Manifold Sampling for Differentiable Uncertainty in Radiance Fields

LINJIE LYU, Max-Planck-Institut für Informatik, Germany
AYUSH TEWARI, MIT CSAIL, USA
MARC HABERMANN, Max-Planck-Institut für Informatik, Germany
SHUNSUKE SAITO, Meta Codec Avatars Lab, USA
MICHAEL ZOLLHÖFER, Meta Codec Avatars Lab, USA
THOMAS LEIMKÜHLER, Max-Planck-Institut für Informatik, Germany
CHRISTIAN THEOBALT, Max-Planck-Institut für Informatik, Germany

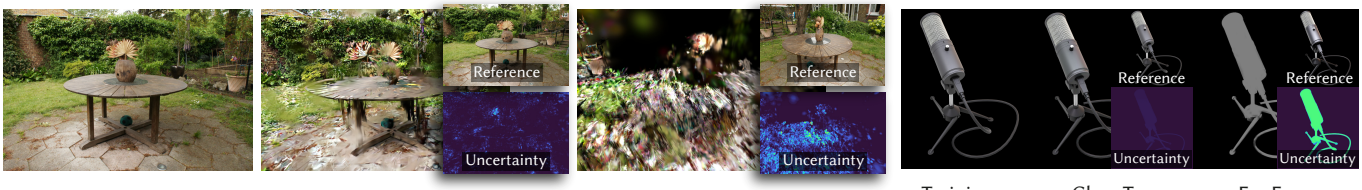

Fig. 1. We propose a method to quantify uncertainty in radiance fields using a novel and efficient manifold sampling strategy. Our approach allows for differentiable optimization of views to reduce reconstruction ambiguities. We demonstrate this capability in next-best view planning (left) and illumination planning (right) tasks.

Radiance fields are powerful and, hence, popular models for representing the appearance of complex scenes. Yet, constructing them based on image observations gives rise to ambiguities and uncertainties. We propose a versatile approach for learning Gaussian radiance fields with explicit and fine-grained uncertainty estimates that impose only little additional cost compared to uncertainty-agnostic training. Our key observation is that uncertainties can be modeled as a low-dimensional manifold in the space of radiance field parameters that is highly amenable to Monte Carlo sampling. Importantly, our uncertainties are differentiable and, thus, allow for gradient-based optimization of subsequent captures that optimally reduce ambiguities. We demonstrate state-of-the-art performance on next-best-view planning tasks, including high-dimensional illumination planning for optimal radiance field relighting quality.

ACM Reference Format:
Linjie Lyu, Ayush Tewari, Marc Habermann, Shunsuke Saito, Michael Zollhöfer, Thomas Leimkühler, and Christian Theobalt. 2024. Manifold Sampling for Differentiable Uncertainty in Radiance Fields. In *SIGGRAPH Asia 2024 Conference Papers (SA Conference Papers'24), December 3–6, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3680528.3687655

1 INTRODUCTION

Recent years have witnessed the overwhelming success of volumetric radiance field representations for scene reconstruction and rendering [Kerbl et al. 2023; Mildenhall et al. 2020; Tewari et al. 2022]. Typically, such algorithms are provided with images observing the scene from multiple viewpoints, but also additional input variation such as time [Park et al. 2021; Xian et al. 2021] or illumination [Martin-Brualla et al. 2021; Srinivasan et al. 2021] have been considered.

The reconstruction of radiance fields from image observations is an instance of an ill-posed inverse problem: Multiple different reconstructions can explain the data [Tarantola 2005]. Therefore, in addition to raw predictions, a complete and robust radiance field model must also provide a measure of its epistemic *uncertainty* [Goli et al. 2024; Hoffman et al. 2023; Jiang et al. 2024; Pan et al. 2022; Shen et al. 2021]. Such an uncertainty estimate should arguably exhibit three key properties: First, it should provide a *high-quality and expressive estimate* of the range of all scene attributes, avoiding proxy computations. Second, the involved computations should be *efficient* to not interfere with high-performance pipelines [Fridovich-Keil et al. 2022; Kerbl et al. 2023; Müller et al. 2022]. Finally, the uncertainty estimate should be *differentiable*. This crucial property turns uncertainty into a functional tool, as it allows to systematically increase model confidence by optimizing for viewpoints, lighting conditions, etc. for subsequent capture – a particularly pressing challenge in resource-constrained applications. Further, differentiability facilitates capture planning for high-dimensional input domains, where plain enumeration and evaluation of candidate capture conditions is infeasible. In this work, we set out to develop a novel, general, and practical approach for uncertainty estimation in radiance fields that exhibits all the above desirable properties.

Existing methods for uncertainty quantification in radiance fields typically come in one of two flavors. Stochastic approaches consider explicit attribute distributions which are optimized during training [Savant et al. 2024; Shen et al. 2022, 2021; Sünderhauf et al. 2023; Yan et al. 2023], often within a variational-inference framework [Blei et al. 2017]. While they directly model distributions of

SA Conference Papers '24, December 3–6, 2024, Tokyo, Japan

parameters, such approaches often require many samples to obtain stable estimates, significantly impairing efficiency. In a second line of work, uncertainty is estimated based on fully trained radiance fields [Goli et al. 2024; Jiang et al. 2024]. Typically, a Laplace approximation [Daxberger et al. 2021; Ritter et al. 2018] is employed, resulting in rather coarse uncertainty proxies. Since these estimates require automatic differentiation, obtaining differentiable uncertainties necessitates the computation of higher-order derivatives, posing significant practical challenges.

Our approach employs a stochastic radiance field [Savant et al. 2024; Shen et al. 2021] with a 3D Gaussian representation [Kerbl et al. 2023] that treats all of its parameters as random variables. Drawing samples from the joint distribution of parameters allows for rendering different realizations of the radiance field. Both, training and uncertainty estimation during inference, require an integral over realizations, which we estimate using Monte Carlo sampling. However, the high dimensionality of radiance field parameters poses a significant challenge, as the joint distribution is typically complex. Incorporating complete covariance matrices across all parameters is intractable. Assuming full independence between parameters [Pan et al. 2022; Savant et al. 2024; Shen et al. 2021], even though common practice for efficiency, leads to excessive variance. Our key contribution is the observation that restricting samples to a *low-dimensional linear manifold in parameter space* is sufficient for an Monte Carlo estimator to efficiently train and estimate uncertainty in a stochastic radiance field. While stochastic sampling is typically considered inefficient and prone to high variance, with our method drawing only very few samples is enough to obtain stable and high-quality gradients for training and uncertainty estimation (Fig. 1).

The benefits of our formulation are threefold. First, explicit modeling of the parameter distribution provides a convenient and interpretable uncertainty measure. Second, the *low-rank approximation* of the covariance matrix results in a smoother objective energy landscape that lends itself to stable optimization. Further, due to the low number of samples required, our strategy is highly efficient in terms of computational performance and memory requirements, both during training and inference. Third, our uncertainties are trivially differentiable, as they simply arise from the mean of rendered radiance field realizations.

We demonstrate that our approach significantly outperforms previous methods across multiple aspects. A main application of our manifold sampling is next-best view planning, which we evaluate on a variety of relevant scenarios. In addition to the optimal selection of camera candidates [Jiang et al. 2024; Kopanas and Drettakis 2023; Pan et al. 2022], an application where we outperform the state of the art, our approach *for the first time* allows for a *differentiable* fine-grained optimization of camera parameters that optimally reduce uncertainty given the current state of the model. We further show that our approach enables differentiable uncertainty estimation in augmented radiance fields, exemplified by the task of relightable reconstruction [Martin-Brualla et al. 2021; Srinivasan et al. 2021]. Concretely, we optimize for the next-best illumination condition that minimizes relighting uncertainty in radiance fields, demonstrating the versatility of our approach and its ability to handle high-dimensional domains.

In summary, our contributions are:

- A novel stochastic radiance field formulation based on manifold sampling that can be efficiently trained.
- A method for differentiable uncertainty estimation that allows for fine-grained optimization of subsequent scene captures.
- The evaluation of our approach on the tasks of next-best viewpoint and illumination planning.

We provide all source code and pre-trained models on https://vcai.mpi-inf.mpg.de/projects/2024-ManifoldUncertainty/.

