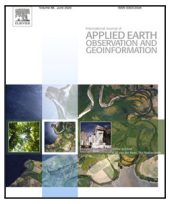
International Journal of Applied Earth Observation and Geoinformation 137 (2025) 104406

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Density uncertainty quantification with NeRF-Ensembles: Impact of data and scene constraints

Miriam Jäger *, Steven Landgraf, Boris Jutzi

Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

ARTICLE INFO

Keywords:
Neural Radiance Fields
Deep Ensembles
NeRF-Ensembles
Density uncertainty
3D reconstruction

ABSTRACT

In the fields of computer graphics, computer vision and photogrammetry, Neural Radiance Fields (NeRFs) are a major topic driving current research and development. However, the quality of NeRF-generated 3D scene reconstructions and subsequent surface reconstructions, heavily relies on the network output, particularly the density. Regarding this critical aspect, we propose to utilize NeRF-Ensembles that provide a density uncertainty estimate alongside the mean density. We demonstrate that data constraints such as low-quality images and poses lead to a degradation of the rendering quality, increased density uncertainty and decreased predicted density. Even with high-quality input data, the density uncertainty varies based on scene constraints such as acquisition constellations, occlusions and material properties. NeRF-Ensembles not only provide a tool for quantifying the uncertainty but exhibit two promising advantages: Enhanced robustness and artifact removal. Through the mean densities, small outliers are removed, yielding a smoother output with improved completeness. Furthermore, applying a density uncertainty-guided artifact removal in post-processing proves effective for the separation of object and artifact areas. We conduct our methodology on 3 different datasets: (i) synthetic benchmark dataset, (ii) real benchmark dataset, (iii) real data under realistic recording conditions and sensors.

1. Introduction

The groundbreaking research on Neural Radiance Fields (NeRFs) (Mildenhall et al., 2020) initiated a new era in computer graphics. NeRFs take a sparse set of images and corresponding camera poses to predict position-dependent density and view-dependent color values, creating new views through the so-called novel-view synthesis. Furthermore, NeRFs provide opportunities in computer vision and photogrammetry by enabling 3D point cloud and mesh reconstructions. However, these density-dependent reconstructions are challenging (Jäger and Jutzi, 2023) and often yield noisy and incomplete surfaces (Li et al., 2023; Wang et al., 2021), while the surfaces are implicit (Lehtola et al., 2022). Unfortunately, the appearance of the resulting geometry in the 3D scene reconstruction depends on the network output and since this density can be interpreted as the differential probability of a ray terminating at an infinitesimal particle at location (Mildenhall et al., 2020), especially the predicted density. Whereas the network optimization is based on the error between training images and rendered images, the density is not controlled with respect to a ground truth information and only learned implicitly. Additionally, each neural network generally leads to a slightly different prediction, which invariably leads to different results.

In this context, the application of techniques to approximate the predictive uncertainty, such as Deep Ensemble (Lakshminarayanan et al., 2017), is of interest. Considering NeRFs, our investigations focus on the impact of data and scene constraints on the density uncertainty in 3D space. For this purpose, we apply NeRF-Ensembles, effectively used in previous work (Sünderhauf et al., 2023), to analyze the impact of data and scene constraints. Such NeRF-Ensembles consist of randomly initialized, independent members of individual NeRFs, providing a quantification of the density uncertainty of the network output, as well as a mean density. Since the density is position-dependent, different NeRFs estimate density values with varying magnitudes, making it possible to calculate the uncertainty of the output from multiple models by using the empirical standard deviation. This makes NeRF-Ensembles straightforward, with minimal modifications to existing NeRF frameworks, without alterations to the training process. Moreover, we not only obtain the estimated uncertainties but also the mean density values, which can contribute to robustness, comparable to random forests (Criminisi and Shotton, 2013).

In this paper, we aim to identify crucial factors for achieving optimal 3D scene reconstructions with NeRFs, forming the foundation for subsequent 3D point cloud and mesh reconstructions. Our

* Corresponding author.
E-mail address: miriam.jaeger@kit.edu (M. Jäger).

https://doi.org/10.1016/j.jag.2025.104406
Received 7 February 2024; Received in revised form 23 January 2025; Accepted 2 February 2025
Available online 24 February 2025

primary focus lies in analyzing the impact of data constraints, such as the quality of input images and camera poses, as well as scene constraints, including acquisition constellations, occlusions, and material properties. We assess quality in terms of rendering performance and density uncertainty, with particular emphasis on understanding how these constraints influence the reliability of the network density estimate. In addition, we provide insights into additional potential for further applications of NeRF-Ensembles. This includes analyzing the 3D reconstruction of NeRF-Ensembles with mean density compared to single NeRFs for enhanced robustness. Furthermore, we propose the use of NeRF-Ensembles for a density uncertainty-guided removal of (foggy) artifacts in post-processing, suggesting a promising approach for addressing fog, ghostly, or floater artifacts in NeRFs, as discussed in Liu et al. (2023), Somraj and Soundararajan (2023), Warburg et al. (2023), Wirth et al. (2023). The code has been released at https://github.com/miriamelie/NeRF-Ensemble.

Our main contributions can be summarized as follows:

- **Data constraints**. Low-quality images and poses have an impact on the training process, as indicated by the Peak Signal-to-Noise Ratio (PSNR) and the density uncertainty of the NeRF output: Image noise σ_{Im} results in lower PSNR, but not necessary higher density uncertainty. While pose noise in translation σ_{t}, rotation σ_{R} and their combination $\sigma_{\text{t,R}}$ results in both lower PSNR and higher density uncertainty. Overall, the density decreases significantly under data constraints, accompanied by the occurrence of (foggy) artifacts and a noisy scene reconstruction.
- **Scene constraints**. Even with high-quality input data the density uncertainty in the scene varies based on acquisition constellations, occlusions and material properties. Areas of the scene that are poorly covered or obscured exhibit high density uncertainty, resulting in the presence of (foggy) artifacts.

In addition to the density uncertainty quantification, NeRF-Ensembles provides further potential and two advantages:

- **Robustness**. The usage of the mean density from NeRF-Ensembles eliminates small outliers in the network output and yields a higher completeness of fine structures and low-textured surfaces, along with a kind of smoothing.
- **Artifact Removal**. Large (foggy) artifacts due to unsuitable acquisition constellations and occlusions can be removed in post-processing by density uncertainty-guided separation of object and artifact areas.

2. Related work

In this section, we summarize related work to our study. Thereby, we give an overview on basic and related research and development on Neural Radiance Fields, Uncertainty Estimation in Deep Learning and Uncertainty Estimation in Neural Radiance Fields.

Neural radiance fields. The foundation for the Neural Radiance Fields (NeRFs) was established by Scene Representation Networks (Sitzmann et al., 2019). Their underlying principle is modeling the scene as a function of 3D coordinates within it. It was followed by the groundbreaking research of Neural Radiance Fields (NeRFs) (Mildenhall et al., 2020). These enable estimation of color values and densities for each 3D coordinate through 6D camera poses and associated 2D images by training a multi-layer perceptron (MLP). The vanilla NeRF was followed by thousands of publications driving research and development in various domains. Camera poses are an essential part of NeRFs. Therefore, some works deal with the efficient training in the case of incorrect or insufficient camera poses and their refinement. BaRF (Lin et al., 2021) relies on bundle block adjustment to train NeRFs with insufficient camera poses. CBARF (Fu et al., 2023) cascades multiple bundle adjustment modules at different levels for stepwise pose refinement. GaRF (Chng et al., 2022) introduces Gaussian-activated radiance field to avoid the drop into local minima during training. Also, embedding pose refinement during training through image data is realized based on residues between pixels of the rendered and training images (Lin et al., 2023). In contrast to the latter mentioned methods, the following do not require proper initial camera poses. GNeRF (Meng et al., 2021) trains a coarse NeRF with random camera poses for approximate poses, which are then refined. NoPe-NeRF (Bian et al., 2023) incorporates undistorted monocular depth priors to constraint the positions of relative poses between images. Since the vanilla NeRF is subject to a training duration of hours to days, some works such as AdaNeRF (Kurz et al., 2022), FastNeRF (Garbin et al., 2021) and Instant NGP (Müller et al., 2022) focus on faster training and rendering. Plenoxels, a non-neural radiance field (Fridovich-Keil et al., 2022) also manages this by representing the scene as a sparse 3D grid. A similar approach is taken by the latest work, 3D Gaussian Splatting (Kerbl et al., 2023). Starting from a sparse point cloud as the initial Gaussian and learning the scene by splitting, growing and expanding Gaussians. However, since we want to investigate the quality of common radiance fields, we use the Instant NGP NeRF implementation in our study, which uses a combination of small MLPs and spatial hash table encoding for real-time training and rendering. Several works are also dedicated to enable artifact-free scene reconstruction, as artifacts often occur in the 3D scene. This is realized for example by visibility analysis that cause artifacts (Somraj and Soundararajan, 2023; Warburg et al., 2023). Whereby Warburg et al. (2023) refines the estimated density and color values in cubes to remove floater artifacts. Another approach performs post-processing removal based on consistency metrics in local neighborhoods to identify floater artifacts (Wirth et al., 2023). Clean-NeRF (Liu et al., 2023) identifies view-dependent appearances e.g. from sparse input views during the training. Current work removes artifacts in the rendered images based on uncertainties from Bayesian Laplace approximation in null spaces due to the ray geometry (Goli et al., 2023).

Uncertainty estimation in deep learning. The significance of uncertainties is continuously and progressively pronounced, not only within the domain of computer vision, since observations and decisions in computer vision are inherently uncertain Meidow (2008), but also extends to neural networks. Because of the large number of model parameters and non-linearities in deep neural networks, the exact posterior probability distribution of a network's output prediction is intractable (Blundell et al., 2015; Loquercio et al., 2020). Therefore, approximative uncertainty quantification approaches like Bayesian Neural Networks (BNN) (MacKay, 1992), as well as Monte Carlo Dropout (MCD) (Gal and Ghahramani, 2016), and Deep Ensembles (DE) (Lakshminarayanan et al., 2017) have been introduced. BNNs provide a mathematical approach where a deterministic neural network is transformed into a stochastic one. This can be achieved by placing probability distributions over the activations or the weights (Jospin et al., 2022). Sampling from these distributions at test time creates an ensemble of models which is used to sample from the posterior distribution of the predictions (Blundell et al., 2015). Gal et al. Gal and Ghahramani (2016) propose MCD as an approximation of a stochastic process to overcome the high computational cost of BNNs. MCD introduces dropout, the well-established regularization method (Srivastava et al., 2014), at test time to sample from the posterior distribution of the predictions. Just like softmax probabilities, MCD uncertainties are not calibrated (Gal and Ghahramani, 2016; Guo et al., 2017). This drawback is overcome by DEs, where an ensemble of trained models produces a sample of predictions at test time (Lakshminarayanan et al., 2017). By a random weight initialization or different data augmentations across the ensemble members, a diverse set of models can be trained (Fort et al., 2020). Due to the introduced randomness across the ensemble members, DEs are well-calibrated (Lakshminarayanan et al., 2017) and outperform other uncertainty quantification methods like softmax probability, BNNs, or MCD (Ovadia et al., 2019; Wursthorn et al., 2022).

