Uncertainty-Informed Volume Visualization using Implicit Neural Representation

Shanu Saklani* Chitwan Goel Shrey Bansal Zhe Wang Soumya Dutta†
IIT Kanpur, India IIT Kanpur, India IIT Kanpur, India Oak Ridge National Lab., USA IIT Kanpur, India

Tushar M. Athawale David Pugmire Christopher R. Johnson
Oak Ridge National Lab., USA Oak Ridge National Lab., USA University of Utah, USA

ABSTRACT

The increasing adoption of Deep Neural Networks (DNNs) has led to their application in many challenging scientific visualization tasks. While advanced DNNs offer impressive generalization capabilities, understanding factors such as model prediction quality, robustness, and uncertainty is crucial. These insights can enable domain scientists to make informed decisions about their data. However, DNNs inherently lack ability to estimate prediction uncertainty, necessitating new research to construct robust uncertainty-aware visualization techniques tailored for various visualization tasks. In this work, we propose uncertainty-aware implicit neural representations to model scalar field data sets effectively and comprehensively study the efficacy and benefits of estimated uncertainty information for volume visualization tasks. We evaluate the effectiveness of two principled deep uncertainty estimation techniques: (1) Deep Ensemble and (2) Monte Carlo Dropout (MC-Dropout). These techniques enable uncertainty-informed volume visualization in scalar field data sets. Our extensive exploration across multiple data sets demonstrates that uncertainty-aware models produce informative volume visualization results. Moreover, integrating prediction uncertainty enhances the trustworthiness of our DNN model, making it suitable for robustly analyzing and visualizing real-world scientific volumetric data sets.

Index Terms: Deep Learning, Uncertainty Quantification, Volume Visualization, Scalar Field Data, Visualization.

1 INTRODUCTION

The scientific visualization community has witnessed myriad applications of deep learning owing to the rapid growth of deep neural network (DNN) research [57]. Among many applications, analysis, visualization, and representation of volumetric data using DNNs have emerged as a promising research domain. DNNs have been used effectively to (1) synthesize volume rendered images given rendering and view parameters [6], (2) generate high-quality super-resolution images [11], (3) perform interactive volume visualization [28], (4) build differentiable rendering models to perform automatic viewpoint and transfer function optimization [62], (5) learn adaptive volume sampling strategies to produce accurate images [61], and (6) learn compressed volume representations [36, 60]. While the above DNN-based approaches facilitate multifaceted exploration of volumetric data, none of the approaches study the impact of prediction uncertainty associated with such models. Model prediction uncertainty, if reliably estimated and communicated to the domain experts, can enable them to make

informed decisions about the data during analysis phase [8, 14]. As a consequence, the DNN-based volume visualization and analysis methods will become more trustworthy and reliable. However, a literature survey reveals that epistemic uncertainty-aware DNN-based volume visualization needs detailed exploration - a gap that this work attempts to fill.

Given the recent success of implicit neural representations (INRs) for compact modeling and representing volumetric data sets [21, 36, 54], in this work, we study the efficacy and applicability of uncertainty-aware INRs for volume visualization task. Since traditional DNNs do not quantify their prediction uncertainty, we augment our INRs with two different deep uncertainty quantification techniques to collect uncertainty estimates along with the predicted data values. We thoroughly analyze and visualize the estimated uncertainty to comprehend model accuracy, trustworthiness, and reliability.

Given the existing deep uncertainty estimation techniques, in this work, we prefer the methods of uncertainty estimation that can be incorporated into an existing DNN with minimal architecture modification so that the visualization community can readily adopt such models to conduct uncertainty-aware volume visualization using DNNs. To that end, we employ Deep Ensembles as our first uncertainty estimation method since Deep Ensembles often outperform other uncertainty estimation methods and produce more accurate predictions [5, 18, 29]. However, the benefits of Deep Ensembles come at the cost of significantly large training time and resource requirements since multiple DNN models need to be trained to produce an ensemble of DNNs. To mitigate this challenge, we study the accuracy and viability of a single-model-based deep uncertainty estimation technique known as Monte Carlo Dropout (MC-Dropout) [14]. Theoretically, the MCDropout is equivalent to approximate inferencing in deep Gaussian processes [10, 14], making it an attractive choice for our work.

We analyze the accuracy, effectiveness, and benefits of uncertainty-aware implicit neural representations (INRs) of volume data augmented with two deep uncertainty estimation techniques: (1) Deep Ensemble and (2) MCDropout method to carry out informative visual analysis of scientific data. Since our methods provide model uncertainty for predicted data values, we are able to quantify fine-grained pixel-wise uncertainty for the volume-rendered images for any user-provided transfer function. As the user changes the transfer function, our pixel-wise uncertainty map also gets updated according to the current transfer function configuration. We can further investigate how uncertainty impacts each color channel while computing the final pixel color using the ray casting method.

---

*e-mail: shanu@cse.iitk.ac.in
†e-mail: soumyad@cse.iitk.ac.in (Corresponding author.)

Finally, we conduct a comprehensive study of the two principled deep uncertainty estimation methods to evaluate their applicability and viability in scientific volume visualization tasks. We examine whether the error and uncertainty maps show any correlation and further advocate that when error computation is not possible due to the unavailability of ground truth data, model uncertainty is still available and can be helpful in assessing the trustworthiness of the volume rendering results. Therefore, our contributions are twofold:

1. We propose the use of uncertainty-aware implicit neural representations for informative and robust visual analysis of volume data, emphasizing the significance of visualizing prediction uncertainty to aid domain scientists in interpreting their results more reliably.

2. We thoroughly compare and contrast between two deep uncertainty estimation techniques: (1) Deep Ensemble, and (2) MCDropout, showcasing their applicability for conducting uncertainty-informed volume visualization.

2 BACKGROUND AND UNCERTAINTY IN DEEP NEURAL NETWORKS

First we summarize research works related to deep learning for scientific visualization and uncertainty visualization in the following section. Then, we briefly introduce different methods for quantifying uncertainty in deep neural networks, followed by a detailed description of two well-known methods of estimating uncertainty in DNNs used in this work.

2.1 Deep Learning for Scientific Visualization

Deep learning has found numerous applications in scientific visualization. Techniques for generating compact neural representations of scientific data are proposed by Lu et al. [36] and Weiss et al. [60]. Hong et al. [28], He et al. [26], and Berger et al. [6] study the visualization of scalar field data using volume-rendered images, and Weiss et al. [59] study volume data with isosurfaces. An adaptive sampling-guided approach is further used by Weiss et al. [61] for efficient volume data visualization using DNNs. Another prominent area of research involves generation of spatiotemporal super-resolution volumes from low-resolution data using deep learning methods [20, 19, 64, 22]. For compressing the volume data, new models are proposed for domain-knowledge-aware latent space generation for scalar data [50]. For the generation of visualization and exploration of parameter spaces for ensemble data, DNNs are also used as surrogates [26, 52, 51]. Han et al. [23] propose a variable-to-variable translation technique for scientific data. A comprehensive review of deep learning applications for scientific visualization are available in the state-of-the-art survey [57].

2.2 Uncertainty Visualization

Visualization of uncertainty continues to be an important and challenging area. Pang et al. [39] give one of the earlier summaries of uncertainty visualization techniques. Visualization methods that are enhanced with tools for uncertainty estimation are given by Brodlie et al. [13]. Non-parametric models are used by Athawale et al. [3] to improve uncertainty in volume rendering. Spatial probability distributions, defined over triangular meshes, are visualized by Potter et al. [44], preceded by a classification of uncertainty visualization techniques [12].

Uncertainty visualization techniques are also explored for iso-contouring methods. The level crossing probability of adjacent points is computed by Pöthkow et al. [42], and to calculate the level crossing probability for each cell, this method is further refined in [43]. In [2], a visual analysis of fiber uncertainty is done. Uncertainty visualization techniques are summed up in a comprehensive survey by Bonneau et al. in [8]. For image processing applications [15] and medical imaging [16], uncertainty visualization

techniques are summarized by Gillmann and colleagues. Whitaker et al. [63] use contour boxplots to examine uncertainty visualization in an ensemble of contours. The latest progress regarding uncertainty in visualization research can be found in [30].