2 RELATED WORK

2.1 Deterministic 3D Reconstruction

3D reconstruction is the problem of taking image observations as input and reconstructing the underlying 3D scene. This problem has been studied for a long time and mature systems such as structure-for-motion (SfM) are widely used to reconstruct 3D geometry and cameras from image observations [Schönberger and Frahm 2016]. Recent years have seen a lot of progress on 3D scene representations that enable high-quality novel view synthesis. Early advances relied on neural scene representations, such as neural radiance fields [Mildenhall et al. 2020], which use neural networks to represent geometry and appearance quantities, and are optimized from the input images using inverse volume rendering. More recently, voxel-based representations [Fridovich-Keil et al. 2023, 2022; Müller et al. 2022; Yu et al. 2021] and 3D Gaussians [Kerbl et al. 2023] have been used to represent the underlying scene for efficient reconstruction and rendering. Similar progress has also been seen for relighting problems, where differentiable physics-based rendering, or image-based lighting formulations have been used to compute 3D reconstructions [Chen et al. 2024; Lyu et al. 2022; Mai et al. 2023; Saito et al. 2024; Zhang et al. 2021a,b].

Most of the progress has focused on *deterministic* 3D reconstruction, where the underlying epistemic uncertainties are not taken into account. Instead, these methods rely on a large number of input observations to minimize the uncertainty. In contrast, we are interested in modeling and reconstructing the uncertainties in the problem.

2.2 Uncertainty in 3D Reconstruction

Modeling uncertainty is an active area of research in machine learning. A common way of estimating uncertainty in deep learning is using variational inference, where a distribution over free parameters is estimated [Blundell et al. 2015; Kendall and Gal 2017; Khan et al. 2018; Zhang et al. 2018]. However, variational inference is often very slow and expensive, e.g., representing the full covariance matrix can be infeasible for larger models. Thus, many methods use Laplace's approximation that estimates the posterior distribution from the optimized MAP solution [Denker and LeCun 1990; Foong et al. 2019; Kristiadi et al. 2020; MacKay 1991; Ritter et al. 2018; Savant et al. 2024]. While Laplace's approximation is more efficient, it can struggle with recovering complex distributions.

Both strategies have also been explored in the context of radiance fields and 3D reconstruction. Variational inference-based approaches compute the posterior distribution. S-NeRF [Shen et al. 2021], ActiveNeRF [Pan et al. 2022], and Savant et al. [2024] assume

independence between all parameters of the 3D representation. CF-NeRF [Shen et al. 2022] uses latent space modeling with normalizing flow models to learn more complex distributions. FisherRF [Jiang et al. 2024] and Bayes' Rays [Goli et al. 2024] use Laplace's approximation to compute the uncertainty in reconstruction. Our method is based on variational inference; however, we show that a simple low-dimensional approximation of the covariance leads to tractable training and inference. Our method outperforms methods that assume independence between all parameters, and at a comparable cost to methods that use Laplace's approximation.

We contrast our approach with popular 3D reconstruction methods that rely on dense training data [Hoffman et al. 2023; Long et al. 2024; Sargent et al. 2023; Tewari et al. 2024; Xu et al. 2023]. These approaches assume large and dense training datasets with no uncertainty, which can then be used to recover epistemic uncertainty at test time from sparse observations. Our method does not rely on any dense training data, and only uses the sparse observations to compute the reconstruction.

2.3 Active 3D Reconstruction

Reconstructing estimates of the uncertainty is important for active scanning applications, where the uncertainty can guide the optimal scene conditions, such as camera poses and lighting. The most common application is camera selection, where the sequence of cameras used for scanning is determined by the current estimate of the uncertainty [Jiang et al. 2024; Kopanas and Drettakis 2023; Pan et al. 2022]. We compare to the state-of-the art methods and demonstrate better performance on this task. Further, we also introduce a differentiable version of the task where cameras are allowed to freely move in 3D, rather than restricted to some candidate cameras. We also demonstrate results on lighting optimization where we reconstruct scene appearance from very few lighting conditions. To the best of our knowledge, we are the first to demonstrate this application.

3 BACKGROUND

In this work, we are concerned with radiance fields $F_{\boldsymbol{\theta}}$ that model a 3D scene using the volumetric representation

$$F_{\boldsymbol{\theta}} : (\mathbf{x}, \mathbf{d}, \boldsymbol{\xi}) \rightarrow (\sigma, \mathbf{c}). \tag{1}$$

The radiance field outputs volumetric density $\sigma \in \mathbb{R}_+$ and RGB color $\mathbf{c} \in \mathbb{R}^3$ as a function of three arguments: $\mathbf{x} \in \mathbb{R}^3$ is a location in the scene, $\mathbf{d} \in \mathbb{S}^2$ is a 3D direction vector, and $\boldsymbol{\xi} \in \mathbb{R}^{n_{\boldsymbol{\xi}}}$ is an (optional) vector of auxiliary parameters that depend on the application. For example, $\boldsymbol{\xi}$ is a scalar time dimension in the case of dynamic reconstruction [Park et al. 2021; Xian et al. 2021], or a high-dimensional vector of illumination conditions in the case of relightable scenes [Lyu et al. 2022; Martin-Brualla et al. 2021; Srinivasan et al. 2021].

How to best represent $F_{\boldsymbol{\theta}}$ remains a subject of active research [Chan et al. 2022; Fridovich-Keil et al. 2022; Kerbl et al. 2023; Mildenhall et al. 2020; Müller et al. 2022; Xu et al. 2022; Yu et al. 2021]. In this work, we focus on the 3D Gaussian Splatting (3DGS) representation [Kerbl et al. 2023], which marks the current state of the art in terms of reconstruction quality and rendering speed. This model represents a radiance field as a Gaussian mixture model, where each primitive has a location in 3D space, an anisotropic covariance, an opacity, and a view-dependent color represented in the spherical harmonics (SH) basis. The detailed workings of this model are not central to our approach. Therefore, we treat $F_{\boldsymbol{\theta}}$ as a black box in our exposition and denote all trainable parameters as $\boldsymbol{\theta} \in \mathbb{R}^{d_{\boldsymbol{\theta}}}$. Typically, a high number of parameters is required to parameterize a radiance field, with $d_{\boldsymbol{\theta}}$ often reaching into the millions.

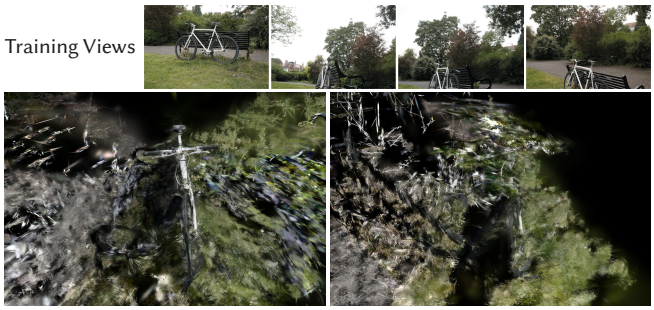

Fig. 2. Using the same four training views (top), different model parameter initializations yield different radiance fields (bottom), reflecting uncertainty.

A radiance field can be rendered to an RGB image $I(\boldsymbol{\xi})$ using emission–absorption volume rendering [Kajiya and Von Herzen 1984]:

$$C_{\boldsymbol{\theta}}(\mathbf{r}, \boldsymbol{\xi}) = \int_{t_n}^{t_f} \exp\left(-\int_{t_n}^{t} \sigma\left(\mathbf{r}(s), \boldsymbol{\xi}\right) \mathrm{d}s\right) \sigma\left(\mathbf{r}(t), \boldsymbol{\xi}\right) \mathbf{c}\left(\mathbf{r}(t), \mathbf{d}, \boldsymbol{\xi}\right) \mathrm{d}t. \tag{2}$$

Here, $C_{\boldsymbol{\theta}}$ is the color of a rendered pixel of I corresponding to a camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ with near and far bounds t_n and t_f, respectively. Again, the subscript $\boldsymbol{\theta}$ denotes its dependence on the trainable parameters of $F_{\boldsymbol{\theta}}$. The primitive-based 3DGS representation solves this integral using an approximation based on rasterization [Zwicker et al. 2001]. Importantly, the rendering process is inherently differentiable, which enables the optimization of radiance field parameters $\boldsymbol{\theta}$ such that renderings match a set of posed training views.

However, regrettably, obtaining a potentially high-dimensional radiance field from image observations alone is an ill-posed problem with nuanced ambiguities: A whole distribution of radiance fields can result in the same rendered image. This is illustrated in Fig. 2, where two static radiance fields are trained using the same four training views (Fig. 2, top) but starting from different initializations. Despite both radiance fields nearly perfectly fitting the training views, rendering the reconstructions from a test view reveals vastly different solutions (Fig. 2, bottom). This problem is even more pronounced in higher-dimensional settings, such as relightable radiance fields, where capturing multiple illumination conditions alongside multiple camera poses is necessary. The degree of ambiguity gradually diminishes as more training views are captured. However, finding an *optimal* capture sequence requires accounting for scene-specific uncertainties. In the following section, we develop an efficient approach to quantify these uncertainties, enabling us to plan the next-best observations that optimally increase model confidence.

