

Bayesian NeRF: Quantifying Uncertainty with Volume Density in Neural Radiance Fields

Sibeak Lee[✉], Kyeongsu Kang[✉], and Hyeonwoo Yu[✉]

Sungkyunkwan University, Suwon, South Korea

{lmjlss, thithin0821, hwyu}@skku.edu

Abstract. We present the Bayesian Neural Radiance Field (NeRF), which explicitly quantifies uncertainty in geometric volume structures without the need for additional networks, making it adept for challenging observations and uncontrolled images. NeRF diverges from traditional geometric methods by offering an enriched scene representation, rendering color and density in 3D space from various viewpoints. However, NeRF encounters limitations in relaxing uncertainties by using geometric structure information, leading to inaccuracies in interpretation under insufficient real-world observations. Recent research efforts aimed at addressing this issue have primarily relied on empirical methods or auxiliary networks. To fundamentally address this issue, we propose a series of formulation extensions to NeRF. By introducing generalized approximations and defining density-related uncertainty, our method seamlessly extends to manage uncertainty not only for RGB but also for depth, without the need for additional networks or empirical assumptions. In experiments we show that our method significantly enhances performance on RGB and depth images in the comprehensive dataset, demonstrating the reliability of the Bayesian NeRF approach to quantifying uncertainty based on the geometric structure.

The code is available at: https://github.com/Lab-of-AI-and-Robotics/Bayesian_NeRF
The video is: https://youtu.be/wp5jW4S_jqo

Keywords: Neural Radiance Fields · Bayesian Networks · Uncertainty Estimation

1 Introduction

The advent of Neural Radiance Fields (NeRF) [20] has significantly advanced the field of novel view synthesis, allowing for the continuous synthesis of views from given images at unseen positions and orientations. As opposed to traditional geometric approaches [25, 27], NeRF leverages a learned neural network model to predict color and density in 3D space from coordinates and view directions, thus better handling complex scenes, intricate textures, and lighting variations. Subsequent research has propelled its real-world application in virtual reality [5], robotics [24, 36], digital twins [16], and autonomous driving [9, 38], underscoring the importance of real-time performance and efficiency when data is limited. However, advancements are hindered by challenges like accurately predicting scenes in unobserved views with limited data availability [40]. This is inherently challenged by the limitations of the sensor’s field of view (FoV) and



Fig. 1: Example of the quantitative results. Ground Truth (Left), prediction result from the baseline model (middle), prediction result of our proposed method (Right). All methods use 4 images as the training set in the NeRF dataset, and predict unobserved viewpoint.

physical occlusions, such as buildings or other vehicles, frequently result in gaps in observable data. Moreover, the intrinsic nature of the real-world ensures that sensor data is invariably accompanied by errors [33]. To overcome these challenges, it is necessary to incorporate uncertainty consideration and this approach is crucial for precise interpretation of and adaptation to the variabilities encountered in real-world environments [1, 4, 11].

Given the importance of considering uncertainty in neural radiance fields, various studies have focused on addressing the variability of color in space. [19] tackled photometric variations to deal with images captured in various environments, while [22] used data selection strategies for enhanced learning. They incorporate an L1 regularization term to mitigate transient density issues. However, considering uncertainty in color presents several critical drawbacks. Firstly, it fails to address uncertainties in other sensor data such as density, which are essential for a comprehensive understanding of 3D spaces. Moreover, predicting pixel variance does not fully capture the uncertainty in the scene’s radiance field and can lead to suboptimal results. In response to these challenges, [28, 29] suggested employing an additional network to consider the density uncertainty. However, the use of an extra network is inefficient for real-time applications and does not guarantee accurate estimations due to the approximation of the probability distribution.

We introduce a series of methodical approaches that allow us to incorporate density-related uncertainty into the NeRF model without altering its network structure. This methodology not only explicitly addresses density uncertainty but also renders the model suitable for real-time applications by leveraging Bayesian techniques to manage and interpret the uncertainties inherent in the data. Our approaches improve rendering in unobserved views, as shown in Fig. 1, where training data from limited viewpoints leads to suboptimal performance. Incorporating a Bayesian framework enhances the model’s capacity to handle unseen data uncertainties, boosting rendering quality from various angles. Furthermore, we confront the limitations imposed by sensors such as Lidar, Thermal Imaging Cameras, and Ultrasonic Sensors, which do not capture color data, rendering traditional methods of uncertainty consideration ineffective. By focusing on density as a universal attribute detectable across different sensor types, our method extends the applicability of NeRF models beyond the realm of RGB cameras. This integration of uncertainty across various sensing modalities not only broadens the flexibility of NeRF applications but also significantly elevates their performance in real-

world scenarios, demonstrating the robustness and adaptability of our enhanced NeRF framework.

2 Related Work

2.1 Novel-view synthesis and NeRF

The field of novel-view synthesis has seen remarkable advancement as computer vision technologies have evolved. Initially, traditional methods were concentrated on understanding the structure of 3D scenes through camera poses, utilizing techniques such as structure-from-motion (SFM) [25] and Multi-View Stereo (MVS) [27]. These methods, based on geometric approaches, aimed to synthesize images from novel viewpoints by accurately modeling 3D structures. The advent of NeRF marked a significant leap forward, enabling the creation of realistic renderings from a smaller set of images through 3D representations learned by neural networks. NeRF estimates volume density and color information based on 3D coordinates and viewing directions, with its network output facilitating image generation via a rendering function. These images reflect observations within the modeled space, informed by the direction of the camera view [19, 22, 28]. Efforts have since been directed at addressing the original NeRF's limitations, spurring a variety of enhancements. For instance, changes to encoding methods have sped up the learning process [3, 8, 21], while large-scale scenes are now depicted using multiple implicit neural networks [41, 43]. Additionally, the introduction of new priors has facilitated learning from sparse data [31, 34]. NeRF's application scope has also expanded, encompassing areas such as scene editing [35, 44], converting text to 3D models [15, 23], and enhancing visual scene-based SLAM technologies [12, 24, 42, 48], demonstrating its versatility and potential in various domains.

