



Article

# Uncertainty Quantification of Neural Reflectance Fields for Underwater Scenes

Haojie Lian <sup>1,2</sup> , Xinhao Li <sup>1</sup>, Leilei Chen <sup>2,\*</sup>, Xin Wen <sup>3</sup>, Mengxi Zhang <sup>4</sup>, Jieyuan Zhang <sup>5</sup> and Yilin Qu <sup>6,7,8,\*</sup> 

- <sup>1</sup> Key Laboratory of In-Situ Property-Improving Mining of Ministry of Education, Taiyuan University of Technology, Taiyuan 030024, China; lianhaojie@tyut.edu.cn (H.L.)
- <sup>2</sup> Henan International Joint Laboratory of Structural Mechanics and Computational Simulation, School of Architecture and Civil Engineering, Huanghuai University, Zhumadian 463000, China
- <sup>3</sup> School of Software, Taiyuan University of Technology, Jinzhong 030600, China; xwen@tyut.edu.cn
- <sup>4</sup> State Key Laboratory of Hydraulic Engineering Intelligent Construction and Operation, Tianjin University, Tianjin 300350, China; zhangmx@tju.edu.cn
- <sup>5</sup> Academy of Military Science, Beijing 100091, China; zhangjynudt@163.com
- <sup>6</sup> School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China
- <sup>7</sup> Unmanned Vehicle Innovation Center, Ningbo Institute of Northwestern Polytechnical University, Ningbo 315103, China
- <sup>8</sup> Key Laboratory of Unmanned Underwater Vehicle Technology of Ministry of Industry and Information Technology, Northwestern Polytechnical University, Xi'an 710072, China
- \* Correspondence: chenlei@mail.ustc.edu.cn (L.C.); quyilin\_95@126.com (Y.Q.)

**Abstract:** Neural radiance fields and neural reflectance fields are novel deep learning methods for generating novel views of 3D scenes from 2D images. To extend the neural scene representation techniques to complex underwater environments, beyond neural reflectance fields underwater (BNU) was proposed, which considers the relighting conditions of on-board light sources by using neural reflectance fields, and approximates the attenuation and backscatter effects of water with an additional constant. Because the quality of the neural representation of underwater scenes is critical to downstream tasks such as marine surveying and mapping, the model reliability should be considered and evaluated. However, current neural reflectance models lack the ability of quantifying the uncertainty of underwater scenes that are not directly observed during training, which hinders their widespread use in the field of underwater unmanned autonomous navigation. To address this issue, we introduce an ensemble strategy to BNU that quantifies cognitive uncertainty in color space and unobserved regions with the expectation and variance of RGB values and termination probabilities along the ray. We also employ a regularization method to smooth the density of the underwater neural reflectance model. The effectiveness of the present method is demonstrated in numerical experiments.

**Keywords:** neural reflectance fields; underwater scenes; uncertainty quantification



**Citation:** Lian, H.; Li, X.; Chen, L.; Wen, X.; Zhang, M.; Zhang, J.; Qu, Y. Uncertainty Quantification of Neural Reflectance Fields for Underwater Scenes. *J. Mar. Sci. Eng.* **2024**, *12*, 349. <https://doi.org/10.3390/jmse12020349>

Academic Editor: Sergei Chernyi

Received: 12 January 2024

Revised: 15 February 2024

Accepted: 16 February 2024

Published: 18 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A 3D reconstruction of underwater scenes is central to marine environmental studies. Neural radiance fields (NeRFs) [1] is a recently emerging method for synthesizing novel views from 2D images based on volume rendering and deep learning. Since its inception, it has drawn tremendous attention in 3D reconstruction area due to its flexibility of processing complex scenes. Nonetheless, the original version of NeRFs was proposed in natural and ambient light in clear air conditions. For underwater environment conditions, which is where the images were acquired by autonomous underwater vehicles (AUVs), two intractable problems arose when using the conventional NeRFs method: (i) the scattering media in the ocean led to attenuation and backscatter effects in the light and distorted the true color; (ii) illumination of on-board light sources resulted in relighting on the scene's appearance. To tackle these challenges, beyond neural reflectance fields underwater

(BNU) [2] was proposed to model the effects of water as a combination of attenuation and inverse scattering. Additionally, it employs neural reflectance fields [3] in place of neural radiance fields to learn the scene's reflection properties and simulate the impact of lighting variations on the scene's appearance. BNU has made a great contribution to neural volume rendering and has paved the way for its application in novel view synthesis from underwater imagery.

Due to the complexity of the underwater environment, as well as due to the limitation of lacking cognitive knowledge about unobserved parts of the scene, there is inherent uncertainty when modeling underwater optical effects. Such cognitive uncertainty existing in underwater neural scene representation directly influences the downstream tasks of AUV navigation, such as the automatic identification of damage in underwater infrastructure, target detection and tracking, mapping and motion planning, etc. [4–6]. By quantifying uncertainty, we can assess the reliability of the model and thus reduce the risks of failures in the aforementioned downstream tasks in real-world underwater exploration conditions that are subject to probabilistic constraints [7].

Uncertainty quantification is a highly active research area in machine learning [8]. Numerous techniques have been proposed such as Bayesian neural networks [9,10], Gaussian procedure, neural processes, Monte-Carlo dropout [11], and ensemble networks [12]. Inspired by the aforementioned research, some scholars have begun to analyze the uncertainties of neural radiance fields [13]. However, to the authors' best knowledge, quantifying the uncertainty associated with neural reflectance fields, particularly for underwater scenes, has not been reported in the existing literature.

To fill the aforementioned research gap, the present study attempted to extend the uncertainty quantification of neural radiance fields [13] to neural reflectance fields [2] based on BNU. We took an ensemble learning approach, which not only averages rendered RGB images, but also introduces an additional cognitive uncertainty term consisting of the termination probability of each ray in order to infer the cognitive uncertainty arising from the absence of cognitive knowledge about the unobserved part of the scene during training [13]. In addition, we employ a regularization method that smooths the density amplitude to test the model's inference uncertainty capability [14].

In this work, we demonstrated that our model can explicitly infer uncertainty in both synthetic and real-world underwater scenes. Besides, the model exhibits superior performance in key metrics related to image quality and reconstruction error. In summary, our research makes the following contributions:

- For the first time, uncertainty quantification has been introduced to neural reflectance fields of underwater scenes, thus enabling us to analyse the reliability and enhance the robustness of the model.
- The regularization proposed in Ref. [14] is incorporated to BNU.
- Our uncertainty quantification framework strictly follows a volume rendering procedure and does not require changes to underlying architectures of the codes.

## 2. Related Works

### 2.1. Underwater Neural Scene Representation

Neural scene representation is a method for encoding and reconstructing 3D scenes based on neural networks. Among them, NeRFs has gained increasing popularity due to its high-quality and realistic image rendering results. Numerous works have followed the steps of NeRFs, and they have expanded the original framework of NeRFs from different perspectives. However, NeRFs and its variants overlook the strong influence of the medium on object appearance in underwater conditions. WaterNeRF [15] takes this as a starting point and utilizes mip-NeRF 360 [16] to simulate underwater scenes, where absorption and backscatter coefficients are learnt by optimizing the Sinkhorn loss between rendered images and histogram-equalized images. Unlike WaterNeRF, SeaThru-NeRF [17] introduces a scattering image formation model to capture the impact of the underwater medium on imaging by respectively assigning color and density parameters to objects and the medium,

to model the effects of shallow water natural ambient light. Additionally, a typical physical-based reconstruction of the optical scene in shallow underwater environment conditions is discussed in [18,19]. They predicted the impact of optical effects on underwater images by inputting the underwater images into an underwater wave propagation model. It is noteworthy that underwater scenes are largely dependent on the depth of the water. Shallow water scenes are influenced by sunlight, while deeper water layers prevent the penetration of sunlight. For deep-sea environments, illumination primarily relies on searchlight beams that are attached to the underwater vehicle, and the influence of natural ambient light can be negligible. Therefore, as the underwater vehicle moves, the scene's appearance will change with different lighting conditions. To adapt to these changes, it is necessary to model the reflection properties of the scene using neural reflectance fields [3]. For example, BNU has utilized neural reflectance field model to learn the albedo, surface normal, and volume density of underwater environments [2]. By jointly learning the medium effects and neural scene representation, it recovers the true colors of underwater images and achieves high-quality rendering under new lighting conditions.

