Conditional-Flow NeRF: Accurate 3D Modelling with Reliable Uncertainty Quantification

Jianxiong Shen[1], Antonio Agudo[1], Francesc Moreno-Noguer[1], and Adria Ruiz[2]

[1] Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain
[2] Seedtag, Spain

Abstract. A critical limitation of current methods based on Neural Radiance Fields (NeRF) is that they are unable to quantify the uncertainty associated with the learned appearance and geometry of the scene. This information is paramount in real applications such as medical diagnosis or autonomous driving where, to reduce potentially catastrophic failures, the confidence on the model outputs must be included into the decision-making process. In this context, we introduce Conditional-Flow NeRF (CF-NeRF), a novel probabilistic framework to incorporate uncertainty quantification into NeRF-based approaches. For this purpose, our method learns a distribution over all possible radiance fields modelling the scene which is used to quantify the uncertainty associated with the modelled scene. In contrast to previous approaches enforcing strong constraints over the radiance field distribution, CF-NeRF learns it in a flexible and fully data-driven manner by coupling Latent Variable Modelling and Conditional Normalizing Flows. This strategy allows to obtain reliable uncertainty estimation while preserving model expressivity. Compared to previous state-of-the-art methods proposed for uncertainty quantification in NeRF, our experiments show that the proposed method achieves significantly lower prediction errors and more reliable uncertainty values for synthetic novel view and depth-map estimation.

Keywords: 3D, scene representation, uncertainty quantification

1 Introduction

Neural fields [62] have recently gained a lot of attention given its ability to encode implicit representations of complex 3D scenes using deep neural networks. Additionally, they have been shown very effective in addressing multiple problems such as 3D reconstruction [32,39], scene rendering [27,31] or human body representation [38,50]. Among different methods built upon this framework [41,5,55], Neural Radiance Fields (NeRF) [33] has obtained impressive results in generating photo-realistic views of 3D scenes and solving downstream tasks such as depth estimation [61], scene editing [26,15] or pose prediction [65,56].

Despite its increasing popularity, a critical limitation of NeRF holding back its application to real-world problems has been recently pointed out by [53]. Concretely, current NeRF-based methods are not able to quantify the uncertainty

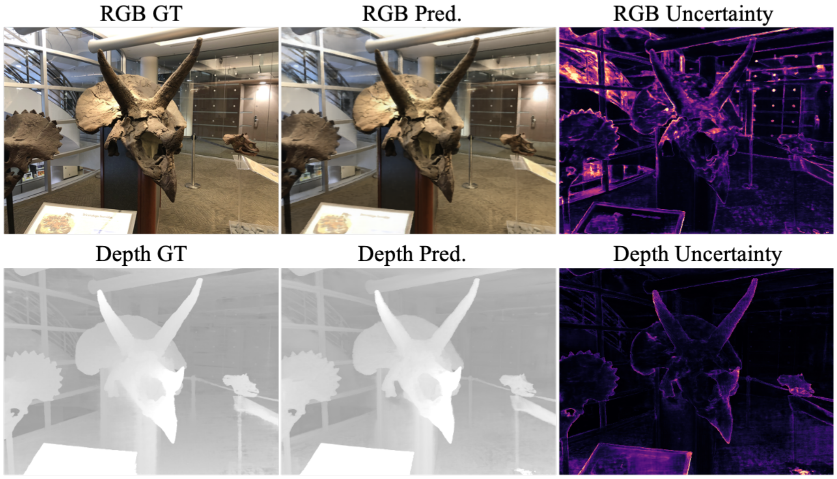


Fig. 1. Ground Truth (Left), Predictions (Center) and Uncertainty Maps (Right) obtained by our proposed CF-NeRF in novel view synthesis and depth-map estimation.

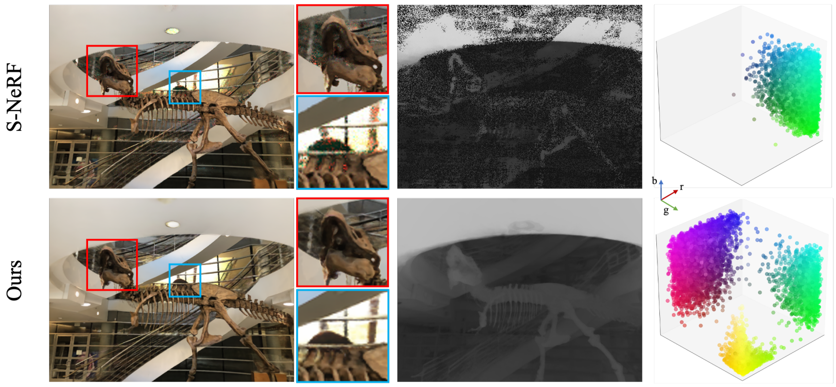
associated with the model estimates. This information is crucial in several scenarios such as robotics [28,22], medical diagnosis [52] or autonomous driving [8] where, to reduce potentially catastrophic failures, the confidence on the model outputs must be included into the decision-making process.

To address this limitation, recent works [31,53] have explored different strategies to incorporate uncertainty quantification into NeRF. Remarkably, Stochastic NeRF [53] (S-NeRF) has recently obtained state-of-the-art results by employing a probabilistic model to learn a simple distribution over radiance fields. However, in order to make this problem tractable, S-NeRF makes strong assumptions over the aforementioned distribution, thus limiting model expressivity and leading to sub-optimal results for complex 3D scenes.

In this context, we propose Conditional-Flow NeRF (CF-NeRF), a novel probabilistic framework to incorporate uncertainty quantification without sacrificing model flexibility. As a result, our approach enables both effective 3D modelling and reliable uncertainty estimates (see Fig. 1). In the following, we summarise the main technical contributions of our method:

Modelling Radiance-Density Distributions with Normalizing Flows: In contrast to S-NeRF where the radiance and density in the scene are assumed to follow simple distributions (see Fig. 2(c-Top)), our method learns them in a flexible and fully data-driven manner using conditional normalizing flows [64]. This allows CF-NeRF to learn arbitrarily complex radiance and density distributions (see Fig. 2(c-Bottom)) which have the capacity to model scenes with complex geometry and appearance.

Latent Variable Modelling for Radiance Fields Distributions: Motivated by De Finetti's Theorem [17], CF-NeRF incorporates a global latent variable in order to efficiently model the joint distribution over the radiance-density vari-



(a) RGB image (b) depth (c) distribution assumption

Fig. 2. Comparison between our CF-NeRF and the previous state-of-the-art method S-NeRF [53] on (a) RGB images and (b) depth-maps. S-NeRF generates noisy results due to the strong assumptions made over the radiance field distribution. In contrast, CF-NeRF renders more realistic and smooth synthetic views. (c) Illustration of the radiance distributions that can be modelled by S-NeRF and CF-NeRF. While the first is only able to represent distributions with a simple form, CF-NeRF can model arbitrary complex distributions by using Conditional Normalizing Flows.

ables for all the spatial locations in the scene. This contrasts with the approach followed by S-NeRF, where these variables are considered independent for different 3D coordinates. Intuitively, the independence assumption violates the fact that changes in the density and radiance between adjacent locations are usually smooth. As a consequence, S-NeRF tends to generate low-quality and noisy predictions (see Fig. 2(a,b)-Top). In contrast, the global latent variable used in CF-NeRF allows to efficiently model the complex joint distribution of these variables, leading to spatially-smooth uncertainty estimates and more realistic results (Fig. 2(a,b)-Bottom).

