MonoInstance: Enhancing Monocular Priors via Multi-view Instance Alignment for Neural Rendering and Reconstruction

Anonymous CVPR submission

Paper ID 12070

Abstract

Monocular depth priors have been widely adopted by neural rendering in multi-view based tasks such as 3D reconstruction and novel view synthesis. However, due to the inconsistent prediction on each view, how to more effectively leverage monocular cues in a multi-view context remains a challenge. Current methods treat the entire estimated depth map indiscriminately, and use it as ground truth supervision, while ignoring the inherent inaccuracy and cross-view inconsistency in monocular priors. To resolve these issues, we propose ***MonoInstance****, a general approach that explores the uncertainty of monocular depths to provide enhanced geometric priors for neural rendering and reconstruction. Our key insight lies in aligning each segmented instance depths from multiple views within a common 3D space, thereby casting the uncertainty estimation of monocular depths into a density measure within noisy point clouds. For high-uncertainty areas where depth priors are unreliable, we further introduce a constraint term that encourages the projected instances to align with corresponding instance masks on nearby views. MonoInstance is a versatile strategy which can be seamlessly integrated into various multi-view neural rendering frameworks. Our experimental results demonstrate that MonoInstance significantly improves the performance in both reconstruction and novel view synthesis under various benchmarks.*

1. Introduction

Learning scene representations from multiple posed RGB images is a foundational task in computer vision and graphics [2, 23, 63, 71], with numerous applications across diverse domains such as virtual reality, robotics and autonomous driving. Bridging the gap between 2D images and 3D representations has become a central challenge in the field. Traditional approaches like Multi-View Stereo (MVS) [59, 69], address this issue by matching features between adjacent views, followed by dense depth estimation and point cloud fusion. Recent methods tackle this problem more effectively through volume rendering. By learning neural representations, either implicit or explicit ones, like NeRF [31] and 3D Gaussians [19], we can conduct volume rendering to rendered these neural representations into images. The rendering results are then supervised by ground truth ones to optimize the neural representations accordingly. Although these methods are capable of generating plausible 3D meshes or novel views [9, 35, 48], they struggle to recover fine-grained geometric details. This limitation arises since that the photometric consistency from color images can not ensure perfect geometric clues, which is further complicated by the shape-radiance ambiguity [66].

To overcome these obstacles, recent solutions typically incorporate monocular priors as additional supervision, such as monocular depths [43, 63, 72] and normals [6, 29, 47]. However, the effectiveness of monocular priors becomes a bottleneck hindering the performance of these methods, primarily due to two factors. One is that the predictions from monocular priors are not perfectly accurate due to domain gaps. The other is that the monocular priors are inferred independently from each RGB image, leading to geometry inconsistency across different viewpoints. MVS-based methods [3, 18, 50] mitigate these issues by deriving the uncertainty through comparing the predicted depths with the projected ones from adjacent views, which is puzzled by view occlusions. While the latest methods [4, 56] incorporate an additional branch within the rendering framework to predict the uncertainty. However, the uncertainty prediction module in these methods is coupled with the rendering branch, and thus its performance is disturbed by the quality of rendering.

To resolve these issues, we introduce MonoInstance to enhance monocular priors for neural rendering frameworks by exploring the inconsistency among each instance depths in monocular cues. Our insight builds on the fact that within the same scene, the monocular priors in 3D space will produce depth inconsistency on different views. Hence, when we back-project the depths of the same object from different views into world coordinate system, we can estimate the uncertainty of a 3D point according to the point density in the neighborhood. Specifically, we first segment multi-view images into consistent instances. For each segmented instance, we then back-project and align the multi-view estimated depth values together to create a noisy point cloud. We then evaluate the density of back-projected depth points from each viewpoint within the fused point cloud as the uncertainty measurement, leading to an uncertainty map on each view to highlight the uncertainty area of the instance. For high-uncertainty regions where the priors do not work well, we introduce an additional constraint term, guide the ray sampling, and reduce the weights for inaccurate supervision to infer the geometry and improve rendering details.

We evaluate MonoInstance upon the state-of-the-art neural representation learning frameworks in dense-view reconstruction, sparse-view reconstruction and novel view synthesis from sparse views under the widely used benchmarks. Experimental results show that our method achieves the state-of-the-art performance in various tasks. Our contributions are listed below.

- We introduce MonoInstance, which detects uncertainty in 3D according to inconsistent clues from monocular priors on multi-view. Our method is a general strategy to enhance monocular priors for various multi-view neural rendering and reconstruction frameworks.
- Based on the uncertainty maps, we introduce novel strategies to reduce the negative impact brought by inconsistent monocular clues and mine more reliable supervision through photometric consistency.
- We show our superiority over the state-of-the-art methods using multi-view neural rendering in 3D reconstruction and novel view synthesis on the widely used benchmarks.

2. Related Work

2.1. Neural 3D Reconstruction with Radiance Fields

Neural Radiance Fields (NeRF) have been a universal technique for multi-view 3D reconstruction. Notable efforts [20, 34, 48] achieve differentiable rendering of neural implicit functions, such as signed distance function [51, 68] and occupancy [15, 34], to infer neural implicit surfaces. Recent approaches introduce various priors as additional supervisions to improve the reconstruction in texture-less areas, such as monocular depth [22, 56, 63], normals [25, 29, 47], semantic segmentations [36, 70]. More recent methods improve the monocular cues by detecting uncertainties through multi-view projection of depths and normals [47, 54], but the projections suffer from view occlusions. Latest methods [4, 45, 56] integrate uncertainty estimation within the neural rendering framework, yet the predicted uncertainties are compromised by the rendering quality, especially in complex structures where RGB rendering fails. Moreover, these techniques are specifically designed for indoor scene reconstruction and not applicable across different multi-view neural rendering frameworks. Since there are often only few available views in real-world scenes, some methods are developed for sparse view reconstruction. These methods either are pre-trained on large-scale datasets and finetuned on test scenes [24, 26, 32, 40, 49], or leverage monocular priors and cross-view features to overfit a single scene [16, 55].

2.2. Novel View Synthesis with Gaussian Splatting

Recently, 3D Gaussian Splatting [19] has become a new paradigm in neural rendering due to its fast rendering speed and outstanding rendering performance. Despite high-quality rendering [27, 52], 3DGS shows poor performance when the number of input views is reduced, due to the overfitted distribution of Gaussians. Recent methods [21, 65, 72] tackle this problem by imposing monocular depth priors. However, the depth priors from pre-trained models often contain significant errors and cannot optimally position the Gaussians. Although monocular depth cues have been widely adopted in multi-view neural rendering and reconstruction frameworks, the uncertainty in depth priors has not been fully explored. To this end, we propose MonoInstance, a universal depth prior enhancement strategy that can seamlessly integrate with various multi-view neural rendering and reconstruction frameworks to improve their performances.