4 METHOD

The central goal of this work is to learn a (multi-dimensional) radiance field $F_{\boldsymbol{\theta}}$ in the form of Eq. 1 from image data that provides a fine-grained estimate of its inherent epistemic uncertainty. We are especially focused on computing uncertainty estimates that are both efficient and differentiable, enabling their application in multi-dimensional next-best-view planning scenarios. To this end, we devise a stochastic radiance field formulation (Sec. 4.1) with a low-dimensional manifold sampler at its core (Sec. 4.2) that allows highly efficient training (Sec. 4.3) and differentiable uncertainty estimation (Sec. 4.4).

We assume we have access to training data in the form of one or multiple posed images, providing per-pixel triplets $\mathcal{T} = \left\{\mathbf{r}_i, \boldsymbol{\xi}_i, C_i\right\}_{i=1}^{N_{\mathcal{T}}}$. Note how this setting is a two-fold generalization of typical radiance field reconstruction tasks. First, the additional auxiliary parameter vector $\boldsymbol{\xi}$ accommodates arbitrary and potentially high-dimensional scene configurations, such as different lighting conditions. Second, we do not necessarily require multiple training images. Our algorithm is capable of handling a *single* image as the initial input and utilizes uncertainty estimates to suggest which images to capture next for optimal successive uncertainty reduction.

4.1 Stochastic Radiance Field

We consider a distribution of radiance fields $p(F_{\boldsymbol{\theta}})$ which we realize using a probabilistic model, i.e., by explicitly modeling the joint distribution $p(\boldsymbol{\theta})$ of all trainable parameters. Drawing a sample $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta})$ gives rise to a specific radiance field realization $F^*_{\boldsymbol{\theta}}$. Our objective is to optimize the posterior distribution $p(F_{\boldsymbol{\theta}}|\mathcal{T})$ by minimizing rendering errors for training views while concurrently maximizing uncertainty for novel views. This approach aims to obtain a posterior distribution that is as spread out as possible while making sure that *each* realization $F^*_{\boldsymbol{\theta}}$ explains the training data $\mathcal{T}$, i.e., when $F^*_{\boldsymbol{\theta}}$ is rendered using inputs $\mathbf{r}_i$ and $\boldsymbol{\xi}_i$ from the training data set, it consistently outputs the corresponding pixel color C_i.

Since the dimensionality $d_{\boldsymbol{\theta}}$ is very high – radiance fields routinely involve millions of parameters –, it is essential to take special care to maintain tractability. We employ a variational-inference framework [Blei et al. 2017] and choose to approximate $p(\boldsymbol{\theta})$ using a continuous *uniform* distribution. In essence, this approximation models a compact hyper-volume $V \subset \mathbb{R}^{d_{\boldsymbol{\theta}}}$ within parameter space with a constant probability density:

$$p(\boldsymbol{\theta}) = \begin{cases} \frac{1}{|V|} & \boldsymbol{\theta} \in V, \\ 0 & \text{else.} \end{cases} \tag{3}$$

As the ground-truth distribution of parameters is generally unknown, we argue that choosing a uniform prior is a reasonable choice. Although a Gaussian prior is commonly favored due to its mathematical convenience [Pan et al. 2022], we do not use a Gaussian prior because it assumes a higher probability around the mean. A counter-example is a single-view case, where Gaussians can move along the depth. Their probability distribution along depth should not be symmetric and centered around any particular depth value. Also, the color of occluded Gaussians should be equally likely to be of any value. In addition, a uniform prior avoids sampling unbounded values such as Gaussian scales. Central to our approach is a novel method for constructing the hyper-volume V, as described in Sec. 4.2.

Using a uniform approximation of $p(\boldsymbol{\theta})$, our training loss for finding an optimal V can be formulated as follows:

$$\mathcal{L} = \sum_{i=1}^{N_{\mathcal{T}}} \int_{\boldsymbol{\theta}^* \in V} \left\| C_{\boldsymbol{\theta}^*}(\mathbf{r}_i, \boldsymbol{\xi}_i) - C_i \right\|_I \mathrm{d}\boldsymbol{\theta}^* - \lambda |V|. \tag{4}$$

Here, the first term encourages that all realizations of $F_{\boldsymbol{\theta}}$ match the training data, while the second term encourages a large hyper-volume. The scalar λ balances the two terms.

Given a trained radiance field, our main application of next-best-view planning requires an uncertainty estimate U per image $I(\boldsymbol{\xi})$. Similar to our training objective in Eq. 4, this can be formalized as an integral over rendered realizations of $F_{\boldsymbol{\theta}}$:

$$U(I(\boldsymbol{\xi})) = \sum_{\mathbf{r} \in I} \int_{\boldsymbol{\theta}^* \in V} \left\| C_{\boldsymbol{\theta}^*}(\mathbf{r}, \boldsymbol{\xi}) - \bar{C}(\mathbf{r}, \boldsymbol{\xi}) \right\|^2 \mathrm{d}\boldsymbol{\theta}^*, \tag{5}$$

where

$$\bar{C}(\mathbf{r}, \boldsymbol{\xi}) = \int_{\boldsymbol{\theta}^* \in V} C_{\boldsymbol{\theta}^*}(\mathbf{r}, \boldsymbol{\xi}) \mathrm{d}\boldsymbol{\theta}^* \tag{6}$$

is the mean pixel color across realizations. A high uncertainty U for a particular $I(\boldsymbol{\xi})$ indicates a good candidate for subsequent capture.

Clearly, the high-dimensional integrals in Eqs. 4-6 over radiance field realizations cannot be evaluated analytically due to the non-linearity of the rendering function. Instead, we opt for a numerical integration scheme. The most general approach that, in principle, scales to high dimensions is Monte Carlo integration, which employs a random sampling of the integration domain. A Monte Carlo estimator of Eq. 4 is given by

$$\mathcal{L}_{\mathrm{MC}} = \sum_{i=1}^{N_{\mathcal{T}}} \frac{1}{M} \sum_{\boldsymbol{\theta}^* \sim V} \left\| C_{\boldsymbol{\theta}^*}(\mathbf{r}_i, \boldsymbol{\xi}_i) - C_i \right\| - \lambda |V|, \tag{7}$$

where $\boldsymbol{\theta}^*$ are now M random samples. Note how Eq. 7 optimizes over the space of generative models that allow to draw samples from and thereby implicitly define V. An analogous estimator for Eq. 5 is

$$U_{\mathrm{MC}}(I(\boldsymbol{\xi})) = \sum_{\mathbf{r} \in I} \frac{1}{M} \sum_{\boldsymbol{\theta}^* \sim V} \left\| C_{\boldsymbol{\theta}^*}(\mathbf{r}, \boldsymbol{\xi}) - \bar{C}(\mathbf{r}, \boldsymbol{\xi}) \right\|^2. \tag{8}$$

The integral in Eq. 6 can be estimated correspondingly.

The benefit of learning $F_{\boldsymbol{\theta}}$ per Eq. 7 and estimating its uncertainty per Eq. 8 is that is only requires a sum of first-order differentiable rendered radiance fields, i.e., different from other approaches [Goli et al. 2024; Jiang et al. 2024], our estimates are derivative-free. This is particularly beneficial in the uncertainty estimate of Eq. 8, as it easily allows to differentiate U_{MC} with respect to camera parameters and/or $\boldsymbol{\xi}$ to perform gradient-based optimization of next-best views, as described in Sec. 4.4. However, regrettably, Monte Carlo estimators typically exhibit excessive variance and require a high number M of samples to provide stable results, in particular in high dimensions. The key technical contribution of this work is the observation that the probability volume V can be modeled as a low-dimensional manifold. This allows stable and high-quality training per Eq. 7 and uncertainty estimation per Eq. 8 using a very low number of samples, as described next.

4.2 Manifold Sampling

We seek to devise a generative model that allows us to sample from a compact radiance-field-specific volume $V \subset \mathbb{R}^{d_\theta}$ of constant probability density that captures the distribution of trainable parameters $\boldsymbol{\theta}$. Ideally, we would aim for maximum expressivity by allowing V to have arbitrary shapes. Powerful generative models of radiance field parameters have been studied [Chan et al. 2021; Gu et al. 2022; Müller et al. 2023] and typically rely on deep neural networks. Unfortunately, our efficiency constraints – training a radiance field needs to be quick while training a deep generative model on top is slow – and the extremely high dimensionality $d_{\boldsymbol{\theta}}$ of the parameter space, make such solutions impractical. Therefore, we impose a more rigid structure.