2.3 Uncertainty in Deep Neural Networks

Despite the widespread popularity of deep learning techniques, there are significant concerns regarding their interpretability, robustness, and generalizability in real-world applications [66]. The inability of conventional DNNs to provide uncertainty estimates can undermine the practical results they offer in fields such as computer vision, natural language processing, scientific visualization, and visual analytics. Developing universal techniques to measure and quantify uncertainty in DNNs remains a challenge, as the types and sources of uncertainty vary greatly across different applications, making it difficult to generalize uncertainty estimation methods. In the following, we succinctly describe some causes of uncertainty in DNNs and various methods proposed to simulate and mitigate them.

Uncertainty in a DNN can be broadly categorized into data (aleatoric) uncertainty and model (epistemic) uncertainty [7, 34]. Data uncertainty arises from errors and noise in measurement systems. In contrast, several factors contribute to model (epistemic) uncertainty. Firstly, current DNN models simplify real-world systems to generate observations, but this simplification introduces errors and uncertainty in predictions. Secondly, many DNNs require careful adjustments, such as implementing dropout [14, 27], experimentally adjusting the model hyperparameters [65], and applying regularization [9]. Variations in these parameter settings can result in different outcomes and associated uncertainty.

2.4 Methods to Model Uncertainty in DNNs

In the following, we discuss various techniques used for modeling uncertainty in DNNs. Then, we discuss two well-known uncertainty estimation methods, MCDropout and Deep Ensembles in detail since these two methods are employed in our work.

Deterministic Methods. Uncertainty estimation capabilities can be integrated into deterministic models by training a network explicitly to quantify uncertainties, as discussed in the work on evidential neural networks [49].

Test-time Augmentation Methods. Data augmentation during testing improves model performance using adversarial examples [4, 58] which determine prediction uncertainty.

Deep Evidential Regression. In deep evidential regression, the network simultaneously learns both parameters and hyperparameters for the corresponding evidential distributions [1]. These evidential distributions are then utilized to estimate model uncertainty.

Uncertainty via Stochastic Data Centering. In this approach, the authors propose that training an ensemble of DNNs with data sets shifted by a constant bias allows for estimating model uncertainty by assessing the variability across predictions from these ensemble members. They further introduce a method to achieve similar uncertainty estimation using a single DNN [55].

Bayesian Methods. Bayesian neural networks (BNNs) utilize prior distributions on the model parameters of DNNs to quantify epistemic uncertainty [17, 34]. Training these networks typically are computatinally expensive and involves methods such as stochastic gradient MCMC [37] and variational inference [24].

2.5 Ensemble Method

Ensemble techniques work on the idea that a group of learners of comparable competence is commonly better than a single learner [47]. Ensemble methods provide an inherent way to measure predictive uncertainty by assessing the differences among various predictions [35, 40], besides just enhancing the generalization
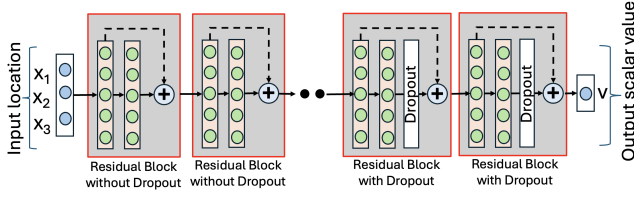
Figure 1: The schematic of the MCDropout-enabled INR model which uses a residual block-based MLP architecture. Dropout layer is added at the last two residual blocks to generate uncertainty estimates during inference time. The INR architecture for the Ensemble method is identical to this, except there are no dropout layers.

error. Therefore, the predictive uncertainty of DNNs can be calculated by taking ensemble methods into account [18, 46]. Deep Ensembles [34] are proposed by Lakshminarayanan et al., where DNNs are designed with two heads that give both the prediction and the associated uncertainty. In essence, Deep Ensemble learning can be equivalent to an approximation of Bayesian averaging [45]. In Bayesian averaging, the prediction of the final model is given as follows:

$$prediction = \int P_W(x)\,\pi(W|D) \tag{1}$$

Here, $P_W(x)$ denotes the probabilistic prediction for sample x, and $\pi(W|D)$ is the posterior probability distribution on the neural network weights, and D is the training data. In practice, estimation the above integral is very challenging, and exploration of all the modes of $\pi(W|D)$ is not needed to calculate this integral precisely [45]. This observation suggests that, from a set of ensemble members, averaging predictions without weights can be viewed as an approximation of Equation 1 [45]. It should be noted that randomizing the training data and initializing parameters randomly introduces enough variety in each learned ensemble member to effectively predict the uncertainty during the training process, a technique adopted in this work to generate the Deep Ensemble model. It is worth mentioning that Ensemble techniques often surpass the techniques that depend on probabilistic back-propagation and Monte Carlo dropout [18, 5, 38, 56], and hence Ensemble methods are often regarded as the state-of-the-art. They also show more resilience to changes in data distribution.

2.6 MCDropout Method

Dropout [14, 27] is primarily used as a regularization technique, which is applied for fine-tuning machine learning models and preventing overfitting by optimizing the adjusted loss function. Gal et al., in their seminal work [14], propose a new viewpoint on how dropout can be efficiently used as a method for approximate Bayesian inference in deep Gaussian processes. By using dropout at test time and running many forward passes with different dropout masks, the model generates a range of predictions rather than a single point estimate and is equivalent to sampling from the Bayesian posterior distribution, $P(W \mid X, Y)$, where W is the weights of the neural network model, X is training data, and Y is target output. The mean of these sampled predictions acts as the expected output of the model. Then, the epistemic uncertainty of the DNN can be conveniently estimated by calculating the standard deviation among these sampled predictions [52]. By collecting Monte Carlo (MC) samples from the network, which is dropout-enabled, such probabilistic predictions are obtained by running multiple forward passes at the time of inference, which is known as the Monte Carlo Dropout (MCDropout) method.

3 UNCERTAINTY-AWARE IMPLICIT NEURAL REPRESENTATION OF SCALAR FIELD DATA

3.1 Implicit Neural Representation

Implicit neural representations (INRs) using periodic activation functions have shown promising results for learning representations of coordinate-based data sets. Input coordinates within the data domain are mapped to their associated output values using such neural networks. By using a sinusoidal activation function in a feedforward neural network, known as SIREN (sinusoidal representation network), INRs can be effectively constructed, as proposed by Sitzmann et al. [53] in their research. Recent works have employed many variations of SIREN to address multiple complex problems in the scientific data visualization community, achieving state-of-the-art results [36, 21, 54, 48]. The achievements of these recent research efforts have inspired us to build our uncertainty-aware model by applying SIRENs as the base neural network architecture.

3.2 Model Architecture

Our objective is to learn a function that represents the mapping from the input data coordinate domain to the corresponding scalar value domain using an implicit neural network (INN). We make our base model a multilayer perceptron for this purpose. It has d input neurons (d can be 2 or 3 depending on the dimensionality of the scalar field being modeled), l hidden layers, and 1 neuron in the output layer. As proposed in [25], we strengthen the foundational SIREN architecture by combining residual blocks and skip connections to enhance the model's learning capacity and ensure steady training of the deep network. The input to our model is a d-dimensional coordinate vector, which corresponds to an output scalar value. Therefore, our implicit neural network learns a function $F(\theta): \mathbb{R}^d \to \mathbb{R}$, where θ represents the parameters of the neural network.

3.3 Uncertainty Quantification Using MCDropout Method

The architecture implemented for the MCDropout method is depicted in Fig. 1. At the last two residual blocks, a post-activation dropout layer is added to make our model dropout-enabled. This is done to conveniently calculate the prediction uncertainty during inference for the MCDropout method. Additionally, during training, dropout layers also help with regularization. In principle, a dropout layer should be incorporated for each residual block to approximate a fully Bayesian neural network [31]. However, Kendall et al. [31] show that adding dropout at each residual block or after each hidden layer can act as a strong regularizer, potentially reducing the overall prediction accuracy [31]. They further suggest that incorporating dropout to the last layer or a small subset of layers is sufficient to generate high-quality predictions along with reliable uncertainty estimates. Hence, we use dropout layer only at the last two residual blocks to approximate a partial Bayesian neural network to produce accurate predictions as well as robust uncertainty estimates. The impact of using different number of dropout layers on the model performance has been further studied in Section 5.