3. Method

Given a set of posed images $\{I_j\}_{j=1}^N$ and the corresponding monocular depth maps $\{D_j\}_{j=1}^N$, we aim to estimate N uncertainty maps $\{U_j\}_{j=1}^N$ according to the inconsistency of monocular depth cues on multi-view images. These uncertainty maps work with our novel strategies to enhance the monocular cues in various neural rendering frameworks to improve the rendering performance and reconstruction accuracy. To achieve this, we introduce a novel scheme to evaluate the uncertainty of 3D points by measuring the point density in a neighborhood. Our novel strategy will use these estimated uncertainty maps to guide the ray sampling, reduce the negative impact brought by the inconsistency, and mine more reliable photometric consistency as a remedy, which thereby enables our method to consistently improve the performance in different neural rendering tasks. An overview of our method is shown in Fig. 1, where we use NeRF-based 3D reconstruction pipeline as an example. The implementation differences when applied to 3DGS can be found in Section 4.3 and the supplementary materials.

3.1. Preliminary

Neural Radiance Fields (NeRF) [31] and 3D Gaussian Splatting (3DGS) [19] have become paradigms for learning 3D representations from multi-view images. By learning a mapping from 3D positions to densities, NeRF is able to render novel views from given viewpoints using volume rendering,

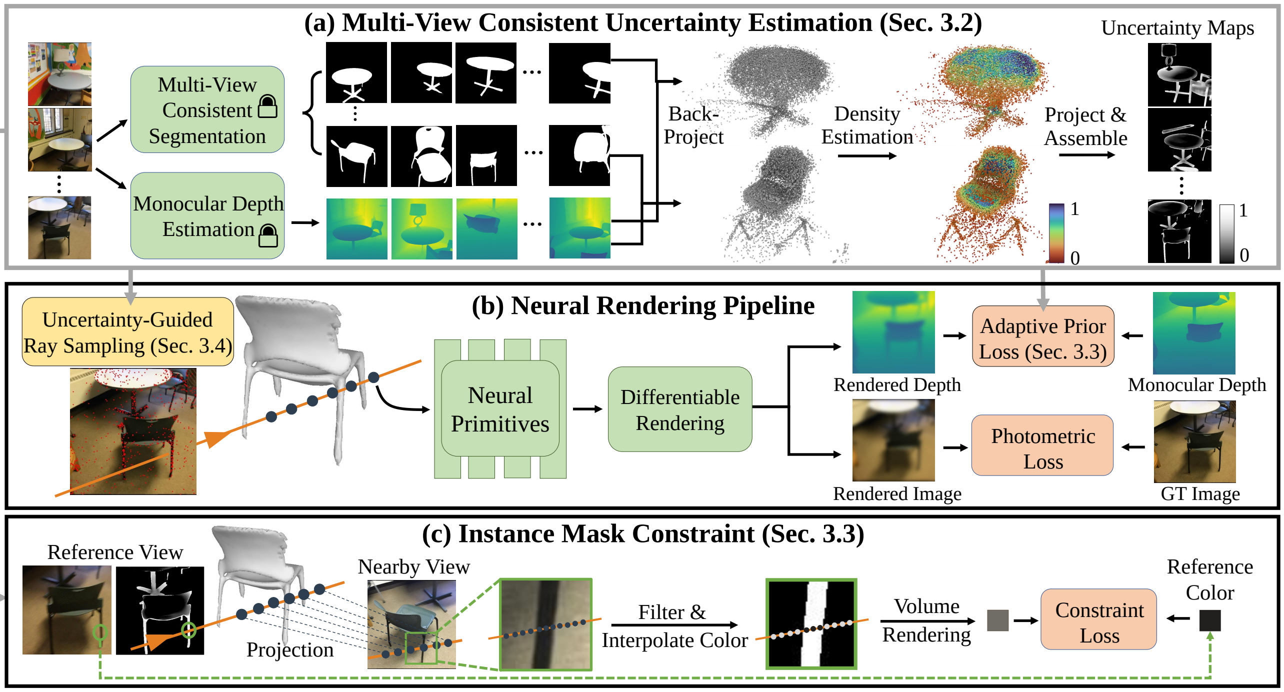

Figure 1. Overview of our method. We take multi-view 3D reconstruction through NeRF based rendering as an example. **(a)** Starting from multi-view consistent instance segmentation and estimated monocular depths, we align the same instance from different viewpoints by back-projecting instance depths into a point cloud. The monocular inconsistent clues across different views become a measurement of density estimation in neighborhood of each point, leading to uncertainty maps (Sec. 3.2). The estimated uncertainty maps are further utilized in **(b)** neural rendering pipeline to guide adaptive depth loss, ray sampling (Sec. 3.4) and **(c)** instance mask constraints (Sec. 3.3).

$$\hat{C}(r) = \sum_{i=1}^{M} \alpha_i T_i c_i, \alpha_i = 1 - \exp(-\sigma_i \delta_i), T_i = \prod_{k=1}^{i-1} (1 - \alpha_k), \tag{1}$$

where $\sigma_i, \delta_i, \alpha_i, c_i$ are the density, sampling interval, opacity and accumulated transmittance at i-th sampled point respectively and $\hat{C}(r)$ is the synthesized color of the ray r. We can also render depth or normal images in a similar way by accumulating the depth or gradient instead of color,

$$\hat{D}(r) = \sum_{i=1}^{M} \alpha_i T_i t_i, \hat{N}(r) = \sum_{i=1}^{M} \alpha_i T_i n_i, \tag{2}$$

where t_i, n_i are the sampling distance and gradient of the i-th sampled point, respectively. Recent methods extract plausible surfaces from radiance fields by modeling a relationship between SDF and volume density,

$$\sigma(s_i) = \begin{cases} \frac{1}{2\beta} \exp(\frac{-s_i}{\beta}) & \text{if } s_i \leq 0 \\ \frac{1}{\beta} - \frac{1}{2\beta} \exp(\frac{s_i}{\beta}) & \text{if } s_i > 0 \end{cases}, \tag{3}$$

where β is a learnable variance parameter and $s_i = \text{SDF}(x_i)$ is the inferred SDF of the sampled point x_i.

Similarly, 3DGS learns 3D Gaussians via differentiable volume rendering for scene modeling,

$$\hat{C}(u,v) = \sum_{i=1}^{M} c_i * o_i * p_i(u,v) \prod_{k=1}^{i-1} (1 - o_k * p_k(u,v)), \tag{4}$$

where $\hat{C}(u,v)$ is the rendered color at the pixel (u,v), $p_i(u,v), c_i, o_i$ denote the Gaussian probability, the color and the opacity of the i-th Gaussian projected onto the pixel (u,v), respectively. The neural primitives such as radiance fields and 3D Gaussians can be optimized by minimizing the rendered color and the GT color,

$$\mathcal{L}_{color} = \sum_{r \in \mathcal{R}} \|\hat{C}(r) - C(r)\|_1. \tag{5}$$

3.2. Uncertainty Estimation from Multi-View Inconsistent Monocular Prior

Monocular depth priors have been widely adopted in neural rendering and reconstruction frameworks. However, under the setting of multi-view, the priors struggle to produce consistent results within the same structures from different viewpoints due to the inherent inaccuracy, which makes the optimization even more complex. This issue inspires us to delve into the monocular uncertainty of scene structures from multi-view to provide a more robust prior for neural rendering. To this end, we introduce a novel manner to evaluate uncertainty by point density in a neighborhood after aligning multi-view instances in a unified 3D space.