In our experiments, we evaluate CF-NeRF over different benchmark datasets containing scenes with increasing complexity. Our qualitative and quantitative results show that our method obtains significantly better uncertainty estimations for rendered views and predicted depth maps compared to S-NeRF and other previous methods for uncertainty quantification. Additionally, CF-NeRF provides better image quality and depth estimation precision. These results confirm that our method is able to incorporate uncertainty quantification into NeRF without sacrificing model expressivity.

2 Related Work

Neural Radiance Fields. NeRF [33] has become a popular approach for 3D scene modelling due to its simplicity and its impressive performance on several tasks. By using a sparse collection of 2D views, NeRF learns a deep neural network encoding a representation of the 3D scene. This network is composed of a set of fully-connected layers that output the volume density and emitted radiance for any input 3D spatial location and 2D viewing direction in the scene. Subsequently, novel views are generated by applying volumetric rendering [16] to the estimated density and radiance values obtained by the network. Recently, several extensions [68,42,36,19,70,54,58,51,66] over NeRF have been proposed. For instance, different works have focused on accelerating NeRF training [25,57,48] or modelling dynamic scenes [44,10,47,23,40]. Despite the advances achieved at different levels, the application of current NeRF-based methods in real scenarios is still limited since they are unable to quantify the uncertainty associated with the rendered views or estimated geometry. Our proposed CF-NeRF explicitly addresses this problem and can be easily combined with most of current NeRF-based approaches.

Uncertainty Estimation in Deep Learning. Uncertainty estimation has been extensively studied in deep learning [13,7,30,59,12] and have been applied to different computer vision tasks [46,3,63]. Early works proposed to use *Bayesian Neural Networks* (BNN) [34,35] to estimate the uncertainty of both network weights and outputs by approximating their marginal distributions. However, training BNNs is typically difficult and computationally expensive. As a consequence, their use in large deep neural networks is limited. More recently, efficient strategies have been proposed to incorporate uncertainty estimation into deep neural networks [49,6]. Among them, MC-Dropout [11] and Deep Ensembles [20] are two of the most popular approaches given that they are agnostic to the specific network architecture [1,14,2,24]. More concretely, MC-Dropout adds stochastic dropout during inference into the intermediate network layers. By doing multiple forward passes over the same input with different dropout configurations, the variance over the obtained set of outputs is used as the output uncertainty. On the other hand, Deep Ensembles applies a similar strategy by computing a set of outputs from multiple neural networks that are independently trained using different parameters initializations. In contrast to MC-Dropout and Deep Ensembles, however, we do not train independent networks to generate the samples. Instead, CF-NeRF explicitly encodes the distribution of the radiance fields into a single probabilistic model.

Uncertainty Estimation in NeRF. Recently, some works [31,53] have attempted to incorporate uncertainty estimation into Neural Radiance Fields. NeRF-in-the-Wild (NeRF-W) [31] modelled uncertainty at pixel-level in order to detect potential transient objects such as cars or pedestrians in modelled scenes. For this purpose, NeRF-W employs a neural network to estimate an uncertainty value for each 3D spatial location. Subsequently, it computes a confidence score for each pixel in a rendered image by using an alpha-compositing strategy similar to the one used to estimate the pixel color in original NeRF. However, this approach is not theoretically grounded since there is no intuitive justification to apply the volume rendering process over uncertainty values. Additionally, NeRF-W is not able to evaluate the confidence associated with estimated depth. On the other hand, S-NeRF [53] has achieved state-of-the-art performance in quantifying the uncertainty of novel rendered views and depth-maps. Instead of learning a single radiance field as in original NeRF [33], S-NeRF models a distribution over all the possible radiance fields explaining the scene. During inference, this distribution is used to sample multiple color or depth predictions and compute a confidence score based on their associated variance. However, S-NeRF imposes strong constraints over the radiance field distribution and, as a consequence, this limits its ability to model scenes with complex appearance and geometry. In contrast, the proposed CF-NeRF combines a conditional normalizing flow and latent variable modelling to learn arbitrary complex radiance fields distributions without any prior assumption.

Complex Distribution Modelling with Normalizing Flows. Normalizing Flows (NF) have been extensively used to model complex distributions with unknown explicit forms. For this purpose, NF uses a sequence of invertible functions to transform samples from a simple known distribution to variables following an arbitrary probability density function. Additionally, the change-of-variables formula can be used to compute the likelihood of different samples according to the transformed distribution. Given its flexibility, Normalizing Flows have been used in several 3D modelling tasks. For instance, [64,45] introduced different flow-based generative models to model 3D point cloud distributions and sample from them. More recently, [37] used a normalizing flow in order to avoid color shifts at unseen viewpoints in the context of NeRF. To the best of our knowledge, our proposed CF-NeRF is the first approach employing Normalizing Flows in order to learn radiance fields distributions for 3D scene modelling.

3 Deterministic and Stochastic Neural Radiance Fields

Deterministic Neural Radiance Fields. Standard NeRF [33] represents a 3D volumetric scene as $\mathcal{F} = \{(\mathbf{r}(\mathbf{x}, \mathbf{d}), \alpha(\mathbf{x})) : \mathbf{x} \in \mathbb{R}^3, \mathbf{d} \in \mathbb{R}^2$, where $\mathcal{F}$ is a set containing the volume density $\alpha(\mathbf{x}) \in \mathbb{R}^+$ and RGB radiance $\mathbf{r}(\mathbf{x}, \mathbf{d}) \in \mathbb{R}^3$ from all the spatial locations $\mathbf{x}$ and view directions $\mathbf{d}$ in the scene.

In order to implicitly model the infinite set $\mathcal{F}$, NeRF employs a neural network $f_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{d})$ with parameters $\boldsymbol{\theta}$ which outputs the density α and radiance $\mathbf{r}$ for any given input location-view pair $\{\mathbf{x}, \mathbf{d}\}$. Using this network, NeRF is able to estimate the color $\mathbf{c}(\mathbf{x}_o, \mathbf{d})$ for any given pixel defined by a 3D camera position $\mathbf{x}_o$ and view direction $\mathbf{d}$ using the volumetric rendering function:

$$\mathbf{c}(\mathbf{x}_o, \mathbf{d}) = \int_{t_n}^{t_f} T(t)\alpha(\mathbf{x}_t)\mathbf{r}(\mathbf{x}_t, \mathbf{d})dt\,, \text{ where } T(t) = \exp(-\int_{t_n}^{t} \alpha(\mathbf{x}_s)ds)\,, \quad (1)$$

and where $\mathbf{x}_t = \mathbf{x}_o + t\mathbf{d}$ corresponds to 3D locations along a ray with direction $\mathbf{d}$ originated at the camera origin and intersecting with the pixel at $\mathbf{x}_o$.

During training, NeRF optimizes the network parameters $\boldsymbol{\theta}$ using Eq. (1) by leveraging Maximum Likelihood Estimation (MLE) over a training set $\mathcal{T}$ containing ground-truth images depicting views of the scene captured from different camera positions. More details about NeRF and its learning procedure can be found in the original paper [33].

Stochastic Neural Radiance Fields. Deterministic NeRF is not able to provide information about the underlying uncertainty associated with the rendered views or estimated depth maps. The reason is that the network $f_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{d})$ is trained using Maximum Likelihood Estimation and thus, the learning process performs a single point estimate over all the plausible radiance fields $\mathcal{F}$ given the training set $\mathcal{T}$. As a consequence, model predictions are deterministic and it is not possible to sample multiple outputs to compute their associated variance.