### 2.2. Uncertainty Estimation in Deep Learning

Uncertainty estimation is a research topic of theoretical and practical importance across many domains [20,21]. Characterizing the uncertainty of neural networks not only enhances the interpretability of model outputs, but also reduces the risk of severe failures occurring. Early research proposed using Bayesian neural networks (BNNs) [22] to introduce probability distributions and estimate the uncertainty of network weights and outputs by probabilistic modeling. However, training BNNs typically require significant modifications to the network architecture and training process, which is computationally expensive. As such, several more low-complexity and practical strategies have been proposed to incorporate uncertainty estimation into deep neural networks, such as Monte-Carlo dropout and deep ensembles. MC-dropout-based methods [23–25] introduce randomness on the intermediate neurons of the network. Such methods perform multiple forward propagations of the model during testing and randomly discard some neurons in each forward propagation to obtain different predictions, which are then aggregated to obtain an uncertainty estimate for a given input. However, the need to perform multiple forward propagations increases the computational cost during the inference process. In deep ensemble strategies [26], by training a finite ensemble of independent deep-learning models, introducing differences such as different initializations, different training data sampling, or different model structures, this kind of method can provide more comprehensive and reliable uncertainty estimation, especially in the case of complex noisy data with better robustness than other methods. Refs. [27–29] adopted model order reduction methods combined with deep learning to accelerate the sampling procedure in uncertainty quantification.

### 2.3. Uncertainty Estimation in Neural Radiance Fields

Recently, several works have explored the possibility of applying uncertainty estimation to NeRFs. The pioneering work on NeRF-W [30] used an auxiliary volumetric radiance field and a data-dependent uncertainty field to model the transient elements and reduce the impact of transient objects on static scene representations. S-NeRF [31] uses a Bayesian model approach to the probability distribution of all possible radiance fields in the scene and quantifies the information uncertainty provided by the model through this distribution. CF-NeRF [32] combines latent variable modeling and conditional normalizing flows to flexibly learn the distribution of radiance fields in a data-driven manner, thereby obtaining reliable uncertainty estimation and maintaining the expressive power of the model. ActiveNeRF [33] proposes an approach based on an active learning scheme, which selects samples that yield the maximum information gain by evaluating the reduction in uncertainty for the entire scene. In this way, the quality of a novel view synthesis can be improved with minimal additional resource investment. However, the aforementioned NeRF-based methods cannot make inferences explicitly in the absence of knowledge about

unknown scene regions, thus leading to high degree of uncertainty [7]. To address this issue, we introduced an additional measure of uncertainty in the unknown space, i.e., we quantified the epistemic uncertainty of unknown scene regions by learning the probability of ray termination in the geometric domain.

### 3. Scientific Background

#### 3.1. Neural Reflectance Fields

Neural reflectance fields (Figure 1) is a method for representing the reflection properties of object surfaces, and it uses MLP to approximate the albedo  $\alpha = (\alpha_r, \alpha_g, \alpha_b)$ , surface normal  $\mathbf{n} = (n_x, n_y, n_z)$ , and volume density  $\sigma$  at a given 3D position  $\mathbf{x} = (x, y, z)$  that maps with the hash encoding  $\gamma$ . It implicitly represents the interaction of rays with the object’s surface. By combining it with a physically based differentiable ray marching framework, it can accurately model the appearance of real scenes with complex geometry and reflection characteristics. The neural reflectance field is oriented toward arbitrary illumination conditions, improving the MLP’s ability to learn high-frequency information and compensating for the shortcomings of the NeRF that only synthesizes novel views under fixed illumination and does not allow for relighting tasks.

The modeling process of the neural reflectance field is expressed in Equation (1) as:

$$(\sigma, \alpha, \mathbf{n}) = \text{MLP}(\gamma(\mathbf{x})). \tag{1}$$

In the following, we use the function  $f_\theta$  implemented to denote a deep neural network MLP with parameters  $\theta$ :

$$(\sigma_\theta, \alpha_\theta, \mathbf{n}_\theta) = f_\theta(\gamma(\mathbf{x})) \tag{2}$$

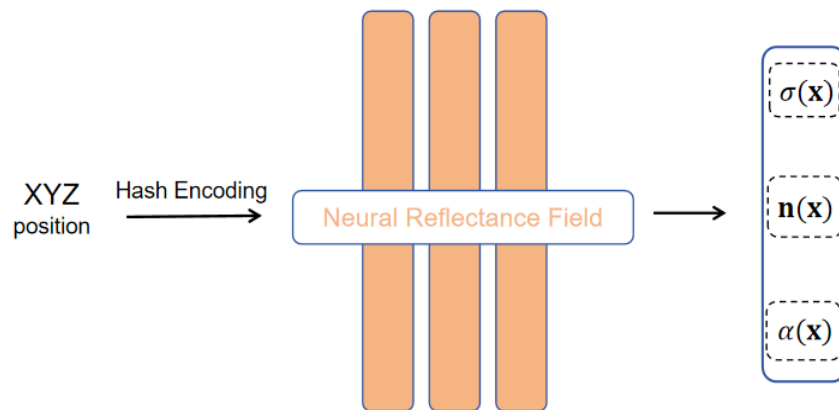


Figure 1. Neural reflectance fields model.

#### 3.2. Beyond NeRFs Underwater (BNU)

As illustrated in Figure 2, BNU considers the effects of optical properties such as attenuation and backscatter on underwater imaging. The radiation  $L_\lambda$  captured by the camera ray  $\mathbf{x} = \mathbf{o} - t\omega$ , which is mapped to the position  $\mathbf{o}$  in the direction  $\omega$ , can be expressed as follows:

$$L_\lambda(\mathbf{x}) = S_\lambda + T_\lambda^n \int_{d_n}^{d_f} T_\lambda(\mathbf{x})\sigma(\mathbf{x})l_\lambda(\mathbf{x})dt, \tag{3}$$

where the integration along the ray is restricted to the near boundary  $d_n$  and the far boundary  $d_f$ ;  $\lambda$  denotes the wavelength;  $S_\lambda$  is the backscatter;  $\sigma(\mathbf{x})$  is the volume density predicted by the MLP at the point  $\mathbf{x}$ ; and  $T_\lambda^n$  is the transmittance between the camera plane and  $d_n$ , which is written as

$$T_\lambda^n = \exp(-\beta_\lambda d_n), \tag{4}$$

where  $\beta_\lambda$  and  $S_\lambda$  stand for the attenuation coefficient and backscatter (which are independent of spatial position throughout the scene reconstruction process), respectively.



Figure 2. BNU model.

The scattered radiance  $l_\lambda$  along the ray from  $\mathbf{x}$  to  $\mathbf{o}$  is expressed as part of the integrand as follows:

$$l_\lambda(\mathbf{x}) = \int_{S^2} I_\lambda(\mathbf{x}, \omega_i) \alpha_\lambda(\mathbf{x}) \cos(\mathbf{n}(\mathbf{x}), \omega_i) d\omega_i, \tag{5}$$

where  $S^2$  represents the spherical domain around point  $\mathbf{x}$  and  $I_\lambda$  is the incident radiance to  $\mathbf{x}$  from direction  $\omega_i$ . The albedo  $\alpha_\lambda(\mathbf{x})$  and normal  $\mathbf{n}(\mathbf{x})$  are the reflection properties at  $\mathbf{x}$  obtained from the neural network prediction.

$T_\lambda(\mathbf{x})$  represents the cumulative transmittance of the ray from  $\mathbf{x}$  to  $d_n$ :

$$T_\lambda(\mathbf{x}) = \exp\left(-\int_{d_n}^t \sigma_\lambda ds\right), \tag{6}$$

where  $\sigma_\lambda(\mathbf{x}) = \sigma(\mathbf{x}) + \beta_\lambda$  denotes the attenuation coefficient.

We used numerical methods to estimate this continuous integral of Equation (3). First,  $[d_n, d_f]$  was divided into  $N$  uniformly distributed intervals through stratified sampling, and then one sample was randomly drawn from each interval as follows:

$$d_i \sim u\left[d_n + \frac{i-1}{N} (d_f - d_n), d_n + \frac{i}{N} (d_f - d_n)\right]. \tag{7}$$

Although we use a set of discrete samples to estimate integration, stratified sampling ensures a continuous scene representation. This is because stratified sampling divides the input space into multiple levels or subregions to comprehensively capture the characteristics of the input space, thus enabling us to evaluate MLP at continuous positions.

