

Evidential Neural Radiance Fields

Ruxiao Duan
Yale University

ruxiao.duan@yale.edu

Alex Wong
Yale University

alex.wong@yale.edu

Abstract

Understanding sources of uncertainty is fundamental to trustworthy three-dimensional scene modeling. While recent advances in neural radiance fields (NeRFs) achieve impressive accuracy in scene reconstruction and novel view synthesis, the lack of uncertainty estimation significantly limits their deployment in safety-critical settings. Existing uncertainty quantification methods for NeRFs fail to separately capture both aleatoric and epistemic uncertainties. Among those that do quantify one or the other, many of them either compromise rendering quality or incur significant computational overhead to obtain uncertainty estimates. To address these issues, we introduce Evidential Neural Radiance Fields, a probabilistic approach that seamlessly integrates with the NeRF rendering process, enabling direct quantification of both aleatoric and epistemic uncertainties from a single forward pass. We compare multiple uncertainty quantification methods on three standardized benchmarks, where our approach demonstrates state-of-the-art scene reconstruction fidelity and uncertainty estimation quality. Code is available at <https://github.com/KerryDRX/EvidentialNeRF>.

1. Introduction

Predictive uncertainty of deep neural networks originates from two different sources, data and model, and the two types of uncertainty are respectively referred to as aleatoric uncertainty (AU) and epistemic uncertainty (EU) [14, 18]. Aleatoric (data) uncertainty stems from intrinsic randomness of the data generation process, while epistemic (model) uncertainty arises from model’s lack of knowledge [9, 12]. Understanding both types of predictive uncertainty is essential for building models that are not only accurate but also reliable and explainable under various conditions.

Neural radiance fields (NeRFs) [26] have achieved remarkable performance in three-dimensional (3D) scene reconstruction and novel view synthesis, yet their incapability of quantifying predictive uncertainty poses significant challenges for their broader adoption in safety-critical domains

such as autonomous driving [11], medical imaging [45], and robotics [44], where precise and prompt quantification of predictive uncertainty remains crucial.

We consider an essential but overlooked research question: *How should a NeRF report what it does not know (epistemic uncertainty) versus what the data cannot resolve (aleatoric uncertainty), without sacrificing rendering fidelity or speed?* Existing uncertainty quantification (UQ) methods for NeRFs generally fall into three categories: closed-form likelihood models, which are unable to capture epistemic uncertainty [24, 30, 35, 36], Bayesian methods, which typically necessitate sampling during inference [7, 37, 38], and ensemble approaches, which require training multiple models [22, 41]. None of these methods quantifies both types of uncertainty with a single forward pass, and many of them have to sacrifice prediction accuracy to obtain uncertainty estimates. Among the likelihood models is the classical Gaussian approach [18] that represents a prediction as the mean and uncertainty as the variance of a normal distribution. This method has been widely adopted in not only NeRFs [24, 30, 31] but also other radiance field frameworks such as Gaussian Splatting [19, 21] and DVGO [39, 40]. While computationally efficient, this paradigm inherently models only aleatoric uncertainty while failing to account for epistemic uncertainty.

To address these limitations, we propose Evidential Neural Radiance Fields (Evidential NeRFs), a probabilistic framework to separately quantify both aleatoric and epistemic uncertainties of NeRFs through a single forward pass (Figure 1). Our method extends the normal probabilistic formulation of radiance modeling by treating the predicted mean and variance of pixel radiance themselves as random variables governed by a higher-order evidential distribution, yielding closed-form predictive uncertainties through the NeRF rendering process. Unlike prior evidential deep learning methods [1, 34], which regress evidential distribution parameters, we adapt evidential reasoning to the volumetric structure of NeRFs, enabling them to predict aleatoric and epistemic uncertainties directly instead of having to reformulate them from evidential parameters.

Beyond the technical limitations of existing approaches,

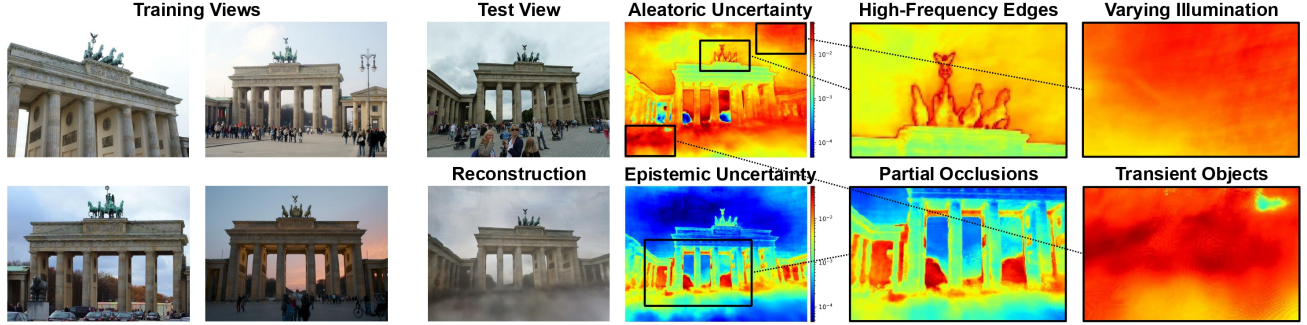


Figure 1. Scene reconstruction along with its aleatoric and epistemic uncertainties rendered by Evidential NeRF from a scene in the wild. AU arises from intrinsic data variability in the training images, including radiance variation (e.g., sky regions due to changing illumination), high-frequency regions (e.g., object edges), and the presence of transient objects (e.g., pedestrians). EU indicates the model’s lack of knowledge, prominently appearing in occluded regions where training supervision is insufficient (e.g., the trees occluded by the gate).

which we address with Evidential NeRF, we also observed that current uncertainty quantification benchmarks lack standardization. Methods are often evaluated using different architectures, data splits, and training setups, making direct comparisons difficult. To mitigate these confounding factors, we establish a new benchmark to isolate and evaluate the UQ methods themselves. Under this standardized evaluation, our method consistently ranks within the top three across all image reconstruction and uncertainty quantification metrics, if not the first. The remaining top-performing methods are mostly ensemble-based, which incur substantial computational cost and are unsuitable for real-time applications. In contrast, our approach is the second-fastest overall, being only 0.04 FPS slower than the fastest method while delivering significantly better performance across all metrics.

Our contributions can be summarized as follows.

- We present a probabilistic framework for neural radiance fields to separately quantify both aleatoric and epistemic uncertainties in 3D scene reconstructions.
- We provide detailed mathematical derivations showing how aleatoric and epistemic uncertainties can be propagated from points to pixels under proper independence assumptions, enabling seamless integration of evidential deep learning with the volumetric rendering paradigm.
- We establish a standardized benchmark for faithful comparison of NeRF uncertainty quantification methods.

2. Related Work

Neural radiance fields. NeRFs [26] represent a 3D scene as a plenoptic function which can be learned from a sparse set of input views. Realized as a deep neural network, the plenoptic function maps spatial locations and viewing directions to their volume densities and radiance. NeRFs produce high-fidelity reconstructions and photorealistic novel view renderings. Numerous variants in architectures, loss functions, and sampling strategies have been proposed to further improve the model’s training efficiency, inference

speed, and rendering quality [2–6, 8, 13, 15, 28, 29, 43, 46].

NeRF uncertainty quantification. UQ methods for NeRFs generally follow three schools of thought.

First, closed-form likelihood models. NeRF-W [24] and ActiveNeRF [30] follow the same probabilistic framework in [18] to model ray colors by normal distributions. MixNeRF [36] and FlipNeRF [35], though not proposed as UQ methods, model ray colors by mixtures of Laplace distributions to represent the multimodal nature of radiance. These approaches employ closed-form probabilistic distributions to model pixel colors, enabling efficient training and uncertainty inference, but overlooking epistemic uncertainty, due to the deterministic nature of the distributions employed.

Second, Bayesian methods. Monte Carlo dropout [7] estimates epistemic uncertainty by interpreting dropout as variational inference and performing multiple stochastic forward passes at test time. S-NeRF [37] and CF-NeRF [38] learn a probability distribution over all the possible radiance fields and approximate the posterior distribution with stochastic variational inference [20] and conditional normalizing flows [32]. While these approaches provide epistemic uncertainty estimates, they typically incur significant computational overhead due to sampling at inference.

Third, ensemble methods. Deep ensembles [22] estimate epistemic uncertainty by training multiple models independently and measuring their prediction variance during inference. Density-aware NeRF Ensembles (DANE) [41] further incorporates a density-based epistemic term that captures uncertainty in scene geometry and appearance, which is added to the uncertainty estimated by standard ensembles. Ensemble-based methods impose the most significant computation and memory overheads, as each model must be trained, stored, and separately evaluated during inference.

Evidential deep learning. Based on the theory of subjective logic [17], evidential deep learning (EDL) [34] is first introduced for classification uncertainty quantification based on a single network and later extended to regression [1]. EDL considers model learning as a process of

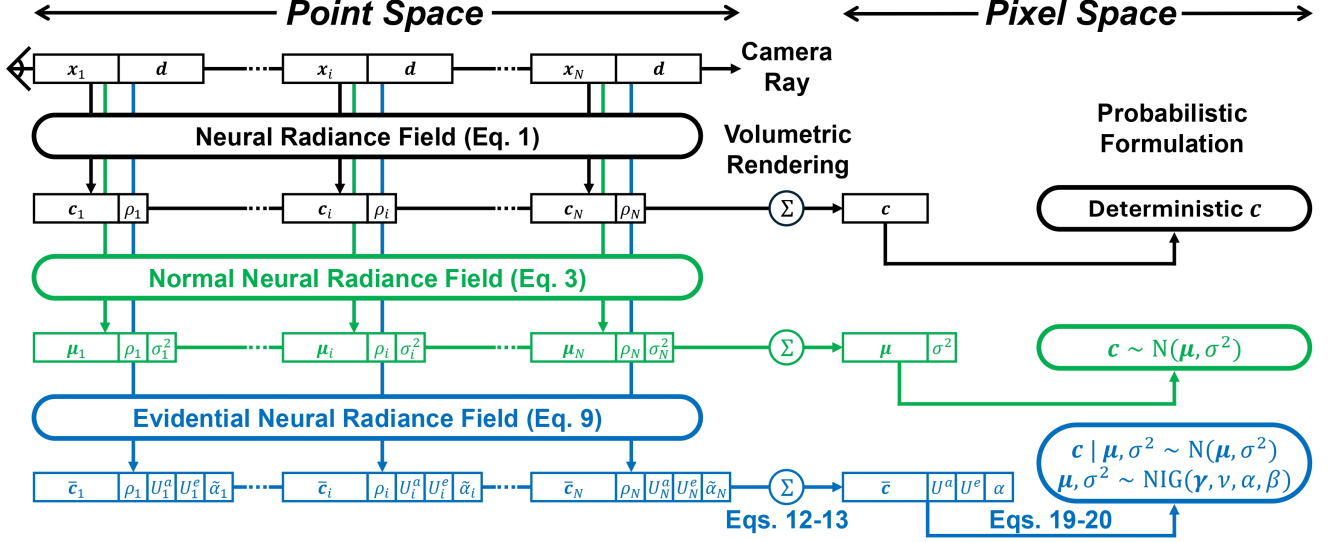


Figure 2. Evolution of NeRF pipeline across three levels of probabilistic formulations. N points sampled along the camera ray give N pairs of spatial location and viewing direction, which are passed to the NeRF model for prediction. Level 1: Vanilla NeRF predicts only point density and color, resulting in a deterministic pixel color without any uncertainty estimate. Level 2: Normal NeRF assumes the point and pixel colors follow normal distributions, quantifying aleatoric uncertainty of rendered color. Level 3: Evidential NeRF assumes the pixel color has random mean and variance following an evidential distribution, quantifying both aleatoric and epistemic uncertainties.

evidence collection, in which each training sample adds support to a higher-order evidential distribution, which is a Dirichlet distribution for classification and a normal-inverse-gamma (NIG) distribution for regression [1, 34]. The probability distribution from which predictions are drawn is assumed to have random parameters sampled from the evidential distribution, and the aleatoric and epistemic predictive uncertainties can be directly obtained by a single forward propagation. Though effective in standard regression, EDL is not naively compatible with NeRF learning paradigm due to its hierarchical volumetric formulation: As NeRFs receive supervision after volumetric rendering, evidential parameters tied to point-level predictions cannot be learned directly from pixel-level observations. Concurrent with our work, ENeRF [47] proposes evidential modeling in NeRFs by adopting NIG distribution at the point level and approximating pixel-level parameters via the mixture of NIGs [23].

3. Preliminaries

Neural radiance fields. NeRF learns a continuous scene representation function f by a multilayer perceptron that maps a pair of spatial location (referred to as point in this paper) $\mathbf{x}_i \in \mathbb{R}^3$ and unit viewing direction $\mathbf{d} \in \mathbb{R}^3$ to its corresponding view-independent volume density $\rho_i > 0$ and directional emitted radiance $\mathbf{c}_i \in [0, 1]^3$ in RGB color. For notational simplicity, consider a scalar color channel $c_i \in [0, 1]$. A NeRF model predicts

$$c_i, \rho_i = f(\mathbf{x}_i, \mathbf{d}). \quad (1)$$

The pixel color of a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ defined by a camera center $\mathbf{o}$ and a viewing direction $\mathbf{d}$ is derived by first sampling N points $\{\mathbf{r}(t_i)\}_{i=1}^N$ along the ray and then computing the weighted sum of point colors via the discretized volumetric rendering equation $c = \sum_{i=1}^N w_i c_i$, where $w_i = \exp\left(-\sum_{j=1}^{i-1} \rho_j \delta_j\right) (1 - \exp(-\rho_i \delta_i))$ is the weight of the i -th point and $\delta_i = t_{i+1} - t_i$ is the distance between two adjacent points.

