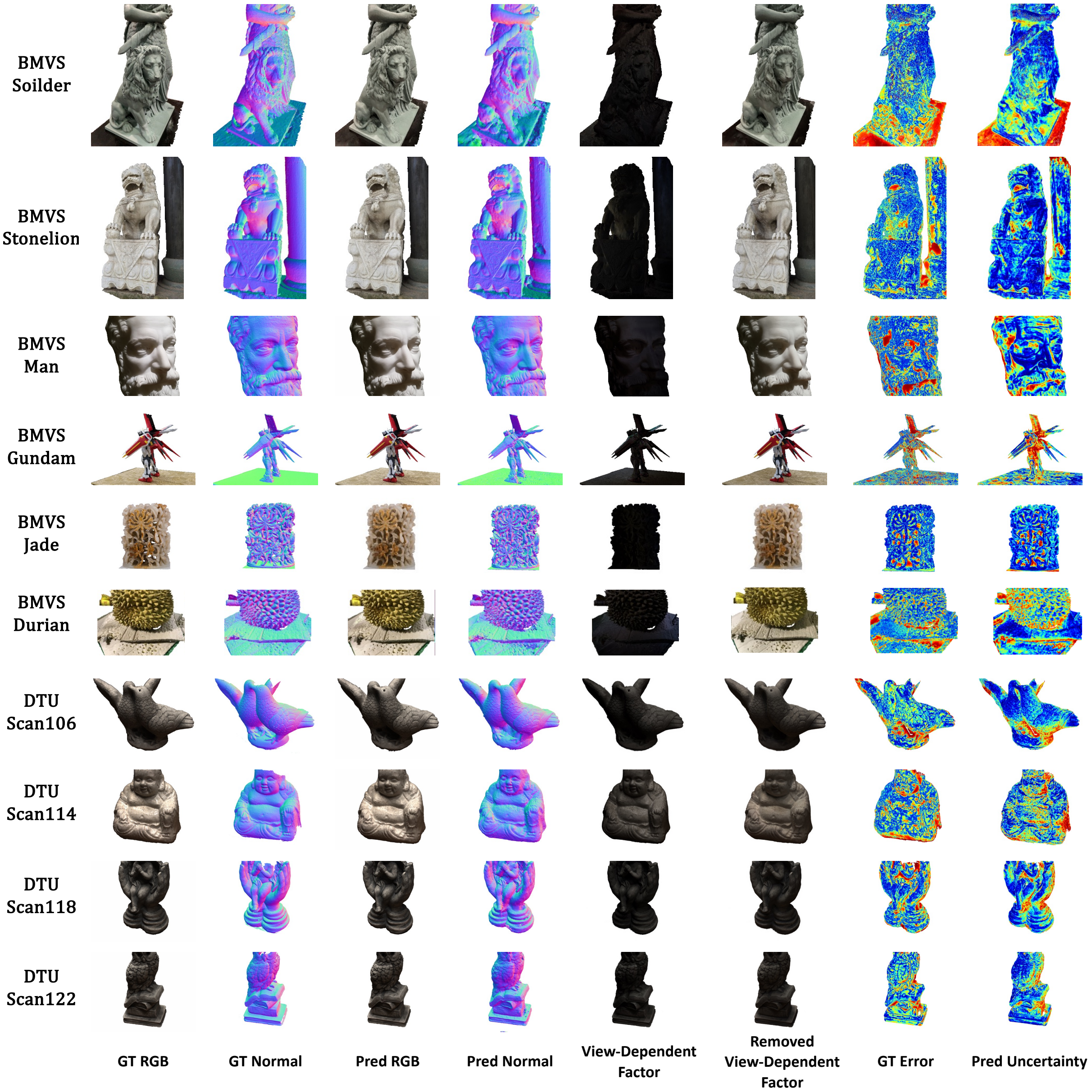

Figure D: **Additional decoupled results and estimated geometric uncertainty.** Our method presents accurate decoupled results for view-dependent factors, and the learned geometric uncertainties align well with the real geometric error.

lighting, and large patch fails to capture geometric consistency in detailed areas such as edges and corners. Although we assume that the region surrounding a point can be treated as a local plane, our method remains applicable to regions such as sharp edges or corners, as the Durian and Jade in Fig. D show.