2.2 Uncertainty estimation in NeRF

Estimating uncertainty is a well-known and important issue in deep learning. Similarly, in NeRF, considering uncertainty is crucial for reducing errors stemming from sensor noise, sensor field of view (FOV), occlusions, and other factors, enabling the representation of space with reliability and robustness. Due to this importance, research efforts are underway to integrate uncertainty estimation into existing NeRF-based techniques. For instance, [19] accurately render scenes from photos taken in various environments, [22] select data for subsequent learning, and [14] consider uncertainty in color to measure the reliability of generated images. While these approaches explicitly estimate uncertainty and do not require more extensive training than baseline models, they have the drawback of not considering uncertainty for volume density. Other recent works [6] represent rendering cleanup by measuring uncertainty in trained models. [30] proposes a method that utilizes an ensemble of multiple NeRF models to quantify uncertainty, which can be applied to next view selection. However, these approaches can only capture epistemic uncertainty and cannot capture inherent aleatoric uncertainty in the data.

Addressing volume density uncertainty, [28, 29] presents a method that employs an additional network to measure depth uncertainty. This method is difficult to apply

in real-time situations due to additional network usage. To address this challenge, we introduce a series of methods based on Bayesian approaches that explicitly estimate volume density uncertainty, tackling the complex issue of depth uncertainty and effectively predicting unobserved views. Our methodologies formulate uncertainty for volume density and offer ways to quantify uncertainty without the need for additional estimation networks.

3 Methodology

3.1 Background

NeRF is a coordinate-based neural representation of the 3D volumetric scene [20]. The goal is to find a continuous non-linear regression $F_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \rho)$, where Θ denotes the network parameters, $\mathbf{x} = (x, y, z)$ and $\mathbf{d} = (\theta, \phi)$ are spatial 3D coordinates in virtual neural field and viewing directions of the camera respectively. For outputs, F_Θ estimates the color $\mathbf{c} = (r, g, b)$ and volume density ρ of the given 3D coordinates in neural field. To perform image rendering, the color of each pixel in the image is estimated by a rendering function, which integrates colors in neural fields over a ray. The expected image color $C(\mathbf{r})$ of camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ that starts from camera origin \mathbf{o} with near and far bound t_n and t_f can be written as:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\rho(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt, \text{ where } T(t) = \exp\left(-\int_{t_n}^t \rho(r(s))ds\right). \quad (1)$$

$T(t)$ denotes the accumulated transmittance along the ray \mathbf{r} . For implementation, a discrete set of samples is utilized to approximate the integral, with stratified sampling ensuring a continuous representation of the scene. The adaptation of the rendering equation for this discrete approximation is articulated as follows:

$$\begin{aligned} \hat{C}(\mathbf{r}) &= \sum_i^N \underbrace{T_i(1 - \exp(-\delta_i\rho_i))}_{\alpha_i} c_i \\ &= \sum_i^N \alpha_i c_i, \text{ where } T_i = \exp\left(-\sum_{j=1}^{i-1} \delta_j \rho_j\right). \end{aligned} \quad (2)$$

Here, $\delta_i = t_{i+1} - t_i$ represents the distance between adjacent samples. The formula for deriving $\hat{C}(\mathbf{r})$ based on (\mathbf{c}_i, ρ_i) is straightforward to differentiate and simplifies to the standard α compositing technique.

3.2 Uncertainty Estimation

Training with limited observations presents a challenge in accurately predicting unobserved parts of the 3D scene. To effectively tackle the problem, it is essential to adopt the Bayesian learning [11] for considering the observation and estimation uncertainty.

For example, [19, 22] simply assume that only the color \mathbf{c} follows a Gaussian distribution $\mathbf{c} \sim \mathcal{N}(\mu_{\mathbf{c}}, \sigma_{\mathbf{c}}^2)$, while treating density ρ and transmittance T as constants. As in Eq. (2), since the cumulated color $C(\mathbf{r})$ along the ray \mathbf{r} is a linear combination of Gaussian random variable \mathbf{c} , it also follows the Gaussian distribution

$$C(\mathbf{r}) \sim \mathcal{N}\left(\sum_{i=1}^{N_s} \mu_i \alpha_i, \sum_{i=1}^{N_s} \sigma_i^2 \alpha_i^2\right). \quad (3)$$

This assumption makes it easy to model the variability in color predictions, facilitating the estimation of color uncertainty and its impact on the rendered scene. However, this color-only approach is limited by considering only RGB images, making it impossible to extend to other sensor values such as depth or intensity. To address the general uncertainty in estimating 3D space from sensors, one can consider incorporating a density field in neural representation. Unfortunately, unlike color-only case, when we assume that density ρ follows a Gaussian distribution $\rho \sim \mathcal{N}(\mu_d, \sigma_d^2)$ the problem becomes extremely complicated. Since $T_i = \exp\left(-\sum_{j=1}^{i-1} \delta_j \rho_j\right)$ as in Eq. (2), we have $T_i \sim \text{Lognormal}(\mu_{T_i}, \sigma_{T_i}^2)$ where $\mu_{T_i} = -\sum_{j=1}^{i-1} \delta_j \mu_j$ and $\sigma_{T_i}^2 = \sum_{j=1}^{i-1} \delta_j^2 \sigma_j^2$. Meanwhile, α in Eq. (2) can be rewritten as:

$$\begin{aligned} \alpha_i &= \exp\left(-\sum_{j=1}^{i-1} \delta_j \rho_j\right) (1 - \exp(-\delta_i \rho_i)) \\ &= \exp\left(-\sum_{j=1}^{i-1} \delta_j \rho_j\right) - \exp\left(-\sum_{j=1}^i \delta_j \rho_j\right) \\ &= T_i - T_{i+1}, \end{aligned} \quad (4)$$

which is the difference of two lognormal random variables T_i and T_{i+1} . It is known to be challenging to have the closed-form distribution of the difference of two *lognormal* random variables, even the two random variables are independent and thus uncorrelated [17, 39]. Therefore, even if we assume that the color \mathbf{c} in Eq. (2) is a constant, $C(\mathbf{r})$ still follows an intractable distribution since $C(\mathbf{r})$ is a linear combination of α . To model this complex distribution, recent works such as [28, 29] utilize generative models and additional network structures, resulting in complex frameworks and challenging learning tasks. In the following, we introduce a series of our analytic but simple and effective approaches with consideration of density uncertainty without the need of additional network structures or trainings.