Towards our solution, we consider a linear model G of the form

$$G(\mathbf{z}) = \bar{\boldsymbol{\theta}} + B\mathbf{z}, \tag{9}$$

where $\bar{\boldsymbol{\theta}} \in \mathbb{R}^{d_\theta}$ is the mean of the distribution, and $\mathbf{z} \sim \mathcal{U}([-1,1]^{d_\theta})$, is uniform-randomly sampled. The generating matrix $B \in \mathbb{R}^{d_\theta \times d_\theta}$ encodes a linear relationship between parameter dimensions and defines the shape of V as a parallelotope (Fig. 3a, top). The covariance matrix of the emerging distribution of parameter samples is given by $\Sigma = BB^T \in \mathbb{R}^{d_\theta \times d_\theta}$ (Fig. 3a, bottom). Trivially, a uniform sampling of $\mathbf{z}$ from the hypercube results in a uniform sampling of V. The most challenging aspect of Eq. 9 lies in representing the generating matrix B, which, in general, involves a number of entries that grows with the *square* of the number of trainable parameters – a completely intractable quantity.

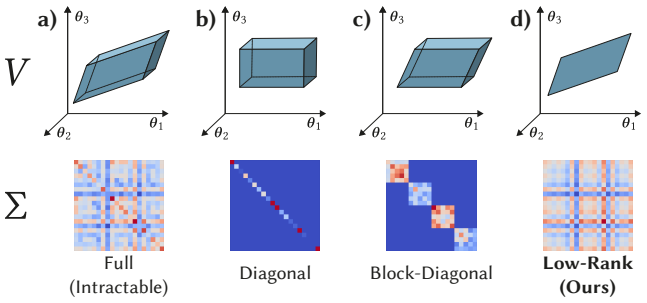

Fig. 3. Different variants to model an uncertainty volume V in the space of radiance field parameters (top row, only three out of millions of parameters are shown) using different covariance matrices Σ (bottom row, 20 dimensions are shown). *(a)* A full Σ is the most expressive solution that leads to an arbitrarily shaped parallelotope, but it suffers from an intractable number of parameters. *(b)* Restricting Σ to a diagonal matrix is a sparse solution, but it can only represent axis-aligned hyper-rectangles. *(c)* A block-diagonal Σ is slightly more expressive, but it requires making representation-specific independence assumptions and small blocks to stay tractable. *(d)* Our solution employs a low-rank covariance matrix, which results in a manifold parallelotope (here a 2D parallelogram). This parameterization is highly efficient to train and results in expressive uncertainty estimates.

A commonly used solution to avoid this quadratic complexity is to presume *independence* among the components of $\boldsymbol{\theta}$ [Goli et al. 2024; Jiang et al. 2024; Pan et al. 2022; Shen et al. 2021]. In this case, B and, consequently, Σ, reduce to a diagonal matrix, and V transforms into an axis-aligned hyper-rectangle (Fig. 3b). Although this strategy notably diminishes complexity, we observe a disproportionate compromise in expressivity and, thus, accuracy. A different, slightly less invasive strategy would be to presume independence between *groups* of variables. In such a model, B and, consequently, Σ, become block-diagonal and parameterize a parallelotope with a restricted distribution of admissible slopes (Fig. 3c). While this solution is strictly more expressive than using a diagonal matrix, we observe that it still suffers from reduced predictive accuracy. Further, it requires making representation-specific independence assumptions between small groups of variables to retain tractability.

Our key observation is that for high-quality and efficient uncertainty quantification in radiance fields, V can be modeled as a linear k-dimensional *manifold* in $d_{\boldsymbol{\theta}}$-dimensional parameter space, where $k \ll d_{\boldsymbol{\theta}}$. Intuitively, the uncertainty volume of radiance fields exhibits significantly fewer degrees of freedom compared to the number of parameters defining the radiance field itself. Concretely, the variation of radiance field parameters $\boldsymbol{\theta}$ can be well explained by a k-dimensional parallelotope (Fig. 3d, top). This translates into a model

$$G(\hat{\mathbf{z}}) = \bar{\boldsymbol{\theta}} + \hat{B}\hat{\mathbf{z}}, \tag{10}$$

where $\hat{B} \in \mathbb{R}^{d_\theta \times k}$ is the generating matrix of the manifold parallelotope, and $\hat{\mathbf{z}} \in \mathbb{R}^k$ are low-dimensional random samples. Note how this construction results in a low-rank covariance matrix $\hat{\Sigma} = \hat{B}\hat{B}^T$ (Fig. 3d, bottom). Surprisingly, we observe that setting $k = 2$ is sufficient for high-quality training and uncertainty estimation in practice, while also enabling fast training.

The advantages of this formulation are three-fold: First, the low-rank approximation results in a compact representation, as we only require $(k+1) \times d_{\boldsymbol{\theta}}$ parameters to represent our full model. Second, Eq. 10 is very efficient to compute. Third, the low effective dimensionality of V makes the Monte Carlo estimators in Eq. 7 and Eq. 8 highly effective and practically noise-free. We provide details on training and uncertainty estimation of our approach in Sec. 4.3 and Sec. 4.4, respectively.

4.3 Training

To train our stochastic radiance field, we incorporate Eq. 10 into Eq. 7 to arrive at our final training objective, which optimizes over the parameters $\bar{\boldsymbol{\theta}}$ and $\hat{B}$ of our generative model:

$$\mathcal{L}_{\text{manif}} = \sum_{i=1}^{N_{\mathcal{T}}} \frac{1}{M} \sum_{\hat{\mathbf{z}} \sim \mathcal{U}([-1,1]^k)} \left\| C_{G(\hat{\mathbf{z}})}(\mathbf{r}_i, \boldsymbol{\xi}_i) - C_i \right\| - \lambda \|\hat{B}\|_1. \tag{11}$$

In the data term, we use the generator G to sample from the parameter manifold. In practice, we use a low-discrepancy Sobol [1967] sequence to sample $\hat{\mathbf{z}}$ for improved stability. We follow Kerbl et al. [2023] and measure pixel differences using a sum of ℓ_1 distance and the SSIM [Wang et al. 2004] metric, while also employing their heuristic densification strategies.

The second term in Eq. 11 is an estimator of the volume of V which we seek to maximize. While the volume of an ordinary parallelotope is easily computed using the determinant of its generating matrix, the volume of a manifold of the shape of a parallelotope in a higher-dimensional ambient space is not straightforward. As a substitute, we employ the entrywise L1 norm of $\hat{B}$, promoting longer sides of the parallelotope, which we found provides higher-quality results

than all alternatives we tested. We choose the L1 norm over L2 because it provides a constant gradient, preventing both gradient vanishing when $\|\hat{B}\|_1$ is small and gradient explosion when $\|\hat{B}\|_1$ is large.

We observe that, for optimal quality, special attention must be given to the parameterization of $\hat{B}$. A naïve parameterization, i.e., an unconstrained optimization of the entries of $\hat{B}$, is prone to producing degenerate results where the columns of $\hat{B}$ are nearly linearly dependent. As a result, the manifold volume can have a lower effective dimensionality than k. In our solution, we prevent such outcomes using a simple strategy: The individual entries of $\hat{B}$ are initialized to low positive values, with a subsequent application of a ReLU activation function. The result is multiplied with a random but fixed sign, i.e., +1 or −1, per matrix entry. This strategy effectively diversifies the directions of the column vectors of $\hat{B}$, preventing degenerate solutions in practice.

Thanks to our manifold sampling, we find that a *single* Monte Carlo sample per training iteration, i.e., $M = 1$, is sufficient for fast and stable convergence. Notably, we observe that our approach does not increase the number of training iterations required to reach a converged result. Consequently, *our formulation imposes only minor extra cost compared to standard radiance field training.*

Except for our loss in Eq. 11 and a $(k + 1)$-fold increase in the number of trainable parameters, the training process follows the standard approach described in Kerbl et al. [2023]. We find that the volume maximization term in Eq. 11 only needs to be applied sporadically. In practice, we set $\lambda = 1$ every ten training iterations and $\lambda = 0$ for all other iterations.