As shown in Fig. F, our method accurately models geometric uncertainty across different numbers of training views, leveraging its strengths in occlusion awareness and robustness to lighting interferences.

Additional results in the transparent region. Reconstructing transparent objects like glass is inherently challenging in multi-view stereo (MVS) tasks. As shown in Fig. H, our method is still able to some extent to overcome it and predict accurate geometric uncertainty in general: for the cases where MVS-based methods fail in reconstruction, ours predicts high uncertainty; for the cases where accurate geometry is recovered using additional constraints (*e.g.* GeoNeuS), ours also achieves accurate uncertainty estimation utilizing the proposed finetuning with decoupled fields to mitigate view-dependent influences.

Additional analysis about pseudo labels. Considering our method computes multi-view consistency as pseudo labels, we conduct further analysis on the settings related to multi-view consistency as shown in Fig. G. In calculating the similarity between projected pixel patches, while Normalized Cross-Correlation (NCC) and Structural Similarity Index Measure (SSIM) are two widely used metrics in multi-view stereo tasks, SSIM considers the structural information of images, making it more robust to variations in lighting, con-

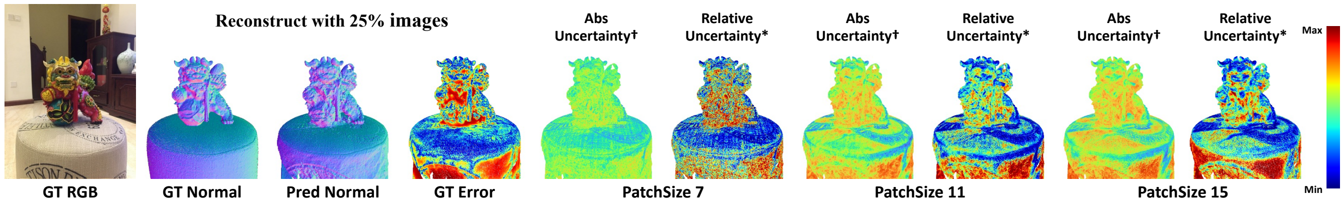

Figure E: **Ablation study of different patch sizes**. Small patch size is sensitive to view-dependent factors, large patch size struggles to capture finer details.

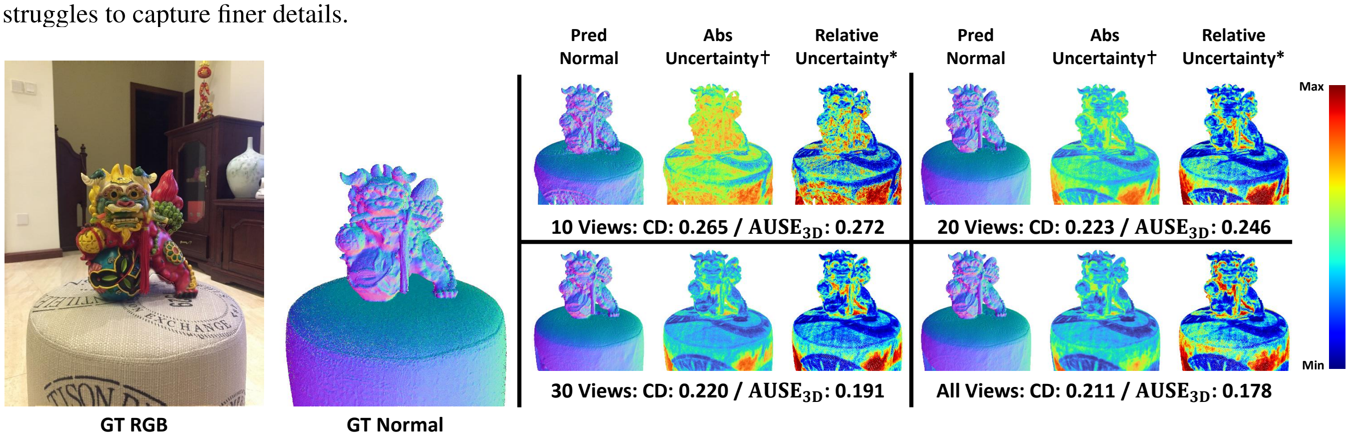
Pred Normal
Abs Uncertainty†
Relative Uncertainty*
Pred Normal
Abs Uncertainty†
Relative Uncertainty*
Max
Min
10 Views: CD: 0.265 / AUSE3D: 0.272
20 Views: CD: 0.223 / AUSE3D: 0.246
30 Views: CD: 0.220 / AUSE3D: 0.191
All Views: CD: 0.211 / AUSE3D: 0.178
GT RGB
GT Normal