As discussed previously, inference using the MCDropout method involves generating a set of Monte Carlo samples by performing multiple forward passes of the dropout-enabled trained network. Hence, we generate m instances (realizations) of the scalar field and then calculate the average scalar field, serving as the predicted (expected) scalar field. Typically, the number of samples needed is decided by checking the convergence of the computed expected (averaged) field, meaning that adding more Monte Carlo samples does not improve the reconstruction quality. The uncertainty associated with each grid point is calculated by measuring the standard deviation among the m scalar values. Note that this uncertainty is measured at each grid point in the spatial domain. However, since our focus in this work is to estimate fine-grained uncertainty in the volume rendered images, we further quantify uncer-

Monte Carlo Dropout Method

| Volume realization 1 | Volume realization 2 | Volume realization 3 |

Deep Ensemble Method

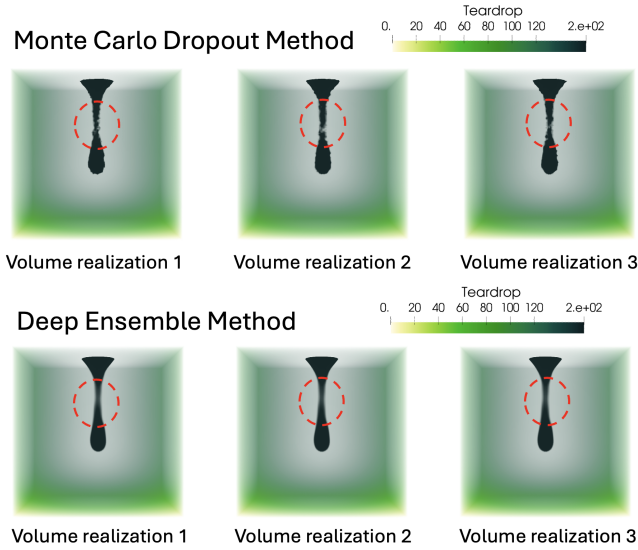| Volume realization 1 | Volume realization 2 | Volume realization 3 |

Figure 2: Volume visualization of Teardrop data set for three representative MC-sampled fields from the MCDropout method (top row) and three representative fields generated from three ensemble members (bottom row) for the Ensemble method. The ground truth is shown in Fig. 3. It is observed that individual MC sampled fields produce inaccurate visualization at the thin central segment of the Teardrop data, highlighted by red dotted circles for the MCDropout method. In contrast, the visualization produced by individual ensemble members are more accurate.

tainty in image space using the volume rendering. To estimate the image space pixel-wise uncertainty for enabling uncertainty-aware volume visualization, we compute the uncertainty of the pixel color values after collecting outputs of ray casting algorithm when applied to each individual MC volume realization (100 in our experiments). Details of this pixel-wise uncertainty calculation are provided in Section 4.1.

3.4 Uncertainty Quantification Using the Ensemble Method

The model architecture for the Ensemble model is the same as shown in Fig.1, except no dropout layers are used. Multiple instances of this model are trained to produce a Deep Ensemble model[34]. To create a robust ensemble model, we train n instances of the SIREN model. The variability across each ensemble member is ensured by randomly shuffling the order of data points at each iteration during training. Once training is completed, the averaged prediction from all n member models at each grid point produces the expected volume field. Note that the number of ensemble members needed is decided by checking the convergence of the computed expected (averaged) volume, meaning that adding more ensemble member models does not improve the reconstruction quality. Similar to the MCDropout method, we compute the data space uncertainty associated with each grid point. Subsequently, image space pixel-wise uncertainty for the Ensemble method is then computed following the same strategy as the MCDropout method, and is further discussed in Section 4.1.

3.5 Loss Function and Hyperparameters

Our uncertainty-aware SIREN model uses 10 residual blocks for MCDropout and Ensemble methods. We use 50 neurons in each hidden layer to generate consistent and comparable results while producing a compact uncertainty-aware representation of the volume data set. Conventional mean squared error loss ($\mathscr{L}_{mse}$) is used to train the network. We use empirical experimentation to identify a

Volume Visualization Results

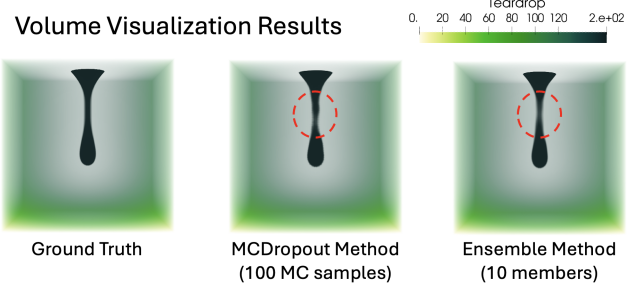| Ground Truth | MCDropout Method (100 MC samples) | Ensemble Method (10 members) |

Figure 3: Visualization of ground truth, expected (averaged) volume visualization by the MCDropout method, and expected (averaged) volume visualization by the Ensemble method of Teardrop data set. We observe that both MCDropout and Ensemble methods produce comparable and high-quality rendering results.

suitable learning rate and batch size combination that produces stable, consistent, and high-quality results across all data sets. We utilize a batch size of 2048 with the Adam optimizer [32], setting the learning rate at 0.00005 and the two Adam optimizer coefficients β_1 and β_2 to their default values at 0.9 and 0.999, respectively. Furthermore, a learning rate decay mechanism is adopted to optimize the training process, with a decay factor of 0.8 and a step size of 15. All the models were trained for 300 epochs to ensure convergence and robustness in performance evaluation. During training, we use a low dropout rate (probability) of $\eta = 0.001$ for Hurricane Isabel and Combustion data sets. This training dropout rate results in a poorly performing MCDropout model for the Teardrop data set. Hence, we use a higher dropout rate of $\eta = 0.05$, which produces a stable MCDropout model for the Teardrop data set. During inference, we use a consistent dropout rate of $\eta = 0.1$ for all the data sets to generate robust uncertainty estimates. No dropout is used for training ensemble models. This consistent approach in model design ensures a rigorous assessment of the effectiveness of uncertainty-aware deep neural networks across various data sets.

Table 1: Description of data sets that are used in the experimentation.

Data set	Dimensionality	Spatial Resolution
Teardrop	3D	$64 \times 64 \times 64$
Isabel Pressure (T=25)	3D	$250 \times 250 \times 50$
Isabel Velocity (T=25)	3D	$250 \times 250 \times 50$
Combustion Mixfrac (T=41)	3D	$240 \times 360 \times 60$

4 UNCERTAINTY-AWARE VOLUME VISUALIZATION

We perform a comprehensive study of the uncertainty-aware INRs using four volume data sets. The dimensionality and spatial resolution of these data sets are reported in Table 1. A GPU server with NVIDIA GeForce GTX 1080Ti GPUs with 12GB GPU memory is used for model training and volume reconstruction. The rendering is done on a MacBook Pro with an Apple M1 Pro chip with 10 CPU and 16 GPU cores and 16GB memory. All the models are implemented in PyTorch [41]. Teardrop data set is generated using a mathematical function [33] sampled on a $64 \times 64 \times 64$ uniform grid. Hurricane Isabel data was produced by the Weather Research and Forecast model, courtesy of NCAR and the U.S. National Science Foundation (NSF). Turbulent Combustion data set is made available by Dr. Jacqueline Chen at Sandia Laboratory through U.S. Department of Energy's SciDAC Institute for Ultrascale Visualization.

4.1 Computation of Pixel-wise Prediction Uncertainty for Volume Visualization

We use the trained INRs to reconstruct the entire volume to comprehensively assess the model's reconstruction quality, prediction
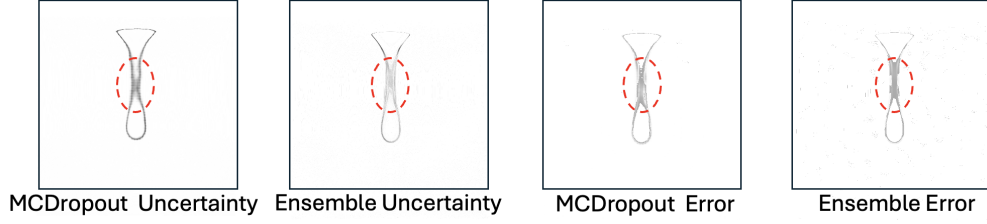
Figure 4: Prediction uncertainty and error maps of Teardrop data. Both methods produce high uncertainty and error at the thin central segment (highlighted by red dotted circles). Such uncertainty information can readily help the users to identify regions where the model is under-confident. We further observe that the teardrop's boundary also shows higher uncertainty and error, indicating that both methods are also unable to confidently predict the sharp boundary regions.

uncertainty, and error. Through empirical experimentation, we observe that 100 MC samples for the MCDropout method and 10 ensemble members for the Ensemble method allow us to produce robust estimates of volume data. A study on how reconstruction quality changes given different numbers of MC samples for the MC-Dropout method and a different number of ensemble members for the Ensemble method is provided later in Section 5. Thus, unless specified otherwise, we use 100 MC samples for MCDropout and 10 ensemble members for the Ensemble method in all experiments.