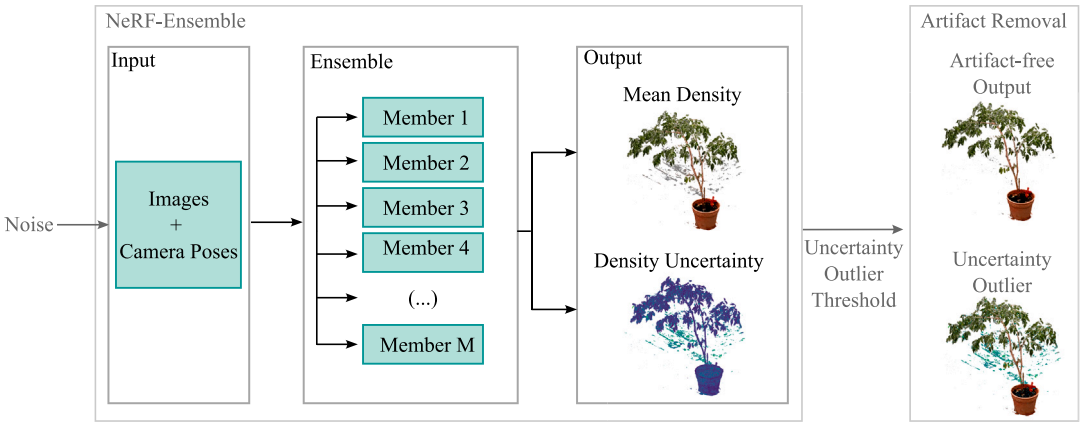


Fig. 1. Flowchart of the methodology: Input data are the images and corresponding camera poses. A NeRF-Ensemble with a total number of M members is trained. The NeRF-Ensemble provides an average network output with mean density values over all NeRFs in the ensemble and a corresponding density uncertainty quantification of the prediction in 3D space. Optionally (shown in gray) we add noise to the input data and investigate the effect of data constraints on the results. In addition, a subsequent removing of high uncertainty points, offers the opportunity to remove (foggy) artifacts.

To overcome the substantial computational cost of DEs, the concept of knowledge distillation (Hinton et al., 2015) has been successfully proposed by multiple prior works (Holder and Shafique, 2021; Landgraf et al., 2024, 2023; Simpson et al., 2022).

Uncertainty estimation in neural radiance fields. Uncertainty estimation techniques can also be applied to NeRFs, where the uncertainty in the rendered images has been addressed in various works. In order to capture potential transient objects in the modeled scene, uncertainty on a pixel-level was addressed within the network by NeRF in the Wild (NeRF-W) (Martin-Brualla et al., 2021). To ensure robustness in case of limited input data, ActiveNeRF (Pan et al., 2022) uses the uncertainty estimation within the training process. Stochastic Neural Radiance Fields (S-NeRF) (Shen et al., 2021) learns a probability distribution over all the possible radiance fields modeling the scene. Conditional-Flow NeRF (CF-NeRF) (Shen et al., 2022) builds up on this, but learns a distribution over all possible complex radiance fields modeling the scene by coupling Latent Variable Modeling and Conditional Normalizing Flows without prior assumptions. The identification of uncertainty caused by a lack of input data is addressed through research on termination probabilities along individual rays by using ensembles (Sünderhauf et al., 2023). Shen et al. (2021, 2022) addresses the limitations of ensembling in NeRF context due to the computationally expensive rendering process for a single NeRF model and its reliability. However, Sünderhauf et al. (2023) shows that ensembling is an effective way to assess uncertainties in NeRFs, and real-time applications such as Instant NGP (Müller et al., 2022) enable real-time training and rendering of multiple NeRFs.

In contrast to previous work, we do not focus on quantifying uncertainty in image space, as this primarily pertains to rendering quality. Instead, we concentrate on the parameter density and its uncertainty, which are essential for accurate geometric representations. Specifically, we employ ensembling by creating NeRF-Ensembles to investigate the impact of data and scene constraints on the density estimates of NeRFs in 3D space. To explore the position-dependent density and its uncertainty across the entire 3D scene, we analyze the implicit continuous field by discretizing it into a dense 3D grid. This enables a comprehensive representation of the density component in 3D space and allows a straightforward estimation of density uncertainty by the empirical standard deviation with the arithmetic mean of the densities. In addition to our analysis of various data and scene constraints, we propose leveraging NeRF-Ensembles to enhance the robustness of 3D reconstruction and a density uncertainty-guided artifact removal during post-processing.

3. Methodology

In this section, we describe our methodology according to Fig. 1, where an ensemble of several NeRFs are trained on the same input data to build a NeRF-Ensemble. The NeRF-Ensemble provides an average network output with mean density over all NeRFs in the ensemble and a corresponding quantification of uncertainty of predicted density uncertainty in 3D space. Section 3.1 outlines the process of the density uncertainty estimation and mean density calculation using NeRF-Ensembles. The input data is given by the images and corresponding camera poses. To investigate the impact of data constraints on the NeRF training process and density uncertainty, we add noise to the input data (Section 3.2), image noise and pose noise.

3.1. Density uncertainty

NeRFs implicitly learn the 3D scene by optimizing the weights and biases of the network. Since the learned field represents a continuous implicit function, we discretize the implicit field by a 3D grid. This is realized by an equidistant fine sampling of positions within a 3D grid of the scene. This allows us to achieve a comprehensive approximated coverage of the entire continuous 3D scene, which would not be possible with several single ray-based samplings or in image space, for instance.

The estimation of the density uncertainty is based on the output of a NeRF-Ensemble. Each NeRF represents a member of an ensemble of NeRFs trained on the same input data. Thereby the varying number of NeRFs are randomly initialized and independently trained. The randomly initialization of the weights (Glorot and Bengio, 2010) of the NeRFs is achieved in the Instant NGP (Müller et al., 2022) implementation. On this basis, the density uncertainty estimation is approximated by the density output of all members in the ensemble. For the position-wise uncertainty estimation, the same position for each NeRF member is regarded.

For each single 3D position X = x,y,z in the 3D density grid, the density uncertainty U_δ is calculated by the empirical standard deviation with the arithmetic mean density $\bar{\delta}$ results from Eqs. (1) and (2):

$$U_\delta(X) = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left(\delta_m(X) - \bar{\delta}(X)\right)^2}, \tag{1}$$

with

$$\bar{\delta}(X) = \frac{1}{M}\sum_{m=1}^{M}\delta_m(X), \tag{2}$$

where M is the total number of members per ensemble and $\bar{\delta}(X)$ is the arithmetic mean value of the density values $\delta(X)$ per position X.

A specification of the mean uncertainty in the 2D image domain is derived from the uncertainty of all pixels (Holder and Shafique, 2021; Landgraf et al., 2023). Since the 3D scene reconstruction is performed in 3D space when it comes to NeRFs, we calculate the mean density uncertainty mU_δ and mean density based on all positions X = x,y,z in a 3D density grid which approximated the scene, respectively, by Eqs. (3) and (4)

$$mU_\delta = \frac{1}{N}\sum_X U_\delta(X), \tag{3}$$

where N is the total number of positions in the 3D grid, and $U_\delta(X)$ is computed using Eq. (1) and

$$m\bar{\delta} = \frac{1}{N}\sum_X \bar{\delta}(X), \tag{4}$$

where $\bar{\delta}(X)$ is computed using Eq. (2).

Note that the density is a unit-less parameter that can be interpreted as the differential probability of a ray terminating at an infinitesimal particle at location (Mildenhall et al., 2020) in 3D to render images with the so-called volume rendering. Accordingly, the density uncertainty is also unit-less.

3.2. Constraints

The output of the network differs within the identical neural network for different training runs with the same input data. Additionally, suggesting that the quality of the input data, due to data constraints, such as image and pose noise, is directly correlated to the density uncertainty of the network output. Furthermore, we assume that increased density uncertainty results from other factors like scene constraints such as acquisition constellation, occlusions, material properties.

Data constraints: Image noise, pose noise. To analyze the impact of data constraints on the rendering quality as well as density and density uncertainty, different types of input data configurations are investigated. Thereby synthetic noise is applied to both the images and camera poses. The following modifications are considered:

- **Image Noise $\boldsymbol{\sigma_{Im}}$**: Random additive Gaussian noise of standard deviations in RGB color space of $\boldsymbol{\sigma_{Im}}$ is added to the images (Russo, 2003). The noisy pixel values are given by:

$$I'(x,y) = I(x,y) + N(x,y), \quad N(x,y) \sim \mathcal{N}(0, \sigma^2_{Im}), \tag{5}$$

  where $I(x,y)$ are the original pixel intensities, and $N(x,y)$ is the normally distributed noise with variance σ^2_{Im}.
- **Pose Noise Translation $\boldsymbol{\sigma_t}$**: Random additive Gaussian noise of standard deviations of $\boldsymbol{\sigma_t}$ is added to the translation component t of the transformation matrix. The noisy translation component t' is computed as:

$$t' = t + N_t, \quad N_t \sim \mathcal{N}(0, \sigma^2_t), \tag{6}$$

  where t is the original translation vector, and N_t is the noise added to the translation.
- **Pose Noise Rotation $\boldsymbol{\sigma_R}$**: Random additive Gaussian noise of standard deviations of $\boldsymbol{\sigma_R}$ is added on the Euler angles of the rotation component R of the transformation matrix. The noisy rotation component R' is calculated as:

$$R' = R + N_R, \quad N_R \sim \mathcal{N}(0, \sigma^2_R), \tag{7}$$

  where R is the original rotation matrix, and N_R is the noise added to the rotation.
- **Pose Noise Translation and Rotation $\boldsymbol{\sigma_{t,R}}$**: Combines the rotation $\boldsymbol{\sigma_t}$ and translation noise $\boldsymbol{\sigma_R}$ as mentioned above, resulting in $\boldsymbol{\sigma_{t,R}}$.

Scene constraints: Acquisition constellation, occlusions, material properties. In addition to data constraints due to low-quality input data, images and camera poses, we aim to investigate other possible sources of an increased density uncertainty. Under real recording conditions, data is not always recorded with a proper acquisition constellation and coverage of the scene. Moreover, the objects in the scene are not always well textured and can reflect complexly in the images depending on the recording situation and material properties. For this reason, we analyze the impact of the following scene constraints on the density uncertainty: Acquisition constellations, occlusions, material properties.

4. Experimental setup

In this section, we introduce the setup for our experiments. We present the datasets which include different scene constraints in Section 4.1, the training metric in Section 4.2, the implementation details in Section 4.3 and the configurations on data constraints in Section 4.4.

4.1. Datasets

Our experiments are based on three different datasets: (i) synthetic benchmark dataset, (ii) real benchmark dataset, (iii) real data under realistic recording conditions and sensor. The datasets differ in their characteristics and thereby include different levels and types of scene constraints such as acquisition constellation, occlusions and material properties.

Synthetic benchmark dataset. Since synthetic data have accurate camera poses and images and cover a wide range of object properties, which are particularly suitable for comprehensive analyses involving their modification. Therefore, we use the NeRF synthetic benchmark dataset (Mildenhall et al., 2020). The dataset consists of synthetic blender images with corresponding ideal camera poses, whereas we use 100 images each. The acquisition constellation provides images in an upper hemisphere or complete sphere that covers a majority of the scene, with a few gaps.

Real benchmark dataset. We use the DTU benchmark dataset (Jensen et al., 2014), which includes scenes featuring real objects, including images and corresponding camera poses. We focus on six scenes, each containing either 49 or 64 RGB images. The acquisition constellation provides images in an upper half hemisphere that covers the scene from one side only.

Real dataset. Additionally, we use real HoloLens data (Jäger et al., 2023) captured with the Microsoft HoloLens generation 2 to investigate the effects of realistic recording conditions and sensors. It consists of 64 images from the 1920 × 1080 photo/video RGB camera of the HoloLens and corresponding internal camera poses of the device, as well as refined poses by Instant NGPs (Müller et al., 2022) pose refinement (Lin et al., 2023). The acquisition constellation provides images in an upper hemisphere that covers the scene from all sides.

4.2. Training metric

The evaluation of the impact of the input data on the rendering quality in the training process is done using the Peak Signal-to-Noise Ratio (PSNR) in dB. It is a common metric to evaluate the quality of the synthetically rendered images compared to the training images, e.g. for NeRFs (Mildenhall et al., 2020) and gives an indication of the training accuracy. It can be defined as follows:

$$\mathrm{PSNR} = -10 \cdot \log_{10}\left(\frac{\mathrm{max\_value}^2}{\mathrm{MSE}}\right), \tag{8}$$

whereas the maximum value max_value is 255 for RGB color images and MSE der Mean Squared Error between the pixel values of two images.