Figure F: **Ablation study of varying numbers of training views**. Both the estimated uncertainty (†) and the ranked uncertainty score (*) are shown. Our method accurately models the geometric uncertainty across different numbers of training views.

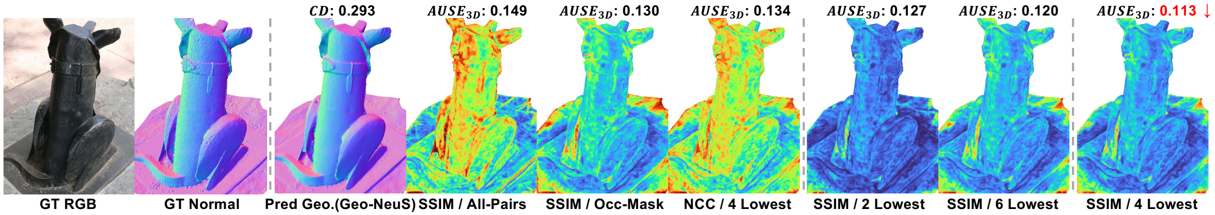

Figure G: **Additional analysis about pseudo labels.** We conduct a detailed analysis of image similarity and patch pair selection strategies to demonstrate the rationale of our scheme.

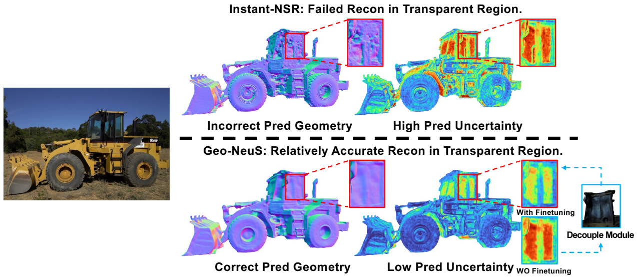

Figure H: **Results in transparent areas.** Our method is able to predict accurate geometric uncertainty in general.

trast, and scale when computing photometric consistency, therefore, we utilize it to better assess the consistency of the projections. We select the four patch pairs with the lowest computed scores to overcome occlusion and lighting disturbances without incurring additional costs. We also test ray casting to detect occlusions between pairs and filter them out when computing consistency as adopted in RefNeuS [10], and as shown in Fig. G, our strategy excels in overcoming disturbances caused by viewing angles and lighting conditions. Besides, selecting the lowest 4 scores is a common practice in traditional MVS [9] and we also demonstrate it achieving better predictions compared to selecting 2 and 6.

D Details of Incremental Reconstruction

In this section, we first visualize the uncertainty estimated by our methods and the reconstructed geometry during incremental reconstruction in the order of training stages. As shown in Figure J, with the increased number of views, the reconstructed mesh achieves higher quality and the uncertainty shows lower scores. Moreover, we conduct qualitative comparisons on reconstruction quality across different NeRF-based next-best-view(NBV) strategies. As shown in Table 4 in the main paper, our geometric uncertainty-guided NBV selection strategy achieves the best reconstruction results under limited views (roughly 30% of the total image). And as shown in Figure I, our strategy outperforms others with more details such as leaves and twigs (in Barn), and lights (in Truck). It is evident that our method demonstrates superior performance in both surface reconstruction and novel view synthesis.