Our goal is to enable visualization and comprehension of fine-grained prediction uncertainty when model-reconstructed scalar data sets are visualized using volume rendering methods for any user-specified transfer functions. As the rendered image depends on transfer functions, the associated uncertainty map should also be updated when the transfer function changes. Both uncertainty estimation methods we use require multiple volume realizations (100 for the MCDropout method and 10 for the Ensemble method) to generate the final volume-rendered image and the associated uncertainty map. We utilize individual volume realizations to estimate the fine-grained pixel-wise uncertainty from the model-predicted results. Given a user-specified transfer function and view direction, our method applies the ray casting algorithm to each volume realization (100 for the MCDropout method and 10 for the Ensemble method) and collects the RGB pixel values for each instance. Then, the final result is obtained by averaging the RGB pixel colors, and the associated uncertainty is estimated by computing the pixel-wise standard deviation. We first estimate each color channel's uncertainty (standard deviation) from the volume rendering results. Then, we compute the final mean pixel uncertainty by averaging the individual color channel uncertainty values. The uncertainty values are then normalized and stored as grayscale images, where darker colors indicate higher uncertainty for the corresponding pixel locations. As the transfer function changes, this process is repeated so that the updated result and corresponding uncertainty map can be generated for visualization.
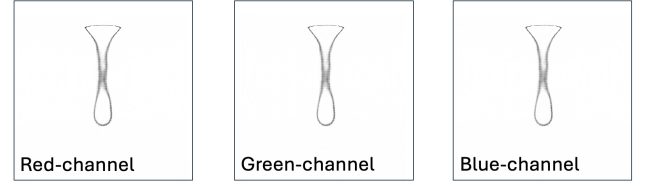
4.2 Uncertainty-Informed Volume Visualization

In the following, we qualitatively and visually study the volume visualization results and the estimated uncertainty patterns for MC-Dropout and Ensemble methods using several volume data sets. Then, in Section 5, we provide a quantitative evaluation of volume reconstruction quality and uncertainty estimates for both methods. Next, we further comprehensively evaluate the proposed methods under varying parameter configurations to assess their applicability, usefulness, and implications.

4.3 Visual Analysis for Teardrop Data

Our first case study uses the Teardrop data set [33]. To study the volume visualization results obtained from the two uncertainty estimation methods, we generate 100 Monte Carlo sample volume reconstructions for the MCDropout method and 10 sample volume

Figure 5: Prediction uncertainty visualization of Teardrop data set for individual RGB color channels. We observe that the prediction uncertainty patterns are comparable across all three color channels for the MCDropout and Ensemble methods.

reconstructions for the ensemble method (as 10 ensemble members are used). In Fig. 2, the top row shows volume rendered results of three individual MC sample volumes for the MCDropout method. Similarly, the bottom row depicts visualization generated by three separate ensemble members. We observe that the individual ensemble members renders the thin central segment region (as highlighted by red dotted circles) more accurately than the individual MCDropout sampled fields when compared against the ground truth rendering shown in Fig. 3. All the volume visualization results use fixed transfer functions, viewpoints, and all other rendering parameters to ensure fair comparison.

Next, in Fig. 3, we show the ground truth rendering, the rendering of the expected (averaged) field constructed using 100 MC sample fields for the MCDropout method and 10 ensemble member reconstructed fields for the Ensemble method, respectively. We observe that the expected field generates accurate and visually similar volume rendering results for both methods. Both MCDropout and Ensemble methods accurately preserve the thin central segment, as red dotted circles show.

In Fig.4, we present the pixel-wise prediction uncertainty and error maps for the Teardrop data set using both the MCDropout and Ensemble methods. The uncertainty represents by the pixel-wise standard deviation, averaged over all three color channels. To estimate the pixel-wise standard deviation, we perform ray casting on 100 MC sampled volumes for the MCDropout method and on 10 ensemble member generated volumes for the Ensemble method. The standard deviation for each color channel is computed from the final pixel colors and then averaged to produce the final uncertainty

image shown in Fig.4. Darker pixels indicate higher uncertainty. The image is generated by first computing pixel-wise average uncertainty values, then mapping these values to a grayscale colormap where darker colors reflect higher uncertainty. The error maps are computed by estimating the channel-wise absolute error between the ground truth pixel colors and the predicted pixel colors. Then the average error is calculated and mapped to a grayscale colormap where darker colors reflect higher error. All uncertainty and error maps are visually comparable as they are generated using a consistent grayscale colormap. Users can compare error and uncertainty between the two uncertainty-estimation methods by comparing the darkness of the pixel intensities. It is observed that both methods exhibit higher prediction uncertainty as well as higher error at the thin segment of the teardrop, indicating that the model is less confident and more erroneous in predicting values in this region. Additionally, higher uncertainty and error is noted at the boundary of the teardrop structure, where a sharp change in color gradient is observed, further indicating reduced confidence in boundary predictions. The results indicate that the uncertainty and error are largely correlated for this data set.

Prediction uncertainty maps for each individual color channel are provided in Fig. 5. The top row displays the results for the MCDropout method, while the bottom row shows the channel-wise uncertainty maps for the Ensemble method. We observe that the estimated uncertainty patterns are identical for each color channel, indicating that each color channel incurs comparable uncertainty estimates in similar spatial regions.

4.4   Visual Analysis for Isabel Pressure Field

Our next case study utilizes the Pressure field from the Hurricane Isabel data set. Both the MCDropout and Ensemble methods produce highly accurate visualizations when compared to the ground truth, as shown in Fig. 6. A close inspection of the uncertainty map generated by the MCDropout method reveals high uncertainty in the Hurricane eye feature region and moderate pixel uncertainty in the surrounding area. In contrast, the Ensemble method produces a cleaner uncertainty map, with higher uncertainty confined to only a few pixels in the land region (bottom right corner). These uncertainty maps indicate that the Ensemble method yields more robust predictions with high confidence, resulting in more reliable volume visualization images compared to the MCDropout method. When the two error maps from the two methods are investigated, it is observed that the error and uncertainty maps convey different information about the models and the MCDropout method produces fewer pixels with higher errors.

In Fig. 7, we present the channel-wise uncertainty maps for the Isabel Pressure field using both the MCDropout (top row) and Ensemble (bottom row) methods. It is observed that the Hurricane eye region exhibits high prediction uncertainty across all three color channels for the MCDropout method, with moderate uncertainty in the rest of the spatial domain. This indicates that model prediction uncertainty can affect individual color channels differently, and displaying channel-wise uncertainty maps allows for a detailed investigation of the impact of uncertainty on each color channel separately. For the Ensemble method, the overall uncertainty for all three color channels is much lower, with minimal variation in the uncertainty patterns among them.

4.5   Visual Analysis for Isabel Velocity Magnitude Field

Fig. 8 illustrates uncertainty-aware volume visualization results for the Velocity Magnitude field of the Hurricane Isabel data set. As with the Pressure field, both the MCDropout and Ensemble methods produce accurate volume visualizations. The uncertainty maps reveal a similar pattern: the MCDropout method shows moderate uncertainty over a broader area around the Hurricane eye feature, while the Ensemble method exhibits higher uncertainty concen-

trated around the eye wall of the Hurricane. The error maps from both methods display similar error patterns, with higher prediction errors concentrated around the Hurricane eye. Examining the uncertainty and error maps of the Ensemble method reveals that, while it predicts the regions around the Hurricane eye with greater confidence, it can still make errors in these areas. This indicates instances where the model is making overconfident predictions.