4.3. Implementation details

For all investigations, the Instant NGP (Müller et al., 2022) NeRF implementation was taken into account, since it enables real-time training and rendering. Regarding the network architecture, the basic NeRF architecture with ReLu activations and hash encoding is selected, while the training incorporates 50 000 training steps on a NVIDIA RTX3090 GPU. Unless otherwise noted, we use 10 member per NeRF-Ensemble for all of our experiments, following prior work (Fort et al., 2020; Lakshminarayanan et al., 2017; Landgraf et al., 2023; Sünderhauf et al., 2023).

The density uncertainty estimation in our application refers to the direct density network output of the NeRF, which is not yet affected by any further activation function than ReLU that takes place for neural rendering. An exponential density activation before the rendering in the Instant NGP (Müller et al., 2022) implementation would cause a non-linear scaling of the data. This increases uncertainties in high density scale ranges between the members of the NeRF-Ensemble non-linearly and would make them less comparable to the density values in low ranges. Furthermore, for part of the analyses we only consider the network output with a density greater than 15. High density values lead to large alpha channels (Mildenhall et al., 2020) in NeRFs. Therefore, points with high density are visible in the rendered image through neural view synthesis, while low density areas hardly have an effect on the scene reconstruction as well as the rendered images. This makes them less relevant in case of noise.

4.4. Data constraints: Image noise, pose noise

Several configurations are investigated in order to study the impact of the data constraints (see Section 3.2) on the NeRF output. For this purpose, the images and corresponding original ground truth camera poses of each dataset are used directly and modified by adding random Gaussian noise on images and poses. For the NeRF synthetic and DTU dataset, in addition to the original input data without noise, the configurations of Table 1 are analyzed. For the ideal, synthetic dataset, we chose several uniform additive noise steps to systematically investigate the impact of noise in regular increments. The image noise σ_{Im}, translation noise σ_{t}, as well as rotation noise σ_{R} and translation and rotation noise $\sigma_{\mathrm{t,R}}$ are added according to Eqs. (5), (6) and (7). The dimension of the NeRF synthetic dataset is unit-less, therefore we add Gaussian noise as a percentage of the circumference of the training camera centers on a (hemi-) sphere, while the noise for DTU is metric-scaled. For the HoloLens data, we apply both the HoloLens' internal camera poses and the poses refined under Instant NGPs pose refinement (Lin et al., 2023) with the same corresponding images.

4.5. Scene constraints: Acquisition constellation, occlusions, material properties

Scene constraints arise from challenging acquisition constellations, occlusions in the scene and their various material properties. The used three datasets are suitable for investigating these constraints, covering spherical, hemispherical, and half-hemispherical acquisition constellations. Occlusions of the objects result from these setups and an uneven coverage of the images. The materials in the datasets, featuring diverse objects, encompass a wide range of material properties, including different fabrics, semi-transparent, and reflective materials.

5. Experimental results

The evaluation is categorized into two main points: Firstly, *Data Constraints: Image Noise, Pose Noise* in Section 5.1. The qualitative variation of the training process and density uncertainty based on different qualities of input data. Secondly, *Scene Constraints: Acquisition constellation, occlusions, material properties* in Section 5.2. The density uncertainty of noticeable anomalies from scene constraints in the results. In addition, this Section provides an outlook on the potential of NeRF-Ensembles in Section 5.3 regarding robustness and removal of (foggy) artifacts. Furthermore, we provide an ablation study in Section 5.4 for the training time of members and ensembles, as well as an analyzation of the impact of the ensemble size.

Table 1
Data Constraints. Configurations of the random additive Gaussian Image and Pose Noise on the NeRF synthetic and DTU dataset. For the NeRF synthetic dataset, the cameras on a (hemi-) sphere with a non-metric circumference of circa 25.33. We add Gaussian noise of a percentage of the circumference for σ_{t}.

	NeRF synthetic	DTU
Image Noise σ_{Im}	5 10 15 20	15
Pose Noise σ_{t}	0.01% 0.02% 0.03% 0.04%	0.1 mm
Pose Noise σ_{R}	0.1° 0.2° 0.3° 0.4°	0.01°
Pose Noise $\sigma_{\mathrm{t,R}}$	0.01%, 0.1° 0.02%, 0.2° 0.03%, 0.3° 0.04%, 0.4°	0.1 mm, 0.01°

5.1. Data constraints: Image noise, pose noise

This Section presents the impact of data constraints. It includes a quantitative evaluation of the training process (Section 5.1.1), as well as a quantitative and qualitative evaluation of the density uncertainty (Section 5.1.2) due to data constraints.

5.1.1. Training

Rendering quality. Concerning the NeRF synthetic dataset, the quantitative results indicate a consistent trend, as shown in the Table 2. There is a rapid drop in PSNR for the lowest additive noise per noise type σ_{Im}, σ_{t}, σ_{R} and $\sigma_{\mathrm{t,R}}$. Image noise $\sigma_{\mathrm{Im}} = 5$ leads to a decrease in PSNR of 1 bis 3 dB. After an initial rapid drop of the PSNR when adding the lowest additive noise, the PSNR per scene continues to decrease for additional noise. Adding noise with a steady magnitude also leads to a steady decrease in PSNR almost continuously in steps of 1 to 4 dB depending on noise type and scene. In addition, no further reduction in PSNR is observed for the combination of σ_{t} and σ_{R} to $\sigma_{\mathrm{t,R}}$. A similar trend is noticed for the DTU dataset, as shown in Table 3. The occurrence of noise leads to a lower PSNR value. Again, the PSNR value for all scenes decreases due to additive Gaussian noise of all types σ_{Im}, σ_{t}, σ_{R} between 1 and 3 dB. The combination of noise in translation and rotation $\sigma_{\mathrm{t,R}}$ again does not lead to a corresponding additive decrease in the PSNR, just as with the NeRF synthetic dataset. The real HoloLens data provides a PSNR of 25 dB with the internal, unrefined poses. Again, the PSNR is lower than for input data where pose refinement takes place with a PSNR of 27.5 dB.

5.1.2. Density uncertainty

Quantitative. The quantitative analysis of the training process using the PSNR has shown a decrease in the quality of the rendered images with increasing data constraints from additive noise on input data. The average density uncertainty mU_{δ} in the 3D grid results show a similar behavior to the PSNR results for different input data configurations. For the synthetic NeRF dataset as well as DTU dataset in Tables 4 and 6, an increase in density uncertainty with increasing noise is noticeable for almost all scenes. It is striking that although noise on the images leads

Table 2
Mean PSNR in dB on the NeRF synthetic dataset per NeRF-Ensemble, where a higher PSNR indicates better rendering quality.

		Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship
Original		37.07	30.68	36.14	38.98	38.44	33.71	38.55	32.85
Image Noise σ_{Im}	5	35.63	30.13	32.35	35.29	35.21	32.34	37.13	30.67
	10	33.28	29.17	31.65	31.36	31.60	30.58	34.91	28.12
	15	30.94	28.06	30.45	28.68	29.19	28.83	32.78	26.22
	20	29.24	27.09	29.55	26.84	27.44	27.32	31.26	24.78
Pose Noise σ_t	0.01%	31.57	28.67	32.98	34.61	31.10	31.62	33.29	29.70
	0.02%	29.51	26.93	30.85	32.46	28.61	29.74	31.63	28.17
	0.03%	28.60	26.27	29.87	31.03	27.44	28.65	30.34	27.47
	0.04%	27.91	25.76	29.10	30.36	26.44	27.88	29.69	26.76
Pose Noise σ_R	0.1°	28.69	26.76	29.96	31.54	27.91	29.14	30.60	27.57
	0.2°	27.12	25.31	28.47	29.62	25.74	27.24	28.55	26.16
	0.3°	26.09	24.56	27.62	27.88	25.10	26.41	28.33	25.60
	0.4°	25.64	23.78	27.34	27.50	24.40	25.53	27.09	25.19
Pose Noise $\sigma_{t,R}$	0.01%, 0.1°	28.71	26.47	29.87	31.12	27.48	28.73	30.27	27.47
	0.02%, 0.2°	27.01	24.90	28.29	29.37	25.80	26.85	28.67	26.23
	0.03%, 0.3°	26.07	24.11	27.47	27.91	24.99	26.11	26.89	25.46
	0.04%, 0.4°	25.59	23.87	27.03	27.16	24.46	25.15	25.94	25.10

Table 3
Mean PSNR in dB on the DTU dataset per NeRF-Ensemble, where a higher PSNR indicates better rendering quality.

		Scan24	Scan37	Scan40	Scan55	Scan63	Scan114
Original		30.27	28.31	28.22	29.17	36.41	32.31
Image Noise σ_{Im}	15	24.54	24.34	24.11	23.79	26.39	24.49
Pose Noise σ_t	0.01 mm	26.83	27.08	26.25	26.16	35.33	29.72
Pose Noise σ_R	0.1°	27.01	27.03	26.42	25.86	35.33	29.89
Pose Noise $\sigma_{t,R}$	0.01 mm, 0.1°	27.05	26.97	26.26	25.36	35.60	29.88

to a lower PSNR, the density uncertainty does not necessarily increase rapidly. Another noticeable aspect in addition to the increase of the density uncertainty from data constraints is the significant decrease of the average mean density $m\bar{\delta}$ (Table 5) in the 3D grid for each scene.

To illustrate the impact of noise on the scenes, Fig. 4 shows the spider charts for the highest noise in each scene. This highlights the overall trend that almost all scenes in the synthetic dataset for non-noisy input data range from 0 to 1. With increasing noise, the uncertainty increases from 1 to 2. For the DTU dataset, results are more balanced across the scenes. A clear trend is evident between the density uncertainty of noisy and original data without noise. The visualization of histograms for points in the entire discretized 3D space in Figs. 2 and 3 visualize the distribution of density uncertainty in the 3D grid, illustrating how data constraints influence the density uncertainty. For the scenes of the NeRF synthetic dataset, it shows that the uncertainty of the density values of the grid are increasingly low for non-noisy data. The values then shift towards higher uncertainty with noise. In addition, the combination of pose noise in the translation and rotation components leads to a further shift towards higher uncertainties. The scenes of the DTU dataset show similar results. Here, peaks with low uncertainties occur more frequently in non-noisy data.

Qualitative. The trends of the quantitative investigations are reflected in the visual results. Fig. 5 shows the reconstructed points in the 3D grid above density of 15 for the NeRF synthetic dataset. Visualized are results from a single NeRF, NeRF-Ensemble as well as the range of the density uncertainty for original input data and input data with additive Gaussian noise. It is evident that a low level of pose noise σ_t = 0.01% (column 4) results in a scene with higher density uncertainty U_δ. Pose noise $\sigma_R = 0.1°$ (column 5) also leads to a visually higher density uncertainty. In addition, the noise in the scene increases, resulting in more artifacts which were not present for data without noise (column 3) at the same density threshold. A similar result is obtained for the DTU dataset in Fig. 6, where even a low level of pose noise $\sigma_t = 0.1$ mm and $\sigma_R = 0.01°$ leads to a significant increase in density uncertainty U_δ. Again, the point clouds are noisier and contain more (foggy) artifacts. In addition, some of the points are missing, which corresponds to the quantitative results showing that the density for low-quality input data, especially poses, decreases significantly. In all scenes, the image noise is barely noticeable in the density uncertainty U_δ and visually only leads to slightly higher noise. The HoloLens data, see Fig. 7, displays a higher density uncertainty U_δ (see Fig. 7(d)) for the internal camera poses, which is particularly evident in the artifacts of the scene. In addition, the density for this configuration is lower, see Fig. 7(c). In contrast, the density for the refined internal camera poses is higher (Fig. 7(g)), while the density uncertainty is lower (see Fig. 7(h)) and there are barely any (foggy) artifacts at all. The analysis make it clear that apparent artifacts or fog have similar high density values as parts of the object and are therefore present in rendered images as well as reconstructed 3D scene. On the other hand, the results (see Figs. 5, 6, 7) show that the density uncertainty of the network is high for these points.