To address this limitation, S-NeRF [53] employs Bayesian Learning to model the posterior distribution $p_{\boldsymbol{\theta}}(\mathcal{F}|\mathcal{T})$ over all the plausible radiance fields explaining the scene given the training set. In this manner, uncertainty estimations can be obtained during inference by computing the variance over multiple predictions obtained from different radiance fields $\mathcal{F}$ sampled from this distribution.

In order to make this problem tractable, S-NeRF uses Variational Inference to approximate the posterior $p_{\boldsymbol{\theta}}(\mathcal{F}|\mathcal{T})$ with a parametric distribution $q_{\boldsymbol{\theta}}(\mathcal{F})$ implemented by a deep neural network. However, S-NeRF imposes two limiting constraints over $q_{\boldsymbol{\theta}}(\mathcal{F})$. Firstly, it models the radiance $\mathbf{r}$ and density α for each location-view pair in the scene as independent variables. In particular, the probability of a radiance field given $\mathcal{T}$ is modelled with a fully-factorized distribution:

$$q_{\boldsymbol{\theta}}(\mathcal{F}) = \prod_{\mathbf{x}\in\mathbb{R}^3} \prod_{\mathbf{d}\in\mathbb{R}^2} q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}) q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}) \,. \tag{2}$$

Whereas this conditional-independence assumption simplifies the optimization process, it ignores the fact that radiance and density values in adjacent spatial locations are usually correlated. As a consequence, S-NeRF tends to render noisy and low-quality images and depth maps (see Fig. 2).

4 Conditional Flow-based Neural Radiance Fields

Our proposed Conditional Flow-based Neural Radiance Fields incorporates uncertainty estimation by using a similar strategy than S-NeRF. In particular, CF-NeRF learns a parametric distribution $q_{\boldsymbol{\theta}}(\mathcal{F})$ approximating the posterior distribution $p(\mathcal{F}|\mathcal{T})$ over all the possible radiance fields given the training views. Different from S-NeRF, however, our method does not assume a simple and fully-factorized distribution for $q_{\boldsymbol{\theta}}(\mathcal{F})$, but it relies on Conditional Normalizing Flows and Latent Variable Modelling to preserve model expressivity. In the following, we discuss the technical details of our approach (Sec. 4.1), the optimizing process following S-NeRF (Sec. 4.2) and inference procedure used to compute the uncertainty associated with novel rendered views and depth maps (Sec. 4.3).

4.1 Modelling Flexible Radiance Field Distributions with CF-NeRF

Radiance Field distribution with Global Latent Variable. As discussed in Sec. 3, S-NeRF formulation assumes that radiance and density variables at

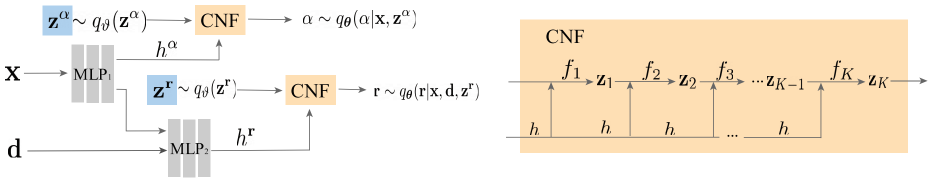

$\mathbf{z}^{\alpha} \sim q_{\vartheta}(\mathbf{z}^{\alpha})$
CNF
$\alpha \sim q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}, \mathbf{z}^{\alpha})$
h^{α}
$\mathbf{x}$
MLP₁
$\mathbf{z}^{\mathbf{r}} \sim q_{\vartheta}(\mathbf{z}^{\mathbf{r}})$
CNF
$\mathbf{r} \sim q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, \mathbf{z}^{\mathbf{r}})$
$\mathbf{d}$
MLP₂
$h^{\mathbf{r}}$
CNF
f_1 $\mathbf{z}_1$ f_2 $\mathbf{z}_2$ f_3 $\cdots \mathbf{z}_{K-1}$ f_K $\mathbf{z}_K$
h h h ... h

Fig. 3. (Left) We use two Conditional Normalizing Flow (CNF) to sample radiance and density values from distributions $q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, \mathbf{z})$ and $q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}, \mathbf{z})$, respectively. Each CNF computes a transformation of a sample from the latent distribution $q_{\psi}(\mathbf{z})$ conditioned to an embedding $\mathbf{h}$. This embedding which is computed by an MLP with the location-view pair $(\mathbf{x}, \mathbf{d})$ as input. (Right) Each CNF is composed by a sequence of invertible transformation functions $f_{1:K}$ conditioned on $\mathbf{h}$.

each spatial location in the field follow a fully-factorized distribution (Eq. (2)). As a consequence, they are forced to be independent from each other. This simplification was motivated by the fact that modelling the joint distribution over the infinite set of radiance and density variables in the field $\mathcal{F}$ is not trivial. To address this problem, our proposed CF-NeRF leverages De Finetti's representation theorem [17]. In particular, the theorem states that any joint distribution over sets of exchangeable variables can be written as a fully-factorized distribution conditioned on a global latent variable. Based on this observation, we define the approximate posterior $q_{\boldsymbol{\theta}}(\mathcal{F})$:

$$q_{\boldsymbol{\theta}}(\mathcal{F}) = \int_{\mathbf{z}} q_{\vartheta}(\mathbf{z}) q_{\boldsymbol{\theta}}(\mathcal{F}|\mathbf{z}) d\mathbf{z} = \int_{\mathbf{z}} q_{\vartheta}(\mathbf{z}) \prod_{\mathbf{x} \in \mathbb{R}^3} \prod_{\mathbf{d} \in \mathbb{R}^2} q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, \mathbf{z}) q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}, \mathbf{z}) d\mathbf{z}\,, \quad (3)$$

where the latent variable $\mathbf{z}$ is sampled from a Gaussian prior $q_{\vartheta}(\mathbf{z}) = \{\mu_z, \sigma_z\}$ with learned mean and variance. In this manner, all the 3D spatial-locations $\mathbf{x}$ and viewing-directions $\mathbf{d}$ generated from the same shared latent variable $\mathbf{z}$ and thus, they are conditionally dependent between them. This approach allows CF-NeRF to efficiently model the distribution over the radiance-density variables in the radiance field without any independence assumption. It is also worth mentioning that De Fineti's theorem has been also used in previous works learning joint distributions over sets for 3D point cloud modelling [18].

Radiance-Density Conditional Flows. Given that the radiance-density distributions can be highly complicated for complex scenes, modelling them with variants of Gaussian distributions as in Stochastic NeRF can lead to sub-optimal results. For this reason, CF-NeRF models $q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, \mathbf{z})$ and $q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}, \mathbf{z})$ using a Conditional Normalizing Flow (CNF) [4]. This allows to learn arbitrarily complex distributions in a fully data-driven manner without any prior assumption.