Gaussian Approximation of Density Uncertainty We propose a method to optimize the NeRF model by considering density uncertainty through a series of assumptions. Using α_i represented in Eq. (4), expected color $C(\mathbf{r})$ and expected depth $D(\mathbf{r})$ of the

ray \mathbf{r} can be derived as [10, 18, 32, 46]:

$$C(\mathbf{r}) = \sum_{i=1}^{N_s} \mathbf{c}_i \alpha_i, \quad (5)$$

$$D(\mathbf{r}) = \sum_{i=1}^{N_s} d_i \alpha_i = \sum_{i=1}^{N_s} \frac{t_{i+1} + t_i}{2} \alpha_i. \quad (6)$$

Here, let us make a slightly strong assumption regarding α , such that $\sum \delta_j \rho_j \ll 1$. In other words, we assume that the intervals between samples are very narrow. Furthermore, neural fields are generally sparse, and the volume density is forced to take values less than 1 using activation functions such as sigmoid. Then we can approximately represent Eq. (4) as the following:

$$\alpha_i \approx \left(1 - \sum_{j=1}^{i-1} \delta_j \rho_j \right) - \left(1 - \sum_{j=1}^i \delta_j \rho_j \right) = \delta_i \rho_i. \quad (7)$$

Now let k 'th volume density ρ_k follows the Gaussian distribution $\rho_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ and the densities are i.i.d.. Assume that the color \mathbf{c} is a constant. Substituting Eq. (7) into Eq. (5), we approximately have the rendered color as $C(\mathbf{r}) \approx \sum_{i=1}^{N_s} \mathbf{c}_i \delta_i \rho_i$. As the rendered color $C(\mathbf{r})$ is now a linear combination of Gaussian random variables, its distribution can be written as:

$$C(\mathbf{r}) \sim \mathcal{N} \left(\sum_{i=1}^{N_s} \mathbf{c}_i \delta_i \mu_i, \sum_{i=1}^{N_s} \mathbf{c}_i^2 \delta_i^2 \sigma_i^2 \right), \quad (8)$$

and we can obtain the approximated solution of the volume density by the following MLE:

$$\{(\hat{\mu}_i, \hat{\sigma}_i)\} = \underset{\{(\hat{\mu}_i, \hat{\sigma}_i)\}}{\operatorname{argmin}} \ln \sum_{i=1}^{N_s} \mathbf{c}_i^2 \delta_i^2 \sigma_i^2 + \frac{(C(\mathbf{r}) - \sum_{i=1}^{N_s} \mathbf{c}_i \delta_i \mu_i)^2}{\sum_{i=1}^{N_s} \mathbf{c}_i^2 \delta_i^2 \sigma_i^2}. \quad (9)$$

Similarly, we have the following MLE for $D(\mathbf{r})$ as:

$$\{(\hat{\mu}_i, \hat{\sigma}_i)\} = \underset{\{(\hat{\mu}_i, \hat{\sigma}_i)\}}{\operatorname{argmin}} \ln \sum_{i=1}^{N_s} d_i^2 \delta_i^2 \sigma_i^2 + \frac{(D(\mathbf{r}) - \sum_{i=1}^{N_s} d_i \delta_i \mu_i)^2}{\sum_{i=1}^{N_s} d_i^2 \delta_i^2 \sigma_i^2}. \quad (10)$$

By using Eq. (9) or Eq. (10), now we can consider the uncertainty of both color or depth, or other sensor information by utilizing the density uncertainty.

Gaussian Approximation of Density and Color Uncertainty In addition to the above method, we also propose a method that considers both color and density by assuming that color follows a Gaussian distribution, rather than being constant. If \mathbf{c}_k also follows the Gaussian distribution $\mathbf{c}_k \sim \mathcal{N}(\mu_{\mathbf{c}_k}, \sigma_{\mathbf{c}_k}^2)$, then Eq. (5) can be seen as a summation

of the product of two Gaussian random variables ρ and \mathbf{c} . The distribution of this product $\rho\mathbf{c}$ is infeasible, however, when $\mu \gg \sigma$ holds for the distributions, we can have the approximated distribution as the following [26]:

$$\sigma_k \mathbf{c}_k \sim \mathcal{N}(\mu_k \mu_{\mathbf{c}_k}, \sigma_k^2 \mu_{\mathbf{c}_k}^2 + \sigma_{\mathbf{c}_k}^2 \mu_k^2 + \sigma_k^2 \sigma_{\mathbf{c}_k}^2).$$

Consequently, the distribution of rendered color $C(\mathbf{r})$ can be approximated as:

$$C(\mathbf{r}) \sim \mathcal{N}\left(\sum_{i=1}^{N_s} \delta_i \mu_i \mu_{\mathbf{c}_i}, \sum_{i=1}^{N_s} \delta_i^2 (\sigma_i^2 \mu_{\mathbf{c}_i}^2 + \sigma_{\mathbf{c}_i}^2 \mu_i^2 + \sigma_i^2 \sigma_{\mathbf{c}_i}^2)\right). \quad (11)$$

Using this approximated distribution, similar to Eq. (9) and Eq. (10), we can have the approximated solution of the volume density and color by the following MLE:

$$\begin{aligned} & (\hat{\mu}_i, \hat{\sigma}_i), (\hat{\mu}_{\mathbf{c}_i}, \hat{\sigma}_{\mathbf{c}_i}) = \\ & \underset{\hat{\mu}, \hat{\sigma}}{\operatorname{argmin}} \ln \left(\sum_{i=1}^{N_s} \delta_i^2 (\sigma_i^2 \mu_{\mathbf{c}_i}^2 + \sigma_{\mathbf{c}_i}^2 \mu_i^2 + \sigma_i^2 \sigma_{\mathbf{c}_i}^2) \right) + \frac{\left(C(\mathbf{r}) - \sum_{i=1}^{N_s} \delta_i \mu_i \mu_{\mathbf{c}_i} \right)^2}{\sum_{i=1}^{N_s} \delta_i^2 (\sigma_i^2 \mu_{\mathbf{c}_i}^2 + \sigma_{\mathbf{c}_i}^2 \mu_i^2 + \sigma_i^2 \sigma_{\mathbf{c}_i}^2)}. \end{aligned} \quad (12)$$