Multi-view consistent segmentation. We first aim to segment every object in the scene to evaluate the uncertainty individually. The reason why we evaluate uncertainty at

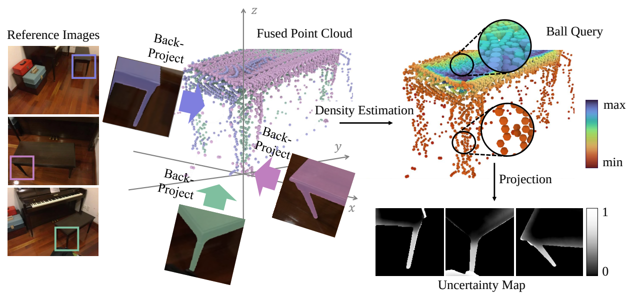

Figure 2. The illustration of uncertainty estimation. Areas with inconsistent depths (chair legs) correspond to more dispersed point cloud areas with low density (few points) in a neighborhood, indicating high uncertainty. In contrast, areas with accurate depths (chair seats) correspond to the points that are densely distributed on the true surface, indicating low uncertainty.

instance object level is to avoid the impact of object scale on density estimation. Inspired by MaskClustering [57], we achieve a consistent segmentation across multi-view through a graph-based clustering algorithm. Specifically, we firstly obtain instance segmentation on each image using [38], and then, we connect pairs of instances from different views with an edge to form a graph, if the back-projected depth point clouds of the two instances are close enough in terms of Chamfer Distance. Graph clustering algorithm [41] is then applied to partition the graph nodes into instance clusters. For indoor scenes, based on the assumption that monocular priors in textureless areas are often reliable [47, 63], we filter out the background instances and set the uncertainty of the them as zero, using GroundedSAM [39] as an identification tool. More implementation details can be found in the supplementary materials.

Uncertainty Estimation. Based on the observation that consistent depth will assemble back-projected points from different views tighter, leading to more certain points, we use the point density in a 3D neighborhood as the uncertainty. This is also a classic idea in point cloud denoising [28, 64]. To this end, we first back-project the monocular depths of each segmented instance from multi-view into world coordinate 3D space to form a point cloud, where the monocular depths are pre-aligned with the rendering depths through scale-shift invariant affine [63]. We observe that the accurate depth points consistently fall on the surface of the instance. In contrast, the noisy points coming from inaccurate predictions are independently distributed along various viewing directions towards the object, thus exhibiting anisotropic distributions with large variance, as illustrated in Fig. 2.

To further evaluate the density, we first downsample the fused point cloud to a fixed number (30,000 in our experiments) to decouple the relationship between the number of the points and the viewpoints. For the segmentation of the instance in each frame, we then back-project the masked monocular depth into 3D points and use ball query [37] to calculate the density of each point in small neighborhood, as

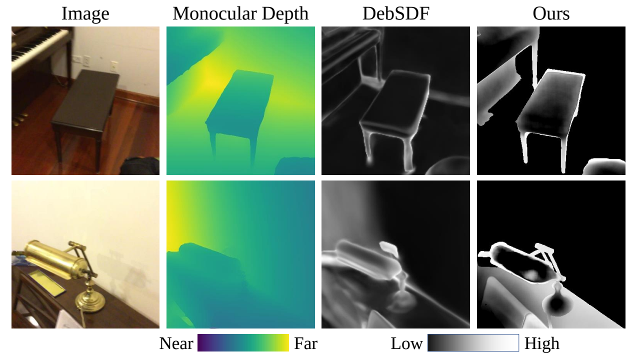

Figure 3. Visual comparison of the estimated uncertainty maps between DebSDF and ours. Our method estimates sharp uncertainty maps that faithfully capture the fine-grained geometric structures.

shown in Fig. 2. The radius for ball query is defined as

$$r = \text{Vol}(B_{opt}(P)) + 0.01, \tag{6}$$

where P is the downsampled fused point cloud, $B_{opt}(P)$ denotes the minimum oriented bounding box of P [1] and Vol denotes the volume of the bounding box. The densities are then normalized across all query points in all frames,

$$d(p(u,v)) = \frac{d(p(u,v))}{\max_{(u,v)\in\mathcal{S}_i} d(p(u,v))}, \tag{7}$$

where $p(u,v)$ is the back-projected 3D point of pixel (u,v), $d(p(u,v))$ is the measured density of that point and $\mathcal{S}_i$ is the segmented pixel area in the i-th image. The normalized densities are back-projected onto the image to obtain the per-pixel uncertainty estimation on the instance,

$$U_i(u,v) = 1 - d(p(u,v)), \tag{8}$$

where $U_i(u,v)$ denotes the uncertainty at the pixel (u,v) of the i-th image. We sequentially estimate the uncertainty for each instance in multi-view, thereby assembling complete uncertainty maps for all views.

3.3. Adaptive Prior Loss and Uncertainty-Based Mask Constraint

With the estimated uncertainty, we aim to reduce the negative impact of the inconsistency from the monocular clues and mine more reliable photo consistency as a remedy. First, we employ the estimated uncertainty maps as weights on the difference between monocular depths and the rendering ones, which filter out the impact brought by inaccurate supervision. This leads to an adaptive prior loss, as shown in Fig. 1.

However, the regions of high-uncertainty, which often contain complex structures, are not effectively recovered by relying solely on photometric loss. To facilitate the learning of these areas, we further introduce an uncertainty-based instance mask constraint, enforcing the alignment of the learned instances within multi-view segmentation, as illustrated in Fig. 1. Specifically, inspired by Pixel Warping [7], for a ray emitted from a high-uncertainty instance region S_r^i in the reference view I_r, we project points $\{p_j\}_{j=1}^K$ sampled on the ray into a nearby view I_n, and filter out the projected points $\{\pi_n(p_j)\}_{j=1}^K$ which fall within the instance mask S_n^i in I_n. We then use the interpolated colors of these filtered projected points on I_n and the corresponding predicted opacities α_j to render the final color,

$$\hat{C}_n^{sil} = \sum_{j=1}^{K} \mathbb{1}_j \cdot I_n[\pi_n(p_j)]\alpha_j \prod_{l<j}(1-\alpha_l),$$
$$\mathbb{1}_j = \begin{cases} 1 & \pi_n(p_j) \in S_n^i \\ 0 & \pi_n(p_j) \notin S_n^i \end{cases}. \quad (9)$$

The rendered color $\hat{C}_n^{sil}$ is compared with the corresponding ground truth color in I_r as additional supervision. Unlike Pixel Warping [7], we discriminately accumulate the projected points that just fall within the instance mask in the nearby view, because we are prompted of which sampling points contribute to the rendering of this instance through multi-view segmentation. This enables us to implicitly constrain these sampling points to align with the object surfaces.