We utilized this set of discrete samples to estimate  $L_\lambda$  by using the integration rule discussed in [34] for volume rendering:

$$\hat{L}_\lambda(\mathbf{x}) = S_\lambda + \sum_{i=0}^N T_\lambda^i T_{\lambda_i} (1 - \exp(-\sigma_i \delta_i)) l_i, \tag{8}$$

$$T_{\lambda_i} = \exp\left(-\sum_{j=0}^i \sigma_{\lambda_j} \delta_j\right). \tag{9}$$

Upon substitution of Equation (9) into Equation (8), we have the following:

$$\hat{L}_\lambda(\mathbf{x}) = S_\lambda + \underbrace{\sum_{i=0}^N \prod_{j=0}^{i-1} T_\lambda^n \exp(-\sigma_{\lambda_j} \delta_j)}_{\text{transmittance}} \underbrace{\left( \overbrace{1 - \exp(-\sigma_i \delta_i)}^{\text{termination probability}} \right)}_{\text{occupancy}} \overbrace{l_i}^{\text{color}}. \quad (10)$$

Equation (3) allows us to compute  $L_\lambda$  for all rays  $\mathbf{x}$  passing through the camera center and the image plane to render an image. Next, we will use the symbol  $L_{\theta_k}$  to denote the estimated radiance of the MLP network  $f_{\theta_k}$  along the light ray  $\mathbf{x}$ .

Also, in a similar manner to the neural radiance fields, the above was primarily divided into coarse rendering and refined rendering. The final rendering came from refined rendering, while the coarse rendering was mainly used for loss calculation. The main difference between them was that coarse rendering utilizes the coarse volume density  $\sigma(\mathbf{x})$  predicted directly by the neural network, while refined rendering uses the exact volume density  $m_o(\mathbf{x})\sigma(\mathbf{x})$ , where  $m_o(\mathbf{x}) = \text{sigmoid}(3(\sigma(\mathbf{x}) - 3))$ .

#### 4. Uncertainty Estimation

##### 4.1. Ensembles for Predictive RGB Uncertainty

We fed the same training data into the neural network model  $\{f_{\theta_k}\}_{k=1\dots M}$  in which the radiance was predicted as follows:

$$L_{\theta_k} = S_\lambda + T_\lambda^n \int_{d_n}^{d_f} T_{\theta_k}(\mathbf{x}) \sigma_{\theta_k} l_{\theta_k}(\mathbf{x}) dt, \quad (11)$$

where

$$T_{\theta_k}(\mathbf{x}) = \exp\left(-\int_{d_n}^t (\sigma_{\theta_k} + \beta_\lambda) ds\right), \quad (12)$$

$$l_{\theta_k} = \int_{S^2} I_\lambda(\mathbf{x}, \omega_i) \alpha_{\theta_k}(\mathbf{x}) \cos(\mathbf{n}_{\theta_k}(\mathbf{x}), \omega_i) d\omega_i. \quad (13)$$

As can be seen through Equation (2), our ensemble model is divided into three main parts: the albedo field, the surface normal field, and the volume density field.

$$(\sigma_{\theta_k}, \alpha_{\theta_k}, \mathbf{n}_{\theta_k}) = f_{\theta_k}(\gamma(\mathbf{x})), \quad (14)$$

where  $k = 1 \dots M$  denotes the number of ensemble model.

By averaging the color of each pixel during the rendering process, the expected radiation of the camera ray  $\mathbf{x}$  is

$$\mu_{\text{RGB}}(\mathbf{x}) = \frac{1}{M} \sum_{k=1}^M L_{\theta_k}(\mathbf{x}). \quad (15)$$

We denote the prediction uncertainty in RGB space as the variance of individual network predictions:

$$\sigma_{\text{RGB}}^2(\mathbf{x}) = \frac{1}{M} \sum_{k=1}^M (\mu(\mathbf{x}) - L_{\theta_k}(\mathbf{x}))^2. \quad (16)$$

$\mu_{\text{RGB}}$  or  $\sigma_{\text{RGB}}$  has three components on the RGB color channels. To simplify the calculations, we assumed that the three color channels R, G, and B are independent of each other, and this was achieved by disregarding their inter-channel relationships and only retaining the primary information. In addition, for the convenience of subsequently combining color variance with other uncertainty terms such as  $\sigma_{\text{epi}}^2(\mathbf{x})$  into a unified uncertainty metric

$\psi^2(\mathbf{x})$ , we merged the variances of the three color channels and used a single value to represent the overall RGB uncertainty.

$$\bar{\sigma}_{\text{RGB}}^2(\mathbf{x}) = \frac{1}{3} \cdot \sum_{L \in \{\text{RGB}\}} \sigma_{\text{RGB},(L)}^2(\mathbf{x}). \tag{17}$$

#### 4.2. Ensembles for Epistemic Uncertainty in Unseen Areas

Using the expectation and variance of RGB values in color space to quantify prediction uncertainty is a simple and partially effective approach. However, these values only reflect the uncertainty of color predictions. They cannot measure cognitive uncertainty about unobserved scenes during training.

From our observations, it was found that since the prediction model lacks relevant data to learn features of unobserved regions during the training process, it does not know the exact shape and color of those regions; thus, it is not able to provide meaningful termination probability predictions. As a result, the model assigns very low termination probabilities to each sample point on ray  $\mathbf{x}$ , and the cumulative sum  $q_{\theta_k}(\mathbf{x})$  of termination probabilities along the ray approaches zero [13]. Hence, we can consider the termination probabilities along the ray  $\mathbf{x}$  as a way for the model to express cognitive uncertainty about the unknown regions of the scene.

$$q_{\theta_k}(\mathbf{x}) = \sum_{i=0}^N \underbrace{\prod_{j=0}^{i-1} T_{\lambda}^n \exp(-\sigma_{\lambda_j} \delta_j)}_{\text{transmittance}} \underbrace{(1 - \exp(-\sigma_i \delta_i))}_{\text{occupancy}} \approx 0. \tag{18}$$

termination probability at sample  $i$

We averaged the termination probabilities along ray  $\mathbf{x}$  across the entire ensemble model as follows:

$$\bar{q}(\mathbf{x}) = \frac{1}{M} \sum_{k=1}^M q_{\theta_k}(\mathbf{x}), \tag{19}$$

where  $\bar{q}(\mathbf{x}) \approx 1$  indicates that the ray intersects with the observed scene structure during training, and  $\bar{q}(\mathbf{x}) \approx 0$  is such if it did not.

In order to capture the prediction uncertainty along ray  $\mathbf{x}$  as comprehensively as possible, we expressed the total uncertainty as a combination of the RGB variance  $\bar{\sigma}_{\text{RGB}}$  and a cognitive uncertainty term  $\sigma_{\text{epi}}$  as follows:

$$\psi^2(\mathbf{x}) = \bar{\sigma}_{\text{RGB}}^2(\mathbf{x}) + \sigma_{\text{epi}}^2(\mathbf{x}), \tag{20}$$

where

$$\sigma_{\text{epi}}^2(\mathbf{x}) = (1 - \bar{q}(\mathbf{x}))^2. \tag{21}$$

Moreover, the predicted colors of the ensemble along ray  $\mathbf{x}$  were modeled as Gaussian distributions with diagonal covariance matrices:

$$\tilde{L}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}_{\text{RGB}}(\mathbf{x}), \mathbf{I}_{3 \times 3} \cdot \psi^2(\mathbf{x})). \tag{22}$$

Simply using the expectation and variance of RGB values in color space would overlook the cognitive knowledge about unobserved scenes during training. However, the termination probability precisely reflects the uncertainty of this process. Therefore, the uncertainty terms  $\bar{\sigma}_{\text{RGB}}^2(\mathbf{x})$  and  $\sigma_{\text{epi}}^2(\mathbf{x})$  were found to be complementary in that they capture the arbitrary and cognitive uncertainty in different aspects of the model.