Markov Approach for Density Uncertainty The approximation in Eq. (7) hardly holds when the ray \mathbf{r} travels a long distance, or the volume density σ takes a large value. As a result, the approximated solutions obtained from Eq. (9) and Eq. (12) become inaccurate. To fundamentally address such approximation errors, we apply the Markov assumption to the traversal of rays as follows: 1) In Eq. (2), α_i represents i 'th random variable involving i 'th density ρ_i and the past densities $\rho_0, \dots, \rho_{i-1}$ expressed as T_i . Since $\rho_0, \dots, \rho_{i-1}$ are variables observed along the path traveled by the ray in temporal order, they can be considered deterministic variables at time step i . Therefore, the probability distribution for α_i can be expressed as follows:

$$p(\alpha_i | \rho_i, \dots, \rho_0) \simeq p(\alpha_i | \rho_i) = p(\alpha_i | o) \quad (13)$$

where $o = 1 - \exp(-\delta_i \rho_i)$ is defined as the occupancy variable, which effectively normalizes density to indicate the presence of material. 2) Similarly, if the summation in the rendering process Eqs. (5) and (6) is performed in temporal order, then α_i can also be regarded as a variable determined in certain time step. In other words, at time step $i+1$, $\alpha_0, \dots, \alpha_i$ are already obtained and determined, making them independent of α_{i+1} . 3) We assume that occupancy o follows a Gaussian distribution $o \sim \mathcal{N}(\mu_o, \sigma_o^2)$. Following these reasonings 1)-3), the rendered color $C(\mathbf{r})$ in Eqs. (5) and (6) can be treated as the linear combination of the Gaussian random variables, simplifying its distribution as follows:

$$C(\mathbf{r}) \sim \mathcal{N}\left(\sum_{i=1}^{N_s} \mathbf{c}_i T_i \mu_{o_i}, \sum_{i=1}^{N_s} \mathbf{c}_i^2 T_i^2 \sigma_{o_i}^2\right). \quad (14)$$

Using the distribution of the rendered color, the approximate solution for volume density can be derived using the following MLE:

$$\{(\hat{\mu}_i, \hat{\sigma}_i)\} = \underset{\{(\hat{\mu}_i, \hat{\sigma}_i)\}}{\operatorname{argmin}} \ln \sum_{i=1}^{N_s} \mathbf{c}_i^2 T_i^2 \sigma_{o_i}^2 + \frac{(C(\mathbf{r}) - \sum_{i=1}^{N_s} \mathbf{c}_i T_i \mu_{o_i})^2}{\sum_{i=1}^{N_s} \mathbf{c}_i^2 T_i^2 \sigma_{o_i}^2}. \quad (15)$$

Note that for $D(\mathbf{r})$, by replacing \mathbf{c}_i to d_i , we can have the similar formulations for the depth estimation since we have the generalized approximation for the occupancy o . We assume the probability variables to be independent with respect to the time sequence and validate the efficiency of such approximation methods through experiments.

4 Experiment

4.1 Experimental Setup

In our experiments, we explore the concept of unobserved and observed views in the context of NeRF. Unobserved views refer to perspectives that the camera has not captured directly, either due to limited camera trajectories or self-occlusion phenomena. In contrast, observed views refer to perspectives close to those captured by the camera. Our method assesses the efficacy of estimating these unobserved views without the necessity for an additional network, focusing on uncertainty related to density. Furthermore, we conduct comparative experiments to predict the observed views, thereby analyzing performance enhancements in rendering unobserved views. We evaluated five different methods in our dataset: (a) *Baseline*, following the traditional NeRF method with reliance solely on photometric loss; (b) *Color*, integrating color uncertainty into NeRF; (c) *Density*, emphasizing density uncertainty during NeRF training; (d) *Color + Density*, addressing uncertainties in both color and density; and (e) *Occupancy*, advancing NeRF by considering occupancy uncertainty based on Markov assumptions.

Dataset We conduct experiments with both RGB and depth images to demonstrate that considering density uncertainty is effective across different image types. For the RGB image dataset, we utilize NeRF datasets [20] comprising both synthetic and real-world scenes. The synthetic dataset was made by rotating around objects at 360 degrees, serving as a benchmark to assess the effectiveness of predicting unobserved views when considering uncertainty. Meanwhile, the Local Light Field Fusion (LLFF) dataset, composed of forward-facing handheld real scenes, is used to evaluate the effectiveness of observed view predictions considering uncertainty. For the depth image dataset, we use the ModelNet10 dataset [37]. This involves extracting depth images from the CAD information through a process that mirrors the NeRF synthetic dataset approach, where we rotate the models every 10 degrees to achieve a comprehensive 360-degree view. The detailed configurations of these datasets are further illustrated in 2, providing an overview of our experimental setup.

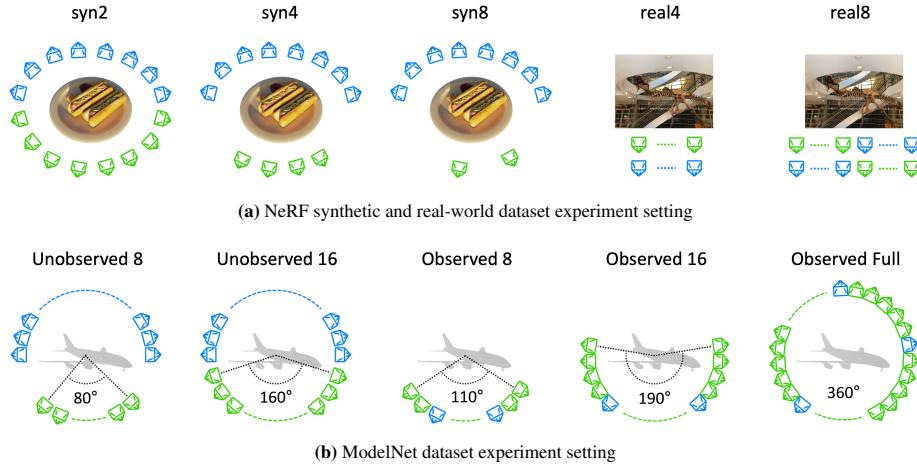


Fig. 2: Dataset Setting. We conduct experiments on the NeRF synthetic, real-world, and ModelNet datasets, with training sets from green cameras and test sets from blue cameras. (a) The NeRF synthetic dataset is used to evaluate unobserved prediction using a limited camera view range for training. In contrast, the real-world dataset, with its forward camera trajectory, randomly assembles train and test sets to assess the observed prediction. (b) The ModelNet dataset evaluates both unobserved and observed predictions, varying its configuration based on the evaluation focus. Images are spaced at 10-degree intervals, totaling 36 images arranged as depicted. (Experiments conducted with the full dataset (a) are not shown.)