4.6   Visual Analysis for Combustion Mixfrac Field

Finally, we present the visualization results for the Mixture Fraction (Mixfrac) field of the Turbulent Combustion data set in Fig. 9. Both methods produce visually identical volume rendering results for the Mixfrac field when compared to the ground truth. However, a comparison of the two uncertainty maps, generated by the MCDropout and Ensemble methods, reveals that the MCDropout method produces a smoother uncertainty map, while the Ensemble method yields a relatively noisier uncertainty map. Additionally, the Ensemble method shows high uncertainty in the turbulent flame regions, whereas the MCDropout method results in lower prediction uncertainty, indicating more robust predictions for such regions. A close examination of the two error maps reveals that both methods produce similar patterns, with higher prediction errors predominantly occurring in the complex burning regions.

5   EVALUATION AND PARAMETER STUDY

5.1   Reconstruction Quality and Prediction Error in Volume Space

Table 2 presents a comparative study between the MCDropout and Ensemble methods in terms of averaged (expected) volume reconstruction quality measured using Peak signal-to-noise-ratio (PSNR). We use predictions from 100 MC samples for MCDropout and 10 ensemble members for Ensemble method to compute the final averaged field. Besides PSNR, we also estimate the Root Mean Squared Error (RMSE) between the reconstructed and ground truth fields. We observe that the reconstruction quality and RMSE between the MCDropout method and a single model trained without any dropout layer are comparable. However, the Ensemble method consistently produces the best reconstruction quality (highest PSNR) with minimum RMSE for all the data sets.

Table 2: PSNR (dB) (↑) and RMSE (↓) values for different methods (No Dropout, MCDropout with 100 samples, and Ensemble with 10 members) applied over data sets. We observe Ensemble method gets higher PSNR and lower RMSE consistently, indicating better performance in reconstruction.

Data set	Method	PSNR (dB)	RMSE
Teardrop	No Dropout	72.009	0.040
	MCDropout (100 MC samples)	71.694	0.042
	Ensemble (10 members)	76.832	0.023
Isabel Pressure	No Dropout	56.935	10.683
	MCDropout (100 MC samples)	56.922	11.116
	Ensemble (10 members)	57.068	10.542
Isabel Velocity	No Dropout	44.324	0.469
	MCDropout (100 MC samples)	44.178	0.475
	Ensemble (10 members)	44.717	0.451
Combustion Mixfrac	No Dropout	48.779	0.004
	MCDropout (100 MC samples)	48.475	0.004
	Ensemble (10 members)	49.434	0.003

5.2   PSNR Study with Varying Number of Ensemble Members

Table 3 illustrates the effect of varying the number of ensemble members on PSNR. Notably, achieving robust predictions requires fewer ensemble members compared to the MCDropout method. The data shows that increasing the number of ensemble members slowly improves the PSNR and essentially leading the PSNR value to become saturated for 10 ensemble members. Nevertheless, for
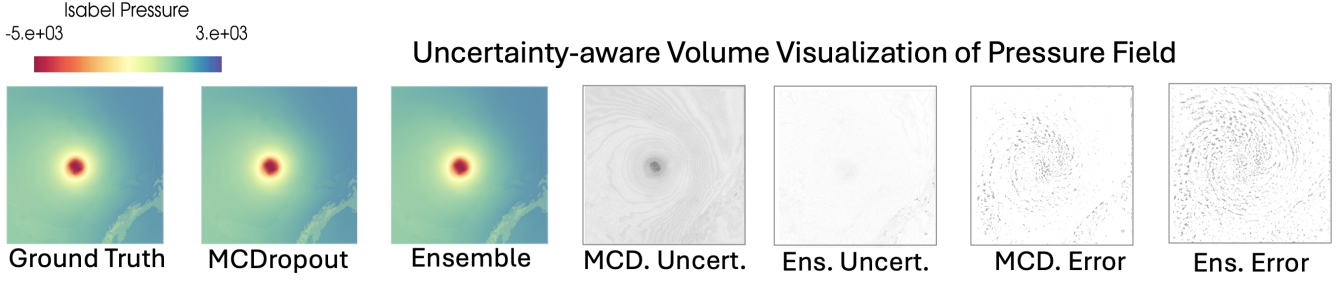
Figure 6: Volume visualization of the Pressure field of Hurricane Isabel data set. We observe that both MCDropout and Ensemble methods produce visually identical volume rendering results when compared against the ground truth. The uncertainty maps indicate higher uncertainty is produced by the MCDropout method at the Hurricane eye region, with moderate uncertainty away from the eye region. However, the Ensemble method produces very confident visualization results with only fewer pixels incurring higher uncertainty near the land region (bottom right corner). The error maps show pixel errors when the predicted image is compared against ground truth. We observe that uncertainty and error maps are not correlated.
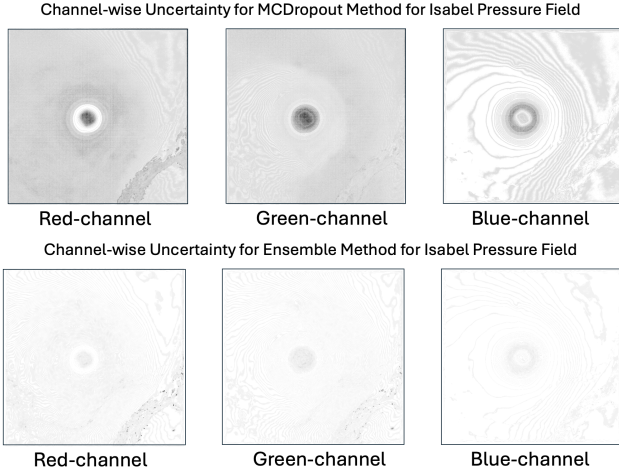

Figure 7: Prediction uncertainty visualization of the Pressure field of Hurricane Isabel data set for individual RGB color channels. We observe that the three channels exhibit different uncertainty patterns, with the higher uncertain regions being concentrated around the Hurricane eye region for the MCDropout method. The uncertainty for all three channels for the Ensemble method is comparatively lower, indicating a more robust prediction.

consistency and robustness across our experiments, we utilize 10 ensemble members.

Table 3: PSNR (dB) (↑) when the varying number of Ensemble members are used for computation in the Ensemble method.

Data set	#Ens mem = 2	#Ens mem = 5	#Ens mem = 7	#Ens mem = 10
Teardrop	76.321	76.682	76.738	76.778
Isabel Pressure	57.838	58.176	58.321	58.425
Isabel Velocity	44.812	45.174	45.275	45.372
Combustion Mixfrac	49.931	51.723	51.997	52.401

5.3 PSNR Study with Varying Number of MC Samples

Table 4 shows the PSNR values of the scalar obtained by averaging different numbers of MC samples. Increasing the number of MC samples up to 100 results in a slight PSNR improvement. Beyond this point, the PSNR gains become marginal, making it impractical to consider more samples. Therefore, to balance computation time and prediction quality, we use 100 MC samples for all experiments involving the MCDropout method.

Table 4: PSNR (dB) (↑) when the varying number of MC samples are used for computation in the MCDropout method.

Data set	#MCSamp =10	#MCSamp =25	#MCSamp =50	#MCSamp =75	#MCSamp =100
Teardrop	69.264	70.749	71.425	71.663	71.782
Isabel Pressure	54.679	55.869	56.356	56.530	56.922
Isabel Velocity	43.671	43.975	44.081	44.117	44.135
Combustion Mixfrac	47.165	47.967	48.270	48.375	48.428

5.4 Impact of Different Number of Ensemble Members on Average Pixel-wise Image space Uncertainty

Table 5 illustrates the impact of varying the number of ensemble members on the average pixel-wise uncertainty value computed in the image space using the volume rendered pixel color values. This value reflects the normalized estimated uncertainty (standard deviation) calculated from the RGB color channels using the Ensemble method. We observe that the uncertainty value gradually saturates as the number of ensemble members increases up to 10.

Table 5: Image space uncertainty values with different number of ensemble members. We observe that by increasing the number of ensemble members, the uncertainty value tends to become saturated.

Data set	#Ens mem = 2	#Ens mem = 5	#Ens mem = 7	#Ens mem = 10
Teardrop	0.018	0.014	0.012	0.010
Isabel Pressure	0.038	0.067	0.073	0.078
Isabel Velocity	0.091	0.155	0.169	0.179
Combustion Mixfrac	0.209	0.329	0.351	0.365

5.5 Impact of Different Number of Dropout Layers on MCDropout Model Performance

We conduct experiments on the Isabel Pressure and Teardrop data sets to assess the detailed impact of varying the number of dropout layers on model performance and prediction accuracy. We present results of averaged volume reconstruction quality results using models with dropout added at the (1) last two residual blocks, (2) last half of the blocks, and (3) all the residual blocks. The results, shown in Table 6, show that with an increased number of dropout layers, the PSNR value drops slowly, as was reported by Kendall et al. [31] since an increased number of dropout layers can act as a strong regularizer. Hence, in our work, we use dropout at the last two residual blocks to produce high-quality volume reconstruction and robust uncertainty estimates.