Normal neural radiance fields. Based on the probabilistic framework in [18], a point color c_i can be modeled by a normal distribution as

$$c_i | \mu_i, \sigma_i^2 \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad (2)$$

with mean $\mu_i \in [0, 1]$ and variance $\sigma_i^2 > 0$ predicted by

$$(\mu_i, \sigma_i^2), \rho_i = f(\mathbf{x}_i, \mathbf{d}). \quad (3)$$

As the pixel color c is the weighted sum of point colors c_i 's, c also follows a normal distribution assuming independence of points:

$$c | \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2), \quad (4)$$

where $\mu := \sum_{i=1}^N w_i \mu_i$ is the mean (prediction) and $\sigma^2 := \sum_{i=1}^N w_i^2 \sigma_i^2$ is the variance (uncertainty) of pixel color. While adjacent points in practice exhibit statistical dependence in radiance, the independence assumption is a common and effective simplification to enable tractable aggregation of point uncertainties into the pixel level. Though efficient, this approach only captures aleatoric uncertainty.

4. Method

Based on Equations (2) and (4), we establish an evidential probabilistic framework for radiance modeling by taking a step further to assume random conditional mean and variance. Different from conventional EDL approaches which train the model to predict evidential parameters, we instead let the model predict aleatoric and epistemic uncertainties directly (Section 4.1), and then propagate the uncertainties from points to pixels compatible with the volumetric rendering paradigm of NeRFs (Section 4.2). Then the evidential distribution can be reformulated from the rendered color and uncertainties (Section 4.3) and learned at the pixel level where supervision is available (Section 4.4).

4.1. Point-level Probabilistic Radiance Modeling

Building upon Equation (2), we further treat the conditional mean and variance of point radiance as random variables instead of point estimates, i.e., $(\mu_i, \sigma_i^2) \sim \pi_i$ where π_i is some probability distribution. Under this formulation, the predictive mean, total, aleatoric, and epistemic uncertainties of the point color c_i can be respectively formulated as

$$\bar{c}_i := \mathbb{E}[c_i] = \mathbb{E}[\mathbb{E}[c_i | \mu_i, \sigma_i^2]] = \mathbb{E}[\mu_i], \quad (5)$$

$$U_i := \text{Var}[c_i] = U_i^{\text{alea}} + U_i^{\text{epis}}, \quad (6)$$

$$U_i^{\text{alea}} := \mathbb{E}[\text{Var}[c_i | \mu_i, \sigma_i^2]] = \mathbb{E}[\sigma_i^2], \quad (7)$$

$$U_i^{\text{epis}} := \text{Var}[\mathbb{E}[c_i | \mu_i, \sigma_i^2]] = \text{Var}[\mu_i]. \quad (8)$$

Note: The classical assumption that μ_i and σ_i^2 are point estimates corresponds to the special case where π_i is a degenerate distribution with a Dirac delta density function concentrated at fixed values of $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ and the uncertainties above are thereby reduced to $U_i^{\text{alea}} = \mathbb{E}[\sigma_i^2] = \hat{\sigma}_i^2$ and $U_i^{\text{epis}} = \text{Var}[\mu_i] = 0$, in which case only aleatoric uncertainty can be captured.

An Evidential NeRF model predicts

$$(\bar{c}_i, U_i^{\text{alea}}, U_i^{\text{epis}}, \tilde{\alpha}_i, \rho_i) = f(\mathbf{x}_i, \mathbf{d}), \quad (9)$$

where $\bar{c}_i \in [0, 1]$ is the point’s mean color, $U_i^{\text{alea}} > 0$ and $U_i^{\text{epis}} > 0$ are aleatoric and epistemic uncertainties, and $\tilde{\alpha}_i > 0$ is a positive shape score which will later be used to derive a pixel-level shape parameter α in Section 4.3. We apply sigmoid activation to predict $\bar{c}_i$ as in original NeRF and softplus activation for all other parameters to enforce range restrictions. The model architecture remains mostly unchanged, except for the three additional output neurons in the last layer to predict $(U_i^{\text{alea}}, U_i^{\text{epis}}, \tilde{\alpha}_i)$.

4.2. Propagation from Points to Pixels

Let the set of conditional means and variances of all point colors along a ray be $\boldsymbol{\theta} := \{(\mu_i, \sigma_i^2)\}_{i=1}^N$, which fully specifies the conditional distribution of the ray color c . Under this setting, the pixel’s predictive mean color, total,

aleatoric, and epistemic uncertainties are defined as

$$\bar{c} := \mathbb{E}[c], \quad U := \text{Var}[c], \quad (10)$$

$$U^{\text{alea}} := \mathbb{E}[\text{Var}[c | \boldsymbol{\theta}]], \quad U^{\text{epis}} := \text{Var}[\mathbb{E}[c | \boldsymbol{\theta}]]. \quad (11)$$

With the point-level predictions from Equation (9) and the independence assumption of points, the pixel color and uncertainties can be derived as

$$\bar{c} = \sum_{i=1}^N w_i \bar{c}_i, \quad U = \sum_{i=1}^N w_i^2 U_i, \quad (12)$$

$$U^{\text{alea}} = \sum_{i=1}^N w_i^2 U_i^{\text{alea}}, \quad U^{\text{epis}} = \sum_{i=1}^N w_i^2 U_i^{\text{epis}}. \quad (13)$$

In other words, pixel-level aleatoric and epistemic uncertainties can be obtained as weighted sums of their point-level counterparts, where the weights are equal to the squared weights for color volumetric rendering. Detailed proofs are provided in Section 8.

4.3. Pixel-level Probabilistic Radiance Modeling

Now we introduce the probabilistic model for pixel colors. Recall that $\mu = \sum_{i=1}^N w_i \mu_i$ and $\sigma^2 = \sum_{i=1}^N w_i^2 \sigma_i^2$ represent the conditional mean and variance of the pixel color c in Equation (4). To quantify aleatoric and epistemic uncertainties of a pixel color in closed form, we model (μ, σ^2) by a normal-inverse-gamma (NIG) distribution

$$\mu, \sigma^2 \sim \text{NIG}(\gamma, \nu, \alpha, \beta), \quad (14)$$

or equivalently, $\mu | \sigma^2 \sim \mathcal{N}(\gamma, \sigma^2/\nu)$ and $\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$, where $\gamma \in [0, 1]$, $\nu > 0$, and $\Gamma^{-1}(\alpha, \beta)$ denotes an inverse-gamma distribution with shape $\alpha > 1$ and scale $\beta > 0$. Consider each pixel color c as being generated through a hierarchical sampling process: Given the higher-order evidential distribution $\text{NIG}(\gamma, \nu, \alpha, \beta)$, drawing a sample (μ, σ^2) from it yields an instance of lower-order normal distribution $\mathcal{N}(\mu, \sigma^2)$, from which the pixel color c is sampled.

With this formulation, the pixel color’s predictive mean, total, aleatoric, and epistemic uncertainties can be directly expressed in terms of the NIG parameters:

$$\bar{c} = \mathbb{E}[c] = \mathbb{E}[\mathbb{E}[c | \mu, \sigma^2]] = \mathbb{E}[\mu] = \gamma, \quad (15)$$

$$U = \text{Var}[c] = U^{\text{alea}} + U^{\text{epis}}, \quad (16)$$

$$U^{\text{alea}} = \mathbb{E}[\text{Var}[c | \mu, \sigma^2]] = \mathbb{E}[\sigma^2] = \frac{\beta}{\alpha - 1}, \quad (17)$$

$$U^{\text{epis}} = \text{Var}[\mathbb{E}[c | \mu, \sigma^2]] = \text{Var}[\mu] = \frac{\beta}{(\alpha - 1)\nu}. \quad (18)$$

Since the Evidential NeRF model directly predicts uncertainties and shape scores instead of NIG parameters, the NIG parameters need to be reformulated as

Dataset	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NLL $\downarrow$	AUSE RMSE $\downarrow$	AUSE MAE $\downarrow$
LF [48]	Baseline	28.5538	0.9172	0.0465			
	Dropout [7]	28.1137	0.9061	0.0569	3.6670	0.0125	0.0049
	Normal [18]	28.0064	0.9165	0.0531	0.4425	0.0090	0.0029
	MoL [36]	28.2200	0.9095	0.0672	-2.5393	0.0084	0.0029
	Ensembles [22]	29.3779	0.9308	0.0411	0.3245	0.0070	0.0026
	DANE [41]	29.3779	0.9308	0.0411	-0.4317	0.0101	0.0039
	Evidential	29.9679	0.9345	0.0359	-2.4491	0.0070	0.0025
LLFF [25]	Baseline	17.5281	0.4720	0.4119			
	Dropout [7]	17.1423	0.4441	0.4468	90.6773	0.0742	0.0426
	Normal [18]	16.5500	0.4135	0.4732	55.3580	0.0674	0.0405
	MoL [36]	16.4218	0.4237	0.4842	2.2470	0.0814	0.0384
	Ensembles [22]	17.9181	0.5109	0.3932	11.1658	0.0513	0.0253
	DANE [41]	17.9181	0.5109	0.3932	9.7273	0.0521	0.0260
	Evidential	17.8793	0.5068	0.3751	0.6765	0.0578	0.0295
RobustNeRF [33]	Baseline	25.2205	0.8296	0.1577			
	Dropout [7]	24.7693	0.8118	0.1745	22.8799	0.0284	0.0160
	Normal [18]	25.2987	0.8522	0.1311	10.4873	0.0250	0.0151
	MoL [36]	23.7874	0.7921	0.2049	-1.3947	0.0304	0.0176
	Ensembles [22]	26.1953	0.8562	0.1438	4.6309	0.0164	0.0098
	DANE [41]	26.1953	0.8562	0.1438	4.1092	0.0283	0.0154
	Evidential	26.2292	0.8641	0.1112	-1.2702	0.0221	0.0138

Table 1. Quantitative results of scene reconstruction and uncertainty quantification on three datasets, averaged over three independent runs. Colored cells denote the **first**, **second**, and **third** best results. See per-scene statistics with standard deviations in supplementary Section 10.

$$\gamma = \bar{c}, \quad \nu = \frac{U^{\text{alea}}}{U^{\text{epis}}}, \quad (19)$$

$$\alpha = 1 + \sum_{i=1}^N \tilde{w} \tilde{\alpha}_i, \quad \beta = U^{\text{alea}}(\alpha - 1), \quad (20)$$

where $\bar{c}$, U^{alea} , and U^{epis} are obtained from the point-to-pixel propagation process in Equations (12) and (13), $\tilde{\alpha}_i$ is the point’s shape score, and $\tilde{w}_i := w_i / \sum_{j=1}^N w_j$ is the normalized weight that determines how much a point’s shape score contributes to the NIG shape parameter α of the pixel.

4.4. Learning

Based on Equations (4) and (14), a pixel color c marginally follows a Student’s t distribution

$$c \sim t\left(\gamma, \frac{\beta(\nu + 1)}{\alpha\nu}, 2\alpha\right), \quad (21)$$

where $t(\mu_t, \sigma_t^2, \nu_t)$ denotes a Student’s t distribution with location μ_t , scale σ_t , and degrees of freedom ν_t . The model is trained via maximum likelihood estimation by minimizing the negative log-likelihood (NLL) of the ground truth

$$\mathcal{L}_{\text{nll}} = -\log p(c^{\text{gt}} | \gamma, \nu, \alpha, \beta) \quad (22)$$

$$= \frac{1}{2} \log \frac{\pi}{\nu} - \alpha \log \Omega + \log \frac{\Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})} + \left(\alpha + \frac{1}{2}\right) \log((c^{\text{gt}} - \gamma)^2 \nu + \Omega), \quad (23)$$

where $c^{\text{gt}} \in [0, 1]$ is the ground truth pixel color, $\Omega = 2\beta(\nu + 1)$, and Γ represents the Gamma function. See derivations in Section 9.

Different values of ν and β may result in the same scale of the t distribution so long as the ratio $\beta(\nu + 1)/\nu$ is fixed. To resolve this ambiguity and also to suppress excessive evidence assigned to inaccurate predictions, a regularizer [1] is introduced to the loss function as

$$\mathcal{L}_{\text{reg}} = |c^{\text{gt}} - \gamma|(2\nu + \alpha), \quad (24)$$

where $|c^{\text{gt}} - \gamma|$ is the absolute error of prediction and $2\nu + \alpha$ is the count of virtual observations representing the evidence [1]. Therefore, the total loss is written as

$$\mathcal{L} = \mathcal{L}_{\text{nll}} + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}}, \quad (25)$$

where $\lambda_{\text{reg}} > 0$ is the regularization coefficient.