Metric In our study, we use three primary metrics for RGB image quality. PSNR [13] evaluates luminance accuracy, SSIM [45] analyzes perceptual quality including luminance, contrast, and structure, and LPIPS [47] uses deep learning to gauge perceptual similarity, reflecting human visual perception. For depth image evaluation, essential for spatial detail, we apply Absolute Relative Error (AbsRel) to measure depth deviation and Root Mean Square Error in log space (RMSElog) for errors closer to the viewer, focusing on the precision of depth estimation. We also incorporate the Average log10 Error (log10), which assesses the average absolute difference in the logarithm base 10 of the estimated versus true depths. Following this, we utilize Threshold Accuracy (δ_i) to assess depth accuracy against a specific standard.

Implementation Details To assess unobserved and observed views, we configure three dataset settings as shown in 2. By varying the number of images used to construct the training set, we evaluate the influence of our method on predicting unobserved views with minimal data: 1) For the synthetic dataset, we use 2, 4, 8, and 100 (full) images for training. The training set is constructed from images captured by green cameras, expanding the camera's observed view range as the number of images increases. The test set consists of 100 images of unobserved views with blue cameras. 2) The LLFF dataset, Unlike the synthetic dataset, is a real-world dataset with a forward-facing camera trajectory. Consequently, without differentiating between observed and unobserved views, we conduct experiments to assess the accuracy of observed view predictions

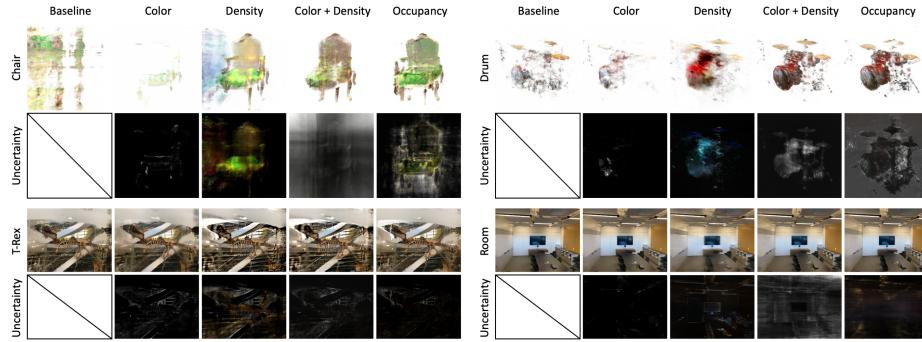


Fig. 3: Qualitative comparisons for the NeRF synthetic and real-world datasets. We compare our proposed methods against the baseline and the color uncertainty method, presenting uncertainty maps for each approach. The chair scene uses two training images and the drum scene employs four training images. Incorporating uncertainty shows a clear performance difference; however, the (b) *Color* method fails to accurately estimate density, resulting in the object appearing faded. The T-rex and room scene employs eight training images. This scene, constructed with a forward camera trajectory, gains relatively less advantage from considering uncertainty.

when considering uncertainty. For training, the dataset is randomly selected, using sets of 4, 8, and the full dataset amount. The notation "full" indicates that all available training data were utilized for learning, reflecting the variance in the amount of training data across different scenes. 3) The ModelNet dataset, generated from CAD information by rotating 10 degrees to produce 36 depth images, varies its configuration based on the prediction accuracy for unobserved and observed views. Given depth images' lack of color, resulting in less detail than RGB images, we configure the training set with a relatively higher number of images to achieve meaningful rendered images. For unobserved view predictions, we use 8 and 16 training images, with respective camera ranges of 80 and 160 degrees as shown in 2. The test set comprises 18 images from entirely different views not used in training, covering a 180-degree range. For observed view predictions, our experiments utilize training sets with 8, 16, and 30 (Full) images. We intersperse 3, 3, and 6 test images within these sets at intervals reflecting the size of each training set.

4.2 Experiment Result and Discussion

NeRF Synthetic and Real-World Dataset In Tab. 1, we show the metric results for RGB images across the five introduced methods. The synthetic dataset shows significant performance improvements, while in the real-world dataset, images taken from the scene's forward direction show modest improvements. Moreover, settings with fewer training images demonstrate greater performance enhancements than those utilizing the full dataset. This outcome indicates that in scenarios where data is limited, the inherent uncertainty within the data becomes more pronounced, making the incorporation of uncertainty into the learning process significantly more impactful.

Although considering uncertainty generally improves performance, the (b) *Color* and (c) *Density* approaches show relatively low performance. Particularly in experi-

Table 1: Quantitative results for the NeRF synthetic and real-world datasets. We compare our proposed methods with the baseline and color uncertainty methods. Experiments are conducted in all scenes within each dataset and average values are provided. Our experiment aims to verify performance in limited scenarios, thus various settings for the number of images used in training are established. Significant performance improvements are observed in the unobserved view setting, with notable gains also in the observed view setting. Both the method considering uncertainty in occupancy and the method accounting for uncertainties in both color and density demonstrate superior performance over other approaches. ("Full" denotes using all training data, with the amount varying by scene.)