5.6 Reconstruction with Different Dropout Probabilities

Table 7 presents a study assessing the quality of volume data reconstruction, measured by PSNR, averaged over 100 MC samples across different test time dropout probabilities. We observe that
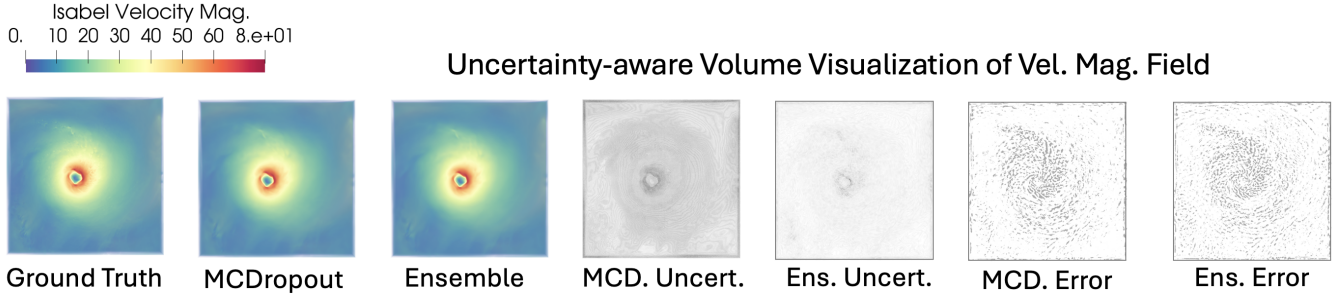
Figure 8: Volume visualization of the Velocity Magnitude field of Hurricane Isabel data set. We observe that both MCDropout and Ensemble methods produce visually identical volume rendering results when compared against the ground truth. The uncertainty map indicates moderate uncertainty in the region surrounding the Hurricane eye feature for the MCDropout method. In contrast, the Ensemble method produces an uncertainty map where fewer pixels show high uncertainty compared to the MCDropout method. The error maps show pixel errors when the predicted image is compared against ground truth. We observe similar error maps for both methods while also note that uncertainty and error maps are not correlated.
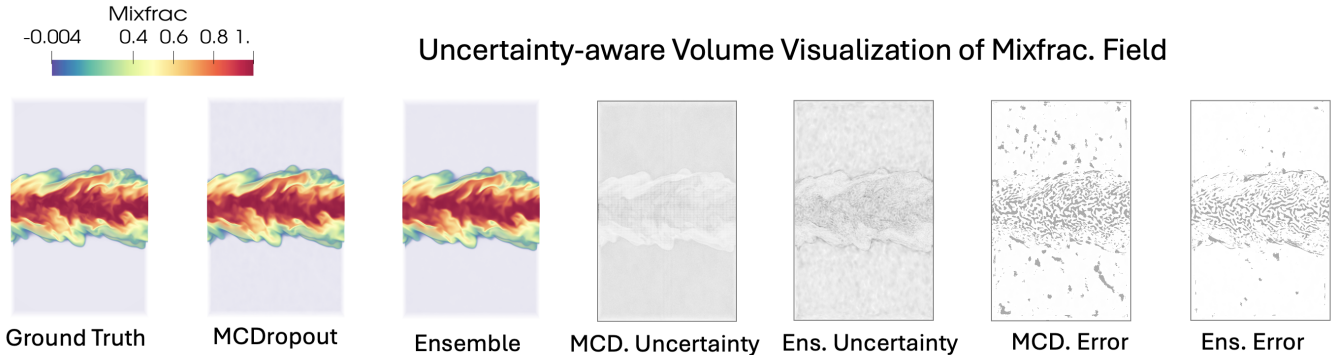

Figure 9: Uncertainty-informed volume visualization of the Mixfrac field of Turbulent Combustion data set. We observe that both MCDropout and Ensemble methods produce visually identical volume rendering results when compared against the ground truth. The uncertainty maps show that while the MCDropout method produces smoother and moderate to low uncertainty, the Ensemble method results in noisier and higher uncertainty inside and around the burning flame region. The two error maps indicate that the MCDropout method incurs more error in pixel values as compared to the Ensemble method.

Table 6: We report how the reconstruction quality (PSNR (↑)) changes when different numbers of Dropout layers are used for the MCDropout method. Three configurations are studied: Dropout used (1) at the last two residual blocks; (2) at the last half of the residual blocks; (3) at all the residual blocks.

Data Set	Last two Res. Block	Last half Res. Block	All Res. Block
Teardrop	71.694	68.048	62.158
Isabel Pressure	56.922	56.784	56.189

Table 7: PSNR (dB) (↑) of 100 MC samples with different test time dropout probabilities during inference.

Data Set	Different Test Time Dropout Probability					
	0.05	0.1	0.2	0.3	0.4	0.5
Teardrop	71.878	71.694	71.282	70.567	69.649	68.437
Isabel Pressure	56.789	56.922	56.084	55.295	54.23	52.728
Isabel Velocity	44.268	44.178	43.924	43.45	42.788	41.843
Combustion Mixfrac	48.658	48.474	47.855	46.848	45.45	43.674

PSNR mostly remains stable up to the dropout probability of 0.2, and as the dropout probability is further increased, expectedly, the PSNR value gradually drops. In their work, Gal et al. [14] suggest that using a small test time dropout probability is sufficient to estimate the model uncertainty robustly. Therefore, in our experiments, we use a dropout probability of 0.1 consistently for all the data sets to generate robust and meaningful uncertainty estimates.

5.7 Impact of Different Number of MC Samples on Average Pixel-wise Image space Uncertainty

Table 8 presents the impact of varying the number of MC samples on the average pixel-wise uncertainty value computed in the image space using the volume rendered pixel color values. This value indicates the normalized estimated uncertainty (standard deviation) calculated from the RGB color channels using the MC-Dropout method. It is seen that the uncertainty value gradually saturates as the number of MC samples reach 100.

5.8 Reconstruction Quality and Error in Image Space

Table 9 provides the PSNR and RMSE values computed using the averaged volume-rendered images for both the methods. The PSNR value reflects the reconstruction quality, while the RMSE indicates the error when the ground truth image pixel values are compared against the averaged image for both methods. The detailed computation of these averaged volume-rendered images has been discussed in Section 4.1. A close inspection of Table 9 reveals that the Ensemble method generally produces higher quality volume-rendered images with lower RMSE values.

5.9 Comparison of Training and Inference Time Between MCDropout and Ensemble Methods

Table 10 presents the training and inference times for both methods. For inference, we report both the volume reconstruction time

Table 8: Image space uncertainty values with different number of MC samples. We observe that by increasing the number of MC samples up to 100, the uncertainty values tend to become saturated.

Data Set	#MCSamp =10	#MCSamp =25	#MCSamp =50	#MCSamp =75	#MCSamp =100
Teardrop	0.213	0.231	0.239	0.239	0.24
Isabel Pressure	0.199	0.213	0.217	0.219	0.22
Isabel Velocity	0.245	0.261	0.266	0.268	0.269
Combustion Mixfrac	0.637	0.671	0.681	0.685	0.686

Table 9: Image space PSNR(dB) (↑) and RMSE (↓) values for MC-Dropout (100 samples) and Ensemble method (10 members). We observe that the Ensemble method generally produce higher PSNR and lower RMSE values.

Data set	Method	PSNR (dB)	RMSE
Teardrop	MC-Dropout (100 MC samples)	48.68	0.882
	Ensemble (10 members)	49.238	0.822
Isabel Pressure	MC-Dropout (100 MC samples)	53.344	0.384
	Ensemble (10 members)	56.050	0.525
Isabel Velocity	MC-Dropout (100 MC samples)	48.122	0.824
	Ensemble (10 members)	49.072	0.739
Combustion Mixfrac	MC-Dropout (100 MC samples)	48.394	0.959
	Ensemble (10 members)	51.431	0.676

and the ray casting time for each method. The Ensemble approach has faster inference times since it uses predictions from only 10 members, whereas MCDropout generates 100 MC samples, resulting in longer inference times. Training a dropout-enabled MC-Dropout model and a single ensemble member takes a comparable amount of time. However, since we train 10 ensemble members to build a robust Ensemble model, the total training time is 10 times greater, making the Ensemble method computationally more expensive than the MCDropout method. Additionally, storing model checkpoints for the Ensemble method requires 10 times more storage than for the MCDropout method. This significant difference in training time often makes MCDropout preferable for achieving timely results with robust uncertainty estimates.