4.4 Differentiable Uncertainty Optimization

To obtain uncertainty estimates per view, we incorporate Eq. 10 into Eq. 8, yielding

$$U_{\text{manif}}(I(\boldsymbol{\xi})) = \sum_{\mathbf{r}\in I} \frac{1}{M} \sum_{\hat{\mathbf{z}}\sim\mathcal{U}([-1,1]^k)} \left\| C_{G(\hat{\mathbf{z}})}(\mathbf{r}, \boldsymbol{\xi}) - \bar{C}(\mathbf{r}, \boldsymbol{\xi}) \right\|^2, \quad (12)$$

with

$$\bar{C}(\mathbf{r}, \boldsymbol{\xi}) = \frac{1}{M} \sum_{\hat{\mathbf{z}}\sim\mathcal{U}([-1,1]^k)} C_{G(\hat{\mathbf{z}})}(\mathbf{r}, \boldsymbol{\xi}). \quad (13)$$

Similar to the training loss in Eq. 11, we use a Sobol [1967] sequence for sampling $\hat{\mathbf{z}}$ in the unit hypercube, but found $M = 2$ necessary for obtaining stable results.

Importantly, Eq. 12 is trivially differentiable with respect to the camera position (via the rays $\mathbf{r}$) and the auxiliary parameters $\boldsymbol{\xi}$. This property allows us to conduct gradient-based optimization with respect to these quantities for finding next-best views, for which we use the Adam [Kingma and Ba 2015] optimizer with default parameters.

5 EXPERIMENTS

In this section, we evaluate our manifold sampling approach for uncertainty estimation, both quantitatively and qualitatively. Since ground-truth uncertainty is unknown, we evaluate performance on two distinct tasks: First, we consider active camera planning (Sec. 5.1), where our objective is to optimize a sequence of cameras to effectively reduce reconstruction uncertainty in ordinary radiance fields. Second, we explore the task of active illumination planning (Sec. 5.2). Here, our aim is to reconstruct a relightable radiance field, and we seek to optimize a sequence of illumination conditions to be utilized for training. We evaluate our uncertainty quantification in Sec. 5.3. Finally, we provide an analysis of several further aspects of our approach in Sec. 5.4.

5.1 Task 1 – Active Camera Planning

Here, we apply our approach to the task of finding a scene-specific sequence of views to be used for training that optimally reduces reconstruction uncertainty in radiance fields. Our evaluation protocol is structured as follows: During the training of a radiance field, we systematically introduce one camera at a time, each added after every 2000 training iterations, starting with a single camera.

We compare different approaches for determining the next-best views. We begin with a baseline method that employs farthest-point sampling. In this approach, a camera is selected from a candidate pool based on its maximum distance from the cameras already in the training set. Then, we consider the two state-of-the-art uncertainty estimation approaches ActiveNeRF [Pan et al. 2022] and FisherRF [Jiang et al. 2024], as well as the active camera placement (ACP) approach of Kopanas and Drettakis [2023]. We adapted all of the aforementioned approaches to the 3DGS representation. Note that none of these approaches support differentiable optimization for the next-best camera; instead, they rely on selecting candidates from a predefined pool. We do not consider the most recent work on radiance field uncertainty estimation of Goli et al. [2024], as they use the same Laplace approximation method as FisherRF [Jiang et al. 2024] but with different representations: NeRF and 3DGS. To ensure fairness, we reimplemented all baselines with 3DGS and found FisherRF is identical to Bayes' Rays with 3DGS in computing radiance field parameter uncertainty. Further, Bayes' Rays does not model the analytical predictive distribution of rendered images. Finally, we consider three variants of our method. In the first variant, Ours (Sel.), we evaluate Eq. 12 for every candidate view and select the one with the highest uncertainty, restricting our method to match the capabilities of the baselines. In the second variant, Ours (Opt. Sel.), we use Eq. 12 to differentiably optimize for the next-best view, initialized from the solution of the first variant. In our final variant, Ours (Opt. Rnd.), we explore the capabilities of our optimization when the process is initialized from a random view.

We conduct our evaluation on the NeRF Synthetic [Mildenhall et al. 2020] and the Mip-NeRF360 [Barron et al. 2022] datasets. In both cases, the candidate pool consists of the full set of training images available (100 and 150, respectively). Since we cannot rely on SfM points, we initialize the 3DGS model using a random point cloud. We evaluate the variants Ours (Opt. Sel.) and Ours (Opt. Rnd.) only on the synthetic datasets, as these strategies require capturing new training images online after each next-best view optimization. For the Mip-NeRF360 dataset, we are limited to the existing training views and hence only evaluate the variant Ours (Sel.). In Tab. 1 and Tab. 2, we provide the results of a quantitative evaluation of reconstruction accuracy on a held-out set of test views. In Tab. 1, we report the mean and standard deviation over ten runs, averaged across eight scenes, to demonstrate the stability of our method

and the chosen sample count. Fig. 4 and Fig. 5 show corresponding qualitative results. Our approach outperforms the baselines across all metrics for both datasets. Candidate selection based on our uncertainty (Ours Sel.) already surpasses the state-of-the-art view synthesis quality. Further, using this selection as a starting point for fine-grained optimization (Ours Opt. Sel.) tends to enhance the quality even more. Starting with a random initialization (Ours Opt. Rnd.) yields nearly equal-quality results, indicating that our optimization method avoids getting stuck in local minima.

In Tab. 3, we compare the upper bound of novel-view synthesis quality between 3DGS [Kerbl et al. 2023] and our method on the NeRF Synthetic dataset, using all training views over 7000 iterations. Our results, achieved through a stochastic training scheme, demonstrate comparable quality to the original 3DGS method.

5.2 Task 2 – Active Illumination Planning

As a second application domain of our approach, we choose the task of training a relightable radiance field [Martin-Brualla et al. 2021; Srinivasan et al. 2021] using controlled illumination. Here, we focus on a scenario where scenes are captured from multiple viewpoints under various illumination conditions, which can be achieved, for instance, through a light stage setup [Debevec et al. 2000]. Rather than optimizing for emitted radiance using a set of m spherical harmonics (SH) coefficients per primitive, as commonly done, we take inspiration from precomputed radiance transfer techniques [Lyu et al. 2022; Saito et al. 2024; Sloan et al. 2002]. In our method, we optimize for a radiance transfer matrix $T \in \mathbb{R}^{m \times n}$ for each primitive in the scene. Distant illumination is parameterized by an SH coefficient vector $\boldsymbol{\xi} \in \mathbb{R}^n$. This vector is multiplied with each transfer matrix T to compute the outgoing radiance per Gaussian that needs to be rendered. This setup allows us to efficiently represent and manipulate illumination conditions for relightable rendering. In our experiments, we set $m = n = 16$, covering the first four SH bands.

Analogous to the setup in Sec. 5.1, we start with a single illumination condition – we choose an angularly uniform illumination, i.e., only the DC coefficient is active – and gradually add new illumination samples every 7000 training iterations. The samples are obtained by differentiably optimizing uncertainty per Eq. 12 with respect to the illumination condition $\boldsymbol{\xi}$. The optimization is initialized from the one-hot parameter vector $\boldsymbol{\xi}$ that results in the highest uncertainty. Note how our gradient-based approach avoids excessive enumeration of candidates in the 16-dimensional parameter space.

In Fig. 6 and Fig. 7 we compare our approach against random and Sobol [1967] sampling of next illumination conditions using three scenes from the NeRF Synthetic dataset. We see that our uncertainty-guided approach significantly outperforms the scene-agnostic baselines.

5.3 Uncertainty Quantification

We evaluate pixel-wise predicted uncertainty in the sparse-view setting by analyzing its correlation with depth estimation error, which serves as an indicator of underlying geometric uncertainty. Specifically, we follow Shen et al. [2022] and use the Area Under Sparsification Error (AUSE) metric on the LF dataset [Yücer et al. 2016]. The results of this analysis are presented in Tab. 4, where we compare our approach against CFNeRF [Shen et al. 2022], Bayes' Rays [Goli et al. 2024], and FisherRF [Jiang et al. 2024]. The results for CFNeRF and Bayes' Rays are taken directly from Goli et al. [2024].

Our method can only be fairly compared against FisherRF, as it is the only baseline using the 3DGS representation. The potential superiority of Bayes' Rays in AUSE may be due to limitations of 3DGS with forward-facing scenes and NeRF's inherent ability to ensure smoothness in the 3D uncertainty field, which leads to smoother depth errors. To enable a fairer comparison, future work will extend our method to NeRF-based architectures. Our stochastic radiance field and training methodology are also readily applicable to discrete volume representations, such as Plenoxels [Fridovich-Keil et al. 2022]. Additionally, future work will explore extending manifold sampling to general continuous neural representations.