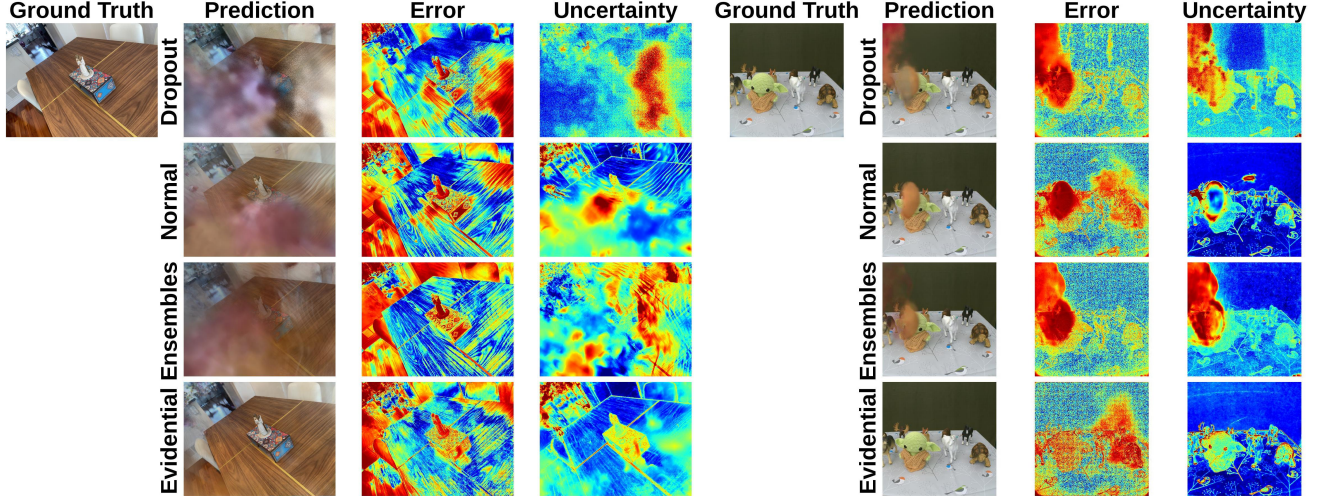


Figure 3. Qualitative comparison on two example scenes, with image reconstructions, error maps, and uncertainty maps. Histogram equalization is conducted on the error maps to highlight the error regions. Our method’s uncertainty is total uncertainty. In general, our method achieves better reconstruction accuracy and produces uncertainty maps that are more consistent with prediction errors.

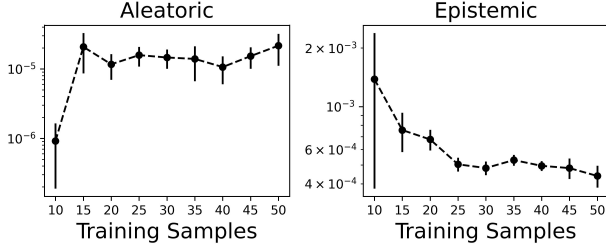


Figure 4. Test uncertainties vs. training sample size on *Android*.

Lastly, we generalize our color channel assumption from a single channel to three for RGB color modeling. We assume that the three color channels have different means but the same uncertainty, due to the high correlation of variance in different color channels of the same pixel [24, 30]. Practically, the output dimension of color prediction head in Equation (9) is set to three, predicting the point’s mean RGB color as $\bar{c}_i \in [0, 1]^3$, and the uncertainty-related parameters are broadcast to all color channels during training.

5. Experiments

5.1. Settings

Datasets. We test the UQ methods on three 3D scene reconstruction datasets: Light Field (LF) [48], Local Light Field Fusion (LLFF) [25], and RobustNeRF [33]. For LF, we test on the four scenes following the setup in [10] with the same train-test split. For LLFF, we employ its eight scenes and adopt the train-test split of [29], training with only three input views to assess UQ performance under sparse-view conditions. For RobustNeRF, we evaluate on its four scenes following their protocol of training on cluttered images only and testing on clean ones.

Metrics. We evaluate the method performance based on the

Mode	Baseline	Dropout	Normal	MoL	Ens.	DANE	Ours
Train/min.↓	11.84	88.54	12.22	12.71	59.22	59.22	13.57
Infer/FPS↑	4.88	0.09	4.71	4.42	0.96	0.96	4.67

Table 2. Average training time per 30k steps and inference FPS of baseline nerfacto and different UQ methods on an A6000 GPU.

accuracy of both the rendered images and the uncertainty estimates. For images, we report PSNR, SSIM, and LPIPS to reflect the image reconstruction quality. For uncertainties, we use negative log-likelihood (NLL) and area under sparsification error (AUSE) with respect to both RMSE and MAE, measuring the quality of uncertainty estimates in terms of distributional fit and error ranking capabilities.

Baselines. We select UQ methods from each of the three categories discussed in Section 2 as baselines. For closed-form likelihood models, we include the normal distribution approach [18] adopted in [24] and [30] and mixture of Laplace distributions (MoL) employed in [36] and [35]. Among Bayesian methods, we choose the classical Monte Carlo dropout approach [7]. For ensemble methods, we consider naive deep ensembles [22] and density-aware NeRF ensembles (DANE) [41].

Implementation. Prior benchmarks on NeRF uncertainty quantification suffer from varying choices of data splits, model architectures, and training schemes. To isolate UQ method effects from engineering confounders, we establish a new standardized benchmark to focus comparison on the underlying UQ approaches themselves. First, for data split, we follow the aforementioned scheme for all models. Second, for architecture, we use nerfacto model in [42] to implement all methods for efficiency. For likelihood models, the only difference is the output layer size and the loss function. For the Bayesian method, Monte Carlo dropout is

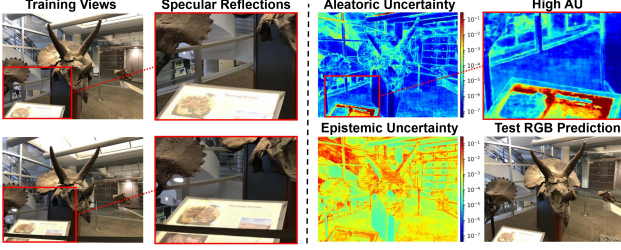


Figure 5. A case where AU dominates EU. The highly reflective surface of the display case in the foreground incurs specular reflections. AU arises due to the presence of data noise caused by the inconsistency of light across different training views.

implemented with a dropout probability 0.2, trained once and sampled five times at inference. For ensembles, both naive ensembles and DANE are implemented by training five models independently. Third, in training, we adopt the default batch size, optimizer, and learning rate scheduler of nerfacto for all methods and train all models for the same number of iterations, chosen based on the convergence speed on each scene. For stability, we run each method independently three times and report the averaged metrics.

5.2. Results

Quantitative results. The quantitative performance of the uncertainty quantification methods is detailed in Table 1 with per-scene statistics provided in Section 10.

For **scene reconstruction**, compared to the baseline nerfacto model, likelihood and Bayesian methods (Normal, MoL, Dropout) struggle to preserve the image reconstruction accuracy, indicating that their predictive performance is compromised in exchange for uncertainty estimation. Ensemble-based methods provide faithful reconstructions in general, but they are extremely inefficient due to their substantial computational cost in both training and inference. In contrast, Evidential NeRF, while being efficient, consistently outperforms the baseline in image reconstruction, demonstrating that our uncertainties need not come at the expense of rendering fidelity. Remarkably, with only a single network trained and a single forward pass required at inference, the evidential approach achieves prediction accuracy comparable to, and in 7 out of 9 image reconstruction metrics even surpassing, the computationally expensive ensemble methods.

For **uncertainty quantification**, in terms of NLL, MoL model achieves the best result overall likely due to its multimodal radiance distribution assumption. Compared to the normal distribution with fixed mean and variance, a normal distribution with random mean and variance increases test data likelihood by factors of 1.8×10^1 , 5.6×10^{23} , and 1.3×10^5 on the three datasets, respectively. This tremendous improvement is because the fixed mean and variance assumption used by classical normal-based UQ

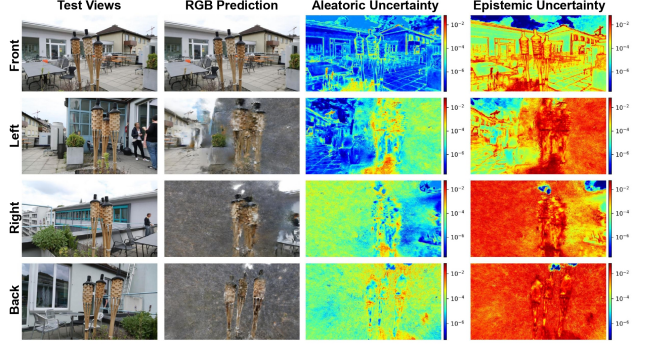


Figure 6. A case where EU dominates AU. The model is trained only by 5 images from the front and asked to render the scene from all the viewing angles. EU arises due to the lack of knowledge during training on the views out of the training distribution.

only captures the aleatoric variations within the training data while overlooking epistemic uncertainty. The likelihood gain is less pronounced on LF because its test images are adjacent to the training views, thus the distribution shift is relatively insignificant. As for AUSE, ensemble-based methods achieve the strongest results in general, thanks to their ability to represent predictive variability through multiple models. Nevertheless, our evidential approach remains highly competitive, often ranking second only to ensembles, demonstrating that jointly modeling aleatoric and epistemic factors yields uncertainty estimates with stronger correlation with errors.

Qualitative results. We qualitatively compare our Evidential NeRF against other methods in Figure 3 and Section 10. A primary goal of UQ is to produce uncertainty maps that accurately identify potential model failures. Previous methods, by neglecting either the aleatoric or epistemic component of predictive uncertainty, often fall short of accurately delimiting regions of predictive inaccuracy. In contrast, by considering both components, our method demonstrates superior alignment with the reconstruction error maps. This suggests that jointly modeling data noise and the model knowledge gap is essential for generating the most reliable uncertainty estimates.

Uncertainty scaling with data. We investigate the effect of training sample size on the magnitude of aleatoric and epistemic uncertainties using an example scene from Robust-NeRF in Figure 4. We reserve a held-out set of views for evaluation and incrementally select 10 to 50 images from the remaining data for training. As the training sample size increases, the test AU increases while the test EU decreases in general, indicating that the additional, potentially more cluttered observations introduce greater data variability, whereas the model’s lack of knowledge is gradually mitigated with more training data.

Aleatoric and epistemic factors in 3D scenes. Several scenarios where aleatoric and epistemic uncertainties emerge

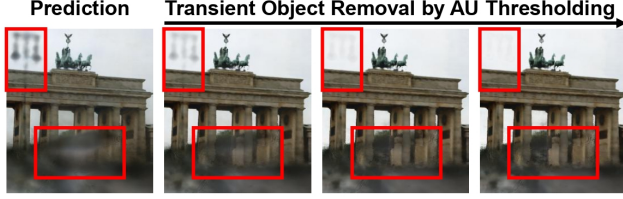


Figure 7. Scene cleaning based on aleatoric uncertainty as a post-processing step for floater removal. Points with AU above a certain threshold have their density reduced to become more transparent. By reducing the threshold, more artifacts can be eliminated.

are respectively presented in Figures 1, 5 and 6. Figure 1 illustrates an in-the-wild scene with several aleatoric factors (varying illumination, high-frequency edge regions, transient objects) and an epistemic factor (partial occlusions). Figure 5 showcases a scenario with a reflective object, where aleatoric light variation is the dominant source of potential rendering error. Figure 6 demonstrates an out-of-distribution example where the model is asked to render the scene from angles that never appeared during training, showing that the lack of model knowledge about these unseen views is the primary cause of predictive failure.

Computational efficiency. We report the total training time and inference frames per second (FPS) of baseline nerfacto and different UQ methods in Table 2. Compared to other UQ methods, in training, our evidential approach is only slightly slower than other two likelihood models while being significantly faster than the ensemble methods. As for inference, since our approach directly predicts uncertainty fields, it is more efficient than the methods that require computing uncertainty from the predicted parameters, such as MoL. Since nerfacto runs on Tiny CUDA Neural Networks [27], which does not support dropout, dropout is implemented without acceleration, making it naturally slower than other methods.

5.3. Applications

We demonstrate two example applications using aleatoric and epistemic uncertainties respectively in Figures 7 and 8.

Scene cleaning. Training NeRFs on unconstrained images often leads to inaccurate predictions due to radiance inconsistencies from uncontrolled lighting changes or transient objects, which are irreducible data noise explained by aleatoric uncertainty. As a robust indicator of geometric artifacts, AU-based scene cleaning can be applied as a post-processing procedure to eliminate noise in the renderings. Figure 7 presents an example where the erroneous floaters of the rendering can be progressively removed by gradually changing the AU threshold to suppress point densities.