5.10 Model vs. Raw Data Size

As Implicit Neural Representations (INRs) like ours have proven effective for compactly representing large-scale volumetric data [36], Table 11 compares the sizes of the models and the raw data. As expected, MCDropout requires approximately ten times less storage than the Ensemble method, which consists of ten individual models, whereas MCDropout is a single model-based approach. Thus, when both data compression and uncertainty estimation are required, the MCDropout method offers a significantly higher compression ratio.

Table 10: Training and inference timings for MCDropout and Ensemble Methods. During reconstruction, we use 100 MC samples for the MCDropout method and 10 ensemble member outputs for the Ensemble method.

Data Set	Training Time (Hours)		Inference time (Seconds)			
			Reconstruction time		Rendering time	
	MCD	ENS	MCD	ENS	MCD	ENS
Teardrop	0.42	0.43	28.95	2.58	49.902	0.838
Isabel Pressure	5.24	5.36	328.14	31.86	24.586	0.779
Isabel Velocity	5.18	5.18	330.91	31.64	38.554	0.945
Combustion Mixfrac	8.71	8.54	545.56	31.63	40.403	1.183

Table 11: Comparison of model size and raw data size.

Data Set	Raw Data Size (KB)	MCDropout Model Size (KB)	Ensemble Model Size (10 members) (KB)
Teardrop	2048	216	2160
Isabel Pressure	12208	220	2200
Isabel Velocity	12208	220	2200
Combustion Mixfrac	20252	220	2200

6 DISCUSSION

Uncertainty-agnostic vs. Uncertainty-aware DNNs. In this work, we advocate the use of deep neural networks (DNNs) that can quantify their prediction uncertainty when employed to perform scientific volumetric data visualization tasks. As demonstrated here, these uncertainty-aware neural networks provide important insights into the reliability of their predictions. By effectively communicating this uncertainty to experts, they can make more informed decisions regarding the data features from the visualization results. Integrating uncertainty is also essential for fostering trust in DNN-predicted results used in scientific visualization research.

MCDropout vs. Ensemble Method. Our research explores two approaches for estimating uncertainty using Implicit Neural Representations (INRs) as the base DNN architecture to visualize volume data sets. Despite its long training times, we employ Deep Ensemble, a well-recognized standard, and investigate the single-model-based MCDropout method to address computational challenges. MCDropout is selected for its theoretical elegance and ease of integration into existing DNN models with minimal modifications. Our results indicate that both methods can offer meaningful prediction uncertainty, while the Ensemble method generally produces more accurate and robust predictions with higher confidence compared to the MCDropout method. It is also observed that the uncertainty maps generated by these two methods could be similar (for the Teardrop data set) as well as different and vary across different data sets. However, considering the extremely high training time required to build a robust ensemble model, MCDropout can be a reliable alternative to the Ensemble method when computational resources are constrained, and prompt uncertainty estimation is essential.

Error vs. Uncertainty. Implicit neural models like ours have been proven to be highly effective in generating compressed representations of large-scale volume data sets [36]. It has been demonstrated that efficient volume visualization can be achieved using only the compressed INR model-reconstructed data, eliminating the need to access the raw volume data. In these applications, quantifying error becomes challenging when the entire visual analysis relies on model-generated data, as the raw data may not be available. Our uncertainty-aware INRs can enhance the trustworthiness of these model-generated results by offering prediction uncertainty and model confidence information to domain experts.

7 CONCLUSIONS AND FUTURE WORK

This paper underscores the significance of understanding uncertainty by applying two deep uncertainty estimation methods. We investigate how prediction uncertainty can benefit DNN-based volume data modeling and visualization tasks. Looking ahead, our research aims to extend to time-varying and multivariate volumetric data sets. Additionally, we plan to explore alternative single-model-based deep uncertainty estimation techniques that are computationally efficient for interactive uncertainty-aware volume visualization. The insights gained from uncertainty estimates can pinpoint areas that require targeted training and highlight model limitations in specific data regions. In critical scenarios, the confidence a model has in its predictions is vital, as acknowledging uncertainties fosters greater trust in the model.

REFERENCES

[1] A. Amini, W. Schwarting, A. Soleimany, and D. Rus. Deep evidential regression. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20. Curran Associates Inc., Red Hook, NY, USA, 2020. 2

[2] T. M. Athawale, C. R. Johnson, S. Sane, and D. Pugmire. Fiber uncertainty visualization for bivariate data with parametric and nonparametric noise models. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):613–623, 2022. 2

[3] T. M. Athawale, B. Ma, E. Sakhaee, C. R. Johnson, and A. Entezari. Direct volume rendering with nonparametric models of uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1797–1807, 2020. doi: doi: 10.1109/TVCG.2020.3030394 2

[4] M. S. Ayhan and P. Berens. Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. In *Medical Imaging with Deep Learning Conference*, 2018. 2

[5] W. H. Beluch, T. Genewein, A. Nurnberger, and J. M. Kohler. The power of ensembles for active learning in image classification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9368–9377, 2018. doi: 10.1109/CVPR.2018.00976 1, 3

[6] M. Berger, J. Li, and J. A. Levine. A generative model for volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 25(4):1636–1650, 2019. doi: 10.1109/TVCG.2018.2816059 1, 2

[7] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. Weight uncertainty in neural networks. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, p. 1613–1622. JMLR.org, 2015. doi: 10.5555/3045118.3045290 2

[8] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. *Overview and State-of-the-Art of Uncertainty Visualization*, pp. 3–27. Springer London, London, 2014. doi: 10.1007/978-1-4471-6497-5_1 1, 2

[9] D. Choi, C. J. Shallue, Z. Nado, J. Lee, C. J. Maddison, and G. E. Dahl. On empirical comparisons of optimizers for deep learning. *arXiv preprint arXiv:1910.05446*, 2019. 2

[10] A. Damianou and N. D. Lawrence. Deep Gaussian processes. In *Proc. of AISTATS*, pp. 207–215, 2013. 1

[11] S. Devkota and S. Pattanaik. Deep learning based super-resolution for medical volume visualization with direct volume rendering. In G. Bebis, B. Li, A. Yao, Y. Liu, Y. Duan, M. Lau, R. Khadka, A. Crisan, and R. Chang, eds., *Advances in Visual Computing*, pp. 103–114. Springer International Publishing, Cham, 2022. doi: 10.1007/978-3-031-20713-6_8 1

[12] A. Dienstfrey and R. Boisvert. *Uncertainty Quantification in Scientific Computing: 10th IFIP WG 2.5 Working Conference, WoCoUQ 2011, Boulder, CO, USA, August 1-4, 2011, Revised Selected Papers*, vol. 377. Springer, 2012. 2

[13] J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong. *Expanding the frontiers of visual analytics and visualization*. Springer, 2012. 2

[14] Y. Gal and Z. Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *Proc. of ICML*, pp. 1050–1059, 2016. 1, 2, 3, 8

[15] C. Gillmann, P. Arbelaez, J. T. Hernandez, H. Hagen, and T. Wischgoll. An uncertainty-aware visual system for image pre-processing. *Journal of Imaging*, 4(9):109, 2018. 2

[16] C. Gillmann, D. Saur, T. Wischgoll, and G. Scheuermann. Uncertainty-aware visualization in medical imaging-a survey. In *Computer Graphics Forum*, vol. 40, pp. 665–689. Wiley Online Library, 2021. 2

[17] E. Goan and C. Fookes. *Bayesian Neural Networks: An Introduction and Survey*, pp. 45–87. Springer International Publishing, Cham, 2020. doi: 10.1007/978-3-030-42553-1_3 2

[18] F. K. Gustafsson, M. Danelljan, and T. B. Schon. Evaluating scalable Bayesian deep learning methods for robust computer vision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 318–319, 2020. 1, 3