Additionally, as indicated in this section, volume density was identified as a significant factor influencing uncertainty in our research. Rendering with coarse volume density  $\sigma(\mathbf{x})$  and exact volume density  $m_o(\mathbf{x})\sigma(\mathbf{x})$  yields different  $\bar{\sigma}_{\text{RGB}}^2(\mathbf{x})$  and  $\sigma_{\text{epi}}^2(\mathbf{x})$  results,

respectively. Therefore, it is necessary to conduct separate experiments for both coarse rendering and refined rendering in a systematic study.

### 5. Numerical Experiment

#### 5.1. Experimental Setup

##### 5.1.1. Dataset

We performed experiments on the scenes introduced in [2]. This process included the synthetic dataset and real dataset. The synthetic images with underwater optical effects were simulated based on the Jaffe–McGlamery model [35]. Ref. [2] used a Sony ILCE-7M3 camera with a 40 mm prime lens and LED lights to collect the real images in a tank at a water depth of 1.3 m. Additionally, the camera poses were simultaneously acquired via COLMAP [36], and the JPEG images were post-processed to ensure high feature quality.

##### 5.1.2. Framework

Our ensemble model primarily consists of three components: the normal field, the density field, and the albedo field (Figure 3). For the choice of  $M$ , we empirically selected  $M = 3$ . The ensemble model sampled 100 rays from an image and 100 points on each ray at each training iteration. The model was trained for 5000 epochs for each scene.

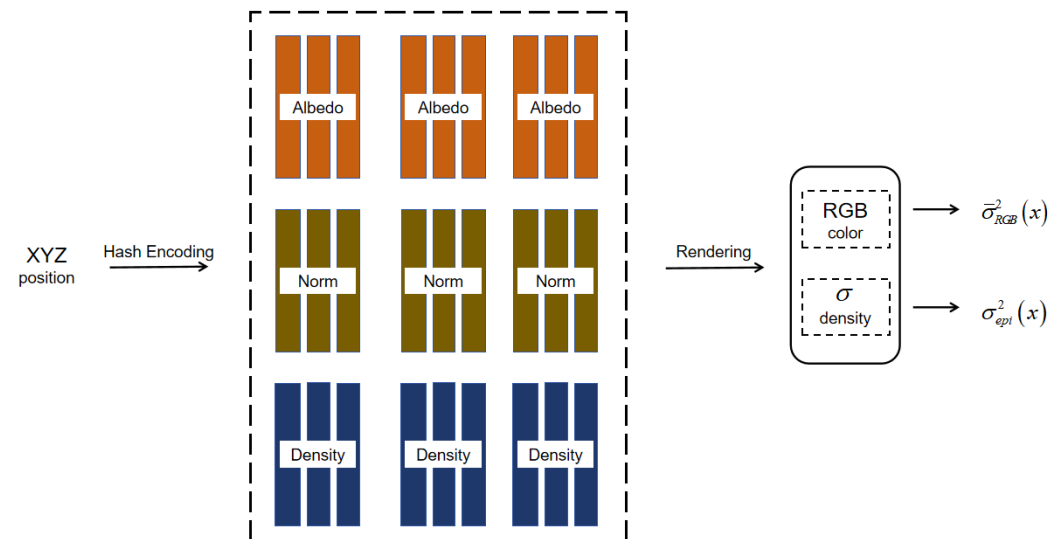


Figure 3. Ensemble model.

##### 5.1.3. Metrics

Herein, we use two main types of metrics: quality metrics and uncertainty quantification. For the quality metrics, we computed the mean absolute error (MAE), mean-square error (MSE), and root-mean-square error (RMSE) to evaluate the reconstruction error, and the peak signal-to-noise ratio (PSNR) was used to characterize the quality of the rendered views.

- The MAE directly calculates the average absolute error between the predicted value of the model and the ground truth. The smaller the MAE value, the more accurate the prediction.

$$MAE(Y_i, f(x_i)) = \frac{1}{n} \sum_{i=0}^{n-1} |Y_i - f(x_i)|. \tag{23}$$

- The MSE is calculated by squaring the differences between the ground truth and predicted values, summing them up, and then taking the average.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - f(x_i))^2. \tag{24}$$



- The RMSE measures the deviation between the predicted values and ground truth, and it is sensitive to outliers in the data.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - f(x_i))^2}. \tag{25}$$

- The PSNR is a metric used to measure image quality. Here,  $\text{MAX}_x$  represents the maximum pixel value of image  $x$ .

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{MAX}_x^2}{\text{MSE}} \right). \tag{26}$$

We used the predicted values obtained through experiments to evaluate the quality of reconstruction. It was evident that the evaluation results of the quality metrics were correlated with the accuracy of the predicted values.

For uncertainty quantification, to assess the reliability of the model prediction uncertainty, we reported two widely used metrics. First, we used the negative log likelihood (NLL) as an evaluation criterion [37]. The using NLL was chosen because our ensemble model provides a Gaussian distribution for predicting rendered colors along rays, where the expectation is equivalent to the estimated color and the variance is a combination of uncertainty measures based on RGB variances and cognitive metrics. In addition, to assess the correlation between prediction error and uncertainty estimates, we reported the area under the sparsification error (AUSE) curve [38–40].

- The log likelihood function was defined as follows:

$$\ln L(\theta|x) = \ln P(x|\theta), \tag{27}$$

where  $\theta$  is the unknown parameters,  $x$  is the observed sample, and  $P(x|\theta)$  is the probability distribution of  $x$  given  $\theta$ . The negative log likelihood (NLL) is defined as the negative of the log likelihood function as follows:

$$\text{NLL}(\theta|x) = -\ln L(\theta|x) = -\ln P(x|\theta). \tag{28}$$

From Section 4.2, the sample of observations  $x$  follows a Gaussian distribution as follows:

$$P(x|\mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2), \tag{29}$$

with the probability density function as

$$p(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} * \exp\left[-(x - \mu)^2 / (2\sigma^2)\right]. \tag{30}$$

Then, the NLL of this Gaussian distribution is the following:

$$\begin{aligned} \text{NLL}(\mu, \sigma^2|x) &= -\ln P(x|\mu, \sigma^2) \\ &= 0.5 \ln(2\pi\sigma^2) + 0.5(x - \mu)^2 / \sigma^2. \end{aligned} \tag{31}$$

- For the AUSE, the prediction error (e.g., RMSE) for each pixel was firstly computed based on the predicted values and ground truth. The uncertainty values obtained from training and the prediction errors for all pixels were merged and sorted. The top 1% of data were removed from the sorted list, and the average error and average uncertainty of the remaining data were treated as the point at 1%. Similarly, the top 2% of data were removed; then, the point at 2% was calculated, and this was so on performed until it reached 100%. This process generates two curves: the prediction error curve and the uncertainty curve. The area enclosed by the two curves is the

AUSE value. A lower AUSE value indicates a higher correlation between uncertainty estimation and true error, thus implying a more reliable uncertainty estimation.

We quantified the uncertainty by means of experimentally obtained predicted values, as well as with  $\bar{\sigma}_{RGB}^2(\mathbf{x})$  and  $\sigma_{epi}^2(\mathbf{x})$ . Based on the above discussion, we could infer that the uncertainty quantification results of the numerical experiments were closely related to the accuracy of these values.

### 5.2. Results

We show the main results of our experiments in Table 1. Our ensemble strategy achieved excellent NLL and AUSE on both synthetic and real datasets. Moreover, it exhibited outstanding performance in terms of image reconstruction error and rendering quality.

**Table 1.** The results of our ensemble model.

			MSE	RMSE	MAE	PSNR	AUSE RMSE	AUSE MAE	NLL
Ours	Synthetic	Coarse	0.0004	0.019	0.0127	35.1	0.0015	0.0015	$-0.0116 \pm 0.0628$
		Refined	0.0004	0.019	0.0127	35.1	0.0014	0.0013	$0.1125 \pm 0.0923$
	Real	Coarse	0.0006	0.021	0.0161	34.1	0.0051	0.0043	$0.5255 \pm 0.103$
		Refined	0.0006	0.021	0.0162	34.1	0.0028	0.0026	$0.5615 \pm 0.0995$

### 5.3. Ablation Study

#### 5.3.1. Influence of the Ensemble

Entire Ensemble Model: As can be seen from Table 2, we outperformed previous works in terms of image reconstruction error and rendering quality.