Active learning. As a proxy of lack of model knowledge, epistemic uncertainty is an ideal metric for active learning. Figure 8 illustrates a next-best-view planning experiment on an LLFF scene. Starting from an initial training set of 5 im-

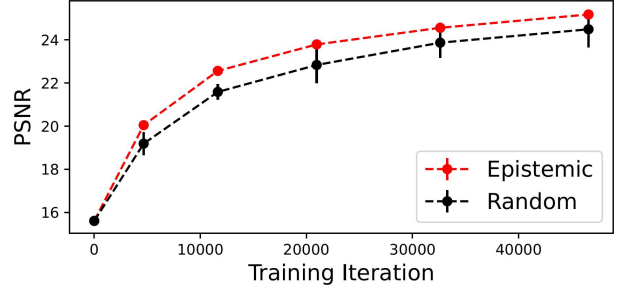


Figure 8. Mean and standard deviation of test PSNR of three runs on *Horns* scene, with two active sampling strategies: EU-based selection and random selection. The samples identified via epistemic uncertainty are more informative for model learning.

ages, an Evidential NeRF is trained iteratively by 5 rounds. In each round, the model is trained by 5 epochs with all the current training images and then tested on a holdout test set, before choosing 5 additional images from the remaining data to add to the training pool. Two sampling strategies for data selection are compared: choosing the images with the most EU and choosing images randomly. The average and standard deviation of three independent runs of each selection scheme are reported. Active selection based on EU gives noticeably higher PSNR than random selection, indicating the samples with higher epistemic uncertainty are more informative for model learning.

6. Conclusion

Despite the efficacy of NeRFs in photorealistic 3D scene reconstruction, the absence of robust predictive uncertainty quantification significantly hinders their deployment in safety-critical domains. While some uncertainty quantification methods have been proposed for NeRFs, none of them provides a mechanism for quantifying both aleatoric and epistemic uncertainties of the scene reconstructions. In this paper, we introduce Evidential Neural Radiance Fields, a principled probabilistic framework that resolves this critical gap. Our approach seamlessly adapts evidential deep learning to the hierarchical nature of NeRFs through a point-to-pixel aleatoric and epistemic uncertainty propagation paradigm. Through extensive experiments on three standardized benchmarks, Evidential NeRF demonstrates its superiority in both image reconstruction fidelity and uncertainty estimation quality compared to other methods, while identifying the aleatoric and epistemic factors that contribute to predictive uncertainty within a variety of 3D scenes.

Limitation. For tractable uncertainty propagation, Evidential NeRF assumes deterministic volumetric density, leaving spatial uncertainty not explicitly modeled. Extending the formulation to capture the uncertainty in scene geometries would enable more comprehensive uncertainty modeling.

7. Acknowledgment

This work is supported by NSF 2112562 Athena AI Institute.

References

- [1] Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in neural information processing systems*, 33:14927–14937, 2020. 1, 2, 3, 5
- [2] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5855–5864, 2021. 2
- [3] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5470–5479, 2022.
- [4] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19697–19705, 2023.
- [5] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European conference on computer vision*, pages 333–350. Springer, 2022.
- [6] Zhiqin Chen, Thomas Funkhouser, Peter Hedman, and Andrea Tagliasacchi. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16569–16578, 2023. 2
- [7] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 1, 2, 5, 6
- [8] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 14346–14355, 2021. 2
- [9] Jakob Gawlikowski, Cedrique Rovile Njieutcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, et al. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 56(Suppl 1):1513–1589, 2023. 1
- [10] Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes’ rays: Uncertainty quantification for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20061–20070, 2024. 6
- [11] Lei He, Leheng Li, Wenchao Sun, Zeyu Han, Yichen Liu, Sifa Zheng, Jianqiang Wang, and Keqiang Li. Neural radiance field in autonomous driving: A survey. *arXiv preprint arXiv:2404.13816*, 2024. 1
- [12] Wenchong He, Zhe Jiang, Tingsong Xiao, Zelin Xu, and Yukun Li. A survey on uncertainty quantification methods for deep learning, 2024. 1
- [13] Tao Hu, Shu Liu, Yilun Chen, Tiancheng Shen, and Jiaya Jia. Efficientnerf efficient neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12902–12911, 2022. 2
- [14] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning*, 110(3):457–506, 2021. 1
- [15] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5885–5894, 2021. 2
- [16] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2):517–547, 2021. 14
- [17] Audun Jsgang. *Subjective Logic: A formalism for reasoning under uncertainty*. Springer Publishing Company, Incorporated, 2018. 2
- [18] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017. 1, 2, 3, 5, 6
- [19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1
- [20] Diederik P Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2
- [21] Jonas Kulhanek, Songyou Peng, Zuzana Kukelova, Marc Pollefeys, and Torsten Sattler. Wildgaussians: 3d gaussian splatting in the wild. *arXiv preprint arXiv:2407.08447*, 2024. 1
- [22] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017. 1, 2, 5, 6
- [23] Huan Ma, Zongbo Han, Changqing Zhang, Huazhu Fu, Joey Tianyi Zhou, and Qinghua Hu. Trustworthy multimodal regression with mixture of normal-inverse gamma distributions. *Advances in Neural Information Processing Systems*, 34:6881–6893, 2021. 3
- [24] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7210–7219, 2021. 1, 2, 6
- [25] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and

- Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (ToG)*, 38(4):1–14, 2019. 5, 6
- [26] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 2
- [27] Thomas Müller. tiny-cuda-nn, 2021. 8
- [28] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 2
- [29] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. Reg-nerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5480–5490, 2022. 2, 6
- [30] Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. Activenerf: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision*, pages 230–246. Springer, 2022. 1, 2, 6
- [31] Weining Ren, Zihan Zhu, Boyang Sun, Jiaqi Chen, Marc Pollefeys, and Songyou Peng. Nerf on-the-go: Exploiting uncertainty for distractor-free nerfs in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8931–8940, 2024. 1
- [32] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015. 2
- [33] Sara Sabour, Suhani Vora, Daniel Duckworth, Ivan Krasin, David J Fleet, and Andrea Tagliasacchi. Robustnerf: Ignoring distractors with robust losses. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20626–20636, 2023. 5, 6
- [34] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018. 1, 2, 3
- [35] Seunghyeon Seo, Yeonjin Chang, and Nojun Kwak. Flipnerf: Flipped reflection rays for few-shot novel view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22883–22893, 2023. 1, 2, 6
- [36] Seunghyeon Seo, Donghoon Han, Yeonjin Chang, and Nojun Kwak. Mixnerf: Modeling a ray with mixture density for novel view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20659–20668, 2023. 1, 2, 5, 6
- [37] Jianxiong Shen, Adria Ruiz, Antonio Agudo, and Francesc Moreno-Noguer. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In *2021 International Conference on 3D Vision (3DV)*, pages 972–981. IEEE, 2021. 1, 2
- [38] Jianxiong Shen, Antonio Agudo, Francesc Moreno-Noguer, and Adria Ruiz. Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In *European Conference on Computer Vision*, pages 540–557. Springer, 2022. 1, 2
- [39] Jianxiong Shen, Ruijie Ren, Adria Ruiz, and Francesc Moreno-Noguer. Estimating 3d uncertainty field: Quantifying uncertainty for neural radiance fields. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2375–2381. IEEE, 2024. 1
- [40] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5459–5469, 2022. 1
- [41] Niko Sünderhauf, Jad Abou-Chakra, and Dmitry Miller. Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9370–9376. IEEE, 2023. 1, 2, 5, 6
- [42] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Justin Kerr, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David McAllister, and Angjoo Kanazawa. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings*, 2023. 6
- [43] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022. 2
- [44] Guangming Wang, Lei Pan, Songyou Peng, Shaohui Liu, Chenfeng Xu, Yanzi Miao, Wei Zhan, Masayoshi Tomizuka, Marc Pollefeys, and Hesheng Wang. Nerf in robotics: A survey. *arXiv preprint arXiv:2405.01333*, 2024. 1
- [45] Xin Wang, Shu Hu, Heng Fan, Hongtu Zhu, and Xin Li. Neural radiance fields in medical imaging: Challenges and next steps. *arXiv preprint arXiv:2402.17797*, 2024. 1
- [46] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Pointnerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5438–5448, 2022. 2
- [47] Xuanchen Xu, Songcheng Han, Yihan Zhou, Wenlong Yin, Feng Zhang, and Lei Chen. Enerf: Evidential neural radiance fields with uncertainty-driven adaptive resampling for sparse-view synthesis. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pages 468–481. Springer, 2025. 3
- [48] Kaan Yücer, Alexander Sorkine-Hornung, Oliver Wang, and Olga Sorkine-Hornung. Efficient 3d object segmentation from densely sampled light fields with applications to 3d reconstruction. *ACM Transactions on Graphics (TOG)*, 35(3):1–15, 2016. 5, 6

Evidential Neural Radiance Fields

Supplementary Material

8. Derivations: Point-to-Pixel Propagation of Radiance and Uncertainty

This section derives the propagation formulae of radiance and uncertainties specified in Equations (12) and (13).

The mean color aggregation formula follows directly from the linearity of expectation, which does not require any independence assumption.

$$\bar{c} = \mathbb{E}[c] = \mathbb{E}\left[\sum_{i=1}^N w_i c_i\right] = \sum_{i=1}^N w_i \mathbb{E}[c_i] = \sum_{i=1}^N w_i \bar{c}_i. \quad (26)$$

Deriving the three uncertainty propagation formulae each requires its own independence assumption. However, only two, and any two, of these three assumptions are necessary to derive all the three formulae.

Assumption 1. *The point colors are independent, i.e., $c_i \perp\!\!\!\perp c_j, \forall i \neq j$.*

Assumption 2. *The point colors are conditionally independent given the conditional means and variances, i.e., $c_i \perp\!\!\!\perp c_j \mid \theta, \forall i \neq j$.*

Assumption 3. *The point colors' conditional means are independent, i.e., $\mu_i \perp\!\!\!\perp \mu_j, \forall i \neq j$.*

With these assumptions, it can be shown that

$$U = \text{Var}[c] = \text{Var}\left[\sum_{i=1}^N w_i c_i\right] \stackrel{\text{A.1}}{=} \sum_{i=1}^N w_i^2 \text{Var}[c_i] = \sum_{i=1}^N w_i^2 U_i, \quad (27)$$

$$\begin{aligned} U^{\text{alea}} &= \mathbb{E}[\text{Var}[c \mid \theta]] = \mathbb{E}\left[\text{Var}\left[\sum_{i=1}^N w_i c_i \mid \theta\right]\right] \stackrel{\text{A.2}}{=} \mathbb{E}\left[\sum_{i=1}^N w_i^2 \text{Var}[c_i \mid \theta]\right] \\ &= \sum_{i=1}^N w_i^2 \mathbb{E}[\text{Var}[c_i \mid \theta]] = \sum_{i=1}^N w_i^2 \mathbb{E}[\text{Var}[c_i \mid \mu_i, \sigma_i^2]] = \sum_{i=1}^N w_i^2 U_i^{\text{alea}}, \end{aligned} \quad (28)$$

$$\begin{aligned} U^{\text{epis}} &= \text{Var}[\mathbb{E}[c \mid \theta]] = \text{Var}\left[\mathbb{E}\left[\sum_{i=1}^N w_i c_i \mid \theta\right]\right] = \text{Var}\left[\sum_{i=1}^N w_i \mathbb{E}[c_i \mid \theta]\right] \\ &\stackrel{\text{A.3}}{=} \sum_{i=1}^N w_i^2 \text{Var}[\mathbb{E}[c_i \mid \theta]] = \sum_{i=1}^N w_i^2 \text{Var}[\mathbb{E}[c_i \mid \mu_i, \sigma_i^2]] = \sum_{i=1}^N w_i^2 U_i^{\text{epis}}, \end{aligned} \quad (29)$$

where $\stackrel{\text{A.}n}{=}$ denotes the step where Assumption n is used.

These three equations are connected by the law of total variance, as

$$\underbrace{\text{Var}[c]}_U = \underbrace{\mathbb{E}[\text{Var}[c \mid \theta]]}_{U^{\text{alea}}} + \underbrace{\text{Var}[\mathbb{E}[c \mid \theta]]}_{U^{\text{epis}}}, \quad (30)$$

$$\underbrace{\text{Var}[c_i]}_{U_i} = \underbrace{\mathbb{E}[\text{Var}[c_i \mid \mu_i, \sigma_i^2]]}_{U_i^{\text{alea}}} + \underbrace{\text{Var}[\mathbb{E}[c_i \mid \mu_i, \sigma_i^2]]}_{U_i^{\text{epis}}}. \quad (31)$$

Therefore, any two of Equations (27) to (29) imply the third, and thus only two of Assumptions 1 to 3 are necessary to derive all the three equations.

9. Derivations: Pixel Radiance Marginal Distribution and Loss Function

This section derives the marginal distribution of pixel radiance and the negative log-likelihood loss in Equations (21) and (23).

Evidential NeRF defines each pixel radiance c by a hierarchical probabilistic model

$$c \mid \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2), \quad \mu \mid \sigma^2 \sim \mathcal{N}(\gamma, \sigma^2/\nu), \quad \sigma^2 \sim \Gamma^{-1}(\alpha, \beta), \quad (32)$$

where $\mathcal{N}(\cdot, \cdot)$ and $\Gamma^{-1}(\cdot, \cdot)$ respectively denote normal distribution and inverse-gamma distribution and $\gamma \in \mathbb{R}$, $\nu > 0$, $\alpha > 1$, and $\beta > 0$ are the evidential NIG parameters. Based on these prerequisites, we derive the marginal distribution and negative log-likelihood of the pixel color c .