3.4. Optimization

Uncertainty-Guided Ray Sampling. We use the estimated uncertainty maps as probabilities to guide the ray sampling, paying more attention to regions with high uncertainty. We first allocate the number of sampling pixels for each instance according to its area in the segmentation. And then we calculate the sampling probabilities according to uncertainty. The probability in i-th view is defined as $prob_i(u,v) = U_i(u,v) + 0.05$, where the additional 0.05 ensures that the sampling is not omitted in areas with zero uncertainty.

Training. Our training process is divided into two stages. In the first stage, we uniformly apply monocular depth priors to learn a coarse representation of the scene. We then render low-resolution depth maps from all viewpoints to align the multi-view monocular depths to the same scale. Subsequently, we evaluate multi-view uncertainty for every segmented instance and assemble them to uncertainty maps of all frames. In the second stage, we integrate the uncertainty maps into the training process to utilize guided ray sampling, adaptive depth loss and instance mask constraints.

Loss Function. The overall loss function is defined as

$$\mathcal{L} = \mathcal{L}_{color} + \lambda_1\mathcal{L}_{eik} + \lambda_2\mathcal{L}_{sil} + \lambda_3\mathcal{L}_d + \lambda_4\mathcal{L}_n, \quad (10)$$

where $\mathcal{L}_{eik}$ is the Eikonal term [60], $\mathcal{L}_{sil}$ is the instance mask constraint introduced in Sec. 3.3, $\mathcal{L}_d$ is the adaptive depth loss and $\mathcal{L}_n$ is an optional adaptive normal loss. λ_{1-4} are hyper-parameters for weighting each term.

4. Experiments

To evaluate the effectiveness of our method, we conduct experiments based on various neural representation learning frameworks using multi-view images, including dense-view 3D reconstruction, sparse-view 3D reconstruction and sparse view synthesis.

4.1. Dense-view 3D Reconstruction

Datasets. We evaluate our performance under two real-world indoor scene datasets, including ScanNet [5] and Replica [44]. We select 4 scenes from ScanNet and all 8 scenes from Replica, following baseline settings [56, 63]. Each scene consists of various numbers of observations from dense viewpoints, ranging from 200 to 400.

Baselines and metrics. We compare our method with the latest indoor scene reconstruction methods including MonoSDF [63], SDF-OCC-Hybrid [29] (shorted for "HybridNeRF"), H2O-SDF [36], DebSDF [56], RS-Recon [61]. Note that the source code of H2O-SDF has not been made publicly available, thus we are unable to obtain its results on Replica dataset. Following baselines [61, 63], we report Chamfer Distance (CD), F-score in ScanNet dataset and additional Normal Consistency (N.C.) in Replica dataset.

Implementation details. We build our code upon the source code of MonoSDF [63]. The hyper-parameters in Eq. (9) are set as $\lambda_1 = 0.1, \lambda_2 = 0.4, \lambda_3 = 0.5, \lambda_4 = 0.05$. Since the monocular normals are homologous with depths which come from the same foundation model, they show similar performances in the same regions of the images. Therefore, we can uniformly utilize the estimated uncertainty map to depth and normal priors. The nearby views used in Sec. 3.3 are selected according to the difference between observation angles. More implementation details are discussed in the supplementary materials.

Comparisons. We report numerical comparisons on ScanNet and Replica datasets in Tab. 1. Our method outperforms all baseline methods on ScanNet dataset and achieves the highest normal consistency on Replica dataset. Visual comparisons in Fig. 4 show that our method is capable of reconstructing fine-grained details of the scene, especially in the small thin structures such as the lamp on the piano, the flowers on the tea table and the chair legs.

4.2. Sparse-view 3D Reconstruction

Datasets. We further evaluate our method in reconstructing 3D shapes from sparse observations on DTU dataset [17]. Following previous methods [16, 62], we report our results on the widely used 15 scenes, each of which shows single object with background from 3 viewpoints with small overlapping.

Baselines and metrics. We compare our method with the latest sparse-view reconstruction approaches including the traditional MVS methods such as COLMAP [42], overfitting-based methods such as NeuSurf [16], generalizing-finetuning methods such as SparseNeuS [26], VolRecon [40], ReTR [24] and UFORecon [32]. We use Chamfer Dis-

Table 1. Averaged dense-view 3D reconstruction metrics on ScanNet and Replica datasets.

Methods	ScanNet					Replica				
	Acc↓	Comp↓	Prec↑	Recall↑	F-score↑	Acc↓	Comp↓	CD↓	N.C.↑	F-score↑
UNISURF [34]	0.554	0.164	0.212	0.362	0.267	0.045	0.053	0.049	0.909	0.789
MonoSDF [63]	0.035	0.048	0.799	0.681	0.733	0.027	0.031	0.029	0.921	0.861
HybridNeRF [29]	0.039	0.041	0.800	0.760	0.779	0.025	0.027	**0.026**	0.934	**0.921**
H2O-SDF [36]	**0.032**	0.037	0.834	0.769	0.799	-	-	-	-	-
DebSDF [56]	0.036	0.040	0.807	0.765	0.785	0.028	0.030	0.029	0.932	0.883
RS-Recon [61]	0.040	0.040	0.809	0.779	0.794	0.027	**0.025**	**0.026**	0.934	0.917
Ours	0.035	**0.032**	**0.846**	**0.824**	**0.834**	**0.024**	0.029	**0.026**	**0.937**	0.918