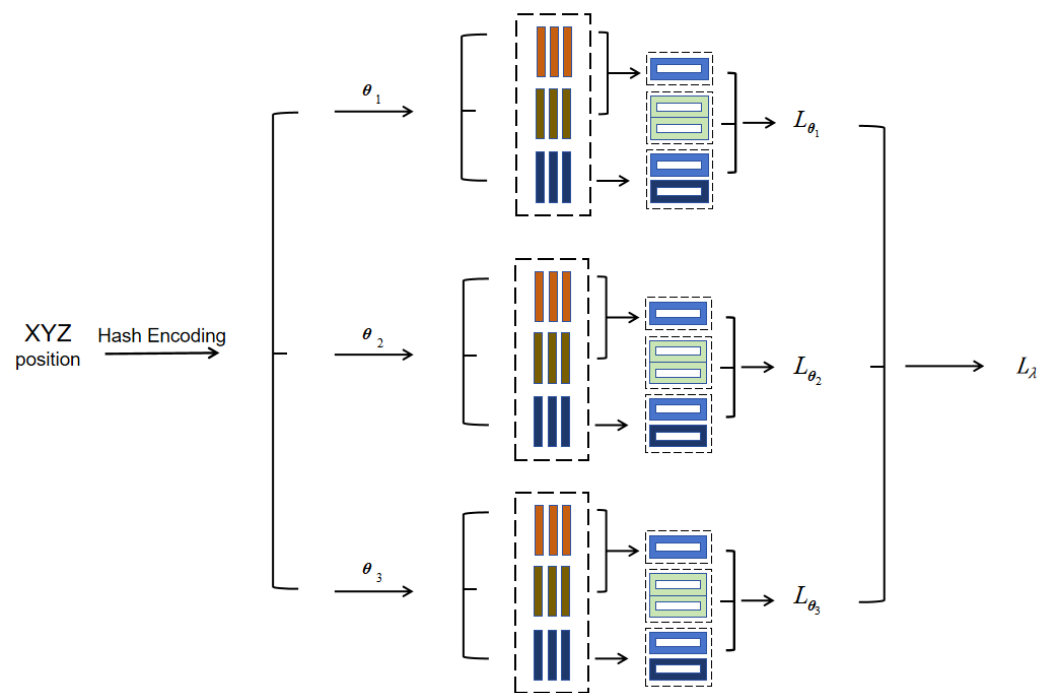
**Table 2.** Influence of the entire ensemble model.

			MSE	RMSE	MAE	PSNR
BNU	Synthetic	Coarse	0.0005	0.02	0.0133	34.6
		Refined	0.0005	0.02	0.0133	34.6
	Real	Coarse	0.0006	0.022	0.0167	33.8
		Refined	0.0006	0.022	0.0167	33.8
Ours	Synthetic	Coarse	0.0004	0.019	0.0127	35.1
		Refined	0.0004	0.019	0.0127	35.1
	Real	Coarse	0.0006	0.021	0.0161	34.1
		Refined	0.0006	0.021	0.0162	34.1

An individually trained network can only learn the information contained in a limited number of training samples, whereas the ensemble model learns features from the expression of features learned through multiple networks with different initialization; as such, it can express richer information.

On the other hand, the ensemble model obtains the final prediction result by averaging the multiple networks (Figure 4), which is equivalent to sampling the dataset several times and then averaging the data. Statistically, after averaging multiple quantification errors, the average error will be smaller or equal to the single quantification error, which means that this horizontal averaging can offset the effect of the larger prediction errors of individual networks to a certain extent.

In short, by using multiple networks to predict the results, the overall prediction error can be reduced to a certain degree, and this improves the quality of image reconstruction so that the PSNR and other metrics can obtain better results than a single model.



**Figure 4.** Ensemble prediction framework. The physical interpretation represented by each color is the same as in Figure 2.

Individual Ensemble Component: As shown in Table 3, we experimented with the results of implementing the ensemble strategy for individual ensemble components (Figure 5).

**Table 3.** Influence of individual ensemble components.

		MSE	RMSE	MAE	PSNR			MSE	RMSE	MAE	PSNR	
BNU	Synthetic	Coarse	0.000491529	0.0200459	0.013337525	34.63339233	Albedo	Coarse	0.000502874	0.020200016	0.013464036	34.58008575
		Refined	0.000492371	0.020073349	0.013330285	34.61220551		Refined	0.000503786	0.020222723	0.013453723	34.56002045
	Real	Coarse	0.000599609	0.022026947	0.01668788	33.81985092		Coarse	0.000589828	0.02165741	0.016421406	33.97740173
		Refined	0.000602939	0.022051068	0.016739015	33.80893326		Refined	0.000590084	0.02164993	0.01644345	33.97847366
Norm	Synthetic	Coarse	0.000459542	0.019584062	0.012990904	34.78876877	Density	Coarse	0.000518398	0.020231744	0.01353504	34.62197495
		Refined	0.000460262	0.019598676	0.012975736	34.77446747		Refined	0.000519	0.020276753	0.013544992	34.58899689
	Real	Coarse	0.000566563	0.021291357	0.016084474	34.09726334		Coarse	0.000555657	0.02125203	0.016046988	34.08520126
		Refined	0.000567068	0.021277852	0.016084943	34.100914		Refined	0.000556773	0.021272138	0.016073255	34.07310486

Firstly, normals are crucial for capturing the surface details and geometric structure of objects. By ensembling multiple norm fields, it is possible to better reproduce the details and shapes of objects in scenes. This allows it to outperform the BNU model in terms of performance on both synthetic and real datasets as it can more accurately capture the appearance characteristics of objects.

Furthermore, by ensembling multiple albedo fields, it is possible to enhance the model’s ability to capture lighting and material variations. However, due to the characteristics of the synthetic data, the contribution of the albedo may be relatively small, thus resulting in a less significant impact from the ensemble of the albedo field, as well as a slightly lower performance when compared to BNU. On the other hand, the contribution of albedo was found to be greater on the real dataset, which captured the appearance features of the object more accurately and therefore outperformed BNU on the real dataset.

Then, the density field plays a key role in the propagation of light within a scene. By ensembling multiple density fields, the effect of light propagation in a scene can be better simulated, thereby providing more accurate rendering results. Since synthetic datasets usually have more idealized lighting conditions, geometric structures, and material



**Table 4.** The uncertainty terms' influence on the entire ensemble framework.

		$\bar{\sigma}_{\text{RGB}}^2(x)$			$\sigma_{\text{epi}}^2(x)$			$\psi^2(x) = \bar{\sigma}_{\text{RGB}}^2(x) + \sigma_{\text{epi}}^2(x)$		
		NLL	AUSE RMSE	AUSE MAE	NLL	AUSE RMSE	AUSE MAE	NLL	AUSE RMSE	AUSE MAE
Synthetic	Coarse	475.5226 ± 15068.98	0.002254584	0.001743699	−0.34395203 ± 0.17283306	0.001247811	0.00123583	−0.011556505 ± 0.062817425	0.001504149	0.001509211
	Refined	474.00775 ± 15063.435	0.002261687	0.001751348	−0.21837942 ± 0.217421	0.002945993	0.002841848	0.11246947 ± 0.09231803	0.001368612	0.001328661
Real	Coarse	1207.923 ± 8092	0.002890238	0.002550809	0.32225832 ± 0.12907358	0.002272962	0.001563942	0.52552766 ± 0.10300194	0.005134132	0.00426048
	Refined	482.2294 ± 4909.587	0.002837191	0.002513715	0.35704166 ± 0.14970568	0.000311964	0.000618934	0.5615406 ± 0.099535696	0.002829015	0.0026005