The probability density functions of the distributions in Equation (32) are respectively given by

$$p(c \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(c - \mu)^2}{2\sigma^2}\right), \quad (33)$$

$$p(\mu \mid \sigma^2) = \frac{\sqrt{\nu}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mu - \gamma)^2\nu}{2\sigma^2}\right), \quad (34)$$

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right). \quad (35)$$

It can therefore be shown that

$$p(c \mid \sigma^2) \quad (36)$$

$$= \int_{-\infty}^{\infty} p(c \mid \mu, \sigma^2) p(\mu \mid \sigma^2) d\mu \quad (37)$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(c - \mu)^2}{2\sigma^2}\right) \frac{\sqrt{\nu}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mu - \gamma)^2\nu}{2\sigma^2}\right) d\mu \quad (38)$$

$$= \int_{-\infty}^{\infty} \frac{\sqrt{\nu}}{2\pi\sigma^2} \exp\left(-\frac{(c - \mu)^2 + (\mu - \gamma)^2\nu}{2\sigma^2}\right) d\mu \quad (39)$$

$$= \int_{-\infty}^{\infty} \frac{\sqrt{\nu}}{2\pi\sigma^2} \exp\left(-\frac{(\nu + 1)\mu^2 - 2(c + \gamma\nu)\mu + (c^2 + \gamma^2\nu)}{2\sigma^2}\right) d\mu \quad (40)$$

$$= \frac{\sqrt{\nu}}{2\pi\sigma^2} \exp\left(-\frac{c^2 + \gamma^2\nu}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(\nu + 1)\mu^2 - 2(c + \gamma\nu)\mu}{2\sigma^2}\right) d\mu \quad (41)$$

$$= \frac{\sqrt{\nu}}{2\pi\sigma^2} \exp\left(-\frac{c^2 + \gamma^2\nu}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(\mu - \frac{c + \gamma\nu}{\nu + 1})^2 - (\frac{c + \gamma\nu}{\nu + 1})^2}{\frac{2\sigma^2}{\nu + 1}}\right) d\mu \quad (42)$$

$$= \frac{\sqrt{\nu}}{2\pi\sigma^2} \exp\left(-\frac{c^2 + \gamma^2\nu}{2\sigma^2}\right) \exp\left(\frac{(c + \gamma\nu)^2}{2\sigma^2(\nu + 1)}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{(\mu - \frac{c + \gamma\nu}{\nu + 1})^2}{\frac{2\sigma^2}{\nu + 1}}\right) d\mu \quad (43)$$

$$= \frac{\sqrt{\nu}}{2\pi\sigma^2} \exp\left(-\frac{c^2 + \gamma^2\nu}{2\sigma^2} + \frac{(c + \gamma\nu)^2}{2\sigma^2(\nu + 1)}\right) \int_{-\infty}^{\infty} \sqrt{2\pi\frac{\sigma^2}{\nu + 1}} \mathcal{N}\left(\mu; \frac{c + \gamma\nu}{\nu + 1}, \frac{\sigma^2}{\nu + 1}\right) d\mu \quad (44)$$

$$= \frac{\sqrt{\nu}}{2\pi\sigma^2} \sqrt{2\pi\frac{\sigma^2}{\nu + 1}} \exp\left(-\frac{(c^2 + \gamma^2\nu)(\nu + 1) - (c + \gamma\nu)^2}{2\sigma^2(\nu + 1)}\right) \int_{-\infty}^{\infty} \mathcal{N}\left(\mu; \frac{c + \gamma\nu}{\nu + 1}, \frac{\sigma^2}{\nu + 1}\right) d\mu \quad (45)$$

$$= \frac{\sqrt{\nu}}{2\pi\sigma^2} \sqrt{2\pi\frac{\sigma^2}{\nu + 1}} \exp\left(-\frac{(c^2 + \gamma^2\nu)(\nu + 1) - (c + \gamma\nu)^2}{2\sigma^2(\nu + 1)}\right) \quad (46)$$

$$= \sqrt{\frac{\nu}{2\pi\sigma^2(\nu + 1)}} \exp\left(-\frac{(c - \gamma)^2\nu}{2\sigma^2(\nu + 1)}\right), \quad (47)$$

i.e.,

$$c \mid \sigma^2 \sim \mathcal{N}\left(\gamma, \frac{\sigma^2(\nu + 1)}{\nu}\right), \quad (48)$$

and thus,

$$p(c) \tag{49}$$

$$= \int_0^\infty p(c | \sigma^2) p(\sigma^2) d\sigma^2 \tag{50}$$

$$= \int_0^\infty \sqrt{\frac{\nu}{2\pi\sigma^2(\nu+1)}} \exp\left(-\frac{(c-\gamma)^2\nu}{2\sigma^2(\nu+1)}\right) \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} \exp\left(-\frac{\beta}{\sigma^2}\right) d\sigma^2 \tag{51}$$

$$= \sqrt{\frac{\nu}{2\pi(\nu+1)}} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \exp\left(-\frac{(c-\gamma)^2\nu}{2\sigma^2(\nu+1)} - \frac{\beta}{\sigma^2}\right) (\sigma^2)^{-\alpha-\frac{3}{2}} d\sigma^2 \tag{52}$$

$$= \sqrt{\frac{\nu}{2\pi(\nu+1)}} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \exp\left(-\frac{\frac{(c-\gamma)^2\nu}{2(\nu+1)} + \beta}{\sigma^2}\right) (\sigma^2)^{-\alpha-\frac{3}{2}} d\sigma^2 \tag{53}$$

$$= \sqrt{\frac{\nu}{2\pi(\nu+1)}} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\Gamma(\alpha + \frac{1}{2})}{\left(\frac{(c-\gamma)^2\nu}{2(\nu+1)} + \beta\right)^{\alpha+\frac{1}{2}}} \Gamma^{-1}\left(\sigma^2; \alpha + \frac{1}{2}, \frac{(c-\gamma)^2\nu}{2(\nu+1)} + \beta\right) d\sigma^2 \tag{54}$$

$$= \sqrt{\frac{\nu}{2\pi(\nu+1)}} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \frac{1}{2})}{\left(\frac{(c-\gamma)^2\nu}{2(\nu+1)} + \beta\right)^{\alpha+\frac{1}{2}}} \int_0^\infty \Gamma^{-1}\left(\sigma^2; \alpha + \frac{1}{2}, \frac{(c-\gamma)^2\nu}{2(\nu+1)} + \beta\right) d\sigma^2 \tag{55}$$

$$= \sqrt{\frac{\nu}{2\pi(\nu+1)}} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \frac{1}{2})}{\left(\frac{(c-\gamma)^2\nu}{2(\nu+1)} + \beta\right)^{\alpha+\frac{1}{2}}} \tag{56}$$

$$= \frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha) \sqrt{2\pi \frac{\beta(\nu+1)}{\nu}}} \left(\frac{(c-\gamma)^2}{2\frac{\beta(\nu+1)}{\nu}} + 1\right)^{-(\alpha+\frac{1}{2})} \tag{57}$$

$$= \frac{\Gamma(\frac{\nu_t+1}{2})}{\Gamma(\frac{\nu_t}{2}) \sqrt{\pi \nu_t \sigma_t^2}} \left(\frac{(c-\mu_t)^2}{\nu_t \sigma_t^2} + 1\right)^{-\frac{\nu_t+1}{2}}, \tag{58}$$

i.e.,

$$c \sim t\left(\mu_t = \gamma, \sigma_t^2 = \frac{\beta(\nu+1)}{\alpha\nu}, \nu_t = 2\alpha\right), \tag{59}$$

where $t(\mu_t, \sigma_t^2, \nu_t)$ denotes a Student's t distribution with location μ_t , scale σ_t , and degrees of freedom ν_t , and $\mathcal{N}(x; \cdot, \cdot)$ and $\Gamma^{-1}(x; \cdot, \cdot)$ represent their probability distribution densities at x . This yields Equation (21).

The negative log-likelihood of c can thereby be derived as

$$-\log p(c) \tag{60}$$

$$= -\log\left(\frac{\Gamma(\alpha + \frac{1}{2})}{\Gamma(\alpha) \sqrt{2\pi \frac{\beta(\nu+1)}{\nu}}} \left(\frac{(c-\gamma)^2}{2\frac{\beta(\nu+1)}{\nu}} + 1\right)^{-(\alpha+\frac{1}{2})}\right) \tag{61}$$

$$= \frac{1}{2} \log\left(2\pi \frac{\beta(\nu+1)}{\nu}\right) + \log \frac{\Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})} + \left(\alpha + \frac{1}{2}\right) \log\left(\frac{(c-\gamma)^2}{2\frac{\beta(\nu+1)}{\nu}} + 1\right) \tag{62}$$

$$= \frac{1}{2} \log \frac{\pi}{\nu} + \frac{1}{2} \log(2\beta(\nu+1)) + \log \frac{\Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})} + \left(\alpha + \frac{1}{2}\right) \log\left(\frac{(c-\gamma)^2\nu}{2\beta(\nu+1)} + 1\right) \tag{63}$$

$$= \frac{1}{2} \log \frac{\pi}{\nu} - \alpha \log(2\beta(\nu+1)) + \log \frac{\Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})} + \left(\alpha + \frac{1}{2}\right) \log((c-\gamma)^2\nu + 2\beta(\nu+1)) \tag{64}$$

$$= \frac{1}{2} \log \frac{\pi}{\nu} - \alpha \log \Omega + \log \frac{\Gamma(\alpha)}{\Gamma(\alpha + \frac{1}{2})} + \left(\alpha + \frac{1}{2}\right) \log((c-\gamma)^2\nu + \Omega), \tag{65}$$

where $\Omega = 2\beta(1+\nu)$. This yields Equation (23).

10. Further Discussions

More quantitative results. In the main paper, quantitative results averaged across all scenes within each dataset are reported. Tables 4 to 7 provide per-scene statistics including the mean and standard deviation of metrics over three independent runs.

More qualitative results. We present the qualitative comparison of the uncertainty methods on LF, LLFF, and RobustNeRF datasets in Figures 11 to 13, respectively. Additionally, we show the aleatoric and epistemic uncertainty maps of more scenes in the wild from Phototoursim [16] in Figure 14.

Hyperparameter selection. The regularization coefficient in the loss function is selected based on the quantitative metrics. Figure 9 illustrates how different coefficients affect image reconstruction and uncertainty estimation. In general, both excessively small and large coefficients lead to suboptimal performance, and the best hyperparameter value is inherently scene-dependent. The specific regularization coefficients utilized to produce the reported results are detailed in Table 3.

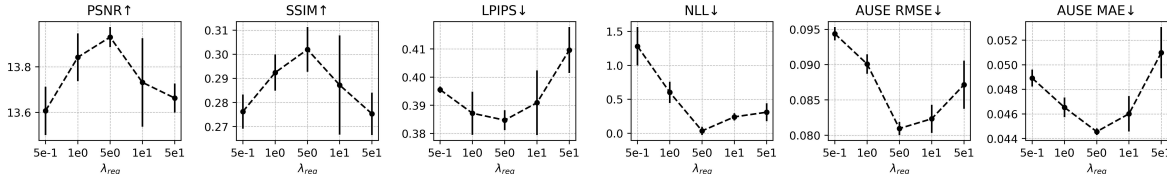


Figure 9. Sensitivity study of regularization coefficient’s effect on all the quantitative metrics of *Leaves* scene.

Scene	Africa	Basket	Statue	Torch	Fern	Flower	Fortress	Horns	Leaves	Orchids	Room	T-Rex	Android	Crab	Statue	Yoda
λ_{reg}	5e-3	1e-1	1e-3	1e0	1e0	5e-2	1e0	5e-1	5e0	1e0	5e-1	5e0	5e-3	5e-4	1e-4	1e-5

Table 3. The regularization coefficients used in each scene of LF, LLFF, and RobustNeRF datasets.

Mutual causes of aleatoric and epistemic uncertainties. Various complex elements in a 3D scene can lead to elevated levels of either aleatoric or epistemic uncertainty. However, attributing each specific factor exclusively to one type of uncertainty is often inappropriate, as many factors affect both uncertainties in different ways and to varying degrees. For example, transient objects increase AU due to color variations introduced by motion, while simultaneously raising EU through partial occlusions. Similarly, edges or high-frequency non-smooth regions tend to exhibit higher AU since their radiance is highly sensitive to input rays, as small inaccuracies in sensing, digitization, or poses can yield large radiance variations, resulting in nearly irreducible data uncertainty; meanwhile, the irregular geometry of such regions obstructs ray coverage and limits supervision signals from those surfaces, thereby increasing EU as well.

Uncertainties of transients. Transient objects can lead to both higher aleatoric and epistemic uncertainties. In practice, the uncertainties of the transient regions depend on the densities assigned to them by the model. Figure 10 shows ten images from two RobustNeRF scenes where the model is trained on images with cluttered objects. It can be observed that when the model cannot disambiguate the transients and the floaters appear in the test renderings, both AU and EU tend to be higher on them, meaning that the model simultaneously receives inconsistent radiance signals (high AU) and lacks sufficient knowledge to determine the presence or geometry of the transients (high EU). When the model resolves the transient objects (by minimizing their densities and removing them from volumetric rendering), the floaters disappear in the test image reconstructions and only AU remains high, indicating that the model no longer lacks the knowledge to determine the presence of the transients but still records the color inconsistency from the training signals as high AU.