For any input location-direction pair $(\mathbf{x}, \mathbf{d})$, we use a sequence of K CNFs that transform the global latent variable $\mathbf{z}$ into samples from the radiance $\mathbf{r}$ or density α distribution $q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, \mathbf{z})$ or $q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}, \mathbf{z})$. More formally, each flow is defined as an invertible parametric function $\mathbf{z}_k = f_k(\mathbf{z}_{k-1}, \mathbf{x}, \mathbf{d})$, where f_k maps

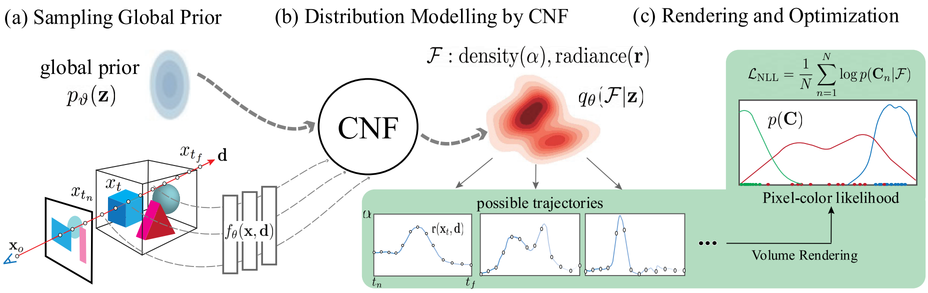

(a) Sampling Global Prior
(b) Distribution Modelling by CNF
(c) Rendering and Optimization
global prior
$p_\vartheta(\mathbf{z})$
CNF
$\mathcal{F}$: density(α), radiance($\mathbf{r}$)
$q_\theta(\mathcal{F}|\mathbf{z})$
$\mathcal{L}_{\text{NLL}} = \frac{1}{N}\sum_{n=1}^{N} \log p(\mathbf{C}_n|\mathcal{F})$
$p(\mathbf{C})$
Pixel-color likelihood
x_{t_f}
$\mathbf{d}$
x_{t_n}
x_t
$\mathbf{x}_o$
$f_\theta(\mathbf{x},\mathbf{d})$
possible trajectories
α
$\mathrm{r}(\mathrm{x}_t,\mathrm{d})$
t_n
t_f
...
Volume Rendering

Fig. 4. Illustration of our pipeline for the inference and computation of the log-likelihood on the pixel color. (a) We sample a set of variables $\mathbf{z}$ from the global latent distribution. (b) Given each spatial location $\mathbf{x}$ along a camera ray with viewing direction $\mathbf{d}$, we can generate a set of density and radiance values by passing each $\mathbf{z}$ variable through our proposed CNF. (c) These values can be represented as a set of different density-radiance trajectories along the ray corresponding to each $\mathbf{z}$, followed by volume rendering techniques to composite each trajectory into a RGB value. Finally, these RGB values are used to compute the log-likelihood for the pixel color and also estimate the model prediction and its associated uncertainty using their mean and variance during inference.

a random variable $\mathbf{z}_{k-1}$ into another one $\mathbf{z}_k$ with a more complex distribution. Note that the each flow f_k is conditioned on the location and view direction $(\mathbf{x}, \mathbf{d})$. Finally, $\mathbf{z}_K$ is followed by a Sigmoid and Softplus activation functions for radiance and density samples, respectively. This process is detailed in Fig. 3.

Using the introduced CNF, radiance and density probabilities $q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, \mathbf{z})$ and $q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}, \mathbf{z})$ can be computed with the change-of-variables formula typically used in normalizing flows as:

$$q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, \mathbf{z}) = q_\vartheta(\mathbf{z}) \left|\det \frac{\partial \mathbf{z}}{\partial \mathbf{r}}\right| = q_\vartheta(\mathbf{z}) \left|\det \frac{\partial \mathbf{r}}{\partial \mathbf{z}_K}\right|^{-1} \prod_{k=1}^{K} \left|\det \frac{\partial \mathbf{z}_k}{\partial \mathbf{z}_{k-1}}\right|^{-1}, \tag{4}$$

where $|\det(\partial \mathbf{z}_k / \partial \mathbf{z}_{k-1})|$ measures the Jacobian determinant of the transformation function f_k. Note that $q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}, \mathbf{z})$ can be computed in a similar manner.

4.2 Optimizing Conditional-Flow NeRF

We adopt a Variational Bayesian approach to learn the parameters of the posterior distribution $q_{\boldsymbol{\theta}}(\mathcal{F})$ defined in Eq. (3). Note that the optimized θ corresponds to the parameters of the CNFs defined in the previous section. More formally, we solve an optimization problem where we minimize the Kullback-Leibler (KL) divergence between $q_{\boldsymbol{\theta}}(\mathcal{F})$ and the true posterior distribution $p(\mathcal{F}|\mathcal{T})$ of radiance fields given the training set:

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \mathbb{KL}(q_{\boldsymbol{\theta}}(\mathcal{F}) || p(\mathcal{F}|\mathcal{T})) \\ &= \min_{\boldsymbol{\theta}} -\mathbb{E}_{q_{\boldsymbol{\theta}}(\mathcal{F})} \log p(\mathcal{T}|\mathcal{F}) + \mathbb{E}_{q_{\boldsymbol{\theta}}(\mathcal{F})} \log q_{\boldsymbol{\theta}}(\mathcal{F}) - \mathbb{E}_{q_{\boldsymbol{\theta}}(\mathcal{F})} \log p(\mathcal{F}) \,. \end{aligned} \tag{5}$$

The first term in Eq. (5) measures the expected log-likelihood of the training set $\mathcal{T}$ over the distribution of radiance fields $q_{\boldsymbol{\theta}}(\mathcal{F})$. The second term indicates the negative Entropy of the approximated posterior. Intuitively, maximizing the Entropy term allows uncertainty estimation by preventing the optimized distribution to degenerate into a deterministic function where all the probability is assigned into a single radiance field $\mathcal{F}$. Finally, the third term corresponds to the cross-entropy between $q_{\boldsymbol{\theta}}(\mathcal{F})$ and a prior over radiance fields $p(\mathcal{F})$. Given that it is hard to choose a proper prior in this case, we assume $p(\mathcal{F})$ to follow uniform distribution and thus, this term can be ignored during optimization. In the following, we detail how the first two terms are computed during training.

Computing the log-likelihood. Assuming that the training set $\mathcal{T}$ is composed by N triplets $\{\mathbf{C}_n, \mathbf{x}^o_n, \mathbf{d}_n\}$ representing the color, camera origin and view direction of a given pixel in a training image, the log-likelihood term in Eq. (5) is equivalent to: $\frac{1}{N}\sum_{n=1}^{N} \log p(\mathbf{C}_n|\mathcal{F})$. In order to compute an individual $p(\mathbf{C}_n|\mathcal{F})$ we use the following procedure.

Firstly, we sample a set of variables $\mathbf{z}^{1:K}$ from the global latent distribution $q_{\vartheta}(\mathbf{z})$. Secondly, given a training triplet $\{\mathbf{C}, \mathbf{x}^o, \mathbf{d}\}$ and a set of 3D spatial locations along a ray defined by $\mathbf{x}_t = \mathbf{x}_o + t\mathbf{d}$, we sample a set of density and radiance values $\{\alpha^{1:K}\}_{t_n:t_f}$ and $\{\mathbf{r}^{1:K}\}_{t_n:t_f}$ by using the Conditional Flow introduced in Sec. 4.1. In particular, we perform a forward pass over the flow with inputs $\mathbf{z}^k$ conditioned to $\mathbf{x}^*_t$ for all t. Subsequently, a set of K color estimates $\hat{\mathbf{C}}_{1:K}$ are obtained by using, for every k, the volume rendering formula in Eq. (1) over the sampled radiance and density values $\{\alpha^k\}_{t_n:t_f}$ and $\{\mathbf{r}^k\}_{t_n:t_f}$, respectively. Finally, with the set of obtained color estimations $\hat{\mathbf{C}}_{1:K}$, we use a non-parametric kernel density estimator as in [53] to approximate the log-likelihood $\log p(\mathbf{C}|\mathcal{F})$ of the pixel color. An illustration of this procedure is shown in Fig. 4.