E Plug-and-play Results with Various Neural Surface Models

Our GURecon serves as a plug-and-play module applicable to various neural surface reconstructions as long as the geometric surface can be computed on the fly. As shown in Fig. K, we integrate GURecon as an additional module into

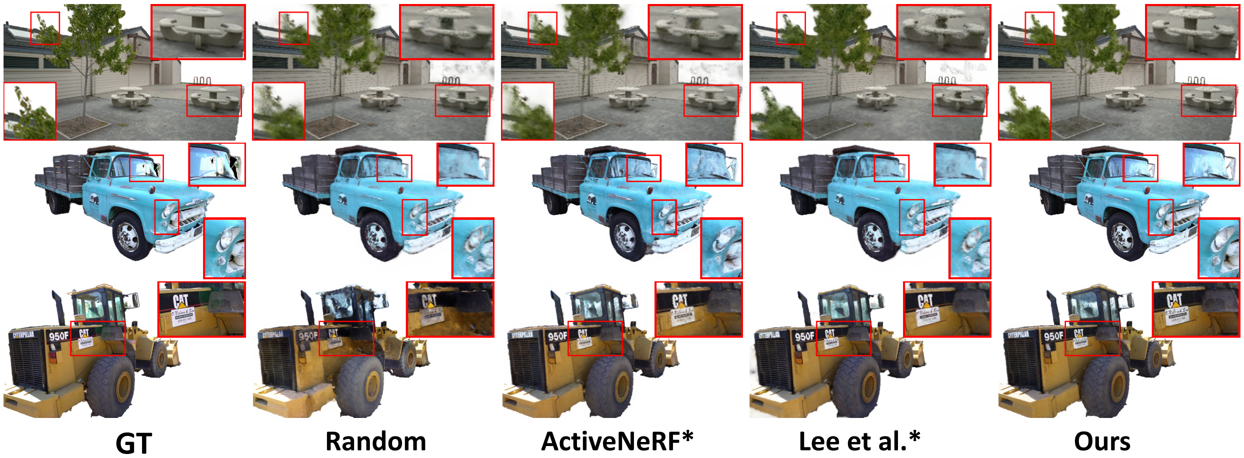

Figure I: **Comparison of novel view synthesis on TNT dataset**. Our strategy outperforms others with more details such as leaves and twigs (in Barn), and lights (in Truck). * corresponds to the SDF-based representation, please refer to Sec. 4.3 in our main paper.