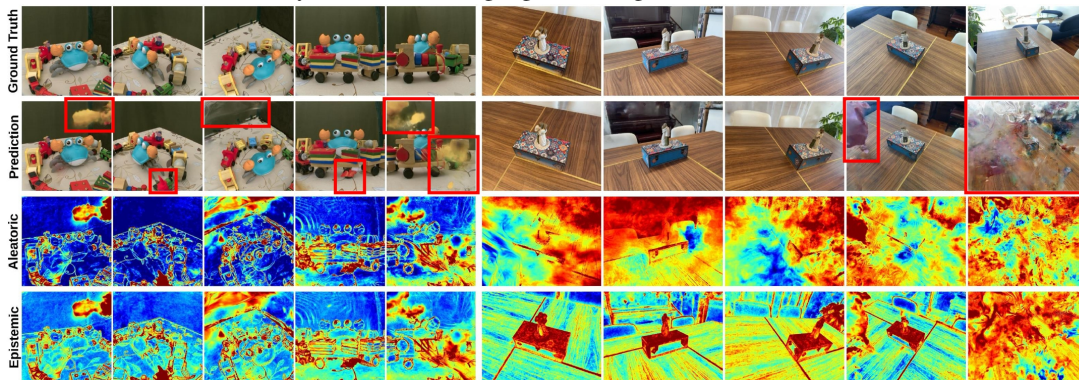


Figure 10. Aleatoric and epistemic uncertainties of scenes with transient objects. The red bounding boxes delineate the erroneous artifacts in test renderings caused by transients in the training views. If the model fails to suppress the floaters, both AU and EU are elevated on the transients; If the model resolves the transience, only AU is higher on the regions where the training transients were once present.

Scene	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NLL $\downarrow$	AUSE RMSE $\downarrow$	AUSE MAE $\downarrow$
Africa	Baseline	26.5406 ± 0.1903	0.9019 ± 0.0009	0.0565 ± 0.0010			
	Dropout	26.6798 ± 0.0205	0.8986 ± 0.0007	0.0590 ± 0.0014	4.0201 ± 0.2676	0.0148 ± 0.0003	0.0064 ± 0.0001
	Normal	27.4766 ± 1.5843	0.9028 ± 0.0192	0.0608 ± 0.0210	-1.4911 ± 0.6587	0.0078 ± 0.0023	0.0033 ± 0.0004
	MoL	28.0663 ± 0.6656	0.9023 ± 0.0082	0.0614 ± 0.0056	-2.3375 ± 0.0088	0.0069 ± 0.0001	0.0035 ± 0.0001
	Ensembles	27.0083 ± 0.0652	0.9171 ± 0.0003	0.0491 ± 0.0006	0.0871 ± 0.1899	0.0087 ± 0.0005	0.0037 ± 0.0001
	DANE	27.0083 ± 0.0652	0.9171 ± 0.0003	0.0491 ± 0.0006	-0.9581 ± 0.1721	0.0139 ± 0.0010	0.0059 ± 0.0003
	Evidential	29.8826 ± 0.0617	0.9331 ± 0.0002	0.0340 ± 0.0015	-2.3643 ± 0.0030	0.0054 ± 0.0000	0.0028 ± 0.0000
Basket	Baseline	28.0171 ± 0.0703	0.9114 ± 0.0019	0.0474 ± 0.0008			
	Dropout	27.3038 ± 0.0176	0.8897 ± 0.0018	0.0610 ± 0.0016	6.2637 ± 0.3747	0.0155 ± 0.0005	0.0054 ± 0.0002
	Normal	27.9387 ± 0.6029	0.9087 ± 0.0097	0.0517 ± 0.0044	5.5852 ± 2.7302	0.0137 ± 0.0033	0.0043 ± 0.0013
	MoL	27.5847 ± 0.2379	0.9044 ± 0.0007	0.0692 ± 0.0015	-2.2614 ± 0.0438	0.0125 ± 0.0016	0.0036 ± 0.0002
	Ensembles	28.9951 ± 0.0715	0.9258 ± 0.0011	0.0427 ± 0.0004	-0.8529 ± 0.3663	0.0058 ± 0.0002	0.0023 ± 0.0001
	DANE	28.9951 ± 0.0715	0.9258 ± 0.0011	0.0427 ± 0.0004	-1.0245 ± 0.3048	0.0087 ± 0.0010	0.0034 ± 0.0004
	Evidential	29.1442 ± 0.1841	0.9263 ± 0.0008	0.0369 ± 0.0002	-2.1757 ± 0.0901	0.0098 ± 0.0008	0.0033 ± 0.0001
Statue	Baseline	32.8018 ± 0.1611	0.9645 ± 0.0007	0.0221 ± 0.0007			
	Dropout	31.8005 ± 0.0787	0.9525 ± 0.0009	0.0377 ± 0.0007	-1.1505 ± 0.0358	0.0057 ± 0.0002	0.0031 ± 0.0001
	Normal	28.8110 ± 1.1896	0.9457 ± 0.0036	0.0424 ± 0.0033	-1.9682 ± 0.4526	0.0055 ± 0.0010	0.0017 ± 0.0001
	MoL	30.6262 ± 0.5385	0.9469 ± 0.0083	0.0545 ± 0.0174	-2.9409 ± 0.0318	0.0030 ± 0.0002	0.0016 ± 0.0001
	Ensembles	33.7663 ± 0.0662	0.9718 ± 0.0001	0.0198 ± 0.0000	-1.9914 ± 0.0994	0.0029 ± 0.0001	0.0017 ± 0.0000
	DANE	33.7663 ± 0.0662	0.9718 ± 0.0001	0.0198 ± 0.0000	-2.1733 ± 0.0538	0.0034 ± 0.0002	0.0020 ± 0.0001
	Evidential	32.8506 ± 0.2748	0.9661 ± 0.0003	0.0221 ± 0.0003	-2.9793 ± 0.0192	0.0031 ± 0.0001	0.0014 ± 0.0000
Torch	Baseline	26.8557 ± 0.0802	0.8911 ± 0.0010	0.0599 ± 0.0014			
	Dropout	26.6709 ± 0.0290	0.8835 ± 0.0005	0.0698 ± 0.0007	5.5346 ± 0.4652	0.0139 ± 0.0005	0.0046 ± 0.0001
	Normal	27.7991 ± 0.1333	0.9087 ± 0.0019	0.0574 ± 0.0015	-0.3559 ± 1.1275	0.0090 ± 0.0022	0.0023 ± 0.0002
	MoL	26.6030 ± 0.1090	0.8845 ± 0.0028	0.0838 ± 0.0042	-2.6173 ± 0.0145	0.0112 ± 0.0005	0.0028 ± 0.0000
	Ensembles	27.7419 ± 0.0417	0.9086 ± 0.0004	0.0528 ± 0.0002	4.0553 ± 1.4211	0.0105 ± 0.0006	0.0026 ± 0.0001
	DANE	27.7419 ± 0.0417	0.9086 ± 0.0004	0.0528 ± 0.0002	2.4292 ± 0.9753	0.0144 ± 0.0013	0.0044 ± 0.0005
	Evidential	27.9943 ± 0.1674	0.9126 ± 0.0017	0.0507 ± 0.0041	-2.2771 ± 0.1350	0.0096 ± 0.0022	0.0025 ± 0.0003

Table 4. Mean and standard deviation of quantitative metrics over three runs on LF.

Scene	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NLL $\downarrow$	AUSE RMSE $\downarrow$	AUSE MAE $\downarrow$
Fern	Baseline	20.1069 ± 0.1190	0.5685 ± 0.0005	0.3617 ± 0.0050			
	Dropout	19.1317 ± 0.1808	0.5153 ± 0.0059	0.4432 ± 0.0060	46.5072 ± 8.3738	0.0517 ± 0.0016	0.0294 ± 0.0020
	Normal	20.1763 ± 0.3308	0.5905 ± 0.0115	0.3273 ± 0.0133	47.5151 ± 1.6352	0.0349 ± 0.0014	0.0193 ± 0.0005
	MoL	18.2953 ± 0.6532	0.5222 ± 0.0298	0.4279 ± 0.0406	1.3336 ± 0.2902	0.0605 ± 0.0157	0.0257 ± 0.0078
	Ensembles	20.7328 ± 0.0591	0.6134 ± 0.0016	0.3462 ± 0.0019	3.2024 ± 0.4579	0.0297 ± 0.0004	0.0148 ± 0.0002
	DANE	20.7328 ± 0.0591	0.6134 ± 0.0016	0.3462 ± 0.0019	2.6334 ± 0.4099	0.0332 ± 0.0002	0.0172 ± 0.0001
	Evidential	20.8095 ± 0.1119	0.6216 ± 0.0074	0.3005 ± 0.0077	-0.4856 ± 0.1133	0.0301 ± 0.0009	0.0160 ± 0.0007
Flower	Baseline	18.3456 ± 0.0046	0.4792 ± 0.0068	0.4230 ± 0.0116			
	Dropout	18.6701 ± 0.1790	0.4990 ± 0.0089	0.3780 ± 0.0057	54.1338 ± 2.6904	0.0554 ± 0.0012	0.0341 ± 0.0012
	Normal	12.6188 ± 4.7635	0.2205 ± 0.1075	0.7529 ± 0.2290	9.0113 ± 6.3991	0.0732 ± 0.0073	0.0590 ± 0.0115
	MoL	18.6338 ± 0.2226	0.5140 ± 0.0055	0.3665 ± 0.0141	1.4129 ± 0.1601	0.0589 ± 0.0023	0.0296 ± 0.0012
	Ensembles	18.7711 ± 0.0344	0.5204 ± 0.0036	0.3993 ± 0.0050	7.7562 ± 0.2043	0.0370 ± 0.0004	0.0173 ± 0.0002
	DANE	18.7711 ± 0.0344	0.5204 ± 0.0036	0.3993 ± 0.0050	7.7131 ± 0.2053	0.0371 ± 0.0004	0.0174 ± 0.0002
	Evidential	19.2606 ± 0.1150	0.5254 ± 0.0126	0.3957 ± 0.0208	2.4728 ± 0.6069	0.0409 ± 0.0022	0.0238 ± 0.0015
Fortress	Baseline	18.3478 ± 0.1598	0.3906 ± 0.0033	0.5240 ± 0.0069			
	Dropout	18.3708 ± 0.5357	0.4063 ± 0.0039	0.5204 ± 0.0077	52.6142 ± 9.0302	0.0518 ± 0.0069	0.0278 ± 0.0022
	Normal	18.3146 ± 0.0871	0.4100 ± 0.0106	0.5060 ± 0.0427	62.8832 ± 20.7391	0.0515 ± 0.0036	0.0248 ± 0.0027
	MoL	17.4740 ± 0.1443	0.3779 ± 0.0064	0.6090 ± 0.0253	1.8754 ± 0.5340	0.0645 ± 0.0009	0.0280 ± 0.0003
	Ensembles	18.8273 ± 0.0496	0.4438 ± 0.0015	0.4517 ± 0.0023	7.5589 ± 0.8103	0.0369 ± 0.0014	0.0188 ± 0.0003
	DANE	18.8273 ± 0.0496	0.4438 ± 0.0015	0.4517 ± 0.0023	7.1596 ± 0.8387	0.0377 ± 0.0017	0.0195 ± 0.0003
	Evidential	18.7310 ± 0.0406	0.4264 ± 0.0065	0.4878 ± 0.0040	0.2943 ± 0.1765	0.0404 ± 0.0011	0.0208 ± 0.0005
Horns	Baseline	15.7052 ± 0.1371	0.4527 ± 0.0105	0.4451 ± 0.0111			
	Dropout	15.4649 ± 0.2658	0.4296 ± 0.0113	0.4729 ± 0.0074	159.8088 ± 22.3326	0.0919 ± 0.0057	0.0482 ± 0.0025
	Normal	13.4923 ± 0.2916	0.2336 ± 0.0089	0.6160 ± 0.0096	22.6808 ± 7.0886	0.1130 ± 0.0047	0.0807 ± 0.0083
	MoL	14.5633 ± 0.1664	0.3638 ± 0.0457	0.5691 ± 0.0891	2.1617 ± 0.2982	0.1101 ± 0.0039	0.0505 ± 0.0008
	Ensembles	15.9843 ± 0.0172	0.5025 ± 0.0033	0.4240 ± 0.0045	22.0785 ± 1.0690	0.0759 ± 0.0013	0.0328 ± 0.0005
	DANE	15.9843 ± 0.0172	0.5025 ± 0.0033	0.4240 ± 0.0045	15.3955 ± 1.4795	0.0750 ± 0.0014	0.0330 ± 0.0006
	Evidential	15.7596 ± 0.1745	0.5034 ± 0.0074	0.3878 ± 0.0075	1.8495 ± 0.1097	0.0941 ± 0.0055	0.0403 ± 0.0023

Table 5. Mean and standard deviation of quantitative metrics over three runs on LLFF.