Computing the Entropy term. In order to compute the entropy term in Eq. (5), we use a Monte Carlo approximation with M samples as:

$$\begin{aligned}\mathbb{E}_{q_{\boldsymbol{\theta}}(\mathcal{F})} \log q_{\boldsymbol{\theta}}(\mathcal{F}) &= \mathbb{E}_{q_{\boldsymbol{\vartheta}}(\mathbf{z})} \prod_{\mathbf{x}\in\mathbb{R}^3} \prod_{\mathbf{d}\in\mathbb{R}^2} \log q_{\boldsymbol{\theta}}(\mathbf{r}|\mathbf{x}, \mathbf{d}, \mathbf{z}) + \mathbb{E}_{q_{\boldsymbol{\vartheta}}(\mathbf{z}^{\alpha})} \prod_{\mathbf{x}\in\mathbb{R}^3} \prod_{\mathbf{d}\in\mathbb{R}^2} \log q_{\boldsymbol{\theta}}(\alpha|\mathbf{x}, \mathbf{z}) \\ &\sim \frac{1}{M} \sum_{m=1}^{M} \Big(\log q_{\boldsymbol{\theta}}(\mathbf{r}_m|\mathbf{x}_m, \mathbf{d}_m, \mathbf{z}_m) + \log q_{\boldsymbol{\theta}}(\alpha_m|\mathbf{x}_m, \mathbf{z}_m) \Big), \qquad (6)\end{aligned}$$

where $\mathbf{z}_m$ are obtained from the latent variable distribution $q_{\boldsymbol{\vartheta}}(\mathbf{z})$ and $\{\mathbf{x}_m, \mathbf{d}_m\}$ are possible 3D locations and view directions in the scene which are randomly sampled. Finally, α_m and $\mathbf{r}_m$ are density and radiance values obtained by applying our conditional flow with inputs $(\mathbf{z}_m,\mathbf{x}_m,\mathbf{d}_m)$. Their probabilities $q_{\boldsymbol{\theta}}(\alpha_m|\mathbf{x}_m, \mathbf{z}_m)$ and $q_{\boldsymbol{\theta}}(\mathbf{r}_m|\mathbf{x}_m, \mathbf{d}_m, \mathbf{z}_m)$ can be computed by using Eq. (4).

4.3 Inference and Uncertainty Estimation

CF-NeRF allows us to quantify the uncertainty associated with rendered images and depth-maps. As described in Sec. 4.2, given any camera pose, we can obtain a set of color estimates $\hat{\mathbf{C}}_{1:K}$ for every pixel in a rendered image. These estimates are obtained by sampling a set of random variables $\mathbf{z}_{1:K}$ from latent distribution $q_{\vartheta}(\mathbf{z})$ and applying CF-NeRF. Finally, the mean and variance of the pixel colors over the K samples are treated as the predicted color and its associated uncertainty, respectively. For depth-map generation, we sample a sequence of density values $\{\alpha_k\}_{t_n:t_f}$ along a camera ray and we compute the expected termination depth of the ray as in [33]. In this way, we obtain a set of depth-values $d_{1:K}$ for each pixel in the depth-map for the rendered image. As in the case of the RGB color, its mean and variance correspond to the estimated depth and its associated uncertainty.

5 Experiments

Datasets. We conduct a set of exhaustive experiments over two benchmark databases typically used to evaluate NeRF-based approaches: the **LLFF** dataset from the original NeRF paper [33] and the Light Field **LF** dataset [68,67]. The first is composed of 8 relatively simple scenes with multiple-view forward-facing images. On the other hand, from the LF dataset we use 4 scenes: *Africa, Basket, Statue* and *Torch*. The evaluation over the LF scenes is motivated by the fact that they have a longer depth range compared to the ones in LLFF and, typically, they present more complicated geometries and appearance. As a consequence, the evaluation over LLFF gives more insights into the ability of NeRF-based methods to model complex 3D scenes. Same as [53], we use a sparse number of scene views (4∼5) for training and the last adjacent views for testing. As discussed in the cited paper, training in an extremely low-data regime is a challenging task and provides an ideal setup to evaluate the ability of the compared methods to quantify the uncertainty associated with the model predictions. While the ground truth of depth is not available in the original datasets, we compute pseudo ground truth by using the method in [61] trained on all the available views per scene.

Baselines. We compare our method with the state-of-the-art S-NeRF [53], NeRF-W [31] and two other methods, Deep-Ensembles(D.E.) [20] and MC-Dropout(Drop.) [11], which are generic approaches for uncertainty quantification in deep learning. For NeRF-W, we remove the latent embedding component of their approach and keep only the uncertainty estimation layers. For Deep-Ensembles, we train 3 different NeRF models in the ensemble in order to have a similar computational cost during training and testing compared to the rest of compared methods. In the case of MC-Dropout, we manually add one dropout layer after each odd layer in the network to sample multiple outputs using random dropout configurations. The described setup and hyper-parameters for the different baselines is the same previously employed in S-NeRF [53].

Implementations details. In order to implement CF-NeRF and the different baselines, we inherit the same network architecture and hyper-parameters than the one employed in the original NeRF[3] implementation. In CF-NeRF,

[3] https://github.com/bmild/nerf

Table 1. Quality and uncertainty quantification metrics on rendered images over the LLFF dataset. Best results are shown in bold with second best underlined.

	quality metrics			uncertainty metrics		
	PSNR↑	SSIM↑	LPIPS↓	AUSE RMSE↓	AUSE MAE↓	NLL↓
D.E. [20]	**22.32**	0.788	0.236	0.0254	0.0122	1.98
Drop. [11]	21.90	0.758	0.248	0.0316	0.0162	1.69
NeRF-W [31]	20.19	0.706	0.291	0.0268	0.0113	2.31
S-NeRF [53]	20.27	0.738	0.229	0.0248	0.0101	0.95
CF-NeRF	21.96	**0.790**	**0.201**	**0.0177**	**0.0078**	**0.57**

we use this architecture to compute the conditional part given $\mathbf{x}$ and $\mathbf{d}$. Since CF-NeRF uses an additional normalizing flow following this network, we add additional layers for the rest of baselines so that they have a similar computational complexity than our method. See the supplementary material for more details on hyper-parameters of our conditional normalizing flow.

Metrics. In our experiments, we address two different tasks using the compared methods: novel view synthesis and depth-map estimation. In particular, we evaluate the quality of generated images/depth-maps and the reliability of the quantified uncertainty using the following metrics:

Quality metrics: We use standard metrics in the literature. Specifically, we report PSNR, SSIM [60], and LPIPS [69] for generated synthetic views. On the other hand, we compute RMSE, MAE and δ-threshold [9] to evaluate the quality of the generated depth-maps.

Uncertainty Quantification: To evaluate the reliability of the uncertainty estimated by each model we report two widely used metrics for this purpose. Firstly, we use the negative log likelihood (NLL) which is a proper scoring rule [21,29] and has been previously used to evaluate the quality of model uncertainty [53]. Secondly, we evaluate the compared methods using sparsification curves [46,3,43]. Concretely, given an error metric (e.g. RMSE), we obtain two lists by sorting the values of all the pixels according to their uncertainty and the computed error. By removing the top $t\%(t = 1 \sim 100)$ of the errors in each vector and repeatedly computing the average of the last subset, we can obtain the sparsification curve and the oracle curve respectively. The area between is the AUSE, which evaluates how much the uncertainty is correlated with the error.