5.4 Analysis

Here, we analyze various aspects of our approach. First, we examine different methods for modeling parameter covariance as shown in Tab. 5. Since a naïve parameterization of the full $d_{\boldsymbol{\theta}} \times d_{\boldsymbol{\theta}}$ covariance matrix is generally intractable (it would require petabytes of memory for a typical scene), we explore the tractable alternatives listed in Fig. 3, including our approach with different ranks k. We consider the task of active camera planning (Sec. 5.1) and measure novel-view synthesis quality on the NeRF Synthetic dataset. Additionally, we report the number of milliseconds required for one training iteration; the total number of training iterations required to reach a converged radiance field is roughly the same across methods.

We see that low-rank approximations consistently outperform the alternatives. Surprisingly, the rank has only a minor impact on result quality, but higher ranks increase training time. Our choice of $k = 2$ strikes a reasonable balance between quality and speed. Specifically, our solution is only 14% slower to train than a vanilla 3DGS model without uncertainty (w/o U), on top of which a Laplace approximation might be run. Asymptotic time complexity can be computed by analyzing the sampling in Eq. 10 and the rasterization of Gaussians. Eq. 10 leads to complexity $O(MNk)$ while rasterization is $O(MN(p + \log N))$, where M is the number of samples, N is the number of Gaussians, k is the rank, and p is the number of pixels.

In Fig. 8, we show representative examples of uncertainty landscapes generated by different methods when a single training view (red dot) is available. FisherRF provides a smooth landscape, but optimizing camera positions with this method is impractical due to the need for higher-order derivatives. ACP's energy landscape is fairly uniform, as it aims to prevent view clustering, which is not a particularly useful measure in a sparse-view setting. Again, this energy cannot be differentiated w.r.t. camera position. Stochastic sampling with a diagonal or block-diagonal covariance matrix introduces excessive noise, whereas our low-rank solutions are smooth by construction. To gain further insights into the optimization dynamics, we start from a random initialization (yellow dot) and differentiably optimize for a view that maximizes uncertainty (green dot). With diagonal and block-diagonal covariance matrices, this optimization

Table 1. Numerical evaluation for novel-view synthesis with **active camera selection** on the **NeRF Synthetic dataset** [Mildenhall et al. 2020].

Method	5 cameras			10 cameras		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Farthest Point	21.91 ± 0.03	0.836 ± 0.001	0.139 ± 0.001	25.77 ± 0.05	0.899 ± 0.001	0.087 ± 0.001
ActiveNeRF	21.87 ± 0.48	0.831 ± 0.010	0.144 ± 0.010	25.59 ± 0.73	0.891 ± 0.011	0.093 ± 0.009
FisherRF	21.24 ± 0.60	0.821 ± 0.013	0.151 ± 0.012	25.89 ± 0.69	0.899 ± 0.008	0.087 ± 0.007
ACP	21.09 ± 0.39	0.818 ± 0.010	0.156 ± 0.009	26.20 ± 0.36	0.902 ± 0.005	0.084 ± 0.004
Ours (Sel.)	22.38 ± 0.36	**0.841** ± 0.007	**0.136** ± 0.006	26.99 ± 0.37	0.911 ± 0.003	0.077 ± 0.003
Ours (Opt. Sel.)	**22.45** ± 0.50	**0.841** ± 0.009	0.140 ± 0.008	**27.52** ± 0.45	**0.917** ± 0.005	**0.076** ± 0.004
Ours (Opt. Rnd.)	22.02 ± 0.70	0.839 ± 0.015	0.147 ± 0.012	26.78 ± 0.52	0.908 ± 0.011	0.082 ± 0.008

Table 2. Numerical evaluation for novel-view synthesis with **active camera selection** on the **Mip-NeRF360 dataset** [Barron et al. 2022].

Method	10 cameras			20 cameras		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
ActiveNeRF	12.50	0.265	0.615	14.38	0.368	0.569
FisherRF	18.02	0.550	0.408	21.38	0.673	0.330
ACP	18.71	0.568	0.399	21.41	0.684	0.320
Ours (Sel.)	**18.87**	**0.578**	**0.388**	**22.27**	**0.696**	**0.312**

Table 3. Numerical evaluation for novel-view synthesis with **all training cameras** on the **NeRF Synthetic dataset** [Mildenhall et al. 2020].

Method	PSNR↑	SSIM↑	LPIPS↓
3DGS	32.58	0.966	0.041
Ours	32.11	0.963	0.045

Table 4. Area Under Sparsification Error (AUSE↓) on the LF dataset [Yücer et al. 2016].

Method	Africa	Basket	Statue	Torch	Avg.
CFNeRF w/ NeRF	0.35	0.31	0.46	0.97	0.52
Bayes' Rays w/ NeRF	**0.27**	**0.28**	**0.17**	**0.22**	0.23
FisherRF w/ 3DGS	0.64	0.54	0.52	0.47	0.54
Ours w/ 3DGS	**0.27**	0.44	0.47	0.44	0.40

eventually converges, but it requires many iterations due the noisy uncertainty landscape. In contrast, our manifold sampling stably converges two to three times faster, even in cases where the total trajectory is longer.

6 CONCLUSION

We proposed a method that explicitly accounts for the epistemic uncertainty in Gaussian radiance fields. We observed that uncertainty can be modeled as a low dimensional manifold in parameter space, which allows efficient Monte Carlo sampling. Importantly, our formulation is fully differentiable, which in contrast to most prior works, allows continuously optimizing scene parameters such as the next-best camera view. We demonstrated the versatility of our formulation through various experiments, with a focus on active camera and illumination planning. In the future, we plan to investigate the applicability of manifold sampling to other radiance field representations and even to different signal modalities.

Table 5. Ablation. We consider different ways to model covariance. In addition to novel-view synthesis quality based on 5 and 10 training views, we report the number of milliseconds required for one training iteration.

Method	Time	5 cameras			10 cameras		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
w/o U	10.8	—	—	—	—	—	—
Diag.	14.0	21.54	0.823	0.151	26.91	0.911	0.077
Block-D.	13.8	21.85	0.830	0.145	26.09	0.900	0.082
Rank 1	11.6	22.02	0.836	0.139	**27.09**	**0.913**	**0.076**
Rank 2	12.3	**22.19**	0.840	0.135	26.69	0.909	0.078
Rank 4	13.5	22.17	**0.842**	**0.133**	26.68	0.907	0.079
Rank 10	15.7	22.12	0.834	0.142	26.08	0.899	0.086

ACKNOWLEDGMENTS

This work was supported by the ERC Consolidator Grant 4DReply (770784).

REFERENCES

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5470–5479.

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. 2017. Variational inference: A review for statisticians. *Journal of the American statistical Association* 112, 518 (2017), 859–877.

Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. 2015. Weight uncertainty in neural network. In *International conference on machine learning*. PMLR, 1613–1622.

Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. 2022. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16123–16133.

Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. 2021. pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5799–5809.

Zhaoxi Chen, Gyeongsik Moon, Kaiwen Guo, Chen Cao, Stanislav Pidhorskyi, Tomas Simon, Rohan Joshi, Yuan Dong, Yichen Xu, Bernardo Pires, He Wen, Lucas Evans, Bo Peng, Julia Buffalini, Autumn Trimble, Kevyn McPhail, Melissa Schoeller, Shoou-I Yu, Javier Romero, Michael Zollhöfer, Yaser Sheikh, Ziwei Liu, and Shunsuke Saito. 2024. URHand: Universal Relightable Hands. In *CVPR*.