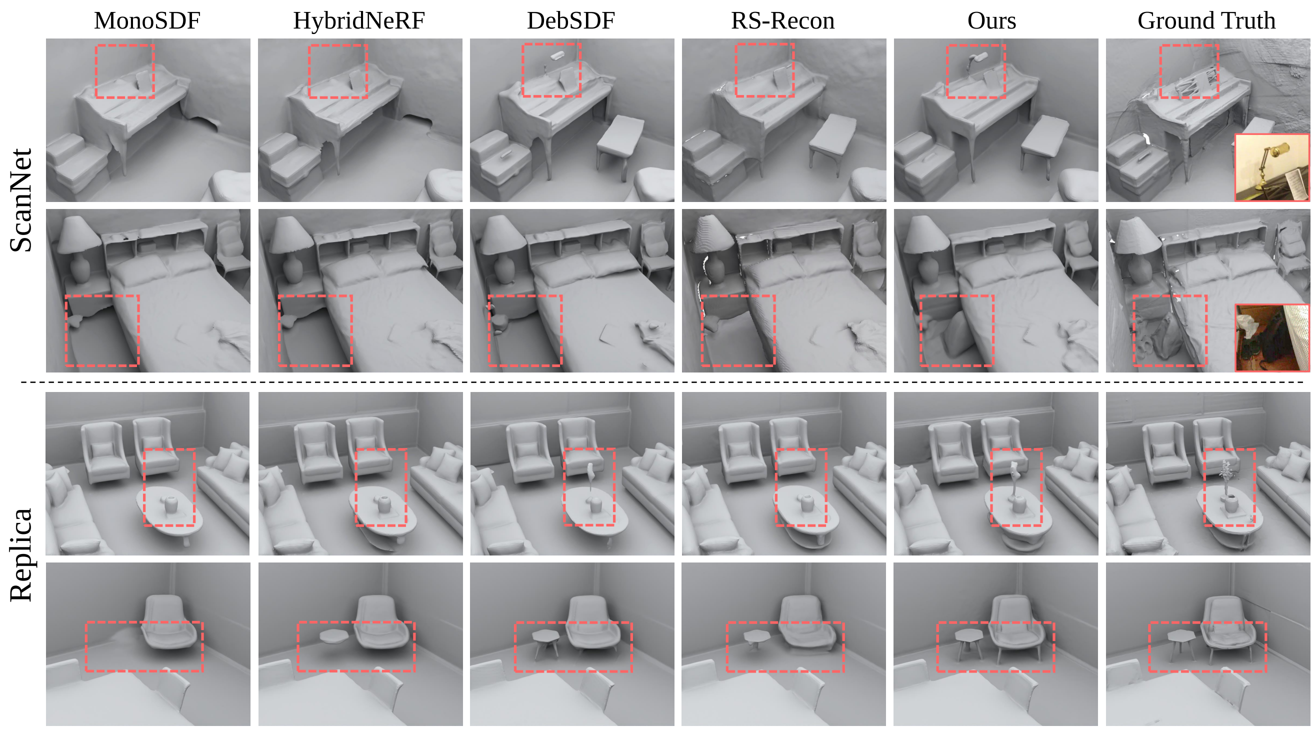

Figure 4. Visual comparisons of dense-view 3D reconstruction on ScanNet and Replica dataset.

tance (CD) between the reconstructed meshes and the real-scanned point clouds as the evaluation metrics, following baselines [16].

Implementation details. We use the official code released by NeuSurf [16] to produce our results of sparse-view reconstruction. The hyper-parameters in Eq. (9) are consistent with those employed in dense-view reconstruction. Since the multi-view images in each DTU scene capture the unique object, there is no need to conduct additional multi-view consistent instance segmentation. In our implementation, we first segment the scene into the object and the background, and then align and compute the uncertainty map only for the center object from various viewpoints.

Comparisons. We report numerical evaluations on DTU dataset in Tab. 2. For fair comparison, we also report the results of NeuSurf with monocular cues (NeuSurf†), which are uniformly applied to all pixels, similar to MonoSDF [63]. The superiority results in terms of CD show the effectiveness of our method. Further comparison between NeuSurf and NeuSurf† reveals that indiscriminately applying monocular depths to all pixels does not significantly improve the performance of NeuSurf. While our method leverages the estimated uncertainty maps to enhance the learning of the high-uncertainty regions, avoiding the misguidance from the inaccurate monocular priors. We showcase our improvements in visual comparison in Fig. 5, where our method consistently produces more complete and smoother surfaces compared to baseline methods.

4.3. Sparse Novel View Synthesis

Datasets. We further evaluate our method on 3DGS-based sparse-input novel view synthesis (NVS) task on LLFF dataset [30]. It contains 8 forward-facing real-world scenes. We select 3 views and downscale their resolutions as 8 to

CVPR
#12070

Table 2. Averaged Chamfer Distance (CD) over the 15 scenes on DTU dataset in reconstructions from sparse views (small overlaps). NeuSurf† means NeuSurf with additional monocular cues.

Methods	COLMAP [42]	SparseNeuS$_{ft}$ [26]	VolRecon [40]	ReTR [24]	NeuSurf [16]	NeuSurf† [16]	UFORecon [32]	Ours
CD ↓	2.61	3.34	3.02	2.65	1.35	1.30	1.43	**1.18**

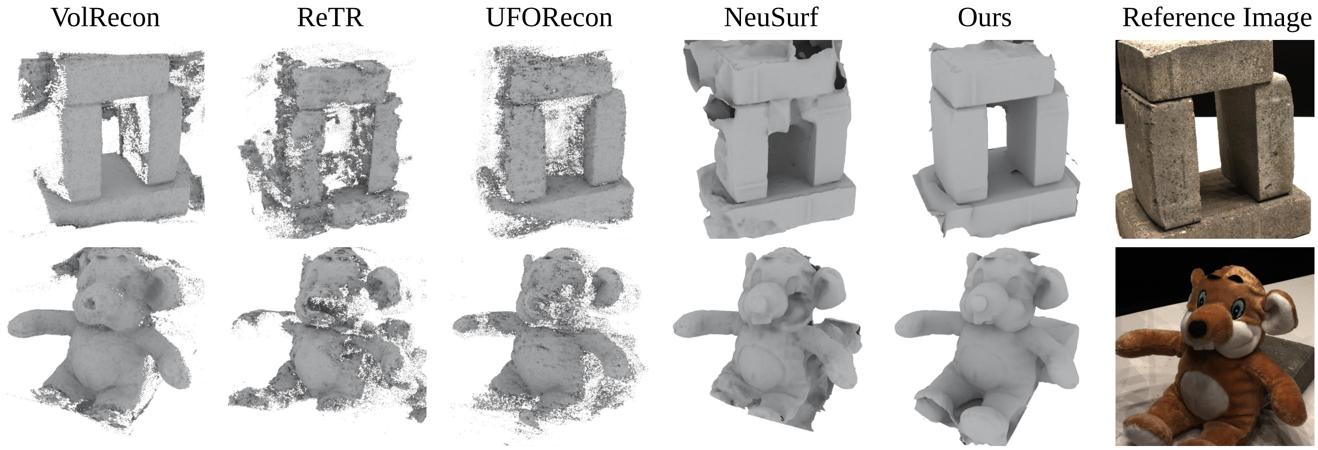

Figure 5. Visual comparisons on DTU dataset under the task of little-overlapping sparse input reconstruction.

Table 3. Quantitative comparison on LLFF dataset in novel view synthesis from sparse views.