5.1 Results over LLFF dataset

Following a similar experimental setup than [53], we firstly evaluate the performance of the compared methods on the simple scenes in the LLFF dataset. All the views in each scene are captured in a restricted setting where all cameras are forward-facing towards a plane, which allows to use Normalized Device Coordinates (NDC) to bound all coordinates to a very small range(0-1). In Table 1, we report the performance of the different evaluated metrics averaged over all the scenes. As can be seen, S-NeRF achieves the second best performance in terms of uncertainty quantification with quality metrics comparable to the rest of baselines. These results are consistent with the results reported in their original

Table 2. Quality and uncertainty quantification metrics on rendered images over LF dataset. Best results are shown in bold with second best underlined.

Scene Type	Methods	Quality Metrics			Uncertainty Metrics			Scene Type	Quality Metrics			Uncertainty Metrics		
		PSNR↑	SSIM↑	LPIPS↓	AUSE RMSE↓	AUSE MAE↓	NLL↓		PSNR↑	SSIM↑	LPIPS↓	AUSE RMSE↓	AUSE MAE↓	NLL↓
Basket	D.E. [20]	25.93	0.87	0.17	0.206	0.161	1.26	Afria	23.14	0.78	0.32	0.385	0.304	2.19
	Drop. [11]	24.00	0.80	0.30	0.324	0.286	1.35		22.13	0.72	0.42	0.396	0.339	1.69
	NeRF-W [31]	19.81	0.73	0.31	0.333	0.161	2.23		21.14	0.71	0.42	0.121	0.089	1.41
	S-NeRF [53]	23.56	0.80	0.23	0.098	0.089	0.26		20.83	0.73	0.35	0.209	0.161	0.56
	CF-NeRF	**26.39**	**0.89**	**0.11**	**0.039**	**0.018**	**-0.90**		**23.84**	**0.83**	**0.23**	**0.077**	**0.054**	**-0.25**
Statue	D.E. [20]	**24.57**	0.85	0.21	0.307	0.214	1.53	Torch	21.49	0.73	0.43	0.153	0.101	1.80
	Drop. [11]	23.91	0.82	0.28	0.297	0.232	1.09		19.23	0.62	0.59	0.226	0.154	3.09
	NeRF-W [31]	19.89	0.73	0.41	0.099	0.071	3.03		15.59	0.56	0.69	0.132	0.131	1.52
	S-NeRF [53]	13.24	0.55	0.59	0.475	0.714	4.56		13.12	0.33	1.02	0.321	0.454	2.29
	CF-NeRF	24.54	**0.87**	**0.16**	**0.040**	**0.019**	**-0.83**		**23.95**	**0.86**	**0.17**	**0.047**	**0.015**	**-0.86**

Table 3. Quality and uncertainty quantification metrics on depth estimation over the LF dataset. All the results are computed from disparity values which are reciprocal to depth. Best results are shown in bold with second best underlined.

Scene Type	Methods	Quality Metrics			Uncertainty Metrics			Scene Type	Quality Metrics			Uncertainty Metrics		
		RMSE↓	MAE↓	δ_3 ↑	AUSE RMSE↓	AUSE MAE↓	NLL↓		RMSE↓	MAE↓	δ_3 ↑	AUSE RMSE↓	AUSE MAE↓	NLL↓
Basket	D.E. [20]	0.221	0.132	0.66	0.480	0.232	7.88	Africa	0.085	**0.039**	0.90	0.218	0.110	3.61
	Drop. [11]	0.241	0.153	0.43	0.297	0.157	10.96		0.216	0.141	0.31	0.544	0.503	10.36
	NeRF-W [31]	0.214	0.179	0.41	-	-	-		0.127	0.079	0.77	-	-	-
	S-NeRF [53]	0.417	0.401	0.23	0.305	0.312	8.75		0.239	0.155	0.36	0.234	0.252	6.55
	CF-NeRF	**0.166**	**0.099**	**0.80**	**0.101**	**0.052**	**6.76**		**0.074**	**0.039**	**0.93**	**0.105**	**0.090**	**2.05**
Statue	D.E. [20]	0.162	0.118	**0.73**	0.115	0.056	7.78	Torch	0.132	0.071	0.71	0.226	0.131	5.76
	Drop. [11]	0.197	0.128	0.59	0.164	0.109	10.89		0.263	0.173	0.06	0.982	0.809	10.84
	NeRF-W [31]	0.276	0.218	0.55	-	-	-		0.271	0.241	0.17	-	-	-
	S-NeRF [53]	0.751	0.709	0.26	0.353	0.386	11.51		0.274	0.236	0.14	0.770	1.013	8.75
	CF-NeRF	**0.122**	**0.098**	**0.73**	**0.069**	**0.053**	**7.38**		**0.110**	**0.061**	**0.78**	**0.164**	**0.089**	**4.15**

paper and demonstrate that the expressivity of S-NeRF on these simple scenes is not severely limited by the imposed strong assumptions over the radiance field distribution. Nonetheless, our CF-NeRF still achieves a significant improvement over most of the metrics, showing the advantages of modelling the radiance field distribution using the proposed latent variable modelling and CNFs.

5.2 Results over LF dataset

To further test the performance of different approaches on larger scenes with more complex appearances and geometries, we evaluate the different methods on the LF dataset. All images in these scenes have a large depth range and are randomly captured towards the target. As a consequence, coordinates cannot be normalized with NDC.

Quantitative results. Results over rendered RGB images and estimated depth-maps for all the scenes and evaluated metrics are shown in Table 2 and Table 3, respectively. As can be seen, S-NeRF obtains poor performance in terms of quality and uncertainty estimation. This clearly shows that the imposed assumptions over the radiance fields distribution are sub-optimal in complex scenes and lead to inaccurate model predictions and unreliable uncertainty estimations. In contrast, our CF-NeRF outperforms all the previous approaches by large margins

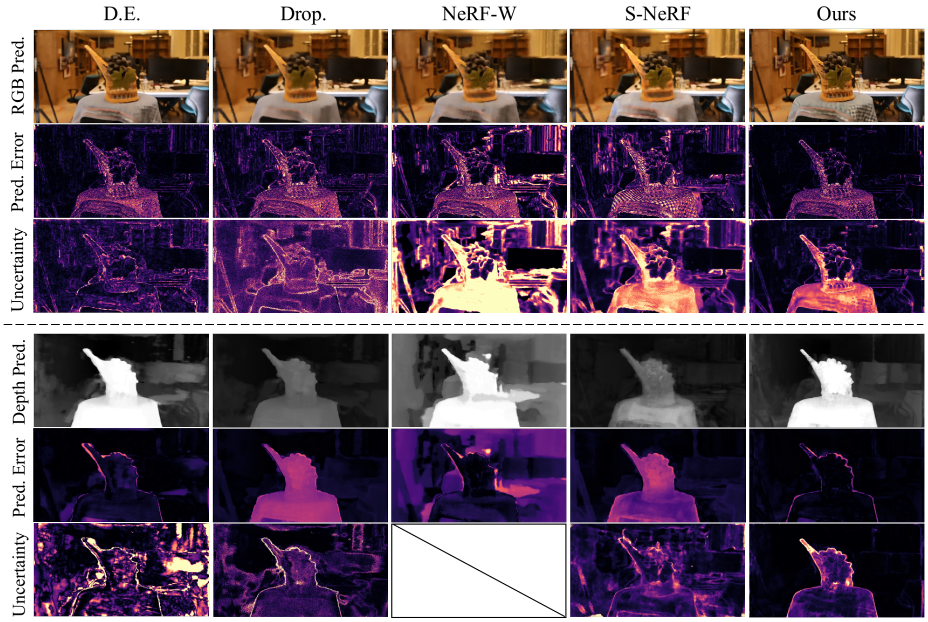


Fig. 5. Qualitative comparison between our CF-NeRF and other baselines over generated images(Top) and depth-maps(Bottom). The model prediction, its computed error with the Ground-Truth and its associated uncertainty estimation are shown respectively in each row. Larger values are shown in yellow while purple and black indicate smaller values of predicted error and uncertainty. NeRF-W is not able to generate multiple depth samples and hence cannot produce uncertainty for depth-maps. Among all these approaches, our CF-NeRF is the only approach that is able to generate both accurately rendered views and depth estimations and visually intuitive uncertainty maps which are highly correlated with the true prediction error. Refer to the supplementary materials for more visualized results.