5.2. Scene constraints: Acquisition constellation, occlusions, material properties

This section presents the impact of the scene constraints and covers a qualitative evaluation of the density uncertainty.

5.2.1. Density uncertainty

The visual results show that the density uncertainty U_δ of the network is varying for different areas of the object. This also applies to the case of non-noisy high-quality input data, which is visualized in column three of Figs. 5–7. In these overview images, high density uncertainties U_δ mainly relate to occluded areas of the scenes that are barely visible. Furthermore, it becomes apparent that the DTU dataset contains high density uncertainties U_δ for many scenes, especially those with low-textured subsurface. For the NeRF synthetic dataset, material properties-dependent uncertainties occur as well.

Fig. 8 provides a detailed representation of some scenes along with the associated acquisition constellations displayed as the three axes of the poses. Hidden object parts due to the acquisition constellation

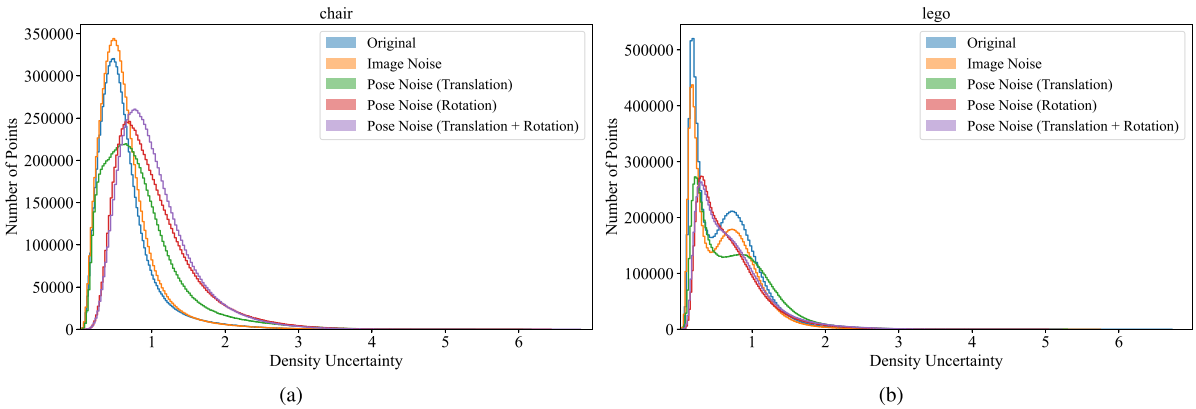


Fig. 2. The histogram illustrates the distribution of the density uncertainty U_δ for points in a 3D grid for the NeRF synthetic. Shown are the scenes under different input data types: Original, image noise σ_{Im}, pose noise σ_t (translation), pose noise σ_R (rotation) and pose noise $\sigma_{t,R}$ (translation + rotation).

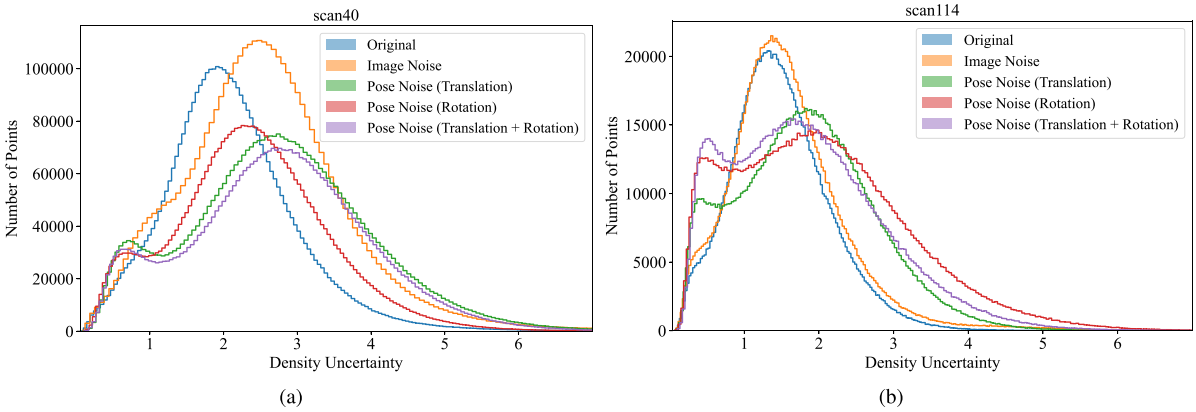


Fig. 3. The histogram illustrates the distribution of the density uncertainty U_δ for points in a 3D grid for the DTU dataset. Shown are the scenes under different input data types: Original, image noise σ_{Im}, pose noise σ_t (translation), pose noise σ_R (rotation) and pose noise $\sigma_{t,R}$ (translation + rotation).

Table 4
Mean **density uncertainty** mU_δ per NeRF-Ensemble of the discretized 3D grid for the NeRF synthetic dataset. A lower mU_δ indicates a more certain estimation of the density across the different NeRF members in the ensemble.

		Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship
Original		1.14	1.33	0.87	1.00	0.86	0.64	1.32	1.04
Image Noise	5	0.64	1.38	0.93	0.78	0.84	0.62	1.23	0.71
σ_{Im}	10	1.13	1.37	0.91	0.70	0.90	0.61	1.21	0.61
	15	1.13	1.35	0.87	0.72	0.96	0.54	1.29	0.60
	20	0.98	1.35	1.04	0.72	1.03	0.55	1.26	0.52
Pose Noise	0.01%	1.57	1.44	0.86	1.07	1.10	0.75	1.32	0.92
σ_t	0.02%	1.60	1.58	0.81	1.08	1.13	0.86	1.31	0.90
	0.03%	1.60	1.64	0.68	1.22	1.12	0.97	1.43	0.88
	0.04%	1.54	1.83	0.63	1.31	1.19	1.03	1.45	0.94
Pose Noise	0.1°	1.58	1.62	0.67	1.26	1.09	0.92	1.38	0.90
σ_R	0.2°	1.06	1.66	0.63	1.28	1.23	1.05	1.41	0.94
	0.3°	1.41	1.95	0.64	1.21	1.26	1.03	1.40	0.93
	0.4°	1.42	1.84	0.69	1.31	1.34	1.10	1.40	1.00
Pose Noise	0.01%, 0.1°	1.51	1.60	0.67	1.31	1.14	0.92	1.37	0.86
$\sigma_{t,R}$	0.02%, 0.2°	1.50	1.93	0.63	1.24	1.20	1.09	1.13	0.89
	0.03%, 0.3°	1.46	1.99	0.62	1.30	1.30	1.10	0.97	0.93
	0.04%, 0.4°	1.41	1.77	0.68	1.21	1.30	1.17	1.87	0.95

and occlusions show higher density uncertainties U_δ. For example, the scene lego, as well as the underside of the scene chair. Uncertainties due to poor coverage of the recordings, which result in occlusions, and the acquisition constellation in general should also be mentioned. This becomes particularly clear with the DTU dataset, whose camera poses are located on a one-sided hemisphere. The back of the object, which is not visible from the camera poses, is therefore not observed. This appears to be the case for subsurfaces without texture. Large white artifacts are reconstructed within the white surfaces in scenes scan24, scan37 and scan40. These are not visible in the rendered images due to the color, but the coordinates of the points contain estimated density values and are therefore present in reconstructed 3D point clouds, which have high uncertainties. This is not the case for scenes scan55 and scan114 with colored, slightly textured subsurface.

Table 5
Mean **mean density** $m\bar{\delta}$ per NeRF-Ensemble of the discretized 3D grid for the NeRF synthetic dataset.

		Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship
Original		1355.76	960.98	273.40	6464.55	1473.02	277.65	375.84	326.37
Image Noise	5	1253.29	719.84	232.71	1508.90	1251.83	202.08	360.59	260.58
σ_{Im}	10	1183.46	630.44	206.77	935.41	1088.10	161.90	364.08	219.51
	15	1076.47	559.19	206.90	547.41	944.62	127.83	318.88	156.06
	20	694.58	521.85	274.78	342.16	779.80	111.82	289.47	88.69
Pose Noise	0.01%	756.92	800.25	228.00	1116.22	721.74	239.66	323.07	248.41
σ_t	0.02%	472.92	620.29	196.26	581.12	354.94	198.60	314.27	185.94
	0.03%	369.09	473.94	140.31	348.64	259.16	195.23	278.88	180.62
	0.04%	294.23	395.66	116.55	321.44	226.63	180.30	256.00	166.25
Pose Noise	0.1°	391.66	513.99	146.19	439.23	276.89	197.37	317.08	177.62
σ_R	0.2°	215.00	365.51	100.87	258.87	173.44	164.31	234.44	148.93
	0.3°	149.26	269.80	86.97	170.35	144.43	132.70	165.83	113.11
	0.4°	132.26	223.46	78.20	151.77	110.57	121.82	142.94	120.88
Pose Noise	0.01%, 0.1°	363.79	524.51	152.87	367.99	262.80	193.13	268.45	175.37
$\sigma_{t,R}$	0.02%, 0.2°	228.43	336.29	96.10	220.46	167.69	143.01	212.43	140.22
	0.03%, 0.3°	163.45	249.97	85.97	161.31	136.88	129.60	136.22	132.94
	0.04%, 0.4°	126.09	225.10	77.69	132.76	117.13	128.76	130.11	107.65

Table 6
Mean **density uncertainty** mU_δ per NeRF-Ensemble of the discretized 3D grid for the DTU dataset. A lower mU_δ indicates a more certain estimation of the density across the different NeRF members in the ensemble.

	Scan24	Scan37	Scan40	Scan55	Scan63	Scan114	
Original		1.52	1.09	1.60	1.46	1.45	1.35
Image Noise σ_{Im}	15	1.25	0.83	1.49	1.47	1.18	1.48
Pose Noise σ_t	0.01 mm	2.13	1.53	1.62	1.50	1.52	1.62
Pose Noise σ_R	0.1°	2.02	1.68	1.40	1.45	1.57	1.40
Pose Noise $\sigma_{t,R}$	0.01 mm, 0.1°	2.22	1.73	1.71	2.30	1.74	1.71

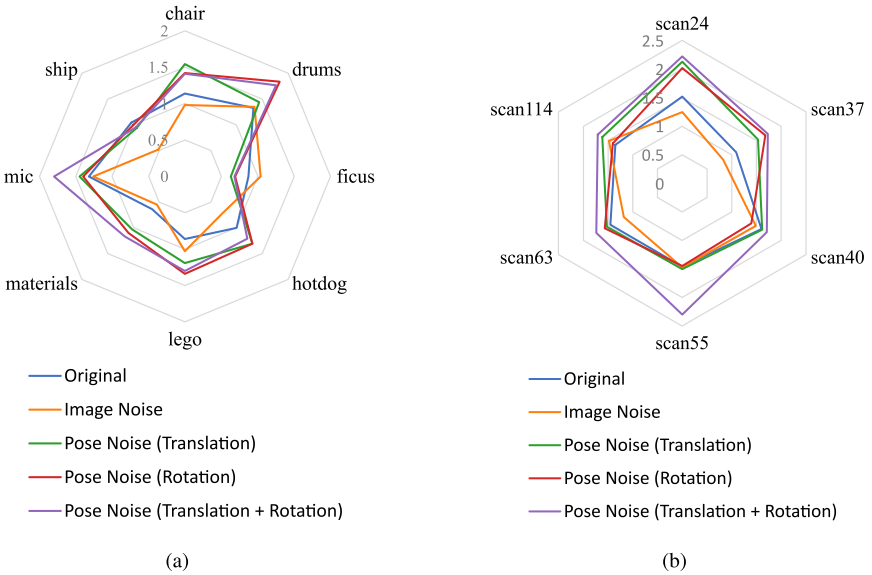


Fig. 4. Spider chart showing the density uncertainty for NeRF synthetic (left) and DTU dataset (right). Displayed are the mean density uncertainty mU_δ in the 3D grid for different types of input data: original, image noise σ_{Im}, pose noise (translation) σ_t, pose noise (rotation) σ_R and pose noise (translation + rotation) $\sigma_{R,t}$.