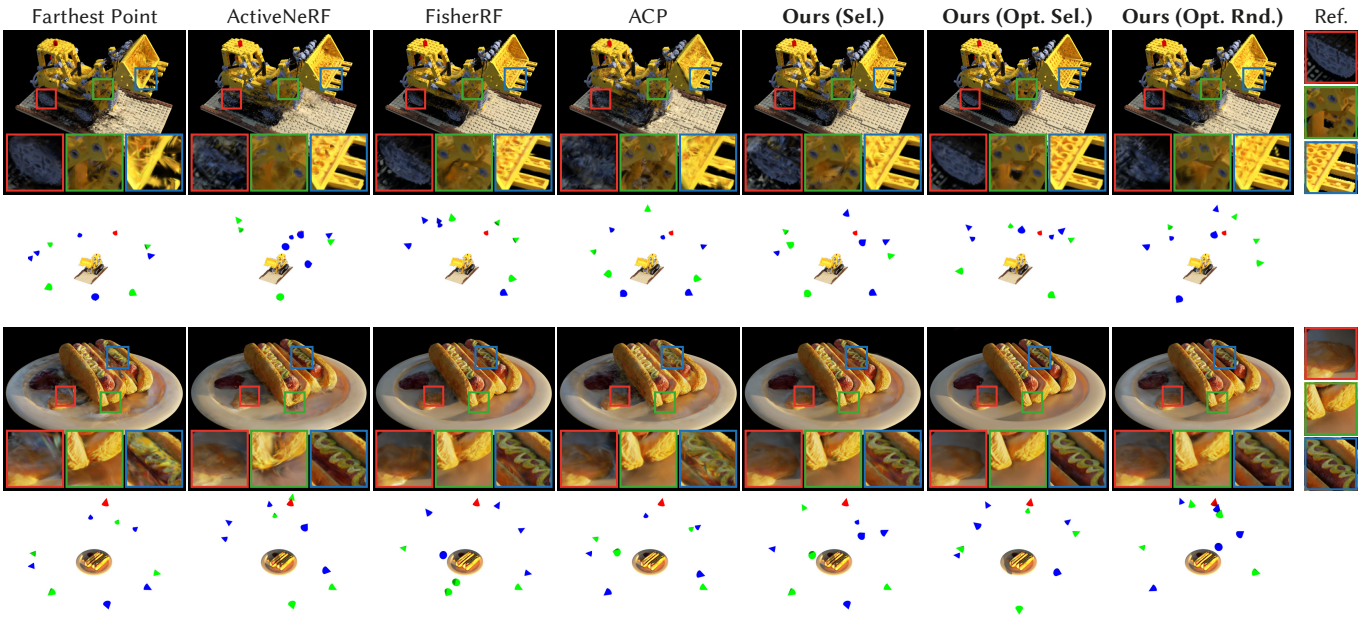

Fig. 4. Active camera planning results on the NeRF Synthetic dataset from different methods (columns) using 10 training views. The first row of each scene shows a novel view, while the second row illustrates the distribution of cameras placed by each method. Red cones represent the initial camera position, green cones represent the first four camera positions, and blue cones represent the final five camera positions.

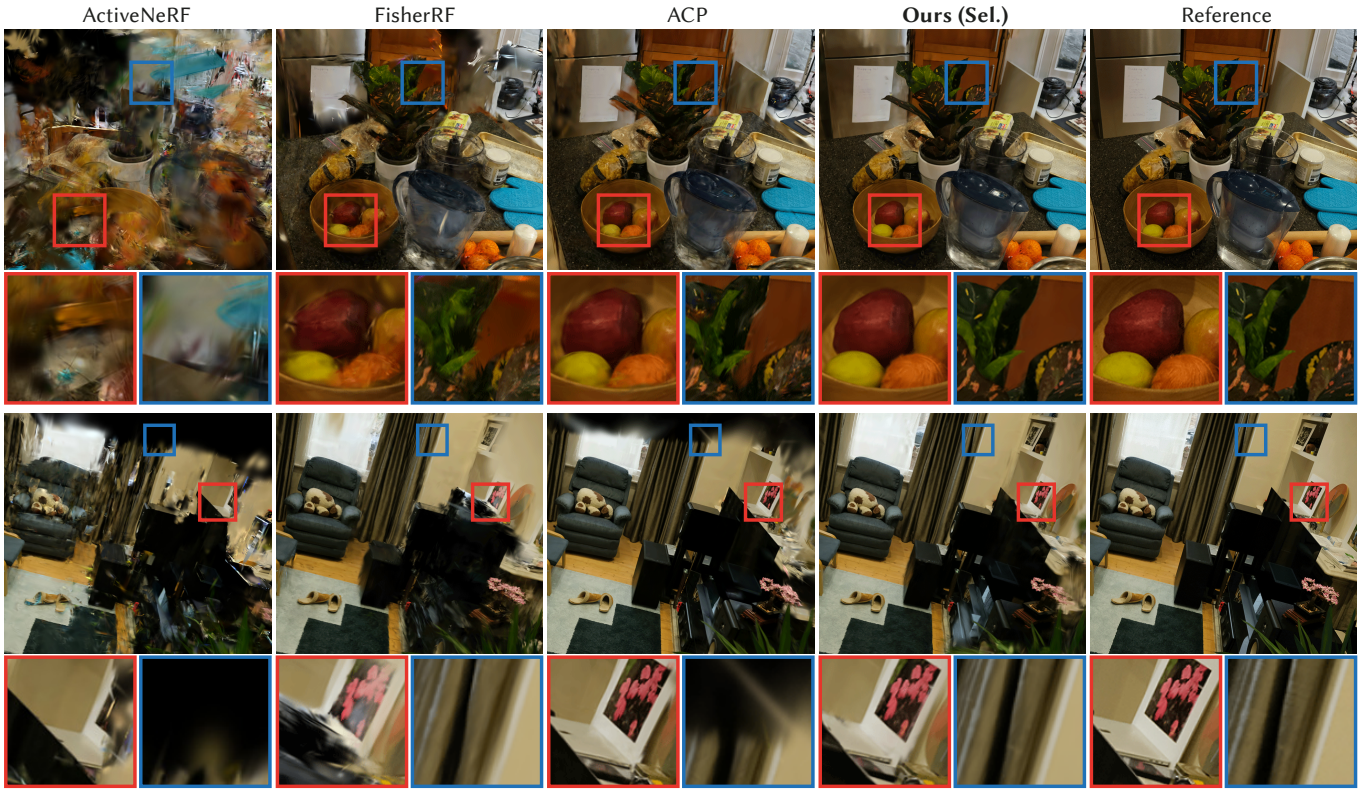

Fig. 5. Active camera planning results on the MipNeRF-360 dataset from different methods (columns) using 20 training views. The images show a novel view.

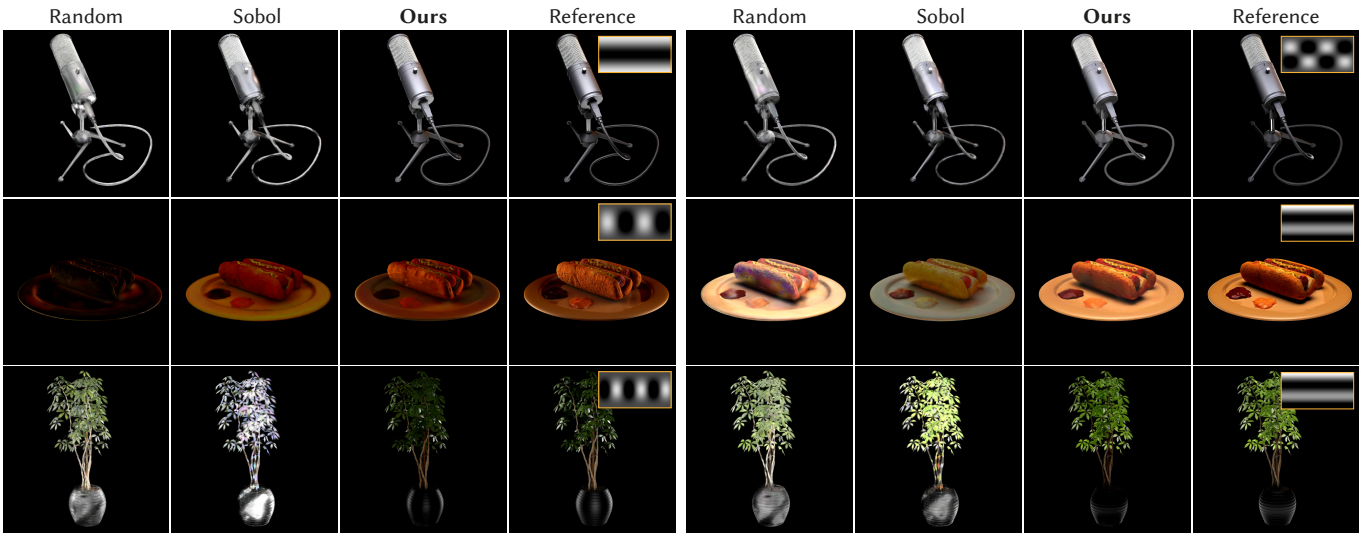

Fig. 6. Active illumination planning results. We demonstrate novel-view relighting results based on 8 training illumination conditions. We display two test illuminations per scene, shown as insets in the respective reference image.

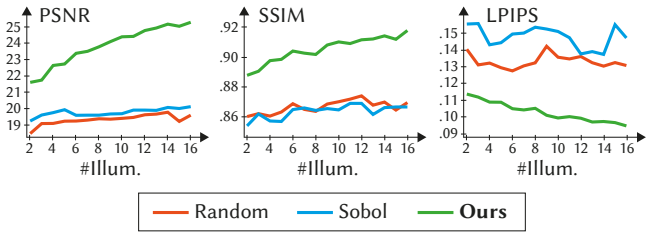

Fig. 7. Quantitative evaluation of relighting quality. We plot image quality metrics against the number of illumination conditions used for training.