across all the scenes and metrics both on image rendering and depth-maps estimation tasks. The better results obtained by our method can be explained because the use of latent variable modelling and conditional normalizing flows allows to learn arbitrarily complex radiance field distributions, which are necessary to model scenes with complicated appearance and geometry. This partly explains that our method achieves better performance, particularly an average 69% improvement compared to the second best method Deep-Ensembles on LPIPS, which measures the perceptual similarity at the image level. On the other hand, the performance of Deep-Ensembles and MC-Dropout is largely limited by the number of the trained models or dropout samples, which cannot be increased infinitely in practical cases. In contrast, the explicitly encoding of the radiance field distribution into a single probabilistic model with CF-NeRF allows to use a higher-number of samples in a more efficient manner.

Qualitative results. In order to give more insights into the advantages of our proposed CF-NeRF, Fig. 5 shows qualitative results obtained by the compared methods on an example testing view of the LF dataset. By looking at the estimated uncertainty for RGB images, we can observe that NeRF-W obtains a higher prediction error which, additionally, is not correlated with the uncertainty estimates. On the other hand, Deep Ensembles and MC-Dropout render images with a lower predictive error compared to NeRF-W. As can be observed, however, their estimated uncertainties are noticeably not correlated with their errors. Interestingly, S-NeRF seems to obtain uncertainty estimates where the values are more correlated with the predictive error. However, the quality of the rendered image is poor, obviously shown in the background. In contrast, our CF-NeRF is the only method able to render both high-quality images and uncertainty estimates that better correlate with the prediction error.

By looking at the results for generated depth-maps, we can see that our method also obtains the most accurate depth estimates, followed by Deep Ensembles. Nonetheless, the latter generates depth maps with obvious errors in the background. Regarding the uncertainty estimation, we can observe a high correlation between the predicted error and the estimated uncertainty maps generated by our CF-NeRF. This contrasts with the rest methods which estimate low confidence values in the background of the scene inconsistent with their low prediction errors on this specific area. In conclusion, our results demonstrate the ability of our proposed method to provide both reliable uncertainty estimates and accurate model predictions.

6 Conclusion

We have presented CF-NeRF, a novel probabilistic model to address the problem of uncertainty quantification in 3D modelling using Neural Radiance Fields. In contrast to previous works employing models with limited expressivity, our method couples Latent Variable Modelling and Conditional Normalizing Flows in order to learn complex radiance fields distributions in a flexible and fully data-driven manner. In our experimental results, we show that this strategy allows CF-NeRF to obtain significantly better results than state-of-the-art approaches, particularly on scenes with complicated appearance and geometry. Finally, it is also worth mentioning that our proposed framework can be easily combined with other NeRF-based architectures in order to provide them with uncertainty quantification capabilities.

Acknowledgements

This work is supported partly by the Chinese Scholarship Council (CSC) under grant (201906120031), by the Spanish government under project MoHuCo PID2020-120049RB-I00 and the Chistera project IPALM PCI2019-103386. We also thank Nvidia for hardware donation under the GPU grant program.

References

1. Aralikatti, R., Margam, D., Sharma, T., Thanda, A., Venkatesan, S.: Global snr estimation of speech signals using entropy and uncertainty estimates from dropout networks. In: INTERSPEECH (2018)
2. Ashukha, A., Lyzhov, A., Molchanov, D., Vetrov, D.P.: Pitfalls of in-domain uncertainty estimation and ensembling in deep learning. In: ICLR (2020)
3. Bae, G., Budvytis, I., Cipolla, R.: Estimating and exploiting the aleatoric uncertainty in surface normal estimation. In: ICCV (2021)
4. van den Berg, R., Hasenclever, L., Tomczak, J., Welling, M.: Sylvester normalizing flows for variational inference. In: proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI) (2018)
5. Bi, S., Xu, Z., Sunkavalli, K., Havsan, M., Hold-Geoffroy, Y., Kriegman, D.J., Ramamoorthi, R.: Deep reflectance volumes: Relightable reconstructions from multiview photometric images. In: ECCV (2020)
6. Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D.: Weight uncertainty in neural networks. In: ICML (2015)
7. Brachmann, E., Michel, F., Krull, A., Yang, M.Y., Gumhold, S., Rother, C.: Uncertainty-driven 6d pose estimation of objects and scenes from a single rgb image. In: CVPR (2016)
8. Djuric, N., Radosavljevic, V., Cui, H., Nguyen, T., Chou, F.C., Lin, T.H., Singh, N., Schneider, J.G.: Uncertainty-aware short-term motion prediction of traffic actors for autonomous driving. In: WACV (2020)
9. Eigen, D., Puhrsch, C., Fergus, R.: Depth map prediction from a single image using a multi-scale deep network. In: NIPS (2014)
10. Gafni, G., Thies, J., Zollhöfer, M., Nießner, M.: Dynamic neural radiance fields for monocular 4D facial avatar reconstruction. In: CVPR (2021)
11. Gal, Y., Ghahramani, Z.: Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: ICML (2016)
12. Geifman, Y., Uziel, G., El-Yaniv, R.: Bias-reduced uncertainty estimation for deep neural classifiers. In: ICLR (2019)
13. Hendrycks, D., Mazeika, M., Kadavath, S., Song, D.X.: Using self-supervised learning can improve model robustness and uncertainty. In: NeurIPS (2019)
14. Hernández, S., Vergara, D., Valdenegro-Toro, M., Jorquera, F.: Improving predictive uncertainty estimation using dropout–hamiltonian monte carlo. Soft Computing **24**, 4307–4322 (2020)
15. Jiakai, Z., Xinhang, L., Xinyi, Y., Fuqiang, Z., Yanshun, Z., Minye, W., Yingliang, Z., Lan, X., Jingyi, Y.: Editable free-viewpoint video using a layered neural representation. In: ACM SIGGRAPH (2021)
16. Kajiya, J., von herzen, B.: Ray tracing volume densities. ACM SIGGRAPH Computer Graphics **18**, 165–174 (1984)
17. Kirsch, W.: An elementary proof of de Finetti's theorem. Statistics & Probability Letters **151**(C), 84–88 (2019)
18. Klokov, R., Boyer, E., Verbeek, J.J.: Discrete point flow networks for efficient point cloud generation. In: ECCV (2020)
19. Kosiorek, A.R., Strathmann, H., Zoran, D., Moreno, P., Schneider, R., Mokr'a, S., Rezende, D.J.: Nerf-vae: A geometry aware 3d scene generative model. In: ICML (2021)
20. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In: NIPS (2017)

21. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In: Adv. Neural Inform. Process. Syst. (2017)
22. Li, Y., Li, S., Sitzmann, V., Agrawal, P., Torralba, A.: 3d neural scene representations for visuomotor control. In: CoRL (2021)
23. Li, Z., Niklaus, S., Snavely, N., Wang, O.: Neural scene flow fields for space-time view synthesis of dynamic scenes. In: CVPR (2021)
24. Liu, J., Paisley, J.W., Kioumourtzoglou, M.A., Coull, B.: Accurate uncertainty estimation and decomposition in ensemble learning. In: NeurIPS (2019)
25. Liu, L., Gu, J., Lin, K.Z., Chua, T.S., Theobalt, C.: Neural sparse voxel fields. In: NeurIPS (2020)
26. Liu, S., Zhang, X., Zhang, Z., Zhang, R., Zhu, J.Y., Russell, B.: Editing conditional radiance fields. In: ICCV (2021)
27. Lombardi, S., Simon, T., Saragih, J., Schwartz, G., Lehrmann, A., Sheikh, Y.: Neural volumes: Learning dynamic renderable volumes from images. ACM Trans. Graph. **38**(4), 65:1–65:14 (Jul 2019)
28. Long, K., Qian, C., Cortes, J., Atanasov, N.: Learning barrier functions with memory for robust safe navigation. IEEE Robotics and Automation Letters **6**(3), 1–1 (2021)
29. Loquercio, A., Segu, M., Scaramuzza, D.: A general framework for uncertainty estimation in deep learning. IEEE Robotics and Automation Letters **5**(2), 3153–3160 (Apr 2020)
30. Malinin, A., Gales, M.J.F.: Predictive uncertainty estimation via prior networks. In: NeurIPS (2018)
31. Martin-Brualla, R., Radwan, N., Sajjadi, M.S.M., Barron, J.T., Dosovitskiy, A., Duckworth, D.: NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In: CVPR (2021)
32. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy Networks: Learning 3d reconstruction in function space. In: CVPR (2019)
33. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)
34. Neal, R.M.: Bayesian Learning for Neural Networks. Springer New York, NY (1995)
35. Neapolitan, R.E.: Learning bayesian networks. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining(KDD) (2007)
36. Neff, T., Stadlbauer, P., Parger, M., Kurz, A., Mueller, J.H., Chaitanya, C.R.A., Kaplanyan, A., Steinberger, M.: Donerf: Towards real-time rendering of compact neural radiance fields using depth oracle networks. Computer Graphics Forum **40**(4), 45–59 (2021)
37. Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S.M., Geiger, A., Radwan, N.: Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. ArXiv **abs/2112.00724** (2021)
38. Niemeyer, M., Mescheder, L., Oechsle, M., Geiger, A.: Occupancy flow: 4d reconstruction by learning particle dynamics. In: ICCV (2019)
39. Park, J.J., Florence, P., Straub, J., Newcombe, R.A., Lovegrove, S.: DeepSDF: Learning continuous signed distance functions for shape representation. In: CVPR (2019)
40. Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable neural radiance fields. ICCV (2021)

41. Peng, L.W., Shamsuddin, S.M.: 3d object reconstruction and representation using neural networks. In: Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia (2004)
42. Peng, S., Dong, J., Wang, Q., Zhang, S.W., Shuai, Q., Bao, H., Zhou, X.: Animatable neural radiance fields for human body modeling. In: CVPR (2021)
43. Poggi, M., Aleotti, F., Tosi, F., Mattoccia, S.: On the uncertainty of self-supervised monocular depth estimation. In: CVPR (2020)
44. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-NeRF: Neural radiance fields for dynamic scenes. In: CVPR (2021)
45. Pumarola, A., Popov, S., Moreno-Noguer, F., Ferrari, V.: C-flow: Conditional generative flow models for images and 3d point clouds. In: CVPR (2020)
46. Qu, C., Liu, W., Taylor, C.J.: Bayesian deep basis fitting for depth completion with uncertainty. In: ICCV (2021)
47. Raj, A., Zollhoefer, M., Simon, T., Saragih, J., Saito, S., Hays, J., Lombardi, S.: PVA: Pixel-aligned volumetric avatars. In: CVPR (2021)
48. Rebain, D., Jiang, W., Yazdani, S., Li, K., Yi, K.M., Tagliasacchi, A.: Derf: Decomposed radiance fields. In: CVPR (2020)
49. Ritter, H., Botev, A., Barber, D.: A scalable laplace approximation for neural networks. In: ICLR (2018)
50. Saito, S., Yang, J., Ma, Q., Black, M.J.: SCANimate: Weakly supervised learning of skinned clothed avatar networks. In: CVPR (2021)
51. Schwarz, K., Liao, Y., Niemeyer, M., Geiger, A.: Graf: Generative radiance fields for 3d-aware image synthesis. In: NIPS (2020)
52. Shamsi, A., Asgharnezhad, H., Jokandan, S.S., Khosravi, A., Kebria, P.M., Nahavandi, D., Nahavandi, S., Srinivasan, D.: An uncertainty-aware transfer learning-based framework for covid-19 diagnosis. IEEE Transactions on Neural Networks and Learning Systems **32**, 1408–1417 (2021)
53. Shen, J., Ruiz, A., Agudo, A., Moreno-Noguer, F.: Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In: 3DV (2021)
54. Sitzmann, V., Martel, J.N.P., Bergman, A.W., Lindell, D.B., Wetzstein, G.: Implicit neural representations with periodic activation functions. In: NIPS (2020)
55. Sitzmann, V., Zollhöfer, M., Wetzstein, G.: Scene representation networks: Continuous 3d-structure-aware neural scene representations. In: NIPS (2019)
56. Su, S.Y., Yu, F., Zollhöfer, M., Rhodin, H.: A-nerf: Articulated neural radiance fields for learning human shape, appearance, and pose. In: NIPS (2021)
57. Tancik, M., Mildenhall, B., Wang, T., Schmidt, D., Srinivasan, P.P., Barron, J.T., Ng, R.: Learned initializations for optimizing coordinate-based neural representations. In: CVPR (2021)
58. Tancik, M., Srinivasan, P.P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J.T., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. In: NIPS (2020)
59. Teye, M., Azizpour, H., Smith, K.: Bayesian uncertainty estimation for batch normalized deep networks. In: ICML (2018)
60. Wang, Z., Simoncelli, E., Bovik, A.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003. vol. 2, pp. 1398–1402 Vol.2 (2003)
61. Wei, Y., Liu, S., Rao, Y., Zhao, W., Lu, J., Zhou, J.: Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo. In: ICCV (2021)
62. Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., Sridhar, S.: Neural fields in visual computing and beyond (2021)

63. Xu, H., Zhou, Z., Wang, Y., Kang, W., Sun, B., Li, H., Qiao, Y.: Digging into uncertainty in self-supervised multi-view stereo. In: ICCV (2021)
64. Yang, G., Huang, X., Hao, Z., Liu, M.Y., Belongie, S.J., Hariharan, B.: Pointflow: 3d point cloud generation with continuous normalizing flows. In: ICCV (2019)
65. Yen-Chen, L., Florence, P., Barron, J.T., Rodriguez, A., Isola, P., Lin, T.Y.: iNeRF: Inverting neural radiance fields for pose estimation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2021)
66. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. In: CVPR (2021)
67. Yücer, K., Sorkine-Hornung, A., Wang, O., Sorkine-Hornung, O.: Efficient 3d object segmentation from densely sampled light fields with applications to 3d reconstruction. ACM Trans. Graph. **35**(3) (2016)
68. Zhang, K., Riegler, G., Snavely, N., Koltun, V.: Nerf++: Analyzing and improving neural radiance fields. ArXiv **abs/2010.07492** (2020)
69. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018)
70. Zhu, L., Mousavian, A., Xiang, Y., Mazhar, H., van Eenbergen, J., Debnath, S., Fox, D.: Rgb-d local implicit function for depth completion of transparent objects. In: CVPR (2021)