**Table 5.** The uncertainty terms' influence on the individual ensemble components.

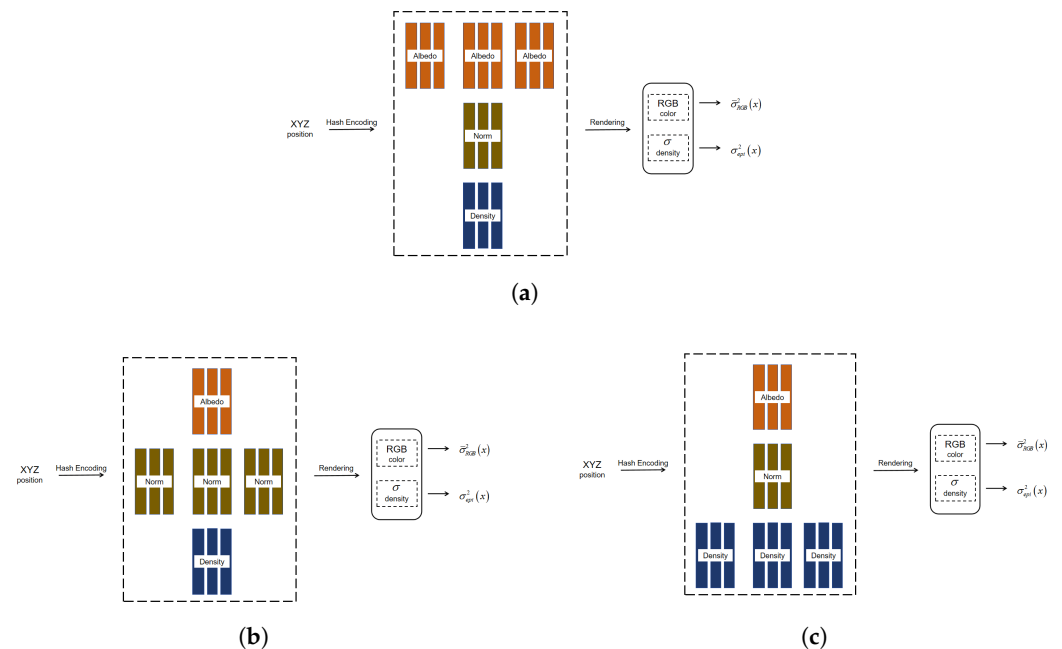
		$\bar{\sigma}_{\text{RGB}}^2(x)$			$\sigma_{\text{epi}}^2(x)$			$\psi^2(x) = \bar{\sigma}_{\text{RGB}}^2(x) + \sigma_{\text{epi}}^2(x)$			
		NLL	AUSE RMSE	AUSE MAE	NLL	AUSE RMSE	AUSE MAE	NLL	AUSE RMSE	AUSE MAE	
Normal	Synthetic	Coarse	450.82037 ± 4871.639	0.000249919	0.000233828	−0.6259125 ± 0.20818105	0.000698834	0.000657816	−0.39603126 ± 0.11323436	0.00241437	0.002293217
		Refined	289.39508 ± 3605.0972	0.00025749	0.000241996	−0.6097659 ± 0.20786819	0.000721777	0.000680518	−0.3866958 ± 0.11498752	0.002479377	0.002356168
	Real	Coarse	1610.0344 ± 8977.543	0.000125807	0.000121349	−0.5150174 ± 0.25648573	0.000450089	0.000362964	−0.38426322 ± 0.14627357	0.000507562	0.000392292
		Refined	968.5724 ± 7319.2705	0.000117436	0.000112676	−0.4824918 ± 0.25163838	0.000414992	0.00030181	−0.35871175 ± 0.1483912	0.000457031	0.000314412
Albedo	Synthetic	Coarse	1070.254 ± 6610.523	0.000510919	0.000523254	−0.6515078 ± 0.18188678	0.000651663	0.000609364	−0.27608374 ± 0.062299836	0.002779543	0.002610347
		Refined	1282.3138 ± 7401.9	0.000609278	0.000620129	−0.6313605 ± 0.18285266	0.000746449	0.000704222	−0.2656904 ± 0.06084825	0.002868676	0.002695229
	Real	Coarse	1639.0918 ± 9054.044	0.000348316	0.000300185	−0.37690923 ± 0.24971825	0.000946304	0.00065761	0.01580802 ± 0.09670887	0.000669593	0.000235055
		Refined	693.6819 ± 5697.16	0.000376567	0.000328893	−0.3352108 ± 0.24189745	0.001014562	0.000690136	0.03544719 ± 0.09218908	0.000787288	0.000305379
Density	Synthetic	Coarse	2052.7996 ± 9721.339	0.002614084	0.002074688	−0.5209125 ± 0.12687543	0.000790859	0.000755849	−0.18053763 ± 0.076944746	0.002355352	0.00182316
		Refined	1352.4849 ± 7712.8203	0.002602652	0.002061267	−0.47604796 ± 0.13382044	0.001606671	0.001540226	−0.13651294 ± 0.076657325	0.001342092	0.001227694
	Real	Coarse	1652.699 ± 8296.807	1.37 × 10 <sup>−5</sup>	9.67 × 10 <sup>−6</sup>	−0.09061173 ± 0.21988438	0.000879356	0.00048239	0.052660644 ± 0.10419571	0.000483531	0.000182983
		Refined	1821.3926 ± 8670.74	2.85 × 10 <sup>−5</sup>	5.22 × 10 <sup>−6</sup>	−0.04039674 ± 0.2547284	0.002096198	0.001341577	0.106006674 ± 0.1293263	0.000180225	0.000390491

Entire Ensemble Model: From Table 4, we can observe that using only  $\bar{\sigma}_{RGB}^2(\mathbf{x})$  as an uncertainty measure results in significantly higher NLL, as well as that metrics such as AUSE are inferior to other cases. This is because color variance in high-frequency regions is strongly influenced by individual pixel outliers. A high deviation in a single pixel can lead to high variance across the entire image, especially in unobserved scenes during training. In contrast, the cognitive uncertainty term  $\sigma_{epi}^2(\mathbf{x})$  achieves excellent results in most scenarios. This is attributed to the fluctuations in individual  $q_{\theta_k}(\mathbf{x})$ , where the ensemble model assigns slightly lower  $q_{\theta_k}(\mathbf{x})$  to rays with higher uncertainty.

Using  $\sigma_{epi}^2(\mathbf{x})$  allows us to differentiate the model’s understanding level between known and unknown regions, effectively capturing cognitive uncertainty in unobserved areas. By comparison with the color variance, termination probability can directly reflect the model’s judgment on the entire ray entering the scene. Hence, it is less affected by individual outliers and more stable reflecting the model’s knowledge level.

The reason for considering the combination of both components, rather than relying on a single one, is that color variance  $\bar{\sigma}_{RGB}^2(\mathbf{x})$  can effectively capture noise uncertainty that is difficult to model caused by factors like sampling and occlusion in known regions. Additionally, in simple scenes with homogeneous color distributions, color variance can better express uncertainty in known regions and can be directly computed in the image pixel space. On the other hand, for unknown regions, utilizing the uncertainty term  $\sigma_{epi}^2(\mathbf{x})$  can supplement the cognitive uncertainty information that cannot be expressed by the color variance well. By combining these two uncertainty terms, a more comprehensive and accurate uncertainty judgment is provided.

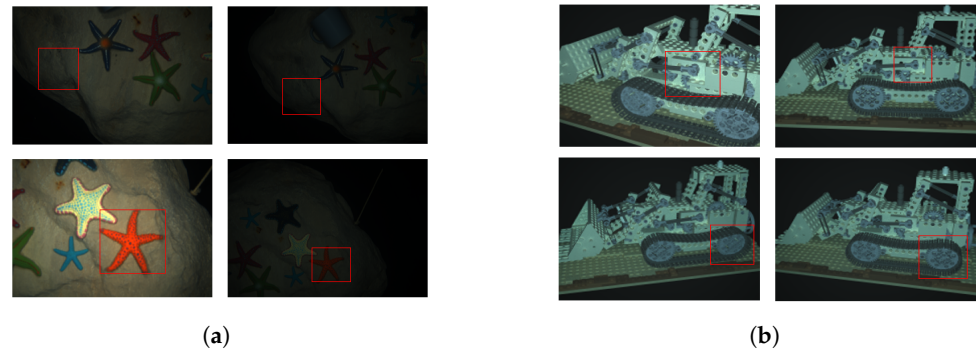
Individual Ensemble Component Frameworks: We experimented with individual ensemble component frameworks (Figure 6). According to Table 5, we can observe that the inferring ability of the uncertainty measure  $\psi^2(\mathbf{x})$  exhibits different performances on the synthetic and real datasets when we focus on a specific ensemble component. Intuitively, performance degradation was observed on the synthetic dataset while the performance enhancement was observed on the real dataset.