Methods	PSNR↑	SSIM↑	LPIPS↓
RegNeRF [33]	19.08	0.587	0.336
FreeNeRF [58]	19.08	0.587	0.336
3DGS [19]	15.52	0.405	0.408
DNGaussian [21]	19.12	0.591	0.294
FSGS [72]	20.31	0.652	0.288
COR-GS [65]	20.45	0.712	0.196
Ours	**20.73**	**0.731**	**0.184**

train, following previous works [33, 72].

Baselines and metrics. We compare our method with latest few-shot NVS methods, including NeRF-based methods, such as RegNeRF [33], FreeNeRF [58] and 3DGS-based methods, such as DNGaussian [21], FSGS [72] and COR-GS [65]. We report PSNR, SSIM [53] and LPIPS scores [67] to evaluate the rendering quality following previous works [46, 72].

Implementation details. Our code in this experiment is built upon FSGS [72], which utilizes monocular depths to enhance the rendering. $\mathcal{L}_{eik}$ and $\mathcal{L}_n$ in Eq. (9) are omitted in our experiment because there is no gradient fields in Gaussian splatting, and the orientation of 3D Gaussians are ambiguous during splatting [11, 12]. Note that 3D Gaussians are directly splatted onto the image plane with no sampled points in the space, thus we design a variant of our instance mask constraint, which encourages the projected instance depth points on the nearby view to move towards the mask of the same instance in nearby view, similar as [13].

Comparisons. The numerical and visual comparison are shown in Tab. 3 and Fig. 6. The visualizations of rendered images and depths further demonstrate our advanced results in recovering complex object details. We further visualize our uncertainty maps across different datasets in Fig. 7. Comparisons among the GT images, monocular depths, and the final results show that our method adaptively captures the inaccuracies in monocular depths, thereby achieving superior results beyond the quality of the priors.

Table 4. Ablation study of each module on ScanNet dataset. Starting from the base model, we progressively add each of our module to reveal the impact of the proposed modules.

	Acc↓	Comp↓	F-score↑
Base	0.039	0.042	0.749
+Mono Uncertainty	0.036	0.039	0.786
+Adaptive Sampling	0.036	0.035	0.805
+Mask Constraint (Full)	**0.035**	**0.032**	**0.834**

Table 5. Ablation study of different monocular priors. The results are averaged F-score across the four ScanNet scenes.

Methods	Omnidata [8]	Metric3D v2 [14]	GeoWizard [10]
MonoSDF	0.733	0.749	0.741
Ours	0.825	0.834	0.829

4.4. Ablation Study

Effectiveness of each module. We conduct ablation studies to justify the effectiveness of the modules in our method on ScanNet dataset. Starting from the base model, which is identical to MonoSDF [63], we progressively add each of our modules to show the improvements of the reconstructed results. These additions include the adaptive monocular prior supervision, the uncertainty-guided ray sampling and the uncertainty-based instance mask constraint, as reported

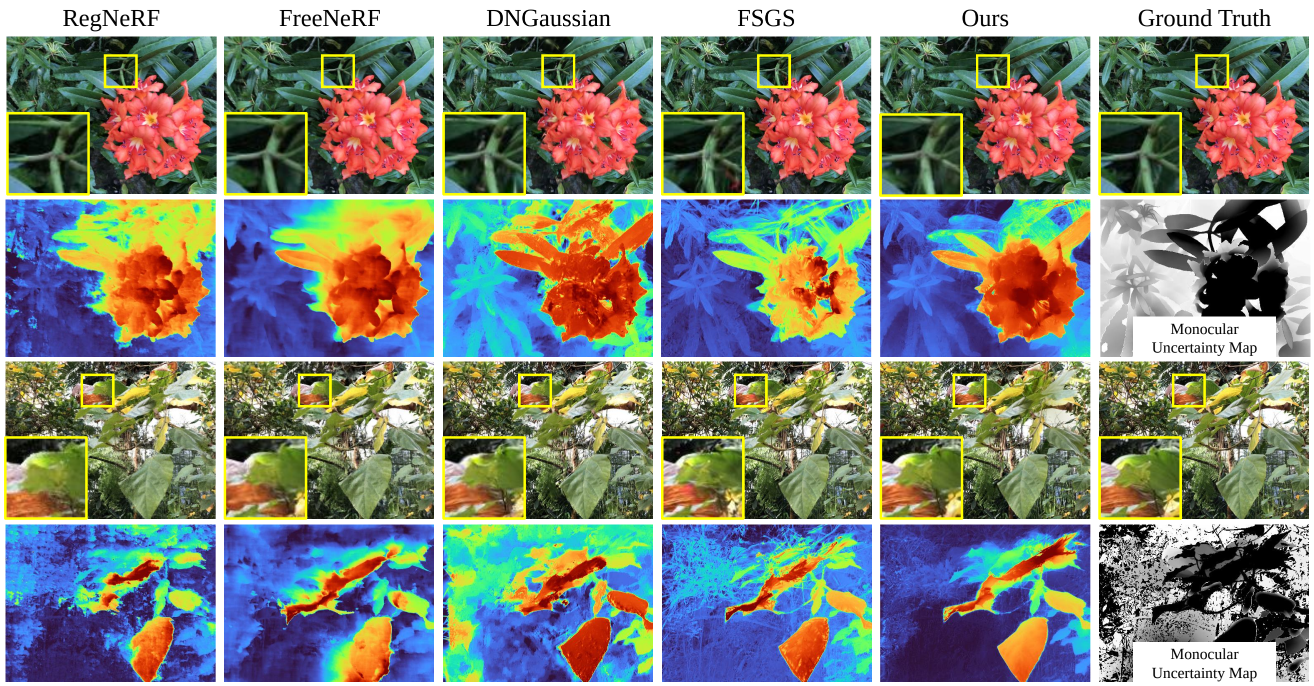

Figure 6. Visual comparisons on LLFF dataset in novel view synthesis from sparse views. In the uncertainty maps, areas that are more white indicate higher uncertainty.

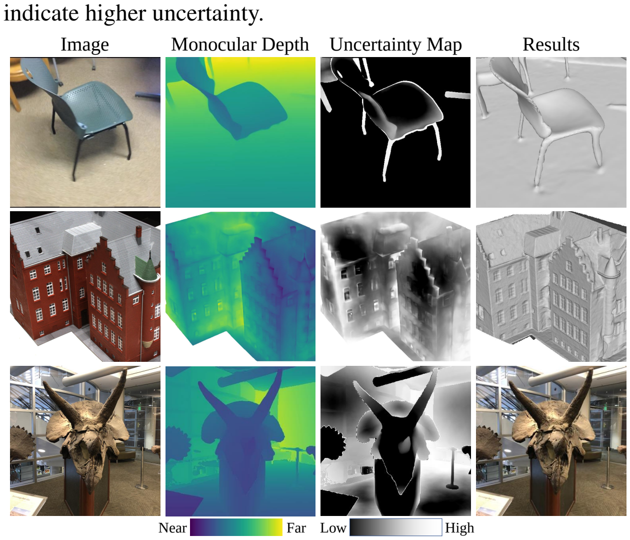

Figure 7. Visualization of our uncertainty maps calculated from monocular depths. Our uncertainties effectively identify the inconsistency across monocular clues on multi-view.