NeRFs predict the continuous density field in the entire 3D space and thus also inside the object, which are always occluded in the images. As the cross-sections in Fig. 9 show, a high density uncertainty U_δ is also present within the objects compared to the object surface.

With regard to specific material properties, the results of the NeRF synthetic dataset in Fig. 5 are representative. It becomes evident that the density uncertainty U_δ varies based on the material properties of the object within the scene. For example, in scene chair, the stripes show higher uncertainties than the rest of the fabric. But also, materials like in scene drums and scene materials show higher uncertainties in areas that are reflective or semi-transparent as shown in detail in Fig. 10. The comparison of the density uncertainty between scan40 and scan63 also indicates that the surface texture affects the uncertainty. While the low-textured subsurface in scan40 shows only a slightly higher uncertainty,

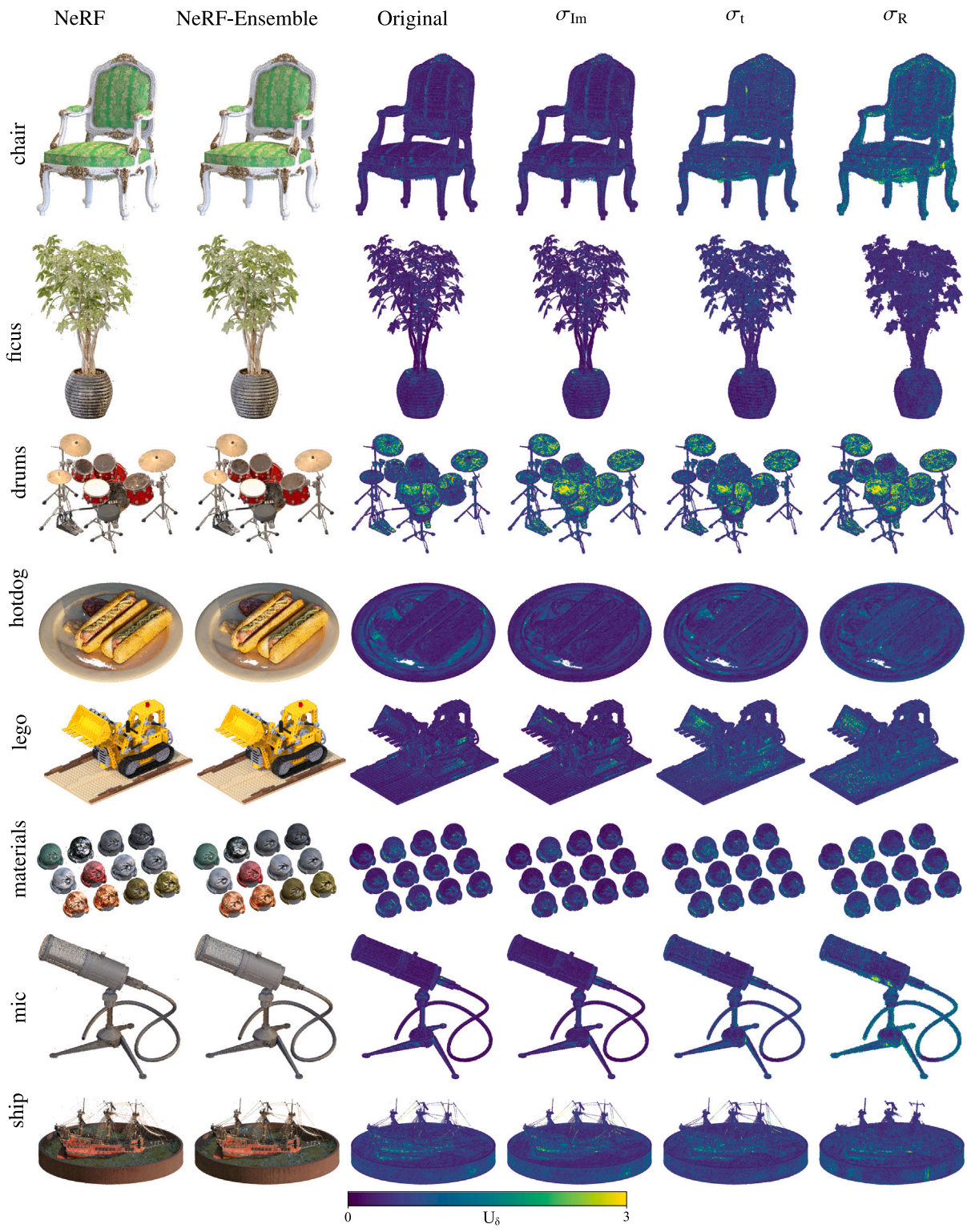


Fig. 5. Qualitative comparison on the NeRF synthetic dataset, showing points in the 3D grid with density above 15. Shown are the individual scenes for single NeRF (column 1), NeRF-Ensemble (column 2) as well as the density uncertainty U_{δ} under different input data configurations: Original (column 3), image noise σ_{Im} (column 4), pose noise σ_{t} (column 5) and pose noise σ_{R} (column 6). Density uncertainty values above 3 are set to 3 for clearer visualization.

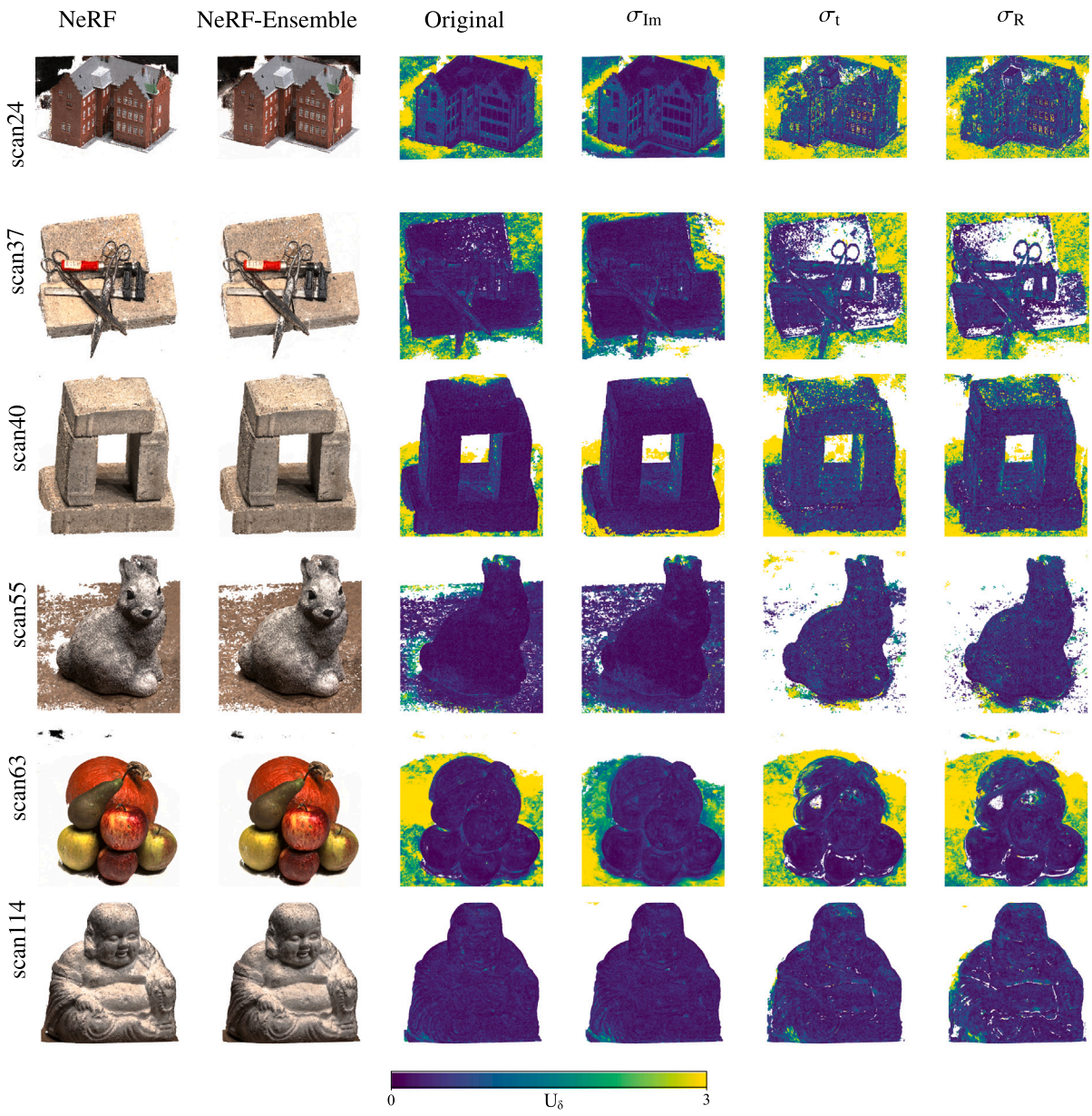


Fig. 6. Qualitative comparison on the DTU dataset, showing points in the 3D grid with density above 15. Shown are the individual scenes for single NeRF (column 1), NeRF-Ensemble (column 2) as well as the density uncertainty U_δ under different input data configurations: Original (column 3), image noise σ_{Im} (column 4), pose noise σ_{t} (column 5) and pose noise σ_{R} (column 6). Density uncertainty values above 3 are set to 3 for clearer visualization.

the uncertainty for the textureless white subsurface of scan63 is greatly increased.

5.3. Potential of NeRF-Ensembles

In addition to the density uncertainty quantification, we present further promising advantages of NeRF-Ensembles. Firstly, increased robustness by the usage of the mean densities from NeRF-Ensembles in Section 5.3.1. Secondly, we provide an approach for a density uncertainty-guided removal of (foggy) artifacts in Section 5.3.2, which inevitably results from our findings.

5.3.1. Robustness

In addition to the quantification of the density uncertainty, NeRF-Ensembles show further advantages. This is shown in the previous qualitative results in Figs. 5, 6, 7 for each dataset by comparing the respective columns 1 for a NeRF and 2 for a NeRF-Ensemble. NeRF-Ensembles can rely on the mean density per position in contrast to the density from a single NeRF. In doing so, NeRF-Ensemble lead to a reduction of small artifacts and outliers, as well as a higher object completeness. Fig. 11 outlines this advantage of NeRF-Ensembles in detail for a scene from NeRF synthetic dataset, DTU dataset and HoloLens data. Small artifacts and outliers are removed (comparison of Fig. 11(a) to 11(b)), a higher completeness of low-textured surfaces is reached (comparison of Fig. 11(c) to 11(d)) as well as a higher completeness of fine structures (comparison of Fig. 11(e) to 11(f)).