Scene	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NLL $\downarrow$	AUSE RMSE $\downarrow$	AUSE MAE $\downarrow$
Leaves	Baseline	13.6976 ± 0.1464	0.2557 ± 0.0155	0.4167 ± 0.0066			
	Dropout	13.9549 ± 0.2228	0.2423 ± 0.0248	0.4235 ± 0.0211	141.2669 ± 12.2890	0.1072 ± 0.0025	0.0618 ± 0.0024
	Normal	13.3743 ± 0.0707	0.2525 ± 0.0111	0.4239 ± 0.0075	104.3151 ± 41.1687	0.1064 ± 0.0013	0.0553 ± 0.0009
	MoL	12.8939 ± 0.1964	0.2408 ± 0.0179	0.4276 ± 0.0162	4.1014 ± 0.2543	0.1266 ± 0.0053	0.0674 ± 0.0044
	Ensembles	14.2397 ± 0.0569	0.2960 ± 0.0069	0.4319 ± 0.0047	12.3460 ± 0.5912	0.0915 ± 0.0010	0.0471 ± 0.0007
	DANE	14.2397 ± 0.0569	0.2960 ± 0.0069	0.4319 ± 0.0047	11.6017 ± 0.1579	0.0928 ± 0.0009	0.0482 ± 0.0006
	Evidential	13.9301 ± 0.0437	0.3020 ± 0.0093	0.3847 ± 0.0035	0.0349 ± 0.0607	0.0809 ± 0.0009	0.0446 ± 0.0003
Orchids	Baseline	14.5292 ± 0.0830	0.3036 ± 0.0057	0.3916 ± 0.0103			
	Dropout	13.3110 ± 0.4149	0.2151 ± 0.0202	0.4709 ± 0.0100	123.9834 ± 10.5705	0.1195 ± 0.0056	0.0762 ± 0.0037
	Normal	14.9466 ± 0.1068	0.3189 ± 0.0126	0.3911 ± 0.0160	25.0742 ± 5.0559	0.0754 ± 0.0028	0.0424 ± 0.0013
	MoL	13.8818 ± 0.0307	0.3242 ± 0.0031	0.3706 ± 0.0047	2.4593 ± 0.0292	0.1058 ± 0.0045	0.0506 ± 0.0021
	Ensembles	14.8488 ± 0.1379	0.3250 ± 0.0081	0.3953 ± 0.0065	9.7556 ± 0.8979	0.0753 ± 0.0037	0.0409 ± 0.0017
	DANE	14.8488 ± 0.1379	0.3250 ± 0.0081	0.3953 ± 0.0065	9.4356 ± 0.7865	0.0746 ± 0.0034	0.0407 ± 0.0016
	Evidential	14.7012 ± 0.1548	0.3253 ± 0.0090	0.3601 ± 0.0103	0.5681 ± 0.2424	0.0843 ± 0.0041	0.0459 ± 0.0017
Room	Baseline	19.7859 ± 0.0325	0.7196 ± 0.0037	0.3864 ± 0.0088			
	Dropout	19.1522 ± 0.0418	0.6704 ± 0.0097	0.4536 ± 0.0085	79.6637 ± 8.4028	0.0637 ± 0.0016	0.0342 ± 0.0006
	Normal	19.8485 ± 0.2959	0.6766 ± 0.0384	0.4178 ± 0.0630	111.9164 ± 13.8320	0.0465 ± 0.0013	0.0232 ± 0.0020
	MoL	17.6348 ± 0.7025	0.5337 ± 0.0754	0.6094 ± 0.0874	3.1864 ± 0.4821	0.0651 ± 0.0042	0.0295 ± 0.0026
	Ensembles	19.9259 ± 0.0347	0.7492 ± 0.0003	0.3656 ± 0.0006	15.8667 ± 0.7134	0.0347 ± 0.0006	0.0161 ± 0.0004
	DANE	19.9259 ± 0.0347	0.7492 ± 0.0003	0.3656 ± 0.0006	13.4646 ± 0.4077	0.0368 ± 0.0004	0.0174 ± 0.0005
	Evidential	19.9548 ± 0.0570	0.7172 ± 0.0181	0.3660 ± 0.0299	1.2947 ± 1.1068	0.0465 ± 0.0027	0.0221 ± 0.0012
T-Rex	Baseline	19.7067 ± 0.0095	0.6063 ± 0.0041	0.3463 ± 0.0047			
	Dropout	19.0824 ± 0.1470	0.5750 ± 0.0059	0.4119 ± 0.0046	67.4407 ± 5.4653	0.0525 ± 0.0006	0.0290 ± 0.0006
	Normal	19.6283 ± 0.1296	0.6055 ± 0.0176	0.3502 ± 0.0238	59.4677 ± 29.4200	0.0379 ± 0.0026	0.0189 ± 0.0008
	MoL	17.9974 ± 0.5803	0.5127 ± 0.0837	0.4935 ± 0.1085	1.4452 ± 0.3356	0.0594 ± 0.0049	0.0255 ± 0.0029
	Ensembles	20.0150 ± 0.0311	0.6373 ± 0.0019	0.3317 ± 0.0028	10.7621 ± 0.1102	0.0293 ± 0.0004	0.0145 ± 0.0002
	DANE	20.0150 ± 0.0311	0.6373 ± 0.0019	0.3317 ± 0.0028	10.4152 ± 0.0855	0.0295 ± 0.0004	0.0146 ± 0.0002
	Evidential	19.8874 ± 0.0246	0.6329 ± 0.0048	0.3177 ± 0.0095	-0.6168 ± 0.0407	0.0455 ± 0.0004	0.0226 ± 0.0006

Table 6. Mean and standard deviation of quantitative metrics over three runs on LLFF.

Scene	Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	NLL $\downarrow$	AUSE RMSE $\downarrow$	AUSE MAE $\downarrow$
Android	Baseline	22.8865 ± 0.0476	0.7691 ± 0.0008	0.1545 ± 0.0011			
	Dropout	22.6094 ± 0.0779	0.7552 ± 0.0001	0.1699 ± 0.0009	17.8212 ± 0.1534	0.0308 ± 0.0002	0.0188 ± 0.0001
	Normal	23.9819 ± 0.0770	0.8146 ± 0.0013	0.1058 ± 0.0021	4.9525 ± 0.7433	0.0227 ± 0.0006	0.0152 ± 0.0003
	MoL	22.1825 ± 0.0186	0.7395 ± 0.0002	0.1905 ± 0.0008	-0.7117 ± 0.0656	0.0273 ± 0.0006	0.0172 ± 0.0005
	Ensembles	23.4866 ± 0.0304	0.8012 ± 0.0006	0.1381 ± 0.0006	8.1690 ± 0.8485	0.0265 ± 0.0001	0.0158 ± 0.0000
	DANE	23.4866 ± 0.0304	0.8012 ± 0.0006	0.1381 ± 0.0006	7.4801 ± 0.8655	0.0274 ± 0.0001	0.0164 ± 0.0000
	Evidential	23.8915 ± 0.0651	0.8116 ± 0.0034	0.1047 ± 0.0007	-1.1616 ± 0.0266	0.0231 ± 0.0003	0.0153 ± 0.0002
Crab	Baseline	28.6708 ± 0.2399	0.9034 ± 0.0038	0.1124 ± 0.0072			
	Dropout	27.5569 ± 0.3792	0.8850 ± 0.0049	0.1186 ± 0.0065	24.1365 ± 2.5685	0.0216 ± 0.0012	0.0074 ± 0.0007
	Normal	27.7154 ± 0.2508	0.9062 ± 0.0016	0.0971 ± 0.0025	2.9771 ± 1.6485	0.0151 ± 0.0007	0.0062 ± 0.0004
	MoL	27.0696 ± 0.2685	0.8767 ± 0.0044	0.1374 ± 0.0061	-2.5692 ± 0.0274	0.0186 ± 0.0013	0.0065 ± 0.0005
	Ensembles	30.0907 ± 0.0696	0.9206 ± 0.0007	0.1072 ± 0.0006	-0.3212 ± 0.4515	0.0048 ± 0.0005	0.0021 ± 0.0001
	DANE	30.0907 ± 0.0696	0.9206 ± 0.0007	0.1072 ± 0.0006	-0.0123 ± 0.2259	0.0234 ± 0.0006	0.0101 ± 0.0000
	Evidential	29.9324 ± 0.0145	0.9160 ± 0.0086	0.0795 ± 0.0154	-2.0303 ± 0.2399	0.0150 ± 0.0059	0.0084 ± 0.0058
Statue	Baseline	20.1787 ± 0.0379	0.7489 ± 0.0046	0.2356 ± 0.0072			
	Dropout	20.0340 ± 0.0841	0.7213 ± 0.0068	0.2772 ± 0.0159	35.4746 ± 2.1971	0.0464 ± 0.0013	0.0304 ± 0.0008
	Normal	20.1453 ± 0.0154	0.7780 ± 0.0039	0.2128 ± 0.0061	17.9184 ± 2.0826	0.0413 ± 0.0008	0.0275 ± 0.0007
	MoL	19.3304 ± 0.1637	0.6834 ± 0.0032	0.3300 ± 0.0069	0.0773 ± 0.0274	0.0571 ± 0.0005	0.0380 ± 0.0007
	Ensembles	20.5674 ± 0.0059	0.7843 ± 0.0008	0.2166 ± 0.0028	9.9047 ± 0.3519	0.0297 ± 0.0003	0.0191 ± 0.0001
	DANE	20.5674 ± 0.0059	0.7843 ± 0.0008	0.2166 ± 0.0028	8.4305 ± 0.2763	0.0367 ± 0.0002	0.0235 ± 0.0000
	Evidential	20.8284 ± 0.0496	0.8017 ± 0.0010	0.1780 ± 0.0019	-0.0313 ± 0.3149	0.0362 ± 0.0008	0.0243 ± 0.0004
Yoda	Baseline	29.1458 ± 0.4079	0.8969 ± 0.0081	0.1284 ± 0.0118			
	Dropout	28.8770 ± 0.1990	0.8857 ± 0.0041	0.1322 ± 0.0020	14.0873 ± 5.1782	0.0148 ± 0.0034	0.0073 ± 0.0014
	Normal	29.3523 ± 0.5839	0.9099 ± 0.0040	0.1088 ± 0.0043	16.1012 ± 8.0429	0.0209 ± 0.0034	0.0115 ± 0.0018
	MoL	26.5671 ± 0.3613	0.8688 ± 0.0051	0.1617 ± 0.0070	-2.3751 ± 0.0676	0.0186 ± 0.0018	0.0088 ± 0.0010
	Ensembles	30.6364 ± 0.1516	0.9188 ± 0.0025	0.1135 ± 0.0029	0.7713 ± 0.1332	0.0044 ± 0.0001	0.0020 ± 0.0001
	DANE	30.6364 ± 0.1516	0.9188 ± 0.0025	0.1135 ± 0.0029	0.5383 ± 0.2742	0.0257 ± 0.0015	0.0116 ± 0.0006
	Evidential	30.2646 ± 0.1857	0.9272 ± 0.0065	0.0826 ± 0.0124	-1.8577 ± 0.2073	0.0139 ± 0.0009	0.0072 ± 0.0012

Table 7. Mean and standard deviation of quantitative metrics over three runs on RobustNeRF.

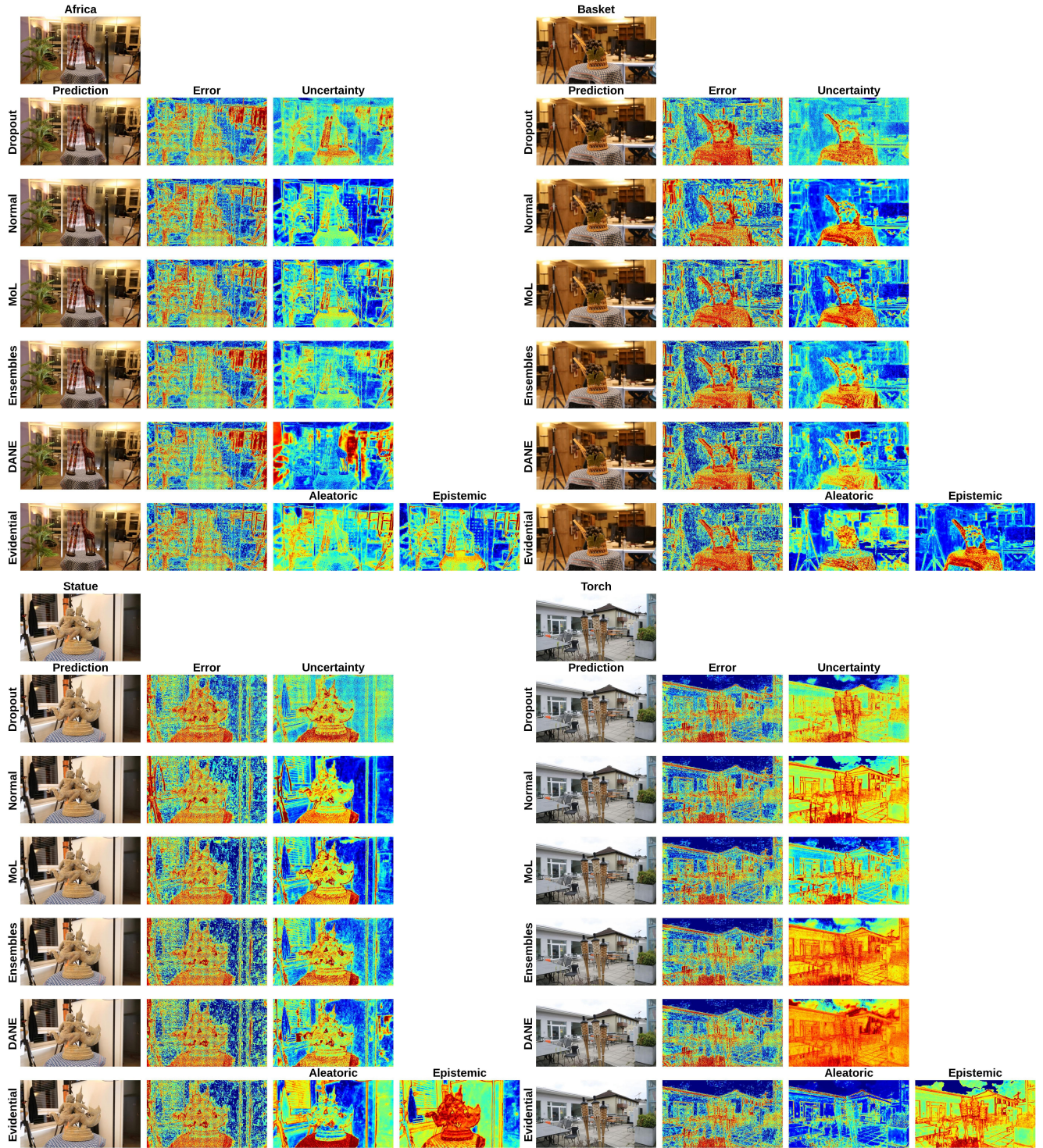


Figure 11. Qualitative comparison on LF.