**Figure 6.** Uncertainty estimation framework of the individual ensemble components: (a) Ensemble albedo field. (b) Ensemble normal field. (c) Ensemble density field.

This difference stems from the complexity of the data itself. Synthetic data scenes are relatively simple with fewer lighting variations. The multi-component ensemble uses multiple sources of information for greater expressiveness, thus leading to more pronounced

advantages. In contrast, in a real dataset, due to the complexity of the data such as in large light variations, the inference capability of uncertainty makes it more dependent on individual components. To provide further illustration, we present the impact of lighting variations on the appearance at the same location in both synthetic and real data in Figure 7. From Figure 7, we can observe that the lighting variations have a significant impact on appearance in the real dataset, while the effect is relatively smaller in the synthetic dataset. A multi-component ensemble approach requires learning resources to be spread evenly across multiple tasks, thus making it difficult to optimize for a particular component in more depth. Therefore, focusing on a specific ensemble component allows for better adaptation to the complex distribution of real data.



**Figure 7.** Real and synthetic images of lighting variations with the same locations being marked by red boxes: (a) Real. (b) Synthetic.

In addition, through observation, we can notice that, when ensembling multiple density fields, the uncertainty measure  $\psi^2(\mathbf{x})$  exhibits a superior performance compared to ensembling multiple norm fields or multiple albedo fields. This is because the termination probability is used to determine whether the propagation of light in the scene is terminated or not. And the termination probability is closely related to density as density reflects the extent of the presence of objects in the scene. By considering the relationship between termination probability and density, the model can better estimate the light propagation within the scene and capture the degree of object presence, thereby effectively handling the uncertainty of the unobserved parts of the scene and quantifying this uncertainty.

#### 5.4. Discussion

From the above study, we can see that the fluctuation caused by a single  $q_{\theta_k}(\mathbf{x})$  can make the cognitive uncertainty term  $\sigma_{\text{epi}}^2(\mathbf{x})$  obtain very good results in most scenes, and the fluctuation caused by a single  $q_{\theta_k}(\mathbf{x})$  can be traced back to the fluctuation caused by a single density. In other words,  $\sigma_{\text{epi}}^2(\mathbf{x})$  depends on a single density prediction along each ray  $\mathbf{x}$ .

However, during the model training process, the density predictions tended to have large oscillations or jumps for various reasons. It was also found that our uncertainty estimation framework is highly sensitive to density, thus making it more likely to capture uncertainty in regions with high oscillations. Considering our goal was to assess the reliability of uncertainty quantification models, we experimented to artificially reduce the density amplitude and fluctuations in the ray sampling region by smoothing the density curve around the camera (Table 6).

We used a regularization method [14] that penalizes density fields in the vicinity of the camera:

$$\mathcal{L}_{occ} = \frac{\sigma_K^\top \cdot \mathbf{n}_K}{K} = \frac{1}{K} \sum_K \sigma_k \cdot n_k, \tag{32}$$

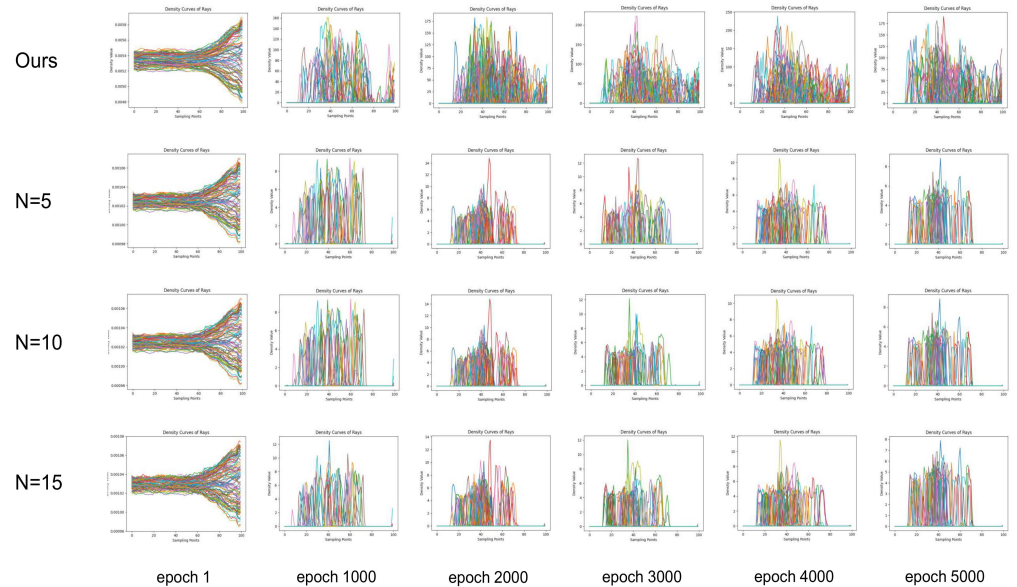
where  $\sigma_K$  denotes the density value of the  $K$  points sampled along the ray  $\mathbf{x}$  from the origin and  $\mathbf{n}_K$  is the binary mask vector that determines whether a sampled point will be penalized or not. We set the value of  $\mathbf{n}_K$  to 1 before the regularization range  $N$ , and we also set the rest of the values to 0. In addition, we used a weight of 0.01 in all experiments.

**Table 6.** Results for the density amplitude and fluctuation interference.

		N = 5			N = 10			N = 15		
		NLL	AUSE RMSE	AUSE MAE	NLL	AUSE RMSE	AUSE MAE	NLL	AUSE RMSE	AUSE MAE
Synthetic	Coarse	$-0.06423895 \pm 0.077842504$	0.012891437	0.011193172	$-0.06423895 \pm 0.077842504$	0.012891437	0.011193172	$-0.06081396 \pm 0.07654448$	0.01272295	0.011058994
	Refined	$-0.06856838 \pm 0.07776483$	0.011792897	0.010343174	$-0.06856838 \pm 0.07776483$	0.011792897	0.010343174	$-0.0646455 \pm 0.0769008$	0.011601113	0.010205722
Real	Coarse	$0.674563 \pm 0.1071525$	0.00481758	0.004217093	$0.6660141 \pm 0.11203171$	0.005699564	0.004884685	$0.674563 \pm 0.1071525$	0.00481758	0.004217093
	Refined	$0.6742599 \pm 0.10600442$	0.004616567	0.003919621	$0.49425298 \pm 0.10282664$	0.004663244	0.00389921	$0.6742599 \pm 0.10600442$	0.004616567	0.003919621



We conducted the above experiment for 100 rays on the same image (Figure 8), and it can be seen that our uncertainty estimation framework performed well even when the density amplitude and fluctuations were artificially reduced (which is sufficient to demonstrate its ability to infer uncertainty).



**Figure 8.** We experimented with 100 rays, which are represented as curves of different colors in the figure, in addition to the vertical axis indicating the density values corresponding to the sampling points, and the horizontal axis indicating the sampling points on the rays. By observation we can find that without the regularization term, the maximum value reached more than 200, but, with the regularization term, the maximum value did not exceed 16.

## 6. Conclusions

In this paper, we introduced an ensemble approach to quantify uncertainty in underwater neural scene representation models with reflection properties. It can quantify prediction uncertainty in RGB space by modeling ensemble rendering errors in the scene and identifying the cognitive uncertainty caused by the lack of cognitive knowledge about unknown scenes through an additional cognitive uncertainty term based on ray termination density. Furthermore, we used a method to artificially reduce its density amplitude to test its ability to capture subtle uncertainties and validate the reliability of the model. The numerical experiments demonstrated that our ensemble model can explicitly infer uncertainty in the model on both synthetic and real scenes, thereby exhibiting superior performance and outperforming previous works in key metrics related to reconstruction error and rendering quality. The proposed algorithm will benefit the downstream tasks of ocean exploration and navigation, such as the automatic identification of damage in underwater infrastructure [18,19], target detection and tracking, mapping and motion planning, etc. The present work is not without limitations. For example, our model assumes that the light source is from onboard light, which is only valid for deep sea conditions. For shallow sea conditions, however, natural light comes into play and should be considered in the model. We will investigate this problem in depth in future work.