5.3.2. Artifact removal

In general, the removal of so-called fog, ghostly, floater or foggy artifacts (Liu et al., 2023; Somraj and Soundararajan, 2023; Warburg

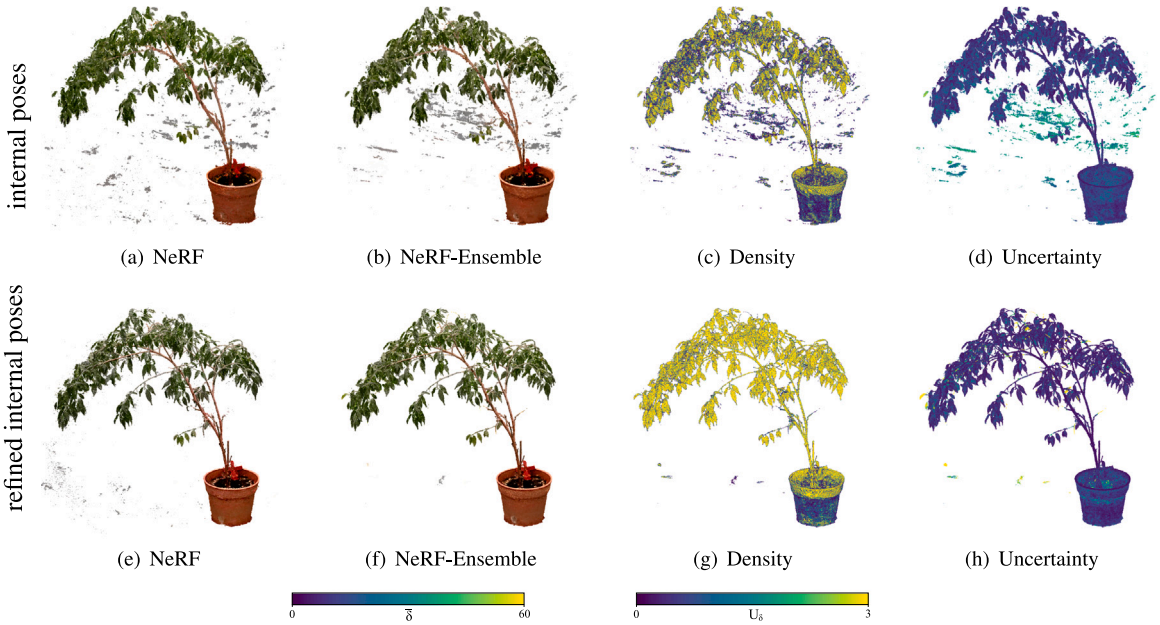


Fig. 7. Qualitative comparison on the HoloLens data, showing points in the 3D grid with density above 15. For the reconstruction from internal HoloLens camera poses: NeRF (a), NeRF-Ensemble (b) and the respective mean density $\overline{\delta}$ (c) and density uncertainty U_δ (d). For the reconstruction from refined internal HoloLens camera poses: NeRF (e), NeRF-Ensemble (f) and the respective mean density $\overline{\delta}$ (g) and density uncertainty U_δ (h).

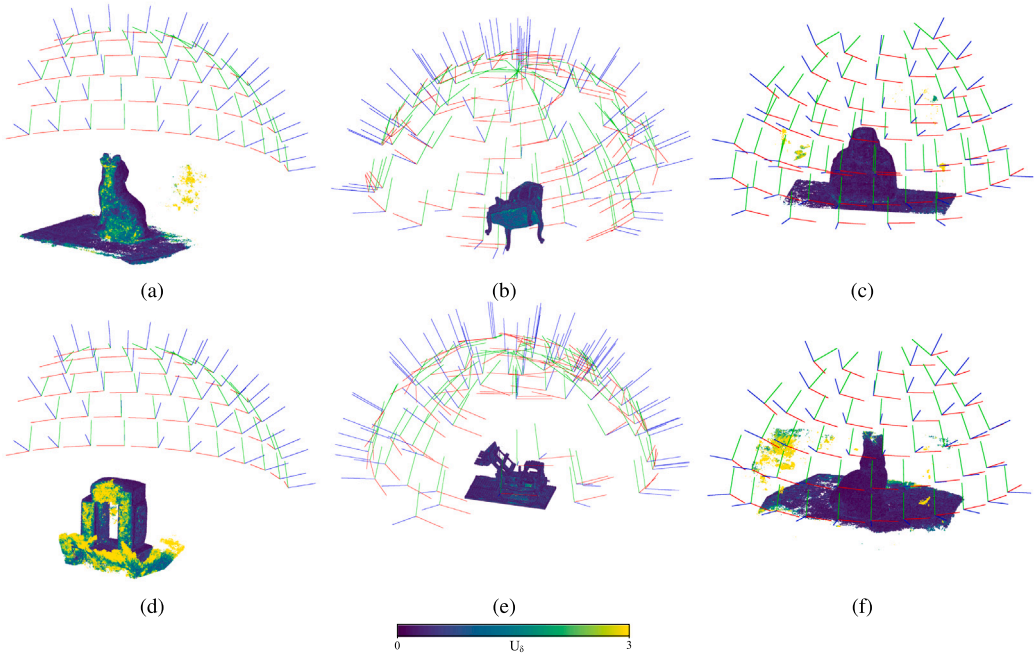


Fig. 8. Acquisition constellation and occlusions. Visualized is the density uncertainty U_δ. The poses of the corresponding images are shown in their constellation with colored orientation of the 3 axes of the poses. Column 1 (a), (d): Back of the object which is occluded and not covered due to the acquisition constellation, which is an upper half hemisphere. Column 2 top (b): Underside of the object that is not covered due to the acquisition constellation, which is an upper hemisphere. Column 2 bottom (e): Object from the side, whereby there are gaps at the sides in the acquisition constellation and areas in the object that are barely visible due to occlusions from the scene geometry. Column 3 (c), (f): (foggy) artifacts with high density uncertainty due to the acquisition constellation, which is an upper half hemisphere.

et al., 2023; Wirth et al., 2023) is of great interest when it comes to NeRFs. Highlighted by our findings, (foggy) artifacts occur due to scene constraints from the acquisition constellation and occlusions (see Section 5.2.1) or in the case of low-quality poses (Jäger et al., 2023) (see Section 5.1.2). Moreover these foggy artifacts change their shape and appearance in the scene during training process. Therefore, we assume that (foggy) artifacts in the 3D scene are subject to high density variations, resulting in a high density uncertainty of the NeRF output.

Building on this, we aim to investigate the uncertainties of those artifacts and propose a possible solution for their removal. Specifically, we introduce a density uncertainty-guided artifact removal approach using NeRF-Ensembles by removing points based on their respective density uncertainties. The objective is to partition the scene points into two distinct distributions and establish a robust threshold for separating object points from artifacts. Similar to the approach in SIFT (Lowe, 2004), which determines whether a keypoint match is

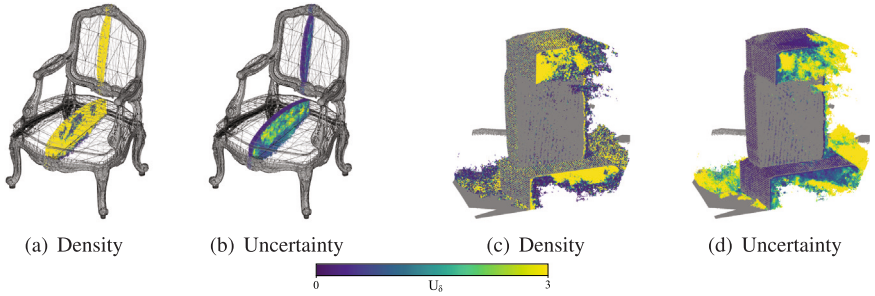


Fig. 9. Occlusions. Cross-sections through the objects. The mean density $U_{\bar{\delta}}$ and density uncertainty U_{δ} are shown in comparison to the reference point clouds and meshes (in gray). The density in the object is usually high, while the density uncertainty U_{δ} is higher in the interior compared to the surface of the object.

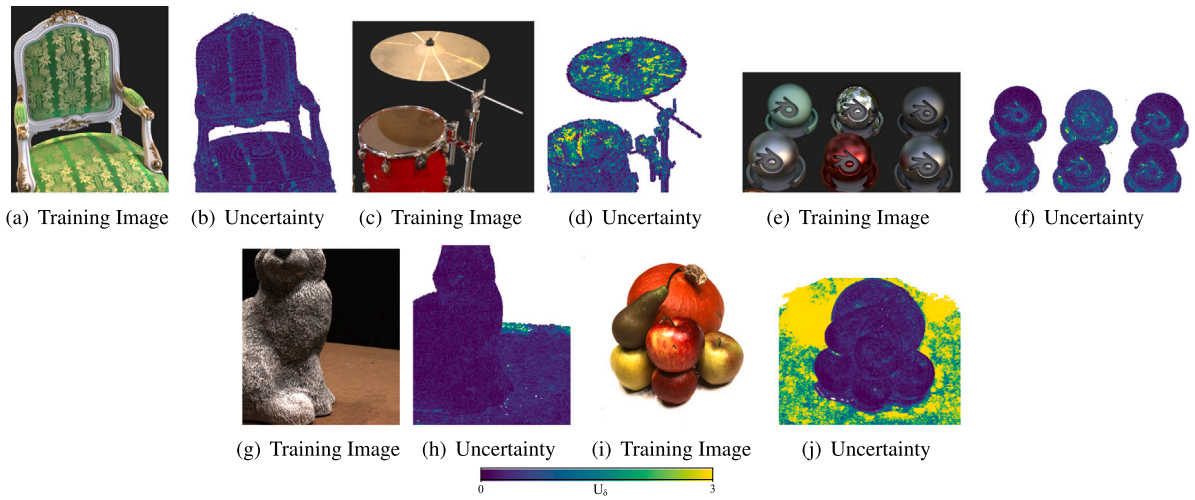


Fig. 10. Material properties. Comparison of a detailed sections of the training images versus the density uncertainty U_{δ}: Scene chair (a) versus (b) with different fabrics, scene drums (c) versus (d) with reflective or semi-transparent material, scene materials (e) versus (f) with reflective material, scene scan40 (g) versus (h) with low-textured material, scene scan63 (i) versus (j) with textureless material.

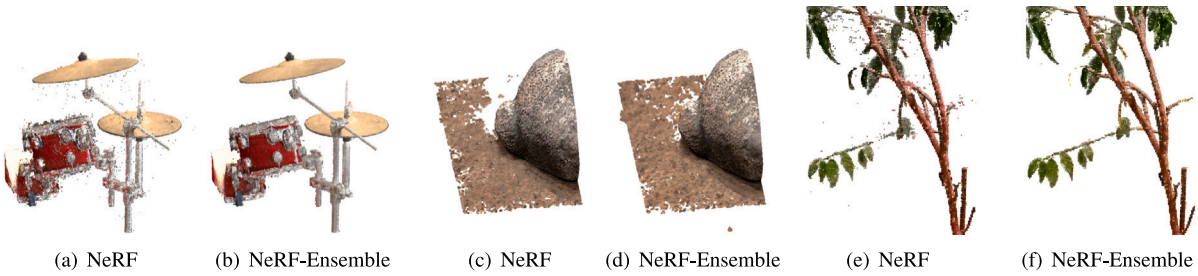
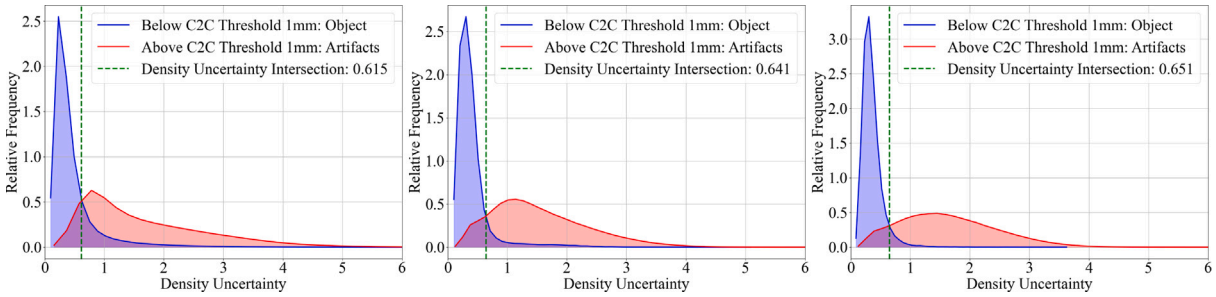


Fig. 11. Comparison of a detailed sections of the 3D point clouds each for a single NeRFs versus a NeRF-Ensemble: NeRF (a) versus NeRF-Ensemble (b) for the nerf synthetic dataset, NeRF (c) versus NeRF-Ensemble (d) for the DTU dataset, as well as NeRF (e) versus NeRF-Ensemble (f) for the HoloLens data with refined poses.

correct or incorrect, we aim to find the intersection of the relative frequency histogram distributions to classify points as objects or artifacts. Relative histograms are employed because the distributions of object and artifact points are assumed to be of unequal sizes. However, our approach works reversely: To determine this threshold, we first compute the cloud-to-cloud (C2C) distances between the scene and the reference point cloud. The resulting points are separated based on a small distance threshold close to zero. This allows us to infer the corresponding density uncertainty, which leads to the separation of the two point distribution sets in the relative histograms.