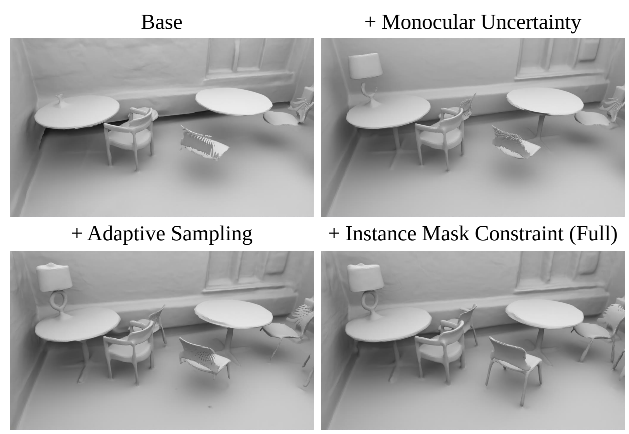

Figure 8. Visualization of ablations on each of our module.

in Tab. 4. The visual comparisons in Fig. 8 indicate that our method, equipped with each proposed module, successfully recovers complete and detailed geometric structures.

Choice of monocular priors. We further evaluate the performance of our method with different prior estimation models, including Omnidata [8], Metric3D v2 [14] and GeoWizard [10]. The improvement of our method beyond MonoSDF [63] indicates that our method consistently enhances the monocular priors obtained from various estimation models. To fully reveal the potential of our approach, we choose Metric3D v2 as our primary prior model.

5. Conclusion

We propose MonoInstance, a novel approach to enhance monocular priors to provide robust monocular cues for multi-view neural rendering frameworks. To this end, we estimate the uncertainty of monocular priors by aligning multi-view instance depths in a unified 3D space and detecting the densities in point clouds. The estimated uncertainty maps can be further utilized in adaptive prior loss, uncertainty-guided ray sampling and instance mask constraint. Our approach can be applied upon different multi-view neural rendering and reconstruction methods to enhance the monocular priors for better neural representation learning. The visual and numerical comparisons with the state-of-the-art methods justify our effectiveness and show our superiority over the latest methods.

References

[1] Gill Barequet and Sariel Har-Peled. Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. *Journal of Algorithms*, 38(1):91–109, 2001. 4

[2] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. *Proceedings of International Conference on Computer Vision (ICCV)*, 2021. 1

[3] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. MVSNeRF: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14124–14133, 2021. 1

[4] Ziyi Chen, Xiaolong Wu, and Yu Zhang. NC-SDF: Enhancing indoor scene reconstruction using neural sdfs with view-dependent normal compensation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5155–5165, 2024. 1, 2

[5] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. 5

[6] Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, and Weiwei Xu. High-quality surface reconstruction using gaussian surfels. In *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery, 2024. 1

[7] François Darmon, Bénédicte Bascle, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. Improving neural implicit surfaces geometry with patch warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6260–6269, 2022. 4, 5

[8] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3D scans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10786–10796, 2021. 7, 8

[9] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-Planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023. 1

[10] Xiao Fu, Wei Yin, Mu Hu, Kaixuan Wang, Yuexin Ma, Ping Tan, Shaojie Shen, Dahua Lin, and Xiaoxiao Long. GeoWizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. In *European Conference on Computer Vision*, pages 241–258. Springer, 2025. 7, 8

[11] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3D Gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. *arXiv preprint arXiv:2311.16043*, 2023. 7

[12] Antoine Guédon and Vincent Lepetit. SuGaR: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 7

[13] Zhizhong Han, Chao Chen, Yu-Shen Liu, and Matthias Zwicker. DRWR: A differentiable renderer without rendering for unsupervised 3D structure learning from silhouette images. In *International Conference on Machine Learning*, 2020. 7

[14] Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3D v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 7, 8

[15] Pengchong Hu and Zhizhong Han. Learning neural implicit through volume rendering with attentive depth fusion priors. In *Advances in Neural Information Processing Systems*, 2023. 2

[16] Han Huang, Yulun Wu, Junsheng Zhou, Ge Gao, Ming Gu, and Yu-Shen Liu. NeuSurf: On-surface priors for neural surface reconstruction from sparse input views. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2312–2320, 2024. 2, 5, 6, 7

[17] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 406–413, 2014. 5

[18] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Uncertainty-aware deep multi-view photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12601–12611, 2022. 1

[19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023. 1, 2, 7

[20] J.J. Leonard and H.F. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *IEEE/RSJ International Workshop on Intelligent Robots and Systems*, pages 1442–1447, 1991. 2

[21] Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. DNGaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20775–20785, 2024. 2, 7

[22] Zizhang Li, Xiaoyang Lyu, Yuanyuan Ding, Mengmeng Wang, Yiyi Liao, and Yong Liu. RICO: Regularizing the unobservable for indoor compositional reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17761–17771, 2023. 2

[23] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8456–8465, 2023. 1

[24] Yixun Liang, Hao He, and Yingcong Chen. ReTR: Modeling rendering via transformer for generalizable neural surface reconstruction. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 5, 7

[25] Zhihao Liang, Zhangjin Huang, Changxing Ding, and Kui Jia. HelixSurf: A robust and efficient neural implicit surface learning of indoor scenes with iterative intertwined regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13165–13174, 2023. 2

[26] Xiaoxiao Long, Cheng Lin, Peng Wang, Taku Komura, and Wenping Wang. SparseNeuS: Fast generalizable neural surface reconstruction from sparse views. In *European Conference on Computer Vision*, pages 210–227. Springer, 2022. 2, 5, 7

[27] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-GS: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20654–20664, 2024. 2

[28] Shitong Luo and Wei Hu. Score-based point cloud denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4583–4592, 2021. 4

[29] Xiaoyang Lyu, Peng Dai, Zizhang Li, Dongyu Yan, Yi Lin, Yifan Peng, and Xiaojuan Qi. Learning a Room with the OCC-SDF Hybrid: Signed distance function mingled with occupancy aids scene representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8940–8950, 2023. 1, 2, 5, 6

[30] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local Light Field Fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (ToG)*, 38(4):1–14, 2019. 6

[31] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, pages 405–421. Springer, 2020. 1, 2

[32] Youngju Na, Woo Jae Kim, Kyu Beom Han, Suhyeon Ha, and Sung-Eui Yoon. UFORecon: Generalizable sparse-view surface reconstruction from arbitrary and unfavorable sets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5094–5104, 2024. 2, 5, 7

[33] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. RegNeRF: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5480–5490, 2022. 7