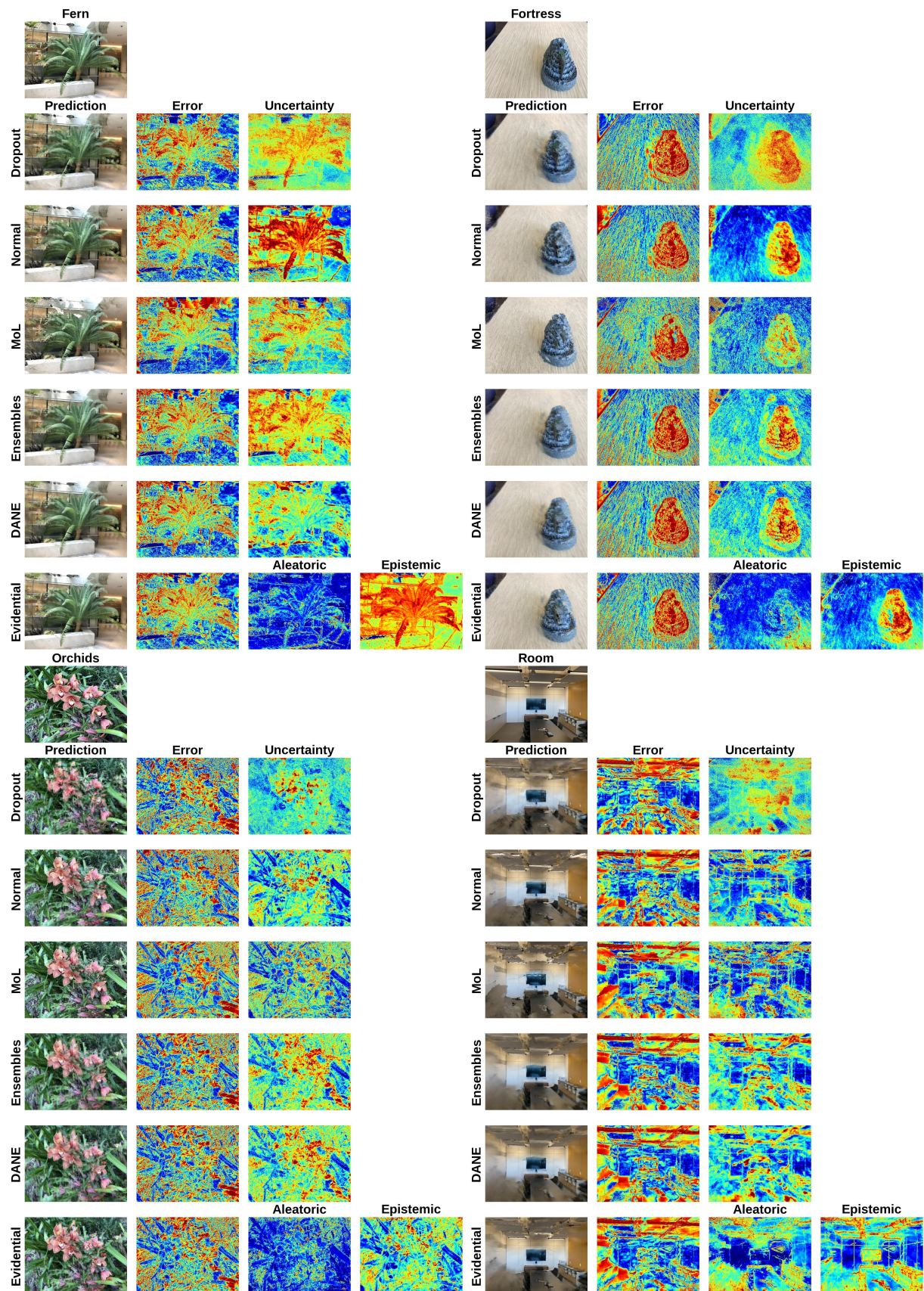


Figure 12. Qualitative comparison on LLFF.

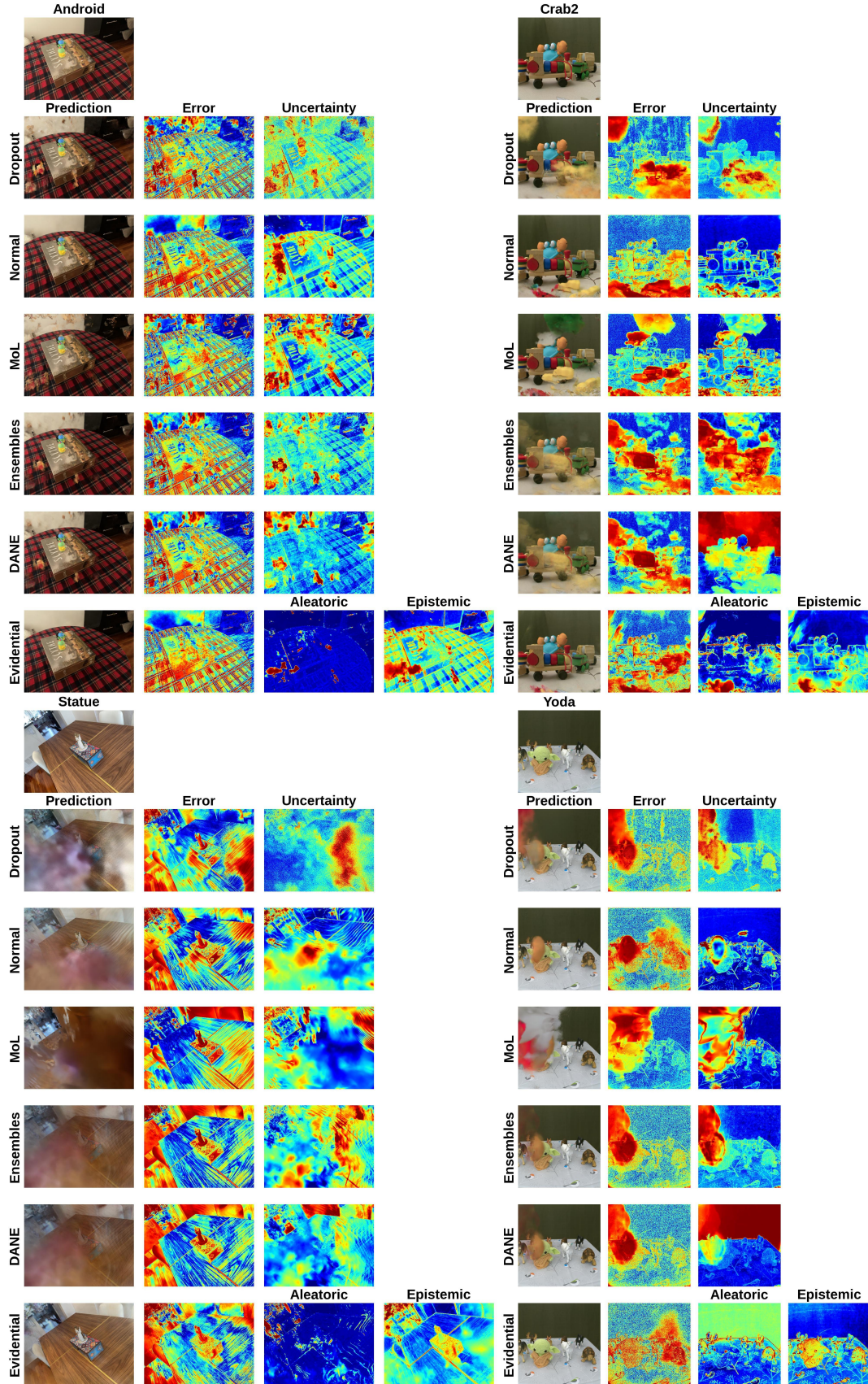


Figure 13. Qualitative comparison on RobustNeRF.

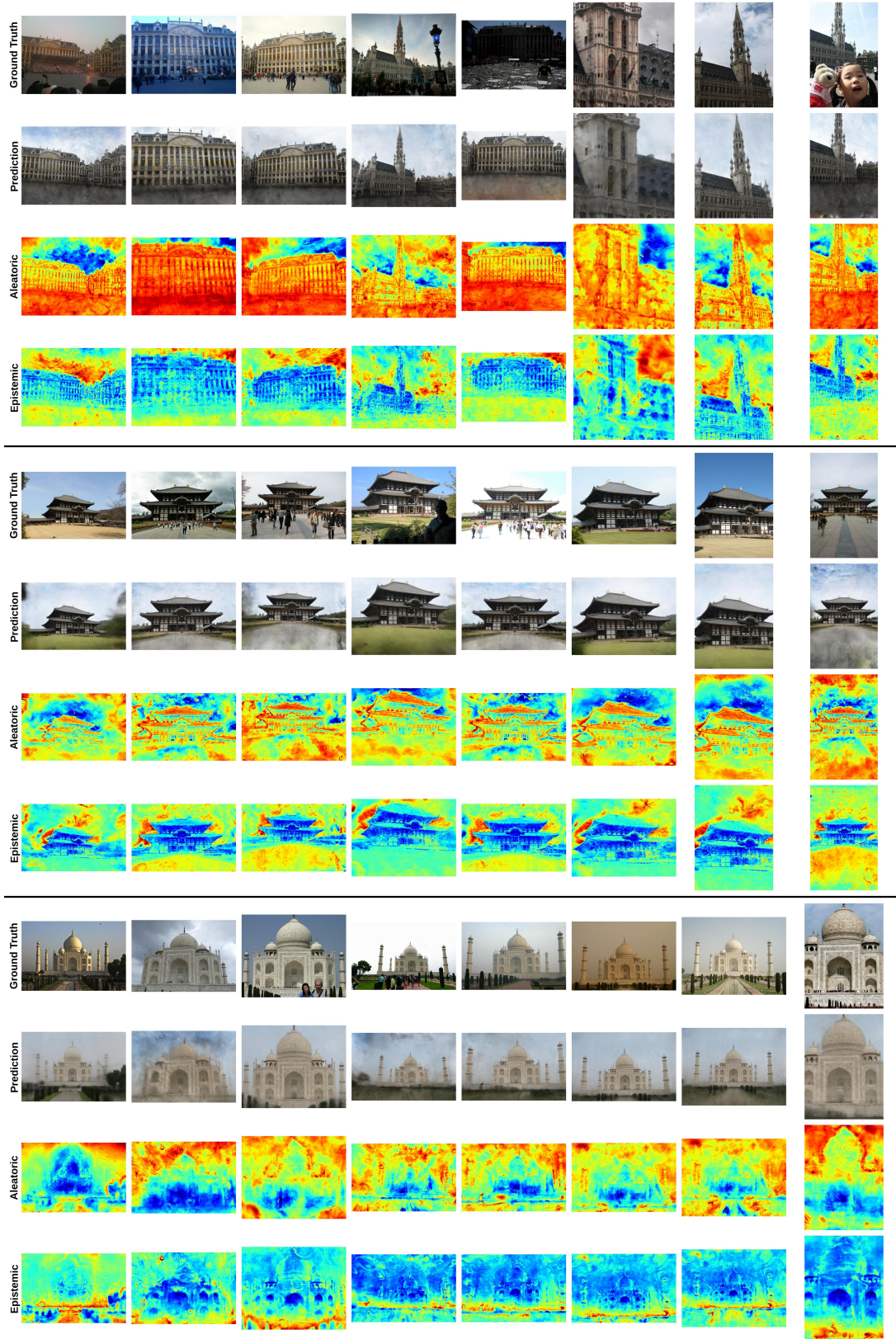


Figure 14. Aleatoric and epistemic uncertainty maps for three in-the-wild scenes from Phototourism. Aleatoric uncertainty peaks in regions with high radiance variance (e.g., sky higher above, building facades, moving figures), whereas epistemic uncertainty is concentrated in areas frequently occluded (e.g., sky directly behind the buildings or objects obscured by pedestrians).