To demonstrate our approach, we focus on scenes using the original, noise-free input data in order to exclude or minimize density uncertainties that may arise from data constraints. A C2C distance threshold of 1 mm is selected to maintain a close approximation to the ground truth object surface. Fig. 12 shows the overlapping relative histograms of the resulting object and artifact distributions. The distributions are relatively bimodal: the object points exhibit density uncertainties predominantly below 1, with their peak around 0.4, while the density uncertainty of artifact points ranges primarily from 0 to 3, with peaks at approximately 0.8, 1 and 1.5. All histograms can be separated at the intersection points, with corresponding thresholds of density uncertainties of 0.615, 0.641 and 0.651. Fig. 13 shows the results of filtering the point clouds using the intersection density uncertainty thresholds. The threshold effectively removes most artifacts; however, small gaps remain in certain areas. As the distributions are not fully disjoint, their complete separation is not achievable without either leaving artifacts in the scene or compromising the object completeness. Nonetheless, we conclude that a density uncertainty threshold in the range of 0.6 to 0.7 represents a suitable initial candidate for artifact removal.



(a) scan40 (b) scan55 (c) scan114

Fig. 12. Overlapping relative histograms for object and artifact point distributions derived from three different scenes (scan40, scan55, scan114). The histograms illustrate the density uncertainty values for object points (blue) and artifact points (red) after applying a 1 mm cloud-to-cloud (C2C) distance threshold. The green dashed lines represent the density uncertainty thresholds derived from the intersection points of the two histograms.

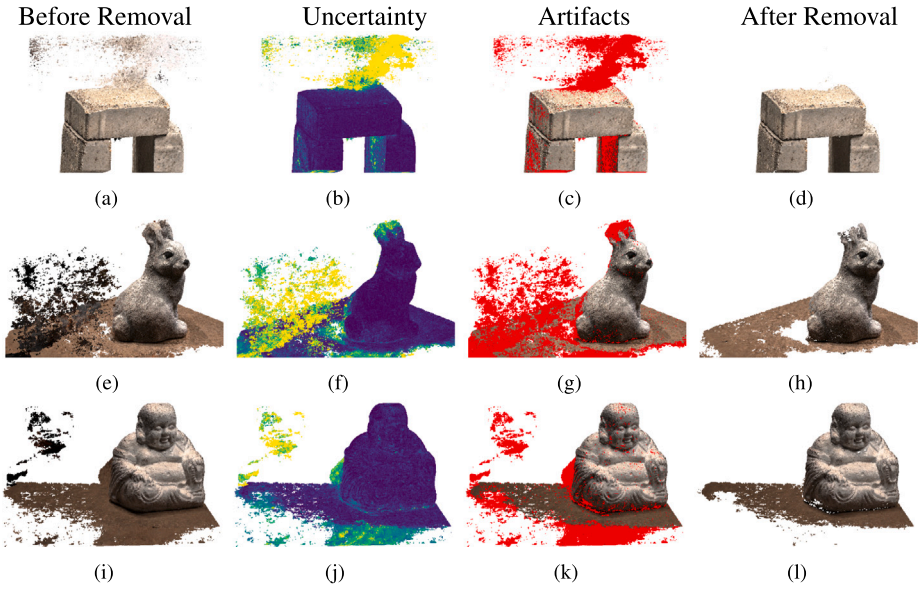


Fig. 13. Artifacts from acquisition constellation. From left to right: Point cloud before artifact removal, density uncertainty U_δ and point cloud with artifacts, point cloud after artifact removal. The respective points with uncertainties above the intersection value of the relative histogram are marked red in the artifacts point cloud.

5.4. Ablation studies: Computation time and convergence

Impact of ensemble size. For training an ensemble, different numbers of members can be chosen to approximate the predictive uncertainty. The number of members influences how reliable the calculated density uncertainty is. Following Lakshminarayanan et al. (2017), we evaluate the predictive uncertainty as a function of ensemble size M. Fig. 14 depict runs on the original scenes of the three datasets with varying numbers of M = 2 to 10 members per ensemble. The results show that the density uncertainty converges and stabilizes after approximately 5 to 6 members per ensemble. This suggests that a minimum of M = 5 members per ensemble is required to ensure that the assessment of density uncertainty in 3D space, discretized by the 3D grid, is stable.

Training time. In this section, we analyze the training time for the NeRF-Ensembles across three different datasets. For all experiments, the Instant NGP (Müller et al., 2022) NeRF implementation with 50 000 iterations on a NVIDIA RTX3090 GPU, was used (see Section 4.3). NeRF-Ensembles require additional computational resources due to the need for multiple models. The computational complexity (Remzan et al., 2024) is approximately $O(M \cdot t)$, where M is the number of ensemble members and t is the training time of a single model. The explicit training times for 50 000 iterations on DTU and NeRF synthetic dataset are shown in Tables 7 and 8. The ensemble training time for a HoloLens member was around 54.00 min, resulting in a average time of 5.40 min for a single member.

6. Discussion

So far, our research using NeRF-Ensembles takes a critical look at the complex interplay between the input data with data constraints, as well as the scene with scene constraints, and the resulting rendering quality as well as the density and density uncertainty.

A key aspect concerns the *data constraints* given by the quality of the input data. Low-quality poses not only significantly degrade the training process, as indicated by a decrease in the PSNR. But also contribute to an increased density uncertainty of the NeRF output. Hereby, pose noise σ_t, σ_R and $\sigma_{t,R}$ all affect the density uncertainty. Along with an increased density uncertainty, a significant decrease in density through low-quality input data was evident. Overall, data constraints lead to a noisier output with large (foggy) artifacts. Beyond the data constraints due to image noise and pose noise, our investigation extended to the impact of *scene constraints* on the density uncertainty. Even with high-quality input data, variations and anomalies exist in the density uncertainty due to other factors. These factors are represented by acquisition constellations, such as a lack of views and occlusions, e.g. when images are captured from one side only. This leads to (foggy) artifacts and a high density uncertainty in these occluded areas. Moreover, material properties of the object, such as different fabric or reflective, are a possible source of higher density uncertainty. It has been observed that reflective surfaces exhibit higher density uncertainty. We attribute this to the fact that such surfaces appear differently

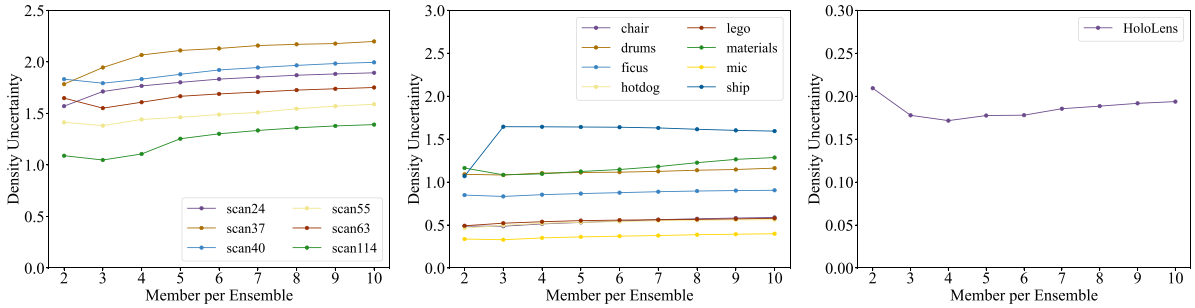


Fig. 14. Predicted mean density uncertainty in 3D space, discretized by the 3D grid, depending on the number of members $M = 2$ to 10 per ensemble . Left: DTU dataset, center: NeRF synthetic dataset, right: Hololens dataset.

Table 7
Training times in minutes for 50,000 iterations: For a NeRF-Ensemble with 10 members, and the average training time for a single NeRF member for NeRF synthetic dataset.

Training time	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship
Ensemble in min	51.70	44.20	40.60	48.70	51.20	41.70	48.10	45.20
mean Member in min	5.17	4.42	4.06	4.87	5.12	4.17	4.81	4.52

Table 8
Training times in minutes for 50,000 iterations: For a NeRF-Ensemble with 10 members, and the average training time for a single NeRF member for DTU dataset.

Training time	Scan24	Scan37	Scan40	Scan55	Scan63	Scan114
Ensemble in min	70.50	60.70	77.00	58.70	63.10	64.10
mean Member in min	7.05	6.07	7.70	5.87	6.31	6.41

from various viewpoints in the training images, making it challenging for the NeRF to estimate the correct density. A similar effect is observed with surface texture. Even surfaces with moderately low texture exhibit increased density uncertainty, while textureless surfaces (e.g., plain white surfaces) show significantly higher density uncertainty. This may be due to the inability to accurately associate the object's surface at these locations due to the lack of texture. As a result, the NeRF estimates varying density values across different viewpoints and networks, which increases the uncertainty in these areas. Moreover, the results of the density uncertainty quantification reveal new potential for the application of NeRF-Ensembles. Firstly, there is an improved robustness from the mean density output of NeRF-Ensembles compared to the density from single NeRFs. The mean density values from ensembles not only eliminates small artifacts and outliers, but also contributes to a more stabilized and higher completeness of the scene. Secondly, NeRF-Ensembles offer promising potential for post-processing, where density uncertainty becomes crucial. The qualitative results indicated that large (foggy) artifact regions tend to have an increased density uncertainty. In our analysis based on overlapping relative histograms of the object and artifact distributions, we investigated which density uncertainty separates the two distributions at their intersection to address these artifacts. The results reveal that the distributions are almost bimodal, leading to consistent intersections across the tested scenes. These intersections occur within a density uncertainty range of 0.6 to 0.7, which constitutes a promising range for initial threshold candidates for artifact removal. However, the distributions are not entirely disjoint, which may result in gaps in the object surface or incomplete artifact removal. For future work, we identify several promising directions: One key area is enhancing the density uncertainty-guided artifact removal. Further research into optimal separation methods for object and artifact areas or complementary filtering strategies could refine our approach and yield more robust results across varying conditions. Another avenue involves density uncertainty-driven pose refinement, on the condition that other constraints on density uncertainty are excluded, e.g., by minimizing the uncertainty while varying the rotation and translation parameters. Additionally, adapting NeRF-Ensembles to dynamic scenes for investigating the density uncertainty under dynamic scene conditions. This involves addressing challenges such as motion blur, object occlusion, and varying perspectives, which introduce temporal changes in the scene and could affect the uncertainty.