Dataset	PSNR \uparrow				
	Baseline	Color	Density	Color+Density	Occupancy
Syn2	11.75	11.29	13.61	14.72	15.12
Syn4	15.40	16.18	17.02	18.01	18.30
Syn8	18.04	18.91	19.19	20.87	20.75
SynFull	25.57	24.20	24.25	25.64	25.68
real4	12.88	12.97	12.88	13.51	13.45
real8	22.60	22.66	22.26	23.01	23.10
realFull	24.81	24.03	23.86	24.78	24.97
Dataset	SSIM \uparrow				
	Baseline	Color	Density	Color+Density	Occupancy
Syn2	0.61	0.61	0.66	0.67	0.69
Syn4	0.74	0.75	0.76	0.78	0.79
Syn8	0.79	0.80	0.79	0.82	0.82
SynFull	0.88	0.86	0.86	0.88	0.88
real4	0.52	0.53	0.52	0.54	0.55
real8	0.69	0.69	0.67	0.69	0.71
realFull	0.76	0.75	0.73	0.76	0.77
Dataset	LPIPS \downarrow				
	Baseline	Color	Density	Color+Density	Occupancy
Syn2	0.44	0.42	0.36	0.34	0.33
Syn4	0.27	0.26	0.24	0.21	0.20
Syn8	0.22	0.21	0.21	0.17	0.17
SynFull	0.11	0.14	0.14	0.11	0.10
real4	0.62	0.63	0.62	0.60	0.59
real8	0.22	0.24	0.25	0.20	0.19
realFull	0.17	0.21	0.22	0.17	0.16

ments with a constrained number of training images, the (b) *Color* method encounters significant challenges in accurately estimating density. This difficulty appears as visual artifacts where objects within the scene, notably illustrated in the chair and drum scene of Fig. 3, appear faded or even vanish. This phenomenon shows the (b) *Color* method's limitations in capturing and reproducing detailed density variations across the scene with limited data. In contrast, methods trained by considering uncertainty related to density exhibit markedly improved robustness. These approaches successfully mitigate such artifacts, ensuring objects retain their visibility and structural integrity, regardless

of the dataset’s size. This distinction emphasizes the critical importance of accounting for uncertainty in geometric structures, particularly in scenarios where the available data does not comprehensively cover the scene’s variability.

In our experiments with the (c) *Density* method, we found that high volume density values in certain datasets pose challenges for accurate density estimation. The accuracy of the method’s approximation declines when the model encounters high density values, resulting in less reliable outcomes from the MLE equations. Consequently, datasets with elevated density values saw a decline in performance due to inaccuracies in the uncertainty approximation. This is particularly evident when comparing performance metrics in Tab. 1, where the (c) *Density* method underperforms compared to (d) *Color + Density* and (e) *Occupancy* approaches.

The problems associated with the (b) *Color* and (c) *Density* methods are effectively resolved by integrating these approaches in the (d) *Color + Density* method, which shows significant performance improvements. However, while this integrated method amalgamates the strengths of both (b) *Color* and (c) *Density* to enhance performance, it occasionally fails to meet the fundamental assumptions inherent in the (c) *Density* approach. This limitation can lead to instances where the integrated method does not perform optimally, particularly in scenarios where the assumptions regarding volume density are not fully satisfied. Such discrepancies underscore the method’s occasional imperfections, despite its overall advancements.

On the other hand, the (e) *Occupancy* approach, derived through reasonable Markov assumptions, exhibits a remarkable level of stability across a wide range of scenarios, including both synthetic and real-world datasets. This stability is not merely a result of its methodological soundness, but also reflects its ability to effectively manage the inherent uncertainty within the data without being overly dependent on the precise fulfillment of specific assumptions.

We have validated a series of methods proposed under a Bayesian framework for RGB images, particularly highlighting the (d) *Color + Density* and (e) *Occupancy* methods for their substantial performance improvements. These approaches, especially when integrated, demonstrate a robust capability to significantly enhance performance, reflecting the strength of applying Bayesian principles to handle uncertainties in neural radiance fields. While our proposed methods show promising results, adopting a probabilistic framework for training may sometimes lead to unstable learning phases, potentially resulting in inaccuracies [2, 7]. Challenges such as parameter initialization and sensitive terms in the loss function can significantly influence the loss value, potentially causing the error function to diverge and leading to issues like image whitening. To overcome these challenges, we employ pre-training to provide a stable foundation and enhance 3D structure processing. By integrating this with NeRF and alternating focus on uncertainty, we streamline training and improve model robustness. This method helps stabilize the loss function, minimizing the risk of divergence and reducing instances of image whitening, thereby improving overall model reliability and performance across various scenarios.

ModelNet Dataset In our ModelNet dataset analysis, as detailed in Tab. 2, we present the metric results for depth images. With depth images absent of RGB information, it is

Table 2: Quantitative results for the ModelNet datasets. Given that depth images lack color information, we compare the (a) *Baseline*, (c) *Density*, and (e) *Occupancy* methods. Experiments were conducted on two instances across all classes, with average values provided. The dataset setting follows Fig. 2, and similar to the RGB dataset, considering uncertainty in the unobserved view prediction setting demonstrates significant performance improvements.

Dataset	Unobserved View Setting			AbsRel \downarrow	RMSE(log) \downarrow	log10 \downarrow
	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$			
original 8	0.497	0.672	0.755	0.537	0.467	0.159
uncert den 8	0.506	0.685	0.809	0.492	0.440	0.151
uncert occu 8	0.543	0.712	0.837	0.451	0.418	0.140
original 16	0.740	0.906	0.962	0.203	0.262	0.076
uncert den 16	0.762	0.911	0.961	0.192	0.257	0.073
uncert occu 16	0.785	0.922	0.964	0.177	0.245	0.068

Dataset	Observed View Setting			AbsRel \downarrow	RMSE(log) \downarrow	log10 \downarrow
	$\delta < 1.25 \uparrow$	$\delta < 1.25^2 \uparrow$	$\delta < 1.25^3 \uparrow$			
original 8	0.880	0.939	0.971	0.102	0.192	0.044
uncert den 8	0.886	0.943	0.973	0.099	0.188	0.043
uncert occu 8	0.891	0.946	0.975	0.096	0.184	0.041
original 16	0.888	0.943	0.974	0.099	0.189	0.043
uncert den 16	0.891	0.946	0.977	0.097	0.186	0.042
uncert occu 16	0.895	0.949	0.978	0.094	0.183	0.041
original full	0.896	0.946	0.976	0.096	0.180	0.041
uncert den full	0.896	0.946	0.977	0.096	0.181	0.041
uncert occu full	0.902	0.949	0.978	0.092	0.176	0.039