[34] Michael Oechsle, Songyou Peng, and Andreas Geiger. UNISURF: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5589–5599, 2021. 2, 6

[35] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. *IEEE International Conference on Computer Vision*, 2021. 1

[36] Minyoung Park, Mirae Do, Yeon Jae Shin, Jaeseok Yoo, Jongkwang Hong, Joongrock Kim, and Chul Lee. H2O-SDF: Two-phase learning for 3d indoor reconstruction using object surface fields. In *The Twelfth International Conference on Learning Representations*, 2023. 2, 5, 6

[37] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 4

[38] Lu Qi, Jason Kuen, Tiancheng Shen, Jiuxiang Gu, Wenbo Li, Weidong Guo, Jiaya Jia, Zhe Lin, and Ming-Hsuan Yang. High quality entity segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4047–4056, 2023. 4

[39] Tianhe Ren, Shilong Liu, Ailing Zeng, Jing Lin, Kunchang Li, He Cao, Jiayu Chen, Xinyu Huang, Yukang Chen, Feng Yan, et al. Grounded SAM: Assembling open-world models for diverse visual tasks. *arXiv preprint arXiv:2401.14159*, 2024. 4

[40] Yufan Ren, Fangjinhua Wang, Tong Zhang, Marc Pollefeys, and Sabine Süsstrunk. VolRecon: Volume rendering of signed ray distance functions for generalizable multi-view reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16685–16695, 2023. 2, 5, 7

[41] Satu Elisa Schaeffer. Graph clustering. *Computer science review*, 1(1):27–64, 2007. 4

[42] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113, 2016. 5, 7

[43] Jiuhn Song, Seonghoon Park, Honggyu An, Seokju Cho, Min-Seop Kwak, Sungjin Cho, and Seungryong Kim. Därf: Boosting radiance fields from sparse input views with monocular depth adaptation. *Advances in Neural Information Processing Systems*, 36:68458–68470, 2023. 1

[44] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, et al. The Replica Dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019. 5

[45] Ziyu Tang, Weicai Ye, Yifan Wang, Di Huang, Hujun Bao, Tong He, and Guofeng Zhang. ND-SDF: Learning normal deflection fields for high-fidelity indoor reconstruction. *arXiv preprint arXiv:2408.12598*, 2024. 2

[46] Guangcong Wang, Zhaoxi Chen, Chen Change Loy, and Ziwei Liu. SparseNeRF: Distilling depth ranking for few-shot novel view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9065–9076, 2023. 7

[47] Jiepeng Wang, Peng Wang, Xiaoxiao Long, Christian Theobalt, Taku Komura, Lingjie Liu, and Wenping Wang. NeuRIS: Neural reconstruction of indoor scenes using normal priors. In *European Conference on Computer Vision*, 2022. 1, 2, 4

[48] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 34, 2021. 1, 2

[49] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. DUSt3R: Geometric 3d vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20697–20709, 2024. 2

[50] Xiaofeng Wang, Zheng Zhu, Guan Huang, Xu Chi, Yun Ye, Ziwei Chen, and Xingang Wang. Crafting monocular cues and velocity guidance for self-supervised multi-frame depth learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2689–2697, 2023. 1

[51] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. HF-NeuS: Improved surface reconstruction using high-frequency details. *Advances in Neural Information Processing Systems*, 35:1966–1978, 2022. 2

[52] Yufei Wang, Zhihao Li, Lanqing Guo, Wenhan Yang, Alex C Kot, and Bihan Wen. ContextGS: Compact 3D Gaussian Splatting with Anchor Level Context Model. *Advances in Neural Information Processing Systems*, 2024. 2

[53] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image Quality Assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 7

[54] Yi Wei, Shaohui Liu, Yongming Rao, Wang Zhao, Jiwen Lu, and Jie Zhou. NerfingMVS: Guided optimization of neural radiance fields for indoor multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5610–5619, 2021. 2

[55] Haoyu Wu, Alexandros Graikos, and Dimitris Samaras. S-VolSDF: Sparse multi-view stereo regularization of neural implicit surfaces. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3556–3568, 2023. 2

[56] Yuting Xiao, Jingwei Xu, Zehao Yu, and Shenghua Gao. DebSDF: Delving into the details and bias of neural indoor scene reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1, 2, 5, 6

[57] Mi Yan, Jiazhao Zhang, Yan Zhu, and He Wang. MaskClustering: View consensus based mask graph clustering for open-vocabulary 3d instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28274–28284, 2024. 4

[58] Jiawei Yang, Marco Pavone, and Yue Wang. Freenerf: Improving few-shot neural rendering with free frequency regularization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8254–8263, 2023. 7

[59] Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. MVSNet: Depth inference for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, pages 767–783, 2018. 1

[60] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33, 2020. 5

[61] Ruihong Yin, Yunlu Chen, Sezer Karaoglu, and Theo Gevers. Ray-distance volume rendering for neural scene reconstruction. *European Conference on Computer Vision*, 2024. 5, 6

[62] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelNeRF: Neural radiance fields from one or few images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 5

[63] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. MonoSDF: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in neural information processing systems*, 35:25018–25032, 2022. 1, 2, 4, 5, 6, 7, 8

[64] Faisal Zaman, Ya Ping Wong, and Boon Yian Ng. Density-based denoising of point cloud. In *9th International Conference on Robotic, Vision, Signal Processing and Power Applications: Empowering Research and Innovation*, pages 287–295. Springer, 2017. 4

[65] Jiawei Zhang, Jiahe Li, Xiaohan Yu, Lei Huang, Lin Gu, Jin Zheng, and Xiao Bai. CoR-GS: sparse-view 3D Gaussian splatting via co-regularization. In *European Conference on Computer Vision*, pages 335–352. Springer, 2024. 2, 7

[66] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. NeRF++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 1

[67] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 7

[68] Yongqiang Zhang, Zhipeng Hu, Haoqian Wu, Minda Zhao, Lincheng Li, Zhengxia Zou, and Changjie Fan. Towards unbiased volume rendering of neural implicit surfaces with geometry priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4359–4368, 2023. 2

[69] Dongxu Zhao, Daniel Lichy, Pierre-Nicolas Perrin, Jan-Michael Frahm, and Soumyadip Sengupta. MVPSNet: Fast generalizable multi-view photometric stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12525–12536, 2023. 1

[70] Xiaowei Zhou, Haoyu Guo, Sida Peng, Yuxi Xiao, Haotong Lin, Qianqian Wang, Guofeng Zhang, and Hujun Bao. Neural 3d scene reconstruction with indoor planar priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2

[71] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R. Oswald, and Marc Pollefeys. NICE-SLAM: Neural implicit scalable encoding for slam. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 1

[72] Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. FSGS: Real-time few-shot view synthesis using gaussian splatting. In *European Conference on Computer Vision*, pages 145–163. Springer, 2024. 1, 2, 7