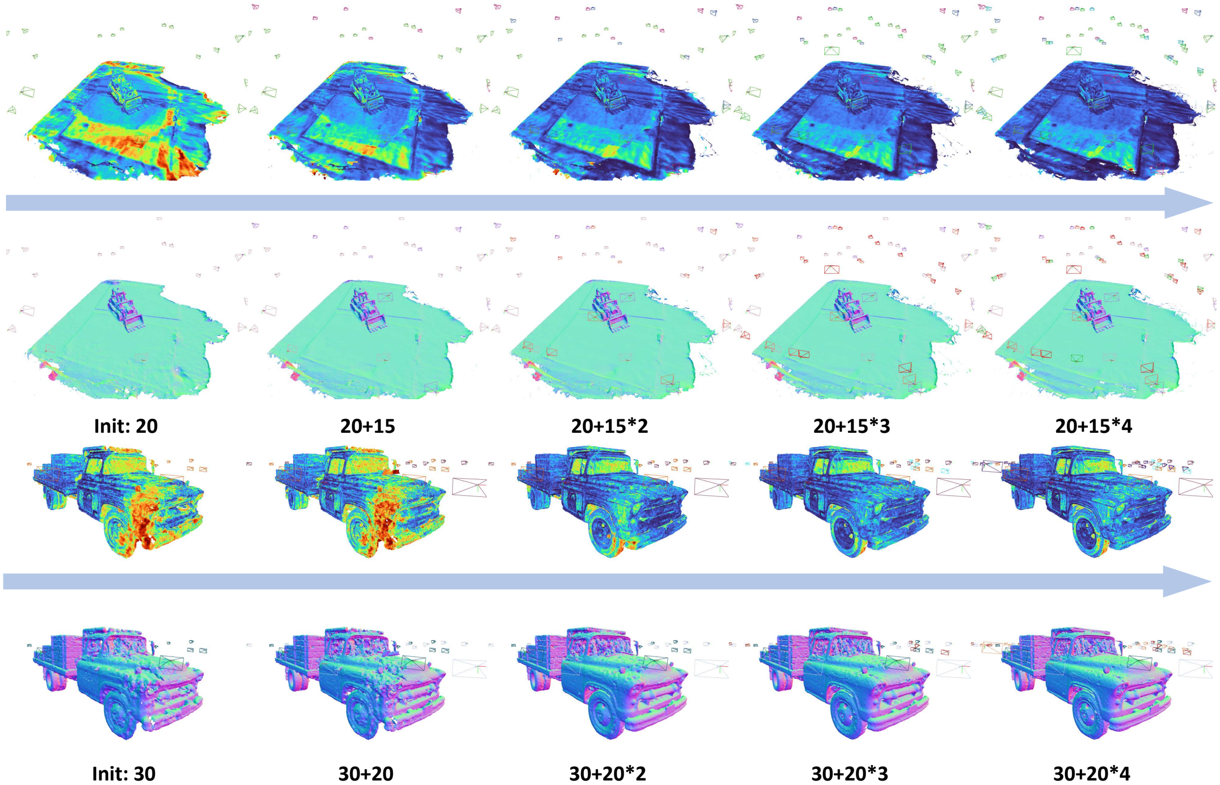

Figure J: **Visualization of incremental reconstruction**. We visualize the geometry and uncertainty estimated by our method during the incremental reconstruction with the changing number of training views.

existing mainstream NeuS approaches: NeuS (Wang et al. 2021), Geo-NeuS (Fu et al. 2022), NeuS-NGP (Zhao et al. 2022) and NeuralAngelo (Li et al. 2023). Our method accurately estimates geometric uncertainty across various models without incurring additional training costs, highlighting its plug-and-play generalizability to other neural surface models.

Detailed description of the plug-and-play extension to 2DGS. We also extend our proposed uncertainty distillation to the latest surface reconstruction work 2DGS (Huang et al. 2024). 2DGS is a state-of-the-art point-based renderer with splendid geometry performance and represents the scene's geometry as a set of 2D Gaussians. A 2D Gaussian is defined in a local tangent plane in world space, parameterized as follows:

$$P(u, v) = \mathbf{p}_k + s_u \mathbf{t}_u u + s_v \mathbf{t}_v v, \tag{5}$$

where $\mathbf{p}_k$ is the central point, $\mathbf{t}_u$ and $\mathbf{t}_v$ are the principal tangential vectors that determine its orientation, and $\mathbf{S} = (s_u, s_v)$ is the scaling vector which controls the variances of the 2D Gaussian distribution. For the point $\mathbf{u} = (u, v)$ in uv space, its 2D Gaussian value can then be evaluated using the standard Gaussian function:

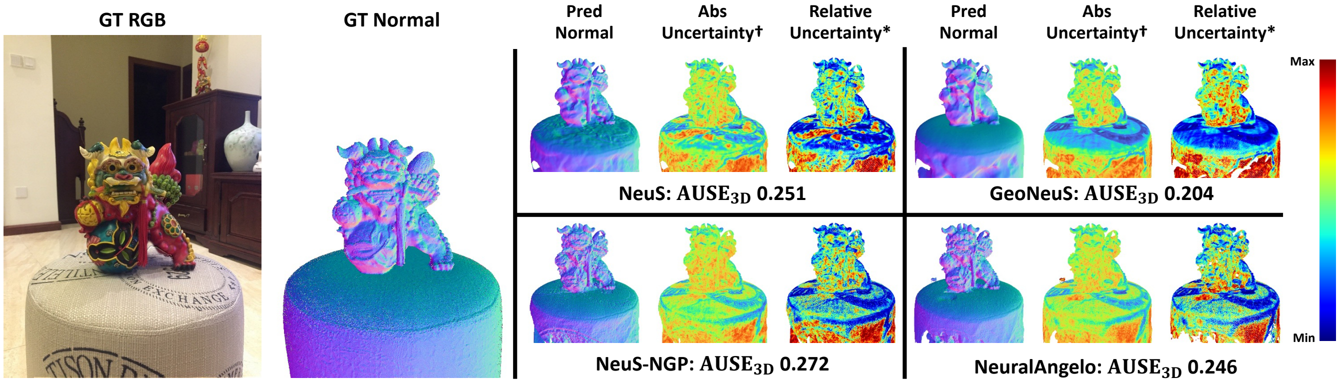

Figure K: **Plug-and-play results in various neural surface models.** Our proposed geometric uncertainty field can be migrated as a plug-and-play module to any neural surface representation, providing an accurate estimation of geometric uncertainty.

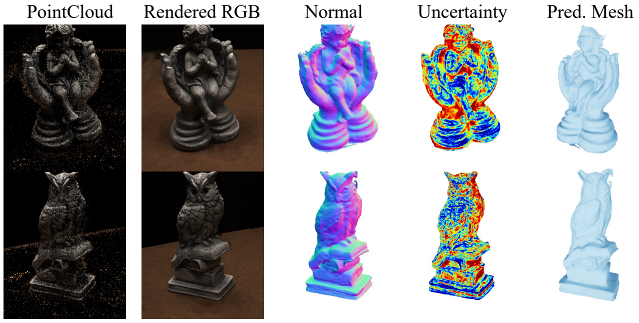

Figure L: **Plug-and-play extension to 2DGS.** We extend our proposed uncertainty distillation into 2DGS.

$$\mathcal{G}(\mathbf{u}) = \exp\left(-\frac{u^2+v^2}{2}\right). \tag{6}$$

The center $\mathbf{p}_k$, scaling (s_u, s_v), and the rotation $(\mathbf{t}_u, \mathbf{t}_v)$ are learnable parameters. Each 2D Gaussian primitive has opacity α and view-dependent appearance $\mathbf{c}$ with spherical harmonics. Through differentiable rasterization, Gaussians are sorted according to their depth value and composed into an image with front-to-back alpha blending:

$$\mathbf{c}(\mathbf{x}) = \sum_{i=1} \mathbf{c}_i \alpha_i \mathcal{G}_i(\mathbf{u}(\mathbf{x})) \prod_{j=1}^{i-1} (1 - \alpha_j \mathcal{G}_j(\mathbf{u}(\mathbf{x}))), \tag{7}$$

where $\mathbf{x}$ represents a homogeneous ray emitted from the camera and passing through uv space.

The main challenge in extending our method into 2DGS is how to obtain the corresponding geometric surface during the training process. Considering the distribution of 2D Gaussian weights corresponding to each pixel is relatively concentrated, we utilize the GS corresponding to the median depth of each pixel where the accumulated opacity reaches 0.5 as the intersection with the actual surface, employ the direction of its shortest axis as the normal, and then substitute them into Eq. 3 in the main paper for homography warping. As shown in Fig. L, with the proposed distillation method, we can supervise an additional attribute of uncertainty for the GS located on the surface.

References

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.

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.

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.

Li, Z.; Müller, T.; Evans, A.; Taylor, R. H.; Unberath, M.; Liu, M.-Y.; and Lin, C.-H. 2023. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 8456–8465.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4): 1–15.

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.

Rosu, R. A.; and Behnke, S. 2023. Permutosdf: Fast multi-view reconstruction with implicit surfaces using permutohedral lattices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8466–8475.

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.

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.*

Wu, T.; Wang, J.; Pan, X.; Xu, X.; Theobalt, C.; Liu, Z.; and Lin, D. 2022. Voxurf: Voxel-based efficient and accurate neural surface reconstruction. *arXiv preprint arXiv:2208.12697*.

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

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.