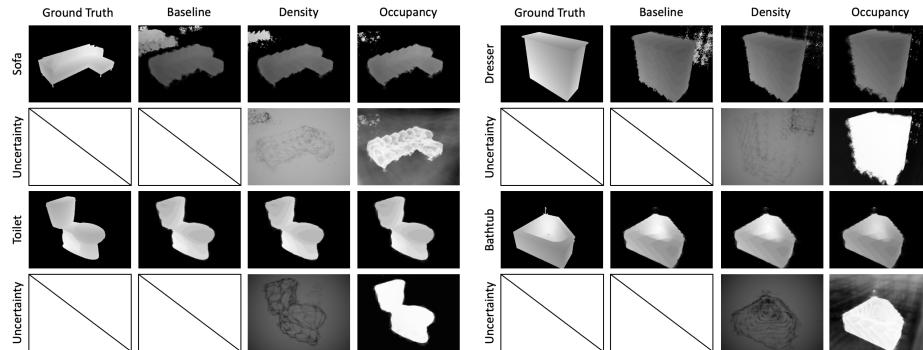


Fig. 4: Qualitative comparisons for the ModelNet dataset. The sofa and dresser scene employs an unobserved view setting and is trained using 8 images. In this scene, the (e) *Occupancy* method effectively addresses the blurring issue. The bath and toilet scene, using the full count of images in an observed view setting, shows clear results.

not feasible to consider uncertainty related to color. Therefore, experiments were conducted using (a) *Baseline*, (c) *Density*, and (e) *Occupancy* approaches. Our findings reveal a consistent trend with what was observed in RGB images: significant performance

improvements in scenarios involving unobserved views and only modest gains in observed view settings. A notable example of this can be seen in the sofa and dresser scene from Fig. 4, where blurring in empty spaces is significantly reduced when uncertainty is factored into the model. This blurring, prevalent across various scenes, is notably mitigated by the (e) *Occupancy* method, showcasing its enhanced capability to refine depth image clarity compared to the (c) *Density* approach. In the toilet and bathtub scene, representing an observed view setting, qualitative differences are subtler due to the comprehensive use of the image set. However, the quantitative analysis presented in Tab. 2 unequivocally demonstrates the better performance of the (e) *Occupancy* method. This suggests that incorporating occupancy into the training of NeRF models for depth images not only enhances performance in specific scenarios, but also indicates broader applicability and effectiveness across both RGB and depth image datasets.

The effectiveness of our proposed methods in depth images underscores their potential for integration with uncertainty across different sensors, emphasizing the scalability of our approach. This adaptability becomes particularly valuable when observational data is abundant and uncertainty within the dataset is less pronounced. Employing data fusion from sensors like RGB, depth, and range allows for the direct modeling of uncertainty, leading to more accurate decision-making. This strategy proves especially advantageous in fields such as robotics and autonomous driving, enhancing model functionality by leveraging the unique strengths of each sensor. Adjusting the interaction weights among these sensors improves the precision of uncertainty estimation, showcasing our approach’s wide applicability and substantial benefits in a variety of scenarios.

Limitations and Future Work Our method encounters challenges due to temporal gaps in the training data, leading to complications in accurately adapting to scenarios. This issue, intensified as data accumulates, necessitates the ongoing recalibration of uncertainty, limiting the approach’s effectiveness in dynamic real-world scenarios. However, directly addressing uncertainty in time-series data presents a viable path to enhance our model’s performance and adaptability. Future efforts will aim to refine this strategy, focusing on more efficient uncertainty management with new data. Such improvements could broaden our method’s utility in practical applications like autonomous driving and robotic navigation, optimizing SLAM technologies for diverse environments.

5 Conclusion

In this work, we have introduced the Bayesian Neural Radiance Field (NeRF) method, significantly advancing the 3D scene representation from diverse viewpoints. Our approach, which quantifies uncertainty in geometric volume structures without extra networks, effectively handles challenging observations. By implementing generalized approximations and defining density-related uncertainty, we have broadened the application of Bayesian NeRF beyond RGB to depth images, achieving enhanced performance across extensive datasets. Our experiments confirm the robustness of this approach, markedly improving rendering quality and demonstrating the reliability of our uncertainty quantification methods. We highlight the critical role of integrating uncertainty

into neural radiance fields, moving beyond the use of additional networks and empirical strategies. Our series of formulation extensions to NeRF presents a scalable solution, managing uncertainty across different sensors and facilitating broader real-world applications. This advancement contributes significantly to fields such as virtual reality, robotics, and autonomous driving, promising efficiency even with limited data availability.

As we progress, tackling challenges such as temporal gaps in training data will be essential for further refining our approach's adaptability. The potential to merge our method with data fusion techniques suggests exciting research and development directions, promising extensive practical applications and bolstering the utility and impact of advanced visualization and navigation technologies in a variety of scenarios.

References

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion* **76**, 243–297 (2021) 2
2. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International conference on machine learning. pp. 214–223. PMLR (2017) 12
3. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Zip-nerf: Anti-aliased grid-based neural radiance fields. arXiv preprint arXiv:2304.06706 (2023) 3
4. Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D.: Weight uncertainty in neural network. In: International conference on machine learning. pp. 1613–1622. PMLR (2015) 2
5. Deng, N., He, Z., Ye, J., Duinkharjav, B., Chakravarthula, P., Yang, X., Sun, Q.: Fov-nerf: Foveated neural radiance fields for virtual reality. *IEEE Transactions on Visualization and Computer Graphics* **28**(11), 3854–3864 (2022) 1
6. Goli, L., Reading, C., Sellán, S., Jacobson, A., Tagliasacchi, A.: Bayes' rays: Uncertainty quantification for neural radiance fields. arXiv preprint arXiv:2309.03185 (2023) 3
7. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. *Advances in neural information processing systems* **30** (2017) 12
8. Hu, W., Wang, Y., Ma, L., Yang, B., Gao, L., Liu, X., Ma, Y.: Tri-miprf: Tri-mip representation for efficient anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 19774–19783 (2023) 3
9. Hu, X., Xiong, G., Zang, Z., Jia, P., Han, Y., Ma, J.: Pc-nerf: Parent-child neural radiance fields under partial sensor data loss in autonomous driving environments. arXiv preprint arXiv:2310.00874 (2023) 1
10. Huang, S., Gojcic, Z., Wang, Z., Williams, F., Kasten, Y., Fidler, S., Schindler, K., Litany, O.: Neural lidar fields for novel view synthesis. arXiv preprint arXiv:2305.01643 (2023) 6
11. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems* **30** (2017) 2, 4
12. Kong, X., Liu, S., Taher, M., Davison, A.J.: vmap: Vectorised object mapping for neural field slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 952–961 (2023) 3
13. Korhonen, J., You, J.: Peak signal-to-noise ratio revisited: Is simple beautiful? In: 2012 Fourth International Workshop on Quality of Multimedia Experience. pp. 37–38. IEEE (2012) 9
14. Lee, M., Kang, K., Yu, H.: Just flip: Flipped observation generation and optimization for neural radiance fields to cover unobserved view. arXiv preprint arXiv:2303.06335 (2023) 3