**Author Contributions:** Conceptualization, H.L., L.C. and Y.Q.; Methodology, H.L., L.C. and Y.Q.; Software, H.L. and X.L.; Validation, X.L., X.W. and M.Z.; Formal Analysis, X.L., X.W. and J.Z.; Investigation, H.L., X.L. and M.Z.; Resources, X.W. and J.Z.; Data Curation, X.L., M.Z. and J.Z.; Writing—Original Draft Preparation, H.L., X.L. and Y.Q.; Writing—Review & Editing, H.L., X.L. and L.C.; Visualization, X.L.; Supervision, H.L., L.C. and Y.Q.; Project Administration, H.L., L.C. and Y.Q.;

Funding Acquisition, H.L. and J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (NSFC) (grant Nos. 62102444, 52274222, and 62206196).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [\[CrossRef\]](#)
- Zhang, T.; Johnson-Roberson, M. Beyond NeRF Underwater: Learning neural reflectance fields for true color correction of marine imagery. *IEEE Robot. Autom. Lett.* **2023**, *8*, 6467–6474. [\[CrossRef\]](#)
- Bi, S.; Xu, Z.; Srinivasan, P.; Mildenhall, B.; Sunkavalli, K.; Hašan, M.; Hold-Geoffroy, Y.; Kriegman, D.; Ramamoorthi, R. Neural reflectance fields for appearance acquisition. *arXiv* **2020**, arXiv:2008.03824.
- Pairet, È.; Hernández, J.D.; Carreras, M.; Petillot, Y.; Lahijanian, M. Online mapping and motion planning under uncertainty for safe navigation in unknown environments. *IEEE Trans. Autom. Sci. Eng.* **2021**, *19*, 3356–3378. [\[CrossRef\]](#)
- Melo, J. AUV position uncertainty and target reacquisition. In Proceedings of the Global Oceans 2020: Singapore–US Gulf Coast, Biloxi, MS, USA, 5–30 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
- Pairet, È.; Hernández, J.D.; Lahijanian, M.; Carreras, M. Uncertainty-based online mapping and motion planning for marine robotics guidance. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2367–2374.
- Shen, J.; Ren, R.; Ruiz, A.; Moreno-Noguer, F. Estimating 3D uncertainty field: Quantifying uncertainty for neural radiance fields. *arXiv* **2023**, arXiv:2008.03824.2311.01815.
- 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. *Inf. Fusion* **2021**, *76*, 243–297. [\[CrossRef\]](#)
- MacKay, D.J. A practical bayesian framework for backpropagation networks. *Neural Comput.* **1992**, *4*, 448–472. [\[CrossRef\]](#)
- Neal, R.M. *Bayesian Learning for Neural Networks*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 118.
- Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.
- Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6402–6413.
- Sünderhauf, N.; Abou-Chakra, J.; Miller, D. Density-aware NeRF ensembles: Quantifying predictive uncertainty in neural radiance fields. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), London, UK, 29 May–2 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 9370–9376.
- 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, Vancouver, BC, Canada, 17–24 June 2023; pp. 8254–8263.
- Sethuraman, A.V.; Ramanagopal, M.S.; Skinner, K.A. WaterNeRF: Neural radiance fields for underwater scenes. In Proceedings of the OCEANS 2023-MTS/IEEE US Gulf Coast, Biloxi, MS, USA, 25–28 September 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–7.
- Barron, J.T.; Mildenhall, B.; Verbin, D.; Srinivasan, P.P.; Hedman, P. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5470–5479.
- Levy, D.; Peleg, A.; Pearl, N.; Rosenbaum, D.; Akkaynak, D.; Korman, S.; Treibitz, T. SeaThru-NeRF: Neural radiance fields in scattering media. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 56–65.
- Orinaitė, U.; Palevičius, P.; Pal, M.; Ragulskis, M. A deep learning-based approach for automatic detection of concrete cracks below the waterline. *Vibroeng. Procedia* **2022**, *44*, 142–148. [\[CrossRef\]](#)
- Orinaitė, U.; Karaliūtė, V.; Pal, M.; Ragulskis, M. Detecting underwater concrete cracks with machine learning: A clear vision of a murky problem. *Appl. Sci.* **2023**, *13*, 7335. [\[CrossRef\]](#)
- Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K.Q. On calibration of modern neural networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1321–1330.

21. Hernández-Lobato, J.M.; Adams, R. Probabilistic backpropagation for scalable learning of bayesian neural networks. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 1861–1869.
22. Neapolitan, R.E. Learning bayesian networks. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007.
23. Aralikatti, R.; Margam, D.; Sharma, T.; Abhinav, T.; Venkatesan, S.M. Global SNR estimation of speech signals using entropy and uncertainty estimates from dropout networks. *arXiv* **2018**, arXiv:1804.04353.
24. Hernández, S.; Vergara, D.; Valdenegro-Toro, M.; Jorquera, F. Improving predictive uncertainty estimation using dropout–Hamiltonian Monte Carlo. *Soft Comput.* **2020**, *24*, 4307–4322. [[CrossRef](#)]
25. Blum, A.; Haghtalab, N.; Procaccia, A.D. Variational dropout and the local reparameterization trick. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–10 December 2015; pp. 2575–2583.
26. Jain, S.; Liu, G.; Mueller, J.; Gifford, D. Maximizing overall diversity for improved uncertainty estimates in deep ensembles. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 4264–4271.
27. Chen, L.; Cheng, R.; Li, S.; Lian, H.; Zheng, C.; Bordas, S.P.A. A sample-efficient deep learning method for multivariate uncertainty qualification of acoustic–vibration interaction problems. *Comput. Methods Appl. Mech. Eng.* **2022**, *393*, 114784. [[CrossRef](#)]
28. Chen, L.; Lian, H.; Xu, Y.; Li, S.; Liu, Z.; Atroshchenko, E.; Kerfriden, P. Generalized isogeometric boundary element method for uncertainty analysis of time-harmonic wave propagation in infinite domains. *Appl. Math. Model.* **2023**, *114*, 360–378. [[CrossRef](#)]
29. Chen, L.; Wang, Z.; Lian, H.; Ma, Y.; Meng, Z.; Li, P.; Ding, C.; Bordas, S.P.A. Reduced order isogeometric boundary element methods for CAD-integrated shape optimization in electromagnetic scattering. *Comput. Methods Appl. Mech. Eng.* **2024**, *419*, 116654. [[CrossRef](#)]
30. 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, Nashville, TN, USA, 20–25 June 2021; pp. 7210–7219.
31. Shen, J.; Ruiz, A.; Agudo, A.; Moreno-Noguer, F. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3D representations. In Proceedings of the 2021 International Conference on 3D Vision (3DV), London, UK, 1–3 December 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 972–981.
32. Shen, J.; Agudo, A.; Moreno-Noguer, F.; Ruiz, A. Conditional-flow NeRF: Accurate 3D modelling with reliable uncertainty quantification. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Cham, Switzerland, 2022; pp. 540–557.
33. Pan, X.; Lai, Z.; Song, S.; Huang, G. ActiveNeRF: Learning where to see with uncertainty estimation. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Cham, Switzerland, 2022; pp. 230–246.
34. Max, N. Optical models for direct volume rendering. *IEEE Trans. Vis. Comput. Graph.* **1995**, *1*, 99–108. [[CrossRef](#)]
35. Song, Y.; Nakath, D.; She, M.; Elibol, F.; Köser, K. Deep sea robotic imaging simulator. In *Pattern Recognition, Proceedings of the ICPR International Workshops and Challenges, Virtual Event, 10–15 January 2021*; Springer: Cham, Switzerland, 2021; pp. 375–389.
36. Schonberger, J.L.; Frahm, J.M. Structure-from-motion revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113.
37. Loquercio, A.; Segu, M.; Scaramuzza, D. A general framework for uncertainty estimation in deep learning. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3153–3160. [[CrossRef](#)]
38. Qu, C.; Liu, W.; Taylor, C.J. Bayesian deep basis fitting for depth completion with uncertainty. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 16147–16157.
39. Bae, G.; Budvytis, I.; Cipolla, R. Estimating and exploiting the aleatoric uncertainty in surface normal estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 13137–13146.
40. Poggi, M.; Aleotti, F.; Tosi, F.; Mattocchia, S. On the uncertainty of self-supervised monocular depth estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3227–3237.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.