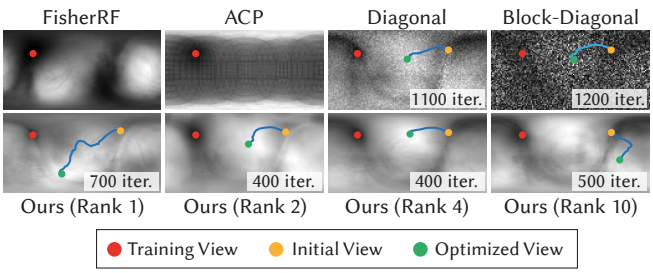

Fig. 8. Uncertainty landscapes for different methods. We display uncertainties associated with cameras positioned on a scene-enclosing sphere using a latitude-longitude parameterization. The only camera in the training pool (red dot) is surrounded by low uncertainty. Where applicable, we optimize for the second view starting from an initialization (yellow dot) to arrive at a view that maximizes the rendered uncertainty (green dot). We list the number of optimization iterations required for each approach.

Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. 2021. Laplace redux-effortless bayesian deep learning. *Advances in Neural Information Processing Systems* 34 (2021), 20089–20103.

Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 145–156.

John Denker and Yann LeCun. 1990. Transforming neural-net output levels to probability distributions. *Advances in neural information processing systems* 3 (1990).

Andrew YK Foong, Yingzhen Li, José Miguel Hernández-Lobato, and Richard E Turner. 2019. 'In-Between'Uncertainty in Bayesian Neural Networks. In *Uncertainty in Deep Learning Workshop*.

Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. 2023. K-Planes: Explicit Radiance Fields in Space, Time, and Appearance. In *CVPR*.

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5501–5510.

Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. 2024. Bayes' Rays: Uncertainty Quantification in Neural Radiance Fields. *CVPR* (2024).

Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. 2022. StyleNeRF: A Style-based 3D Aware Generator for High-resolution Image Synthesis. In *International Conference on Learning Representations*. https://openreview.net/forum?id=iUuzzTMUw9K

Matthew D Hoffman, Tuan Anh Le, Pavel Sountsov, Christopher Suter, Ben Lee, Vikash K Mansinghka, and Rif A Saurous. 2023. Probnerf: Uncertainty-aware inference of 3d shapes from 2d images. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 10425–10444.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. 2024. FisherRF: Active View Selection and Uncertainty Quantification for Radiance Fields using Fisher Information. In *ECCV*.

James T Kajiya and Brian P Von Herzen. 1984. Ray tracing volume densities. *ACM SIGGRAPH computer graphics* 18, 3 (1984), 165–174.

Alex Kendall and Yarin Gal. 2017. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems* 30 (2017).

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023). https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

Mohammad Khan, Didrik Nielsen, Voot Tangkaratt, Wu Lin, Yarin Gal, and Akash Srivastava. 2018. Fast and scalable bayesian deep learning by weight-perturbation in adam. In *International conference on machine learning*. PMLR, 2611–2620.

Diederik Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*. San Diega, CA, USA.

Georgios Kopanas and George Drettakis. 2023. Improving NeRF Quality by Progressive Camera Placement for Free-Viewpoint Navigation. In *Vision, Modeling, and Visualization*, Michael Guthe and Thorsten Grosch (Eds.). The Eurographics Association. https://doi.org/10.2312/vmv.20231222

Agustinus Kristiadi, Matthias Hein, and Philipp Hennig. 2020. Being bayesian, even just a bit, fixes overconfidence in relu networks. In *International conference on machine learning*. PMLR, 5436–5446.

Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, et al. 2024. Wonder3d: Single image to 3d using cross-domain diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9970–9980.

Linjie Lyu, Ayush Tewari, Thomas Leimkuehler, Marc Habermann, and Christian Theobalt. 2022. Neural Radiance Transfer Fields for Relightable Novel-view Synthesis with Global Illumination. In *ECCV*.

David MacKay. 1991. Bayesian model comparison and backprop nets. *Advances in neural information processing systems* 4 (1991).

Alexander Mai, Dor Verbin, Falko Kuester, and Sara Fridovich-Keil. 2023. Neural Microfacet Fields for Inverse Rendering.

Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. 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*. 7210–7219.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.

Norman Müller, Yawar Siddiqui, Lorenzo Porzi, Samuel Rota Bulo, Peter Kontschieder, and Matthias Nießner. 2023. Diffrf: Rendering-guided 3d radiance field diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4328–4338.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)* 41, 4 (2022), 1–15.

Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. 2022. Activenerf: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision*. Springer, 230–246.

Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. 2021. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5865–5874.

Hippolyt Ritter, Aleksandar Botev, and David Barber. 2018. A scalable laplace approximation for neural networks. In *6th international conference on learning representations, ICLR 2018-conference track proceedings*, Vol. 6. International Conference on Representation Learning.

Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. 2024. Relightable Gaussian Codec Avatars. In *CVPR*.

Kyle Sargent, Zizhang Li, Tanmay Shah, Charles Herrmann, Hong-Xing Yu, Yunzhi Zhang, Eric Ryan Chan, Dmitry Lagun, Li Fei-Fei, Deqing Sun, and Jiajun Wu. 2023. ZeroNVS: Zero-Shot 360-Degree View Synthesis from a Single Real Image. *CVPR, 2024* (2023).

Luca Savant, Diego Valsesia, and Enrico Magli. 2024. Modeling uncertainty for Gaussian Splatting. *arXiv preprint arXiv:2403.18476* (2024).

Johannes Lutz Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited.

Jianxiong Shen, Antonio Agudo, Francesc Moreno-Noguer, and Adria Ruiz. 2022. Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In *European Conference on Computer Vision*. Springer, 540–557.

Jianxiong Shen, Adria Ruiz, Antonio Agudo, and Francesc Moreno-Noguer. 2021. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In *2021 International Conference on 3D Vision (3DV)*. IEEE, 972–981.

Peter-Pike Sloan, Jan Kautz, and John Snyder. 2002. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Trans. Graph.* 21, 3 (jul 2002), 527–536. https://doi.org/10.1145/566654.566612

Ilya Meerovich Sobol. 1967. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* 7, 4 (1967), 784–802.

Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. 2021. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In *CVPR*.

Niko Sünderhauf, Jad Abou-Chakra, and Dimity Miller. 2023. Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 9370–9376.

Albert Tarantola. 2005. *Inverse problem theory and methods for model parameter estimation*. SIAM.

Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Wang Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. 2022. Advances in Neural Rendering. *Computer Graphics Forum* 41, 2 (2022), 703–735.

Ayush Tewari, Tianwei Yin, George Cazenavette, Semon Rezchikov, Josh Tenenbaum, Frédo Durand, Bill Freeman, and Vincent Sitzmann. 2024. Diffusion with forward models: Solving stochastic inverse problems without direct supervision. *Advances in Neural Information Processing Systems* 36 (2024).

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

Wenqi Xian, Jia-Bin Huang, Johannes Kopf, and Changil Kim. 2021. Space-time neural irradiance fields for free-viewpoint video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9421–9431.

Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. 2022. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5438–5448.

Yinghao Xu, Hao Tan, Fujun Luan, Sai Bi, Peng Wang, Jiahao Li, Zifan Shi, Kalyan Sunkavalli, Gordon Wetzstein, Zexiang Xu, et al. 2023. Dmv3d: Denoising multi-view diffusion using 3d large reconstruction model. *arXiv preprint arXiv:2311.09217* (2023).

Zike Yan, Haoxiang Yang, and Hongbin Zha. 2023. Active Neural Mapping. In *Intl. Conf. on Computer Vision (ICCV)*.

Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. 2021. Plenoctrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5752–5761.

Kaan Yücer, Alexander Sorkine-Hornung, Oliver Wang, and Olga Sorkine-Hornung. 2016. Efficient 3D object segmentation from densely sampled light fields with applications to 3D reconstruction. *ACM Transactions on Graphics (TOG)* 35, 3 (2016), 1–15.

Guodong Zhang, Shengyang Sun, David Duvenaud, and Roger Grosse. 2018. Noisy natural gradient as variational inference. In *International conference on machine learning*. PMLR, 5852–5861.

Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. 2021a. Neural light transport for relighting and view synthesis. *ACM Transactions on Graphics (TOG)* 40, 1 (2021), 1–17.

Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021b. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)* 40, 6 (2021), 1–18.

Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. 2001. EWA volume splatting. In *Proceedings of the conference on Visualization'01*. 29–36.