15. Liu, R., Wu, R., Van Hoorick, B., Tokmakov, P., Zakharov, S., Vondrick, C.: Zero-1-to-3: Zero-shot one image to 3d object. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9298–9309 (2023) [3](#)
16. Liu, Y., Tu, X., Chen, D., Han, K., Altintas, O., Wang, H., Xie, J.: Visualization of mobility digital twin: Framework design, case study, and future challenges. In: 2023 IEEE 20th International Conference on Mobile Ad Hoc and Smart Systems (MASS). pp. 170–177. IEEE (2023) [1](#)
17. Lo, C.F., et al.: The sum and difference of two lognormal random variables. *Journal of Applied Mathematics* **2012** (2012) [5](#)
18. Malik, A., Mirdehghan, P., Nousias, S., Kutulakos, K.N., Lindell, D.B.: Transient neural radiance fields for lidar view synthesis and 3d reconstruction. arXiv preprint arXiv:2307.09555 (2023) [6](#)
19. Martin-Brualla, R., Radwan, N., Sajjadi, M.S., Barron, J.T., Dosovitskiy, A., Duckworth, D.: Nerf in the wild: Neural radiance fields for unconstrained photo collections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7210–7219 (2021) [2, 3, 5](#)
20. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1), 99–106 (2021) [1, 4, 8](#)
21. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)* **41**(4), 1–15 (2022) [3](#)
22. Pan, X., Lai, Z., Song, S., Huang, G.: Activenerf: Learning where to see with uncertainty estimation. In: European Conference on Computer Vision. pp. 230–246. Springer (2022) [2, 3, 5](#)
23. Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: Dreamfusion: Text-to-3d using 2d diffusion. arXiv preprint arXiv:2209.14988 (2022) [3](#)
24. Rosinol, A., Leonard, J.J., Carbone, L.: Nerf-slam: Real-time dense monocular slam with neural radiance fields. In: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 3437–3444. IEEE (2023) [1, 3](#)
25. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4104–4113 (2016) [1, 3](#)
26. Seijas-Macías, A., Oliveira, A.: An approach to distribution of the product of two normal variables. *Discussiones Mathematicae Probability and Statistics* **32**(1-2), 87–99 (2012) [7](#)
27. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06). vol. 1, pp. 519–528. IEEE (2006) [1, 3](#)
28. Shen, J., Agudo, A., Moreno-Noguer, F., Ruiz, A.: Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In: European Conference on Computer Vision. pp. 540–557. Springer (2022) [2, 3, 5](#)
29. Shen, J., Ruiz, A., Agudo, A., Moreno-Noguer, F.: Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In: 2021 International Conference on 3D Vision (3DV). pp. 972–981. IEEE (2021) [2, 3, 5](#)
30. Sünderhauf, N., Abou-Chakra, J., Miller, D.: Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 9370–9376. IEEE (2023) [3](#)
31. Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P.P., Barron, J.T., Kretzschmar, H.: Block-nerf: Scalable large scene neural view synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8248–8258 (2022) [3](#)

32. Tao, T., Gao, L., Wang, G., Chen, P., Hao, D., Liang, X., Salzmann, M., Yu, K.: Lidar-nerf: Novel lidar view synthesis via neural radiance fields. arXiv preprint arXiv:2304.10406 (2023) 6
33. Thrun, S.: Probabilistic robotics. Communications of the ACM **45**(3), 52–57 (2002) 2
34. Turki, H., Ramanan, D., Satyanarayanan, M.: Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12922–12931 (June 2022) 3
35. Wang, C., Chai, M., He, M., Chen, D., Liao, J.: Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3835–3844 (2022) 3
36. Wang, H., Wang, J., Agapito, L.: Co-slam: Joint coordinate and sparse parametric encodings for neural real-time slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13293–13302 (2023) 1
37. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1912–1920 (2015) 8
38. Wu, Z., Liu, T., Luo, L., Zhong, Z., Chen, J., Xiao, H., Hou, C., Lou, H., Chen, Y., Yang, R., et al.: Mars: An instance-aware, modular and realistic simulator for autonomous driving. In: CAAI International Conference on Artificial Intelligence. pp. 3–15. Springer (2023) 1
39. Wutzler, T.: Functions for the lognormal distribution in R, howpublished = <https://github.com/bgctw/lognorm>, note = Accessed: 2023-09-19 5
40. Xiang, Y., Choi, W., Lin, Y., Savarese, S.: Data-driven 3d voxel patterns for object category recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1903–1911 (2015) 1
41. Yang, J., Pavone, M., Wang, Y.: Freenerf: Improving few-shot neural rendering with free frequency regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8254–8263 (2023) 3
42. Yang, X., Li, H., Zhai, H., Ming, Y., Liu, Y., Zhang, G.: Vox-fusion: Dense tracking and mapping with voxel-based neural implicit representation. In: 2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). pp. 499–507. IEEE (2022) 3
43. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4578–4587 (2021) 3
44. Yuan, Y.J., Sun, Y.T., Lai, Y.K., Ma, Y., Jia, R., Gao, L.: Nerf-editing: geometry editing of neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18353–18364 (2022) 3
45. Yue, J.C., Clayton, M.K.: A similarity measure based on species proportions. Communications in Statistics-theory and Methods **34**(11), 2123–2131 (2005) 9
46. Zhang, J., Zhang, F., Kuang, S., Zhang, L.: Nerf-lidar: Generating realistic lidar point clouds with neural radiance fields. arXiv preprint arXiv:2304.14811 (2023) 6
47. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018) 9
48. Zhu, Z., Peng, S., Larsson, V., Xu, W., Bao, H., Cui, Z., Oswald, M.R., Pollefeys, M.: Nice-slam: Neural implicit scalable encoding for slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12786–12796 (2022) 3