Limitations. Nonetheless, building NeRF-Ensembles for uncertainty quantification requires training multiple individual NeRF networks, which increases the computational time. Firstly, faster methods than the original, time-consuming NeRF—such such as the Instant NGP used in this study—can help mitigate computational demands. In our case, training an ensemble of 10 members requires a total of approximately 54.68 min (about 0.91 h), with each individual NeRF taking an average of 5.47 min. Secondly, our investigations on the impact of ensemble size show that density uncertainty stabilizes with approximately 5 to 6 members per ensemble, indicating that using a smaller ensemble size beyond this threshold can reduce computational demands without compromising reliability. Thirdly, approaches such as parallelized training processes, early stopping or knowledge distillation (Holder and Shafique, 2021; Landgraf et al., 2023; Simpson et al., 2022) can be employed. Since NeRF-Ensembles offer benefits beyond density uncertainty quantification, such as enhanced robustness and potential for artifact removal, they can justify the computational demand. Moreover, NeRF-Ensembles are straightforward to implement, requiring minimal modifications to existing NeRF frameworks.

Altogether, we can summarize the following main aspects to enable optimal 3D scene reconstructions with a high training performance regarding the novel view synthesis and a low density uncertainty of the network output: 1) High-quality input data, referring to images and poses. High-quality poses seem to be more important than non-noisy images in terms of the density uncertainty, i.e. the quality of the volumetric density. This is a crucial aspect, when generating 3D point cloud and mesh reconstructions from NeRFs. Nonetheless, high-quality images are important for proper rendering. 2) Acquisition constellations and occlusions: Convergent acquisitions, spherically distributed around the object with as few gaps as possible between the views. 3) Well-textured subsurface and background. 4) The application of NeRF-Ensembles over a single NeRF to achieve a quantification of density uncertainty, to eliminate small outliers by mean density values as well as the density uncertainty-guided removal of (foggy) artifacts in post-processing.

7. Conclusion

In summary, we provided a comprehensive investigation of the density uncertainty of NeRFs using NeRF-Ensembles. Our work examines the interplay of data constraints (such quality of images and poses) and scene constraints (such as occlusions, acquisition constellations and material properties), along with the rendering quality during training and the resulting density and density uncertainty. Moreover, we highlight further potential of NeRF-Ensembles for enhanced robustness and artifact removal.

The results emphasize the crucial impact of data constraints with the quality of the input data, which includes both images and poses. We observe a substantial impact on training performance, as measured by the PSNR. Furthermore, we reveal a correlation between data constraints and density uncertainty. Even with high-quality input data, variations in density uncertainty persist. These variations result from scene constraints, including acquisition constellations, occlusions, and material properties. This highlights the relevance of a detailed density uncertainty quantification when using NeRFs. Moreover, NeRF-Ensembles are suitable for enhancing the robustness due to their multiple votes from several NeRFs, comparable to well-established techniques like random forests (Criminisi and Shotton, 2013). The mean density eliminates small artifacts and outliers and contributes to the reconstruction with higher completeness. Moreover, NeRF-Ensembles offer practical applications in post-processing for the removal of large (foggy) artifacts through a density uncertainty-guided separation of object and artifact areas.

Looking ahead, our findings contribute valuable insights by demonstrating how data and scene constraints affect the uncertainty in NeRF density estimation by using NeRF-Ensembles. We show that the density, which controls the geometric representation in scene reconstruction, and its uncertainty are strongly dependent on the quality and constraints of the data and the scene. Additionally, two further advantages of NeRF-Ensembles beyond uncertainty quantification were addressed: on one hand, improved robustness through the use of the mean density from the ensemble, and on the other hand, artifact removal based on the density uncertainty. As NeRFs are increasingly used for 3D point cloud and mesh reconstructions, our findings on the impact of various constraints on the density uncertainty can offer a framework for the optimal application of NeRFs, opening new opportunities in the fields of computer graphics, computer vision, and photogrammetry.

CRediT authorship contribution statement

Miriam Jäger: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Steven Landgraf:** Methodology. **Boris Jutzi:** Supervision, Resources, Project administration, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Bian, W., Wang, Z., Li, K., Bian, J.-W., Prisacariu, V.A., 2023. NoPe-NeRF: Optimising neural radiance field with no pose prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 4160–4169.

Blundell, C., Cornebise, J., Kavukcuoglu, K., Wierstra, D., 2015. Weight uncertainty in neural network. In: Proceedings of the 32nd International Conference on Machine Learning. Vol. 37, pp. 1613–1622.

Chng, S.-F., Ramasinghe, S., Sherrah, J., Lucey, S., 2022. Gaussian activated neural radiance fields for high fidelity reconstruction and pose estimation. In: European Conference on Computer Vision. Springer, pp. 264–280.

Criminisi, A., Shotton, J., 2013. Decision Forests for Computer Vision and Medical Image Analysis. Springer Science & Business Media.

Fort, S., Hu, H., Lakshminarayanan, B., 2020. Deep ensembles: A loss landscape perspective. arXiv:1912.02757.

Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A., 2022. Plenoxels: Radiance fields without neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5501–5510.

Fu, H., Yu, X., Li, L., Zhang, L., 2023. CBARF: Cascaded bundle-adjusting neural radiance fields from imperfect camera poses. arXiv:2310.09776.

Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning. In: Proceedings of Machine Learning Research, Vol. 48, pp. 1050–1059.

Garbin, S.J., Kowalski, M., Johnson, M., Shotton, J., Valentin, J., 2021. Fastnerf: High-fidelity neural rendering at 200fps. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14346–14355.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, pp. 249–256.

Goli, L., Reading, C., Sellán, S., Jacobson, A., Tagliasacchi, A., 2023. Bayes' rays: Uncertainty quantification for neural radiance fields. arXiv:2309.03185.

Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On calibration of modern neural networks. In: Precup, D., Teh, Y.W. (Eds.), Proceedings of the 34th International Conference on Machine Learning. In: Proceedings of Machine Learning Research, Vol. 70, PMLR, pp. 1321–1330.

Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. In: NIPS Deep Learning and Representation Learning Workshop.

Holder, C.J., Shafique, M., 2021. Efficient uncertainty estimation in semantic segmentation via distillation. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). IEEE, Montreal, BC, Canada, pp. 3080–3087.

Jäger, M., Hübner, P., Haitz, D., Jutzi, B., 2023. A comparative neural radiance field (NeRF) 3D analysis of camera poses from HoloLens trajectories and structure from motion. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLVIII-1/W1-2023, 207–213.

Jäger, M., Jutzi, B., 2023. 3D density-gradient based edge detection on neural radiance fields (NeRFs) for geometric reconstruction. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLVIII-1/W3-2023, 71–78.

Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanæs, H., 2014. Large scale multi-view stereopsis evaluation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 406–413.

Jospin, L.V., Laga, H., Boussaid, F., Buntine, W., Bennamoun, M., 2022. Hands-on Bayesian neural networks—A tutorial for deep learning users. IEEE Comput. Intell. Mag. 17 (2), 29–48.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. 42 (4), 1–14.

Kurz, A., Neff, T., Lv, Z., Zollhöfer, M., Steinberger, M., 2022. Adanerf: Adaptive sampling for real-time rendering of neural radiance fields. In: European Conference on Computer Vision. Springer, pp. 254–270.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Vol. 30, Curran Associates, Inc..

Landgraf, S., Hillemann, M., Kapler, T., Ulrich, M., 2024. Efficient multi-task uncertainties for joint semantic segmentation and monocular depth estimation. In: DAGM German Conference on Pattern Recognition. GCPR, Springer.

Landgraf, S., Wursthorn, K., Hillemann, M., Ulrich, M., 2023. DUDES: Deep uncertainty distillation using ensembles for semantic segmentation. arXiv:2303.09843.

Lehtola, V.V., Koeva, M., Elberink, S.O., Raposo, P., Virtanen, J.-P., Vahdatikhaki, F., Borsci, S., 2022. Digital twin of a city: Review of technology serving city needs. Int. J. Appl. Earth Obs. Geoinf. 102915.

Li, Z., Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.-Y., Lin, C.-H., 2023. Neuralangelo: High-fidelity neural surface reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Lin, C.-H., Ma, W.-C., Torralba, A., Lucey, S., 2021. Barf: Bundle-adjusting neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5741–5751.

Lin, Y., Müller, T., Tremblay, J., Wen, B., Tyree, S., Evans, A., Vela, P.A., Birchfield, S., 2023. Parallel inversion of neural radiance fields for robust pose estimation. In: 2023 IEEE International Conference on Robotics and Automation. ICRA, pp. 9377–9384. http://dx.doi.org/10.1109/ICRA48891.2023.10161117.

Liu, X., Tai, Y.-W., Tang, C.-K., 2023. Clean-NeRF: Reformulating NeRF to account for view-dependent observations. arXiv:2303.14707.

Loquercio, A., Segu, M., Scaramuzza, D., 2020. A general framework for uncertainty estimation in deep learning. IEEE Robot. Autom. Lett. 5 (2), 3153–3160.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

MacKay, D.J.C., 1992. A practical Bayesian framework for backpropagation networks. Neural Comput. 4 (3), 448–472.

Martin-Brualla, R., Radwan, N., Sajjadi, M.S., Barron, J.T., Dosovitskiy, A., Duckworth, D., 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7210–7219.

Meidow, J., 2008. Consideration of uncertainty in computer vision: Necessity and chance. Pattern Recognit. Image Anal. 18, 216–221.

Meng, Q., Chen, A., Luo, H., Wu, M., Su, H., Xu, L., He, X., Yu, J., 2021. Gnerf: Gan-based neural radiance field without posed camera. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6351–6361.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2020. NeRF: Representing scenes as neural radiance fields for view synthesis. In: European Conference on Computer Vision. ECCV, pp. 405–421.

Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. (ToG) 41 (4), 1–15.

Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J., Lakshminarayanan, B., Snoek, J., 2019. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Vol. 32, Curran Associates, Inc..

Pan, X., Lai, Z., Song, S., Huang, G., 2022. ActiveNeRF: Learning where to see with uncertainty estimation. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII. Springer, pp. 230–246.

Remzan, N., Hachimi, Y.E., Tahiry, K., Farchi, A., 2024. Ensemble learning based-features extraction for brain mr images classification with machine learning classifiers. Multimedia Tools Appl. 83 (19), 57661–57684.

Russo, F., 2003. A method for estimation and filtering of Gaussian noise in images. IEEE Trans. Instrum. Meas. 52 (4), 1148–1154.

Shen, J., Agudo, A., Moreno-Noguer, F., Ruiz, A., 2022. Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In: Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III. Springer, pp. 540–557.

Shen, J., Ruiz, A., Agudo, A., Moreno-Noguer, F., 2021. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In: 2021 International Conference on 3D Vision (3DV). IEEE, pp. 972–981.

Simpson, I.J.A., Vicente, S., Campbell, N.D.F., 2022. Learning structured Gaussians to approximate deep ensembles. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 366–374.

Sitzmann, V., Zollhoefer, M., Wetzstein, G., 2019. Scene representation networks: Continuous 3D-structure-aware neural scene representations. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Vol. 32, Curran Associates, Inc..

Somraj, N., Soundararajan, R., 2023. ViP-NeRF: Visibility prior for sparse input neural radiance fields. In: ACM Special Interest Group on Computer Graphics and Interactive Techniques. SIGGRAPH.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Sünderhauf, N., Abou-Chakra, J., Miller, D., 2023. Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. In: 2023 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 9370–9376.

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W., 2021. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. Adv. Neural Inf. Process. Syst. 34, 27171–27183.

Warburg, F., Weber, E., Tancik, M., yński, A.H., Kanazawa, A., 2023. Nerfbusters: Removing ghostly artifacts from casually captured nerfs. In: International Conference on Computer Vision. ICCV.

Wirth, T., Rak, A., Knauthe, V., Fellner, D.W., 2023. A post processing technique to automatically remove floater artifacts in neural radiance fields. In: Computer Graphics Forum. Wiley Online Library, e14977.

Wursthorn, K., Hillemann, M., Ulrich, M., 2022. Comparison of uncertainty quantification methods for CNN-based regression. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLIII-B2-2022, 721–728.