[19] J. Han and C. Wang. TSR-TVD: Temporal super-resolution for time-varying data analysis and visualization. *IEEE transactions on visualization and computer graphics*, 26(1):205–215, 2019. 2

[20] J. Han and C. Wang. SSR-TVD: Spatial super-resolution for time-varying data analysis and visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2445–2456, 2020. 2

[21] J. Han and C. Wang. CoordNet: Data generation and visualization generation for time-varying volumes via a coordinate-based neural network. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):4951–4963, 2023. doi: 10.1109/TVCG.2022.3197203 1, 3

[22] J. Han, H. Zheng, D. Z. Chen, and C. Wang. Stnet: An end-to-end generative framework for synthesizing spatiotemporal super-resolution volumes. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):270–280, 2021. 2

[23] J. Han, H. Zheng, Y. Xing, D. Z. Chen, and C. Wang. V2V: A deep learning approach to variable-to-variable selection and translation for multivariate time-varying data. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1290–1300, 2021. 2

[24] A. Hasanzadeh, E. Hajiramezanali, K. Narayanan, N. Duffield, M. Zhou, and X. Qian. Semi-implicit graph variational auto-encoders. In *Proc. of NIPS*, pp. 10712–10723, 2019. 2

[25] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. 3

[26] W. He, J. Wang, H. Guo, K.-C. Wang, H.-W. Shen, M. Raj, Y. S. G. Nashed, and T. Peterka. Insitunet: Deep image synthesis for parameter space exploration of ensemble simulations. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):23–33, 2020. 2

[27] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012. 2, 3

[28] F. Hong, C. Liu, and X. Yuan. DNN-VolVis: Interactive volume visualization supported by deep neural network. In *2019 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 282–291, 2019. 1, 2

[29] Z. Hu, J. Zhang, H. Wang, S. Liu, and S. Liang. Leveraging relational graph neural network for transductive model ensemble. In *Proc. of KDD*, p. 775–787, 2023. 1

[30] A. Kamal, P. Dhakal, A. Javaid, V. Devabhaktuni, D. Kaur, J. Zaientz, and R. Marinier III. Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization*, 24:1–30, 05 2021. doi: 10.1007/s12650-021-00755-1 2

[31] A. Kendall, V. Badrinarayanan, and R. Cipolla. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. In *British Machine Vision Conference 2017, BMVC 2017, London, UK, September 4-7, 2017*. BMVA Press, 2017. 3, 7

[32] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[33] A. Knoll, Y. Hijazi, A. Kensler, M. Schott, C. Hansen, and H. Hagen. Fast ray tracing of arbitrary implicit surfaces with interval and affine arithmetic. *Computer Graphics Forum*, 28(1):26–40, Feb. 2009. doi: 10.1111/j.1467-8659.2008.01189.x 4, 5

[34] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in neural information processing systems*, pp. 6402–6413, 2017. 2, 3, 4

[35] M. Leutbecher and T. N. Palmer. Ensemble forecasting. *Journal of computational physics*, 227(7):3515–3539, 2008. 2

[36] Y. Lu, K. Jiang, J. A. Levine, and M. Berger. Compressive neural representations of volumetric scalar fields. *Computer Graphics Forum*, 40(3):135–146, 2021. doi: 10.1111/cgf.14295 1, 2, 3, 9

[37] Y. Ma, N. J. Foti, and E. B. Fox. Stochastic gradient MCMC methods for hidden M arkov models. In *Proc. of ICML*, pp. 2265–2274, 2017. 2

[38] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. V. Dillon, B. Lakshminarayanan, and J. Snoek. Can you trust your

model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, 2019. doi: 10.5555/3454287.3455541 3

[39] A. T. Pang, C. M. Wittenbrink, S. K. Lodha, et al. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1997. 2

[40] W. S. Parker. Ensemble modeling, uncertainty and robust predictions. *Wiley interdisciplinary reviews: Climate change*, 4(3):213–223, 2013. 2

[41] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. 4

[42] K. Pöthkow and H.-C. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Trans. on Vis. and Comp. Graphics*, 17:1393–1406, 2011. 2

[43] K. Pöthkow, B. Weber, and H.-C. Hege. Probabilistic marching cubes. In *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'11, pp. 931–940, 2011. 2

[44] K. Potter, J. Krüger, and C. Johnson. Towards the visualization of multi-dimensional stochastic distribution data. In *Proceedings of The International Conference on Computer Graphics and Visualization (IADIS) 2008*, 2008. 2

[45] R. Rahaman and A. H. Thiery. Uncertainty quantification and deep ensembles. In *Neural Information Processing Systems*, 2020. 3

[46] A. Renda, M. Barsacchi, A. Bechini, and F. Marcelloni. Comparing ensemble strategies for deep learning: An application to facial expression recognition. *Expert Systems with Applications*, 136:1–11, 2019. 3

[47] O. Sagi and L. Rokach. Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 8(4):e1249, 2018. 2

[48] S. Sahoo, Y. Lu, and M. Berger. Neural Flow Map Reconstruction. *Computer Graphics Forum*, 41(3):391–402, 2022. doi: 10.1111/cgf. 14549 3

[49] M. Sensoy, L. Kaplan, and M. Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems*, pp. 3179–3189, 2018. 2

[50] J. Shen, H. Li, J. Xu, A. Biswas, and H.-W. Shen. IDLat: An importance-driven latent generation method for scientific data. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):679–689, 2022. 2

[51] N. Shi, J. Xu, H. Li, H. Guo, J. Woodring, and H.-W. Shen. VDL-Surrogate: A view-dependent latent-based model for parameter space exploration of ensemble simulations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):820–830, 2023. doi: 10.1109/TVCG.2022.3209413 2

[52] N. Shi, J. Xu, S. W. Wurster, H. Guo, J. Woodring, L. P. Van Roekel, and H.-W. Shen. GNN-Surrogate: A hierarchical and adaptive graph neural network for parameter space exploration of unstructured-mesh ocean simulations. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2301–2313, 2022. 2, 3

[53] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. In *Proc. NeurIPS*, 2020. 3

[54] K. Tang and C. Wang. STSR-INR: Spatiotemporal super-resolution for multivariate time-varying volumetric data via implicit neural representation. *Computers & Graphics*, 119:103874, 2024. doi: 10.1016/j.cag.2024.01.001 1, 3

[55] J. Thiagarajan, R. Anirudh, V. S. Narayanaswamy, and T. Bremer. Single model uncertainty estimation via stochastic data centering. In *Advances in Neural Information Processing Systems*, vol. 35, pp. 8662–8674. Curran Associates, Inc., 2022. 2

[56] A. Vyas, N. Jammalamadaka, X. Zhu, D. Das, B. Kaul, and T. L. Willke. Out-of-distribution detection using an ensemble of self supervised leave-out classifiers. In *Proceedings of the European Conference on Computer Vision*, pp. 550–564, 2018. 3

[57] C. Wang and J. Han. DL4SciVis: A state-of-the-art survey on deep learning for scientific visualization. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2022. 1, 2

[58] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, and T. Vercauteren. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing*, 338:34–45, 2019. 2

[59] S. Weiss, M. Chu, N. Thuerey, and R. Westermann. Volumetric isosurface rendering with deep learning-based super-resolution. *IEEE Transactions on Visualization and Computer Graphics*, 27(6):3064–3078, 2019. 2

[60] S. Weiss, P. Hermüller, and R. Westermann. Fast neural representations for direct volume rendering. *Computer Graphics Forum*, 41(6):196–211, 2022. 1, 2

[61] S. Weiss, M. Işlk, J. Thies, and R. Westermann. Learning adaptive sampling and reconstruction for volume visualization. *IEEE Transactions on Visualization and Computer Graphics*, 28(7):2654–2667, 2022. 1, 2

[62] S. Weiss and R. Westermann. Differentiable direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):562–572, 2022. doi: 10.1109/TVCG.2021.3114769 1

[63] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour Boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2713–2722, 2013. 2

[64] S. W. Wurster, H. Guo, H.-W. Shen, T. Peterka, and J. Xu. Deep hierarchical super resolution for scientific data. *IEEE Transactions on Visualization and Computer Graphics*, 29(12):5483–5495, 2022. 2

[65] Y. You, J. Hseu, C. Ying, J. Demmel, K. Keutzer, and C.-J. Hsieh. Large-batch training for LSTM and beyond. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–16, 2019. 2

[66] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021. 2